Downloaded from orbit.dtu.dk on: May 11, 2024

DTU Library

Synthetic biology tools to control growth and production in Escherichia coli

Landberg, Jenny Marie

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Landberg, J. M. (2020). *Synthetic biology tools to control growth and production in Escherichia coli*.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

DTU Biosustain
The Novo Nordisk Foundation Center for Biosustainability

Synthetic biology tools to control growth and production in *Escherichia coli*

Jenny Landberg

PhD thesis

Synthetic biology tools to control growth and production in *Escherichia coli*

PhD Thesis

Jenny Landberg

The Novo Nordisk Foundation Center for Biosustainability
Technical University of Denmark

August 2020

Synthetic biology tools to control growth and production in *Escherichia coli*
PhD thesis written by Jenny Landberg
Supervised by Professor Alex Toftgaard Nielsen

© PhD Thesis 2020 Jenny Landberg
Novo Nordisk Center for Biosustainability
Technical University of Denmark
All rights reserved

PREFACE

This thesis is written as partial fulfillment of the requirements to obtain a PhD degree at the Technical University of Denmark. The work presented in this thesis was carried out between September 2017 and August 2020 at the Novo Nordisk Foundation Center for Biosustainability at the Technical University of Denmark, Kongens Lyngby, Denmark, and during a research stay from May 2019 to October 2019 in the laboratory of Jay Keasling at the Joint BioEnergy Institute at Lawrence Berkeley National Laboratory in Emeryville, California, USA. The work was supervised by Professor Alex Toftgaard Nielsen and funding was provided by the Novo Nordisk Foundation through the Copenhagen Bioscience PhD Programme, grant no. NNF16CC0020908. The external stay was partially funded by Kaj and Hermilla Ostenfelds Foundation, Reinholdt W. Jorck and Hustrus Foundation, Otto Mønsteds Foundation and Markussens Student Foundation.

Jenny Landberg
Kgs. Lyngby, August 2020

ABSTRACT

Climate change is a growing concern worldwide. With the amount of energy and resources we use today, we would need 1.7 Earths to cover our consumption in a sustainable way. The main contributing factor to the changing climate is the extraction and use of fossil resources for production of fuels and chemicals. This releases high amounts of carbon dioxide to the atmosphere, which drives global warming. To minimize the environmental impact, it is critical that we find new, sustainable options to oil-based manufacturing.

For centuries, humans have used microorganisms to make everything from beer and wine to bread, yogurt, cheese, and kimchi. In the recent decades, scientific breakthroughs within gene technology and DNA sequencing have enabled the use of microbes in novel applications. By rewiring microbial metabolism, we can establish cell factories that utilize renewable resources to sustainably produce many of the chemicals, fuels and materials that are currently made from fossil resources. However, most of these bioproducts are more expensive to produce than oil-derived products. In order to make cell factories economically feasible, it is important to find new ways to optimize the yield, titer and productivity of bioproduction processes.

This thesis presents new tools and methods that can be used to increase product yields and control growth and production in the microbial cell factory *Escherichia coli*. A screening of the *E. coli* genome was carried out to identify genes and intergenic regions that, when inhibited, decouples growth and production. Identified targets that were shown to stop cell growth and increase production capacity were further used to improve production of a single-domain antibody. Using proteomics, it was found that the growth decoupling strains were metabolically active and did not exhibit a typical stationary phase response. Finally, an autoinducible gene expression system based on the tryptophan operon and the T7 RNA polymerase was developed and applied for production of a protein and a biochemical. The system relies on tryptophan depletion and does not require addition of expensive inducers.

DANSK SAMMENFATNING

Klimaændringer er en voksende bekymring verden over. Vi bruger i dag så megen energi og resurser, at vi har behov for 1.7 jordkloder for at dække vores forbrug på en bæredygtig måde. Den primære bidragende faktor til klimaændringerne er vores anvendelse af fossile resurser til produktion af brændstoffer og kemikalier. Dette frigiver store mængder kuldioxid til atmosfæren, hvilket er med til at skabe den globale opvarmning. For at minimere den fremtidige miljøpåvirkning er det kritisk, at vi finde nye, bæredygtige alternativer til de oliebaserede produktionsmetoder.

Mikroorganismer har i mange århundreder hjulpet mennesket at fremstille alt fra øl og vin til brød, yoghurt, ost og kimchi. I de seneste årtier har videnskabelige gennembrud inden for genteknologi og DNA-sekvensering gjort det muligt at bruge mikrober til nye formål. Ved at modificere deres metabolisme, kan vi skabe cellefabrikker, der bruger fornybare ressourcer til bæredygtig produktion af de kemikalier, brændstoffer og materialer, der i dag produceres fra fossile ressourcer. De fleste af disse bioprodukter er dog dyrere end tilsvarande produkter fra olie. For at etablere økonomisk konkurrencedygtige cellefabrikker, er det vigtigt at udvikle nye metoder, der kan bruges til at forbedre og optimere bioproduktionsprocesser.

Denne afhandling præsenterer nye værktøjer og metoder, der kan bruges til at øge produktudbyttet og kontrollere vækst og produktion i den mikrobielle cellefabrik *Escherichia coli*. En screening af *E. coli*-genomet blev udført for at identificere gener og intergeniske regioner, der afkobler vækst og produktion når de inhiberes. Identificerede gener, der stoppede cellevækst og øgede produktionskapaciteten, blev derefter brugt til at forbedre produktionen af et antistof. En proteomics analyse viste, at celler med syntetisk inhiberet vækst, fortsat var metaboliskt aktive, og ikke gik i stationær fase, når tilvæksten blev inhiberedet. Til sidst udvikledes et autoinducerbart genekspressionssystem baseret på tryptofan-operonet og T7 RNA polymerase. Systemet kan bruges til produktion af proteiner of biokemikalier, og er ikke afhængig af dyre inducermolekyler.

ACKNOWLEDGEMENTS

Doing a PhD has been a wonderful journey full of ups and downs. I have a lot of people to thank for the support, guidance, encouragement and friendship they have provided during these years.

First of all, I would like to thank my supervisor Alex Toftgaard Nielsen. You have always been positive, supportive and full of great ideas. Thank you for all the enjoyable scientific and non-scientific discussions, for giving me freedom to carry out my own ideas and for giving me great advice whenever I needed it. I would like to thank all the old and new members of the Nielsen BCFO lab. Torbjørn, thank you for your endless support, advice and lots of laughs and good times. Elleke, thank you for all the encouragement and for inspiring me with your scientific dedication. Hemanshu and Christian, thank you for great collaborations, and for introducing me to the lab and to your research projects. A big thank you to Amalie, Arrate, Ivan, Philip, Regiane, Rosa, Sheila, Stephanie, Viviënne and Yixin for making BCFO such a great group to be in. I would also like to thank the wonderful students I have supervised during my time as a PhD student; Amanda, Cleménce and Marie – it was such a pleasure to work with all of you and I wish you the best of luck in the future!

During my time at Cfb, I have met many people who I am thankful to. Helén, Silvia and Sophia, thank you for wonderful friendship and for lots of fun times together. I look forward to future trips to Italy with you all. Maja, you are the best person imaginable to go through a start-up adventure with. Thank you for being a great friend and for sharing all the Mycropt ups and downs with me. Thank you to all the other people who I have had the pleasure of getting to know; a special thanks to Anja, Carolyn, Cristina, Ida, Kristoffer, Maja, Morten, Pernille, Sophia and the rest of the Nørholm group for being the best lunch company; Elsayed, Lasse, Markus, Naia, Niko, Sheila, Songyuan and Tune for fruitful scientific collaboration; all the members of the PhD club for organizing fun and inspiring events and activities; and Davíd, Daria, Ekaterina, Gonzalo and the rest of the students

at the Novo Nordisk Foundation Copenhagen Bioscience PhD Programme for sharing the PhD journey with me. A big thanks to Amelia and Moreno for coordinating the program in an excellent way. I would also like to thank Rebeca, Susanne and all the other administrative, analytic, scientific and supporting staff, which Cfb would not function without.

I am very grateful to Jay Keasling for welcoming me to his lab, and for his valuable scientific input and advice during my external research stay at the Joint BioEnergy Institute. Amin, thank you for being such an inspiring and enthusiastic supervisor and friend, and for helping me regain my passion for science. Amanda, thank you for being a wonderful student, and for inspiring me with your curiosity and excitement. Elias, thank you for being the best friend I could have asked for during my time at JBEI. Thanks to all the other wonderful people who made my stay in Berkeley so amazing, especially Alli, Andrew, Cissi, Disa, Kavitha, Jing, Paul, Matthias, Yuzhong and the little Zargar group; Amanda, Jessica, Ravi and Sami.

I would like to send a special thanks to Suzanne, James, Kurt, Farbod, Virginia and Ivan for a fantastic research stay at University of California, Irvine during the last year of my Master's. You inspired me to start my PhD journey.

I am grateful for my friends and family in Sweden for always being there to support me. Ett stort tack till mamma för att du alltid ställer upp och tror på mig. Tack mormor och morfar för att ni alltid finns där för mig. Tack till pappa, Johan, Julia och hela släkten i Växjö. Last but not least, thank you Sebastian for being my neverending source of happiness, encouragement and inspiration. I could not have done this without you.

PUBLICATIONS

I. Li S[§], Jendresen CB[§], Landberg J[§], Pedersen LE, Sonnenschein N, Jensen SI, Nielsen AT. Genome-wide CRISPRi-based identification of targets for decoupling growth from production. ACS Synthetic Biology (2020) 9(5); 1030-1040. doi: 10.1021/acssynbio.9b00143

II. Landberg J, Risager Wrigh, N, Wulff T, Herrgård M, Nielsen AT. CRISPR interference of nucleotide biosynthesis improves production of a single-domain antibody in *Escherichia coli*. Biotechnology and Bioengineering (2020). doi: 10.1002/bit.27536

III. Landberg J, Mundhada H, Nielsen AT. An autoinducible trp-T7 expression system for production of proteins and biochemicals in *Escherichia coli*. Biotechnology and Bioengineering (2020) 117(5); 1513-1524. doi: 10.1002/bit.27297

[§] These authors contributed equally.

Publications not included in this thesis

I. Landberg J§, Zargar A§, Hernandez A, Wehrs M, Lal R, Chang S, Wang J, Baker J, Nielsen AT, Mukhopadyay A, Keasling J. An autoinducible gene expression system for *Saccharomyces cerevisiae*. *Manuscript in preparation.*

II. Mohamed ET, Mundhada H, Landberg J, Cann I, Mackie RI, Nielsen AT, Herrgård MJ, Feist AM. Generation of an *E. coli* platform strain for improved sucrose utilization using adaptive laboratory evolution. Microbial Cell Factories (2019) 18(116). doi: 10.1186/s12934-019-1165-2

III. Yu J, Landberg J, Shavarebi F, Bilanchone V, Okerlund A, Wanninayake U, Zhao L, Kraus G, Sandmeyer S. Bioengineering triacetic acid lactone production in *Yarrowia lipolytica* for pogostone synthesis. Biotechnolology and Bioengineering (2018) 115(9); 2383-2388. doi: 10.1002/bit.26733

IV. Patterson K§, Yu J§, Landberg J, Chang I, Shavarebi F, Bilanchone V, Sandmeyer S. Functional genomics for the oleaginous yeast *Yarrowia lipolytica*. Metabolic Engineering (2018) 7(48); 184-196. doi: 10.1016/j.ymben.2018.05.008

§ These authors contributed equally.

Patents not included in this thesis

I. Li S., Jendresen CB., Pedersen LE., Landberg J., Falkenberg KB., Mundhada H., Nielsen AT. (2018) IPC No. C12N1/21; C12N15/63; C12N9/88; C12P13/22; C12P21/02; C12P7/42. Methods for decoupling cell growth from production of biochemicals and recombinant polypeptides. (Patent No. WO2018020012).

CONTENT

Preface	IV
Abstract	V
Dansk sammenfatning	VI
Acknowledgements	VII
Publications	IX
Content	1
PhD thesis outline	2
Chapter 1 | Introduction	4
Chapter 2 | Genome-wide CRISPRi-based identification of targets for decoupling growth and production	62
Chapter 3 | CRISPR interference of nucleotide biosynthesis improves production of a single-domain antibody in *Escherichia coli*	99
Chapter 4 | An autoinducible trp-T7 expression system for production of proteins and biochemicals in *Escherichia coli*	141
Chapter 5 | Conclusions and outlook	172

PHD THESIS OUTLINE

Already in 1892, Svante Arrhenius predicted that the extraction of fossil resources would lead to global warming. Today, we know that the exploitation of our planet's resources has had an enormous impact on the environment, with effects that will last through the century and beyond. Global warming, rise of sea level, ocean acidification, extreme weather, shrinking ice sheets and mass extinction of plants and animals are just some of the issues caused by our extensive use of fossil fuels. Besides the considerable climate impact that fossil fuel usage has had, shrinking oil reserves makes it critical to find new and sustainable solutions for the production of fuels, chemicals and materials.

Bio-based manufacturing of chemicals and fuels using microbial cell factories could constitute a key step to reduce our dependence of petroleum. Although a number of bioprocesses have been commercialized, development is hampered due to limitations in existing conversion technology, lack of political incentives, and high costs compared to oil-based production. Addressing these limitations and improving the economic feasibility of cell factories will allow for eco-efficient, bio-based production methods to replace oil refineries. It will also enable the production of new types of compounds with novel and improved properties.

Traditional cell factory engineering efforts have usually been limited to pathway balancing and static rewiring of metabolic flux. The emergence of synthetic biology and novel gene regulation methods has led to the development of new approaches that can be used to further increase yields, titers and production rates. This includes dynamic regulation of growth and production, where growth and competing pathway fluxes are downregulated, and product pathways are upregulated, at a certain time-point of the production process.

The aim with this PhD thesis was to develop synthetic biology tools that can be used to improve product yields by regulating growth and production in

the cell factory *Escherichia coli*. Controlling growth and production enables efficient use of the provided substrates and nutrients, and allows for the fermentation process to be divided into two phases, where biomass is accumulated in the first phase and product is formed in the second phase. This can significantly improve cell factory performance, as well as the overall economic feasibility of bio-based production.

Chapter 1 gives a general introduction to industrial biotechnology and microbial cell factories. The chapter describes how metabolic engineering and synthetic biology can be used to improve cell factories, and gives an overview of the pathways and regulatory mechanisms that are relevant for the thesis.

Chapter 2 describes the genome-wide CRISPRi screening and identification of targets for decoupling growth from production. In this study, 12,000 sgRNAs were constructed and used to find targets that inhibited growth while protein production was maintained or increased.

Chapter 3 describes the application of the growth decoupling targets identified in chapter 2. CRISPRi was used to inhibit genes in the nucleotide biosynthesis pathway, which significantly improved the yield and titer of a single domain antibody. The metabolic state of the growth decoupled cells was investigated using proteomics.

Chapter 4 describes the design and development of an autoinducible expression system based on the tryptophan promoter and the T7 RNA polymerase. The system was applied for the production of a toxic protein and the amino acid L-serine.

Chapter 5 gives a summary and perspective on the work conducted in this thesis, and an outlook on the challenges and opportunities in the field.

Chapter 1

INTRODUCTION

For thousands of years, humankind has used microbial fermentation to conserve and produce food and beverages. One of the earliest findings date back to the Natufian people, who brewed beer-like alcoholic beverages in boulder mortars in Israel almost 13,000 years ago.[1] In the Neolithic village of Jiahu in Northern China, scientists have discovered pottery jars with 9,000 year old traces of fermented beverages containing rice, fruit and honey,[2] while the earliest grape wine is believed to have been made near what is now Tbilisi, Georgia, around 8,000 years ago.[3] Throughout history, the use of fermentation for preparation and preservation of food and drinks has given rise to a variety of products including bread, cheese, yoghurt and kombucha. However, it was not until the middle of the 19th century that Louis Pasteur demonstrated how fermentation works. Through a set of experiments, he showed that yeast grown under anaerobic conditions consumes sugars and nutrients and converts them into alcohol, carbon dioxide and other metabolites.[4] Soon after, he discovered the role of lactic acid bacteria in lactic acid fermentation after investigating contaminations creating sour-tasting alcohol at a sugar beet distillery.[5]

Pasteur's work has been followed by numerous scientific breakthroughs within recombinant DNA technology, DNA sequencing and synthesis, and genome engineering, which has moved the biotechnology field forward. Together, these findings have laid ground for our understanding of microbial metabolism and growth, and how we can harness and optimize inherent enzymatic reactions and pathways as well as complement them with novel metabolic functions in order to create designer cell factories where the metabolic flux is redirected toward a specific product. These cell factories can be used in industrial biotechnology to help us produce not only food and beverages, but a range of other compounds useful for our society. Hopefully, future research efforts will enable sustainable cell factories to replace oil refineries for the production of materials, chemicals and fuels.

1.1 Industrial biotechnology

Industrial biotechnology can be defined as "the application of biotechnology for the industrial processing and production of chemicals, materials and fuels. It includes the practice of using microorganisms or enzymes to generate industrially useful products in a more efficient way, or generate substances and chemical building blocks with specific capabilities that conventional petrochemical processes cannot provide".[6] Generally, bacteria, yeast or fungi are used to convert renewable feedstock into value-added products, such as biochemicals and proteins. Compared to traditional petroleum-based manufacturing methods, biotechnological production generates less waste and toxic by-products, emits lower amounts of greenhouse gases, and reduces the overall need of fossil resources for manufacturing of fuels, chemicals and materials.[7] Eco-efficient production methods combined with increased product diversity means industrial biotechnology can help companies reduce production costs and reach new markets. Through its sustainable and innovative solutions to complex environmental challenges, the sector is and will continue to be a key driver toward green growth and a circular bio-economy.[7,8]

The first large scale industrial biotechnology process was invented by Chaim Weismann in the early 20th century, almost 150 years after Pasteur's work on fermentation. Weismann discovered the acetone-butanol-ethanol fermentation process in *Clostridium acetobutylicum*, and used the organism for large scale production of acetone that was needed for the British smoke propellants during World War I.[9] In the middle of World War II, a decade after Alexander Fleming's discovery of this bacteria-killing compound secreted by mold, penicillin was the next industrial biotechnology product produced in large scale.[10] This was followed by other notable examples moving the field forward, including the development of enzymes to replace harmful phosphates in detergents in the 1960s, and the emergence of the bioethanol industry in the US and Brazil due to escalating oil prices in the 1970s. The first was an excellent example of how biotechnological products can be used to solve environmental issues in an efficient and sustainable way, by both reducing phosphate pollution of lakes and streams, and saving energy by maintaining wash efficiency at lower wash temperatures.[11]

Entering into the 21st century, the biotechnology sector has witnessed an immense growth. From 2000 to 2019, the annual revenue from the US industrial biotechnology sector increased from less than $20 billion to over $147 billion, with an annual growth rate exceeding 10% throughout the last decade.[12,13] Biochemicals are already 0.5% of the total US GDP, meaning that 17-25% of the overall revenue from fine chemicals produced in the US comes from industrial biotechnology.[14] In Europe, the sector was valued at €31.5 billion in 2013, and expected to reach €57.5-€99 by 2030 provided new political incentives focusing on sustainable and green solutions are implemented.[6] The most prevalent product groups currently on the market are industrial enzymes, biofuels, fine chemicals and platform intermediate chemicals.[7] Besides growth in the existing product segments, new compounds such as bulk chemicals, bioplastics, biochemicals and biomaterials with novel functionalities are expected to enter the market in the coming decade.[7,15] However, there are still hurdles to overcome to enable the transition from oil- to bio-based production[7,8] (Fig. 1.1).

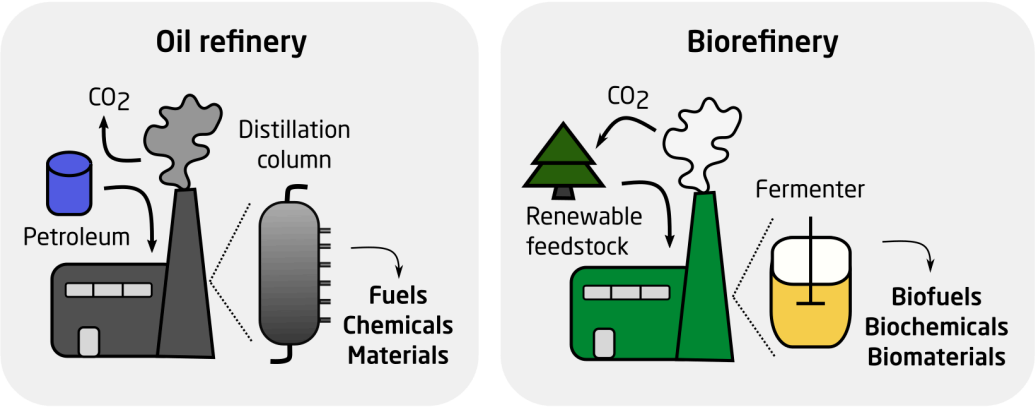


Figure 1.1. Transitioning from oil-based to bio-based manufacturing. The current use of petroleum for production of fuels, chemicals and materials generates toxic byproducts, releases CO_2 and has multiple detrimental effects on the environment. It is possible to replace a substantial part of oil-based products with sustainable bio-based products. These are manufactured from renewable feedstock that is converted into biochemicals, biofuels and biomaterials by microorganisms or enzymes.

Feedstock diversification. A large portion of the currently used renewable feedstock is derived from corn, sugar beets and sugar cane grown on arable land.[16] The use of alternative feedstock such as agricultural residues, lignocellulosic biomass, waste vegetable oils or gas should be put into practice in order to avoid competition with food supply and improve sustainability and economic feasibility of bioproducts.[16,17]

Novel organisms and chemical transformations. Domesticating and developing engineering tools for non-conventional microorganisms, and streamlining the identification of new enzymes and pathways, is required to enable utilization of all types and parts of the feedstock and facilitate manufacturing of new types of functionalized biologics.[18,19] This will further advance production of compounds currently on the market, promote feedstock diversification, and enable chemical and biological co-processing where platform biochemicals are functionalized through chemical manufacturing, and *vice versa*.[7]

Bioprocess feasibility. Fermentation‡ and processing constitute a major expense in bioproduction, where cost of substrate (feedstock) is often more than 60% of the total cost.[17,20] Developing new fermentation methods and engineering efficient cell factories capable of high substrate-to-product conversion rates and production yields will improve the overall economic feasibility of production and bring more bioproducts to the market.[18]

Policy making and funding. Political incentives are required to support research, innovation and commercialization activities, develop new infrastructure, standardize and harmonize bioproduct labelling and simplify regulatory procedures.[21,22] In order for the EU to reach its goal of reduced CO_2 emissions, substantial funding is needed to establish new biorefineries. Furthermore, the public needs to be educated about the advantages of industrial biotechnology in order to create a market-pull for bioproducts.[7,22]

‡Although fermentation in its strict scientific meaning refers to anaerobic extraction of energy from an organic molecule using an organic molecule as the final electron acceptor, the term will be used throughout this thesis to refer to any small or large scale microbial production process taking place in the presence or absence of oxygen.

The bioproduction process

The general industrial bioproduction process (Fig. 1.2) can be divided into two main stages; upstream processing, consisting of media preparation, strain inoculation and fermentation, and downstream processing, consisting of separation, purification and final processing. During media preparation, the culture media is mixed, hydrated and, if needed, sterilized. This is followed by inoculation and propagation of the selected microbial production strain in increasing culture volumes. When the desired cell density and culture volume has been reached, the cells are transferred to a large fermenter where the microbes convert the carbon and nutrients in the media to the desired product. Temperature, pH, aeration and agitation rate is usually controlled throughout the fermentation, which can be operated in batch, fed-batch or continuous mode. After fermentation, the cells are separated from the media and the product is extracted, purified and processed in several steps dependent on the product characteristics and purity requirements.

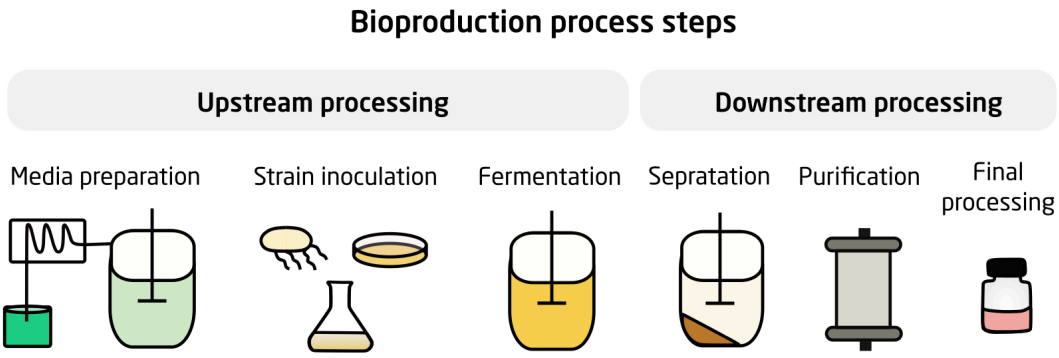


Figure 1.2. Processing steps of an industrial biotechnology production process. Upstream processing consists of media hydration, mixing and sterilization followed by strain propagation and fermentation. Downstream processing consists of separation of cell mass and media and recovery of the product from the cell mass or broth. This is followed by one or more purification and processing steps, which are dependent on the type of product and the required product purity.

1.2 Microbial cell factories

Microbial cell factories are the catalysts of industrial biotechnology, responsible of taking low-value feedstock and turning it into value-added bioproducts. The bioproduct can be expressed in an organism which naturally produce the compound, or in a heterologous host.[23] As new genetic tools are emerging, non-model organisms such as *Yarrowia lipolytica*,[24] *Kluyveromyces lactis*[19] and *Moorella thermoacetica*,[25] to name a few, are gaining more attention for their novel metabolic functions, ability to metabolize diverse substrates and tolerance to high temperatures.[26] However, model organisms *Escherichia coli* and *Saccharomyces cerevisiae* are still the two most commonly used cell factories due to their well understood metabolism and physiology, availability of metabolic models and ease of genetic engineering.

Escherichia coli. Named after physician and bacteriologist Theodor Escherich who discovered the bacteria in 1885, *E. coli* is the most well studied organisms in the field of microbiology.[27] It is a Gram-negative, facultative anaerobic chemoheterotroph that is naturally present in our gut microbiota.[28] The ease of culturing and propagating *E. coli* has made it the preferred host for a wide range of microbiological applications. The most common laboratory strains are MG1655 and W3110 (both derivatives of laboratory strain K-12), and DH5α, BL21, C, W and Crooks. Out of 6,626 unique proteins found in these strains, 50% are shared between all strains, 23% are shared between 2-6 strains and as many as 27% are unique to a single strain.[29] Therefore, it is not unexpected that every strain has its own specific characteristics in terms of growth rate, metabolic flux distribution, byproduct profile and gene expression and regulation. The most suitable strain for production of a certain compound can be selected by reviewing genomic sequences, predicting metabolic fluxes and evaluating gene expression values. For example, MG1655 has a high predicted flux through a range of heterologous pathways, making it a suitable host for the production of many types of biochemicals,[29] while BL21 is commonly used for recombinant protein production as it lacks two main proteases.[30] The widespread use of *E. coli* as a cell factory is easily understood as advantages include fast growth rate, substrate and metabolic versatility, and an extensive

genetic toolbox. However, the function of one third of the (on average) 4,400 genes is still not known,[31] and the organism has some disadvantages including toxin expression, overgrowth competing with production, need for expensive chemical inducers and genetic instability at high production levels.[32]

Saccharomyces cerevisiae. Just as *E. coli* in prokaryotic biology, *S cerevisiae* is one of the most well-studied organisms in eukaryotic cell biology. It has been used to study yeast mating, physiology and genetics, as well as eukaryotic signal transduction, regulation and the cell cycle. In 1996, *S. cerevisiae* strain S288C was the first eukaryotic organism to have its full genome sequenced.[33] The natural habitat of the yeast includes oak trees, cactuses and grape peels; strain EM93, from which S288c is derived, was isolated in 1938 from a rotting fig tree in California.[34] *S. cerevisiae* has traditionally has been used for various food applications such as winemaking, beer brewing and baking.[35,36] Today, it is widely used in industrial biotechnology for production of proteins, industrial enzymes and bio-ethanol.[37] Some of the most common laboratory strains beside S288C include CEN.PK, W303, Sigma1278b and FL100, which all have small variations in their genetic and metabolic properties.[38,39] Laboratory strains are usually haploid with mating type a or α, and auxotrophic for various amino acids to facilitate gene integration and plasmid transformation. Industrial strains are mostly diploid, aneuploid or even polyploid, which renders large genetic diversity and unique characteristics, but complicates strain engineering efforts.[37] Overall, *S. cerevisiae* is robust, stress tolerant, capable of expressing heterologous proteins and pathways, easy to engineer and resistant to phage contamination.[40] However, the yeast has a lower growth rate compared to many bacteria, and there is a need for better synthetic biology tools to advance the production of novel chemicals and address issues with co-factor imbalance.[41]

Production of proteins

Microbial cell factories are used for the production of proteins in different parts of the biotechnology industry. Products range from industrial enzymes such as proteases, lipases and amylases used in laundry detergents and food and beer processing, to antibodies, cytokines and other recombinant proteins produced in the pharmaceutical industry. The protein production host

organism is commonly selected based on the characteristics of the product. Prokaryotic hosts are often preferred due to their fast growth, short bioprocess times and low-cost media, and *E. coli* is the go-to host for expression of many of the new and established products.[42] However, *E. coli* limitations in secretion capacity and folding stability makes alternative production hosts with better traits, such as high secretion capability, an attractive alternative.[32] Many proteins are also produced in eukaryotic hosts, especially those requiring post-translation modifications. Depending on the nature of the modification, yeast, filamentous fungi, insect cells or mammalian cells can be used. For example, insulin is produced in yeast while monoclonal antibodies are expressed in mammalian cells.[42]

Single-domain antibodies. Immunoglobulin antibodies are a part of the adaptive immune system in all jawed vertebrates.[43] They have been extensively researched and are today used in diagnostics, treatment of disease, and in various research applications. In humans, the most common variant is immunoglobulin G (IgG).[44] It consists of light and heavy chains that are linked to each other in a tetrameric quaternary structure, where the antigen-binding fragment (Fab) is responsible for sensing and binding the epitope of the antigen (Fig. 1.3). To improve and facilitate research applications and pharmacokinetic properties of antibodies, scientists have used the Fab part, or made small fusion proteins consisting of the variable heavy and light chains connected with a linker peptide (single-chain variable fragment, scFv) (Fig. 1.3).

In 1989, scientists at the Free University of Brussels made a fascinating discovery; dromedaries have, in addition to the conventional immunoglobulins, a different type of antibody that is devoid of light chains (Fig. 1.3).[45] A few years later, scientists at the University of Miami discovered that sharks also express their own version of heavy chain antibodies (Fig. 1.3).[46] This led to the development of single-domain antibodies (sdAbs), which are engineered from the variable region of heavy-domain antibodies of camelids (VHH) and cartilaginous fishes (VNAR). The small size compared to Fabs and scFvs brings many advantages; they are easier to produce, better at penetrating tissue, more stable at high temperatures and low pH, and still maintain high antigen specificity.[47] sdAbs are being developed for diverse di-

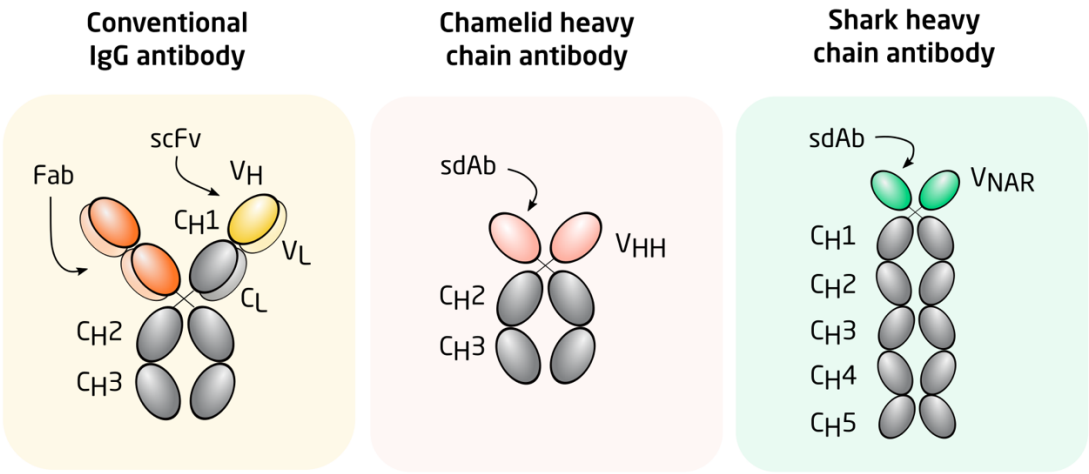


Figure 1.3. An IgG antibody, a *Camelidae* heavy chain antibody and a shark heavy chain antibody. IgG; immunoglobulin G, Fab; antigen-binding fragment, scFv; single-chain variable fragment, sdAb; single domain antibody, CH; heavy chain constant domain, VH; heavy chain variable domain, CL; light chain constant domain, VL; light chain variable domain, VHH; heavy chain variable domain of the camelid heavy chain antibody, VNAR; heavy chain variable domain of the shark heavy chain antibody.

agnostics and therapeutics applications, and the first sdAb-based drug for treatment of a rare blood clotting disorder was just launched on the market.[48]

Heavy chain antibodies are expressed not only in dromedaries and sharks, but also in llamas, camels, skates, rays and other animals belonging to the *Camelidae* family or the Chondrichthyes class of animals.[49,50] New sdAbs are commonly generated in llamas by injecting the animal with an antigen and harvesting the produced immunoglobulins. After ensuring antigen specificity, scale-up and production can take place. sdAbs typically lack glycosylation sites and can thus be expressed in bacteria, which means they are easier and cheaper to produce at large scale compared to conventional antibodies that are expressed in mammalian cells.[51] Therefore, they are of interest not only as diagnostics and therapeutics, but also for industrial biotechnology applications such as replacement of antibiotics in animal feed.[52,53] In Chapter 3 of this thesis, bacterial production of an sdAb was improved using growth decoupling.

Production of biochemicals

Biochemicals are more complex to produce than proteins, as they often require coordinated rewiring of metabolic flux through complex pathways. Compared to proteins, which are difficult to manufacture through chemical synthesis, many biochemicals are in tough competition with cheaper, chemically synthesized petrochemicals. However, there are still several compounds on the market, including biofuels, biopolymers, organic acids and amino acids.[54] Beside decreasing the carbon footprint compared to oil-derived chemicals, utilizing microbial metabolism for biochemical manufacturing paves the way for new products that are particularly difficult to synthesize chemically, and have novel properties and areas of application. Interestingly, all compounds that can be produced in nature are derived from only 12 precursor metabolites.[55] Therefore, establishing platform strains where metabolic flux has been optimized toward these different precursor compounds is a promising approach for speeding up cell factory construction.

Biochemicals can be produced by a single host or by a microbial consortia, where two or more organisms are used to convert the substrate into the final product.[56] Semi-synthetic manufacturing, where a platform chemical is synthesized in a host organism followed by further functionalization through chemical synthesis, is also a very promising approach.[57,58] The range of compounds that can be produced microbially is wide, and includes everything from simple primary metabolites to unnatural amino acids and more complex products such as terpenoids, flavonoids, alkaloids and polyketides.[59,60] With further research yielding new bioinformatics methods, metabolic engineering strategies and synthetic biology tools, the possibilities will continue to expand.

L-serine. The amino acid L-serine is currently used in the food, feed, cosmetics and pharmaceutical industry, but has great potential for extended use since it can be applied as a building block for numerous chemicals.[61] Today, it is produced by extraction from hydrolysates, through chemical synthesis, or by enzyme or whole cell catalysis from glycine. Production of L-serine in microbial cell factories will enable the expansion to novel application areas by reducing the production cost and environmental impact,

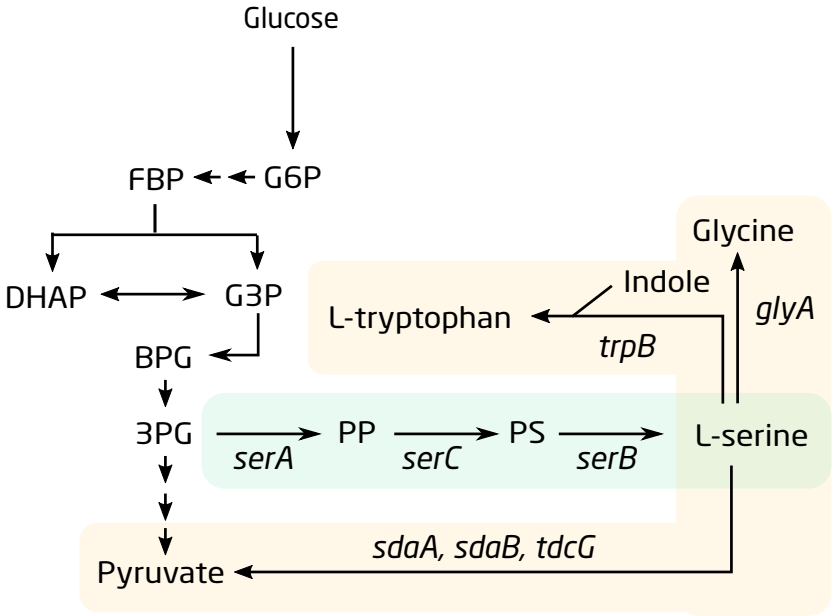


Figure 1.4. The L-serine synthesis, degradation and conversion pathways of *E. coli*. Glycolysis (no background), L-serine synthesis (green) and L-serine degradation and conversion pathways (yellow). G6P; glucose 6-phosphate, FBP; fructose 1,6-bisphosphate, DHAP; dihydroxyacetone phosphate, G3P; glyceraldehyde 3-phosphate, BPG; 1,3-bisphosphoglycerate, 3PG; 3-phosphoglycerate, PP; phosphohydroxy-pyruvate, PS; phosphoserine, TCA; tricarboxylic acid cycle. *serA* encodes for D-3-phosphoglycerate dehydrogenase; *serC* encodes for phosphoserine aminotransferase; *serB* encodes for phosphoserine phosphatase; *glyA* encodes for serine hydroxymethyltransferase; *trpB* encodes for tryptophan synthase (ß-chain); *sdaA*, *sdaB*, *tdcG* encodes for L-serine dehydratase.

and ensure feasible scale-up of production so that an increased market demand can be met.[62] As L-serine is a medium-value compound with low profit margin, it is however important to ensure that economically feasible production can be achieved by optimizing product titer and yield, and lowering process costs.

In *E. coli*, L-serine is synthesized from 3-phosphoglycerate (3PG) via a three-enzyme-pathway consisting of D-3-phosphoglycerate dehydrogenase encoded by *serA*, phosphoserine aminotransferase encoded by *serC* and phosphoserine phosphatase encoded by *serB* (Fig. 1.4). L-serine can be

degraded to pyruvate by L-serine dehydratase, or converted to glycine by serine hydroxymethyltransferase. Together with indole, L-serine is also used for the synthesis of tryptophan. The yield and titer of L-serine has been improved significantly through systematic metabolic engineering of *E. coli*. The degradation pathways have been deleted; L-serine feedback inhibition has been removed by random mutagenesis of *serA*; L-serine secretion has been improved by exporter overexpression; and adaptive laboratory evolution (ALE) has been used to improve L-serine tolerance.[63,64] In this highly optimized strain, a randomized ribosome binding site (RBS) library of the L-serine pathway was integrated into the genome, and screened to find high-producing variants.[65] This resulted in L-serine titers of 50 g/L with a yield of 0.36 g/g glucose in lab-scale fed batch fermentation, which is the highest reported titer for *E. coli* so far.[65] In Chapter 4 of this thesis, we engineered an autoinducible expression system to replace IPTG-based induction of the L-serine pathway in this strain, in order to reduce production costs. Beside *E. coli*, the main production organism to be metabolically engineered for L-serine production is *Corynebacterium glutamicum*, with a highest reported titer of 43.9 g/L and yield of 0.44 g/g sucrose.[66]

Regardless of the selected product, the cell factory must be efficient at producing large amounts of it. The efficiency is assessed by measuring the product titer, rate and yield (TRY).[54] TRY can be improved through different metabolic engineering and synthetic biology strategies that increase product expression, decrease byproduct formation, and dynamically balance microbial growth and production.[54]

Metabolic engineering of cell factories

Microorganisms have the innate ability to build hundreds of different compounds; however, they are generally not evolved to overproduce a certain protein or metabolite. In order to establish the advanced microbial cell factories used in bioproduction, the metabolic flux of the organism needs to be rewired.[67,68] There are different approaches to redirect metabolism and improve production. Biochemical productivity can be enhanced by over-expressing biochemical pathway enzymes, increasing precursor and co-factor supply, decreasing byproduct formation, deleting product-degrading enzymes, removing product feedback inhibition and co-localizing pathway

enzymes using synthetic scaffolds.[23,69] Compartmentalization of certain enzymatic reactions has been useful for engineering eukaryotic organisms.[70] For heterologous protein production, promoter selection, codon optimization and improvement of secretion pathways has proven sucessful.[23] Other metabolic engineering effort include increasing the substrate range and improving tolerance toward different metabolites.

Metabolic engineering has enabled the commercialization of a range of products besides industrial enzymes, amino acids and bioethanol. By introducing a twelve-step-pathway consisting of heterologous and native enzymes to *S. cerevisiae*, scientists at the University of California, Berkeley, managed to produce[71] and scale-up[72] production of the malarial medicine precursor artemisinic acid. Although the success of semi-synthetic artemisinin did not happen due to market saturation and market resistance,[73] American biotechnology company Amyris has further developed the terpenoid platform strain for commercialization of other products. Through a collaborative effort between Genencor and DuPont, large scale manufacturing of polymer building block 1,3-propanediol was commercialized in 2006, and in 2013, Genomatica and BASF launched their commercial scale production of solvent 1,4-butanediol.[74]

Developing commercial-scale cell factories is expensive and time-consuming. On average, it takes 6-8 years and costs >$50 mio.[54] Besides academic efforts at universities and research centers over the world, metabolic engineering and synthetic biology companies such as Ginkgo Bioworks and Zymergen are working on bringing down costs and speeding up the most common approach to cell factory engineering; the Design-Build-Test-Learn (DBTL) cycle (Fig. 1.5). DBTL is applied not only for cell factory construction, but also for tool development and other types of biological engineering.

Design. In this phase, the cell factory design for the chosen product is laid out. The host organism is chosen based on capability of feedstock utilization and inherent ability of expressing the product, tolerance to the product and the bioprocess conditions, strain physiology, and availability of genetic tools for strain engineering. The enzymatic pathway required for product expression can be identified in literature or in databases such as BRENDA[75]

and KEGG.[76] If a pathway is unknown, there are computational tools that can be used to find biosynthetic gene clusters or identify product pathways through a retrosynthetic approach.[77] Expression of the pathway genes needs to be balanced, and large efforts have been made toward standardizing biological parts for predictable design of synthetic constructs.[78] The Registry of Standard Biological Parts was founded at Massachusetts Institute of Technology in 2003 in combination with the start of the iGEM (Internationally Genetically Engineered Machine) competition and contains, among other items, characterized promoters, terminators, ribosome binding sites and plasmid backbones that can be used for cell factory design and diverse synthetic biology applications.[79] Last but not least, there is an increasing amount of useful computational metabolic models that can be applied to identify mutations that improve production.[80,81]

Build. In the build phase, the biological parts are synthesized and assembled according to the intended design. The decreasing cost of DNA synthesis and sequencing, and the continuous advances in genetic engineering technology is constantly speeding up this phase; today, it is even possible to synthesize whole synthetic genomes.[82] The build process includes cloning and introduction of heterologous enzymes or pathways, knock-out of native genes or pathways, and precision-based genome engineering with removal or insertion of nucleotides into the genome of the host strain. Further protein engineering may be required if the pathway enzymes are unstable at the intended bioprocess temperature or pH.[83] Even though standards are improving and the number of characterized parts are growing, the behavior of standardized biological parts is not always predictable when introduced to control novel enzymes and pathways. Therefore, it is oftentimes necessary to generate large libraries of constructs from where the best variant can be selected.[84] Today, it possible to assemble expression and strain libraries with millions of variants. Plate-based oligonucleotide synthesis is an affordable option where hundreds of thousands of oligos can be generated and cloned onto plasmids or introduced into the genome. This approach was used to generate a genome-wide sgRNA library for *E. coli* in the study in Chapter 2 of this thesis. Multiplex Automated Genomic Engineering (MAGE) can introduce genetic diversity at multiple sites in the chromosome, using degenerate nucleotides targeting different loci.[85] Clustered Regularly Inter-

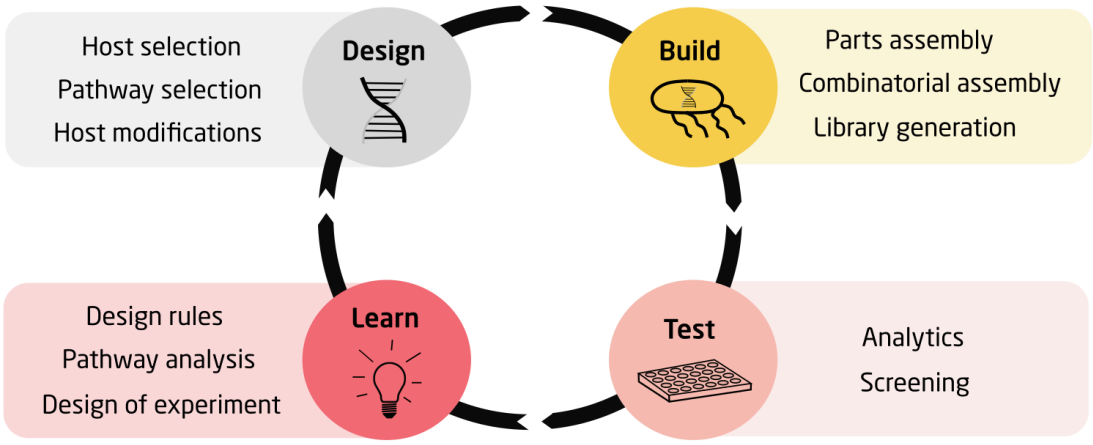


Figure 1.5. The Design-Build-Test-Learn (DBTL) cycle for biological engineering. Adapted from Petzold et al.[77]

spaced Short Palindromic Repeats (CRISPR)-based engineering[86] has revolutionized the field, with applications in the build phase ranging from basepair switching[87] and multigene editing[88] to knock-out[89] and genome integration of large DNA parts.[90] Besides these targeted methodologies, cell factories can be evolved to tolerate toxic compounds or grow (better) on new substrates using ALE, where selection pressure sorts out tolerant or fast-growing mutants.[91]

Test. In the test phase, the output from the constructed system is measured. Depending on the product, the test phase may pose a significant bottleneck in the DBTL cycle. Metabolite analysis using High Performance Liquid Chromatography (HPLC), Mass Spectrometry (MS) and Gas- or Liquid Chromatography-Mass Spectrometry (GC- or LC-MS) is highly accurate and enables monitoring of the target molecule, pathway intermediates, byproducts and substrates. However, the run-time is long, which makes it unsuitable when investigating thousands of variants.[77] Microfluidics platforms where metabolites are detected using surface-enhanced Raman scattering (SERS) are a promising option for high-throughput screenings, and has been developed for the detection of several different metabolites, including *p*-coumaric acid and cinnamic acid.[92,93] Transcriptomics, proteomics and metabolomics analysis can give a detailed and holistic picture

of the state of the cell and lay ground rules for the next cycle iteration. Although high-throughput and automated 'omics workflows are developing, the technology does not allow for screening of thousands of variants. High-throughput screens can instead be carried out by coupling product level to survival (e.g. antibiotic resistance) or to a fluorescent output (creating a biosensor).[94,95] The latter is commonly done by engineering a riboswitch or a (synthetic) transcriptional regulator to be activated by product binding.[96,97] In this way, high-producing variants can be selected using microtiter plate screens or fluorescence-activated cell sorting (FACS). However, not all metabolites can be coupled to a fluorescent output. Developing a biosensor for a new metabolite is cumbersome, and the final construct has to be optimized in terms of dynamic range, sensitivity and specificity. In some cases, fluorophore expression can instead be used as a direct measure of productivity, as in Chapter 2 and Chapter 3 of this thesis, where green fluorescent protein (GFP) was used to screen a genome-wide and a small sgRNA library for promising growth decoupling targets.

Learn. The learn phase is the last, and, unfortunately, the most weakly supported step in the DBTL cycle.[54] Here, data from the test phase is thoroughly analyzed to identify gaps and shortcomings that needs to be addressed for a successful next cycle iteration. The information can be used to improve metabolic models, generate new design rules and identify any modifications that should be done to the host organism or production pathway. Further development in high-throughput analytics, metabolic modelling and synthetic biology design standards will make it easier to identify the flaws and shortcomings in strain and construct design, enabling a better supported learn step.[77]

Synthetic biology for cell factories

In synthetic biology, scientist use existing or synthetic biological parts to redesign or build new-to-nature biological systems. It is applied not only in microbial manufacturing, but also in food, agricultural, environmental and health biotechnology. Major events in the history of synthetic biology have been beautifully summarized in a review by Cameron et al;[98] they range from Jacob and Monod's discovery of cellular network regulation to the establishment of the first synthetic circuits, the engineering of synthetic

chromosomes, the construction of a bacteria with a synthetic genome and the use of synthetic biology to commercialize bioproducts and establish cell factories under synthetic dynamic regulation. The last few years, the synthetic biology field has been growing rapidly, with an exponentially increasing number of scientific papers being published each year.[99] Start-ups in the field raised almost $4 billion in 2018, summing up to $12.4 billion the last 10 years.[100]

In order to build predictable synthetic biological systems, promoters, terminators, ribosome binding sites, protein coding sequences, plasmid backbones and other biological parts that are used to make the synthetic constructs need to be well characterized. Therefore, standardization is and has been central to advance synthetic biology. Besides The Registry of Standard Biological Parts and iGEM, standardization efforts include the Synthetic Biology Open Language (SBOL),[101] containing standards for *in silico* and graphic representation of biological designs, BIOFAB (International Open Facility Advancing Biotechnology),[102] a research collaboration that has resulted in the characterization of thousands of genetic elements, and new data models to facilitate the registration of biological part performance.[103] Biofoundries, located all over the world, generate large amounts of data that can be used to implement new design rules and improve standardization measures. They are often operating with the aim of speeding up the DBTL cycle through the use of high-throughput methods, automatization, artificial intelligence and machine learning approaches.[104] Knowledge sharing is further supported by open source biology resources such as OpenWetWare and EcoliWiki.

In the light of industrial biotechnology, synthetic biology has allowed biological engineers to approach cell factory optimization in new ways. On a DNA level, codon optimization has been used to improve protein production, and the development of unnatural base pairs has expanded the genetic code and enabled synthesis of proteins and peptides with novel properties.[60,105,106] Advances in DNA synthesis, reading and editing has sped up the cell factory build process, and enabled the generation of large strain libraries. For regulation of genes and pathways, hybrid and synthetic promoters have been established for predictable control of gene expression,

both in a constitutive and inducible manner.[107,108] Numerous synthetic circuits have been engineered to do everything from counting events[109] and generate synchronized oscilliations[110] to dynamically control metabolic flux[111] and decouple cell growth from metabolite production.[112] On an RNA level, synthetic small RNAs (sRNA)[113] and riboswitches[114] have been used to regulate gene expression and improve biochemical production, and several other promising applications are being investigated.[115] On the protein level, protein domains with novel catalytic functions have been engineered to construct non-natural synthetic pathways,[116] and unnatural amino acids are being used to develop new types of therapeutics.[60,117] Forward engineering metabolic pathways has both enabled and improved the production of various compounds.[98]

Unsurprisingly, inserting all these types of synthetic devices into microbes can result in significant metabolic burden, genetic instability, and disruption of endogenous metabolic functions.[118] Measures to circumvent this can be taken by model-based prediction of host interaction with synthetic constructs and introduction of burden-regulated feedback mechanisms.[119,120]

An ultimate goal of synthetic biology would be to forward engineer whole microbes for a desired purpose. In the future, it might be possible to design minimal cell factories optimized for the production of a specific biochemical or protein. Although still far from reach, significant progress has been made. This includes the synthesis of whole synthetic chromosomes in yeast[121] and the creation of a bacteria with a synthetic,[82] minimal[122] and recoded[123] genome. Naturally, the advancement of synthetic biology has raised discussions on ethics and scientific code of conduct, as the construction of synthetic organisms may have implications for the environment and for human and animal health. Spread of genetically modified organisms (GMO) in nature could negatively affect biodiversity, and microbes or viruses could be designed to be used as bioweapons.[124] With all opportunities offered by synthetic biology, especially from the growing number of applications made possible by CRISPR-based engineering, it is important to take ethics into consideration and set up regulatory systems and policies that are adapted to the field.[125]

CRISPR-Cas. The emergence of CRISPR-based technologies has revolutionized the synthetic biology and metabolic engineering field. CRISPR-Cas is a bacterial adaptive immune system, of which the function and structure was discovered and thoroughly investigated in the 90's and 00's.[126–130] Since then, it has been repurposed as a precise engineering tool for bacteria, yeast, mammalian cells and plants.[131–134]

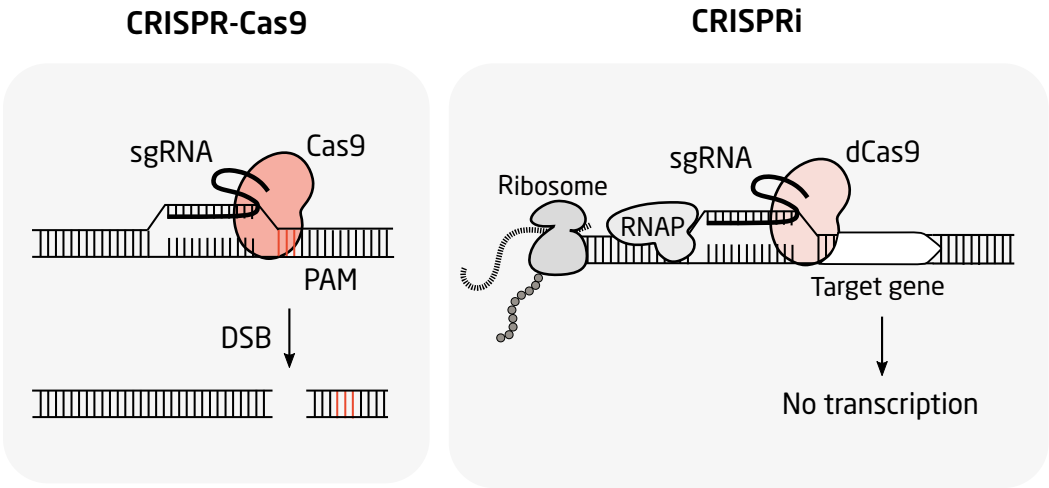


Figure 1.6. CRISPR-Cas9 and CRISPRi. Left panel: Cas9 and sgRNA pairs up, and the sgRNA guides the dCas9 to the DNA sequence complementary to the initial 20 nucleotide sequence of the sgRNA. Cas9 cleaves the DNA, causing a double-stranded break. The break can be repaired vid non-homologous end joining, or homologous recombination using a donor DNA. Cas9 can only cleave DNA in the presence of a PAM site. The PAM site sequence of Cas9 is 5'-NGG-3', where N can be any nucleotide. Right panel: guiding dCas9 to a target gene blocks transcription of the gene via steric hindrance. Binding within the target gene blocks RNAP elongation, while binding to the promoter region prevents initiation of elongation. Abbreviations: sgRNA; single guide RNA, DSB; double strand break, PAM; protospacer adjacent motif, Cas9; CRISPR-associated protein 9, dCas9; catalytically dead Cas9, RNAP; RNA polymerase.

CRISPR-Cas systems are found in around half of the sequenced bacterial genomes.[135] They are divided into two main classes based on the CRISPR associated protein (Cas); class 1 uses multiple Cas proteins and class 2 uses a single Cas protein. The Cas proteins are nucleases that are guided by small RNAs to cut DNA target sequences. The RNAs are expressed from CRIPSR loci that contain residue DNA from phages and other foreign genetic elements that have previously infected the cell. CRISPR-Cas confers resistant to invasion of phages with similar DNA structure, as the RNAs will recognize and guide Cas-mediated double strand break of the foreign DNA. In order to cut the DNA, most Cas proteins require a protospacer adjacent motif (PAM) to be present next to the RNA binding site. The PAM sequence is usually 2-6 bp, and is dependent on the specific Cas protein.[136]

The *Streptococcus pyogenes* Cas9 is the most commonly used Cas enzyme for engineering purposes. It is guided by an engineered single guide RNA (sgRNA), which can be designed to bind and induce Cas9-mediated strand break at a designated locus (Fig. 1.6). CRISPR-Cas9 has been used for a range of applications, including genome engineering, construction of genome-wide libraries, and directed basepair exchange.[86,87,137] By mutating active site residues histidine 10 and aspartic acid 840 of the Cas9 enzyme into alanine, CRISPR has further been repurposed into an efficient regulator of gene expression; CRISPR interference (CRISPRi) (Fig. 1.6) can be used to inhibit transcription,[138] and CRISPR activation (CRISPRa),[139] where dCas9 is fused with a transcriptional activator, can be used to enhance transcription. Inducible control of sgRNA and dCas9 expression allows silencing of a gene or cellular function to take place at a desired timepoint, which can be highly useful for dynamic metabolic engineering purposes.[138] In Chapter 2 and Chapter 3 of this thesis, CRISPRi was used to screen for gene targets that could function as growth decoupling switches, and to increase sdAb production through inhibition of nucleotide biosynthesis, respectively.

1.3 Regulation of growth and metabolism

The growth and metabolism of microbes is regulated through networks of complex interactions. External cues, such as pH, temperature and nutrient availability, are sensed and transformed to cellular responses through transcriptional, translational and proteolytic changes that affect metabolism, morphology, growth and other processes.[140] A brief overview of the regulation of growth and metabolic pathways that are relevant for this thesis is given below.

Inducible promoters – the tryptophan operon

Microbes can use inducible promoters to respond to nutrient availability and environmental changes by coupling presence or absence of a metabolite or physical input to the expression or repression of a certain gene, operon or pathway. The response can for example be mediated by binding of the metabolite to a regulator protein, resulting in binding or dissociation of the protein from its DNA binding sequence, and subsequent repression or derepression of the promoter. One of the most well studied examples is the tryptophan operon.

In 1953, Monod and Cohen-Bazire concluded that biosynthesis of L-tryptophan in *E. coli* is repressed in presence of tryptophan, and induced when tryptophan is depleted.[141] The regulatory mechanisms of the operon have been extensively studied since then. There are five structural genes that encode the enzymes responsible for converting chorismate to L-tryptophan; *trpE*, *trpD*, *trpC*, *trpB* and *trpA* (Fig. 1.7).[142] The expression of these genes is regulated by the trp repressor (TrpR). When cells are grown in media with tryptophan, TrpR binds the amino acid as well as the tryptophan operator (*trpO*), which hinders expression of the *trpE-A* genes. In absence of tryptophan, the conformation of TrpR changes and it disassociates from *trpO*, enabling transcription of the *trpE-A* genes.[143] Regulation is further enhanced by a mechanism called attenuation; in front of *trpE-A*, there is a leader sequence (*trpL*) with an attenuator.[144] The leader sequence also contains two tryptophan-encoding codons in a row. This is unusual, as only

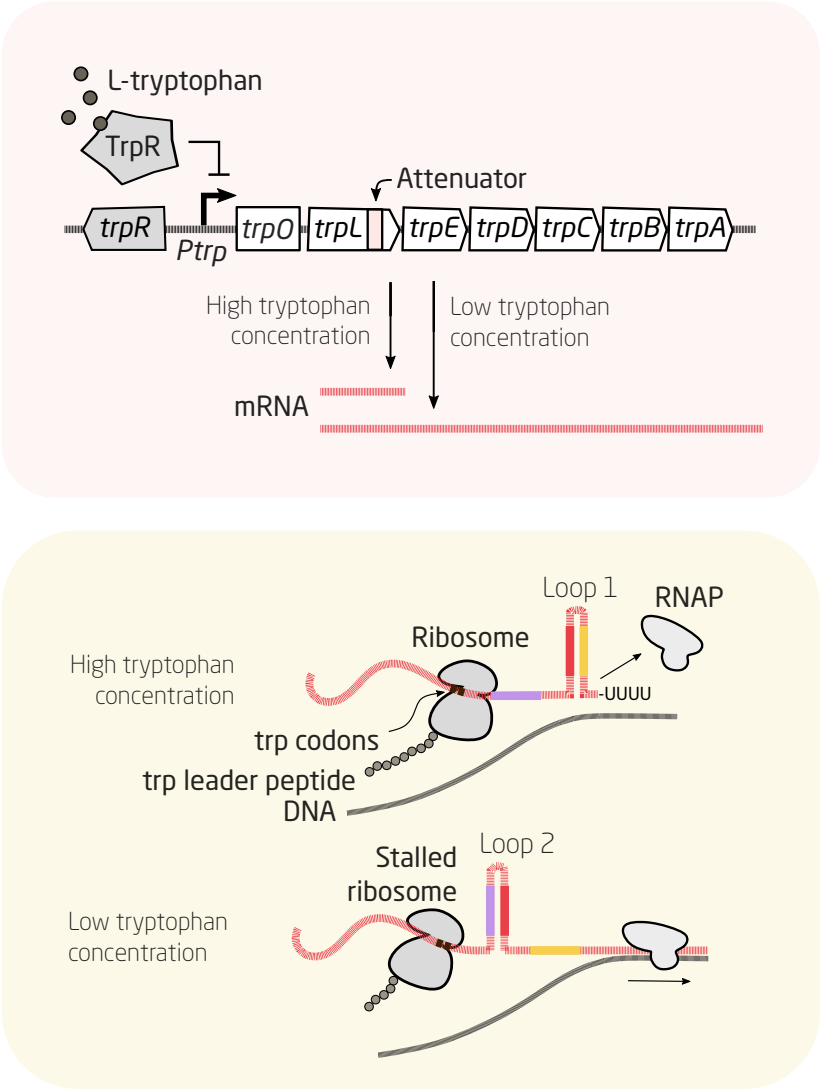


Figure 1.7. Regulation of the tryptophan operon in *E. coli*. Upper panel: the structural genes *trpE*, *trpD*, *trpC*, *trpB* and *trpA* are regulated by the tryptophan repressor (TrpR), which binds operator *trpO* in presence of tryptophan. Lower panel: expression of the structural genes is further regulated by attenuation; at high tryptophan levels, the ribosome translates the leader peptide at full speed, leading to the formation of stem loop 1 at the attenuation site of the leader sequence *trpL*. This causes early transcription termination. At low tryptophan concentrations, the ribosome stalls at the dual tryptophan codons, leading to the formation of stem loop 2. The transcription of the structural genes can proceed without interruption. Abbreviations: RNAP; RNA polymerase, trp; tryptophan.

one out of one hundred amino acids are L-tryptophan in an average *E. coli* protein.[145] Depending on the tryptophan availability, the trp leader sequence can form different stem loops that affect the transcription of the *trp* mRNA. In low tryptophan concentrations, the ribosome is stalled at the double tryptophan codon site; this will lead to the formation of a stem loop that allows continued transcription of the *trp* mRNA by the RNAP. At high tryptophan concentrations, the ribosome will not stall while translating the mRNA. This leads to the formation of another *trpL* stem loop that works as a RNAP termination signal, which stops transcription (Fig. 1.7).[146] Together with TrpR repression, attenuation of *trpL* enables tight regulation of tryptophan biosynthesis. The intrinsic regulatory properties of the tryptophan operon, as well as of other native inducible systems, has been widely exploited and engineered when creating inducible expression systems to optimize for microbial production of proteins and biochemicals. Chapter 4 in this thesis describes the development of a tryptophan-regulated gene expression system.

Nucleotide biosynthesis

Nucleotides are the building blocks for DNA and RNA, and are essential for many other metabolic processes and signaling pathways. They can be synthesized *de novo* from various precursor molecules, or through salvage pathways using extracellular metabolites or intracellular degradation products of DNA and RNA as start material.[147] Degradation of nucleotides is primarily induced when cells are grown under nitrogen starvation.[148]

De novo synthesis of purines (adenine and guanine) starts with the formation 5-phosphoribosyl-1-amine (PRA) by PRPP amidotransferase (PurF) (Fig. 1.8). The activity of PurF is allosterically regulated by nucleotide monophosphates (NMPs) adenosine monophosphate (AMP) and guanosine monophosphate (GMP).[147] Expression of *purF* is negatively regulated by the purine repressor (PurR), which, upon binding corepressors guanine and hypoxanthine, regulates the expression of most genes in the purine *de novo* biosynthesis pathway.[149] The end products of the purine pathway, AMP and GMP, can be further converted into nucleotide- and deoxyribonucleotide triphosphates (NTPs and dNTPs) and used as RNA and DNA building blocks. Synthesis of NMPs via the purine salvage pathway is carried out through phosphorol-

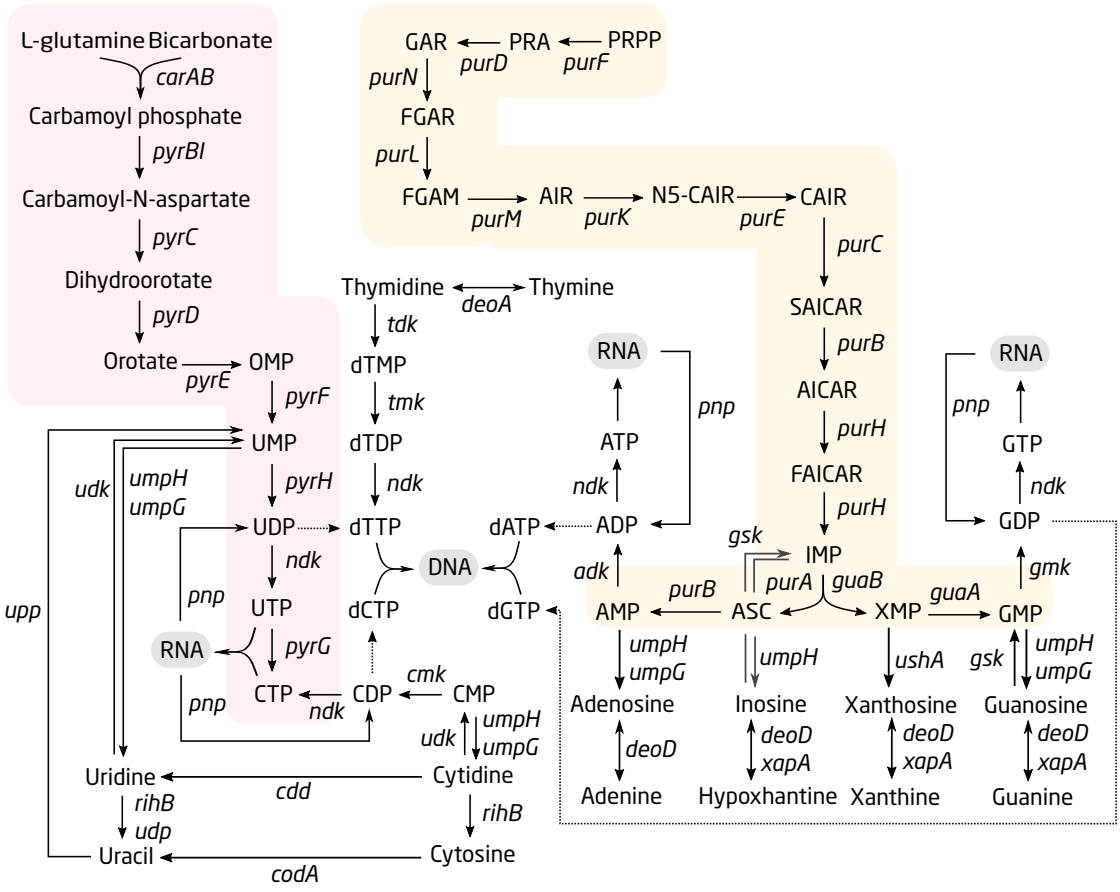


Figure 1.8. Nucleotide biosynthesis. An overview the *de novo* pyrimidine (pink) and purine (yellow) biosynthesis pathways together with a selection of the nucleotide salvage and interconversion pathways of *E. coli*. Constructed based on information from KEGG, Martinussen et al.[147] and Jensen et al.[148] Please note that the scheme only depicts selected reactions in the salvage and interconversion pathways. Dashed arrows indicate several intermediate reactions. Abbreviations: PRPP; Phosphoribosyl diphosphate, PRA; 5-Phosphoribosyl-1-amine, GAR; 5-Phosphoribosyl-1-*N*-glycinamide, FGAR; 5-Phosphoribosyl-1-*N*-formylglycinamide, FGAM; 5-Phosphoribosyl-1-*N*-formylglycinamidine, AIR; 5´-Phosphoribosyl-1´-*N*-(5-amino)imidazole, N5-CAIR; 5´-Phosphoribosyl-1´-*N*-(5-amino)imidazole-5-*N*-carboxylate; CAIR; 5´-Phosphoribosyl-1´-*N*-(5-amino)imidazole-4-carboxylate, SAICAR; 5´-Phosphoribosyl-1´-*N*-(5-amino)imidazole-4-*N*-succinocarboxamide, FAICAR; 5´-Phosphoribosyl-1´-N-(5-amino)imidazole-4-N-carboxamide, IMP; Inosine 5´-monophosphate,

Figure 1.8 continued. ASC; Adenylosuccinate, AMP; Adenosine 5′-monophos-phate, ADP; Adenosine 5′-diphosphate, ATP; Adenosine 5′-triphosphate, dATP; 2′-deoxyadenosine 5′-triphosphate; XMP; Xanthosine 5′-monophosphate; GMP; Guanosine 5′-monophosphate, GDP; Guanosine 5′-diphosphate, GTP; Guanosine 5′-triphosphate, dGTP; 2-deoxyguanosine 5′-triphosphate.OMP; Orotidine 5′-monophosphate, UMP; Uridine 5′-monophosphate, UDP; Uridine 5′-diphosphate, UTP; Uridine 5′-triphosphate, CMP; Cytidine 5′-monophosphate, CDP; Cytidine 5′-diphosphate, CTP; Cytidine 5′-triphosphate, dCTP; 2′-deoxycytidine 5′-triphosphate; dTMP; deoxythymidine monophosphate, dTDP; deoxythymidine diphosphate, dTTP; 2′-deoxythymidine 5′-triphosphate.

ytic cleavage and phosphoribosylation of ribonucleosides and deoxy-ribonucleosides (Fig. 1.8).[147]

De novo synthesis of pyrimidines (cytosine, thymine and uracil) starts with the formation of carbamoyl phosphate from bicarbonate and L-glutamine (or ammonia) (Fig. 1.8). This step is carried out by carbamoyl phosphate synthase (CarAB). Expression of *carAB* is repressed by PurR and negatively regulated by high intracellular levels of uridine triphosphate (UTP) and L-arginine.[150,151] The subsequent steps of the pyrimidine pathway are subject to regulation by several other fascinating mechanisms; expression of aspartate carbamoyltransferase (*pyrBI*) and orotate phosphoribosyl-transferase (*pyrE*) is downregulated by attenuation in response to high concentrations of UTP;[152] translation of dihydroorotase (*pyrC*) and dihydroorotate dehydrogenase (*pyrD*) is inhibited at high CTP concentrations through start-site switching;[153] and cytidine triphosphate (CTP) synthase (PyrG) is allosterically regulated by guanosine triphosphate (GTP) and CTP.[148] The *de novo* pathway products CTP and UTP can be incorporated into RNA, while DNA deoxyribonucleotides dTTP and dCTP are synthesized from uridine diphosphate (UDP) and cytidine diphosphate (CDP), respectively. Salvage of pyrimidines is carried out through phosphorylation of uridine and cytidine into their respective NMPs (Fig. 1.8).[147]

Stationary phase and stress response

The stationary phase and stress response of *E. coli* is activated in the post-exponential and stationary growth phase, as well as upon exposure to a number of different stresses including starvation, heat shock, low pH, DNA damage and high osmolarity.[154,155] It is essential for the survival in prolonged stationary phase and stress conditions, and is mediated by RNA polymerase sigma S (RpoS, σ^S or σ^{38}), which regulates around 10% of the genes in *E. coli*.[156] RpoS is one of seven sigma factors that binds the RNA polymerase (RNAP) core enzyme and enables transcription from promoters with sigma factor specific sequence motifs.

The cellular levels of RpoS are tightly controlled. Transcriptionally, gene expression is regulated by (at least) two transcription factors; cAMP-CRP (cyclic adenosine monophosphate-cAMP receptor protein), which can induce or repress *rpoS* depending on the growth conditions; and ArcA (anaerobic respiratory control protein), which represses *rpoS*, possibly in response to high ratios of $NADH/NAD^+$.[155] The intracellular signaling molecule guanosine tetraphosphate (ppGpp) involved in the stringent response also plays an important role in *rpoS* expression, as RpoS levels and activity are significantly reduced in mutants not able to produce this molecule.[157] On a translational level, *rpoS* mRNA is regulated by RNA-binding proteins[158] and small regulatory RNAs that, for example, prevent the formation of secondary structures in order to enable efficient translation.[159] On the protein level, RpoS is subject to proteolytic degradation. During stress-free conditions and exponential growth, RssB (regulator of RpoS) binds to the sigma factor, enabling ATP-dependent degradation by the ClpXP protease.[160] In response to stress or starvation, RssB is inhibited by anti-adapter (Ira) proteins and RpoS rapidly accumulates in the cell.[161] As levels of RpoS increase, it outcompetes the *E. coli* housekeeping sigma factor RNA polymerase sigma D (RpoD or σ^{70}) for binding the RNAP core enzyme.

The expression of RpoS-dependent genes reshapes the protein landscape in a way that enables survival in stress and starvation conditions. Although different genes are induced depending on the specific stress that the cells are exposed to,[162] the bacteria generally enter an energy-saving mode. The cell

volume is decreased and there is a downregulation of genes involved in aerobic metabolism and respiration.[162] Synthesis of flagellar proteins is reduced and the cells become less motile.[163] The ribosomal content, which is linearly correlated to growth rate,[164] is reduced as ribosomal RNA synthesis is downregulated. This leads to a reduction of the protein synthesis rate to less than 20% of the rate of exponentially growing cells.[165] At the same time, genes involved in anaerobic respiration and fermentative metabolism are upregulated.[155] The cellular membrane composition is altered, and expression of transporters and other membrane proteins is increased.[162]

Stringent response

The stringent response is induced by various stress conditions, such as starvation of amino acids, carbon, phosphate and nitrogen.[166] It is closely intertwined with the stationary phase and stress response; the stringent response is mediated by the stress alarmone ppGpp, which also induces expression of *rpoS* and stabilizes RpoS binding to the RNAP core enzyme.[157] The levels of ppGpp and RpoS are linearly correlated in most *E. coli* strains.[167] Besides RpoS, ppGpp accumulation also induce or enhance the activity of some genes under control of RpoD, RNA polymerase sigma E (RpoE or σ^{24}), RNA polymerase sigma N (RpoN or σ^{54}) and RNA polymerase sigma H (RpoH or σ^{32}).[168,169] Transcriptional profiling has shown that 500 genes in *E. coli* K-12 MG1655 are affected by the stringent response.[170]

When cells are starved for amino acids, binding of uncharged transfer-RNA (tRNA) to the ribosomes induces synthesis of guanosine pentaphosphate (pppGpp) by GTP pyrophosphokinase (RelA). pppGpp is then hydrolyzed to ppGpp. SpoT, the other main *E. coli* ppGpp synthase, is responsible for ppGpp synthesis in response to starvation of other nutrients.[171] As ppGpp accumulates in the cell, it binds to RNAP together with RNAP-binding transcription factor DksA. This destabilizes RNAP binding to rRNA promoters, and results in significant downregulation of rRNA transcription and subsequent reduction of protein synthesis.[171] Growth-related processes are inhibited, while the expression of genes involved in stress survival, proteolysis and amino acid transport and biosynthesis is upregulated.[171]

1.4 Balancing growth and production

In nature, microbes rarely encounter the ideal growth conditions that are provided in the laboratory. Nutrients are scarce, and if available, most bacteria or yeast are evolved to quickly utilize them to make more cells and secrete byproducts that will give them a competitive advantage toward other microorganisms later on.[155] Unless cell mass or byproducts are the desired end-products, this intrinsic capability can be disadvantageous when using microbes for bioproduction. Excessive growth drains cellular resources from product formation, leading to reduced product TRY.[172] High cellular densities also increase the need for energy-intensive stirring, cooling and (for production under aerobic conditions) oxygenation. The fact that the supplied carbon source is a major expense associated with bioproduction further highlights the importance of channeling the metabolic flux toward production.[17] However, strong constitutive expression of a protein or product pathway already from the start of the fermentation may interfere with or even inhibit cellular growth, which also reduces productivity. Thus, it is important to cleverly balance the distribution of carbon and nutrients between cell mass and product.[173]

Ideally, a microbial bioproduction process should consist of two phases (Fig. 1.9).[174] In the first phase, the growth phase, there is little or no production of the desired protein or metabolite. Instead, cells are allowed to rapidly accumulate until the optimal amount of catalytic biomass is reached. In the second phase, the production phase, growth is stalled and the carbon flux is rewired toward expression of the protein or the biochemical that should be produced. Decoupling growth and production in this way can increase product yield and titer, and further improve the bioprocess economy by decreasing the cost for stirring, cooling and downstream processing. It can be achieved by different means of engineering the microorganism and/or modifying the bioprocess conditions phases (Fig. 1.9). The shift from the growth to the production phase should preferably not require addition of expensive chemical inducers, or induce a stress response which lowers the metabolic activity or capacity for protein synthesis.

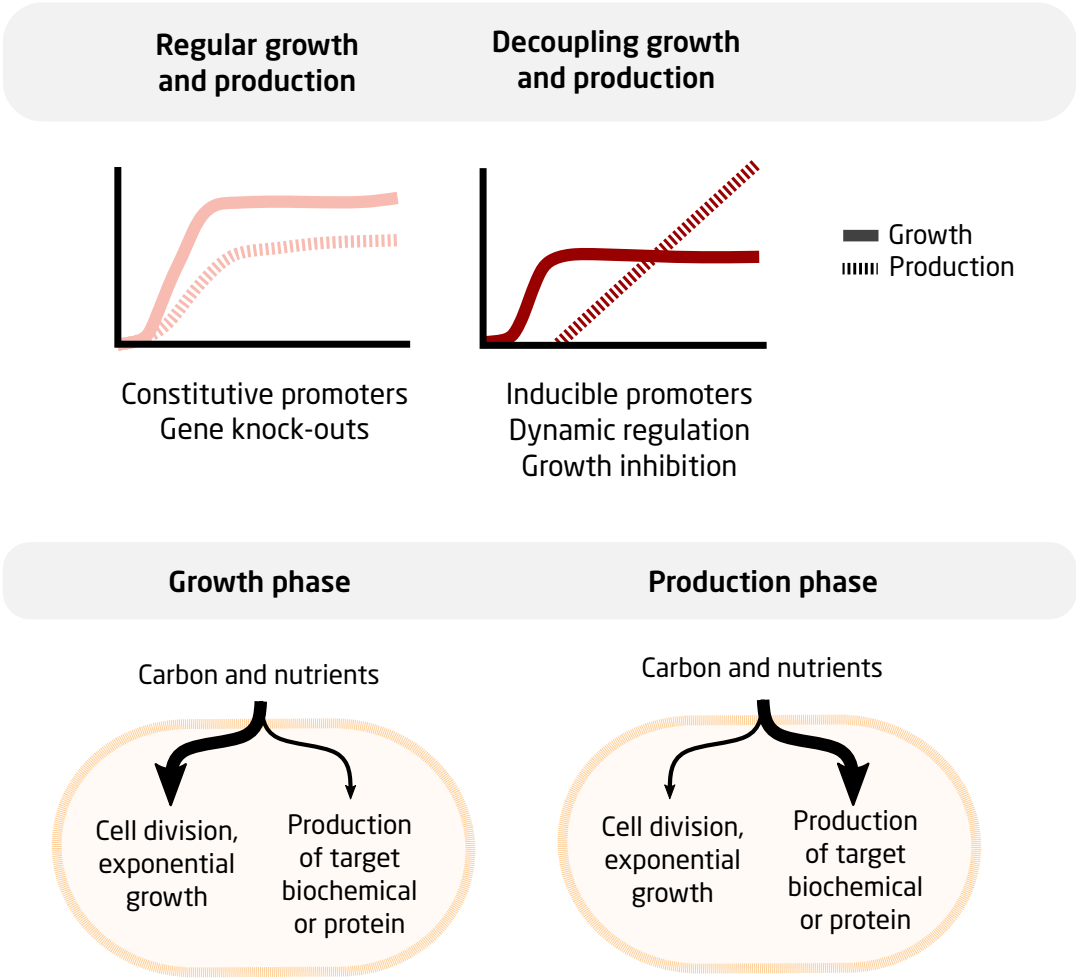


Figure 1.9. Decoupling growth and production. Upper panel: schemes showing regular and decoupled growth and production, and different methods to achieve it. Lower panel: desired carbon and nutrient flux in the growth and production phase.

Inducible expression systems

Ever since Monod and Jacob's discovery of the regulatory mechanisms of operons,[175] scientists have worked on engineering inducible promoters and expression systems that can be used for the production of proteins and biochemicals. New systems are continuously being discovered and developed.[176] Inducible promoters are especially useful for production of toxic proteins, and for biochemicals produced via complex pathways. The

expression and activities of long and complex product pathways oftentimes competes with native metabolic processes, and can lead to depletion of energy, decreased co-factor availability or inhibition of growth by toxic byproducts or pathway intermediates.[177] In order to minimize these effects, inducible promoters can be used to start expression of the pathway or protein when the desired cell density has been reached. There are many inducible expression systems available for *E. coli*, and some of the most commonly used ones are listed in Table 1.1.

Several aspects should be taken into consideration when engineering inducible promoters for microbial cell factories. The ideal promoter should have a low basal level expression (i.e. low "leakiness") when in the uninduced state, be capable of high expression when in the maximum induced state, and be tunable (over a wide dynamic range) so that expression strength can be adapted and optimized for the specific product or pathway enzyme. Low basal level expression and increased maximum expression levels can for example be engineered by random mutagenesis of the promoter and/or operator site,[178,179] or through the construction of hybrid promoters.[180] Modulation of the -10 and -35 RNAP binding sites using saturation mutagenesis or thermodynamical models predicting RNAP-DNA interaction has been used to establish synthetic promoters with improved dynamic range.[181–184] Modifying the host strain for constitutive inducer uptake, for example by knock-out of specialized transporters, has proven useful for enabling or improving system tunability.[185,186] Finally, engineering the inducer-binding transcription factors by hybridization of different repressors and/or mutagenesis of selected regions can enhance repression and enable induction by novel inducers.[187,188]

Another appealing trait for inducible systems is autoinducibility, i.e. induction of the promoter without the addition of chemical inducers, and without manual changes of fermentation conditions. Autoinducible systems can decrease fermentation complexity and the risk of contamination, and decrease the overall process costs as inducers are usually expensive to add to large scale fermentations. In this thesis, an autoinducible trp-T7 expression system that responds to tryptophan depletion was engineered and tested for

Table 1.1. An overview of some of the commonly used inducible expression systems in *E. coli*.

Promoter/ regulator	Mode of induction	Expression strength	Comments	Ref.
Plac/ LacI	Addition of IPTG or growth on lactose	Low	Relatively expensive inducer (IPTG), auto-inducible in glucose/lactose media. Repressed by glucose. The mutated version *PlacUV5* enables higher expression levels.	189,190
Ptrp/ TrpR	Tryptophan starvation or addition of 3-indoleacrylic acid	Intermediate	Low basal level expression, autoinducible in low tryptophan media. Engineered leaderless versions (without *trpL*) enable higher expression levels.	191
Ptac, Ptrc/ LacI	Addition of IPTG or growth on lactose	Intermediate	Two hybrid promoters of *Ptrp*/*Plac* with lower basal- and higher expression level than *Plac*. Relatively expensive inducer (IPTG), autoinducible in glucose/ lactose media. Titratable in *lacY*- strains.	180,183, 185
P_{T7}/ T7 RNAP	Various§	Very high	Phage-derived T7 RNAP is usually integrated in prophage λDE3 in the genome of the production host; commonly under control of *PlacUV5*/LacI. Extensively used for industrial protein production.	192

araP$_{BAD}$/ AraC	Addition of arabinose	Intermediate to high	Titratable and tight control of expression, inexpensive inducer. The arabinose pathway should be deleted to avoid inducer degradation.	186,193
rhaP$_{BAD}$/ RhaS	Addition of rhamnose	Intermediate to high	Tight control of expression, expensive inducer. The rhamnose pathway should be deleted to avoid inducer degradation.	194
PprpB/ PrpR	Addition of propionate	Intermediate	Titratable control of expression, inexpensive inducer.	195
Pm/ XylS	Addition of *m*-toluate	Low to high	Titratable and tight control of expression, inexpensive inducer. The mutated promoter derivative *Pm* ML1-17 enables higher expression levels.	196,197
p_L/ cI^{857}	Increase in temperature to 42 °C	High	Promoter and repressor derived from the λ phage. Low basal level expression. Repressor cI harbors a heat-sensitive mutation and is subject to thermal inactivation.	198
PLtetO-1/ TetR	Addition of tetracycline or derivative	Intermediate to high	Titratable promoter with low basal level expression. Relatively inexpensive inducer.	199

§ Expression of T7 RNAP is usually controlled by the *Plac*-derived *lacUV5* promoter and induced by addition of IPTG. Several other T7-based expression systems have been engineered, including temperature-, tryptophan-, rhamnose- and arabinose-inducible constructs.[200–203]

production of a toxic protein and the biochemical L-serine (Chapter 4). The time of induction could be modified by varying the start concentration of tryptohan (or tryptophan-containing yeast extract) in the media.[200] Promoters responsive to stress,[204] oxygen limitation[205] or starvation of other nutrients, such as phosphate[206] and glucose,[207] have also been used to engineer autoinducible expression systems. The LacI-based promoters (Table 1.1) can be induced by growth on lactose, and autoinducible production using these promoters can be carried out in batch fermentation by modifying the ratio of glucose/lactose in the start media.[208] However, due to the low solubility of lactose in water, this is unfeasible for large-scale fed-batch fermentations.

The application of quorum sensing (QS) circuits is another interesting approach for engineering autoinducible pathway induction. QS is used by bacteria to communicate and regulate various functions, such as biofilm formation and virulence, via the expression and sensing of autoinducers.[209] The luciferase (lux) operon from *Vibrio fischeri* is a well-studied example. The lux acyl-homoserine lactone (AHL) autoinducer N-(3-oxohexanoyl)-homoserine lactone (3OC6-HSL) is synthesized by AHL synthase (LuxI). As the *V. fischeri* population density increases, so does the concentration of the autoinducer. When the autoinducer has reached a certain threshold concentration, 3OC6-HSL is bound by transcriptional regulator LuxR. This leads to expression of the luciferase-encoding genes and production of light.[209] LuxR/I expression systems have been engineered to control autoinducible production of bisabolene,[210] and applied for dynamic regulation of metabolic flux.[111]

Dynamic regulation

Dynamic regulation of metabolic flux is a promising approach for optimizing production of a range of different compounds, and can be used to further improve "statically" engineered microbial cell factories. Strategies can be independent of, or specific to, the production pathway. They generally aim to dynamically redirect flux from growth or byproduct formation toward production, to improve precursor availability, or to balance product pathway expression (Fig. 1.10). Unlike static metabolic engineering that includes knock-out of competing pathways, integration of heterologous enzymes and

gene expression from constitutive promoters, dynamic engineering allows microbes to dynamically adapt their metabolism to the changing conditions in the cell or the fermenter, which minimizes metabolic burden, enables better allocation of cellular resources and balances growth and production.[173] Although dynamic regulatory systems are often complex to engineer, they can be highly advantageous when expressing complex heterologous pathways, and for production in larger scale where there is a heterologous distribution of nutrients and oxygen.[211] They further allow for the downregulation of enzymes, fluxes or competing pathways that are essential to the cell.[212]

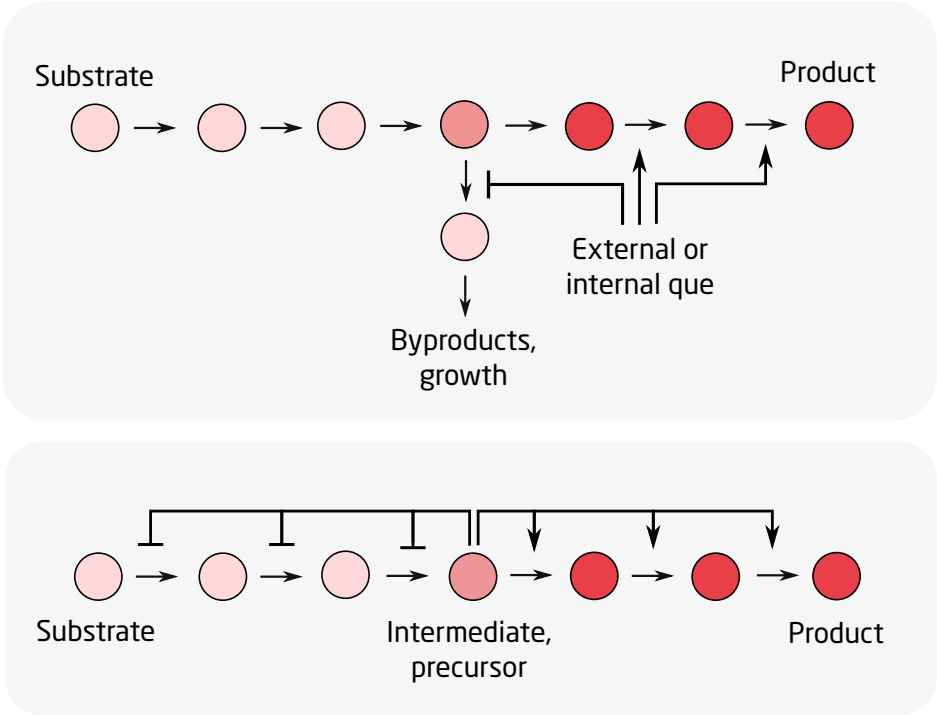


Figure 1.10. Examples of dynamic regulation. Upper panel: Dynamic regulation of growth, product and byproduct expression based on sensing of an external or internal que, such as temperature shift, quorum sensing molecules or inducers. Lower panel: Dynamic regulation of product pathway expression based on sensing of an intermediate or precursor that activate and/or repress pathway enzymes to balance metabolic flux.

There are several examples where promoters or transcription factors responsive to metabolic activity, precursors, pathway intermediates, or signaling molecules have been identified, engineered and applied for dynamic regulation of biochemical production.[204,213] Transcription factor-based dynamic regulation of malonyl-CoA, an important precursor molecule for fatty acid and polyketide synthesis, was achieved by constructing a feedback loop that adjusted the upstream synthesis and downstream consumption of this key metabolite.[214] This improved fatty acid production more than 2-fold.[214] In another study, two enzymes in the lycopene biosynthesis pathway were placed under control of a regulator activated by acetyl phosphate (AP).[215] AP serves as an indicator of glucose availability and glycolytic flux. As cells were engineered to only produce lycopene in response to high AP concentrations (i.e. in excess glucose conditions), the metabolic burden was minimized and lycopene production increased significantly.[215] In yeast, a cell density responsive synthetic QS system based on the yeast mating pathway was repurposed to simultaneously induce pathway gene expression and silence byproduct formation, significantly increasing the yield and titer of para-hydroxybenzoic acid.[216,217]

Dynamic redirection of flux can be done by regulating key enzymes that are involved in growth-related or byproduct-forming processes, or are part of or acting upstream of the production pathway.[218,219] Key enzymes can be engineered to respond to environmental or synthetic input such as degradation tags or temperature, or be regulated by inducible promoters.[218,220] Heterologous QS systems have for example been used to control the expression of phosphofructokinase-I (Pfk-I), which enabled cell density dependent, autonomous and dynamic regulation of metabolic flux into product pathways, and increased the titer of *myo*-inositol more than 5-fold.[111] Isopropanol titers were increased 4-fold in an *E. coli* strain where carbon flux was redirected to product formation through inducible inhibition of a TCA cycle key enzyme,[221] and targeted proteolysis of Pfk-I increased the *myo*-inositol yield two-fold.[218]

There are numerous tools available for engineering dynamic regulation. Inducible promoters, discussed in the previous section, are useful for regulating key enzymes and re-directing pathway fluxes. CRIPSRi is efficient

at downregulating expression of competing reactions, and can be multiplexed to inhibit several targets simultaneously.[222] The development of CRIPSRi[138] and CRISPRa[139,223] has enabled the use of CRISPR for simultaneous repression and activation of different gene targets.[224] Inhibition of key enzymes can also be carried out using various degradation tags that enable inducible protein degradation.[219,225] Most of these tools are also useful for engineering growth inhibition-based decoupling.

Growth inhibition

Inhibition of cell growth can be used to decouple growth and production, as growth-inhibited cells are forced to utilize available carbon and nutrients for product formation instead of growth. This strategy often overlaps and makes use of dynamic regulation efforts that involves inhibiting essential key enzymes or steering flux away from pathways essential for growth.[212] Generally, growth inhibition can be carried out by starving the cells for nutrients or engineering synthetic circuits that, at a certain timepoint, stop the cells from growing in response to internal or external input.

Nutrient starvation. Although nutrient depletion generally leads to a stationary phase and stress response that decrease cell productivity, starvation of nutrients, metals and trace elements has successfully been applied to increase production of various biochemicals. The cell behavior and metabolic adjustments in response to starvation is dependent on the specific nutrient that is depleted; for example, the glucose consumption rate of *E. coli* is less affected by starvation of magnesium or sulfate compared to starvation of nitrogen.[172] It has been shown that the availability of acetyl-CoA, an important precursor molecule for numerous biochemicals, is improved in sulfur-starved cells,[226] and that the production of many acetyl-CoA-derived metabolites including limonene, naringenin and mevalonate can be improved by magnesium starvation.[227–229] Limiting thiamine in an *E. coli* thiamine auxotroph increased the D-lactate titer almost 20%,[230] and double limitation of magnesium and nitrogen has been shown to significantly improve the production of fatty acids in yeast.[231]

It is also possible to utilize changes in oxygen availability or pH to switch from a growth to a production phase. Although this strategy is not generally

applicable, it has been shown to work very well for selected products. "Starving" cells for oxygen in two-stage fermentations with an aerobic growth phase and an anaerobic production phase has been used to improve the production of both lactic acid and succinate in *Lactobacillus* and *Corynebacterium*, respectively.[232,233] Production of 2-ketogluconic acid in *Klebsiella pneumoniae* reached a remarkable titer of 186 g/L in a similar two-stage fermentation where the growth-to-production shift was induced by a drop in pH.[234]

It can be a challenge to maintain an active metabolic state when using nutrient starvation to inhibit growth. Metabolic rates decrease, and the general protein synthesis and degradation rate in stationary phase cells drops to 0.5-20% of that in exponentially growing cells.[174] Different approaches have been taken to increase or maintain the glycolytic flux in stationary phase cells. Overexpressing phosphoenolpyruvate protein phosphotransferase (PtsI), involved in glucose uptake and normally inhibited during nitrogen starvation, improved metabolic rates up to 4-fold in nitrogen-limited stationary phase cells.[235] Altering ppGpp levels by mutation of stringent response enzymes, and increasing ATP consumption by overexpression of the F_1 part of ATP synthase, has also been shown to boost glycolytic flux in slow-growing or growth arrested cells.[236,237]

Most of the promoters used for protein and biochemical production are based on RpoD-recognition.[238] Since RpoS is the dominating sigma factor in stationary phase, expression of product pathways and proteins under RpoD promoters can decrease when cells encounter starvation. Synthetic inhibition of growth, where cells stop growing without experiencing a typical stationary phase response, can be a promising approach to overcome this.

Synthetic growth inhibition. Decoupling growth from production by synthetic growth inhibition often involves degradation of essential enzymes, or the inhibition of essential genes or cellular functions. Dynamic regulatory strategies are often used to inhibit essential enzymes or downregulate metabolic pathways needed for growth. Decoupling can be carried out through transcriptional, translational or posttranslational regulatory strategies (Fig. 1.11).

On a transcriptional level, growth can be modulated by direct control of the transcription machinery. By putting RNAP under control of an inducible promoter, the growth rate could be tuned from zero to maximum depending on the inducer concentration.[239] The non-growing cells were still metabolically active and produced near maximum theoretical yield of glycerol from glucose. Using inducible inhibition of the native *E. coli* RNAP, and T7 RNAP-based expression of target protein, the yield of glycotransferases could be increased 12-fold.[240] Further scale-up and testing of the system in large scale fed-batch fermentation showed that target protein content reached >55% of the total cell dry weight.[241] There are also many interesting examples of genetic circuits that inhibit growth and induce production, for example in response to glucose starvation,[242] sensing of nutrients,[112] addition of inducers[243] and inducer depletion.[244] The latter was used to increase wax ester yield >3-fold by enabling a gradual, dynamic transition from biomass to production state in response to degradation of arabinose.[244] Placing the essential gene isocitrate dehydrogenase (*icd*) under control of p_L/cI^{857} enabled a temperature-dependent switch from growth to production; this two-stage approach increased productivity of itaconic acid almost 50%.[245]

Transcriptional inhibition using CRIPSRi can be applied to modulate growth and production. (Fig. 1.11) CRISPRi-based repression of essential genes has for example been used to arrest growth and improve production of biofuels in cyanobacteria and naringenin in *E. coli*.[246,247] In *E. coli*, CRISPRi-based repression of four genes (*dnaA*, *oriC*, *pyrF* and *thyA*) involved in DNA replication and nucleotide synthesis has been shown to inhibit biomass formation and increase the yield of GFP and mevalonate.[248] Using single-cell microfluidics, it was further shown that *pyrF*-inhibited cells stopped dividing, but continued to produce GFP for up to 40 h.[248,249] As the rational design of functional growth decoupling targets can be challenging, a follow-up study presented in Chapter 2 of this thesis used a library-based approach to screen for additional promising targets.[250] Interestingly, the genome-wide approach confirmed that inhibition of nucleotide synthesis is a suitable growth decoupling strategy. Thus, in Chapter 3 of this thesis, twenty-one targets in the purine and pyrimidine biosynthesis pathways were

subsequently screened and the top targets were used to improve production of a single-domain antibody. Using proteomics, it was further shown that the repression of these targets does not result in a typical RpoS-mediated stationary phase response, even though growth is stalled.

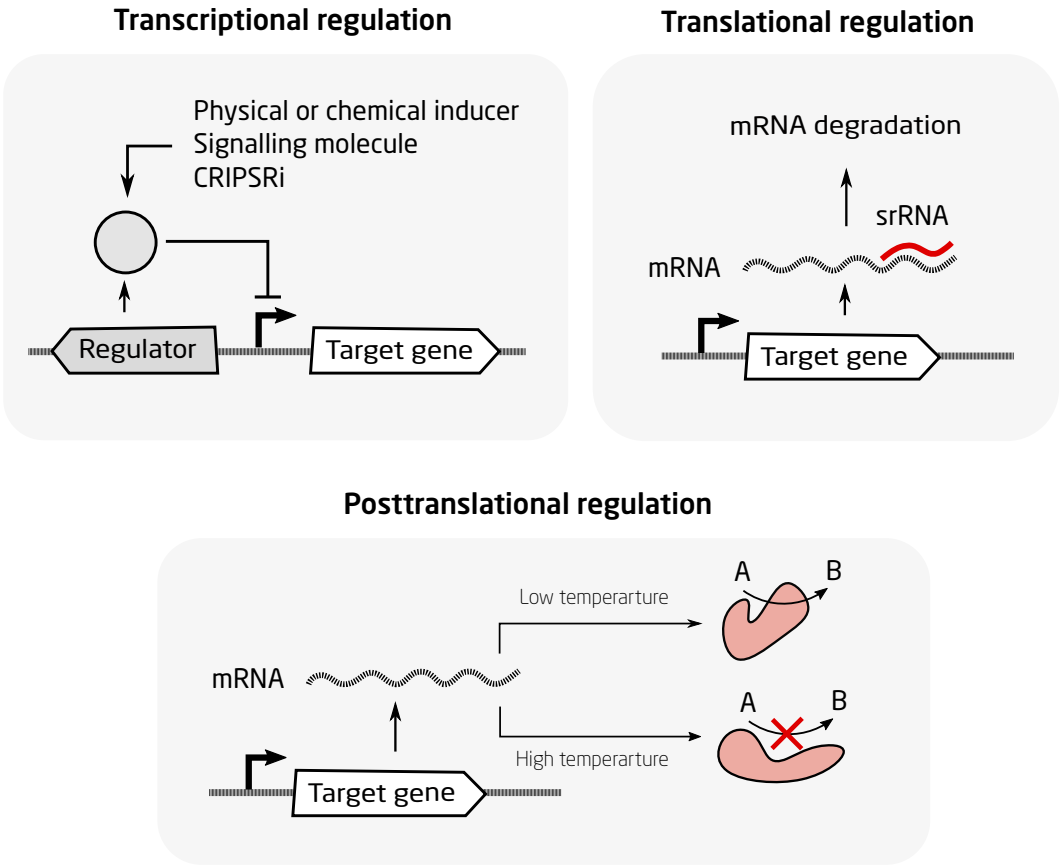


Figure 1.11. Inhibiting growth to decouple growth and production. Left upper panel: transcriptional regulation of target gene expression mediated by a regulator that is induced by an inducer or signaling molecule. Right upper panel: translational regulation of a target gene through srRNA-mediated degradation of target mRNA. Lower panel: posttranslational regulation through the use of a temperature-sensitive target enzyme. At lower temperatures, enzyme activity is normal. When the temperature is increase, the enzyme misfolds and is inactivated. Abbreviations: CRIPSRi; CRISPR interference, srRNA; small regulatory RNA, mRNA; messenger RNA.

On a translational level, growth inhibition can be engineered by targeting mRNAs or modifying ribosomal activity (Fig. 1.11). Synthetic small regulatory RNA and antisense RNA has for example been applied to improve production of cadaverine and pyruvate, respectively.[113,251] It is also possible to uncouple growth and production using synthetic, orthogonal ribosomes, from where production genes are translated.[252]

Growth inhibition based on posttranslational regulation can be carried out using inducible degradation or inactivation of key enzymes.[219] An interesting approach is the use of heat-sensitive enzymes (Fig. 1.11). These can be established by random mutation approaches, such as error-prone PCR, and selected in high-throughput screens based on their ability to slow down growth at higher temperatures.[84,253] At lower temperatures, the enzyme is fully functional, and cells can utilize the substrate and nutrients for biomass accumulation. When it is time to enter the production phase, the temperature is increased and the enzyme will be deactivated. Depending on what enzyme is targeted, this will inhibit biomass and/or metabolic flux through the enzyme pathway.[253] This strategy has for example been used to improve production of citrulline using a temperature sensitive argininosuccinate synthetase,[253] and increase 3-hydroxypropinoate titers using a temperature sensitive enoyl-[acyl carrier protein]-reductase.[254]

Overall, growth decoupling and dynamic balancing of growth and production has clearly contributed to the advancement of cell factory development. Although many challenges still need to be addressed, this approach has helped us overcome some of the obstacles and limitations with microbial production. Future research efforts in this area may help us take another step forward on the way toward creating economically feasible cell factories that can work as a sustainable replacement of fossil fuels and chemicals.

1.5 References for Chapter 1

1. Liu, L. *et al.* Fermented beverage and food storage in 13,000 y-old stone mortars at Raqefet Cave, Israel: Investigating Natufian ritual feasting. *J. Archaeol. Sci. Reports* **21**, 783–793 (2018).
2. McGovern, P. E. *et al.* Fermented beverages of pre- and proto-historic China. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 17593–17598 (2004).
3. McGovern, P. *et al.* Early Neolithic wine of Georgia in the South Caucasus. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E10309–E10318 (2017).
4. Pasteur, L. Mémoire sur la fermentation alcoolique. in *Annales de chimie et de physique* (ed. Masson, V.) **52**, 323–426 (1857).
5. Pasteur, L. Mémoire sur la fermentation appelée lactique. in *Annales de Chimie et de Physique* (ed. Masson, V.) **53**, 404–418 (1858).
6. EuropaBio. *Jobs and growth generated by industrial biotechnology in europe*. (2016).
7. National Research Council. *Industrialization of Biology: A Roadmap to Accelerate the Advanced Manufacturing of Chemicals Committee*. *The National Academies Press* (2015). doi:https://doi.org/10.17226/1900
8. Milken Institute & Financial Innovations Lab. *Unleashing the power of the bio-economy*. (2013).
9. Weizmann, C. & Hamlyn George, A. Fermentation process for the production of acetone and butyl alcohol. (1920).
10. Gaynes, R. The discovery of penicillin - new insights after more than 75 years of clinical use. *Emerg. Infect. Dis.* **23**, 849–853 (2017).
11. Dybdahl Hede, P. A beginner's guide to enzymes in detergents. *International News on Fats, Oils and Related Materials* (2014). Available at: https://engage.novozymes.com/insights/beginners-guide-enzymes-detergents. (Accessed: 28th April 2020)
12. Carlson, R. Estimating the biotech sector's contribution to the US economy. *Nat. Biotechnol.* **34**, 247–255 (2016).
13. Carlson, R. Bioeconomy Dashboard: Economic Metrics. *Bioeconomy capital* (2019). Available at: http://www.bioeconomycapital.com/bioeconomy-dashboard. (Accessed: 1st May 2020)
14. Carlson, R. Seeing The End Of Oil. *Bioeconomy Capital* (2019). Available at: http://www.bioeconomycapital.com/posts. (Accessed: 1st May 2020)
15. Sherpa Group & European Commission KET High Level Group. *KET – Industrial Biotechnology*. (2011).
16. Naik, S. N., Goud, V. V., Rout, P. K. & Dalai, A. K. Production of

first and second generation biofuels: A comprehensive review. *Renew. Sustain. Energy Rev.* **14**, 578–597 (2010).
17. Dürre, P. & Eikmanns, B. J. C1-carbon sources for chemical and fuel production by microbial gas fermentation. *Curr. Opin. Biotechnol.* **35**, 63–72 (2015).
18. E4tech (UK) Ltd, RE-CORD & SDLO Wageningen University and Resarch Centre. *From the Sugar Platform to biofuels and biochemicals*. *Final report for the European Commission Directorate-General Energy* (2015).
19. Rebello, S. *et al.* Non-conventional yeast cell factories for sustainable bioprocesses. *FEMS Microbiol. Lett.* **365**, 1–10 (2018).
20. Dale, B. E. & Linden, J. C. Fermentation Substrates and Economics. in *Annual Reports on Fermentation Processes* **7**, 107–134 (ACADEMIC PRESS, INC., 1984).
21. BIO-TIC. *The bioeconomy enabled - A roadmap to a thriving industrial biotechnology sector in Europe*. (2015).
22. Philp, J. & Winickoff, D. E. Clusters in Industrial Biotechnology and Bioeconomy: The Roles of the Public Sector. *Trends Biotechnol.* **35**, 682–686 (2017).
23. Davy, A. M., Kildegaard, H. F. & Andersen, M. R. Cell Factory Engineering. *Cell Syst.* **4**, 262–275 (2017).
24. Patterson, K., Yu, J., Landberg, J., Chang, I. & Shavarebi, F. Functional genomics for the oleaginous yeast Yarrowia lipolytica. *Metab. Eng.* **48**, 184–196 (2018).
25. Redl, S. *et al.* Thermodynamics and economic feasibility of acetone production from syngas using the thermophilic production host Moorella thermoacetica. *Biotechnol. Biofuels* **10**, 1–17 (2017).
26. Yan, Q. & Fong, S. S. Challenges and advances for genetic engineering of non-model bacteria and uses in consolidated bioprocessing. *Front. Microbiol.* **8**, 1–16 (2017).
27. Friedmann, H. C. Escherich and Escherichia. *EcoSal Plus* **6**, (2014).
28. Eckburg, P. B. *et al.* Microbiology: Diversity of the human intestinal microbial flora. *Science* **308**, 1635–1638 (2005).
29. Monk, J. M. *et al.* Multi-omics Quantification of Species Variation of Escherichia coli Links Molecular Features with Strain Phenotypes. *Cell Syst.* **3**, 238-251.e12 (2016).
30. Ratelade, J. *et al.* Production of recombinant proteins in the lon-deficient BL21(DE3) strain of Escherichia coli in the absence of the DnaK chaperone. *Appl. Environ. Microbiol.* **75**, 3803–3807 (2009).
31. Danchin, A. & Fang, G. Unknown unknowns: essential genes in quest for function. *Microb. Biotechnol.* **9**, 530–540 (2016).
32. Ferrer-Miralles, N. & Villaverde, A. Bacterial cell factories for

recombinant protein production; expanding the catalogue. *Microb. Cell Fact.* **12**, 2–5 (2013).
33. Goffeau, A. *et al.* Life with 6000 Genes. *Science* **274**, 546–567 (1996).
34. Mortimer, R. K. & Johnston, J. R. Genealogy of principal strains of the yeast genetic stock center. *Genetics* **113**, 35–43 (1986).
35. Mortimer, R. K. Evolution and variation of the yeast (Saccharomyces) genome. *Genome Res.* **10**, 403–409 (2000).
36. Liti, G. The fascinating and secret wild life of the budding yeast S. cerevisiae. *Elife* **4**, 1–9 (2015).
37. de Winde, J. *Functional Genetics of Industrial Yeasts*. *Functional Genetics of Industrial Yeasts* (Springer Berlin Heidelberg, 2003).
38. Engel, S. R. & Cherry, J. M. The new modern era of yeast genomics: Community sequencing and the resulting annotation of multiple Saccharomyces cerevisiae strains at the Saccharomyces Genome Database. *Database* **2013**, 1–9 (2013).
39. Daran-Lapujade, P. *et al.* Comparative genotyping of the Saccharomyces cerevisiae laboratory strains S288C and CEN.PK113-7D using oligonucleotide microarrays. *FEMS Yeast Res.* **4**, 259–269 (2003).
40. Hong, K. K. & Nielsen, J. Metabolic engineering of Saccharomyces cerevisiae: A key cell factory platform for future biorefineries. *Cell. Mol. Life Sci.* **69**, 2671–2690 (2012).
41. Kavšček, M., Stražar, M., Curk, T., Natter, K. & Petrovič, U. Yeast as a cell factory: Current state and perspectives. *Microb. Cell Fact.* **14**, 1–10 (2015).
42. Overton, T. W. Recombinant protein production in bacterial hosts. *Drug Discov. Today* **19**, 590–601 (2014).
43. Klein, J. & Nikolaidis, N. The descent of the antibody-based immune system by gradual evolution. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 169–174 (2005).
44. Vidarsson, G., Dekkers, G. & Rispens, T. IgG subclasses and allotypes: From structure to effector functions. *Front. Immunol.* **5**, 1–17 (2014).
45. Hamer-Casterman Atarchouch, T, C. *et al.* Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446–448 (1993).
46. Greenberg, A. S. *et al.* A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature* **374**, 168–173 (1995).
47. Muyldermans, S. Nanobodies: Natural Single-Domain Antibodies. *Annu. Rev. Biochem.* **82**, 775–797 (2013).
48. Sheridan, C. Llama-inspired antibody fragment approved for rare blood disorder. *Nature biotechnology* **37**, 333–334 (2019).

49. Arbabi-Ghahroudi, M. Camelid single-domain antibodies: Historical perspective and future outlook. *Front. Immunol.* **8**, 1–8 (2017).
50. Dooley, H. & Flajnik, M. F. Antibody repertoire development in cartilaginous fish. *Dev. Comp. Immunol.* **30**, 43–56 (2006).
51. Fernandes, C. F. C. *et al.* Camelid single-domain antibodies as an alternative to overcome challenges related to the prevention, detection, and control of neglected tropical diseases. *Front. Immunol.* **8**, 1–8 (2017).
52. Bactolife. 11 mio DKK grant for develpment of Ablacto+. (2020). Available at: http://www.bactolife.com/project/bactolife-grant-product-development-ablacto/. (Accessed: 18th May 2020)
53. Novobind. Targeted antimicrobials. (2018). Available at: https://www.novobind.com/. (Accessed: 18th May 2020)
54. Nielsen, J. & Keasling, J. D. Engineering Cellular Metabolism. *Cell* **164**, 1185–1197 (2016).
55. Nielsen, J. It Is All about Metabolic Fluxes. *J. Bacteriol.* **185**, 7031–7035 (2003).
56. Sabra, W., Dietz, D., Tjahjasari, D. & Zeng, A. P. Biosystems analysis and engineering of microbial consortia for industrial biotechnology. *Eng. Life Sci.* **10**, 407–421 (2010).
57. Markham, K. A. *et al.* Rewiring *Yarrowia lipolytica* toward triacetic acid lactone for materials generation. *Proc. Natl. Acad. Sci.* 201721203 (2018). doi:10.1073/pnas.1721203115
58. Yu, J. *et al.* Bioengineering triacetic acid lactone production in Yarrowia lipolytica for pogostone synthesis. *Biotechnol. Bioeng.* **115**, 2383–2388 (2018).
59. Awan, A. R., Shaw, W. M. & Ellis, T. Biosynthesis of therapeutic natural products using synthetic biology. *Adv. Drug Deliv. Rev.* **105**, 96–106 (2016).
60. Zhang, Y. *et al.* A Semi-Synthetic Organism that Stores and Retrieves Increased Genetic Information. *Nature* **551**, 644–647 (2017).
61. Werpy, T. & Petersen, G. *Top Value Added Chemicals from Biomass Volume I - Results of Screening for Potential Candidates from Sugars and Synthesis Gas*. *US Department of Energy Efficiency and Renewable Energy* (2004). doi:10.2172/15008859
62. Zhang, X., Xu, G., Shi, J., Koffas, M. A. G. & Xu, Z. Microbial Production of L-Serine from Renewable Feedstocks. *Trends in Biotechnology* **36**, 700–712 (2018).
63. Mundhada, H., Schneider, K., Christensen, H. B. & Nielsen, A. T. Engineering of high yield production of L-serine in Escherichia coli. *Biotechnol. Bioeng.* **113**, 807–816 (2016).
64. Mundhada, H. *et al.* Increased production of L-serine in Escherichia

coli through Adaptive Laboratory Evolution. *Metab. Eng.* **39**, 141–150 (2017).
65. Rennig, M. *et al.* Industrializing a Bacterial Strain for L-Serine Production through Translation Initiation Optimization. *ACS Synth. Biol.* **8**, 2347–2358 (2019).
66. Zhang, X. *et al.* High-yield production of L-serine through a novel identified exporter combined with synthetic pathway in Corynebacterium glutamicum. *Microb. Cell Fact.* **19**, 115 (2020).
67. Bailey, J. E. Toward a science of metabolic engineering. *Science* **252**, 1668–1675 (1991).
68. Stephanopoulos, G. & Vallino, J. J. Network rigidity and metabolic engineering in metabolite overproduction. *Science* **252**, 1675–1681 (1991).
69. Lee, H., DeLoache, W. C. & Dueber, J. E. Spatial organization of enzymes for metabolic engineering. *Metab. Eng.* **14**, 242–251 (2012).
70. Avalos, J. L., Fink, G. R. & Stephanopoulos, G. Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. *Nat. Biotechnol.* **31**, 335–341 (2013).
71. Ro, D. K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).
72. Paddon, C. J. *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**, 528–532 (2013).
73. Peplow, M. Synthetic biology's first malaria drug meets market resistance. *Nature* **530**, 389–390 (2016).
74. Clarke, L. & Kitney, R. Developing synthetic biology for industrial biotechnology applications. *Biochem. Soc. Trans.* **48**, 113–122 (2020).
75. BRENDA - Enzyme Database. (2020). Available at: https://www.brenda-enzymes.org/. (Accessed: 11th May 2020)
76. Kanehisa, M. & Goto, S. KEGG : Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
77. Petzold, C. J., Chan, L. J. G., Nhan, M. & Adams, P. D. Analytics for metabolic engineering. *Front. Bioeng. Biotechnol.* **3**, 1–11 (2015).
78. National Academy of Sciences Engineering and Medicine. *Preparing for Future Products of Biotechnology*. *Preparing for Future Products of Biotechnology* (2017). doi:10.17226/24605
79. iGEM. Registry of Standard Biological Parts. (2020). Available at: http://parts.igem.org/Main_Page. (Accessed: 12th May 2020)
80. Gu, C., Kim, G. B., Kim, W. J., Kim, H. U. & Lee, S. Y. Current status and applications of genome-scale metabolic models. *Genome Biol.* **20**, 1–18 (2019).
81. Lloyd, C. J. *et al.* COBRAme: A computational framework for genome-

scale models of metabolism and gene expression. *PLoS Comput. Biol.* **14**, 1–14 (2018).
82. Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010).
83. Brannigan, J. A. & Wilkinson, A. J. Protein engineering 20 years on. *Nature Reviews Molecular Cell Biology* **3**, 964–970 (2002).
84. Beuter, D. *et al.* Selective Enrichment of Slow-Growing Bacteria in a Metabolism- Wide CRISPRi Library with a TIMER Protein. *ACS Synth. Biol.* **7**, 2775–2782 (2018).
85. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, (2009).
86. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–824 (2013).
87. Gaudelli, N. M. *et al.* Programmable base editing of T to G C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
88. Jiang, Y. *et al.* Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **81**, 2506–2514 (2015).
89. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–7 (2014).
90. Chung, M. E. *et al.* Enhanced integration of large DNA into E. coli chromosome by CRISPR/Cas9. *Biotechnol. Bioeng.* **114**, 172–183 (2017).
91. Conrad, T. M., Lewis, N. E. & Palsson, B. O. Microbial laboratory evolution in the era of genome-scale science. *Molecular Systems Biology* **7**, (2011).
92. Morelli, L. *et al.* Injection molded lab-on-a-disc platform for screening of genetically modified: E. coli using liquid-liquid extraction and surface enhanced Raman scattering. *Lab Chip* **18**, 869–877 (2018).
93. Morelli, L. *et al.* Simultaneous quantification of multiple bacterial metabolites using surface-enhanced Raman scattering. *Analyst* **144**, 1600–1607 (2019).
94. Rennig, M. *et al.* TARSyn: Tunable Antibiotic Resistance Devices Enabling Bacterial Synthetic Evolution and Protein Production. *ACS Synth. Biol.* **7**, 432–442 (2018).
95. Zhang, J., Jensen, M. K. & Keasling, J. D. Development of biosensors and their application in metabolic engineering. *Curr. Opin. Chem. Biol.* **28**, 1–8 (2015).
96. Michener, J. K. & Smolke, C. D. High-throughput enzyme evolution in Saccharomyces cerevisiae using a synthetic RNA switch. *Metab. Eng.* **14**, 306–316 (2012).
97. Skjoedt, M. L. *et al.* Engineering prokaryotic transcriptional activators

as metabolite biosensors in yeast. *Nat. Chem. Biol.* **12**, 951–958 (2016).
98. Cameron, D. E., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. *Nat. Rev. Microbiol.* **12**, 381–390 (2014).
99. Calero, P. & Nikel, P. I. Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms. *Microb. Biotechnol.* **12**, 98–124 (2019).
100. Cumbers, J. Synthetic Biology Has Raised $ 12.4 Billion - Here Are Five Sectors It Will Soon Disrupt. *SynBioBeta* (2019). Available at: https://synbiobeta.com/synthetic-biology-has-raised-12-4-billion-here-are-five-sectors-it-will-soon-disrupt/. (Accessed: 18th May 2020)
101. Galdzicki, M. *et al.* The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nat. Biotechnol.* **32**, 545–550 (2014).
102. Sanders, R. NSF grant to launch world's first open - source genetic parts production facility. *UC Berkeley News Center* (2010). Available at: http://newscenter.berkeley.edu/2010/01/20/biofab_synthetic_biology/. (Accessed: 3rd June 2020)
103. Sainz de Murieta, I., Bultelle, M. & Kitney, R. I. Data model for biopart datasheets. *Eng. Biol.* **2**, 7–18 (2018).
104. Hillson, N. *et al.* Building a global alliance of biofoundries. *Nat. Commun.* **10**, 1038–1041 (2019).
105. Hirao, I. *et al.* An unnatural base pair for incorporating amino acid analogs into proteins. *Nat. Biotechnol.* **20**, 177–182 (2002).
106. Malyshev, D. A. *et al.* A Semi-Synthetic Organism with an Expanded Genetic Alphabet. **509**, 385–388 (2014).
107. Blazeck, J. & Alper, H. S. Promoter engineering: Recent advances in controlling transcription at the most fundamental level. *Biotechnol. J.* **8**, 46–58 (2013).
108. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **10**, 354–360 (2013).
109. Friedland, A. *et al.* Synthetic gene networks that count. *Science* **324**, 1199–1202 (2009).
110. Danino, T., Mondragón-Palomino, O., Tsimring, L. & Hasty, J. A synchronized quorum of genetic clocks. *Nature* **463**, 326–330 (2010).
111. Gupta, A., Reizman, I. M. B., Reisch, C. R. & Prather, K. L. J. Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. *Nat. Biotechnol.* **35**, 273–279 (2017).
112. Lo, T. M., Chng, S. H., Teo, W. S., Cho, H. S. & Chang, M. W. A Two-Layer Gene Circuit for Decoupling Cell Growth from Metabolite

Production. *Cell Syst.* **3**, 133–143 (2016).
113. Na, D. *et al.* Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. *Nat. Biotechnol.* **31**, 170–174 (2013).
114. Maung, N. W. & Smolke, C. D. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 14283–14288 (2007).
115. Chappell, J., Watters, K. E., Takahashi, M. K. & Lucks, J. B. A renaissance in RNA synthetic biology: New mechanisms, applications and tools for the future. *Curr. Opin. Chem. Biol.* **28**, 47–56 (2015).
116. Mak, W. S. *et al.* Integrative genomic mining for enzyme function to enable engineering of a non-natural biosynthetic pathway. *Nat. Commun.* **6**, 1–9 (2015).
117. Synthorx. Synthorx - Technology. Available at: https://synthorx.com/technology/. (Accessed: 9th June 2020)
118. Qian, Y., Huang, H. H., Jiménez, J. I. & Del Vecchio, D. Resource Competition Shapes the Response of Genetic Circuits. *ACS Synth. Biol.* **6**, 1263–1272 (2017).
119. Ceroni, F. *et al.* Burden-driven feedback control of gene expression. *Nat. Methods* **15**, 387–393 (2018).
120. Boo, A., Ellis, T. & Stan, G. B. Host-aware synthetic biology. *Curr. Opin. Syst. Biol.* **14**, 66–72 (2019).
121. Dymond, J. S. *et al.* Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471–476 (2011).
122. Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science* **351**, 1414 (2016).
123. Fredens, J. *et al.* Total synthesis of Escherichia coli with a recoded genome. *Nature* **569**, 514–518 (2019).
124. Trump, B. D. Synthetic biology regulation and governance: Lessons from TAPIC for the United States, European Union, and Singapore. *Health Policy (New. York).* **121**, 1139–1146 (2017).
125. Wang, F. & Zhang, W. Synthetic biology: Recent progress, biosafety and biosecurity concerns, and possible solutions. *J. Biosaf. Biosecurity* **1**, 22–30 (2019).
126. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182 (2005).
127. Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**, 653–663 (2005).
128. Bolotin, A., Quinquis, B., Sorokin, A. & Dusko Ehrlich, S. Clustered

regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).
129. Brouns, S. *et al.* Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes Stan. *Science* **321**, 960–965 (2013).
130. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
131. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–822 (2012).
132. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci.* **109**, 15539–15540 (2012).
133. Gao, C. *et al.* Programmable biomolecular switches for rewiring flux in Escherichia coli. *Nat. Commun.* **10**, 1–12 (2019).
134. Zhang, Y., Malzahn, A. A., Sretenovic, S. & Qi, Y. The emerging and uncultivated potential of CRISPR technology in plant science. *Nat. Plants* **5**, 778–794 (2019).
135. Hille, F. *et al.* The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**, 1239–1259 (2018).
136. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
137. Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).
138. Larson, M. H. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).
139. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
140. Hayes, C. S. & Low, D. A. Signals of growth regulation in bacteria. *Curr. Opin. Microbiol.* **12**, 667–673 (2009).
141. Monod, J. & Cohen-Bazire, G. L'effet d'inhibition spécifique dans la biosynthèse de la tryptophane-demase chez aerobacter aerogenes. *Comptes Rendus Hebd. Des Seances L Acad. Des Sci.* **236**, 530–532 (1953).
142. Yanofsky, C. *et al.* The complete nucleotide sequence of the tryptophan operon of Escherichia coli. *Nucleic Acids Res.* **9**, 6647–6668 (1981).
143. Yanofsky, C., Horn, V. & Gollnick, P. Physiological studies of

tryptophan transport and tryptophanase operon induction in Escherichia coli. *J. Bacteriol.* **173**, 6009–6017 (1991).
144. Lee, F. & Yanofsky, C. Transcription termination at the trp operon attenuators of Escherichia coli and Salmonella typhimurium: RNA secondary structure and regulation of termination. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 4365–4369 (1977).
145. Weickert, M. J. & Apostol, I. High-fidelity translation of recombinant human hemoglobin in Escherichia coli. *Appl. Environ. Microbiol.* **64**, 1589–1593 (1998).
146. Yanofsky, C. Attenuation in the control of expression of bacterial operons. *Nature* **289**, 751–758 (1981).
147. Martinussen, J., Willemoës, M. & Kilstrup, M. Nucleotide Metabolism. in *Comprehensive Biotechnology* (ed. Murray Moo-Young) **1**, 91–108 (Elsevier B.V., 2011).
148. Jensen, K. F., Dandanell, G., Hove-Jensen, B. & Willemoës, M. Nucleotides, Nucleosides, and Nucleobases. *EcoSal Plus* (2013). doi:10.1128/ecosalplus.3.6.2
149. Houlberg, U. & Jensen, K. F. Role of hypoxanthine and guanine in regulation of Salmonella typhimurium pur gene expression. *J. Bacteriol.* **153**, 837–845 (1983).
150. Han, X. & Turnbough, C. L. Regulation of carAB expression in Escherichia coli occurs in part through UTP-sensitive reiterative transcription. *J. Bacteriol.* **180**, 705–713 (1998).
151. Pierard, A. *et al.* Repression of Escherichia coli carbamoylphosphate synthase: relationships with enzyme synthesis in the arginine and pyrimidine pathways. *J. Bacteriol.* **127**, 291–301 (1976).
152. Turnbough, C. L., Hicks, K. L. & Donahue, J. P. Attenuation control of pyrBI operon expression in Escherichia coli K-12. *Proc. Natl. Acad. Sci.* **80**, 368–372 (1983).
153. Sørensen, K. I., Baker, K. E., Kelln, R. & Neuhard, J. Nucleotide Pool-Sensitive Selection of the Transcriptional Start Site In Vivo at the Salmonella typhimunium pyrC and pyrD Promoters. *J. Bacteriol.* **175**, 4137–4144 (1993).
154. Lange, R. & Hengge-Aronis, R. Identification of a central regulator of stationary-phase gene expression in Escherichia coli. *Mol. Microbiol.* **5**, 49–59 (1991).
155. Hengge, R. Stationary-Phase Gene Regulation in Escherichia coli. *EcoSal Plus* **4**, (2011).
156. Hengge, R. Proteolysis of σS (RpoS) and the general stress response in Escherichia coli. *Res. Microbiol.* **160**, 667–676 (2009).
157. Gentry, D. R., Hernandez, V. J., Nguyen, L. H., Jensen, D. B. &

Cashel, M. Synthesis of the stationary-phase sigma factor σ(s) is positively regulated by ppGpp. *J. Bacteriol.* **175**, 7982–7989 (1993).
158. Muffler, A., Fischer, D. & Hengge-Aronis, R. The RNA-binding protein HF-I, known as a host factor for phage Qβ RNA replication, is essential for rpoS translation in Escherichia coli. *Genes Dev.* **10**, 1143–1151 (1996).
159. Storz, G., Altuvia, S. & Wassarman, K. M. An Abundance of RNA Regulators. *Annu. Rev. Biochem.* **74**, 199–217 (2005).
160. Muffler, A., Fischer, D., Altuvia, S., Storz, G. & Hengge-Aronis, R. The response regulator RssB controls stability of the sigma(S) subunit of RNA polymerase in Escherichia coli. *EMBO J.* **15**, 1333–1339 (1996).
161. Battesti, A. *et al.* Anti-adaptors provide multiple modes for regulation of the RssB adaptor protein. *Genes Dev.* **27**, 2722–2735 (2013).
162. Weber, H., Polen, T., Heuveling, J., Wendisch, V. F. & Hengge, R. Genome-Wide Analysis of the General Stress Response Network in Escherichia coli: RpoS-Dependent Genes, Promoters, and Sigma Factor Selectivity. *J. Bacteriol.* **187**, 1591–1603 (2005).
163. Makinoshima, H. *et al.* Growth phase-coupled alterations in cell structure and function of Escherichia coli. *J. Bacteriol.* **185**, 1338–1345 (2003).
164. Miura, A., Krueger, J. H., Itoh, S., de Boer, H. A. & Nomura, M. Growth-rate-dependent regulation of ribosome synthesis in E. coli: Expression of the lacZ and galK genes fused to ribosomal promoters. *Cell* **25**, 773–782 (1981).
165. Reeve, C. A., Amy, P. S. & Matin, A. Role of protein synthesis in the survival of carbon-starved Escherichia coli K-12. *J. Bacteriol.* **160**, 1041–1046 (1984).
166. Brown, D. R., Barton, G., Pan, Z., Buck, M. & Wigneshweraraj, S. Nitrogen stress response and stringent response are coupled in Escherichia coli. *Nat. Commun.* **5**, 1–8 (2014).
167. Ferenci, T., Galbiati, H. F., Betteridge, T., Phan, K. & Spira, B. The constancy of global regulation across a species: The concentrations of ppGpp and RpoS are strain-specific in Escherichia coli. *BMC Microbiol.* **11**, (2011).
168. Jishage, M., Kvint, K., Shingler, V. & Nyström, T. Regulation of sigma factor competition by the alarmone ppGpp. *Genes Dev.* **70**, 1260–1270 (2002).
169. Costanzo, A. & Ades, S. E. Growth phase-dependent regulation of the extracytoplasmic stress factor, σE, by guanosine 3′,5′-bispyrophosphate (ppGpp). *J. Bacteriol.* **188**, 4627–4634 (2006).

170. Durfee, T., Hansen, A. M., Zhi, H., Blattner, F. R. & Ding, J. J. Transcription profiling of the stringent response in Escherichia coli. *J. Bacteriol.* **190**, 1084–1096 (2008).
171. Magnusson, L. U., Farewell, A. & Nyström, T. ppGpp: A global regulator in Escherichia coli. *Trends Microbiol.* **13**, 236–242 (2005).
172. Chubukov, V. & Sauer, U. Environmental dependence of stationary-phase metabolism in Bacillus subtilis and Escherichia coli. *Appl. Environ. Microbiol.* **80**, 2901–2909 (2014).
173. Tan, S. Z. & Prather, K. L. Dynamic pathway regulation: recent advances and methods of construction. *Curr. Opin. Chem. Biol.* **41**, 28–35 (2017).
174. Burg, J. M. *et al.* Large-scale bioprocess competitiveness: the potential of dynamic metabolic control in two-stage fermentations. *Curr. Opin. Chem. Eng.* **14**, 121–136 (2016).
175. Jacob, F., Perrin, D., Sánchez, C. & Monod, J. L'opéron: groupe de gènes à expression coordonnée par un opérateur. *C. R. Hebd. Seances Acad. Sci.* **250**, 1727–1729 (1960).
176. Ruegg, T. L. *et al.* Jungle Express is a versatile repressor system for tight transcriptional control. *Nat. Commun.* **9**, 1–13 (2018).
177. Holtz, W. J. & Keasling, J. D. Engineering Static and Dynamic Control of Synthetic Pathways. *Cell* **140**, 19–23 (2010).
178. Rohlhill, J., Sandoval, N. R. & Papoutsakis, E. T. Sort-Seq Approach to Engineering a Formaldehyde-Inducible Promoter for Dynamically Regulated Escherichia coli Growth on Methanol. *ACS Synth. Biol.* **6**, 1584–1595 (2017).
179. Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. Tuning genetic control through promoter engineering. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 12678–12683 (2005).
180. de Boer, H. A., Comstock, L. J. & Vasser, M. The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proc. Natl. Acad. Sci.* **80**, 21–25 (1983).
181. Chen, Y. *et al.* Tuning the dynamic range of bacterial promoters regulated by ligand-inducible transcription factors. *Nat. Commun.* **8**, (2018).
182. Brewster, R. C., Jones, D. L. & Phillips, R. Tuning Promoter Strength through RNA Polymerase Binding Site Design in Escherichia coli. *PLoS Comput. Biol.* **8**, (2012).
183. Brosius, J., Erfle, M. & Storella, J. Spacing of the -10 and -35 regions in the tac promoter. Effect on its in vivo activity. *J. Biol. Chem.* **260**, 3539–3541 (1985).
184. Solem, C. & Jensen, P. R. Modulation of gene expression made easy.

Appl. Environ. Microbiol. **68**, 2397–2403 (2002).
185. Khlebnikov, A. & Keasling, J. D. Effect of lacY expression on homogeneity of induction from the Ptac and Ptrc promoters by natural and synthetic inducers. *Biotechnol. Prog.* **18**, 672–674 (2002).
186. Khlebnikov, A., Skaug, T. & Keasling, J. D. Modulation of gene expression from the arabinose-inducible araBAD promoter. *J. Ind. Microbiol. Biotechnol.* **29**, 34–37 (2002).
187. Taylor, N. D. *et al.* Engineering an allosteric transcription factor to respond to new ligands. *Nat. Methods* **13**, 177–183 (2016).
188. Meinhardt, S. *et al.* Novel insights from hybrid LacI/GalR proteins: Family-wide functional attributes and biologically significant variation in transcription repression. *Nucleic Acids Res.* **40**, 11139–11154 (2012).
189. Gronenborn, B. Overproduction of phage Lambda repressor under control of the lac promotor of Escherichia coli. *MGG Mol. Gen. Genet.* **148**, 243–250 (1976).
190. Silverstone, A. E., Arditti, R. R. & Magasanik, B. Catabolite-insensitive revertants of lac promoter mutants. *Proc. Natl. Acad. Sci. U. S. A.* **66**, 773–779 (1970).
191. Hallewell, R. & Emtage, S. Plasmid vectors containing the tryptophan operon promoter suitable for efficient regulated expression of foreign genes. *Gene* **9**, 27–47 (1980).
192. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130 (1986).
193. Guzman, L.-M., Belin, D., Carson, M. J. & Beckwith, J. Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose PBAD Promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
194. Haldimann, A., Daniels, L. L. & Wanner, B. L. Use of new methods for construction of tightly regulated arabinose and rhamnose promoter fusions in studies of the Escherichia coli phosphate regulon. *J. Bacteriol.* **180**, 1277–1286 (1998).
195. Lee, S. K. & Keasling, J. D. A propionate-inducible expression system for enteric bacteria. *Appl. Environ. Microbiol.* **71**, 6856–6862 (2005).
196. Blatny, J., Brautaset, T., Winther-Larsen, H., Haugan, K. & Valla, S. Construction and use of a versatile set of broad-host-range cloning and expression vectors based on the RK2 replicon. *Appl. Environ. Microbiol.* **63**, 370–379 (1997).
197. Bakke, I. *et al.* Random mutagenesis of the Pm promoter as a powerful strategy for improvement of recombinant-gene expression. *Appl. Environ. Microbiol.* **75**, 2002–2011 (2009).

198. Elvin, C. M. *et al.* Modified bacteriophage lambda promoter vectors for overproduction of proteins in Escherichia coli. *Gene* **87**, 123–126 (1990).
199. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
200. Landberg, J., Mundhada, H. & Nielsen, A. T. An autoinducible trp-T7 expression system for production of proteins and biochemicals in Escherichia coli. *Biotechnol. Bioeng.* **117**, 1513–1524 (2020).
201. Chao, P., Law, S., Chen, P. T. & Hung, B. High production of heterologous proteins in Escherichia coli using the thermo-regulated T7 expression system. *Appl. Microbiol. Biotechnol.* **58**, 446–453 (2002).
202. Mckinney, J., Guerrier-Takada, C., Gala, J. & Altman, S. Tightly Regulated Gene Expression System in Salmonella enterica Serovar Typhimurium. *J. Bacteriol.* **184**, 6056–6059 (2002).
203. Wagner, S. *et al.* Tuning Escherichia coli for membrane protein overexpression. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 14371–14376 (2008).
204. Dahl, R. H. *et al.* Engineering dynamic pathway regulation using stress-response promoters. *Nat. Biotechnol.* **31**, 1039–1046 (2013).
205. Lara, A. R. *et al.* Characterization of Endogenous and Reduced Promoters for Oxygen-Limited Processes Using Escherichia coli. *ACS Synth. Biol.* **6**, 344–356 (2017).
206. Shin, P. K. & Seo, J.-H. Analysis of E. coli phoA-lacZ fusion gene expression inserted into a multicopy plasmid and host cell's chromosome. *Biotechnol. Bioeng.* **36**, 1097–1104 (1990).
207. Maury, J. *et al.* Glucose-dependent promoters for dynamic regulation of metabolic pathways. *Front. Bioeng. Biotechnol.* **6**, 1–12 (2018).
208. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
209. Waters, C. M. & Bassler, B. L. Quorum Sensing: Cell-to-Cell Communication in Bacteria. *Annu. Rev. Cell Dev. Biol.* **21**, 319–346 (2005).
210. Kim, E. M. *et al.* Autonomous control of metabolic state by a quorum sensing (QS)-mediated regulator for bisabolene production in engineered E. coli. *Metab. Eng.* **44**, 325–336 (2017).
211. Brockman, I. M. & Prather, K. L. J. Dynamic metabolic engineering: New strategies for developing responsive cell factories. *Biotechnol. J.* **10**, 1360–1369 (2015).
212. Lynch, M. D. Into new territory: improved microbial synthesis through engineering of the essential metabolic network. *Curr. Opin. Biotechnol.*

38, 106–111 (2016).
213. Zhang, F., Carothers, J. M. & Keasling, J. D. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat. Biotechnol.* **30**, 354–359 (2012).
214. Xu, P., Li, L., Zhang, F., Stephanopoulos, G. & Koffas, M. Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 11299–11304 (2014).
215. Farmer, W. R. & Liao, J. C. Improving lycopene production in Escherichia coli by engineering metabolic control. *Nat. Biotechnol.* **18**, 533–537 (2000).
216. Williams, T. C., Nielsen, L. K. & Vickers, C. E. Engineered Quorum Sensing Using Pheromone-Mediated Cell-to-Cell Communication in Saccharomyces cerevisiae. *ACS Synth. Biol.* **2**, 136–149 (2013).
217. Williams, T. C. *et al.* Quorum-sensing linked RNA interference for dynamic metabolic pathway control in Saccharomyces cerevisiae. *Metab. Eng.* **29**, 124–134 (2015).
218. Brockman, I. M. & Prather, K. L. J. Dynamic knockdown of E. coli central metabolism for redirecting fluxes of primary metabolites. *Metab. Eng.* **28**, 104–113 (2015).
219. Durante-Rodríguez, G., De Lorenzo, V. & Nikel, P. I. A Post-translational Metabolic Switch Enables Complete Decoupling of Bacterial Growth from Biopolymer Production in Engineered Escherichia coli. *ACS Synth. Biol.* **7**, 2686–2697 (2018).
220. Moser, F. *et al.* Dynamic control of endogenous metabolism with combinatorial logic circuits. *Mol. Syst. Biol.* **14**, 1–18 (2018).
221. Soma, Y., Tsuruno, K., Wada, M., Yokota, A. & Hanai, T. Metabolic flux redirection from a central metabolic pathway toward a synthetic pathway using a metabolic toggle switch. *Metab. Eng.* **23**, 175–184 (2014).
222. Lv, L., Ren, Y. L., Chen, J. C., Wu, Q. & Chen, G. Q. Application of CRISPRi for prokaryotic metabolic engineering involving multiple genes, a case study: Controllable P(3HB-co-4HB) biosynthesis. *Metab. Eng.* **29**, 160–168 (2015).
223. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41**, 7429–7437 (2013).
224. Lian, J., HamediRad, M., Hu, S. & Huimin, Z. orthogonal tri-functional CRISPR system. *Nat. Commun.* **8**, 1688 (2017).
225. Cameron, D. E. & Collins, J. J. Tunable protein degradation in bacteria. *Nat. Biotechnol.* **32**, 1276–1281 (2014).

226. Masuda, A., Toya, Y. & Shimizu, H. Metabolic impact of nutrient starvation in mevalonate-producing Escherichia coli. *Bioresour. Technol.* **245**, 1634–1640 (2017).
227. Tokuyama, K. *et al.* Magnesium starvation improves production of malonyl-CoA-derived metabolites in Escherichia coli. *Metab. Eng.* **52**, 215–223 (2019).
228. Willrodt, C., Hoschek, A., Bühler, B., Schmid, A. & Julsing, M. K. Decoupling production from growth by magnesium sulfate limitation boosts de novo limonene production. *Biotechnol. Bioeng.* **113**, 1305–1314 (2016).
229. Li, S., Jendresen, C. B. & Nielsen, A. T. Increasing production yield of tyrosine and mevalonate through inhibition of biomass formation. *Process Biochem.* **51**, 1992–2000 (2016).
230. Tian, K. *et al.* Limitation of thiamine pyrophosphate supply to growing Escherichia coli switches metabolism to efficient d-lactate formation. *Biotechnol. Bioeng.* **113**, 182–188 (2016).
231. Bellou, S., Triantaphyllidou, I. E., Mizerakis, P. & Aggelis, G. High lipid accumulation in Yarrowia lipolytica cultivated under double limitation of nitrogen and magnesium. *J. Biotechnol.* **234**, 116–126 (2016).
232. Liu, W., Zheng, P., Yu, F. & Yang, Q. A two-stage process for succinate production using genetically engineered Corynebacterium acetoacidophilum. *Process Biochem.* **50**, 1692–1700 (2015).
233. Ge, X. Y., Xu, Y., Chen, X. & Zhang, L. Y. Regulation of metabolic flux in Lactobacillus casei for lactic acid production by overexpressed ldhL gene with two-stage oxygen supply strategy. *J. Microbiol. Biotechnol.* **25**, 81–88 (2015).
234. Sun, Y. *et al.* Two-stage fermentation for 2-ketogluconic acid production by Klebsiella pneumoniae. *J. Microbiol. Biotechnol.* **24**, 781–787 (2014).
235. Chubukov, V. *et al.* Engineering glucose metabolism of Escherichia coli under nitrogen starvation. *npj Syst. Biol. Appl.* **3**, 16035 (2017).
236. Michalowski, A., Siemann-Herzberg, M. & Takors, R. Escherichia coli HGT: Engineered for high glucose throughput even under slowly growing or resting conditions. *Metab. Eng.* **40**, 93–103 (2017).
237. Koebmann, B. J., Westerhoff, H. V, Snoep, J. L., Nilsson, D. & Jensen, P. R. The Glycolytic Flux in Escherichia coli Is Controlled by the Demand for ATP Brian. *J. Bacteriol.* **184**, 3909–3916 (2002).
238. Registry of Standard Biological Parts. Catalog of constitutive promoters. Available at: http://parts.igem.org/Promoters/Catalog/Constitutive. (Accessed:

8th July 2020)
239. Izard, J. *et al.* A synthetic growth switch based on controlled expression of RNA polymerase. *Mol. Syst. Biol.* **11**, 840 (2015).
240. Lemmerer, M. *et al.* Decoupling of recombinant protein production from Escherichia coli cell growth enhances functional expression of plant Leloir glycosyltransferases. *Biotechnol. Bioeng.* **116**, 1259–1268 (2019).
241. Stargardt, P., Feuchtenhofer, L., Cserjan-Puschmann, M., Striedner, G. & Mairhofer, J. Bacteriophage Inspired Growth-Decoupled Recombinant Protein Production in Escherichia coli. *ACS Synth. Biol.* (2020). doi:10.1021/acssynbio.0c00028
242. Bothfeld, W., Kapov, G. & Tyo, K. E. J. A Glucose-Sensing Toggle Switch for Autonomous, High Productivity Genetic Control. *ACS Synth. Biol.* **6**, 1296–1304 (2017).
243. Venayak, N., Raj, K., Jaydeep, R. & Mahadevan, R. An Optimized Bistable Metabolic Switch to Decouple Phenotypic States during Anaerobic Fermentation. *ACS Synth. Biol.* (2018). doi:10.1021/acssynbio.8b00284
244. Santala, S., Efimova, E. & Santala, V. Dynamic decoupling of biomass and wax ester biosynthesis in Acinetobacter baylyi by an autonomously regulated switch. *Metab. Eng. Commun.* **7**, (2018).
245. Harder, B. J., Bettenbrock, K. & Klamt, S. Temperature-dependent dynamic control of the TCA cycle increases volumetric productivity of itaconic acid production by Escherichia coli. *Biotechnol. Bioeng.* **115**, 156–164 (2018).
246. Shabestary, K. *et al.* Targeted Repression of Essential Genes To Arrest Growth and Increase Carbon Partitioning and Biofuel Titers in Cyanobacteria. *ACS Synth. Biol.* **7**, 1669–1675 (2018).
247. Wu, J., Du, G., Chen, J. & Zhou, J. Enhancing flavonoid production by systematically tuning the central metabolic pathways based on a CRISPR interference system in Escherichia coli. *Sci. Rep.* **5**, 1–14 (2015).
248. Li, S. *et al.* Enhanced protein and biochemical production using CRISPRi-based growth switches. *Metab. Eng.* **38**, 274–284 (2016).
249. Li, S. *et al.* Methods for decoupling microbial cell growth from production of biochemicals and recombinant proteins. *PCT Int. Appl.* 202pp. (2018).
250. Li, S. *et al.* Genome-wide CRISPRi-based identification of targets for decoupling growth from production. *ACS Synth. Biol.* **9**, (2020).
251. Nakashima, N., Tamura, T. & Good, L. Paired termini stabilize antisense RNAs and enhance conditional gene silencing in Escherichia

coli. *Nucleic Acids Res.* **34**, e138–e138 (2006).
252. Darlington, A. P. S., Kim, J., Jiménez, J. I. & Bates, D. G. Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes. *Nat. Commun.* **9**, (2018).
253. Schramm, T. *et al.* High-throughput enrichment of temperature-sensitive argininosuccinate synthetase for two-stage citrulline production in E. coli. *Metab. Eng.* **60**, 14–24 (2020).
254. Lynch, M. D., Gill, R. T. & Lipscomb, T. E. Method for producing 3-hydroxypropionic acid and other products. 179 (2014).

Chapter 2

GENOME-WIDE CRISPRI-BASED IDENTIFICATION OF TARGETS FOR DECOUPLING GROWTH AND PRODUCTION

Songyuan Li[§†], Christian Bille Jendresen[§†‡], Jenny Landberg[§†], Lasse Ebdrup Pedersen[†], Nikolaus Sonnenschein[†], Sheila Ingemann Jensen[†], Alex Toftgaard Nielsen[†‡*]

[†] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet 220, 2800 Kongens Lyngby, Denmark.
[‡] CysBio ApS, Agern Allé 1, 2970 Hørsholm, Denmark.

[§] These authors contributed equally to the study

***Corresponding Author**
Alex Toftgaard Nielsen
The Novo Nordisk Foundation Center for Biosustainability
Technical University of Denmark
Kemitorvet, DK-2800 Kongens Lyngby
Denmark
Email: atn@biosustain.dtu.dk

Abstract

Growth decoupling can be used to optimize microbial production of biobased compounds by inhibiting excess biomass formation and redirect carbon flux from growth to product formation. However, identifying suitable genetic targets through rational design is challenging. Here, we conduct a genome-wide CRISPRi screen to discover growth switches suitable for decoupling growth and production. Using an sgRNA library covering 12,238 loci in the *Escherichia coli* genome, we screen for targets that inhibit growth while allowing for continued protein production. In total, we identify 1332 sgRNAs that simultaneously decrease growth and maintain or increase accumulation of GFP. The top target *sibB/ibsB* shows more than 5-fold increase in GFP accumulation and 45% decrease in biomass formation. Overall, our genome-wide CRISPRi screen provides key targets for growth decoupling, and the approach can be applied to improve biobased production in other microorganisms.

2.1 Introduction

Methods for inhibiting bacterial growth are of interest for the biotechnology industry as well as for the identification of novel antibiotic targets. Bacteria, such as *Escherichia coli*, are widely used as cell factories to produce biochemicals or proteins.[1] One of the main challenges during production of such compounds is to achieve a high product yield, i.e. conversion of substrate to product, since substrate is one of the main cost contributors of a fermentation process. However, as bacteria have evolved to grow exponentially when environmental conditions allow, cell factories typically exhibit overgrowth during fermentation. During production of biofuels and biochemicals, a high cell density results in increased cost of aeration and a decreased product yield as the supplied carbon source is used to accumulate biomass instead of the desired product. A low concentration of product compared to biomass furthermore results in increased cost of downstream processing. An ideal production scenario during a fed-batch fermentation would consist of two phases; one where biomass is quickly accumulated, followed by one where the production organism primarily accumulates product. A general strategy for improving the performance of cell factories is to apply a growth switch that controls cell growth while high production of proteins or biochemicals is maintained.[2–6]

Nutrient limitation is an option to achieve a biphasic growth-production scenario, where growth will be limited when a selected nutrient is exhausted. This, however, may elicit unfavorable cellular responses, such as stringent response, that may result in decreased production. Alternatively, addition of an inhibitor of a cellular process may be employed,[7] and similarly, a complete deletion of a gene involved in *de novo* synthesis of cell building blocks may be employed in a similar run-out strategy. Different approaches can be employed to identify suitable targets for growth inhibition. A common approach is based on gene knock-out, where genes essential for cell growth can be identified by the failure of constructing knock-out mutants or by conditional lethality with or without a certain compound. Genome-wide single gene deletion study performed in *E. coli* K12 identified 303 genes as essential even in rich media[8], while non-essential genes also may have

reduced fitness.[9] Temporary silencing of genes may be an attractive alternative to gene deletions, and small RNAs, including *cis*-encoded antisense RNA (asRNA) and *trans*-encoded Hfq-dependent small RNA, have for example been engineered and applied for gene inhibition studies.[10,11] Genome-wide screening focused on identifying targets involved in drug sensitivity has previously been performed using asRNAs.[12] Using CRISPR interference (CRISPRi),[13] which employs a catalytically inactive Cas9 (dCas9), transcription of a coding sequence targeted by the single guide RNA (sgRNA) can be prevented.[13] As opposed to inhibition based on small RNAs, the CRISPRi system inhibits transcription instead of translation, which is expected to provide a more robust inhibition of genes with varying mRNA concentrations, without the cost of producing excess mRNA. Expression of the system can be tightly controlled and has been shown to be highly efficient and specific when applied in bacteria,[14] and it is therefore an excellent tool for investigating gene targets relevant to cell growth. Targeting CRISPRi toward different regions only requires modification of the 20 nt sgRNA target sequence. This enables a simple approach for genome-wide studies, especially as the emergence of on-chip DNA synthesis and next generation sequencing has reduced the cost of library synthesis and analysis.[15,16] Several studies have already been carried out using CRISPR or CRISPR-derived approaches for genome-scale screening to identify essential genes or genes associated with drug sensitivity.[17–20] The first studies were conducted in mammalian cells, however, recent studies have employed the system to identify essential genes in *Bacillus subtillis*[21], *Streptococcus pneumoniae*[22] and *E. coli*,[23,24] and to investigate genome-scale functional genomics in *E. coli*[25] and *Vibrio natrigenens*[26]. Recently, a modular suite of CRISPRi systems applicable for studying gene essentiality and functional genomics in non-model bacteria was developed.[27] Additionally, genome-wide CRISPRi has been applied in combination with single-cell time-lapse imaging in order to determine how CRISPRi gene knock-down impact cell cycle coordination in *E. coli*.[28]

Here we take advantage of CRISPRi for temporal control and ability to multiplex identification of growth-determining genes. As growth inhibition may be detrimental to production of a protein of interest, we additionally identified specific targets that increase the production of a protein, exemplified by GFP. A library of 12,238 sgRNAs designed to target open

reading frames and non-coding regions was designed and used for the screening experiment, which resulted in the identification of a range of sgRNAs that can be applied as efficient growth switches. Two of the selected sgRNAs, *sibB/ibsB* and *yheV*, displayed more than 5-fold increase in GFP production. The genome-wide CRISPRi-approach will be of interest for the biotech industry, as it can be used to find genomic loci that, when repressed, increase the product yield of a desired compound. Further evaluation of targets that display a strong growth-inhibiting phenotype may additionally be of interest for development of novel antimicrobial agents.

2.2 Materials and methods

Strains, plasmids and media

Lysogeny broth (LB, 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) medium and LB agar plates with appropriate antibiotics were used for cultivation and screening during cloning. The CRISPRi library screening experiments were performed in M9 minimal medium with glucose and yeast extract as indicated in the text, and with appropriate antibiotics. Carbenicillin, ampicillin, and chloramphenicol were used with working concentrations of 100 µg/mL, 100 µg/mL, and 25 µg/mL, respectively. *E. coli* NEB 5-alpha from New England Biolabs was used as a cloning strain. *E. coli* Sij17 was used as the parental fluorescent strain for the CRISPRi library screening. The strain was obtained by integration of a GFP expression cassette and a kanamycin marker, 9 basepairs downstream of *glmS* in *E. coli* MG1655, as previously described.[29] The CRISPRi system was expressed from plasmids pdCas9-bacteria and pSLQ1236.[14] Upon induction with anhydrotetracycline (aTc), the dCas9 was expressed from pdCas9-bacteria and the sgRNA was expressed from pSLQ1236, activating the CRISPRi system. Strains and plasmids are listed in Supplementary Table S2.10.

Design and preparation of the CRISPRi library

A total of 12,238 sgRNAs were designed to target locations across the genome, with two sgRNAs for each gene coding sequence and 3497 sgRNAs distributed evenly in the non-coding regions. All designed target sequences

are listed in Supplementary Tables S2.1 (genes) and S2.2 (intergenic sequences). sgRNAs targeting gene coding sequences were designed to bind the non-template strand near the start codon region. CRISPy++ was used to estimate off-target efficiency of each sgRNA, and sgRNAs with low off-target efficiency (scores < 5000) were preferred.[30]

Designed oligonucleotides (CTCCCTATCAGTGATAGAGAAAAGAC TAGT-N_{20}-GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT AGTCCGTTATCAACTTGAAAAAG) were ordered as pooled libraries (CustomArray Inc), where N_{20} indicated the target sequence. The library was amplified using primers SON172 and SON173 (Supplementary Table S2.9), and was inserted into plasmid pSLQ1236 (obtained as a gift from Professor Stanley Qi, Stanford University), which was amplified using primers SON178 and SON179 (Supplementary Table S2.9). The two products were assembled using NEBuilder® HiFi DNA Assembly Cloning Kit (New England Biolabs) before transformation into and NEB 5-alpha Electrocompetent *E. coli* cells (New England Biolabs). In order to ensure the representation of the entire library, colonies corresponding to 150-fold library coverage were obtained after transformation of the library into the cloning strain. The pooled library was grown overnight and plasmids were prepared using GeneJET Plasmid Miniprep Kit (Thermo Scientific). Fluorescent strain Sij17/pdCas9-bacteria was transformed with the pooled plasmid sgRNA library. Colonies corresponding to a 60-fold coverage of the library size were obtained to ensure library representation. These transformants were pooled, grown up overnight, and aliquoted for glycerol stocks as the original CRISPRi cell library. Plasmids of the original library were extracted from a 5 mL overnight culture and submitted for next generation sequencing.

Growth and fluorescence sorting experiments

Targets affecting cell growth were identified by comparing the presence of sgRNAs in induced and uninduced cultures. An aliquot of the original CRISPRi cell library was inoculated in M9 media supplemented with 0.5% (w/v) glucose and 0.02% (w/v) yeast extract (M9G0.5YE), and cultivated overnight at 37 °C and 250 rpm. The pre-cultures were diluted 100 times into fresh M9 medium with 0.5% (w/v) glucose (M9G0.5). Six 20 mL parallel

cultures were prepared in baffled shake flasks, and cultivated at 37 °C and 250 rpm. Half of the cultures were induced with 200 ng/mL aTc after 1 h. Twenty four hours after inoculation, a 5 mL sample was taken from each culture and used to extract plasmids that were subsequently submitted for next generation sequencing. Targets that improved protein production were identified by analyzing the enrichment of sgRNAs in libraries of cells with high fluorescence. Three induced cultures were analyzed and sorted using a FACS Aria (Becton Dickinson, San Jose, USA). Forward-scatter and side-scatter were detected as small- and large-angle scatters of the 488 nm laser, respectively. GFP fluorescence was detected with a 488 nm long-pass and a 530/30 nm band-pass filter set. The top 1% of cells with the highest fluorescence (FITC) were sorted and collected. The threshold for fluorescence intensity used in this sorting was applied for the following rounds of sorting. In total, 33,000 cells were collected from each of the three cultures. The three sorted populations were recovered in 1 mL SOC media for 2 h, and subsequently grown in M9G0.5YE overnight. These overnight cultures were used as pre-cultures for the next sorting round. Three rounds of sorting were performed using the same settings, with the exception that 50,000 cells were collected in the last sorting. For each round of sorting, plasmids were extracted from 5 mL of medium from each culture and submitted for next generation sequencing.

Next generation sequencing

The target regions of prepared plasmid extracts were amplified by two rounds of PCR and used for next generation sequencing. For the first round, 40 ng of DNA was added to a 20 µL standard Phusion Hot Start II DNA Polymerase (Thermo Fisher Scientific) reaction mix. The DNA was amplified with primers SON233 and SON234. The PCR program used was 98 °C for 5 min followed by 25 cycles of 98 °C for 30 s, 65 °C for 30 s, and 72 °C for 30 s, with a final elongation step at 72 °C for 7 min. Each PCR product was purified using AMPure XP beads (Beckman Coulter, CA), and diluted in 50 µL Tris-buffer (pH 8.5). For the second PCR round, 5 µL purified product was added to a 20 µL standard Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific) reaction mix and amplified with Nextera XT Index primers (Illumina no. FC-131-1001). The PCR program used was 98 °C for 5 min followed by 25 cycles of 98 °C for 30 s, 65 °C for

30 s, and 72 °C for 30 s, with a final elongation step at 72 °C for 7 min. Each PCR product was again purified using AMPure XP beads (Beckman Coulter, CA). Each prepared sample was verified and quantified using a 2100 Bioanalyzer (Agilent) and a Qubit 2.0 Fluorometer (Thermo Fisher Scientific), diluted to an appropriate concentration, and analyzed by next generation sequencing. Sequencing was performed on a NextSeq 500 desktop sequencer (Illumina, San Diego, CA) using 75 bp single read.

Data analysis

Counts of sgRNAs for each sample were extracted from sequencing files using python codes. The Tag Count Comparison (TCC) method was used to normalize and analyze sgRNA frequency for the growth screening and determination of significantly enriched genes.[31] A growth score was calculated for each sgRNAs as $\log_2$(mean frequency of $sgRNAs_{induced\ cultures}$/mean frequency of $sgRNAs_{uninduced\ cultures}$). Counts obtained from the GFP enrichment experiment were normalized by the total counts of all sgRNAs. The GFP score for each sgRNA was calculated as the mean value of normalized counts in three parallel samples.

Gene set enrichment analysis and data visualization

For each gene, the sgRNA with the lowest growth score was selected for GSEA and biological pathway data visualization. Intergenic targets were discarded. GSEA was performed using GSEAPY, a python wrapper for GSEA and Enrichr developed by Subramanian et al.[32,33] It is used to calculate an enrichment score (ES) of the experimental gene set for each term in the given reference pathway gene set. ES is calculated on the basis of the ranking in the experimental gene set, and represents the degree of overrepresentation of a specific term at the top or bottom of the data set (thus, ES can be positive or negative). To adjust for multiple hypothesis testing, a normalized ES (NES) and a corresponding false discovery rate (FDR) is calculated for each term. In this study, significantly enriched metabolic pathways (p-value < 0.05) were identified with the GSEAPY prerank module by submitting the list with growth inhibiting (growth score < 1) gene targets and their corresponding growth scores together with either a BIGG, EcoCyc or KEGG reference pathway gene sets (Supplementary data).

To visualize the growth and fluorescence data from the library screen, the data sets were applied in Escher, a tool for building and visualizing metabolic pathways.[34] For mapping of growth score data, reactions catalyzed by more than one functional homologue and reactions catalyzed by a protein consisting of several subunits were represented by the homologue or subunit with the lowest growth score for any sgRNA. For mapping of enrichment in the fluorescence sorting, the fluorescence score of the same sgRNA target with the lowest growth score was selected to represent the reaction on the metabolic map.

Confirmation of targets

Top targets were selected according to the analysis of the deep sequencing results. Candidate sgRNAs were assembled into plasmid pSLQ1236 using primers SON203-SON232 (Supplementary Table S2.9). A control plasmid without sgRNA expression was assembled using pSLQ1236 as the template with primers SON176 and SON177 (Supplementary Table S2.9) followed by USER assembly of the overlapping ends. Standard reagents and methods described above were used for this assembly. Obtained plasmids were transformed into strain Sij17 with pdCas9-bacteria separately. For each tested strain, an end-point experiment and a growth profiling were experiment was performed to characterize the effect of selected targets on cell growth and GFP production. Cell cultures were prepared as follows. Biological triplicates were inoculated in M9G0.5YE and grown overnight at 37 °C and 250 rpm in an orbital shaker. The pre-cultures were diluted 100 times in fresh M9G0.5 medium; for each strain, six cultures were prepared, and half of them were induced 1 h after inoculation. For the growth profiling experiment, cell cultures were prepared and 150 µL was transferred into 96-well microtiter plates. Plates were cultivated in an ELx808 plate reader (BioTek, USA) at 37 °C with medium shaking. The optical density (OD) of each culture was measured at 630 nm every 10 min for 37 h. For fluorescence and end-point OD measurements, cell cultures were prepared, and 800 µL was transferred into 96-well deep-well plates and cultivated at 37°C and 300 rpm. After 24 h, the cultures were sampled and diluted appropriately, and the OD and fluorescence were measured using a Synergy Mx plate reader (BioTek, USA). The OD was measured at 630 nm and the GFP fluorescence was measured using an excitation at 485 nm and emission at 535 nm with a

gain set to 80. Samples were also taken for flow cytometry analysis, which was carried out using the same settings as described above.

2.3 Results

Design and construction of the CRISPRi library

Bacteria have evolved complicated regulatory networks for coordinating growth with other cellular activities. Our understanding of these networks is currently not sufficient for rationally constructing an ideal growth switch, which can be used to block growth while allowing for continued production of biochemicals and proteins. Therefore, we designed a genome wide library of sgRNAs to search the entire *E. coli* genome for suitable targets (Fig. 2.1). A total of 12,238 computationally optimized sgRNAs were designed to target locations across the genome, with two sgRNA targets for each gene coding sequence (Supplementary Table S2.1) and 3497 sgRNAs distributed evenly in non-coding regions (Supplementary Table S2.2). A total of 8974 sgRNAs were designed to target 4441 coding sequences, as some genes are present in multiple loci in the genome and, for a few short coding sequences, only one appropriate sgRNA target sequence could be found (Supplementary Table S2.1). Notably, 255 sgRNAs were mistakenly duplicated, resulting in a library with a total of 12,217 unique sgRNA sequences. sgRNAs targeting gene coding sequences were designed to bind the non-template strand near the start codon region, as this has previously been shown to be the most efficient location for inhibition of the expression of genes using CRISPRi.[13,35,36] Crispy++ was used to score sgRNA off-target tendency and in cases with multiple sgRNAs in the desired region of the gene, the sgRNA with the lowest off-target score was picked.[30] A pool of oligonucleotides embodying all targets were synthesized and cloned into pSLQ1236, a plasmid used for sgRNA expression.[14] In order to prepare the cell library, the cloned library of sgRNAs and the dCas9-expressing plasmid pdCas9-bacteria was transformed into an *E. coli* strain constitutively expressing a single-copy integrated GFP under control of constitutive promoter J23100 (strain Sij17). The GFP was integrated 9 bp downstream of *glmS*. Expression of dCas9 and sgRNA was tightly controlled using a tetracycline inducible

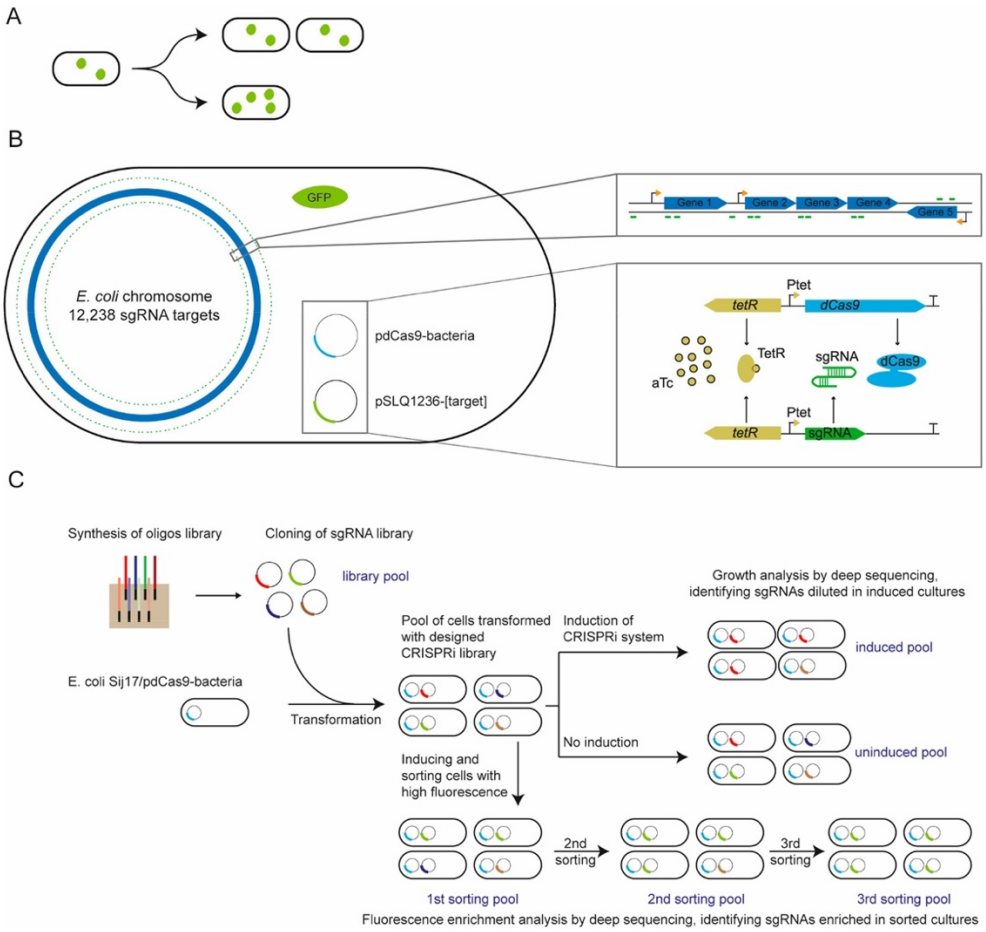


Figure 2.1. Overview of design and construction of the sgRNA library for the genome-wide functional screening. (a) The aim is to identify mutants that sacrifice growth for increased production of a target of interest, here green fluorescent protein (GFP). (b) A library of sgRNAs were designed to target genes or intergenic regions across the genome of an E. coli cell that expresses GFP under a constitutive promoter. The CRISPRi system is encoded on two plasmids: dCas9 and the library of sgRNAs on another. The system was activated by addition of aTc, inducing the expression of both dCas9 and sgRNA that repress the gene or intergenic target by blockage. (b) Overview of the construction of the sgRNA library and two screening experiments that combined enable the identification of genetic switches that repress growth while allowing for continued production of protein.

promoter. The growth switch was induced by addition of 200 ng/µL anhydrotetracycline (aTc) (Fig. 2.1b). As determined by deep sequencing

analysis, at least 99.91% of the designed sgRNAs were present in the constructed library and there was only an approximately 3-fold difference in frequency between the 90 percentile and the 10 percentile (Fig. 2.2a).

Genome-wide functional screening for growth repression and protein enrichment

The frequencies of sgRNAs that effectively repressed growth were assumed to decrease in the cell library during growth when the CRISPRi system had been activated. A growth screen, in which we compared the frequencies of sgRNAs in the induced cell library with those in the uninduced library, was therefore performed in order to identify promising targets for repressing growth (Fig. 2.1c).

The distribution of frequencies of sgRNAs in the library was dramatically changed in induced compared with uninduced cultures (Fig. 2.2b). The distribution histogram flattened out for the induced culture, with more sgRNAs with frequencies above and below average (Fig. 2.2b). Of the designed sgRNAs covered in the growth analysis, 71.5% (8736) were changed (false discovery rate < 0.01), of which 44.4% were diluted and 27.1% were enriched (Fig. 2.2b, Supplementary Table S2.3). The enrichment may be due to faster growth among the 27.1% sgRNAs, slower growth of other targets, or, most likely, a combination of both. Confidence for the enrichment scores can be seen in Supplementary Table S2.4, and the reads mapped per sample and per sgRNA per sample can be seen in Supplementary Fig. S2.2a and S2.2b, respectively.

To test the validity of the growth score, we examined genes previously shown to be essential in M9 media[37] and found that for 87.6% of the genes at least one corresponding sgRNA gave a significant repression of growth (Fig. 2.2c, means and standard deviations in Supplementary Table S2.5). Also including genes found to be essential in rich media[8], 89.4% had a sgRNAs resulting in significant repression of growth (Supplementary Table S2.6). Analysis of all sgRNA pairs targeting the same gene showed that, of the sgRNAs being enriched or depleted, 64% of sgRNA pairs had the same effect on growth, while 36% were oppositely affecting growth. The opposite effect

may be due to differences in sgRNA efficiency,[38] or due to mutations in the sgRNA or dCas9.

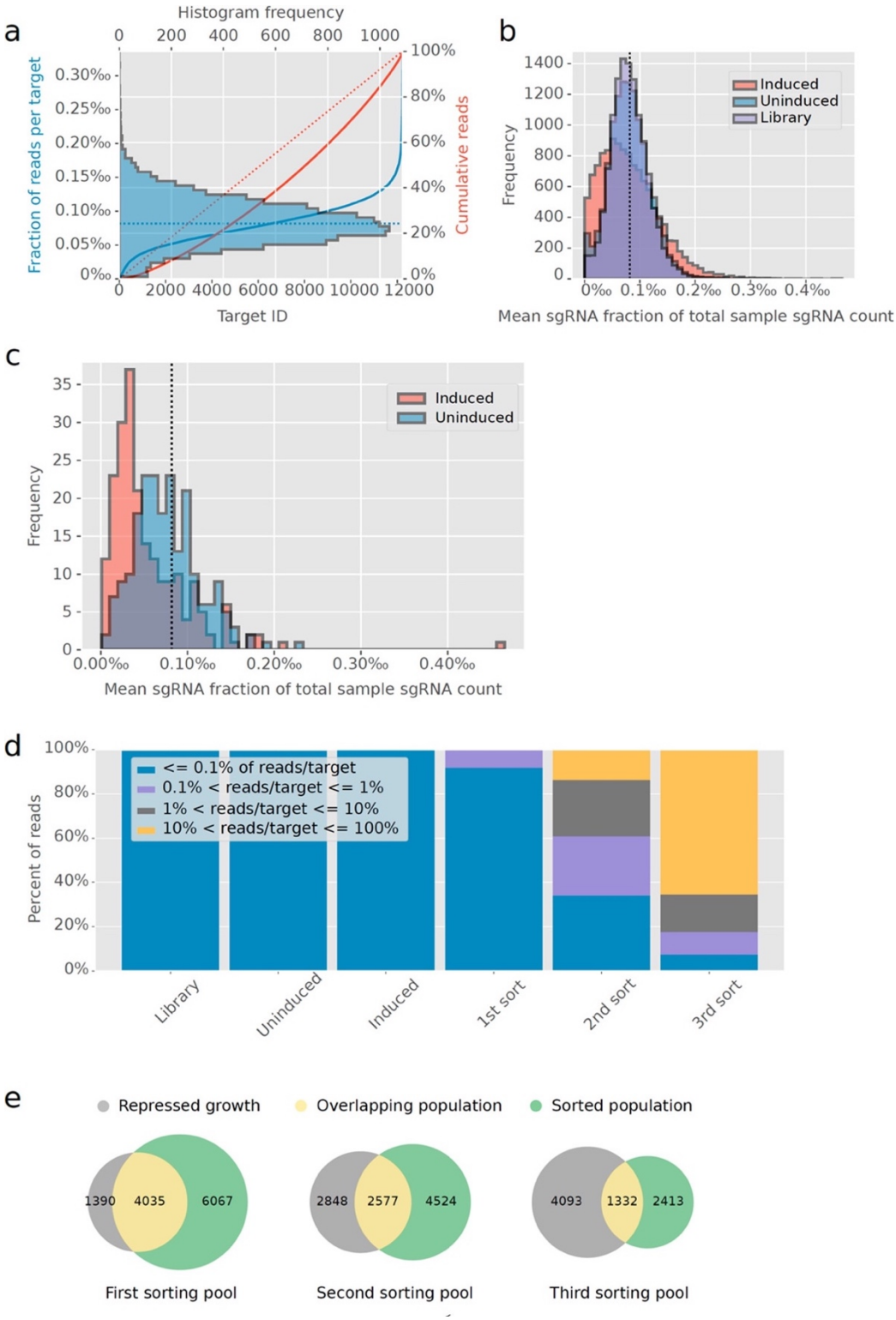


Figure 2.2. Genome-wide functional screening for growth and protein enrichment. (a) Analysis of the constructed sgRNA library. sgRNAs were ranked by the number of reads from minimum to maximum across the x-axis, and the fraction of reads of each sgRNA (per mille) in the library is shown as the blue solid line. The red solid line shows the cumulative number of reads. The dashed lines show the ideal situation if each sgRNA occupy an even amount of reads in the library. The histogram shows the distribution of sgRNA reads (bin-width 0.0067 per mille). (b) Distribution of reads of sgRNAs in the library. Violet represents the distribution of the original library. The distribution of the cell library with and without induction is shown in red and blue, respectively (bin-width 0.0093 per mille). (c) Distribution of read frequencies of sgRNAs with essential gene targets is shown with and without induction in red and blue, respectively (bin-width 0.0101 per mille). (d) Change in distribution of read frequencies of sgRNAs during the genome-wide functional sorting for increased GFP accumulation. (e) Numbers of sgRNA targets identified in the genome-wide functional screening as growth repressing, enriched in GFP-accumulation, and both, shown in gray, green and yellow, respectively. The first, second and third sorting populations are shown from left to right. Both the growth screen and the fluorescence sorting were conducted in three biological replicates. The presented numbers represent the average of the biological replicates.

Next, we carried out a screen in order to identify targets that increase protein production, exemplified by GFP, in which cells with high fluorescence were isolated through three rounds of FACS sorting (Fig. 2.1c). A total of at least 3.3 million cells were analyzed, and of these the 1% with the highest fluorescence were collected, resulting in a library coverage of at least 270 fold. The consecutive enrichment sorting resulted in a reduction in the pool of sgRNAs from 10,102 after the first sorting to 3,745 after the third sorting. 65% of the final reads belonged to only two sgRNAs (2.46 million reads on average of a total of 3.71 million reads on average), targeting *uup* and *ygaQ*, and the 40 most dominant sgRNA sequences accounted for 93% of the reads (Fig. 2.2d, sgRNA sequences and number of reads listed in Supplementary Table S2.7). Meanwhile, the overlap between the pool of sgRNAs that down-regulated growth and the pool that were identified during the production

sorting was reduced, leaving 1,332 sgRNAs after the last round of sorting (Fig. 2.2e, Supplementary Table S2.8).

GSEA and metabolic pathway visualization

In order to investigate and visualize the effect of gene repression on metabolic pathways, gene set enrichment analysis (GSEA)[32,33] was performed using the prerank module of GSEAPY. Reference pathway gene sets from KEGG, BIGG and EcoCyc were used for the analysis. Expectedly, the analysis showed significant enrichment of sgRNAs designed to target genes in metabolic pathways known to be essential for growth, such as ribosome function, peptidoglycan biosynthesis and DNA replication (Table 2.1, KEGG reference set). Since cells were grown in minimal medium, different parts of the amino acid metabolism including aminoacyl-tRNA biosynthesis, glycine, serine and threonine metabolism (Table 2.1) and L-lysine, L-threonine and L-methionine biosynthesis (Supplementary Table S2.12, EcoCyc reference set) also showed significant enrichment. Comparison of the results from KEGG, BIGG and EcoCyc databases identified purine and pyrimidine biosynthesis and the TCA cycle as the common enriched metabolic pathways in all three reference data sets (Table 2.1, BiGG reference set in Supplementary Table S2.11, Ecocyc reference set in Supplementary Table S2.12). The screening data was sorted and visualized in two separate biological pathway maps using the visualization tool Escher;[34] one displaying the growth score (Supplementary Fig. S2.1a) and one displaying the abundance after fluorescence sorting (Supplementary Fig. S2.1b). The metabolic pathways significantly enriched in all three reference pathway data sets (purine and pyrimidine biosynthesis and TCA cycle) are shown on maps of the nucleotide (Fig. 2.3a) and core metabolism (Fig. 2.3b). Targets of interest for growth decoupling can be identified as reactions that are both marked blue in the fluorescence maps and red in the growth maps. Although the sgRNA read abundance of the fluorescence sorting is not a direct measure of the protein abundance, the results are in line with our previous results, where inhibition of the nucleotide biosynthesis by targeting *pyrF,* encoding orotidine 5'-phosphate decarboxylase, leads to a decrease in cell growth and increase in expression of GFP, a property that was also applied to increase production of a biochemical compound of interest.[39]

Table 2.1. GSEA of KEGG metabolic pathways for growth inhibiting sgRNA targets (growth score < 1).

Metabolic pathway	NES[a]	p-value	FDR
Ribosome	2.258	0.0	0.0
Peptidoglycan biosynthesis	1.734	0.001	0.020
Aminoacyl-tRNA biosynthesis	1.685	0.0	0.026
Propanoate metabolism	1.674	0.002	0.023
Fatty acid metabolism	1.657	0.003	0.023
Protein export	1.638	0.009	0.025
Pyruvate metabolism	1.620	0.001	0.027
Homologous recombination	1.618	0.002	0.024
Fatty acid biosynthesis	1.614	0.006	0.022
Bacterial secretion system	1.526	0.013	0.057
Glycine, serine and threonine metabolism	1.501	0.008	0.066
Carbon metabolism	1.457	0.002	0.098
Citrate cycle (TCA cycle)	1.448	0.021	0.099
Pentose and glucuronate interconversions	-1.433	0.0	0.125
Pyrimidine metabolism	1.414	0.010	0.129
DNA replication	1.396	0.043	0.142
Purine metabolism	1.331	0.011	0.213
Biosynthesis of amino acids	1.318	0.005	0.224
Biosynthesis of antibiotics	1.288	0.005	0.262
Biosynthesis of secondary metabolites	1.273	0.001	0.282
Metabolic pathways	1.251	0.0	0.292

[a]NES represents the normalized enrichment score for each gene set; FDR shows the estimated probability of a false positive finding within a gene set with a given NES.

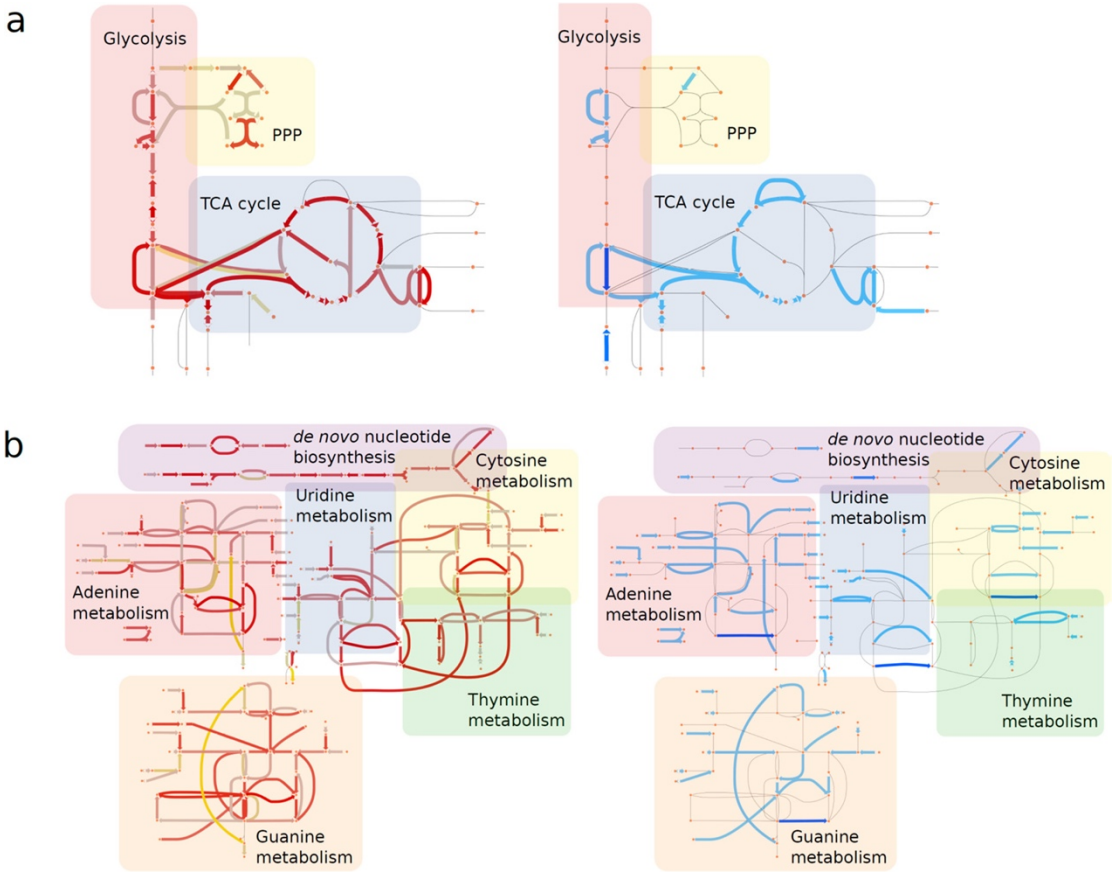


Figure 2.3. Visualization of the effect of repression of genes on either growth or increased protein production. The growth score was calculated for each sgRNA as $\log_2$(mean frequency of sgRNAs_induced cultures/mean frequency of sgRNAs_uninduced cultures). The data is mapped in a red to yellow scale, where a darker shade represents lower (red) or higher (yellow) growth score. No growth effect (growth score = 1) is shown in gray. The fluorescence data from the third fluorescence sorting is mapped in blue, where a darker shade represents higher mean of the normalized count in the sorted population. For all maps, reactions with no data are shown as black thin lines. The approximate limits for the different metabolic processes in the maps are highlighted in different colors. (a) Growth score and fluorescence data mapped to *E. coli* core metabolism. Background shading indicates glycolysis (red), pentose phosphate pathway (PPP, yellow), and TCA cycle (blue). (b) Growth score and fluorescence data mapped to *E. coli* nucleotide metabolism. Background shading indicates reactions belonging to *de novo* nucleotide biosynthesis (purple), adenosine (red), uridine (blue), cytosine (yellow) thymine (green) and guanine (orange) metabolism.

Validation of growth and GFP production of selected targets

In order to verify the effect of the identified targets, we selected and resynthesized 15 sgRNAs from the overlapping pool of sgRNAs found to inhibit growth while also being isolated during the sorting for increased GFP fluorescence (Table 2.2). None of the selected sgRNAs were designed to target a loci near the integrated GFP (Fig. S2.4). Of the 15 targets, 10 were selected because they had the lowest estimated growth score (< -3) (Fig. 2.4a, circled in green, Supplementary Table S2.13), while 5 were selected due to their high GFP score (>10.000) combined with a medium-low growth score (-1 to -2) (Fig. 2.4a, circled in red, Supplementary Table S2.13). The 15 sgRNAs were tested by reconstructing the sgRNA-expressing plasmids, which were then introduced into the same genetic background as the library, compared with a control strain harboring the empty sgRNA plasmids pSLQ1236-*blank*. The cell growth, cell size and the GFP production was quantified with or without induction of the CRISPRi system for each selected sgRNA. Cell size and specific GFP production was measured with FACS and cell growth was measured as OD_{600}.

The majority of selected targets were found to exhibit the expected phenotype, with a reduction of cell growth and an increase in GFP production (Fig. 2.4b). Growth profiling showed that significant growth inhibition was observed for strains with sgRNA targeting *sibB/ibsB, ygaQ, malZ, yjeN yaiY(p), ydiB, lpxC, glcA* and *casC*. For the majority of these strains, the growth rate was slowly reduced until the cells finally entered stationary phase (Supplementary Fig. S2.3).

End-point measurements of OD and fluorescence showed that 12 strains exhibited more than 20% increase in specific fluorescence and 9 strains showed more than 20% reduction in cell density when comparing the induced and the uninduced cultures (Fig. 2.4a). Strains expressing sgRNAs targeting *sibB/ibsB, ygaQ, malZ, yjeN, yaiY(p), ydiB, lpxC* and *glcA* exhibited both phenotypes. Seven strains (*sibB/ibsB, ygaQ, malZ, yjeN, yaiY(p), ydiB* and *lpxC*) showed more than 2-fold increase of specific fluorescence upon induction, and four of these (*sibB/ibsB, ygaQ, ydiB* and *glcA*) showed a signifi-

Table 2.2. Target genes or locations of the 15 sgRNAs selected for validation.

Target	Essentiality	Function	Operon
ruvA	Not essential	Branch migration of Holliday structures; repair	*ruvA,ruvB*
rcsB-rcsC, intergenic	Not essential	Repetitive Extragenic Palindrome (REP) element	
gyrA	Essential	DNA gyrase, subunit A	
yfjW	Not essential	CP4-57 prophage; predicted inner membrane protein	*yfjW*
lpxC	Essential	UDP-3-O-acyl-N-acetylglucosamine deacetylase	*lpxC*
casC	Not essential	Cascade subunit C	*casA, casB, casC*
yaiY(p), intergenic	Not essential	Inner membrane protein	
ddpD	Not essential	D,D-dipeptide ABC transporter ATPase	*ddpX, A, B, C, D, F*
ydiB	Not essential	Shikimate dehydrogenase / quinate dehydrogenase	*ydiN, ydiB, aroD*
sibB/ibsB	No information	Toxin-antitoxin locus	
yheV	Not essential	Predicted protein	*kefG, kefB, yheV*
ygaQ	Not essential	Conserved protein	
glcA	Not essential	Glycolate / lactate:H+ symporter	
yjeN	Not essential	Predicted protein	*yjeN, yjeO*
malZ	Not essential	Maltodextrin glucosidase	

cantly higher homogeneous shift of cellular fluorescence. With the exception of the strain targeting *lpxC*, all strains had less than 1.5-fold increase in cell

size, indicating that the increase in specific fluorescence was mainly due to elevated GFP density. Among the selected sgRNAs, *yheV* and *sibB/ibsB* displayed the highest specific fluorescence with a 5.5- and 5.1-fold increase, respectively. The sgRNA targeting *yheV* did not display growth arrest, while the sgRNA targeting *sibB/ibsB* showed a 45% decrease in cell density (Fig. 2.4b).

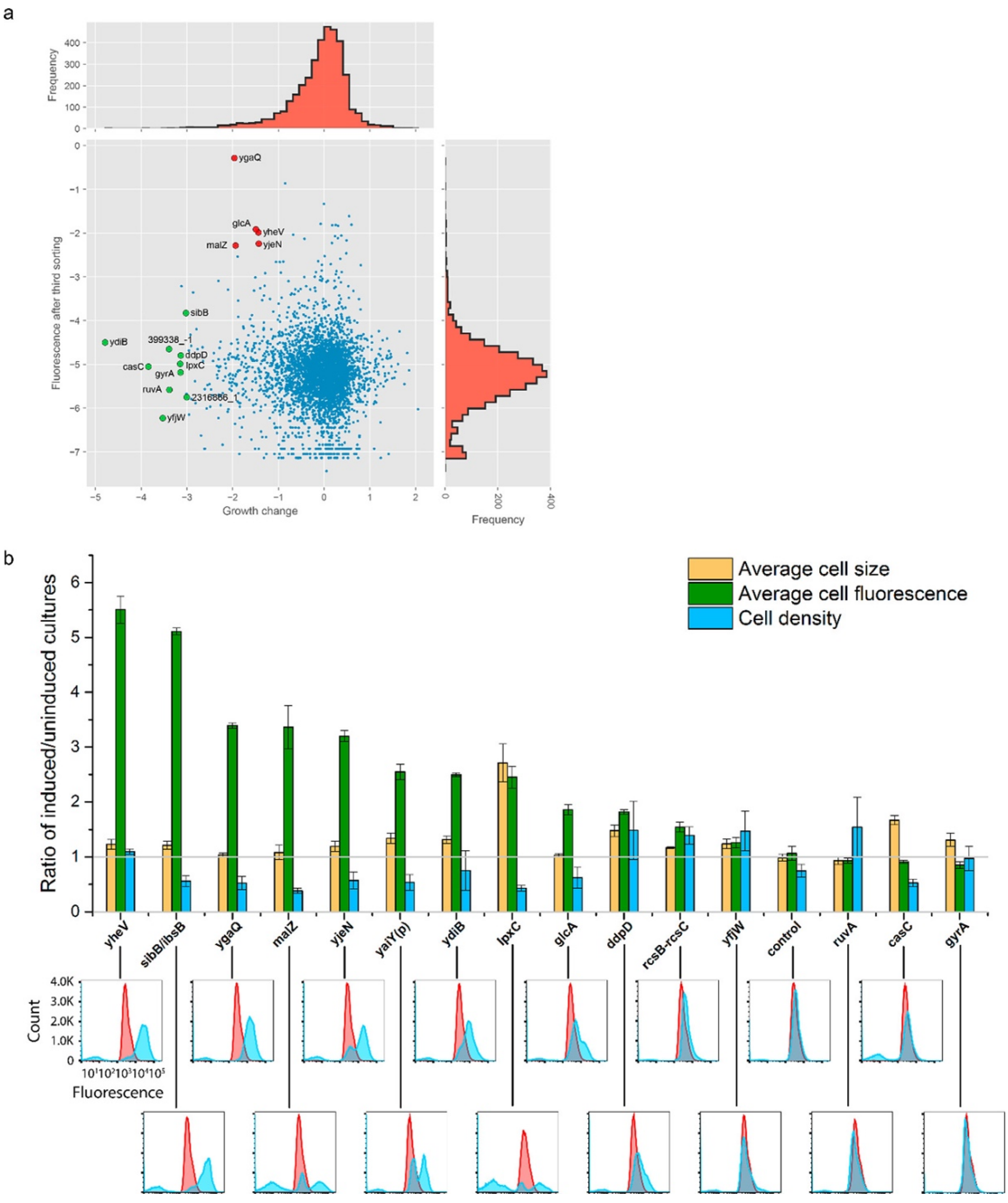


Figure 2.4. Selection and validation of genetic growth switches that enables continued protein production. (a) The pool of sgRNAs that were present in the last round of sorting, with their effect on cell growth and GFP enrichment. The growth score was calculated as log2(mean frequency of sgRNAsinduced cultures/mean frequency of sgRNAsuninduced cultures), and the GFP score was calculated as the mean normalized reads in the last round of sorting. sgRNAs circled with green and red were selected for further validation. (b) Characterization of the growth inhibition and GFP production for the selected targets. The upper diagram shows the ratio of specific fluorescence, cell size, and cell density of induced/uninduced cultures for each tested target. The average value represents means of three independent experiments. The diagrams below show the corresponding fluorescence distribution for induced (red) and uninduced (blue) cultures. Average cell fluorescence (FITC-A) and average cell size (FSC-A) was analyzed by flow cytometry after 24 h of cultivation.

2.4 Discussion

In this study, we performed a genome-wide CRISPRi-based repression screen in order to identify sgRNA targets that, when activated, cause a decrease in cell growth and/or an increase in protein production. An inducible system for gene repression such as this expands the applicability of this method to all essential genes that cannot be investigated through knock-out. The reported high efficiency as well as specificity of the CRISPRi system gives this method substantial advantages over other functional screening methods.[11,12] We have previously used rational targeting of biomass inhibition in order to increase product yield; either by using inhibitors, nutrient limitation,[7] or by targeting known essential genes by mutation or CRISPRi.[39]

However, scanning the entire bacterial genome can for example identify genes or locations suitable as novel antibiotic drug targets for inhibiting pathogen growth, or provide methods for improving the performance of microbial cell factories.

Previous studies have for example applied genome-wide CRISPRi screens in *E. coli* for functional genomics[25] and to identify essential genes.[23] These studies applied a several-fold larger library with either 60,000 sgRNAs designed to target the start of ORFs[25] or a start pool of 92,000 sgRNAs designed to target positions randomly along the chromosome.[23] Similar to this study, they reported that 93%[25] and 79%[23] of genes previously identified as essential[8] were found to be essential or significantly impaired cell growth when targeted with CRISPRi. It is difficult to directly compare the essential genes identified in the different studies given that different media composition was used. However, both studies observed a significant depletion of growth inhibiting targets among essential cellular processes including ribosome assembly, peptidoglycan biosynthesis, cell wall synthesis and tRNA metabolism, which is consistent with our findings.

In our analysis, 5,425 sgRNAs were significantly depleted during growth following induction, and sgRNAs targeting *acpP*, *dnaK*, *fumA*, *gyrA*, *gyrB*, *holA*, *infA*, *rho*, *rplJ*, *rrfF*, and *yeaP* had growth scores below -3 for both sgRNAs. While there is a chance of unexpected sgRNA-directed inhibitory effects,[36] essential genes were strongly enriched among the sgRNA with lowest growth scores. Furthermore, the library was enriched for production of GFP in three consecutive rounds of fluorescence-based cell sorting, resulting in isolation of fewer and fewer sgRNAs, with only 40 sgRNAs making up 93% of sequences ultimately.

We furthermore performed GSEA to identify metabolic pathways enriched in growth inhibiting sgRNAs. As expected, processes involved in cellular proliferation were identified as significantly enriched in the growth inhibited-part of the library. Next, the growth score and fluorescence screening data was visualized on metabolic pathway maps to allow for easy identification of possible growth decoupling targets within specific categories of metabolic processes.

We validated the initial screening data by characterization of 15 high-performing sgRNAs in order to identify growth switches that would repress cell growth while maintaining or increasing protein production. The majority of the selected sgRNAs showed phenotypes consistent with the

initial screening, and most of these genes or genomic locations the sgRNAs were designed to target have not been previously characterized or linked to this specific phenotype (Table 2.2). Among the effective sgRNAs was one targeting *sibB/ibsB*, a known toxin-antitoxin system that induces cell stasis when disrupted,[40] and here it was found to significantly inhibit growth and increase the expression of GFP more than 5-fold. Several other sgRNAs, including *ygaQ, malZ, yjeN yaiY(p), ydiB, lpxC* and *glcA*, showed decreased growth and increased GFP production, making them promising candidates for functional growth switches.

Because of unexpected inhibitory or off-target effects, expressing an sgRNA targeting a specific gene will not always correspond to the effect seen from silencing or knocking out that gene. This screen does not consider any unexpected inhibitory effects caused by sgRNA expression, but it simply identifies sgRNAs that work as growth switches. Thus, for a portion of sgRNAs identified in the library, the corresponding gene may not work as a growth decoupling target.

Possible unexpected sgRNA inhibitory effects could include for example sgRNAs reducing proteolytic degradation of GFP, stabilizing mRNA or improving protein maturation. As in the example of lpxC, where cell size was increased almost 3-fold (Fig. 2.4b), inhibition of certain genes may result in an increased cell size, which in itself leads to a higher content of protein. Cells may also accumulate higher amounts of protein by simply not dividing and diluting the protein pool. However, an increase in protein content could also be seen for targets that did not inhibit growth, such as *yheV*. Downstream polar effects for unknown genes and genes located in operons were not considered in this work and are likely present in the library. Silencing a gene in an operon may affect and inhibit downstream operon genes, unless there is an internal promoter present controlling the downstream gene.[25]

Depending on the specific target or target group, there are probably various growth decoupling mechanisms of action in play. Inhibition of certain essential proteins or metabolic functions may lead to the upregulation of ribosome production to compensate for the missing protein, which could generally upregulate the protein synthesis rate. Previously, we have seen that

CRISPRi of *oriC* and *pyrF* leads to 46-86% of the excess carbon being redistributed from biomass to product formation,[39] indicating that inhibition of growth leads to a shift in carbon flux from growth to production. Reducing biomass accumulation while maintaining carbon flux and metabolic activity seems to result in a redistribution of carbon toward other active metabolic processes.

2.5 Conclusions

The sgRNAs found in our CRISPRi screen can be applied to enhance protein production, but they may also have the potential to be used to increase production of biochemicals. Biochemical production, similar to protein synthesis, requires continuous synthesis of different metabolic precursors as well as energy generation. The results validate the use of CRISPRi-based genome-wide screening for identifying novel functions of characterized or uncharacterized genes for the purpose of fundamental or application based studies.

2.6 Acknowledgements

We would like to thank Anna Koza for help with Next Generation Sequencing. S.L. was funded by a grant from the People Programme (Marie Curie Actions) of the European Union Seventh Framework Programme FP7-People-2012-ITN, under Grant agreement no. 317058, "BACTORY". C.B.J was funded by grant no. NNF15OC0015246 from the Novo Nordisk Foundation. J.L. was funded by grant no. NNF16CC0020908 from the Novo Nordisk Foundation.

2.7 References for Chapter 2

1. Na, D., Kim, T. Y. & Lee, S. Y. Construction and optimization of synthetic pathways in metabolic engineering. *Curr. Opin. Microbiol.* **13**, 363–370 (2010).
2. Soma, Y., Tsuruno, K., Wada, M., Yokota, A. & Hanai, T. Metabolic flux redirection from a central metabolic pathway toward a synthetic pathway using a metabolic toggle switch. *Metab. Eng.* **23**, 175–184 (2014).
3. Brockman, I. M. & Prather, K. L. J. Dynamic knockdown of E. coli central metabolism for redirecting fluxes of primary metabolites. *Metab. Eng.* **28**, 104–113 (2015).
4. Izard, J. *et al.* A synthetic growth switch based on controlled expression of RNA polymerase. *Mol. Syst. Biol.* **11**, 840 (2015).
5. Li, S. *et al.* Enhanced protein and biochemical production using CRISPRi-based growth switches. *Metab. Eng.* **38**, 274–284 (2016).
6. Lo, T. M., Chng, S. H., Teo, W. S., Cho, H. S. & Chang, M. W. A Two-Layer Gene Circuit for Decoupling Cell Growth from Metabolite Production. *Cell Syst.* **3**, 133–143 (2016).
7. Li, S., Jendresen, C. B. & Nielsen, A. T. Increasing production yield of tyrosine and mevalonate through inhibition of biomass formation. *Process Biochem.* **51**, 1992–2000 (2016).
8. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
9. Koo, B. M. *et al.* Construction and Analysis of Two Genome-Scale Deletion Libraries for Bacillus subtilis. *Cell Syst.* **4**, 291-305.e7 (2017).
10. Nakashima, N., Tamura, T. & Good, L. Paired termini stabilize antisense RNAs and enhance conditional gene silencing in Escherichia coli. *Nucleic Acids Res.* **34**, e138–e138 (2006).
11. Na, D. *et al.* Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. *Nat. Biotechnol.* **31**, 170–174 (2013).
12. Meng, J. *et al.* A genome-wide inducible phenotypic screen identifies antisense RNA constructs silencing Escherichia coli essential genes. *FEMS Microbiol. Lett.* **329**, 45–53 (2012).
13. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).
14. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-

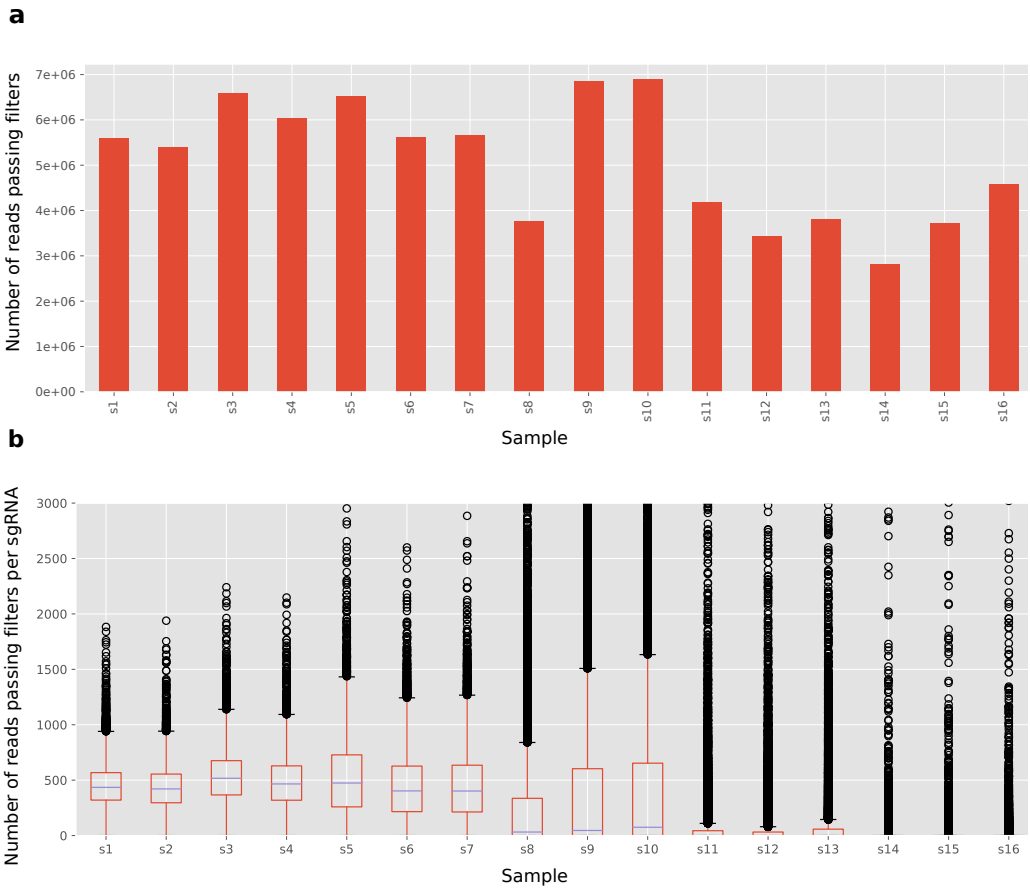
specific control of gene expression. *Nat. Protoc.* **8**, 2180–2196 (2013).
15. Quan, J. *et al.* Parallel on-chip gene synthesis and application to optimization of protein expression. *Nat. Biotechnol.* **29**, 449–452 (2011).
16. Metzker, M. L. Sequencing technologies - the next generation. *Nat. Rev. Genet.* **11**, 31–46 (2010).
17. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science* **343**, 80–84 (2014).
18. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–7 (2014).
19. Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).
20. Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* (2014). doi:10.1016/j.cell.2014.09.029
21. Peters, J. M. *et al.* A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. *Cell* **165**, 1493–506 (2016).
22. Liu, X. *et al.* High-throughput CRISPRi phenotyping identifies new essential genes in Streptococcus pneumoniae. *Mol. Syst. Biol.* **13**, 931 (2017).
23. Rousset, F. *et al.* Genome-wide CRISPR-dCas9 screens in E. coli identify essential genes and phage host factors. *PLoS Genet.* **14**, 1–28 (2018).
24. Hawkins, J. S. *et al.* Modulated efficacy CRISPRi reveals evolutionary conservation of essential gene expression-fitness relationships in bacteria. *bioRxiv* (2019). doi:10.1101/805333
25. Wang, T. *et al.* Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance-net. *Nat. Commun.* **9**, (2018).
26. Lee, H. H. *et al.* Functional genomics of the rapidly replicating bacterium Vibrio natriegens by CRISPRi. *Nat. Microbiol.* **4**, 1105–1113 (2019).
27. Peters, J. M. *et al.* Enabling genetic analysis of diverse bacteria with Mobile-CRISPRi. *Nature Microbiology* (2019). doi:10.1038/s41564-018-0327-z
28. Camsund, D. *et al.* Time-resolved imaging-based CRISPRi screening. *Nat. Methods* (2019). doi:10.1038/s41592-019-0629-y
29. Bonde, M. T. *et al.* Predictable tuning of protein expression in bacteria. *Nat. Methods* **13**, 233–236 (2016).
30. Ronda, C. *et al.* Accelerating genome editing in CHO cells using

CRISPR Cas9 and CRISPy, a web-based target finding tool. *Biotechnol. Bioeng.* **111**, 1604–16 (2014).
31. Sun, J., Nishiyama, T., Shimizu, K. & Kadota, K. TCC: an R package for comparing tag count data with robust normalization strategies. *BMC Bioinformatics* **14**, 219 (2013).
32. Mootha, V. K. *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).
33. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–50 (2005).
34. King, Z. A. *et al.* Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *PLOS Comput. Biol.* **11**, e1004321 (2015).
35. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41**, 7429–7437 (2013).
36. Vigouroux, A., Rousset, F., Varet, H., Khanna, V. & Bikard, D. A CRISPRi screen in E. coli reveals sequence-specific toxicity of dCas9. *Nat. Commun.* 1–10 doi:10.1038/s41467-018-04209-5
37. Patrick, W. M., Quandt, E. M., Swartzlander, D. B. & Matsumura, I. Multicopy suppression underpins metabolic evolvability. *Mol. Biol. Evol.* (2007). doi:10.1093/molbev/msm204
38. Guo, J. *et al.* Improved sgRNA design in bacteria via genome-wide activity profiling. *Nucleic Acids Res.* (2018). doi:10.1093/nar/gky572
39. Li, S. *et al.* Enhanced protein and biochemical production using CRISPRi-based growth switches. *Metab. Eng.* **38**, 274–284 (2016).
40. Fozo, E. M. New type I toxin-antitoxin families from 'wild' and laboratory strains of E. coli: Ibs-Sib, ShoB-OhsC and Zor-Orz. *RNA Biology* **9**, 1504–1512 (2012).

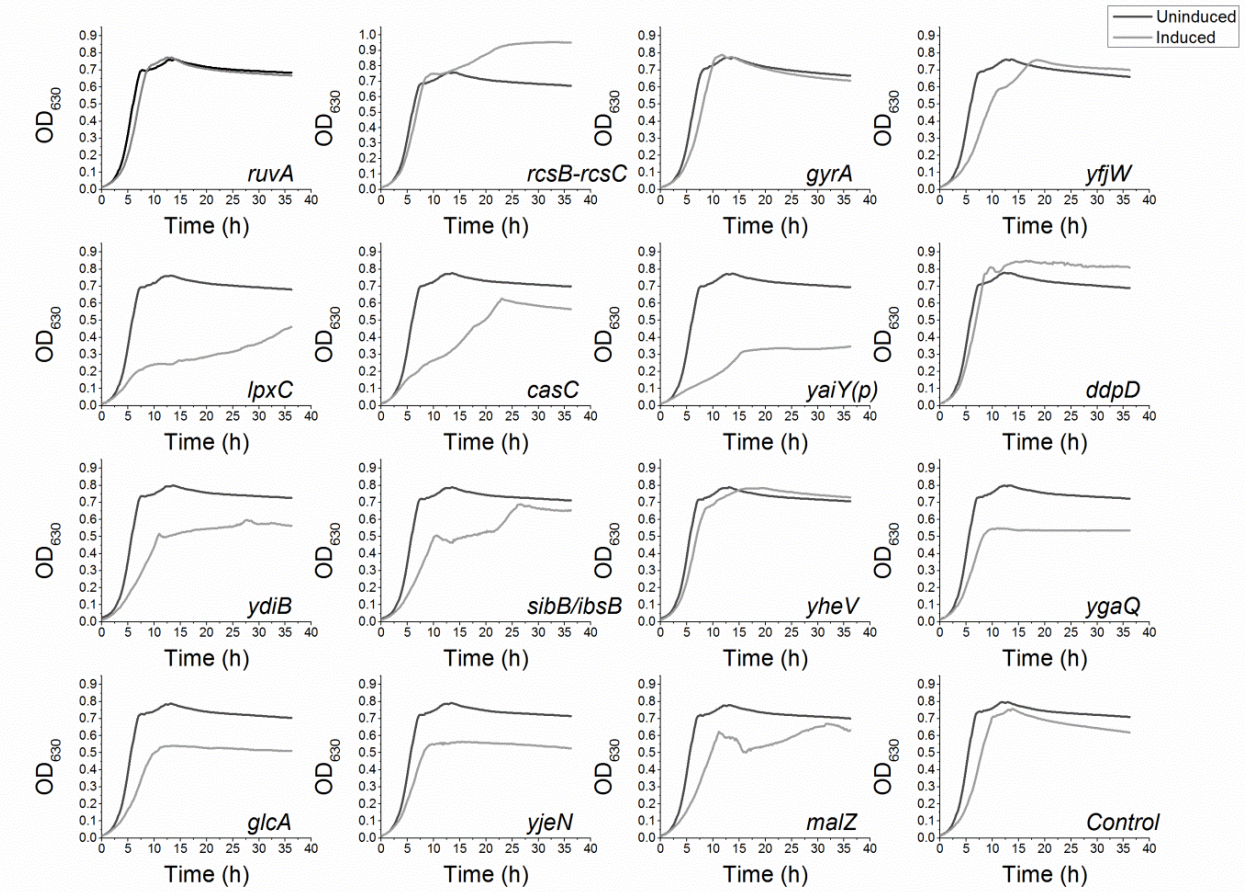
2.8 Supplementary materials

Supplementary Figure S2.1 can be found at https://dx.doi.org/10.1021/acssynbio.9b00143.

Supplementary Figure S2.1. Visualization of the fluorescence and growth score data from the CRISPRi library screen. (a) The growth score was calculated for each sgRNA as $\log_2$(mean frequency of sgRNAs$_{\text{induced cultures}}$/mean frequency of sgRNAs$_{\text{uninduced cultures}}$). The data was mapped in a red to blue scale, where a darker shade represents lower or higher growth score for red and blue, respectively. No growth effect (growth score = 1) is shown in grey. Reactions with no data are shown in black. (b) The fluorescence data from the 3rd sorting is mapped in green, where a darker shade represents higher mean of normalized count in the sorting. Reactions with no data are shown in black.



Supplementary Figure S2.2. Sequencing depth. (a) Total reads per samples. (b) Reads per sgRNA per sample. s1 is the initial sgRNA library; s2-4 is the sgRNA library harvested from cells before the growth experiment (in triplicates); s5-7 is the sgRNA library harvested from cells after the growth experiment (in triplicates); s8-10 is the sgRNA library harvested after the first FACS sorting (in triplicates); s11-13 is the sgRNA library harvested after the second FACS sorting (in triplicates); s14-16 is the sgRNA library harvested after the first FACS sorting (in triplicates).



Supplementary Figure S2.3. Growth profiling of strains with sgRNA targeting selected genes or genomic locations. Cell growth was monitored for 36 h after induction with anhydrotetracycline.

Supplementary Tables S2.1 to S2.8 are presented as Excel files, and can be found at https://dx.doi.org/10.1021/acssynbio.9b00143.

Supplementary Table S2.1. The designed sgRNA target sequences targeting genes. The table lists the starting position of the coding sequence, the strand (1 or -1), the gene length and for each target sequence the starting position and strand (1 or -1) relative the coding sequence, as well as the CRISPY++ score.[30]

Supplementary Table S2.2. The designed sgRNA target sequences targeting intergenic regions. The table list the position of the target sequence, the strand (1 or -1), as well as the CRISPY++ score.[30]

Supplementary Table S2.3. Calculated growth score for all sgRNA targets.

Supplementary Table S2.4. Confidence scores (q- and p-values) for the enrichment analysis.

Supplementary Table S2.5. Mean values and standard deviation for the histograms in Figure 2.2c.

Supplementary Table S2.6. Essential gene list with the corresponding effect of CRISPRi repression.

Supplementary Table S2.7. Mean of normalized counts of each sgRNA present in the first, second and third round of sorting for GFP enrichment.

Supplementary Table S2.8. List of the 1332 targets that were present in the third round of fluorescence sorting, while also resulting in significant growth inhibition upon induction of the library.

Supplementary Table S2.9. Primers used in the study.

Primers	**Sequence (5' → 3')**	**Purpose and reference**
SON119 (Sg-Seq-R)	GCAGCGAGTCAGTGAGCGAG	Sequencing primer for sgRNA[5]
SON143 (sg_tet_PCR_F)	TGTTGAATACTCATACTCTTCC	Sequencing primer for sgRNA[5]
SON172	CTCCCTATCAGTGATAGAGAAAAGACT	Amplification of library

SON173	CGGGCCCAAGCTTCAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGAC	Amplification of library
SON176 (1236empty_F)	AGTCGGU GCTTTTTTTGAAG	Amplification of empty pSLQ1236 for USER cloning[5]
SON177 (1236empty_R)	ACCGACU ACTAGTCTTTTCTCTATCACTG	Amplification of empty pSLQ1236 for USER cloning[5]
SON178	AGTCGGTGCTTTTTTTGAAG	Amplification of pSLQ1236
SON179	ACTAGTCTTTTCTCTATCACTGATAGGGAG	Amplification of pSLQ1236
SON203	ACTTCAATTAACACCAGCGG GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Amplification of *ruvA* target
SON204	CCGCTGGTGTTAATTGAAGT ACTAGTCTTTTCTCTATCACTGATAGGGA	Amplification of *ruvA* target
SON205	AGACGCGTTAGTGTCTTATC GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Amplification of rcsB-rcsC, intergenic target
SON206	GATAAGACACTAACGCGTCT ACTAGTCTTTTCTCTATCACTGATAGGGA	Amplification of rcsB-rcsC, intergenic target
SON207	AGCTCTTCCTCAATGTTGAC GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Amplification of *gyrA* target
SON208	GTCAACATTGAGGAAGAGCT ACTAGTCTTTTCTCTATCACTGATAGGGA	Amplification of *gyrA* target
SON209	ATTACCTTTTGTGAAGGCAG GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Amplification of *yfjW* target
SON210	CTGCCTTCACAAAAGGTAAT ACTAGTCTTTTCTCTATCACTGATAGGGA	Amplification of *yfjW* target
SON211	CGTCAGGGTGACTTTCTTGC GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Amplification of *lpxC* target

SON212	GCAAGAAAGTCACCCTGACG ACTAGTCTTTTCTCTATCACTGATAGGG A	Amplification of *lpxC* target
SON213	CGTCGCGGTTCAGACATGAA GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGC	Amplification of *casC* target
SON214	TTCATGTCTGAACCGCGACG ACTAGTCTTTTCTCTATCACTGATAGGG A	Amplification of *casC* target
SON215	GAGAAGCAGATGACTTCCGG GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGC	Amplification of *yaiY*(p) intergenic target
SON216	CCGGAAGTCATCTGCTTCTC ACTAGTCTTTTCTCTATCACTGATAGGG A	Amplification of *yaiY*(p) intergenic target
SON217	GTTGAATGTCCAGAACGGGT GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGC	Amplification of *ddpD* target
SON218	ACCCGTTCTGGACATTCAAC ACTAGTCTTTTCTCTATCACTGATAGGG A	Amplification of *ddpD* target
SON219	TAAACTGTGGCGGATAGGAT GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGC	Amplification of *ydiB* target
SON220	ATCCTATCCGCCACAGTTTA ACTAGTCTTTTCTCTATCACTGATAGGG A	Amplification of *ydiB* target
SON221	TACTAAGACTACCAGGGCGG GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGC	Amplification of *sibB/ibsB* target
SON222	CCGCCCTGGTAGTCTTAGTA ACTAGTCTTTTCTCTATCACTGATAGGG A	Amplification of *sibB/ibsB* target
SON223	AATCCTGCGCCTGACAGGCC GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGC	Amplification of *yheV* target
SON224	GGCCTGTCAGGCGCAGGATT ACTAGTCTTTTCTCTATCACTGATAGGG A	Amplification of *yheV* target
SON225	ATAGGTAAATTTCTGGGTCC GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGC	Amplification of *ygaQ* target

SON226	GGACCCAGAAATTTACCTAT ACTAGTCTTTTCTCTATCACTGATAGGG A	Amplification of *ygaQ* target
SON227	CGGCATATACATTTGGGTCC GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGC	Amplification of *glcA* target
SON228	GGACCCAAATGTATATGCCG ACTAGTCTTTTCTCTATCACTGATAGGG A	Amplification of *glcA* target
SON229	CTTCGGTTATTGCCGGGTCC GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGC	Amplification of *yjeN* target
SON230	GGACCCGGCAATAACCGAAG ACTAGTCTTTTCTCTATCACTGATAGGG A	Amplification of *yjeN* target
SON231	TGTTTAACAAATGGGGGCAC GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGC	Amplification of *malZ* target
SON232	GTGCCCCCATTTGTTAAACA ACTAGTCTTTTCTCTATCACTGATAGGG A	Amplification of *malZ* target
SON233	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAG CACTCCCTATCAGTGATAGAGAAAAG	Amplification of library for sequencing
SON234	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGATTCAGATCCTCTTCTGAGATG AG	Amplification of library for sequencing

Supplementary Table S2.10. Strains and plasmids.

Strain	**Description**	**Source**
E.coli NEB 5-alpha	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17* (cloning strain)	NEB
E.coli MG1655	F- lambda- *ilvG- rfb-50 rph-1*	This study
E.coli Sij17	*E.coli* MG1655 BBJ23100-*gfp*::KmR	[29]
Sij17[dCas9-6blank]	*E.coli* Sij17 with pdCas9-bacteria and pSLQ1236-*blank*	This study

Plasmids	Description	Reference/ source
pdCas9-bacteria	dCas9 expression plasmid, dCas9 was expressed under a tetracycline inducible promoter; CamR	[13]
pSLQ1236	sgRNA expression plasmid, sgRNA was expressed under tetracycline inducible promoter; AmpR	[14]
p1236-blank	pSLQ1236 without sgRNA sequence; AmpR	This study

Supplementary Table S2.11. GSEA of BIGG biological process terms for growth inhibiting sgRNA targets (growth score < 1). NES represents the normalized enrichment score for each gene set; FDR shows the estimated probability of a false positive finding within a gene set with a given NES.

Biological process term	NES	*p*-value	FDR
Lipopolysaccharide Biosynthesis / Recycling	1.335	0.031	0.263
Citric Acid Cycle	1.413	0.048	0.167
Transport, Outer Membrane	1.465	0.013	0.126
Cell Envelope Biosynthesis	1.501	0.007	0.131
Purine and Pyrimidine Biosynthesis	1.557	0.005	0.120

Supplementary Table S2.12. GSEA of EcoCyc biological process terms for growth inhibiting sgRNA targets (growth score < 1). NES represents the normalized enrichment score for each gene set; FDR shows the estimated probability of a false positive finding within a gene set with a given NES.

Biological process term	NES	*p*-value	FDR
Superpathway of glyoxylate bypass and TCA	1.365	0.046	0.220
Aspartate superpathway	1.379	0.041	0.246
tRNA charging	1.452	0.022	0.150

Superpathway of histidine, purine, and pyrimidine biosynthesis	1.472	0.004	0.145
Superpathway of L-lysine, L-threonine and L-methionine biosynthesis I	1.491	0.025	0.145
Superpathway of purine nucleotides *de novo* biosynthesis II	1.541	0.006	0.110
N-10-formyl-tetrahydrofolate biosynthesis	1.591	0.007	0.080
Superpathway of pyrimidine deoxyribonucleotides *de novo* biosynthesis	1.725	0.003	0.021
Peptidoglycan biosynthesis I (*meso*-diaminopimelate containing)	1.977	0.0	0.0

Supplementary Table S2.13. Data analysis of the sgRNAs selected for further validation.

Target	sgRNA sequence	Group	Count	Rel. gRNA pos.	Rel. gRNA strand	Target pos.	Rel. count	Fluor. after 3rd sort.	Growth change
2316866_1	AGACGCGTTAGTGTCTTATC	Induced	73	0	True	2316866	0.000012	-5.751755	-3.00337.
		Uninduced	533	0	True	2316866	0.000089		
399338_-1	GAGAAGCAGATGACTTCCGG	Induced	49	0	False	399338	0.000008	-4.651215	-3.38918.
		Uninduced	470	0	False	399338	0.000078		
casC	CGTCGCGGTTCAGACATGAA	Induced	48	41	False	2881050	0.000008	-5.052785	-3.84057.
		Uninduced	635	41	False	2881050	0.000106		

ddpD	GTTGAATGTCCAGAACGGGT	Induced	76	5	False	1558030	0.000013	-4.798511	-3.13351.
		Uninduced	618	5	False	1558030	0.000102		
glcA	CGGCATATACATTTGGGTCC	Induced	184	7	False	3119596	0.000031	-1.916694	-1.49174.
		Uninduced	473	7	False	3119596	0.000079		
gyrA	AGCTCTTCCTCAATGTTGAC	Induced	41	27	False	2336792	0.000007	-5.183951	-3.13922.
		Uninduced	333	27	False	2336792	0.000056		
lpxC	CGTCAGGGTGACTTTCTTGC	Induced	64	58	False	106556	0.000011	-4.988975	-3.14681.
		Uninduced	520	58	False	106556	0.000087		
malZ	TGTTTAACAAATGGGGGCAC	Induced	96	21	False	422517	0.000016	-2.285522	-1.93609.
		Uninduced	337	21	False	422517	0.000056		
ruvA	ACTTCAATTAACACCAGCGG	Induced	24	39	False	1945364	0.000004	-5.581891	-3.38208.
		Uninduced	227	39	False	1945364	0.000038		
sibB	TACTAAGACTACCAGGGCGG	Induced	38	20	False	2153643	0.000006	-3.828478	-3.01953.
		Uninduced	281	20	False	2153643	0.000047		
ydiB	TAAACTGTGGCGGATAGGAT	Induced	22	40	False	1773788	0.000004	-4.499705	-4.78875.
		Uninduced	562	40	False	1773788	0.000093		
yfjW	ATTACCTTTTGTGAAGGCAG	Induced	63	133	False	2773317	0.000011	-6.228876	-3.528.
		Uninduced	669	133	False	2773317	0.000111		
ygaQ	ATAGGTAAATTTCTGGGTCC	Induced	133	15	False	2786396	0.000022	-0.286271	-1.96165.
		Uninduced	478	15	False	2786396	0.000079		
yheV	AATCCTGCGCCTGACAGGCC	Induced	308	38	False	3478591	0.000052	-1.985745	-1.43647.
		Uninduced	766	38	False	3478591	0.000128		

yjeN	CTTCGGTTATTGCCGGGTCC	Induced	188	14	False	4385392	0.000032	-2.242079	-1.42819.
		Uninduced	462	14	False	4385392	0.000077		

Chapter 3

CRISPRI INTERFERENCE OF NUCLEOTIDE BIOSYNTHESIS IMPROVES PRODUCTION OF A SINGLE-DOMAIN ANTIBODY IN *ESCHERICHIA COLI*

Jenny Landberg[†], Naia Risager Wright[†], Tune Wulff[†], Markus Herrgård[†‡], Alex Toftgaard Nielsen[†*]

[†] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet 220, 2800 Kongens Lyngby, Denmark.
[‡] Current address: BioInnovation Institute, Ole Maaløes Vej 3, 2200 Copenhagen, Denmark.

***Corresponding Author**
Alex Toftgaard Nielsen
The Novo Nordisk Foundation Center for Biosustainability
Technical University of Denmark
Kemitorvet, DK-2800 Kongens Lyngby
Denmark
Email: atn@biosustain.dtu.dk

Abstract

Growth decoupling can be used to optimize production of biochemicals and proteins in cell factories. Inhibition of excess biomass formation allows for carbon to be utilized efficiently for product formation instead of growth, resulting in increased product yields and titers. Here, we used CRISPR interference (CRISPRi) to increase production of a single domain antibody (sdAb) by inhibiting growth during production. First, we screened 21 sgRNA targets in the purine and pyrimidine biosynthesis pathways, and found that repression of 11 pathway genes led to increased GFP production and decreased growth. The sgRNA targets *pyrF*, *pyrG*, and *cmk* were selected and further used to improve production of two versions of an expression-optimized sdAb. Proteomics analysis of the sdAb-producing *pyrF*, *pyrG*, and *cmk* growth decoupling strains showed significantly decreased RpoS levels and an increase of ribosome-associated proteins, indicating that the growth decoupling strains do not enter stationary phase and maintain their capacity for protein synthesis upon growth inhibition. Finally, sdAb production was scaled up to shake-flask fermentation where the product yield was improved 2.6-fold compared to the control strain with no sgRNA target sequence. An sdAb content of 14.6% was reached in the best-performing *pyrG* growth decoupling strain.

3.1 Introduction

Decoupling growth and production can significantly improve yield, titer and productivity by dividing the bioproduction process into two phases; a growth phase where substrate is transformed into catalytic biomass, followed by a production phase where growth is stalled and substrate is utilized for product formation.[1] Growth decoupling has been shown to increase production of both biochemicals[2–4] and proteins,[5,6] and a recently published strain design algorithm further highlights the possibilities of using growth decoupling to improve production of a large number of small molecules in *Escherichia coli*.[7]

Decoupling is generally achieved by natural or synthetic regulation of growth and/or induction of product expression. Unless necessary for the specific product, it is important that the decoupled cells do not enter stationary phase in order to maintain cellular activity and production capacity. Natural regulation includes starvation for nutrients or other essential compounds,[4,8,9] as well as regulation of environmental cues such as oxygen[10] or pH.[11] Synthetic regulation usually involves synthetic circuits that interfere with growth and metabolic flux, and can be combined with sensing of environmental inputs. For example, temperature-based decoupling can be achieved by coupling expression of a flux node to a temperature-inducible promoter,[12] or by using heat-sensitive enzymes that shut down flux through a competing pathway upon temperature shift.[13]

Most commonly, synthetic regulation takes place on a translational or transcriptional level. However, post-translational regulation of pathway proteins has also been shown to efficiently decouple growth and production. Tagging the first enzyme in a product-forming pathway with a degradation tag that is cleaved off upon induction resulted in complete uncoupling of growth and production and rapid accumulation of high amounts of the biopolymer poly-3-hydroxybutarate.[3] In another study, inducible degradation of a glycolytic enzyme was used to re-direct flux toward product formation and increase titer of *myo*-inositol two-fold.[14]

Translational regulation can be achieved through the use of orthogonal ribosomes for translation of the product or product pathway,[15] or by stalling growth by inhibiting endogenous ribosomes.[16] Combining inhibition of the native *E. coli* ribosomes with glycotransferase expression from T7 polymerase increased the glycotransferase product yield almost 12-fold.[5]

Regulation of transcription has successfully been applied to inhibit cell accumulation and turn on production, often through synthetic genetic circuits.[17] By controlling isocitrate lyase expression with a degradable inducer, carbon flux could gradually be routed toward wax ester accumulation, improving wax ester yields almost 4-fold during growth on acetate.[18] CRISPRi is an excellent tool for regulating gene expression on a transcriptional level.[19] It can be induced to target gene(s) or cellular function(s) at a desired time point in order to increase precursor supply,[20] redirect metabolic flux toward production and away from byproduct formation,[21,22] or to induce growth arrest by inhibition of essential genes.[2,23] Partial CRISPRi-based inhibition of citrate synthase GltA increased productivity of butanol in the cyanobacteria *Synechocystis*,[23] and CRISPRi-based inhibition of *pyrF* improved the yield of mevalonate almost 5-fold in *E. coli*.[2] In order to screen the *E. coli* genome for additional promising growth decoupling targets, we established a genome-wide sgRNA library and screened it to identify several targets that increased GFP production and inhibited growth.[24] The results from this study indicated an enrichment of promising targets among genes involved in biosynthesis of purines and pyrimidines.

Here, we construct CRISPRi-based growth switches targeting the purine and pyrimidine biosynthesis pathway in *E. coli*, and screen for improved protein production. More than half of the screened growth switch targets display significant growth decoupling effects, with simultaneous growth inhibition and increase in GFP production. We apply the three top-performing targets for production of two different expression-optimized versions of a single-domain antibody (sdAb or Nanobody®).[25] sdAbs are antibodies derived from camelids or cartilaginous fishes with potential use in various biotechnological applications, including as therapeutics.[26] Here, we see a significant increase in sdAb production for cultures with activated

CRISPRi. To detect proteome-wide changes induced by expression of the CRISPRi system and inhibition of the specific targets, we conduct a proteomics analysis of all sdAb-producing cultures. Proteomics shows that RpoS levels are significantly downregulated and ribosome-associated proteins are significantly upregulated in the growth decoupling CRISPRi-strains compared to the sdAb-expressing control strains after 24 h of growth. We conclude that, although CRISPRi-based repression of nucleotide biosynthesis stalls growth, it does not induce stationary phase response. Furthermore, the maintained ribosome content in the growth decoupled strains may explain the increase in sdAb accumulation. Finally, we apply the two top targets in shake flask fermentation and show that inhibition of *pyrG* increases sdAb yield 2.6-fold compared to the control strain without sgRNA target, with sdAb levels reaching 14.6% of the total protein content.

3.2 Materials and methods

Media and materials

Lysogeny broth agar plates (LB, 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) and 2xYT (16 g/L bactotryptone, 10 g/L yeast extract, 5 g/L NaCl) medium with appropriate antibiotics were used for cultivation and screening during cloning. Ampicillin, chloramphenicol and kanamycin was used with working concentrations of 100 µg/mL, 50 µg/mL and 50 µg/mL, respectively. Growth and production was carried out in M9 minimal medium with 0.1 mM $CaCl_2$, 2.0 mM $MgSO_4$, 1x M9 salts, 1x trace element solution and 1x vitamin solution, which was supplemented with glucose and appropriate antibiotics. The 10x concentrated stock solution of M9 salts consisted of 6.8 g/L Na_2HPO_4 anhydrous, 3 g/L KH_2PO_4, 5 g/L NaCl and 1 g/L NH_4Cl, which had been dissolved in double-distilled water and autoclaved. The 1000x concentrated stock solution of trace elements consisted of 15g/L EDTA(Na_2)*$2H_2O$, 4.5 g/L $ZnSO_4$*$7H_2O$, 0.7 g/L $MnCl_2$*$4H_2O$, 0.3 g/L $CoCl_2$*$6H_2O$, 0.2 g/L $CuSO_4$*$2H_2O$, 0.4 g/L $NaMoO_4$*$2H_2O$, 4.5 g/L $CaCl_2$*$2H_2O$, 3 g/L $FeSO_4$*H_2O, 1 g/L H_3BO_3 and 0.1 g/L KI, which had been dissolved in double-distilled water and sterile filtered. The 1000x concentrated stock solution of vitamins consisted of 10 mg/L pyridoxine HCl, 5 mg/L thiamine HCl, 5 mg/L riboflavin, 5

mg/L nicotinic acid, 5 mg/L calcium D-pantothenate, 5 mg/L 4-aminobenzoic acid, 5 mg/L lipoic acid, 2mg/L biotin, 2 mg/L folic acid and 0.1 mg/L vitamin B12, which had been dissolved in double-distilled water and sterile filtered. Chemicals that were used in the study were purchased from Sigma-Aldrich (Taufkirchen, Germany) and restriction enzymes and PCR polymerases were purchased from ThermoScientific (Waltham, MA, USA). USER enzyme was purchased from BioNordika (Herlev, Denmark).

Plasmid and strain construction

The primers used in this study were ordered from Integrated DNA Technologies (Leuven, Belgium). All primers are listed in Supplementary Table S3.1. Plasmid purification was carried out with the Machery Nagel plasmid purification kit (Dure, Germany) and cell transformation was carried out using the TSS buffer method.[27] *Escherichia coli* DH5α was used for cloning and propagation. All strains and plasmids used in the study are listed in Table S3.2.

The sgRNA plasmids were constructed by Gibson assembly.[28] Primers containing a 20-nucleotide target sgRNA sequence specific for each target gene were used to amplify pSLQ1236.[19] The linear fragment was then assembled according to standard Gibson assembly protocol. The sgRNA sequences were designed using CRISPy-web[29] and are shown in bold in the primer list (Table S3.1). The psdAb-TIR plasmids were constructed by USER cloning.[30,31] The translation-optimized sdAb expression plasmids pET28a-Nanobody®-TIR^{SynEvo1} and pET28a-Nanobody®-TIR^{SynEvo2} and the pClodF13 origin of replication from pCDFDuet (Novagen) was amplified with Phusion U polymerase using primers jl130/131 and jl154/155, respectively. The PCR products were mixed with USER enzyme and incubated for 20 minutes at 37 °C and 20 min at 25 °C, followed by transformation to competent cells. The tetracycline-inducible dCas9 was integrated into the attB-186(O) site in the *E.* coli genome using the Clonetegration method from St-Pierre et al.[32] The selection marker was excised using FLP recombinase and integration was confirmed using colony PCR.

Screening of sgRNA targets

Pre-cultures were prepared by inoculation of biological triplicates of each sgRNA-target strain in a 96-deep well plate (96-DWP) with 800 µL M9 medium supplemented with 0.5% glucose and 0.02% yeast extract (YE), and were grown overnight at 37 °C, 250 rpm. The overnight cultures were inoculated with a 1:100 inoculum ratio (start OD of ~0.03) in two duplicate 96-DWPs with 800 µL M9 medium supplemented with 0.5% glucose, and were grown at 37 °C, 250 rpm for 24 h. After 1 h of growth, 200 ng/mL of anhydrotetracycline (aTc) was added to one of the duplicate 96-DWP to induce the CRISPRi system. Optical density (OD) and fluorescence was measured after 12 and 24 h of growth. OD was measured at 600 nm. For flow cytometry, samples were diluted appropriately and analyzed with a LSRFortessa (Becton Dickinson, San Jose, USA). Expression of GFP was detected using a 488 nm long-pass and a 530/30 nm band-pass filter setting. The forward-scatter and side-scatter was detected as small- and large-angle scatters of the 488 nm laser, respectively. The results were analyzed with FlowJo (Becton, Dickinson & Company, Franklin Lakes, NJ, USA).

Single-domain antibody production

Pre-cultures were prepared by inoculation of biological triplicates of cells transformed with the sdAb and the sgRNA plasmids. Pre-cultures were grown overnight at 37 °C, 250 rpm in 24-DWP in 2.5 mL of M9 medium with 0.5% glucose and 0.02% YE. For the small scale sdAb production experiment, overnight cultures were inoculated in duplicates to an OD of 0.03 in 24-DWPs with 2.5 mL M9 medium with 0.5% glucose. For sdAb production in shake flasks, overnight cultures were inoculated in duplicates to an OD of 0.03 in 250 mL shake flasks with 50 mL M9 medium with 1% glucose. The CRISPRi system was induced in half of the cultures after 1 h using 200 ng/mL aTc. sdAb production was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at OD 0.4. After 24 h, 1 OD unit of culture was harvested and submitted for proteomics analysis. All samples from the deep well plate and the shake flask fermentation experiments, respectively, were run in the same proteomics analysis round, where label-free quantification (LFQ) values of the measurable proteins present in the cell

were determined. The sdAb content (%) for each strain was calculated by dividing the LFQ value of the sdAb with the total LFQ value.

Sample preparation for proteomics analysis

Frozen cells were kept at −80 °C until processing of samples. Thawing of the cells were done on ice and any remaining supernatant was removed after centrifugation at 15,000g for 10 min. While kept on ice, two 3-mm zirconium oxide beads (Glen Mills, NJ, USA) were added to the samples. Immediately after moving the samples away from ice, 100 µl of 95 °C Guanidinium HCl (6 M Guanidinium hydrochloride (GuHCl), 5 mM tris (2-carboxyethyl) phosphine (TCEP), 10 mM chloroacetamide (CAA), 100 mM Tris–HCl pH 8.5) was added to the samples. Cells were disrupted in a Mixer Mill (MM 400 Retsch, Haan, Germany) set at 25 Hz for 5 min at room temperature, followed by 10 min in thermo mixer at 95 ° at 2000 rpm. Any remaining cell debris was removed by centrifugation at 15,000g for 10 min, after which 50 µl of supernatant was collected and diluted with 50 µl of 50 mM ammonium bicarbonate. Based on protein concentration measurements (BSA), 100 µg protein was used for tryptic digestion. Tryptic digestion was carried out at constant shaking (400 rpm) for 8 h, after which 10 µl of 10% TFA was added and samples were ready for StageTipping using C18 as resin (Empore, 3M, USA).

For analysis of the samples, a CapLC system (Thermo Scientific) coupled to an Orbitrap Q-exactive HF-X mass spectrometer (Thermo Scientific) was used. First, samples were captured at a flow of 10 µl/min on a precolumn (µ-precolumn C18 PepMap 100, 5µm, 100Å) and then at a flow of 1.2 µl/min the peptides were separated on a 15 cm C18 easy spray column (PepMap RSLC C18 2µm, 100Å, 150 µmx15cm). The applied gradient went from 4% acetonitrile in water to 76% over a total of 60 minutes. While spraying the samples into the mass spectrometer, the instrument operated in data dependent mode using the following settings: MS-level scans were performed with Orbitrap resolution set to 60,000; AGC Target 3.0e6; maximum injection time 50 ms; intensity threshold 5.0e3; dynamic exclusion 25 sec. Data dependent MS2 selection was performed in Top 20 Speed mode with HCD collision energy set to 28% (AGC target 1.0e4, maximum injection time 22 ms, Isolation window 1.2 m/z).

Proteomics data analysis

For analysis of the thermo rawfiles, Proteome discoverer 2.3 was used with the following settings: Fixed modifications: Carbamidomethyl (C) and Variable modifications: oxidation of methionine residues. First search mass tolerance 20 ppm and a MS/MS tolerance of 20 ppm. Trypsin as enzyme and allowing one missed cleavage. FDR was set at 0.1%. The Match between runs window was set to 0.7 min. Quantification was only based on unique peptides and normalization between samples was based on total peptide amount. For the searches, a protein database consisting of the reference *E. coli* proteome UP000000625 and the sequences of the sdAb[25] and dCas9[19] were used.

Computational analysis and visualization of proteomics data

For further processing and data analysis of the proteome dataset, only proteins with measurements in all samples were used (1,739 proteins for the DWP experiment). Differential expression analysis was performed using the EdgeR package.[33] Gene Ontology (GO) terms[34,35] were obtained from current.geneontology.org/annotations/ecocyc.gaf.gz on September 2nd 2019, and GO analysis was performed by means of the *Piano* package using the method *Stouffer*.[36] *p* values were adjusted for multiple testing using the *Benjamini/Hochberg* approach.

3.3 Results and discussion

Construction and screening of growth decoupling strains targeting purine and pyrimidine biosynthesis

A total of 22 different genes in the nucleotide biosynthesis pathway were selected as targets to investigate the potential of using purine and pyrimidine biosynthesis genes as CRISPRi-based growth switches (Fig. 3.1a). The chosen targets are part of *de novo* purine biosynthesis (*purA*, *purB*, *purC*, *purD*, *purE*, *purF*, *purH*, *purK*, *purL*, *purM*, *purN*, *guaA*, *guaB*), *de novo* pyrimidine biosynthesis (*pyrB*, *pyrC*, *pyrD*, *pyrE*, *pyrF*, *pyrG*, *pyrH*, *ndk*) or the pyrimidine

salvage pathway (*cmk*).[37] sgRNAs targeting the different genes were cloned onto plasmid pSLQ1236 using Gibson cloning, resulting in 21 plasmids (we were not successful at obtaining a cloning construct for the sgRNA targeting *pyrC*). Each sgRNA plasmid was transformed together with pdCas9 into strain MG1655-gfp, harboring a genome integrated GFP under constitutive promoter J23100 inserted 9 bp downstream of *glmS*.[38] An empty sgRNA plasmid with no insert sequence as well as a wild-type *E. coli* strain were used as controls. To compare growth and production of samples with the CRISPRi system induced or uninduced, overnight precultures were split in two and one was induced with anhydrotetracycline (aTc) after 1 h of growth. Samples for measuring growth and fluorescence were taken after 12 and 24 h (Fig. 3.1b).

Results showed that after 12 h, 17 out of 21 induced targets displayed increased GFP production compared to the respective uninduced control, with a fold change between 1.1- and 5.7-fold (Fig 3.1c, upper plot and Fig. S3.1a). Background fluorescence from the wild type control was negligible (data not shown). Inhibition of *pyrG* and *cmk* resulted in the highest GFP production levels. Growth inhibition could be seen in 19 out of the 21 targets, and 15 targets displayed simultaneous inhibition of growth and increase in GFP production (Fig. 3.1c, Fig. S3.1b). After 24 h, 12 out of 21 targets still showed a significant increase in production, with the fold-change decreasing slightly to a range between 1.1- and 4.5-fold (Fig 3.1c, lower plot and Fig. S3.1c). CRISPRi-based repression of *pyrG* and *cmk* still resulted in the highest production. In most cultures, the fluorescence had decreased compared to the 12 h time point. Only 4 out of 21 CRISPRi-induced strains had a lower OD compared to the respective uninduced control (Fig S3.1d); however, this can to large extent be explained by the decrease in OD that uninduced strains displayed between the 12 and 24 h sample points. This decrease could also be seen for the wild-type control strain (data not shown).

Cell size could have a significant impact on protein accumulation, as larger cells can contain higher amounts of protein. However, flow cytometry data of the CRISPRi-induced strains showed that of all the screening targets, only *purK* had an (around 2-fold) increase in cells size compared to the respective uninduced control.

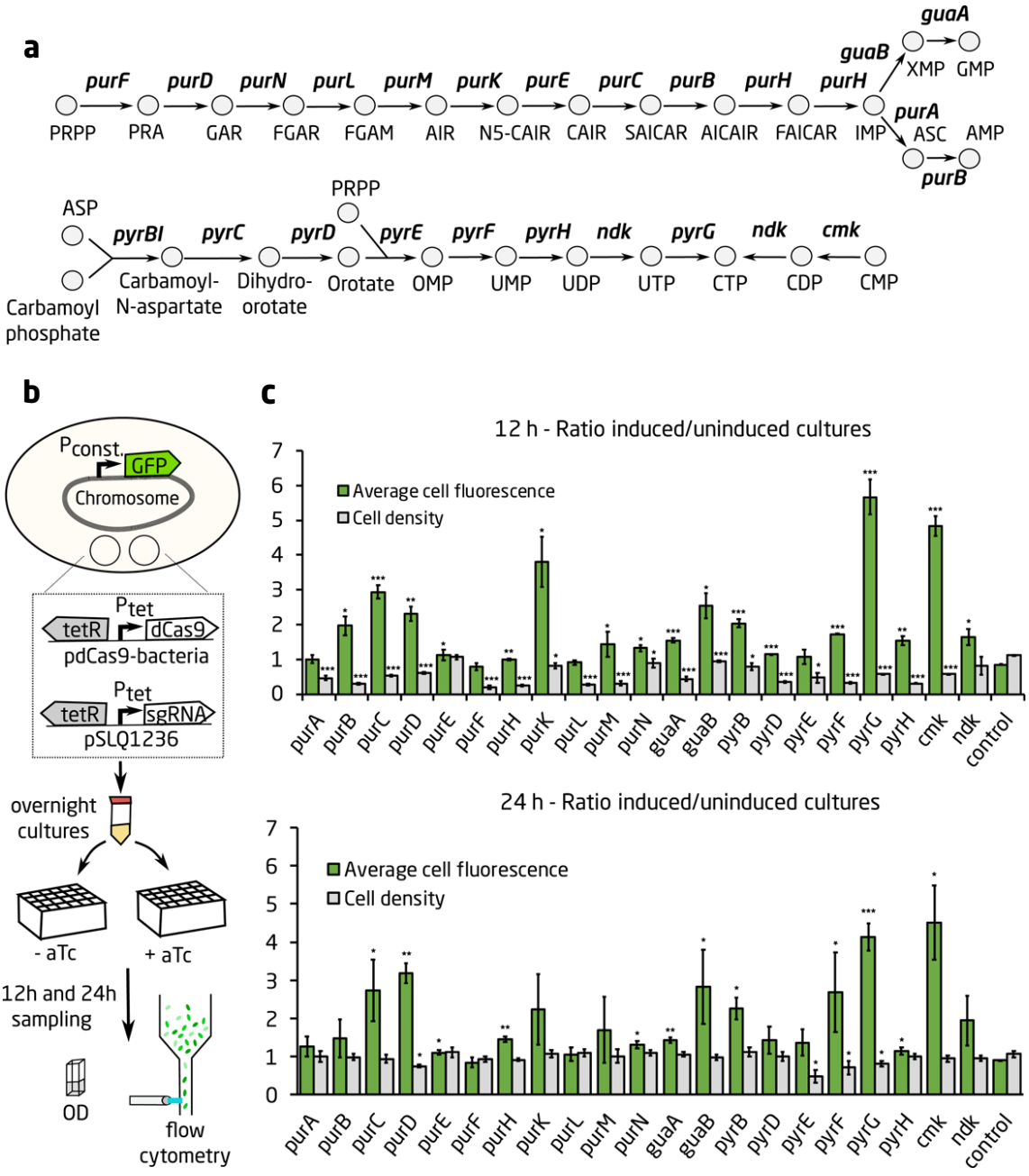


Figure 3.1. Screening the purine and pyrimidine biosynthesis pathways for growth decoupling targets. (a) The sgRNA target genes in the purine and pyrimidine biosynthesis pathways of Escherichia coli. (b) Experimental overview. (c) Ratio of OD and of GFP fluorescence in CRISPRi-induced and uninduced cultures for each of the screened sgRNA targets after 12 h and 24 h of growth. OD and fluorescence were calculated as the average of three biological replicates. Standard deviations between replicates are shown as error bars. A two-tailed t-test was used to check for significant difference between the growth decoupling and control strains; $p < .05$ (*);

Figure 3.1 continued. $p < .001$ (**); $p < .0001$ (***). Abbreviations: PRPP; Phosphoribosyl diphosphate, PRA; 5-Phosphoribosyl-1-amine, GAR; 5-Phosphoribosyl-1-N-glycinamide, FGAR; 5-Phosphoribosyl-1-N-formylglycinamide, FGAM; 5-Phosphoribosyl-1-N-formylglycinamidine, AIR; 5´-Phosphoribosyl-1´-N-(5-amino)imidazole, N5-CAIR; 5´-Phosphoribosyl-1´-N-(5-amino)imidazole-5-N-carboxylate; CAIR; 5´-Phosphoribosyl-1´-N-(5-amino)imidazole-4-carboxylate, SAICAR; 5´-Phosphoribosyl-1´-N-(5-amino)imidazole-4-N-succinocarboxamide, AICAR; 5´-Phosphoribosyl-1´-N-(5-amino)imidazole-4-N-carboxamide, FAICAR; 5´-Phosphoribosyl-1´-N-(5-formylamino)imidazole-4-N-carboxamide, IMP; Inosine 5´-monophosphate, XMP; Xanthosine 5´-monophosphate; GMP; Guanosine 5´-monophosphate, ASC; Adenylosuccinate, AMP; Adenosine 5´-monophosphate, OMP; Orotidine 5´-monophosphate, UMP; Uridine 5´-monophosphate, UDP; Uridine 5´-diphosphate, UTP; Uridine 5´-triphosphate, CTP; Cytidine 5´-triphosphate, CDP; Cytidine 5´-diphosphate, CMP; Cytidine 5´-monophosphate.

Overall, the growth switch targets *pyrG* and *cmk* were the best-performing targets in the screen (Fig. 3.1c, Fig. S3.1e). They were selected for further testing together with *pyrF*, which has previously been shown to work as an efficient growth switch for both protein and biochemical production.[2] Flow cytometry analysis of these strains revealed that the CRISPRi-induced *pyrF* and *cmk* populations had a unimodal fluorescence distribution after 12 h, with *pyrF* showing signs of a slight shift toward bimodality after 24 h (Fig. S3.1e). On the other hand, the *pyrG* population had a bimodal fluorescence distribution at both 12 h and 24 h, where part of the population produced GFP in similar levels as the control, and part of the population produced very high amounts of GFP (Fig. S3.1e). This indicates that the *pyrG* strain is divided into two populations after CRISPR induction, where one consists of a growth-stalled, high-producing cells and the other consists of regularly growing and producing cells.

It is also worth noting that the strain used in our study, MG1655, has a mutation in *rph1* that interferes with expression of *pyrE*, which is located downstream or *rph1*.[39] Therefore, MG1655 is under pyrimidine limitation at

higher growth rates,[39] which could potentially strengthen the growth inhibition efficient of CRISPRi when targeting pyrimidine biosynthesis.

Improving sdAb production using growth decoupling

Next, we applied the *pyrF*, *pyrG* and *cmk* targets for improving production of a commercially relevant protein. Single-domain antibodies (sdAbs or Nanobody®) (Fig. 3.2a) are derived from immunoglobulin-γ antibodies found in camelids.[40] They possess various interesting features compared to the commonly used monoclonal antibodies, such as smaller size, higher solubility and increased stability.[26] sdAbs are commonly produced in *E. coli* as they generally do not require posttranslational modifications.[41] They can be used as they are or fused to chemicals or protein domains, and have a great potential for applications within research, diagnostics and as therapeutics.[26] The first sdAb on the therapeutics market was recently approved for treatment of a blood disorder.[42] We selected an sdAb for which the expression had previously been optimized in a study by Rennig et al.[25] They developed two different translation-optimized versions of the sdAb (pET28a-Nanobody®-TIR^{SynEvo1} and pET28a-Nanobody®-TIR^{SynEvo2}). Both harbored changes in the six nucleotides upstream of the start codon, which significantly improved expression compared to the original construct.[25] To facilitate culturing, the tetR-pTet-dCas9 cassette was integrated into the phage 186 integration site in the genome of MG1655-DE3[43] using pOSIP[32], resulting in strain MG1655-DE3-dCas9. To avoid plasmid incompatibility between the sdAb and sgRNA plasmids, the origin of replication for pET28a-Nanobody®-TIR^{SynEvo1} and pET28a-Nanobody®-TIR^{SynEvo2} were changed to ClodF13, resulting in psdAb-TIR1 and psdAb-TIR2, respectively. MG1655-DE3-dCas9 was transformed with psdAb-TIR1 or psdAb-TIR2 and sgRNA plasmids with targets *pyrF*, *pyrG*, *cmk*. An sgRNA vector without targeting sequence was used a control. Precultures were grown in a 24-DWP with 2.5 mL media overnight. The precultures were inoculated in duplicates into two 24-DWPs with 2.5 mL fresh media. One of these was induced with aTc after 1 h of growth. sdAb production was induced in all cultures at an OD of 0.4 using 1 mM IPTG. Samples were collected for OD and proteomics after 24 h (Fig. 3.2b). The proteomics data for the deep well plate experiment can be found in Supplementary File S3.1.

The CRISPRi-induced cultures showed significant growth inhibition, as the OD reached around half the OD of the uninduced controls (Fig. 3.2c-d, panel 1). The uninduced cultures harboring sgRNA targets all grew to a similar OD as the control strains. Upon induction of the respective target sgRNA, protein levels of PyrF, PyrG and Cmk were decreased to 6%, 35% and 10% respectively, compared to the uninduced control strain in the sdAb-TIR1 strains (Fig 3.3, Fig. S3.2a). For sdAb-TIR2, protein levels of PyrF, PyrG and Cmk were decreased to 7%, 42% and 11%, respectively, compared to the uninduced control strain upon induction of the respective target sgRNA (Fig. S3.2b). This implies that gene silencing was efficient in the *pyrF* and *cmk* strains, but not in the *pyrG* strain. The relatively high levels of *pyrG* expression seen in the induced strains indicates that the bimodally distributed population in Fig. S3.1e could consist of a growth-stalled, high-producing population with little to no expression of *pyrG*, and a regular population that has escaped *pyrG* repression and produces normal levels of PyrG and sdAb. This might for example depend on an inefficient sgRNA design, and/or on that the high metabolic burden of high-producing cells creates a strong selection pressure allowing "CRISPR escapers" to take over the population.

sdAb production was significantly in four of the growth decoupling strains compared to the CRISPRi-induced control strain ($p = .023$, and .003 for *pyrF* and *cmk* in psdAb-TIR1, respectively; $p = .01$, and $< .001$ for *pyrF* and *pyrG* in psdAb-TIR2, respectively) (Fig. 3.2c-d, panel 2). The best-performing target for psdAb-TIR2 was *pyrG* with a 2.7-fold increase in sdAb per OD and an sdAb content of 6.3% of the total protein content, compared to 2.4% in the induced control strain (Fig. 3.2d, panel 2). For psdAb-TIR1, *cmk* had a 2.3-fold increase in sdAb per OD, and a final content of 4.8% sdAb compared to 2.1% in the control strain.

It is not completely clear why different sgRNAs worked better for psdAb-TIR1 and psdAb-TIR2. Inhibition of pyrimidine biosynthesis will lead to alterations of the UTP and CTP pools, and these fluctuations will be different depending on whether *pyrF*, *pyrG* or *cmk* is inhibited (see results in "Pyrimidine pathway expression"). Increases and decreases of the UTP and CTP pools can affect expression of the gene encoding the protein to be

produced, especially if the nucleotide sequence upstream the gene contains T and C residues. Thus, the efficiency of the *pyrF*, *pyrG* and *cmk* targets may vary as the six nucleotides upstream the start codon are different for sdAb-TIR1 and sAb-TIR2 (TGGTAA and GAATAT for sdAb-TIR1 and sAb-TIR2, respectively). This is worth considering when using nucleotide biosynthesis inhibition to increase production of proteins.

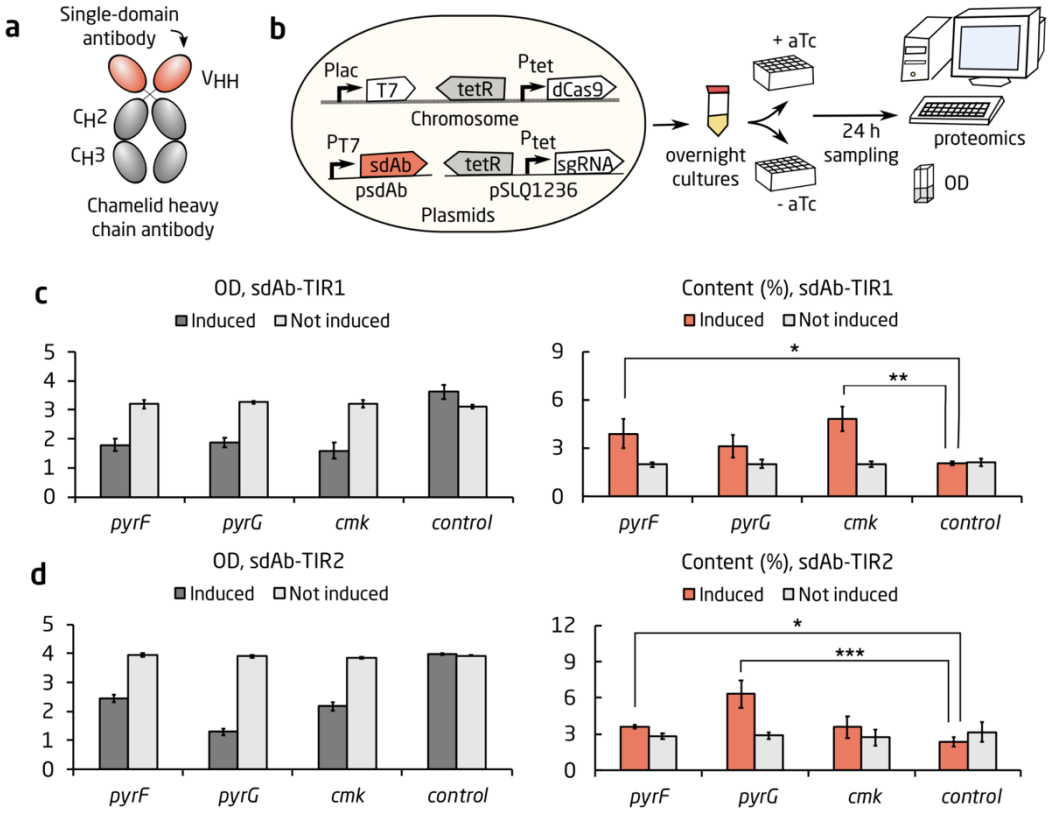


Figure 3.2. sdAb production in deep well plate. Application of the top-performing growth switches pyrF, pyrG and cmk for production of two expression-optimized sdAbs with different translation initiation regions (TIR). (a) sdAbs are derived from the heavy chain of an antibody. (b) Experimental overview. (c) Growth and sdAb production after 24 h for strains harboring psdAb-TIR1 and sgRNA plasmids targeting pyrF, pyrG and cmk. (d) Growth and sdAb production after 24 h for strains harboring psdAb-TIR2 and sgRNA plasmids targeting pyrF, pyrG and cmk. For (c) and (d), the first bar graph shows OD and the second bar graph shows percent sdAb content Cultures where the CRISPRi system was induced are

Figure 3.2 continued. shown in dark grey (OD) or red (sdAb content). Uninduced cultures are shown in bright grey. The values were calculated as an average of three biological replicates. Error bars represent standard deviation of the replicates. A two-tailed t-test was used to check for significant difference between the strains; $p < .05$ (*); $p < .001$ (**); $p < .0001$ (***).

Proteomics analysis of sdAb-producing growth decoupling strains

Samples taken at the 24 h timepoint were used for proteomics analysis of all strains. One OD unit of each culture was harvested and analyzed as described in the Material and methods section. The resulting proteome data set was subjected to differential expression analysis and further analyzed to identify enriched gene ontology (GO) terms. The data set can be found in Supplementary File S3.2.

Pyrimidine pathway expression

Up- and downregulation of genes in the pyrimidine biosynthesis pathway was determined by comparing the sdAb-producing CRISPRi-induced strains harboring an sgRNA plasmid (i.e. growth decoupling strains) to the sdAb-producing CRISPRi-induced strain with the empty control sgRNA plasmid (i.e. control strains) (Table S3.3). The analysis showed that protein levels were differently regulated depending on the specific sgRNA target. Generally, gene expression in the upper part of the pyrimidine biosynthesis pathway was upregulated upon inhibition of *pyrF*, downregulated upon inhibition of *cmk*, and up- or downregulated upon inhibition of *pyrG* (Fig. 3.3). Pyrimidine biosynthesis is known to be controlled by the nucleotide pool through sensing of intracellular levels of UTP and CTP.[44] These pools are expected to vary depending on the specific sgRNA target. In *pyrF*-inhibited strains, both UTP and CTP pools are expected to decrease as PyrF is operating upstream of UTP and CTP synthesis (Fig. 3.1a). Inhibition of *pyrG* and *cmk* should, on the other hand, result in reduced CTP levels and an increase (or maintenance) of the UTP pool,[45] as these genes encode enzymes responsible for converting UTP to CTP (*pyrG*), or are active in the salvage pathway of pyrimidine synthesis (*cmk*) (Fig. 3.1a).

Expression of *pyrBI* and *pyrE* is transcriptionally regulated by intracellular UTP levels through transcription pausing and attenuation.[44,46] When UTP concentrations are high, the UTP-rich transcription pause sites are rapidly transcribed by RNA polymerase, allowing the *pyrBI* and *pyrE* attenuators to form translation-terminating attenuation loops. *pyrBI* is further controlled by reiterative transcription, resulting in an even higher degree of repression and derepression by the UTP pool.[47] As a result, *pyrBI* and *pyrE* expression is upregulated at low concentrations of UTP, and *vice versa*.[44] Analysis of the proteomics data showed a significant upregulation (>2-fold) of both PyrBI and PyrE upon inhibition of *pyrF*, in agreement with the expected decrease in the UTP pool in *pyrF* strains. In the *pyrG* strains, PyrBI was downregulated and PyrE was unaffected, while both proteins were downregulated in the *cmk* strains. As UTP levels have been shown to increase in a *cmk* mutant strain,[45] it is expected that expression of *pyrBI* and *pyrE* decrease when *cmk* is inhibited. The downregulated PyrBI and maintained PyrE levels in the *pyrG* strains indicates that blocking this gene may lead to an increase in UTP, but not enough to enhance *pyrE* expression. PyrBI was generally more strongly induced and repressed compared to PyrE, most likely due to the extra level of regulation that the *pyrBI* operon is under.[47] Expression of *pyrF* has also been shown to increase at low UTP levels, and is likely regulated by UTP-sensitive reiterative transcription.[48]

pyrC and *pyrD* are transcriptionally and translationally regulated by the intracellular CTP pool.[44] When the intracellular ratio of GTP/CTP is low, the initiating transcript nucleotide of *pyrC* and *pyrD* is shifted to a CTP and the mRNAs will form a hairpin loop that prevents the ribosome from binding and translating the genes.[49,50] PyrC and PyrD levels were upregulated more than 2-fold in the *pyrF* strains, and between 1.3-1.8-fold in the *pyrG* strains. Unexpectedly, both PyrC and PyrD were slightly downregulated in the *cmk* strains. This may indicate that inhibition of *cmk* has less impact on CTP levels compared to inhibition of *pyrG*, or that other regulation factors such as PurR-based repression or GTP pool-dependent regulation of *pyrC* and *pyrD* is activate during *pyrG* but not *cmk* repression.[51,52]

Expression of PyrG was upregulated in the *pyrF* strains, while no effect on expression could be seen in the *cmk* strains. Regulation of *pyrG* has not been fully elucidated in *E. coli*, however, the gene seems to be regulated by the CTP pool through start-site switching similar to *pyrC* and *pyrD*.[44] Cmk levels were increased in the *pyrF* and *pyrG*-inhibited strains, which could indicate that expression of the gene is affected by UTP and CTP levels. Not much is known about the transcriptional regulation of *cmk*, except that it is co-transcribed with ribosomal protein S1 (*rpsA*), which is transcriptionally repressed by its own protein product.[51,53]

Overall, the specific regulation pattern seen in the pyrimidine biosynthesis pathway is consistent with existing literature, and in combination with the significantly reduced expression levels of PyrF, PyrG and Cmk in their respective target strains (Fig. 3.3, Fig. S3.2), it shows that the sgRNA and dCas9 are efficiently inhibiting expression of their specific gene target.

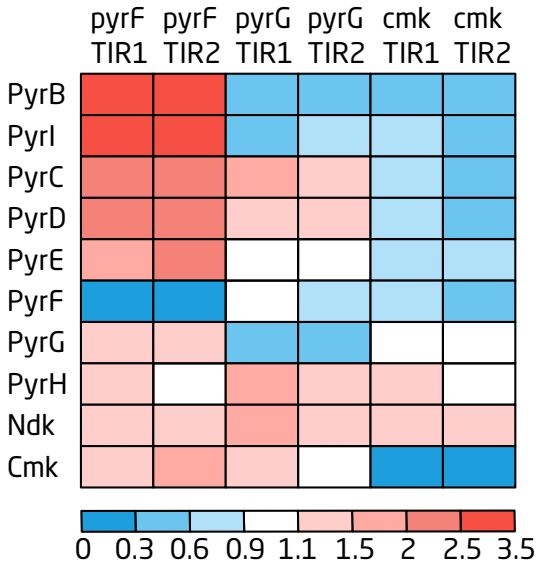


Figure 3.3. Heatmap of the fold-change of protein levels in the pyrimidine biosynthesis pathway of the CRISPRi-induced *pyrF*, *pyrG* and *cmk* growth decoupling strains expressing psdAb-TIR1 and psdAb-TIR2. Fold-change for the growth decoupling strains was calculated by dividing the CRISPRi-induced *pyrF*, *pyrG* and *cmk* strains with the respective (TIR1 or TIR2) CRISPRi-induced control strain harboring the sgRNA control plasmid. The data can be found in Table S3.3.

It is worth noting that a comparison of induced and uninduced dCas9 expression in cells harboring the control sgRNA plasmid revealed that expression of dCas9 and sdAb did not significantly affect protein levels in the pyrimidine pathway. The only significant exception was *pyrE* ($p = 0.0015$), which was slightly downregulated in the control strain harboring psdAb-TIR2 (Table S3.3). A previous study did also not report differential expression of the pyrimidine pathway in CRISPRi-expressing strains with no sgRNA target sequence.[54]

Gene ontology enrichment analysis

Comparison of sdAb-producing CRISPRi-induced strains harboring an sgRNA plasmid (i.e. growth decoupling strains) to the sdAb-producing CRISPRi-induced strain with the empty control sgRNA plasmid (i.e. control strains) revealed that several GO process and compartment terms were significantly up- or downregulated in the growth decoupling strains (Supplementary File S3.2). In total, 1739 proteins were detected in all samples. Of those, 639, 624, 824, 858, 827, and 516 proteins were differentially expressed in *pyrF* sdAb-TIR1, *pyrF* sdAb-TIR2, *pyrG* sdAb -TIR1, *pyrG* sdAb -TIR2, *cmk* sdAb -TIR1 and *cmk* sdAb -TIR2, respectively, compared to the control strain. ($q < .05$). Interestingly, ribosome-associated terms such as ribosomal assembly and cytosolic ribosomal subunit were upregulated in all growth decoupling strains except *pyrF* harboring psdAb-TIR2 (Fig. 3.4). It is well known that ribosome content is closely correlated with growth rate in *E. coli* [55]. As cells reach stationary phase, ribosome content decreases drastically and the protein synthesis rate is reduced to around 20% of the rate during exponential growth.[56] The GO enrichment analysis indicates that, while the control strains have reached stationary phase after 24 h, the growth decoupled strains do not enter stationary phase upon growth inhibition, even though they are no longer growing exponentially. This hypothesis is further corroborated by the relatively low levels of RNA polymerase sigma factor (RpoS or σ^{38}) found in the growth decoupling strains. *rpoS* expression is normally activated in post-exponential and stationary phase in response to a number of inputs, including high cell density, energy limitation, starvation of carbon and nutrients and changes in osmolarity and pH.[57] In the CRISPRi-induced *pyrF*, *pyrG* and *cmk* strains,

RpoS levels were only 7-15% compared to the control strains (Table S3.4). Several proteins known to be under control of RpoS were significantly downregulated compared to the stationary phase control, including for example pyruvate oxidase (PoxB), peroxiredoxin (OsmC) and DNA-protecting starvation protein (Dps)[58] (Table S3.4). As inhibition of growth decrease the amount of catalytic biomass, the overall glucose uptake rate will decrease even if the specific glucose rate is maintained. This will delay glucose depletion and starvation response, which could explain why no stationary phase response is seen in the *pyrF*, *pyrG* and *cmk* strains.

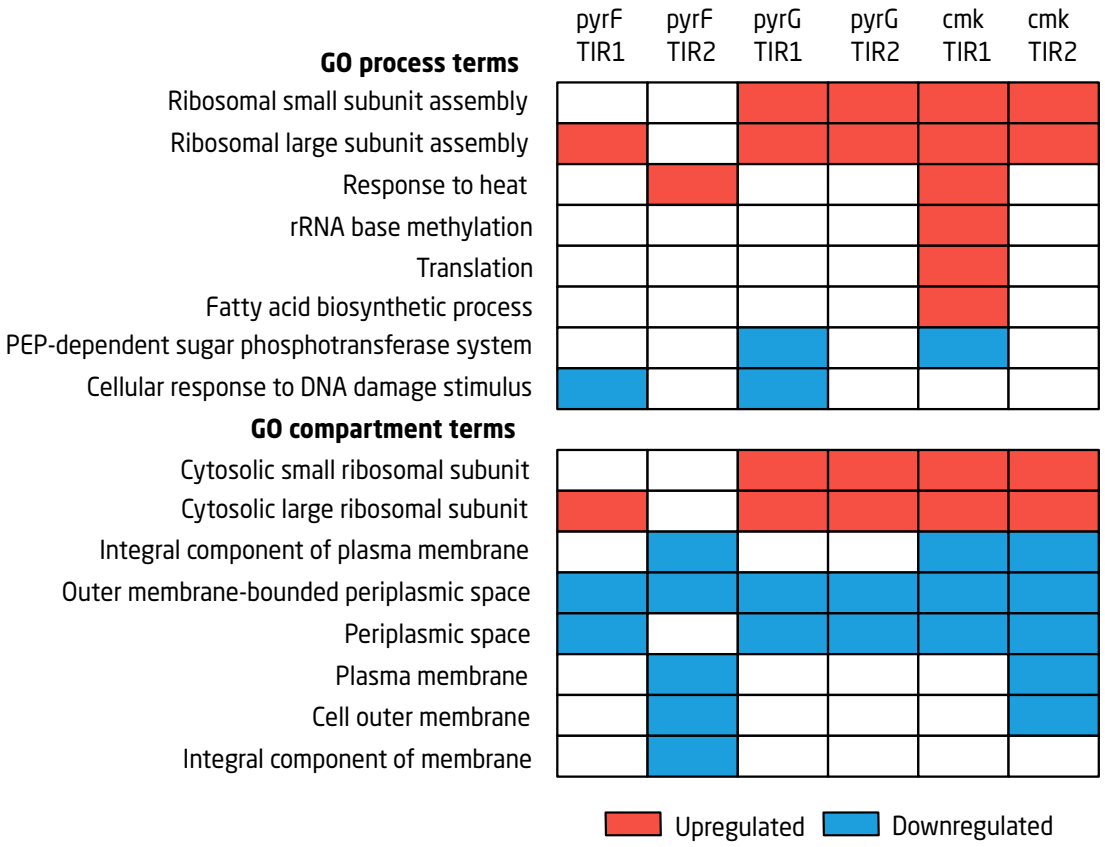


Figure 3.4. Gene ontology enrichment analysis. Gene ontology process and compartment terms found to be significantly upregulated (red, $p < .05$) or downregulated (blue, $p < .05$) in the *pyrF*, *pyrG* and *cmk* growth decoupling strains expressing sdAB-TIR1 or sdAb-TIR2. The differential expression and GO enrichment analysis can be found in Supplementary File 1. Abbreviations: PEP; Phosphoenolpyruvate.

The relatively increased ribosome content may provide an explanation for the increase in sdAb production seen in the growth decoupling strains. With a maintained ribosome availability, growth-inhibited cells maintain their capacity for protein synthesis during an extended amount of time compared to strains entering stationary phase. Upon CRISPRi-based inhibition, the cells continue to produce proteins, but cannot divide due to limited nucleotide availability. This hypothesis is corroborated by a previous study, where we used single cell microfluidics to show that GFP is continuously produced in *pyrF*-inhibited cells.[2] Furthermore, as growth is inhibited, glucose cannot not be used for biomass accumulation during this time, but is available for other metabolic processes.

Nearly all GO terms that were decreased in the CRISPRi-induced *pyrF*, *pyrG* and *cmk* strains are associated with cellular membrane and periplasmic compartments (Fig. 3.4). There may be several reasons for this. First of all, around 14% of RpoS-regulated genes encode membrane proteins.[57] Downregulated proteins under control of RpoS include for example transport proteins PotF and UgpB of the ABC superfamily, and putative transport- and membrane proteins YdcS and YeaY[58] (Table S3.4). Expression of membrane-associated proteins may also be directly altered by the decrease in pyrimidine supply, as these nucleotides are required for phospholipid synthesis. It has previously been shown that a *cmk* mutant strain with decreased dCTP and CTP pools becomes cold sensitive and displays altered expression of outer membrane porins *ompC* and *ompF*.[45] Furthermore, overexpression of dCas9 can be toxic and alter gene expression in *E. coli*,[54] which could affect the GO enrichment analysis. A recent study found that expression of dCas9 leads to significant downregulation of cell and membrane biogenesis and translation, and an upregulation of transcription and amino acid and carbohydrate metabolism.[54] However, a comparison of the CRISPRi-induced and non-induced control strains in our study did not yield any enriched GO terms (Supplementary File S3.2), and we could not identify any overlap in significantly up- or downregulated genes and proteins between our data and the data set from Cho et al. Our GO enrichment analysis was based on a comparison of growth decoupling and control strains that both express dCas9, which should normalize for effects occurring due to

dCas9 expression. Nevertheless, there is still a possibility of unexpected dCas9 effects when there is no sgRNA present to guide DNA binding.[59]

It should be noted that membrane and periplasmic proteins are not always reliably quantified with the proteomics method used in this study. However, the fact that membrane protein expression is normally upregulated in stationary phase cells, and that levels of measured membrane proteins is similar within each replicate (Table S3.4), implies that there is a real decrease when comparing the growth decoupling and the control strains.

Inhibition of *pyrG* was not as efficient as that of *pyrF* or *cmk*. Considering the limited decrease of PyrG protein in the cells, and the fact that GFP distribution in this strain was clearly bimodal (Fig. S3.1e), the proteomics analysis may not give the full picture of up- and downregulation of proteins that would occur from complete inhibition of *pyrG* expression. However, the GO terms that were identified overlapped with the ones identified for the *pyrF* and *cmk* strains, and we do not consider the limited *pyrG* inhibition to be crucial for the outcome of the proteomics analysis.

Polar effects

Expression of dCas9 and sgRNAs may result in unspecific binding and off-target effects. Polar effects due to CRISPRi inhibition of *pyrF* resulted in significant downregulation of YciH protein levels (Table S3.5). Knock-out of *yciH* has previously been shown to increase expression of 66 genes, and decrease expression of 20 genes in *E. coli*;[60] however none of these proteins were found to be significantly up- or downregulated in our study. Polar effects of *pyrG* inhibition resulted in downregulation of *eno*, located downstream of *pyrG* (Table S3.5). *eno* encodes enolase, which catalyzes conversion of 2-phosphoglycerate to phosphoenolpyruvate, and is also involved in processing of RNA and degradation of mRNA.[61] It is difficult to predict exactly how these polar effects might have impacted the results of this study. However, in our previous library screen, repression of *eno* and *yciH* did not result in significant inhibition of growth or an increase in GFP production, indicating that the results seen for *pyrF* and *pyrG* inhibition are not affected by downstream *eno* and *yciH* inhibition, respectively.[24]

Scale-up of sdAb production to shake flask fermentation

We applied the best-performing targets *cmk* for psdAb-TIR1 and *pyrG* for psdAb-TIR2 for scale-up of sdAb production to small batch fermentation in shake flasks. The experiment was carried out as shown in Fig. 3.2a, but cultures were grown in 250 mL shake flasks with 50 mL media. OD was measured continuously during the experiment (Fig. S3.3a). sdAb expression was induced in all cultures at an OD of 0.4, and samples for proteomics were taken after 24 h.

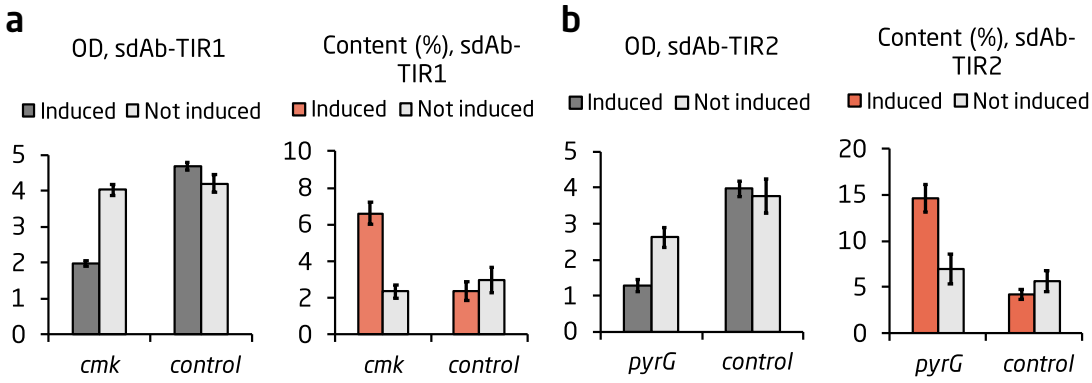


Figure 3.5. sdAb production in small batch fermentation. (a) Growth and sdAb production after 24 h for strains harboring psdAb-TIR1 and the *cmk* sgRNA plasmid. (b) Growth and sdAb production after 24 h for strains harboring psdAb-TIR2 and the *pyrG* sgRNA plasmid. Cultures where the CRISPRi system was induced are shown in dark grey (OD), red (sdAb content). Uninduced cultures are shown in bright grey. The values were calculated as an average of three biological replicates. Error bars represent standard deviation of the replicates.

The replicates with induced CRISPRi expression showed significant growth inhibition with a final OD around half of that in the control strains (Fig. 3.5a-b, panel 1). Protein levels of Cmk and PyrG in the strains with sgRNAs targeting *pyrG* and *cmk* were decreased to 2.7% and 18%, respectively, compared to the respective induced control strains (Fig. S3.3b). sdAb production was significantly improved in the growth inhibited strains ($p < .001$ for both *cmk* and *pyrG* compared to the induced control), with a 2.2-

and 2.6-fold increase in sdAb per OD in *cmk* and *pyrG*, respectively (Fig. 3.5b, panel 2). The final sdAb content reached 6.6% in the *cmk* strain harboring psdAb-TIR1, and 14.6% in the *pyrG* strain harboring psdAb-TIR2 (Fig. 3.5b, panel 2).

3.4 Conclusions

The nucleotide biosynthesis pathway is an unexpected CRISPRi target for improving protein production, as it provides precursor nucleotides for RNA synthesis. However, inhibiting the essential *de novo* pyrimidine biosynthesis after initial biomass accumulation still enables cells to supply nucleotides from turnover of RNA and DNA through the pyrimidine salvage pathway, and inhibition of the pyrimidine salvage pathway enables supply of nucleotides from the *de novo* pyrimidine biosynthesis pathway. CRISPRi-based inhibition of *pyrF* and *thyA* has previously been shown to inhibit growth and improve production of both GFP and mevalonate,[2] and in that study it was further shown that growth decoupled strains remain growth inhibited and metabolically active for up to 48 h.

In this study, we showed that sdAb production per OD can be increased up to 2.6-fold upon CRISPRi-based inhibition of nucleotide biosynthesis. This means that the overall titer of sdAb in shake flask fermentation of the best-performing growth decoupling strain *cmk* was almost doubled, even though the OD was half of that in the control. The maintained high capacity for protein synthesis and lack of stationary phase response shows that inhibition of nucleotide biosynthesis is a useful approach to increase protein production. Besides the maintained ribosome availability, there are probably other underlying mechanisms in play that enables the growth decoupling strains to maintain or increase production although growth is inhibited. It is possible that the nucleotides supplied from RNA turnover are sufficient to support continued protein synthesis, but not to support DNA replication and cell growth. It would be highly interesting to elucidate metabolic adjustments upstream of protein synthesis, such as changes in metabolic flux through glycolysis and other pathways, using ^{13}C metabolic flux analysis and metabolomics. The occurrence of complete growth and production

decoupling could further be confirmed by ribosomal activity assays of cells in the growth-inhibited state.[62]

Future efforts should also focus on generating strains that are more industrially applicable, where growth decoupling can be achieved in an autoinducible manner without the use of CRISPRi. This can for example be done by engineering temperature-sensitive versions of the PyrF, PyrG and Cmk enzymes, so that growth can be inhibited by changing cultivation temperature, or by controlling expression of *pyrF*, *pyrG* and *cmk* with promoters that automatically turn off when the desired cell density has been reached.

3.5 Acknowledgements

The plasmids pSLQ1236 and pdCas9-bacteria were a kind gift from Stanley Qi. The pOSIP KO and the PE-FLP plasmids were a gift from Drew Endy & Keith Shearwin (Addgene plasmids # 45985 and 45978, respectively). The sdAb production plasmids were a gift from Morten Nørholm. We would like to thank the Novo Nordisk Foundation for funding through grant no. NNF10CC1016517. JL was funded by the Novo Nordisk Foundation Copenhagen Bioscience PhD Programme, grant no. NNF16CC0020908. NRW was funded by the Innovation Foundation Denmark, case no. 7038-00165B and the Novo Nordisk R&D STAR Fellowship programme.

3.6 References for Chapter 3

1. Burg, J. M. *et al.* Large-scale bioprocess competitiveness: the potential of dynamic metabolic control in two-stage fermentations. *Curr. Opin. Chem. Eng.* **14**, 121–136 (2016).
2. Li, S. *et al.* Enhanced protein and biochemical production using CRISPRi-based growth switches. *Metab. Eng.* **38**, 274–284 (2016).
3. Durante-Rodríguez, G., De Lorenzo, V. & Nikel, P. I. A Post-translational Metabolic Switch Enables Complete Decoupling of Bacterial Growth from Biopolymer Production in Engineered Escherichia coli. *ACS Synth. Biol.* **7**, 2686–2697 (2018).
4. Willrodt, C., Hoschek, A., Bühler, B., Schmid, A. & Julsing, M. K. Decoupling production from growth by magnesium sulfate limitation boosts de novo limonene production. *Biotechnol. Bioeng.* **113**, 1305–1314 (2016).
5. Lemmerer, M. *et al.* Decoupling of recombinant protein production from Escherichia coli cell growth enhances functional expression of plant Leloir glycosyltransferases. *Biotechnol. Bioeng.* **116**, 1259–1268 (2019).
6. Huber, R., Roth, S., Rahmen, N. & Büchs, J. Utilizing high-throughput experimentation to enhance specific productivity of an E.coli T7 expression system by phosphate limitation. *BMC Biotechnol.* **11**, (2011).
7. Venayak, N., von Kamp, A., Klamt, S. & Mahadevan, R. MoVE identifies metabolic valves to switch between phenotypic states. *Nat. Commun.* **9**, 1–9 (2018).
8. Li, S., Jendresen, C. B. & Nielsen, A. T. Increasing production yield of tyrosine and mevalonate through inhibition of biomass formation. *Process Biochem.* **51**, 1992–2000 (2016).
9. Matsuda, F. *et al.* Magnesium starvation improves production of malonyl-CoA-derived metabolites in Escherichia coli. *Metab. Eng.* **52**, 215–223 (2018).
10. Ge, X. Y., Xu, Y., Chen, X. & Zhang, L. Y. Regulation of metabolic flux in Lactobacillus casei for lactic acid production by overexpressed ldhL gene with two-stage oxygen supply strategy. *J. Microbiol. Biotechnol.* **25**, 81–88 (2015).
11. Sun, Y. *et al.* Two-stage fermentation for 2-ketogluconic acid production by Klebsiella pneumoniae. *J. Microbiol. Biotechnol.* **24**, 781–787 (2014).
12. Harder, B. J., Bettenbrock, K. & Klamt, S. Temperature-dependent

dynamic control of the TCA cycle increases volumetric productivity of itaconic acid production by Escherichia coli. *Biotechnol. Bioeng.* **115**, 156–164 (2018).
13. Lynch, M. D. Into new territory: improved microbial synthesis through engineering of the essential metabolic network. *Curr. Opin. Biotechnol.* **38**, 106–111 (2016).
14. Brockman, I. M. & Prather, K. L. J. Dynamic knockdown of E. coli central metabolism for redirecting fluxes of primary metabolites. *Metab. Eng.* **28**, 104–113 (2015).
15. Darlington, A. P. S., Kim, J., Jiménez, J. I. & Bates, D. G. Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes. *Nat. Commun.* **9**, (2018).
16. Mairhofer, J., Striedner, G., Grabherr, R. & Wilde, M. Patent no. WO 2016/174105 A1 Uncoupling growth and protein production. (2016).
17. Lo, T. M., Chng, S. H., Teo, W. S., Cho, H. S. & Chang, M. W. A Two-Layer Gene Circuit for Decoupling Cell Growth from Metabolite Production. *Cell Syst.* **3**, 133–143 (2016).
18. Santala, S., Efimova, E. & Santala, V. Dynamic decoupling of biomass and wax ester biosynthesis in Acinetobacter baylyi by an autonomously regulated switch. *Metab. Eng. Commun.* **7**, (2018).
19. Larson, M. H. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).
20. Cress, B. F. *et al.* CRISPRi-mediated metabolic engineering of E. coli for O-methylated anthocyanin production. *Microb. Cell Fact.* **16**, 1–14 (2017).
21. Chang, Y., Su, T., Qi, Q. & Liang, Q. Easy regulation of metabolic flux in Escherichia coli using an endogenous type I-E CRISPR-Cas system. *Microb. Cell Fact.* **15**, (2016).
22. Tian, T., Kang, J. W., Kang, A. & Lee, T. S. Redirecting Metabolic Flux via Combinatorial Multiplex CRISPRi-Mediated Repression for Isopentenol Production in Escherichia coli. *ACS Synth. Biol.* **8**, 391–402 (2019).
23. Shabestary, K. *et al.* Targeted Repression of Essential Genes To Arrest Growth and Increase Carbon Partitioning and Biofuel Titers in Cyanobacteria. *ACS Synth. Biol.* **7**, 1669–1675 (2018).
24. Li, S. *et al.* Genome-wide CRISPRi-based identification of targets for decoupling growth from production. *ACS Synth. Biol.* **9**, (2020).
25. Rennig, M. *et al.* TARSyn: Tunable Antibiotic Resistance Devices Enabling Bacterial Synthetic Evolution and Protein Production. *ACS Synth. Biol.* **7**, 432–442 (2018).

26. Wesolowski, J. *et al.* Single domain antibodies: Promising experimental and therapeutic tools in infection and immunity. *Med. Microbiol. Immunol.* **198**, 157–174 (2009).
27. Chung, C. T., Niemela, S. L. & Miller, R. H. One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci.* **86**, 2172–2175 (2006).
28. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
29. Blin, K., Ebdrup, L., Weber, T. & Lee, S. Y. CRISPy-web : An online resource to design sgRNAs for CRISPR applications. *Synth. Syst. Biotechnol.* **1**, 118–121 (2016).
30. Cavaleiro, A. M., Kim, S. H., Seppälä, S., Nielsen, M. T. & Nørholm, M. H. H. Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning. *ACS Synth. Biol.* **4**, 1042–1046 (2015).
31. Nour-Eldin, H. H., Hansen, B. G., Nørholm, M. H. H., Jensen, J. K. & Halkier, B. A. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res.* **34**, 1–8 (2006).
32. St-Pierre, F. *et al.* Complete Methods for 'One-step cloning and chromosomal integration of DNA'. *ACS Synth. Biol.* **2**, 537–541 (2013).
33. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
34. Ashburner, M. *et al.* Gene Ontology : tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
35. The Gene Ontology Consortium. The Gene Ontology Resource : 20 years and still GOing strong. *Nucleic Acids Res.* **47**, 330–338 (2019).
36. Väremo, L., Nielsen, J. & Nookaew, I. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Res.* **41**, 4378–4391 (2013).
37. Martinussen, J., Willemoës, M. & Kilstrup, M. Nucleotide Metabolism. in *Comprehensive Biotechnology* (ed. Murray Moo-Young) **1**, 91–108 (Elsevier B.V., 2011).
38. Bonde, M. T. *et al.* Predictable tuning of protein expression in bacteria. *Nat. Methods* **13**, 233–236 (2016).
39. Jensen, K. F. The Escherichia coli K-12 'wild types' W3110 and MG1655 have an rph frameshift mutation that leads to pyrimidine starvation due to low pyrE expression levels. *J. Bacteriol.* **175**, 3401–3407 (1993).

40. Hamer-Casterman Atarchouch, T, C. *et al.* Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446–448 (1993).
41. Fernandes, C. F. C. *et al.* Camelid single-domain antibodies as an alternative to overcome challenges related to the prevention, detection, and control of neglected tropical diseases. *Front. Immunol.* **8**, 1–8 (2017).
42. Chanier, T. & Chames, P. Nanobody Engineering: Toward Next Generation Immunotherapies and Immunoimaging of Cancer. *Antibodies* **8**, 13 (2019).
43. Mundhada, H., Schneider, K., Christensen, H. B. & Nielsen, A. T. Engineering of high yield production of L-serine in Escherichia coli. *Biotechnol. Bioeng.* **113**, 807–816 (2016).
44. Turnbough, C. L. & Switzer, R. L. Regulation of Pyrimidine Biosynthetic Gene Expression in Bacteria: Repression without Repressors. *Microbiol. Mol. Biol. Rev.* **72**, 266–300 (2008).
45. Fricke, J., Neuhard, J., Kelln, R. A. & Pedersen, S. The cmk gene encoding cytidine monophosphate kinase is located in the rpsA operon and is required for normal replication rate in Escherichia coli. *J. Bacteriol.* **177**, 517–523 (1995).
46. Turnbough, C. L., Hicks, K. L. & Donahue, J. P. Attenuation control of pyrBI operon expression in Escherichia coli K-12. *Proc. Natl. Acad. Sci.* **80**, 368–372 (1983).
47. Roland, K. L., Liu, C. & Turnbough, C. L. Role of the ribosome in suppressing transcriptional termination at the pyrBI attenuator of Escherichia coli K-12. *Proc. Natl. Acad. Sci.* **85**, 7149–7153 (1988).
48. Liu, C., Heath, L. S. & Turnbough, C. L. Regulation of pyrBI operon expression in Escherichia coli by UTP-sensitive reiterative RNA synthesis during transcriptional initiation. *Genes Dev.* **8**, 2904–2912 (1994).
49. Wilson, H. R., Archer, C. D., Liu, J. & Turnbough, C. L. Translational Control of pyrC Expression Mediated by Nucleotide- Sensitive Selection of Transcriptional Start Sites in Escherichia coli. *J. Bacteriol.* **174**, 514–524 (1992).
50. Sørensen, K. I., Baker, K. E., Kelln, R. & Neuhard, J. Nucleotide Pool-Sensitive Selection of the Transcriptional Start Site In Vivo at the Salmonella typhimunium pyrC and pyrD Promoters. *J. Bacteriol.* **175**, 4137–4144 (1993).
51. Jensen, K. F., Dandanell, G., Hove-Jensen, B. & Willemoës, M. Nucleotides, Nucleosides, and Nucleobases. *EcoSal Plus* (2013). doi:10.1128/ecosalplus.3.6.2
52. Jensen, K. F. Regulation of Salmonella typhimurium pyr Gene Expression : Effect of Changing Both Purine and Pyrimidine

Nucleotide Pools. *J. Gen. Microbiol.* **135**, 805–815 (1989).
53. Skouv, J., Schnier, J., Rasmussen, M. D., Subramanian, A. R. & Pedersen, S. Ribosomal protein S1 of Escherichia coli is the effector for the regulation of its own synthesis. *J. Biol. Chem.* **265**, 17044–17049 (1990).
54. Cho, S. *et al.* High-Level dCas9 Expression Induces Abnormal Cell Morphology in Escherichia coli. *ACS Synth. Biol.* **7**, 1085–1094 (2018).
55. Miura, A., Krueger, J. H., Itoh, S., de Boer, H. A. & Nomura, M. Growth-rate-dependent regulation of ribosome synthesis in E. coli: Expression of the lacZ and galK genes fused to ribosomal promoters. *Cell* **25**, 773–782 (1981).
56. Reeve, C. A., Amy, P. S. & Matin, A. Role of protein synthesis in the survival of carbon-starved Escherichia coli K-12. *J. Bacteriol.* **160**, 1041–1046 (1984).
57. Hengge, R. Stationary-Phase Gene Regulation in Escherichia coli. *EcoSal Plus* **4**, (2011).
58. Weber, H., Polen, T., Heuveling, J., Wendisch, V. F. & Hengge, R. Genome-Wide Analysis of the General Stress Response Network in Escherichia coli: RpoS-Dependent Genes, Promoters, and Sigma Factor Selectivity. *J. Bacteriol.* **187**, 1591–1603 (2005).
59. Zhang, S. & Voigt, C. A. Engineered dCas9 with reduced toxicity in bacteria: Implications for genetic circuit design. *Nucleic Acids Res.* **46**, 11115–11125 (2018).
60. Osterman, I. A. *et al.* A bacterial homolog YciH of eukaryotic translation initiation factor eIF1 regulates stress-related gene expression and is unlikely to be involved in translation initiation fidelity A bacterial homolog YciH of eukaryotic translation initiation factor eIF1. *RNA Biol.* **12**, 966–971 (2015).
61. Nurmohamed, S., Adam, R., Robinson, V. & Luisi, B. F. Molecular recognition between Escherichia coli enolase and ribonuclease E. *Acta Crystallogr.* **66**, 1036–1040 (2010).
62. Li, G. W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**, 624–635 (2014).

3.7 Supplementary materials

Supplementary Files S3.1 and S3.2 are presented as Excel files, and can be found at https://doi.org/10.1002/bit.27536.

Supplementary File S3.1. Proteome data set for the sdAb production experiments.

Supplementary File S3.2. Differential expression analysis of the sdAb production in deep well plates.

Table S3.1. Primers used in the study. sgRNA target sequence is marked in bold.

Nr.	Primer sequence 5'- 3'	Description
jl37	**TTTACCTTCGTCACCCCATT**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Forward insert of purA sgRNA into pSLQ1236
jl38	**AATGGGGTGACGAAGGTAAA**ACTAGTCTTTTTCTCTATCACTGATAGGGA	Reverse insert of purA sgRNA into pSLQ1236
jl39	**TCCATCGACAGGGGAAACGG**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Forward insert of purB sgRNA into pSLQ1236
jl40	**CCGTTTCCCCTGTCGATGGA**ACTAGTCTTTTTCTCTATCACTGATAGGGA	Reverse insert of purB sgRNA into pSLQ1236
jl41	**CGGGTTTTCCGTGCTGTATA**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Forward insert of purC sgRNA into pSLQ1236
jl42	**TATACAGCACGGAAAACCCG**ACTAGTCTTTTTCTCTATCACTGATAGGGA	Reverse insert of purC sgRNA into pSLQ1236
jl43	**CGGCGACTGGGCCGCTTTCC**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Forward insert of purD sgRNA into pSLQ1236
jl44	**GGAAAGCGGCCCAGTCGCCG**ACTAGTCTTTTTCTCTATCACTGATAGGGA	Reverse insert of purD sgRNA into pSLQ1236

jl45	**GACACGCGCCGGATTATTGC**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Forward insert of purE sgRNA into pSLQ1236
jl46	**GCAATAATCCGGCGCGTGTC**ACTAGTCTTTTCTCTATCACTGATAGGGA	Reverse insert of purE sgRNA into pSLQ1236
jl47	**TCATAAATCGACTGGTTAAC**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Forward insert of purF sgRNA into pSLQ1236
jl48	**GTTAACCAGTCGATTTATGA**ACTAGTCTTTTCTCTATCACTGATAGGGA	Reverse insert of purF sgRNA into pSLQ1236
jl49	**AAACACTGAGCAGAGCGCGG**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Forward insert of purH sgRNA into pSLQ1236
jl50	**CCGCGCTCTGCTCAGTGTTT**ACTAGTCTTTTCTCTATCACTGATAGGGA	Reverse insert of purH sgRNA into pSLQ1236
jl51	**GGCCTAACTGCCCGTTACCG**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Forward insert of purK sgRNA into pSLQ1236
jl52	**CGGTAACGGGCAGTTAGGCC**ACTAGTCTTTTCTCTATCACTGATAGGGA	Reverse insert of purK sgRNA into pSLQ1236
jl53	**ATTCGGAATGCCGACAGTGC**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Forward insert of purL sgRNA into pSLQ1236
jl54	**GCACTGTCGGCATTCCGAAT**ACTAGTCTTTTCTCTATCACTGATAGGGA	Reverse insert of purL sgRNA into pSLQ1236
jl55	**GGCATCTTTGTAGCTAAGAG**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Forward insert of purM sgRNA into pSLQ1236
jl56	**CTCTTAGCTACAAAGATGCC**ACTAGTCTTTTCTCTATCACTGATAGGGA	Reverse insert of purM sgRNA into pSLQ1236
jl57	**CTGTAAATTACTTCCGTTGC**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Forward insert of purN sgRNA into pSLQ1236
jl58	**GCAACGGAAGTAATTTACAG**ACTAGTCTTTTCTCTATCACTGATAGGGA	Reverse insert of purN sgRNA into pSLQ1236
jl59	**GAGAACCGAAGTCCAGAATG**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Forward insert of guaA sgRNA into pSLQ1236

jl60	**CATTCTGGACTTCGGTTCTC**ACTAG TCTTTTCTCTATCACTGATAGGGA	Reverse insert of guaA sgRNA into pSLQ1236
jl61	**CGGTAGAGTGAGCAGGAACG**GTTTT AGAGCTAGAAATAGCAAGTTAAAAT AAGGC	Forward insert of guaB sgRNA into pSLQ1236
jl62	**CGTTCCTGCTCACTCTACCG**ACTAG TCTTTTCTCTATCACTGATAGGGA	Reverse insert of guaB sgRNA into pSLQ1236
jl63	**ATGATATGTTTCTGATATAG**GTTTT AGAGCTAGAAATAGCAAGTTAAAAT AAGGC	Forward insert of pyrB sgRNA into pSLQ1236
jl64	**CTATATCAGAAACATATCAT**ACTAGT CTTTTCTCTATCACTGATAGGGA	Reverse insert of pyrB sgRNA into pSLQ1236
jl65	**GCGGCGGATCTTTAATACCT**GTTTT AGAGCTAGAAATAGCAAGTTAAAAT AAGGC	Forward insert of pyrC sgRNA into pSLQ1236
jl66	**AGGTATTAAAGATCCGCCGC**ACTAG TCTTTTCTCTATCACTGATAGGGA	Reverse insert of pyrC sgRNA into pSLQ1236
jl67	**AAAAGGGCTTTACGAACGAA**GTTTT AGAGCTAGAAATAGCAAGTTAAAAT AAGGC	Forward insert of pyrD sgRNA into pSLQ1236
jl68	**TTCGTTCGTAAAGCCCTTTT**ACTAG TCTTTTCTCTATCACTGATAGGGA	Reverse insert of pyrD sgRNA into pSLQ1236
jl69	**TAAGCGCAAATTCAATAAAC**GTTTTA GAGCTAGAAATAGCAAGTTAAAATA AGGC	Forward insert of pyrE sgRNA into pSLQ1236
jl70	**GTTTATTGAATTTGCGCTTA**ACTAG TCTTTTCTCTATCACTGATAGGGA	Reverse insert of pyrE sgRNA into pSLQ1236
jl71	**AGGAGAATTCGTAACAGCGC**GTTTT AGAGCTAGAAATAGCAAGTTAAAAT AAGGC	Forward insert of pyrF sgRNA into pSLQ1236
jl72	**GCGCTGTTACGAATTCTCCT**ACTAG TCTTTTCTCTATCACTGATAGGGA	Reverse insert of pyrF sgRNA into pSLQ1236
jl73	**GGCAATGCCTTTACCCAGAG**GTTTT AGAGCTAGAAATAGCAAGTTAAAAT AAGGC	Forward insert of pyrG sgRNA into pSLQ1236
jl74	**CTCTGGGTAAAGGCATTGCC**ACTAG TCTTTTCTCTATCACTGATAGGGA	Reverse insert of pyrG sgRNA into pSLQ1236

jl75	**TTTATAGACGGGTTTTGCAT**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Forward insert of pyrH sgRNA into pSLQ1236
jl76	**ATGCAAAACCCGTCTATAAA**ACTAGTCTTTTCTCTATCACTGATAGGGA	Reverse insert of pyrH sgRNA into pSLQ1236
jl77	**GGCCATCAATGGTAATAACC**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Forward insert of cmk sgRNA into pSLQ1236
jl78	**GGTTATTACCATTGATGGCC**ACTAGTCTTTTCTCTATCACTGATAGGGA	Reverse insert of cmk sgRNA into pSLQ1236
jl79	**ACGTTTTTTGCTACCGCGTT**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Forward insert of ndk sgRNA into pSLQ1236
jl80	**AACGCGGTAGCAAAAAACGT**ACTAGTCTTTTCTCTATCACTGATAGGGA	Reverse insert of ndk sgRNA into pSLQ1236
jl130	ACATGTTC**U**TTCCTGCGTTATCCC	Forward primer for amplifying psdAb-TIR1/2 to change ori to CloDF13
jl131	ATCTTTTCUACGGGGTCTGACGC	Reverse primer for amplifying psdAb-TIR1/2 to change ori to CloDF13
jl154	AGAAAAGAUCAAATAGCTAGCTCACTCGG	Forward primer for amplifying pCDF-Duet CloDf13 origin
jl155	AGAACATGUAAATCTAGAGCGGTTCAGTAG	Reverse primer for amplifying pCDF-Duet CloDf13 origin

Table S3.2. Strains and plasmids used in the study. The name used to refer to a specific strain or plasmid in the main text is written in bold.

Strain	Description	Reference
E. coli NEB 5-alpha	*fhuA2* Δ*(argF-lacZ)U169 phoA glnV44 Φ80* Δ*(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*; cloning strain	New England Biolabs
E. coli K12 MG1655	F- lambda- *ilvG- rfb-50 rph-1*	Lab collection
SIJ19	**MG1655-gfp**: MG1655 Bb_J23100 *gfp* ::FRT	(Bonde et al., 2016)
E.coli MG1655 (DE3)	**MG1655-DE3**: MG1655 λ*(DE3)*	(Mundhada et al., 2016)
JL100	**MG1655-DE3-dCas9**: MG1655 λ*(DE3)* attB-186(O): tetR-Ptet-dCas9	This study
JL86	MG1655-gfp with psgRNA-purA and pdCas9	This study
JL87	MG1655-gfp with psgRNA-purB and pdCas9	This study
JL88	MG1655-gfp with psgRNA-purC and pdCas9	This study
JL89	MG1655-gfp with psgRNA-purD and pdCas9	This study
JL90	MG1655-gfp with psgRNA-purE and pdCas9	This study
JL91	MG1655-gfp with psgRNA-purF and pdCas9	This study
JL92	MG1655-gfp with psgRNA-purH and pdCas9	This study
JL93	MG1655-gfp with psgRNA-purK and pdCas9	This study
JL94	MG1655-gfp with psgRNA-purL and pdCas9	This study
JL95	MG1655-gfp with psgRNA-purM and pdCas9	This study
JL96	MG1655-gfp with psgRNA-purN and pdCas9	This study
JL97	MG1655-gfp with psgRNA-guaA and pdCas9	This study
JL98	MG1655-gfp with psgRNA-guaB and pdCas9	This study
JL99	MG1655-gfp with psgRNA-pyrB and pdCas9	This study
JL101	MG1655-gfp with psgRNA-pyrD and pdCas9	This study

JL102	MG1655-gfp with psgRNA-pyrE and pdCas9	This study
JL103	MG1655-gfp with psgRNA-pyrF and pdCas9	This study
JL104	MG1655-gfp with psgRNA-pyrG and pdCas9	This study
JL105	MG1655-gfp with psgRNA-pyrH and pdCas9	This study
JL214	MG1655-DE3-dCas9 with psgRNA-pyrF and psdAb-TIR1	This study
JL215	MG1655-DE3-dCas9 with psgRNA-pyrG and psdAb-TIR1	This study
JL216	MG1655-DE3-dCas9 with psgRNA-cmk and psdAb-TIR1	This study
JL217	MG1655-DE3-dCas9 with psgRNA-blank and psdAb-TIR1	This study
JL218	MG1655-DE3-dCas9 with psgRNA-pyrF and psdAb-TIR2	This study
JL219	MG1655-DE3-dCas9 with psgRNA-pyrG and psdAb-TIR2	This study
JL220	MG1655-DE3-dCas9 with psgRNA-cmk and psdAb-TIR2	This study
JL221	MG1655-DE3-dCas9 with psgRNA-blank and psdAb-TIR2	This study

Plasmid	**Description**	**Reference**
pCDF-Duet	Cloning and expression vector; $Spec^R$	Novagen
pdCas9-bacteria	**pdCas9**: p15A vector with aTc-inducible expression of dCas9; $Chlor^R$	(Larson et al., 2013)
pSLQ1236	**pSLQ1236**: pColE1 vector with aTc-inducible expression of sgRNA; Amp^R	(Larson et al., 2013)
pOSIP KO	Cloning-integration plasmid, integrates at the phage 186 site; Kan^R	(St-Pierre et al., 2013)
pOSIP KO-dCas9	Plasmid for genomic integration of aTc-inducible dCas9; Kan^R	(Li et al., 2016)
pE-FLP	Plasmid for expressing FLP recombinase to excise the integration plasmid backbone; Amp^R	(St-Pierre et al., 2013)

pET28a-Nanobody®-TIR^{SynEvo1}	pBR322/ROP vector with Nanobody® expressed from the T7 promoter. Synthetically evolved translation initiation region (TIR) for optimized expression; Kan^R	(Rennig et al., 2018)
pET28a-Nanobody®-TIR^{SynEvo2}	pBR322/ROP vector with Nanobody® expressed from the T7 promoter. Synthetically evolved TIR for optimized expression; Kan^R	(Rennig et al., 2018)
psgRNA-blank	**control**: pSLQ1236 without sgRNA sequence; Amp^R	(Li et al., 2016)
pJL45	**purA**: pSLQ1236 with sgRNA targeting purA; Amp^R	This study
pJL46	**purB**: pSLQ1236 with sgRNA targeting purB; Amp^R	This study
pJL51	**purC**: pSLQ1236 with sgRNA targeting purC; Amp^R	This study
pJL52	**purD**: pSLQ1236 with sgRNA targeting purD; Amp^R	This study
pJL53	**purE**: pSLQ1236 with sgRNA targeting purE; Amp^R	This study
pJL54	**purF**: pSLQ1236 with sgRNA targeting purF; Amp^R	This study
pJL55	**purH**: pSLQ1236 with sgRNA targeting purH; Amp^R	This study
pJL56	**purK**: pSLQ1236 with sgRNA targeting purK; Amp^R	This study
pJL57	**purL**: pSLQ1236 with sgRNA targeting purL; Amp^R	This study
pJL58	**purM**: pSLQ1236 with sgRNA targeting purM; Amp^R	This study
pJL59	**purN**: pSLQ1236 with sgRNA targeting purN; Amp^R	This study
pJL60	**guaA**: pSLQ1236 with sgRNA targeting guaA; Amp^R	This study
pJL61	**guaB**: pSLQ1236 with sgRNA targeting guaB; Amp^R	This study
pJL62	**pyrB**: pSLQ1236 with sgRNA targeting pyrB; Amp^R	This study

pJL64	**pyrD**: pSLQ1236 with sgRNA targeting pyrD; Amp^R	This study
pJL65	**pyrE**: pSLQ1236 with sgRNA targeting pyrE; Amp^R	This study
pJL66	**pyrF**: pSLQ1236 with sgRNA targeting pyrF; Amp^R	This study
pJL67	**pyrG**: pSLQ1236 with sgRNA targeting pyrG; Amp^R	This study
pJL68	**pyrH**: pSLQ1236 with sgRNA targeting pyrH; Amp^R	This study
pJL69	**cmk**: pSLQ1236 with sgRNA targeting cmk; Amp^R	This study
pJL70	**ndk**: pSLQ1236 with sgRNA targeting ndk; Amp^R	This study
pJL97	**psdAb-TIR1**: pET28a-Nanobody®-$TIR^{SynEvo1}$ plasmid with origin of replication changed to ClodF13; Kan^R	This study
pJL98	**psdAb-TIR2**: pET28a-Nanobody®-$TIR^{SynEvo1}$ plasmid with origin of replication changed to ClodF13; Kan^R	This study

Supplementary Tables S3.3-S3.5 are presented as Excel files, and can be found at https://doi.org/10.1002/bit.27536.

Supplementary Table S3.3. Analysis of proteins in the pyrimidine biosynthesis pathway.

Supplementary Table S3.4. Analysis of RpoS and proteins associated with the RpoS response.

Supplementary Table S3.5. Analysis of polar effects.

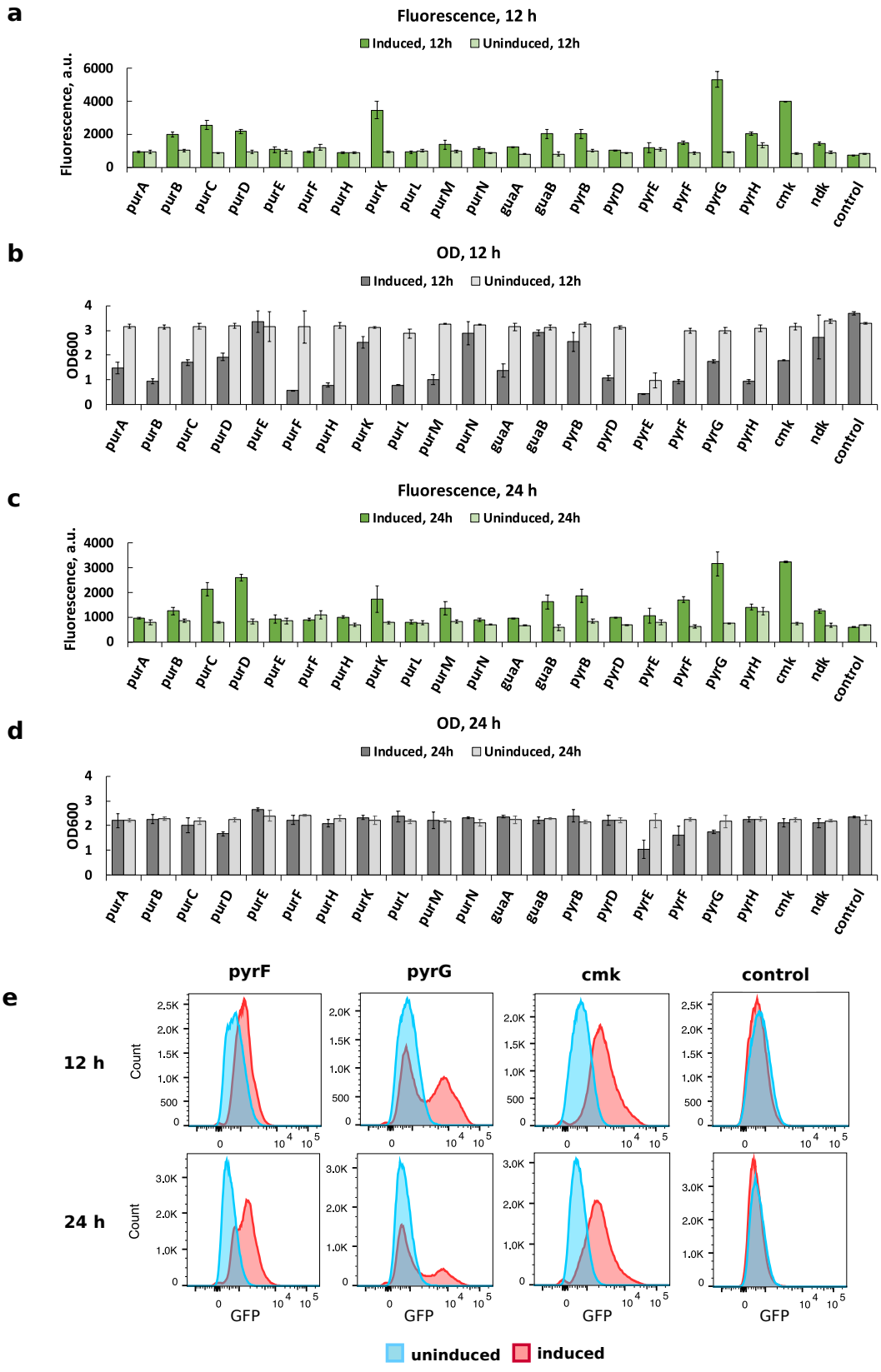

a
Fluorescence, 12 h
Induced, 12h
Uninduced, 12h
Fluorescence, a.u.
6000
4000
2000
0
purA purB purC purD purE purF purH purK purL purM purN guaA guaB pyrB pyrD pyrE pyrF pyrG pyrH cmk ndk control
b
OD, 12 h
Induced, 12h
Uninduced, 12h
OD600
4
3
2
1
0
purA purB purC purD purE purF purH purK purL purM purN guaA guaB pyrB pyrD pyrE pyrF pyrG pyrH cmk ndk control
c
Fluorescence, 24 h
Induced, 24h
Uninduced, 24h
Fluorescence, a.u.
4000
3000
2000
1000
0
purA purB purC purD purE purF purH purK purL purM purN guaA guaB pyrB pyrD pyrE pyrF pyrG pyrH cmk ndk control
d
OD, 24 h
Induced, 24h
Uninduced, 24h
OD600
4
3
2
1
0
purA purB purC purD purE purF purH purK purL purM purN guaA guaB pyrB pyrD pyrE pyrF pyrG pyrH cmk ndk control
e
pyrF
pyrG
cmk
control
12 h
24 h
Count
2.5K
2.0K
1.5K
1.0K
500
0
3.0K
2.0K
1.0K
0
10^4 10^5
GFP
uninduced
induced

Figure S3.1. Screening the purine and pyrimidine biosynthesis pathway for growth decoupling targets. (a) GFP fluorescence in CRISPRi-induced and uninduced strains after 12 h of growth. (b) OD of CRISPRi-induced and uninduced strains after 12 h of growth. (c) GFP fluorescence in CRISPRi-induced and uninduced strains after 24 h of growth. (d) OD of CRISPRi-induced and uninduced strains after 24 h of growth. (e) Histograms over 12 h and 24 h GFP fluorescence in CRISPRi-induced and not induced cultures harboring sgRNA plasmids targeting *pyrF*, *pyrG*, *cmk* or an empty control vector. The OD and fluorescence were calculated as the average of three biological replicates. Standard deviations between replicates are shown as error bars. CRISPRi-induced strains are shown in dark green (fluorescence in bar plots), dark grey (OD) or red (fluorescence in histograms), and uninduced strains are shown in bright green (fluorescence in bar plots), bright grey (OD) or blue (fluorescence in histograms).

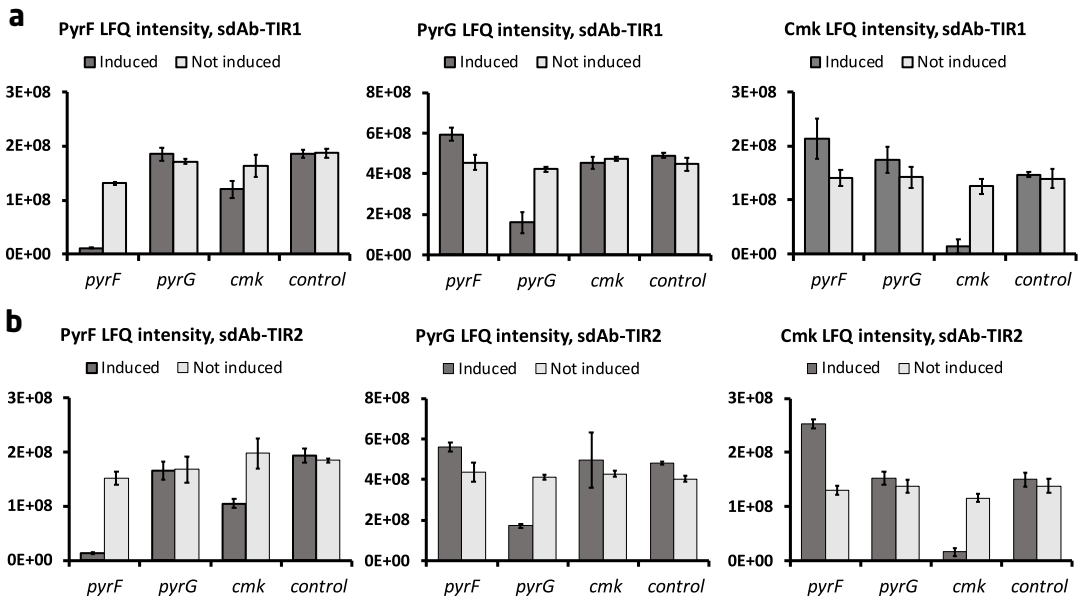


Figure S3.2. Protein levels of PyrF, PyrG and Cmk in sdAb production strains with or without induction of the CRISPRi growth decoupling system. (a) Protein levels of PyrF (panel 1), PyrG (panel 2) and Cmk (panel 3) in strains harboring the psdAb-TIR1 plasmid and sgRNA plasmids targeting *pyrF*, *pyrG*, *cmk* or an empty control vector. (a) Protein levels of PyrF (panel 1), PyrG (panel 2) and Cmk (panel 3) in strains harboring the psdAb-TIR2 plasmid and sgRNA plasmids targeting *pyrF*, *pyrG*, *cmk* or an empty control vector. CRISPRi-induced strains are shown in dark grey and uninduced strains are shown in bright grey. Protein LFQ intensities are normalized to OD and were calculated as the average of three biological replicates. Standard deviations between replicates are shown as error bars. The proteome data set can be found in Supplementary File S3.1.

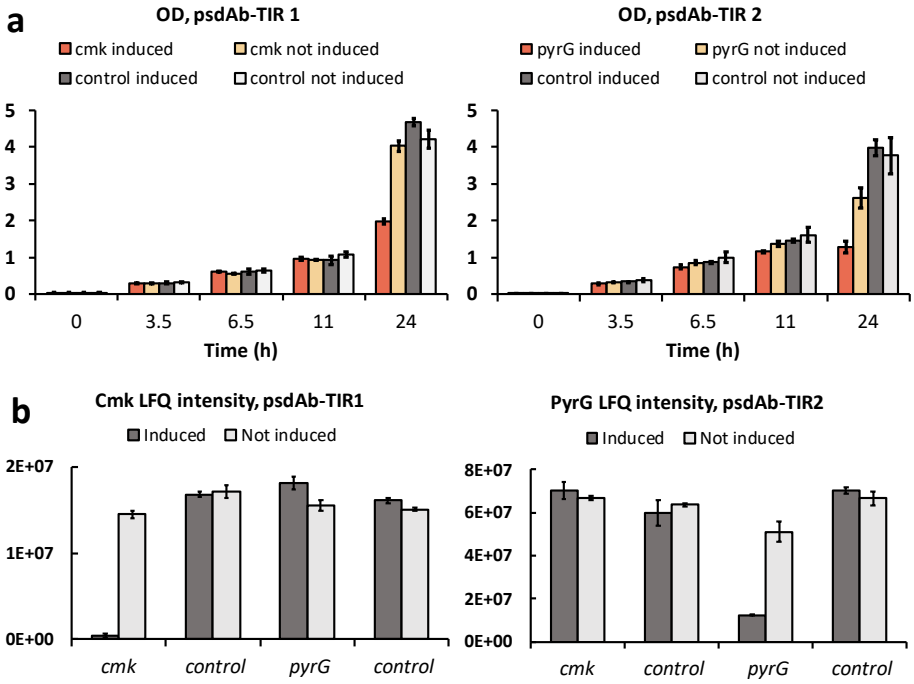


Figure S3.3. Growth and protein levels during production of sdAb in shake flask fermentation. (a) OD of CRISPRi-induced and not induced strains harboring psdAb-TIR1 and an sgRNA plasmid targeting *cmk* or an empty control vector (panel 1), and psdAb-TIR2 and an sgRNA plasmid targeting *pyrG* or an empty control vector (panel 2). CRISPRi-induced strains are shown in dark orange (*cmk* and *pyrG*) or dark grey (control), and uninduced strains are shown in bright orange (*cmk* and *pyrG*) or bright grey (control). (b) Protein levels of Cmk in strains harboring the psdAb-TIR1 plasmid and sgRNA plasmids targeting *cmk* or an empty control vector (panel 1). Protein levels of PyrG in strains harboring the psdAb-TIR2 plasmid and sgRNA plasmids targeting *pyrG* or an empty control vector (panel 2). CRISPRi-induced strains are shown in dark grey and uninduced strains are shown in bright grey. LFQ intensities are normalized to OD. The OD and protein LFQ intensities were calculated as the average of three biological replicates. Standard deviations between replicates are shown as error bars. The proteome data set can be found in Supplementary File S3.1.

Chapter 4

AN AUTOINDUCIBLE TRP-T7 EXPRESSION SYSTEM FOR PRODUCTION OF PROTEINS AND BIOCHEMICALS IN *ESCHERICHIA COLI*

Jenny Landberg[†], Hemanshu Mundhada[†‡*], Alex Toftgaard Nielsen[†]

[†] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet 220, 2800 Kongens Lyngby, Denmark.
[‡] CysBio ApS, Agern Allé 1, 2970 Hørsholm, Denmark.

***Corresponding Author**
Hemanshu Mundhada
CysBio ApS
Agern Allé 1, 2970 Hørsholm
Denmark
Email: hemu@cysbio.com

Abstract

Inducible expression systems can be applied to control the expression of proteins or biochemical pathways in cell factories. However, several of the established systems require addition of expensive inducers, making them unfeasible for large scale production. Here, we establish a genome integrated trp-T7 expression system where tryptophan can be used to control the induction of a gene or a metabolic pathway. We show that the initiation of gene expression from low- and high-copy vectors can be tuned by varying the initial concentration of tryptophan or yeast extract, and that expression is tightly regulated and homogenous when compared to the commonly used lac-T7 system. Finally, we apply the trp-T7 expression system for production of L-serine, where we reach titers of 26 g/L in fed-batch fermentation.

4.1 Introduction

Cell factories are a promising option for sustainable production of bio-based compounds such as proteins and biochemicals.[1] While some products have been commercialized, it is generally difficult to produce them at a price that is competitive with petroleum-derived chemicals.[2] The complex (heterologous) proteins and pathways that are used in cell factories oftentimes compete with native metabolic processes for cellular resources, and pathway expression at an early growth phase can drain the cell for energy or result in redox imbalance.[3,4] Furthermore, if the end-product is a toxic metabolite or protein, production at the beginning of the fermentation results in growth inhibition or poor growth.[5] Expressing the product pathway from a promoter that is induced at a specific timepoint enables a two-phase production process with biomass buildup followed by product accumulation. This approach can increase yield, titer, productivity and improve process economics.[6]

There are several inducible promoters and expression systems available for use in *Escherichia coli.*[7] One of the most commonly used systems is the lac-T7 (DE3) system.[8] It consists of a T7 RNA polymerase (T7 polymerase) controlled by the *lacUV5* promoter and repressed by the lactose operon repressor (LacI). The protein or biochemical pathway is placed under control of the T7 promoter (*pT7*) and expression is induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) or by growth on lactose.[8,9] Gene expression by T7 polymerase enables very high expression levels, which has led to extensive use of the system for protein production.[10] Proteins or biochemical pathways are also commonly expressed directly from the lac promoter (*Plac*)[11] or from the engineered lac and tryptophan fusion promoters Tac[12] and Trc,[13] which have lower basal- and higher expression levels compared to *Plac*. All these systems require addition of IPTG for induction, which is expensive and unfeasible during production of low- or medium-value compounds.[14] Although IPTG-based systems can also be induced by switching to lactose-based growth,[9] feeding lactose as a sole carbon source to a fed-batch or continuous fermentation is not feasible due to its cost and relatively low solubility in water.

Inducible promoters can be tuned to improve their function and dynamic range. Common approaches to minimize basal level expression and tune the induced expression level include engineering/mutating the repressor, operator and upstream elements, or the -10 and -35 binding sites of the RNA polymerase as well as the spacing in between.[13,15–17] It has also been shown that novel, synthetic, ligand-responsive promoter sequences can be used to improve expression levels up to 100-fold compared to native sequences.[18] If the promoter function is unknown, characterization of important promoter elements can be done by establishing mutant promoter libraries and screening those in a high-throughput manner.[19] Using information gained in a promoter library screen, the repressor binding site of the formaldehyde-inducible promoter could be identified and the information could be used to engineer mutated promoter variants with lower basal- and higher induced level expression and a significantly improved dynamic range.[19] Thermodynamic models are also helpful in the design of inducible promoters, especially to elucidate and predict the interaction between RNA polymerase and DNA.[15,20] In a recent study, Chen et al.[15] developed a method that can be used to engineer the dynamic range of ligand-inducible promoters. Using a thermodynamic model, they established a promoter library with different -10 and -35 sites and applied the optimal site combinations to build promoters regulated by different combinations of transcription factors, resulting in predictable expression with a variable dynamic range.

Several attempts have been made to find affordable and easy-to-apply (auto)inducible expression systems with low levels of basal expression and high, tunable levels of gene expression. Such systems do for example include those based on sensing of temperature, cell signaling, or depletion/addition of a compound that is naturally occurring in the cultivation medium. For temperature-based induction, the p_L and p_R phage promoters can be used to induce gene expression, usually by shifting temperature from below 37 to up to 42 °C.[21,22] The system has been coupled to the T7 polymerase for high-level protein production of protein.[23] Although an inexpensive method of control, increased cultivation temperatures can induce cellular stress responses and alter protein folding.[24] Additionally, exact and homogenous temperature control can be difficult to achieve in large-scale vessels.[25] For cell signaling-based induction, the native bacterial quorum-sensing system

has been applied to establish autoinducible production by coupling gene expression to the concentration of secreted quorum signaling molecules.[26,27] This type of system has also been shown to work when coupled to the T7 polymerase for further amplification of protein output.[28] While expression is not always tightly regulated and the system might require tuning for each desired compound, further engineering could provide a system suitable for large scale fermentation.

The tryptophan promoter (*Ptrp*) is induced as tryptophan is depleted from the medium, or by addition of 3-indoleacrylic acid. It has been extensively applied for protein and polypeptide production[29] and there are several expression vector systems based on the promoter.[10,30] Generally, the native or leaderless (without the trp leader sequence *trpL*) promoter and the trp repressor (*trpR*) have been inserted into an expression vector for direct control of a gene of interest (GOI). The promoter is tightly regulated, but does not provide as high expression levels as lac-T7 and plasmid instability is prevalent.[31]

In this study, we engineered a genome integrated auto-inducible expression system by combining the tightly regulated tryptophan promoter with the high-expressing T7 polymerase. We showed that induction of gene expression could be tuned by the initial tryptophan concentration for low- and high-copy vectors, and that expression strength was dependent on the use of the native or leaderless promoter constructs. By cloning the toxin MazF under control of our promoter constructs, we verified that the promoters are tightly regulated and suitable even for toxic protein production. Using flow cytometry, we further demonstrated that population homogeneity was improved compared to the commonly used lac-T7 system. Finally, we showed that the trp-T7 expression system can be used for biochemical production by applying it for batch and fed-batch fermentation of L-serine.

4.2 Materials and methods

Strains, media and materials

For cultivation and screening during cloning, 2xYT (16 g/L bactotryptone, 10 g/L yeast extract, 5 g/L NaCl) medium and lysogeny broth (LB, 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) agar plates with appropriate antibiotics were used. Ampicillin, kanamycin, chloramphenicol and spectinomycin was used with working concentrations of 100 µg/mL, 50 µg/mL, 50 µg/mL and 50 µg/mL respectively. For cloning and screening of plasmids carrying a tryptophan repressible gene, 0.5 mM tryptophan was added to agar plates and growth medium to repress gene expression. Growth and production experiments were performed in M9 minimal medium with 0.1 mM $CaCl_2$, 2.0 mM $MgSO_4$, 1x trace element solution and 1x M9 salts supplemented with glucose, appropriate antibiotics and other compounds as described below. The stock solution of trace elements was 1000x and consisted of 15g/L EDTA(Na_2)*$2H_2O$, 4.5 g/L $ZnSO_4$*$7H_2O$, 0.7 g/L $MnCl_2$*$4H_2O$, 0.3 g/L $CoCl_2$*$6H_2O$, 0.2 g/L $CuSO_4$*$2H_2O$, 0.4 g/L $NaMoO_4$*$2H_2O$, 4.5 g/L $CaCl_2$*$2H_2O$, 3 g/L Fe SO_4*H_2O, 1 g/L H_3BO_3 and 0.1 g/L KI dissolved in double-distilled water and sterile filtered. The stock solution of M9 salts was 10x and consisted of 6.8 g/L Na_2HPO_4 anhydrous, 3 g/L KH_2PO_4, 5 g/L NaCl and 1 g/L NH_4Cl dissolved in double-distilled water and autoclaved. The strains and plasmids that were used are listed in Table 4.1. Chemicals were purchased from Sigma-Aldrich (Taufkirchen, Germany). Restriction enzymes and PCR polymerases were purchased from ThermoScientific (Waltham, MA, USA). USER enzyme was purchased from BioNordika (Herlev, Denmark).

Plasmid and strain construction

Plasmids were purified using the Machery Nagel plasmid purification kit (Dure, Germany). Cells were transformed using the TSS buffer method. 40 mL exponentially growing cells (OD ~ 0.4-0.8) were spun down and resuspended in 0.5-1 mL TSS buffer (100 g/L PEG 8000, 30 mM $MgCl_2$, 20 mL DMSO in LB). Cells were mixed with 100 ng plasmid and put on ice for 15 min. The plasmid-cell mix was subjected to heatshock (42 °C, 1 min)

and put on ice for 2 min. Cells were incubated for 1 h at 37 °C before plating on LB agar with appropriate antibiotics. All plasmids were constructed by USER cloning.[32,33] Briefly, template DNA was amplified using Phusion U polymerase and primers containing a uracil complementary overhang. PCR products were purified and subjected to *DpnI* digestion. The backbone and the template fragments were mixed with USER enzyme and buffer to a volume of 10 µL and incubated for 20 minutes at 37 °C followed by 20 min at 25 °C. The mix was diluted with 10 µL water and 10 µL of the diluted mix was used for transformation into competent cells. All primers used in the study were ordered from Integrated DNA Technologies (Leuven, Belgium) and are listed in Suppl. Table 4.1. The tryptophan repressor, operator and promoter and the *mazF* gene were amplified from the *E. coli* genome. Integration of trp-T7 and trp*-T7 was carried out according to the protocol by St. Pierre et al.[34] Briefly, the cassettes were cloned on a pOSIP vector, followed by integration at the attB-186(O) site. The antibiotic cassette was excised using a FLP recombinase.

Table 4.1. Strains and plasmids used in the study. Names used to refer to strains and plasmids in the main text are written in bold.

Strain	Description	Reference
E. coli NEB 5-alpha	*fhuA2* Δ*(argF-lacZ)U169 phoA glnV44 Φ80* Δ*(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*; cloning strain	New England Biolabs
E. coli K12 MG1655	F- lambda- *ilvG- rfb-50 rph-1*	Lab collection
E. coli MG1655 (DE3)	lac-T7: MG1655 λ *(DE3)*	[35]
ALE5	MG1655 Δ*sdaA*Δ*sdaB*Δ*tdcG*Δ*glyA*; serine tolerant strain	[36]
ALE5	MG1655 Δ*sdaA*Δ*sdaB*Δ*tdcG*Δ*glyA* λ*(DE3)*; serine tolerant strain	[36]
JL13	trp-T7: MG1655 attB-186(O): P_{trp}; tryptophan repressor and promoter controlling T7 polymerase expression	This study
JL44	trp*-T7: MG1655 attB-186(O): P_{trp*}; tryptophan repressor and truncated promoter controlling T7 polymerase expression	This study

JL45	trp-T7 with pSEVA27-mCherry	This study
JL46	trp*-T7 with pSEVA27-mCherry	This study
JL84	trp-T7-ser: ALE5 attB-186(O): P_{trp}; tryptophan repressor and promoter controlling T7 polymerase expression	This study
JL141	trp-T7 with pCDF-mCherry	This study
JL142	trp*-T7 with pCDF-mCherry	This study
JL151	lac-T7 with pSEVA-mCherry	This study
JL153	lac-T7 with pCDF-mCherry	This study
JL154	trp-T7-ser with pSEVA27-serABC	This study
JL155	trp*-T7-ser with pSEVA27-serABC	This study
JL160	trp-T7 with pSEVA27	This study
JL161	trp-T7 with pSEVA27-mazF	This study
JL162	trp-T7 with pCDF	This study
JL163	trp-T7 with pCDF-mazF	This study
JL164	trp*-T7 with pSEVA27	This study
JL165	trp*-T7 with pSEVA27-mazF	This study
JL166	trp*-T7 with pCDF	This study
JL167	trp*-T7 with pCDF-mazF	This study

Plasmid	**Description**	**Reference**
pOSIP KO	Cloning-integration plasmid, integrates at phage 186 site; Km^R	[34]
pOSIP KO-lacI-T7	Plasmid for integrating T7 polymerase under control of LacI; Km^R	[36]
pE-FLP	Plasmid for expressing FLP recombinase to excise the integration plasmid backbone; Amp^R	[34]
pSEVA27-sl3	pSEVA27: Low copy number plasmid pSC101 origin (~5 copies per cell); MCS under T7 promoter; Km^R	[37]
pCDF-1b	pCDF: High copy number plasmid CloDF13 origin (~20-40 copies per cell); MCS under T7 promoter; Sp^R	Novagen

pJL3	pOSIP-trp*-T7: pOSIP KO harboring a P_{trp*} integration cassette controlling the T7 polymerase; Km^R	This study
pJL5	pSEVA-mCherry: SEVA27-sl3 plasmid with mCherry expressed from the T7 promoter; Km^R	This study
pJL9	pOSIP-trp-T7: pOSIP KO harboring a P_{trp} integration cassette controlling the T7 polymerase; Km^R	This study
pJL42	pCDF-1b plasmid with mCherry expressed from the T7 promoter; Sp^R	This study
pJL83	pCDF-mCherry: pCDF-1b with *lacI* removed and mCherry expressed from the T7 promoter; Sp^R	This study
pJL88	pSEVA-mazF: pSEVA27-sl3 plasmid with *mazF* expressed from the T7 promoter; Km^R	This study
pJL89	pCDF-mazF: pCDF-1b plasmid with *lacI* removed and *mazF* expressed from the T7 promoter; Sp^R	This study
pSEVA27-sl-*serA*^{mut}*CB* TIR 1	pSEVA-serACB: Expression of the serine operon from the T7 promoter; Km^R	[38]

Microtiter plate cultivations

Single transformants were inoculated in 96-deep well plates in 500 µL M9 medium supplemented with 0.4% glucose, 0.02% yeast extract and 0.5 mM tryptophan. Cultures were grown overnight at 37 °C, 250 rpm. Cells were washed and inoculated with a 1:100 inoculum ratio in microtiter plates with 150 µL M9 medium supplemented with 0.5% glucose and a specified amount of tryptophan or yeast extract. The plates were sealed with oxygen permeable film (Breathe-Easy sealing membrane, Sigma-Aldrich) and incubated at medium shaking, 37 °C in an ELx808 Absorbance Reader (BioTek, Winooski, VT, USA) for OD_{630} measurements, or in a Synergy H1 Hybrid Multi-Mode Reader (BioTek, Winooski, VT, USA) for OD_{630} and fluorescence (excitation 587 nm, emission 617 nm) measurements.

Flow cytometry

Single transformants were inoculated in 2 mL M9 medium supplemented with 0.4% glucose, 0.02% yeast extract and 0.5 mM tryptophan. Cultures were grown overnight at 37 °C, 250 rpm. Cells were washed and inoculated

with an inoculum ratio of 1:100 in a 24 deep well plate in 2.5 mL M9 medium supplemented with 0.4% glucose. The MG1655(DE3) strains were induced with 1 mM IPTG at OD_{630} ~ 0.3-0.4. Samples were diluted appropriately and analyzed in a MACSQuant® VYB flow cytometer (Miltenyi Biotec, Cologne, Germany). Expression of mCherry was detected using a yellow laser (561 nm) and the 615/20 nm Y2 filter. Forward (FSC) and side (SSC) scatter was detected with the yellow laser and a 561/10 nm filter. The results were analyzed with FlowJo (Becton, Dickinson & Company, Franklin Lakes, NJ, USA).

Serine production in small batch fermentation

Pre-inoculums were prepared by inoculating single transformants in 2.5 mL M9 medium supplemented with 0.4% glucose, 0.02% yeast extract, 0.5 mM tryptophan and 2 mM glycine. Cultures were grown overnight at 37 °C, 250 rpm. Cells were washed and inoculated to an OD of 0.05 in 50 mL M9 medium supplemented with 0.4% glucose, 2 mM glycine and 0.04 or 0.5 mM tryptophan.

Serine production in fed-batch fermentation

Medium for fed-batch fermentation was prepared as previously described,[36] with an addition of 0.125% YE instead of 0.2 % YE to the batch medium. A preculture was grown overnight at 37 °C, 250 rpm, in 2xYT medium with 0.5 mM trp and 0.1% glucose. Cells were washed and the bioreactor was inoculated to an OD of 0.1. Oxygen saturation was maintained at 40% throughout the fermentation. Samples were taken regularly for analysis. A correlation of 0.374 between cell dry weight (CDW) and OD was used for CDW estimation and subsequent yield calculations.[35]

Analytical methods

Samples from the small batch and fed-batch fermentations were filtered and diluted appropriately. Glucose and byproducts were analyzed with HPLC as previously described.[39] Serine was analyzed with LC-MS as previously described.[35]

4.3 Results

Design and construction of a trp-T7 expression system

To establish a high-expressing, inexpensive inducible system with low basal level expression, we wanted to combine the high level of gene expression achievable from the T7 polymerase with the tightly regulated trp operon. Using such a system, we hypothesized that the onset of gene expression could be tuned to a certain timepoint by regulating the amount of tryptophan added to the medium at the start of the cultivation (Fig. 4.1a-b).

The trp operon consists of five structural genes encoding tryptophan synthase, and is regulated by the tryptophan repressor (TrpR). In growth medium with excess tryptophan, TrpR binds the tryptophan operator (*trpO*), blocking transcription of the operon from the trp promoter (*Ptrp*). The repression is further (~10-fold) enhanced by attenuation, a mechanism where the trp leader sequence (*trpL*) forms RNA stem loops that blocks transcription in the presence of tryptophan.[40] We constructed two integrative cassettes where the *E. coli* tryptophan regulation machinery was designed to control the T7 polymerase (Fig. 4.1c). The first variant contained the tryptophan repressor and the native tryptophan leader sequence, promoter and operator. This variant is hereafter referred to as trp-T7. The second variant contained the tryptophan repressor, the tryptophan promoter and operator, but not the tryptophan leader sequence. This variant is hereafter referred to as trp*-T7. The cassettes were cloned on a pOSIP integration vector[34] and integrated into the attB-186(O) site in *E. coli* K12 MG1655, resulting in the two strains hereafter referred to as trp-T7 and a trp*-T7.

Low-copy plasmid pSEVA27 (hereafter pSEVA) and high-copy plasmid pCDF-1b (hereafter pCDF) harboring *pT7* controlled by *lacO* were used as expression vectors to evaluate expression homogeneity and basal level expression, timing of induction in different concentrations of tryptophan, and biochemical production from the expression system (Fig. 4.1d). We decided to not replace the *lacO* site with a *trpO* site in order to facilitate cloning during future use of the system, as most T7-based expression vectors have a *lacO* site present. We further hypothesized that *trpR*-based repression

of the genomic T7 polymerase would provide sufficient regulation of the system, as the trp operon is inherently under more strict regulation when compared to the lac operon. However, keeping *lacO* meant that we had to remove *lacI* from pCDF to enable target gene expression during growth on glucose.

Evaluation of expression profiles in different concentrations of tryptophan and yeast extract

For evaluation of the promoter system, strains with integrated trp-T7 and trp*-T7 (that contains or lacks the *trpL*, respectively), were transformed with a low- or high-copy plasmid harboring mCherry under control of *pT7*, resulting in four different strains (Table 4.1, Fig. 4.1c-d). The induction profile of each strain was investigated using different starting concentrations of tryptophan and (tryptophan-containing) yeast extract (YE) with mCherry as a reporter gene (Fig. 4.2a). Strains were grown in 0-0.5 mM tryptophan or 0-1% YE, and OD and fluorescence was monitored for 20 h. Overall, the initial concentration of tryptophan and YE was inversely correlated to the fluorescence per OD and the initiation of mCherry expression (Fig. 4.2c-d, Suppl. Fig. 4.1), showing that tryptophan and YE could be used to control the induction of the gene of interest (GOI) in both versions of the trp-T7 strains, and for both the high- and low-copy vector. The fluorescence and fluorescence per OD at different time points were generally higher with decreasing concentrations of tryptophan or yeast extract, since expression was induced earlier (Fig. 4.2b-c). An exception was trp-T7 with pCDF-mCherry, where fluorescence per OD was similar for all YE concentrations between 0-0.25% (Fig. 4.2c, Suppl. Fig. 4.1). As expected, growth was dependent on the amount of yeast extract added to the medium (Suppl. Fig. 4.1). For strains carrying the low-copy plasmid pSEVA-mCherry (Fig. 4.2b), high initial concentrations of tryptophan and YE (0.5, 0.25 mM, and 1, 0.5 and 0.25%, respectively) completely repressed mCherry expression in trp-T7, while low levels of expression could be seen after ~10-13 h in trp*-T7. This is expected, as the *trpL*-containing trp-T7 version of the promoter should exhibit a higher level of repression than trp*-T7. With most concentrations of tryptophan and YE, the fluorescence per OD was only marginally higher in trp*-T7 compared to trp-T7. However, for the strains

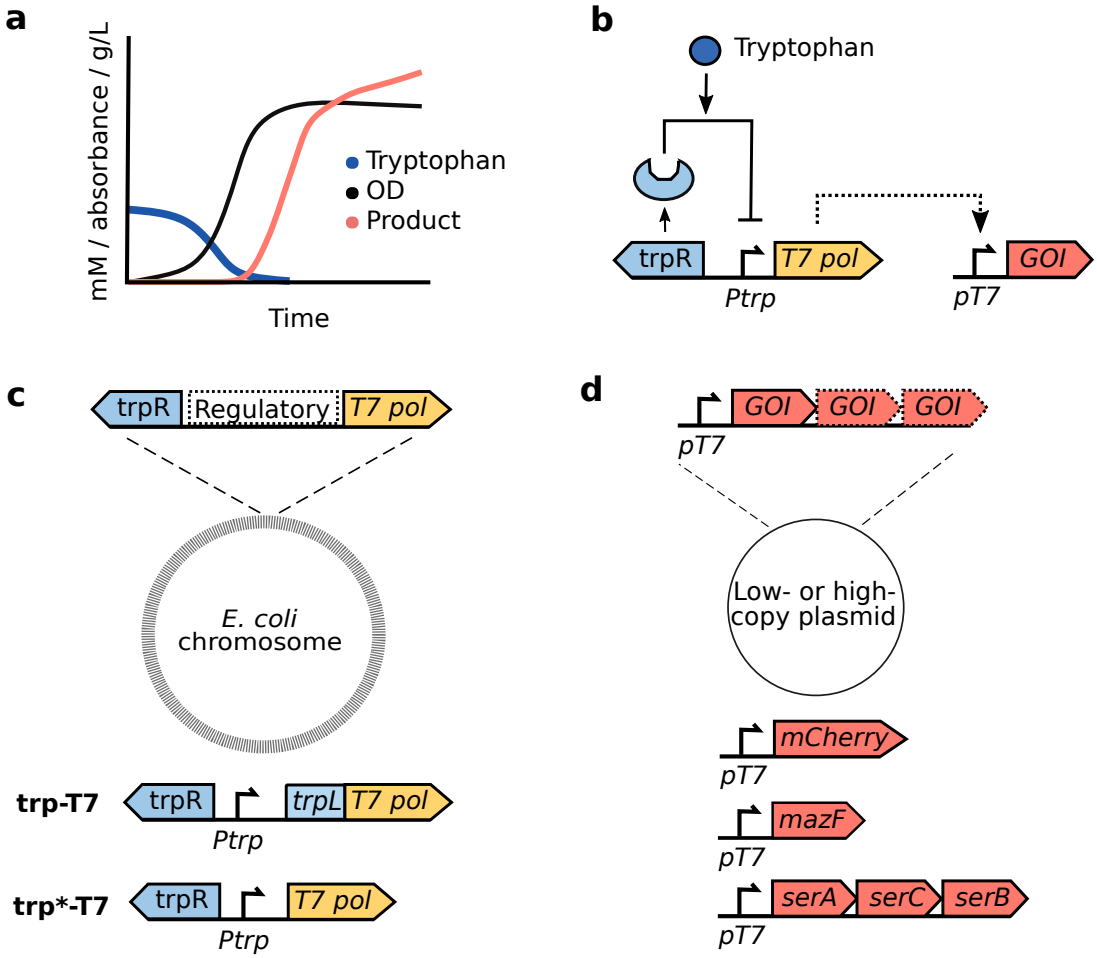


Figure 4.1. The layout and function of the trp-T7 expression system. (a) Tryptophan-based repression and induction for production of a protein or biochemical. (b) Repression and induction of the trp-T7 expression system is controlled by tryptophan concentration. When tryptophan is present in the growth medium it is bound by TrpR, which binds trpO and inhibits expression of T7 polymerase and the GOI. After tryptophan is consumed, the conformation of TrpR changes, leading to dissociation from trpO. T7 polymerase is produced and transcribes the plasmid-based GOI from pT7. (c) Layout of the genome integrated cassettes trp-T7 and trp*-T7. (d) The plasmids with the different T7-transcribed genes that were used to test trp-T7 based protein and biochemical production in this study.

carrying the high-copy plasmid pCDF-mCherry (Fig. 4.2c), the fluorescence per OD was significantly higher in trp*-T7, compared to trp-T7, for both trp and YE cultures. Low but significant levels of mCherry were observed in trp*-T7 even at high initial concentrations of tryptophan and YE, while, for trp-T7, repression was similar to that seen in the low-copy plasmid strains. Interestingly, the fluorescence per OD in strains harboring the trp-T7 expression cassette was higher when mCherry was expressed from the low-copy pSEVA plasmid than from high-copy pCDF plasmid. This may be due to higher metabolic burden resulting from maintaining multiple plasmid copies.[41]

Comparing expression strength and homogeneity to lac-T7

Next, we sought to investigate the expression heterogeneity of both promoter constructs and compare it to the commonly used lac-T7 expression system. A lac-T7 strain (*E.coli* MG1655 (DE3))[35] was transformed with pSEVA-mCherry or pCDF-mCherry. Overnight cultures were inoculated in minimal medium without tryptophan. The lac-T7 cultures were induced with 1 mM IPTG at OD ~ 0.3-0.35. Expression levels were investigated with flow cytometry 8 and 24 h after inoculation (Fig. 4.3a).

For the pSEVA-mCherry strains, fluorescence was slightly lower for trp-T7 compared to trp*-T7 after 8 h ($p = 0.027$), while no significant difference was seen after 24 h ($p = 0.12$) (Fig. 4.3b, Suppl. Fig. 4.2a). Fluorescence of the lac-T7 strain was on average 3.9- and 2.7-fold higher compared to the trp-T7 and trp*T7 strain, respectively. The lac-T7 strain also had a narrower distribution of fluorescence at the single cell level. The broad fluorescence profile seen in the trp strains could be the result of heterogenous transport and degradation rates of tryptophan, resulting in differences in the trp operon induction response,[42] and such population heterogeneity could be improved by for example engineering constitutive expression of tryptophan permeases. After 24 h, the difference between the trp and lac strains had decreased, with the lac-T7 strain having an average fluorescence that was 1.8- and 1.3-fold higher compared to the trp-T7 and trp*T7 strain, respectively. There was a subpopulation of low- and non-expressing cells forming in the lac-T7 strain, while expression in the trp strains was homogenous (Fig. 4.3b, Suppl. Fig 4.2c).

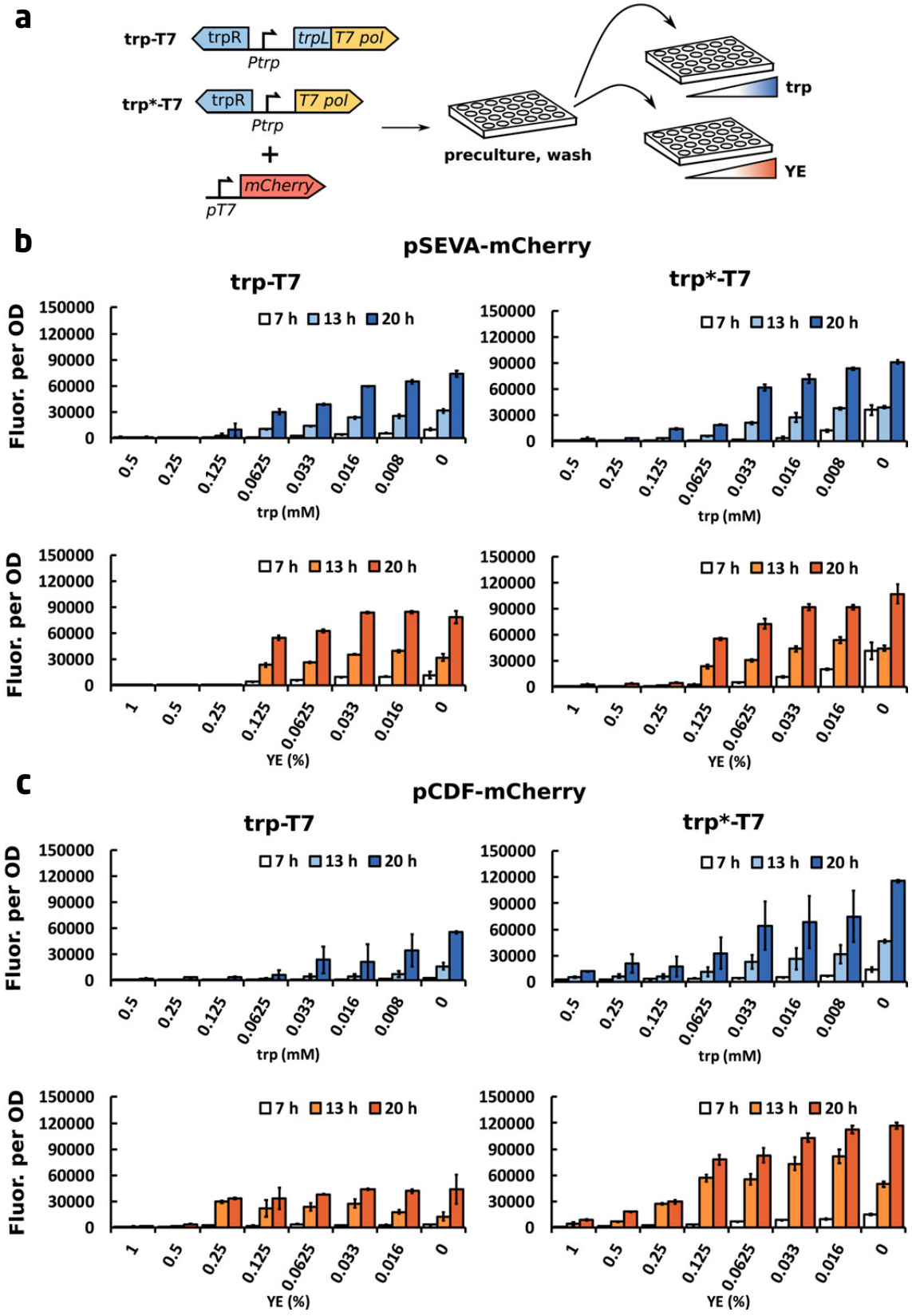

a
trp-T7
trpR
trpL
T7 pol
Ptrp
trp*-T7
trpR
T7 pol
Ptrp
+
mCherry
pT7
preculture, wash
trp
YE
b
pSEVA-mCherry
trp-T7
trp*-T7
Fluor. per OD
7 h
13 h
20 h
150000
120000
90000
60000
30000
0
0.5
0.25
0.125
0.0625
0.033
0.016
0.008
0
trp (mM)
1
0.5
0.25
0.125
0.0625
0.033
0.016
0
YE (%)
c
pCDF-mCherry
trp-T7
trp*-T7
Fluor. per OD
7 h
13 h
20 h
trp (mM)
YE (%)

Figure 4.2. Experiment layout and fluorescence per OD after 7, 13 and 20 h for strains harboring the trp-T7 or trp*-T7 expression system. The strains were grown with different starting concentrations of tryptophan (trp) or yeast extract (YE). (a) Experiment scheme. (b) Fluorescence per OD for strains trp-T7 and trp*-T7 with mCherry expressed from low-copy plasmid pSEVA. (c) Fluorescence per OD for strains trp-T7 and trp*-T7 with mCherry expressed from high-copy plasmid pCDF. The different time points are shown in bright (7 h), medium (13 h) or dark (20 h) shades of blue or orange for strains grown with tryptophan or YE, respectively. The fluorescence per OD was calculated as the average of three biological replicates. The standard deviations are shown as error bars.

For the pCDF-mCherry strains, fluorescence was significantly higher for the trp*-T7 compared to the trp-T7 strain after 8 h of growth ($p = 0.009$) (Fig. 4.3c, Suppl. Fig. 4.2b). The lac-T7 was expressing higher levels of mCherry than both trp strains, with a 17.9- and 5.5-fold higher average fluorescence compared to trp-T7 and trp*-T7, respectively. Similar to the pSEVA-mCherry cultures, the lac-T7 strain had a narrower expression profile compared to both trp strains. After 24 h, the difference between the trp and lac strains had decreased slightly, with lac-T7 having 8.2- and 3.5-fold higher average fluorescence levels compared to trp-T7 and trp*-T7, respectively (Fig. 4.3c, Suppl. Fig 4.2b). However, mCherry expression in the lac-T7 strain was highly heterogenous, with a large population not expressing or expressing only low amounts of the protein (Fig. 4.3c, Suppl. Fig 4.2d). The expression heterogeneity of the lac-T7 system constitutes a significant problem when using the system in fed-batch or continuous production processes. It results in decreased productivity, especially as the subpopulation that stops producing the compound of interest usually quickly outgrows the producing population.[43] Application of a less leaky version of the lac-T7 system or application of a constitutive lac permease could potentially improve population homogeneity of the IPTG- induced system, without affecting the expression levels negatively.[44]

Overall, the average fluorescence was significantly lower for trp-T7 compared to the lac-T7 strains (Fig. 4.3b-c, Suppl. Fig. 4.2a-b), which may partly be explained by the tight regulation of the trp expression system.

Additionally, due to the *lacO* site present on the expression plasmids, mCherry expression is also being repressed by endogenously expressed *lacI*. Comparison of mCherry fluorescence levels for trp-T7 and trp*-T7 with or without IPTG induction showed that fluorescence per OD increased approximately 10% after IPTG addition for both pSEVA- and pCDF-based expression vectors (data not shown). Thus, higher protein expression levels could potentially be achieved by deleting the endogenous *lacI* gene.

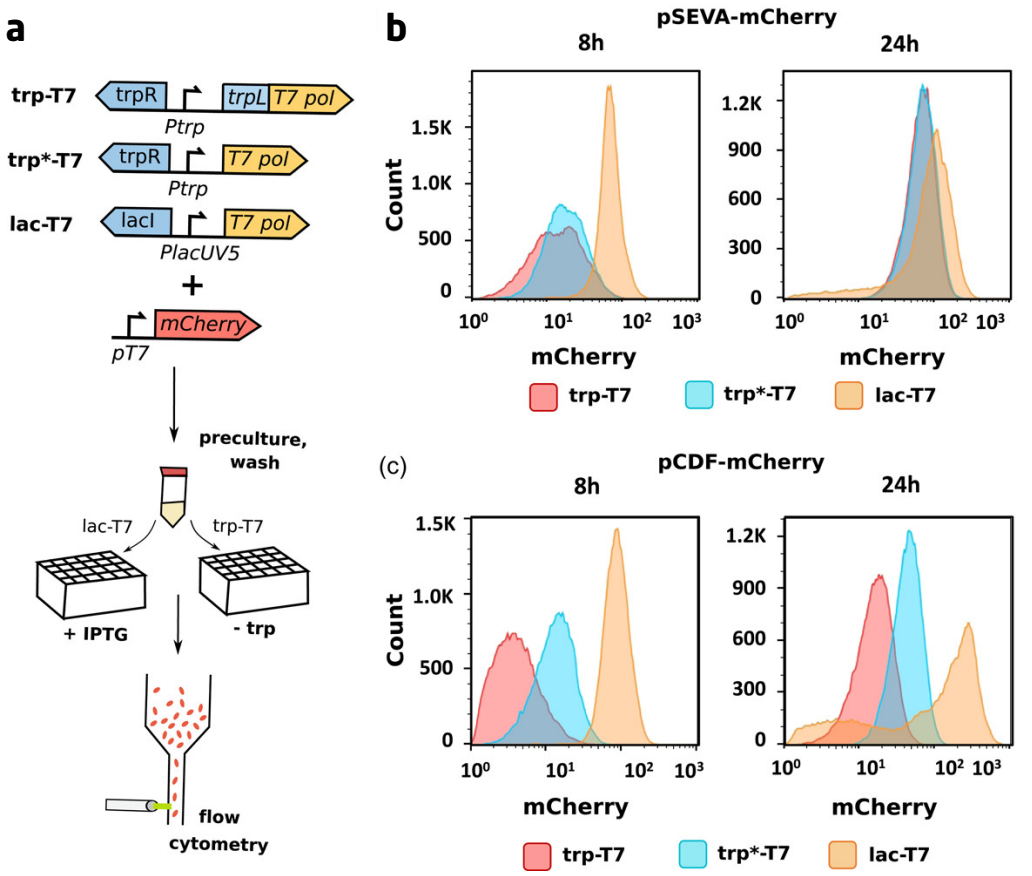


Figure 4.3. Histograms showing the fluorescence of strains trp-T7, trp*-T7 and lac-T7 after 8 and 24 h of growth. (a) Experiment scheme. (b) Fluorescence of strains with mCherry expressed from low-copy vector pSEVA. (c) Fluorescence of strains with mCherry expressed from high-copy vector pCDF. mCherry was induced by growth in tryptophan deplete medium (trp-T7, trp*-T7) or by addition of 1 mM IPTG (lac-T7). Each histogram shown is a representative sample out of three biological replicates.

Evaluating the system for toxic protein production

Production of toxic proteins is challenging, as they can inhibit cell growth. An expression system with low basal level expression is preferred to avoid growth inhibition and decreased productivity when expressing such proteins. To investigate whether our constructs are suitable for toxic protein production, the *mazF* gene was cloned behind *pT7* on pSEVA and pCDF, resulting in plasmids pSEVA-mazF and pCDF-mazF. MazF is an ACA-specific endoribonuclease that stalls cell growth by ribosome independent mRNA cleavage.[45,46] The toxin has previously been used to demonstrate the suitability of an inducible expression systems in *E. coli*.[17] The *mazF* plasmids and respective control plasmids (no gene after *pT7*) were transformed to trp-T7 and trp*-T7, resulting in eight different strains.

The experiment was carried out similar as shown in Fig. 4.2a; overnight cultures were inoculated in microtiter plates with 0-0.25 mM trp or 0-0.5% YE and growth was monitored for 24 h (Fig. 4.4a-b). At the highest tryptophan and the two highest YE concentrations, no significant difference could be seen in growth rate and final OD between the control strains and the strains harboring the *mazF*-plasmids, with the exception of trp*-T7 harboring pCDF-mazF where a slight decrease in growth rate could be seen also for the highest trp and YE concentrations (Fig. 4.4b). At the lower concentrations, the MazF-expressing strains did not grow, grew with a much lower growth rate or reached a lower OD compared to the control strain for all constructs. Overall, the results show that the trp-T7 expression system is tight enough to be suitable for expression of toxic proteins from high- and low-copy vectors in *E. coli*.

Interestingly, growth was resumed in the toxin-expressing strains toward the end of the microtiter plate cultivation. This phenomenon has been observed for expression of other toxins in our lab, and is possibly a result of single cells overcoming the toxic effect of MazF by mutating the toxin gene, the T7 polymerase or possibly by overexpressing the native antitoxin *mazE*. It is worth noting that growth was initiated at approximately the same timepoint in all cultures where growth was completely inhibited by toxin expression, indicating some common underlying mechanism.

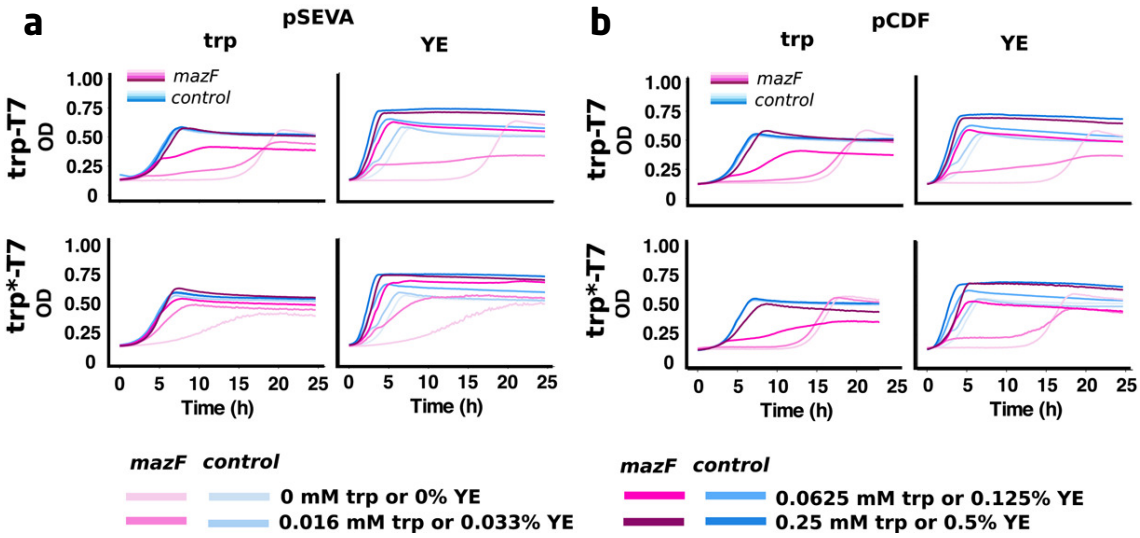


Figure 4.4. Growth curves of strains harboring the trp-T7 expression systems and a mazF-expressing plasmid (pink lines) or the corresponding control plasmid (blue lines). A darker shade of each color represents a higher concentration of trp or YE added to the medium at the beginning of the cultivation. a) Growth curves of strains trp-T7 and trp*-T7 harboring *mazF* on pSEVA. b) Growth curves of strains trp-T7 and trp*-T7 harboring *mazF* on pCDF. The growth curves were calculated as an average of three biological replicates.

Applying trp-T7 for production of L-serine

L-serine is currently used in the pharmaceutical and cosmetics industry, but has potential for extended industrial use as a building block for numerous chemicals.[47] In previous studies, the L-serine pathway has been expressed from *pT7* by an integrated *lacUV5*-T7 polymerase induced by addition of IPTG.[35,36] To test L-serine production from our tryptophan repressible expression system the trp-T7 cassette was integrated into an L-serine-tolerant strain of *Escherichia coli*.[36] We chose to only apply one of the cassettes, as mCherry production levels of trp-T7 and trp*-T7 were similar according to flow cytometry (Fig. 4.3b). The strain was transformed with pSEVA27-sl-*serA*^{mut}*CB* TIR1 (hereafter pSEVA-serACB) harboring an expression-optimized serine operon under control of *pT7*,[38] resulting in strain trp-T7-ser.

Overnight cultures of trp-T7-ser were inoculated in shake flasks with either 0.04 or 0.5 mM tryptophan to test inducible and repressible function of the system, respectively. OD and serine was measured continuously for 24 h. At the end of cultivation, serine titers reached 0.55 g/L (Fig. 4.5a), slightly lower than the 0.62 g/L achieved when using lac-T7.[38] A decrease in growth rate and final OD could be seen for the serine-producing strain, which may be due to carbon being utilized for serine production instead of growth.

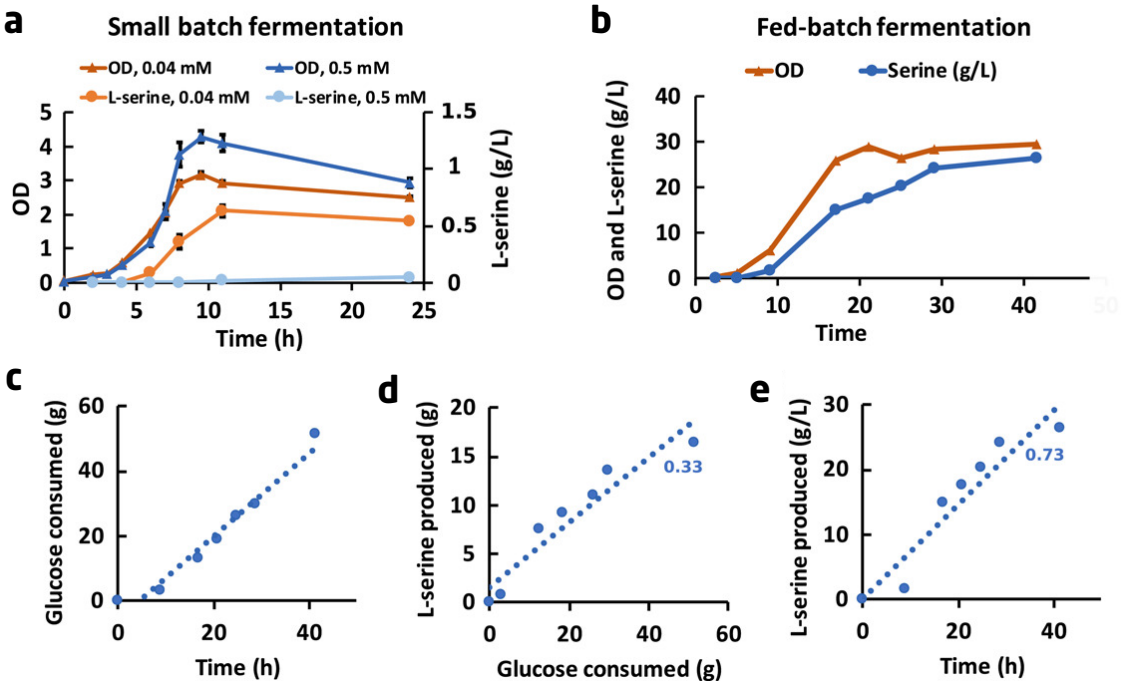


Figure 4.5. Serine production in trp-T7-ser harboring the trp-T7 expression system and plasmid pSEVA-serACB (a) Small batch fermentation. Growth of trp-T7-ser in 0.04 mM tryptophan (triangles, dark orange) and 0.5 mM tryptophan (triangles, blue). L-serine production of trp-T7-ser in 0.04 mM tryptophan (circles, orange) and 0.5 mM tryptophan (circles, bright blue). OD and serine concentrations were calculated as the average of three biological replicates. The standard deviations are shown as error bars. (b) Growth (orange) and production (blue) curves for fed-batch fermentation of trp-T7-ser. (c) Glucose consumed during the fed-batch fermentation of trp-T7-ser. (d) L-serine yield during the fed-batch fermentation of trp-T7-ser. (e) Productivity of the fed-batch fermentation of trp-T7-ser.

To investigate the scalability of the process, we further tested the strain using fed-batch fermentation. Tryptophan was replaced by yeast extract for initial repression of the serine pathway, as this is a more plausible alternative for large scale fermentation. After calculating the tryptophan concentration in the yeast extract and taking the larger cultivation volume in the bioreactor into account, 0.125% YE was added to the medium before start of cultivation. An overnight culture was inoculated in a 1 L bioreactor containing 250 mL medium supplemented with 0.125% YE. The glucose feed was started after 9 h (at OD ~ 9) and growth and serine production was monitored for ~ 42 h (Fig. 4.5b).

At the end of the fermentation, 26 g/L L-serine had been produced, approximately half of the 48 g/L reached when using lac-T7.[38] The productivity was 0.73 g/Lh (Fig. 4.5e), approximately 70 % of that reached with the lac-T7 system. However, the same yield of L-serine per glucose (0.33 and 0.30 g/g for trp-T7 and lac-T7, respectively) was achieved; the lower titer and productivity may partly be explained by a decreased glucose uptake and growth rate in our trp-T7 strain (Fig. 4.5c-d). Increasing the concentration of YE to 0.2%, which was used in the lac-T7 fermentation, might improve growth rate and glucose uptake but may also affect the timing of induction of the L-serine pathway. Another contributing factor could also be the lower expression levels from trp-T7 compared to lac-T7 that was observed previously (Fig. 4.3b).

4.4 Conclusions

Timing the induction of protein and product pathway expression can improve bioprocess performance and can serve as an important mode of control during fermentation. Low-cost expression systems are of special interest for large-scale fermentations and for production of low-or medium-value compounds. In this study, we demonstrated that the *E. coli* tryptophan promoter can be used to control T7-based expression of a gene of interest and that induction can be tuned by varying tryptophan or yeast extract concentration in the start medium. Tryptophan depletion did not seem to impact growth in our experiments. However, for production of other types

of biochemicals or proteins, the growth rate and tryptophan uptake rate may be affected, which would impact the dynamics of the expression system. This could especially be the case if the expressed (heterologous) pathway enzymes or products have a high tryptophan content. In those cases, engineering approaches may have to be taken to tune system and adapt to the new conditions.

Production levels were lower compared to using the lac-T7 expression system for both protein and biochemical production. Further improvement of the system could be made by engineering constitutive tryptophan uptake or using a l*acI* knock-out strain to improve induction homogeneity, or engineering the dynamic range using a mutant promoter library approach.[19] Overall, the low-cost and tight regulation of the trp-T7 system makes it an attractive alternative for large scale production of biochemicals and proteins.

4.5 Acknowledgements

The pOSIP KO and the PE-FLP plasmids were a gift from Drew Endy & Keith Shearwin (Addgene plasmids # 45985 and 45978, respectively). We would like to thank Hanne Bjerre Christensen and Lars Schrübbers for conducting the serine analysis. The authors would like to thank the Novo Nordisk Foundation (grant number NNF16CC0020908 and NNF10CC1016517) for funding.

4.6 References for Chapter 4

1. Nielsen, J. & Keasling, J. D. Engineering Cellular Metabolism. *Cell* **164**, 1185–1197 (2016).
2. Gustavsson, M. & Lee, S. Y. Prospects of microbial cell factories developed through systems metabolic engineering. *Microb. Biotechnol.* **9**, 610–617 (2016).
3. Brockman, I. M. & Prather, K. L. J. Dynamic metabolic engineering: New strategies for developing responsive cell factories. *Biotechnol. J.* **10**, 1360–1369 (2015).
4. Blazeck, J. & Alper, H. S. Promoter engineering: Recent advances in controlling transcription at the most fundamental level. *Biotechnol. J.* **8**, 46–58 (2013).
5. Saida, F., Uzan, M., Odaert, B. & Bontems, F. Expression of Highly Toxic Genes in E. coli: Special Strategies and Genetic Tools. *Curr. Protein Pept. Sci.* **7**, 47–56 (2006).
6. Burg, J. M. *et al.* Large-scale bioprocess competitiveness: the potential of dynamic metabolic control in two-stage fermentations. *Curr. Opin. Chem. Eng.* **14**, 121–136 (2016).
7. Terpe, K. Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **72**, 211–222 (2006).
8. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130 (1986).
9. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
10. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: Advances and challenges. *Front. Microbiol.* **5**, 1–17 (2014).
11. Gronenborn, B. Overproduction of phage Lambda repressor under control of the lac promotor of Escherichia coli. *MGG Mol. Gen. Genet.* **148**, 243–250 (1976).
12. de Boer, H. A., Comstock, L. J. & Vasser, M. The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proc. Natl. Acad. Sci.* **80**, 21–25 (1983).
13. Brosius, J., Erfle, M. & Storella, J. Spacing of the -10 and -35 regions in the tac promoter. Effect on its in vivo activity. *J. Biol. Chem.* **260**, 3539–3541 (1985).
14. Keasling, J. D. Synthetic biology and the development of tools for

metabolic engineering. *Metab. Eng.* **14**, 189–195 (2012).
15. Chen, Y. *et al.* Tuning the dynamic range of bacterial promoters regulated by ligand-inducible transcription factors. *Nat. Commun.* **8**, (2018).
16. Ellefson, J. W., Ledbetter, M. P. & Ellington, A. D. Directed evolution of a synthetic phylogeny of programmable Trp repressors article. *Nat. Chem. Biol.* **14**, 361–367 (2018).
17. Guan, L., Liu, Q., Li, C. & Zhang, Y. Development of a Fur-dependent and tightly regulated expression system in Escherichia coli for toxic protein synthesis. *BMC Biotechnol* **13**, 25 (2013).
18. Liu, X. *et al.* De novo design of programmable inducible promoters. *Nucleic Acids Res.* **47**, 10452–10463 (2019).
19. Rohlhill, J., Sandoval, N. R. & Papoutsakis, E. T. Sort-Seq Approach to Engineering a Formaldehyde-Inducible Promoter for Dynamically Regulated Escherichia coli Growth on Methanol. *ACS Synth. Biol.* **6**, 1584–1595 (2017).
20. Brewster, R. C., Jones, D. L. & Phillips, R. Tuning Promoter Strength through RNA Polymerase Binding Site Design in Escherichia coli. *PLoS Comput. Biol.* **8**, (2012).
21. Elvin, C. M. *et al.* Modified bacteriophage lambda promoter vectors for overproduction of proteins in Escherichia coli. *Gene* **87**, 123–126 (1990).
22. Kincade, J. M. & deHaseth, P. L. Bacteriophage lambda promoters pLand pRsequence determinants of in vivo activity and of sensitivity to the DNA gyrase inhibitor, coumermycin. *Gene* **97**, 7–12 (1991).
23. Chao, P., Law, S., Chen, P. T. & Hung, B. High production of heterologous proteins in Escherichia coli using the thermo-regulated T7 expression system. *Appl. Microbiol. Biotechnol.* **58**, 446–453 (2002).
24. Valdez-cruz, N. A., Caspeta, L., Pérez, N. O., Ramírez, O. T. & Trujillo-roldán, M. A. Production of recombinant proteins in E . coli by the heat inducible expression system based on the phage lambda pL and / or pR promoters. *Microb. Cell Fact.* **9**, 1–16 (2010).
25. Gvazdaitjs, B. G. *et al.* Temperature control in fermenters: Application of neural nets and feedback control in breweries. *J. Inst. Brew.* **100**, 99–104 (1994).
26. Gupta, A., Reizman, I. M. B., Reisch, C. R. & Prather, K. L. J. Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. *Nat. Biotechnol.* **35**, 273–279 (2017).
27. Kim, E. M. *et al.* Autonomous control of metabolic state by a quorum sensing (QS)-mediated regulator for bisabolene production in

engineered E. coli. *Metab. Eng.* **44**, 325–336 (2017).
28. Zargar, A., Quan, D. N. & Bentley, W. E. Enhancing Intercellular Coordination: Rewiring Quorum Sensing Networks for Increased Protein Expression through Autonomous Induction. *ACS Synth. Biol.* **5**, 923–928 (2016).
29. Bass, S. H. & Yansura, D. G. Application of the E. coli trp promoter. *Appl. Biochem. Biotechnol. - Part B Mol. Biotechnol.* **16**, 253–260 (2000).
30. Chevalet, L., Robert, A., Gueneau, F., Bonnefoy, J. Y. & Nguyen, T. Recombinant protein production driven by the tryptophan promoter is tightly controlled in ICONE 200, a new genetically engineered E. coli mutant. *Biotechnol. Bioeng.* **69**, 351–358 (2000).
31. Somerville, R. L. The trp promoter of escherichia coli and its exploitation in the design of efficient protein production systems. *Biotechnol. Genet. Eng. Rev.* **6**, 1–41 (1988).
32. Cavaleiro, A. M., Kim, S. H., Seppälä, S., Nielsen, M. T. & Nørholm, M. H. H. Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning. *ACS Synth. Biol.* **4**, 1042–1046 (2015).
33. Nour-Eldin, H. H., Hansen, B. G., Nørholm, M. H. H., Jensen, J. K. & Halkier, B. A. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res.* **34**, 1–8 (2006).
34. St-Pierre, F. *et al.* Complete Methods for 'One-step cloning and chromosomal integration of DNA'. *ACS Synth. Biol.* **2**, 537–541 (2013).
35. Mundhada, H., Schneider, K., Christensen, H. B. & Nielsen, A. T. Engineering of high yield production of L-serine in Escherichia coli. *Biotechnol. Bioeng.* **113**, 807–816 (2016).
36. Mundhada, H. *et al.* Increased production of L-serine in Escherichia coli through Adaptive Laboratory Evolution. *Metab. Eng.* **39**, 141–150 (2017).
37. Kim, S. H., Cavaleiro, A. M., Rennig, M. & Nørholm, M. H. H. SEVA Linkers: A Versatile and Automatable DNA Backbone Exchange Standard for Synthetic Biology. *ACS Synth. Biol.* **5**, 1177–1181 (2016).
38. Rennig, M. *et al.* Industrializing a Bacterial Strain for l -Serine Production through Translation Initiation Optimization. *ACS Synth. Biol.* **8**, 2347–2358 (2019).
39. Kildegaard, K. R. *et al.* Evolution reveals a glutathione-dependent mechanism of 3-hydroxypropionic acid tolerance. *Metab. Eng.* **26**, 57–66 (2014).
40. Yanofsky, C. Attenuation in the control of expression of bacterial operons. *Nature* **289**, 751–758 (1981).

41. Jones, K. L., Kim, S. & Keasling, J. D. Low-Copy Plasmids can Perform as Well as or Better Than High-Copy Plasmids for Metabolic Engineering of Bacteria. *Metab. Eng.* **338**, 328–338 (2000).
42. Yanofsky, C., Horn, V. & Gollnick, P. Physiological studies of tryptophan transport and tryptophanase operon induction in Escherichia coli. *J. Bacteriol.* **173**, 6009–6017 (1991).
43. Rugbjerg, P., Sarup-Lytzen, K., Nagy, M. & Sommer, M. Synthetic addiction extends the productive life time of engineered Escherichia coli populations. *Proc. Natl. Acad. Sci.* **115**, 2347–2352 (2018).
44. Khlebnikov, A. & Keasling, J. D. Effect of lacY Expression on Homogeneity of Induction from the P tac and P trc Promoters by Natural and Synthetic Inducers. 672–674 (2002).
45. Santos-sierra, S., Berzal-herranz, A. & Mu, A. J. Insights into the specificity of RNA cleavage by the Escherichia coli MazF toxin. **567**, 316–320 (2004).
46. Zhang, Y. *et al.* MazF Cleaves Cellular mRNAs Specifically at ACA to Block Protein Synthesis in Escherichia coli. *Mol. Cell* **12**, 913–923 (2003).
47. Werpy, T. & Petersen, G. *Top Value Added Chemicals from Biomass Volume I - Results of Screening for Potential Candidates from Sugars and Synthesis Gas*. *US Department of Energy Efficiency and Renewable Energy* (2004). doi:10.2172/15008859

4.7 Supplementary materials

Table S4.1. Primers used in this study.

Nr.	Primer sequence 5'- 3'	Description
2197	AGCTGAGGUCGCCTCAG C	Forward primer to amplify pSEVA27together with 2733 for mCherry and *mazF* insertion
2733	AAACTGGTCUCCTTCTTA AAGTTAAACAAAATTATT TCTAGAG	Reverse primer to amplify pSEVA27 together with 2197 for mCherry and *mazF* insertion
11707	AGATGCAUGGCGCCTAA CC	Forward primer to amplify pOSIP-trp-T7 together with jl26 to remove *trpL*
11708	AGAGGAUCCCCGGGTAC	Reverse primer to amplify pOSIP-KO-T7 together with 11834 for *trpR*+*trpO* insertion
11726	ACA GTA ATU AAT TAA CCT AGG CTG CTG CCA CC	Forward primer to amplify pCDF-1b together with 11795 for mCherry insertion
11795	ATG GTA TAU CTC CTT ATT AAA GTT AAA CAA AAT TAT TTC TAC AGG G	Reverse primer to amplify pCDF-1b together with 11726 for mCherry insertion
11830	ATC GTA CUC TTT AGC GAG TAC AAC CGG G	Reverse primer for amplifying *trpR* together with 11831 to clone with *trpO* on pOSIP-KO-T7
11831	AGA GGG CUTTA TCA ATC GCT TTT CAG CAA CAC CTC	Forward primer for amplifying *trpR* together with 11830 to clone with *trpO* on pOSIP-KO-T7
11832	AGT ACG AUGAG CTG TTG ACA ATT AAT CAT CGA AC	Forward primer for amplifying *trpO* together with 11833 to clone with *trpR* on pOSIP-KO-T7
11833	ATC GTG TTC AUTGT TAT TCT CTA ATT TTG TTC AAA AAA AAG CC	Reverse primer for amplifying *trpO* together with 11832 to clone with *trpR* on pOSIP-KO-T7

11834	ATG AAC ACG AUT AAC ATC GCT AAG AAC G	Foward primer to amplify pOSIP-KO-T7 together with 11708 for *trpR+trpO* insertion
11910	AGA CCA GTT UATG AGC AAG GGC GAG GAG GA	Forward primer for amplifying mCherry togther with 11911 to clone on pSEVA27
11911	ACC TCA GCUCTA CTT GTA CAG CTC GTC CAT GCC	Reverse primer for amplifying mCherry together with 11910 to clone on pSEVA27
jl17	ACT GCC UCGAAAGGTTTTGCGCC A	Forward primer to amplify pCDF-mCherry and pCDF-mazF together with jl18 to remove *lacI*
jl18	AGG CAG UGA GCG CAA CGC AAT TAA TG	Reverse primer to amplify pCDF-mCherry and pCDF-mazF together with jl17 to remove *lacI*
jl19	ATA TAC CAUG AGC AAG GGC GAG GAG GA	Forward primer for amplifying mCherry togther with jl35 to clone on pCDF-1b
jl26	ATT TCT TAT CCA UTGT TAT TCT CTA ATT TTG TTC TGT CGA TAC C	Reverse primer to amplify pOSIP-trp-T7 together with pOSIP_UF to remove *trpL*
jl35	AAT TAC TGUCTA CTT GTA CAG CTC GTC CAT GCC	Reverse primer for amplifying mCherry togther with JL19 to clone on pCDF-1b
jl106	ACCTCAGCUCTACCCAAT CAGTACGTTAATTTTGGC	Forward primer to amplify *mazF* from the genome together with jl107 to clone on pSEVA27
jl107	AGACCAGTTUATGGTAA GCCGATACGTACCC	Reverse primer to amplify *mazF* from the genome together with jl106 to clone on pSEVA27
jl108	AGATATACCAUGGTAAGC CGATACGTACCC	Forward primer to amplify *mazF* from the genome together with jl109 to clone on pCDF-1b
jl109	TAGGTTAATUCTACCCAA TCAGTACGTTAATTTTGG C	Reverse primer to amplify *mazF* from the genome together with jl108 to clone on pCDF-1b
jl110	AATTAACCUAGGCTGCT GCC	Forward primer to amplif*y* pCDF-1b together with jl110 for *mazF* insertion

jl111	ATGGTATATCUCCTTATT AAAGTTAAACAAAATTAT TTCTAC	Reverse primer to amplify pCDF-1b together with jl111 for mazF insertion

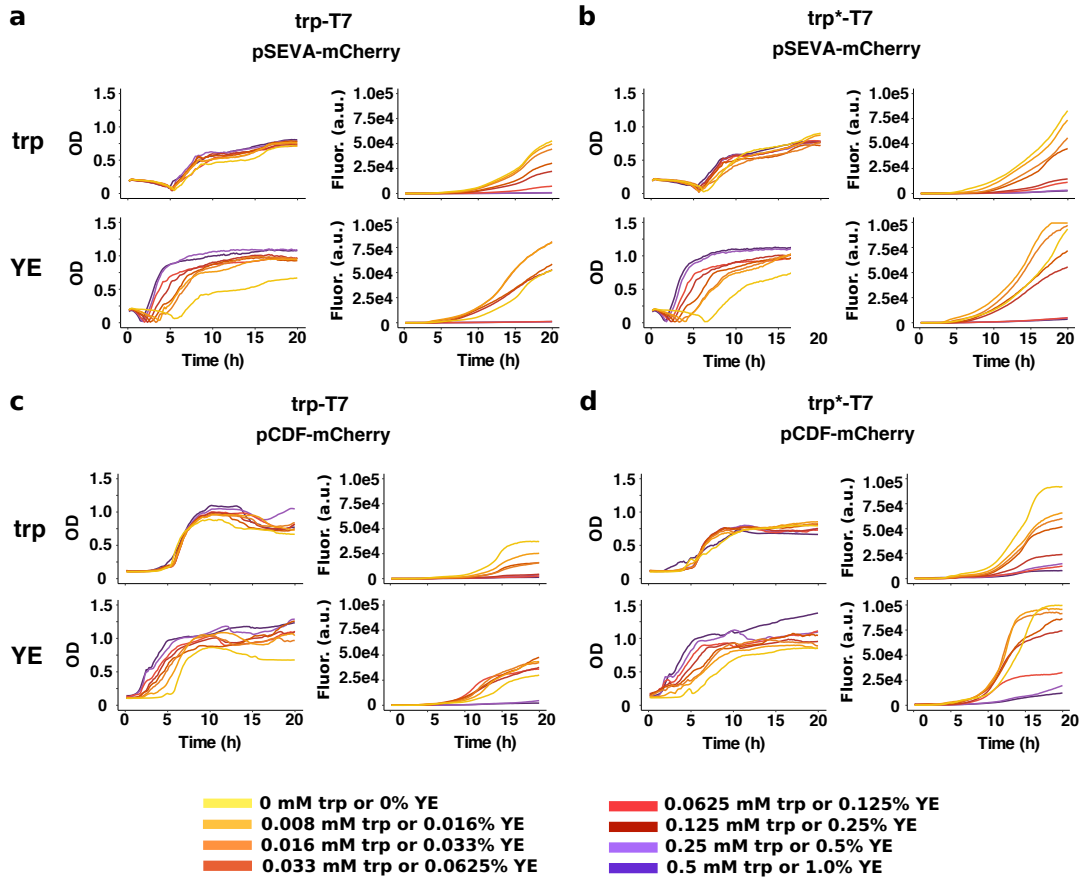


Figure S4.1. Growth and fluorescence curves in different starting concentrations of tryptophan (trp) or yeast extract (YE). (a) OD and fluorescence for trp-T7 with pSEVA-mCherry. (b) OD and fluorescence for trp*-T7 with pSEVA-mCherry. (c) OD and fluorescence for trp-T7 and pCDF-mCherry. (d) OD and fluorescence for trp*-T7 and pCDF-mCherry. The growth and fluorescence were calculated as the average of three biological replicates.

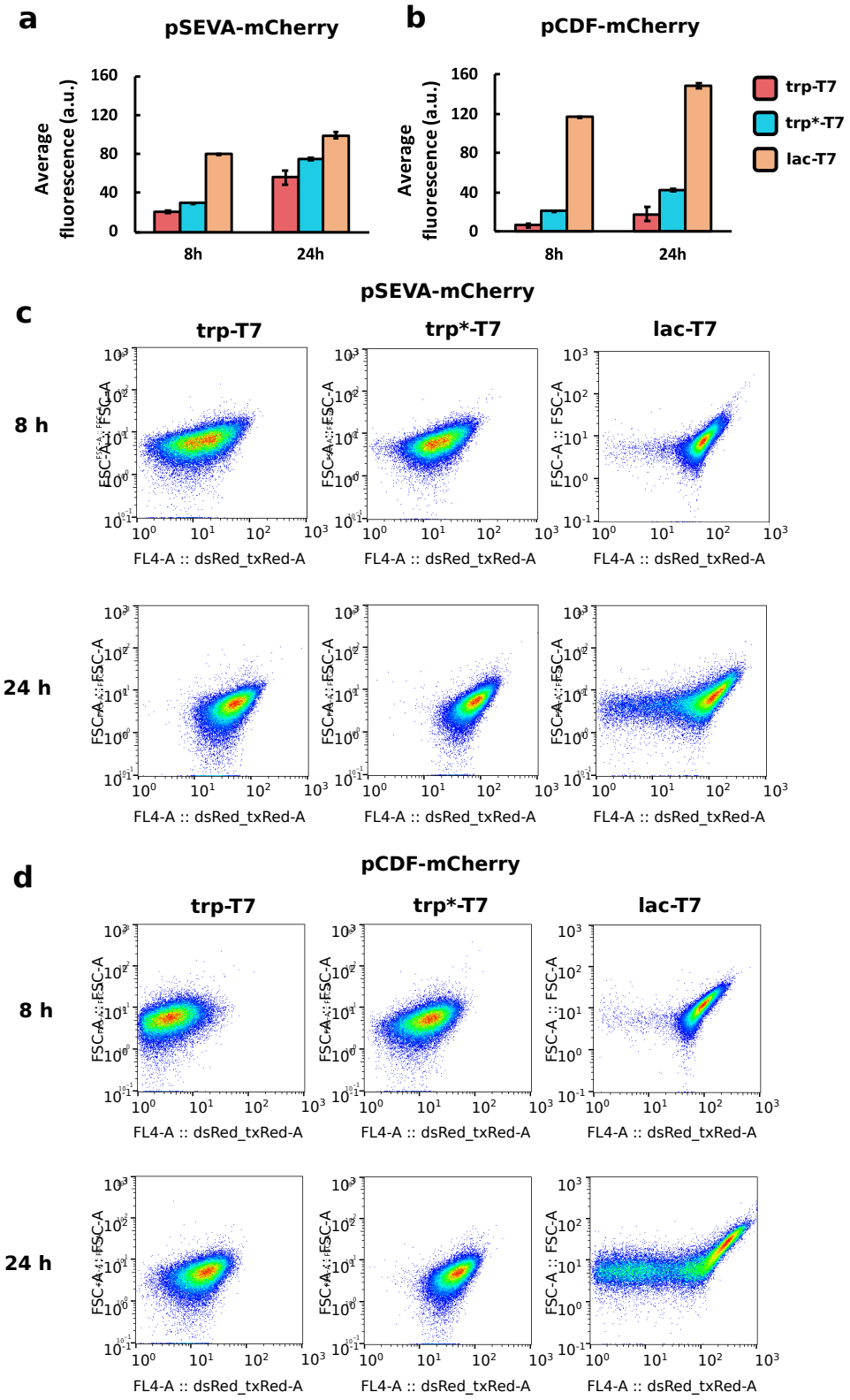

a
pSEVA-mCherry
Average fluorescence (a.u.)
160
120
80
40
0
8h
24h
b
pCDF-mCherry
trp-T7
trp*-T7
lac-T7
c
pSEVA-mCherry
trp-T7
trp*-T7
lac-T7
8 h
24 h
FSC-A :: FSC-A
FL4-A :: dsRed_txRed-A
d
pCDF-mCherry

Figure S4.2. Flow cytometry data from 8 and 24 h of growth in minimal medium. (a) Average fluorescence of pSEVA-mCherry cultures when expressed from the trp-T7, trp*-T7 or lac-T7 expression system. (b) Average fluorescence of pCDF-mCherry cultures when expressed from the trp-T7, trp*-T7 or lac-T7 expression system. (c) Scatterplots showing forward scatter (y-axis) and mCherry fluorescence (x-axis) from pSEVA-mCherry using the trp-T7, trp*-T7 or lac-T7 expression system. (d) Scatterplots showing forward scatter (y-axis) and mCherry fluorescence (x-axis) from pCDF-mCherry using the trp-T7, trp*-T7 or lac-T7 expression system. The trp strains were induced by growth in tryptophan deplete medium, and the lac-T7 system was induced by addition of 1 mM IPTG at OD ~0.3-0.35. OD was similar for all strains at the sampling time points. For (a) and (b), the average values and standard deviations were calculated from three biological replicates. For (c) and (d), each histogram shown is a representative sample out of three biological replicates.

Chapter 5

CONCLUSIONS AND OUTLOOK

With microbial cell factories, we have the opportunity to re-think and re-invent the way we produce chemicals, fuels and materials. Using renewable feedstock, we can replace today's polluting and hazardous production methods with sustainable and green solutions that are part of a circular bioeconomy. To reach the stage where microorganisms are the go-to production vessels, a significant amount of research and development is needed.

This thesis presents a small step forward on the way toward superior cell factories. It has focused on methods that can be used to increase product yields by controlling microbial growth and production. This includes the screening, identification and application of gene targets that improved protein yields by decoupling growth and production, and the design and construction of an autoinducible expression system for inexpensive induction of protein or product pathway expression.

In the first and second study of this thesis, the CRISPRi-based, genome-wide screen was shown to be a fast and efficient tool for finding suitable targets for growth decoupling. This approach is directly translatable to other microorganisms. The identified targets were furthermore shown to inhibit growth, but not metabolic activity, which is a desirable trait for these types of growth switches. In the third study of this thesis, a system enabling low cost auto-induction of gene expression during fermentation was developed. The autoinducible expression system was shown to be suitable for production of both proteins and biochemicals. It could serve as an interesting alternative to other T7-based expression systems for small and large-scale bio-based production.

Clearly, further improvement is needed for feasible application of the growth decoupling system in large-scale processes. CRISPRi induction should be carried out using an autoinducible or temperature-controlled induction system, avoiding the addition of expensive inducers to the media; here, the trp-T7 expression system could be an interesting candidate. Another way forward would be to develop temperature-sensitive variants of the selected target enzymes, and run the fermentation in a two-stage process where temperature is increased when switching from growth to production. This is an interesting approach that has been demonstrated to work for other target proteins in small and large scale.

Together with classical metabolic engineering strategies, growth decoupling and dynamic regulation will play a central role in the development of tomorrow's cell factories. Balancing of resources and downregulation of growth and competing essential pathways is an important strategy to maximize the amount of product that can be made in a bioprocess.

To further advance the field of cell factory development, it is critical to decrease cell factory construction costs and increase the speed of the design-build-test-learn cycle. The advancement of (automated) high-throughput DNA assembly techniques and decreased costs of DNA synthesis and sequencing has already sped up the build step. New high-throughput screening methods and further standardization of biological parts is key to reduce the duration of the test and design steps. In addition, improved metabolic models and new big data approaches using artificial intelligence and machine learning will play a central role in supporting the learn and design steps. Altogether, this will lead the way forward when we develop the next generation of microbial cell factories.