Development of synthetic biology tools for growth decoupled production and protein expression

Rong, Yixin

Publication date:
2023

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

DTU Biosustain
The Novo Nordisk Foundation Center for Biosustainability

PhD thesis

Development of synthetic biology tools for growth decoupled production and protein expression

Yixin Rong

Development of synthetic biology tools for growth decoupled production and protein expression

PhD thesis

Yixin Rong

The Novo Nordisk Foundation Center for Biosustainability

Technical University of Denmark

July 2022

Preface

This thesis is written as part of the fulfilment of the requirements to obtain a PhD degree at the Technical University of Denmark. The work presented is conducted between March 1st 2019 and July 15th 2022 in the Bacterial Cell Factories research group at the Novo Nordisk Foundation Center for Biosustain, Technical University of Denmark, Kongens Lyngby, Denmark. The work was supervised by Professor Alex Toftgaard Nielsen. The project was funded by the Novo Nordisk Foundation within the framework of the Fermentation-based Biomanufacturing Initiative (FBM), grant number: NNF17SA0031362.

Abstract

With each passing day, the need for humankind to transition how our economy operates becomes clearer and clearer. One of the most important transitions is moving away from using non-renewable fossil fuels and towards renewable and sustainable technologies. While windmills and electric cars can go far in addressing fossil fuel consumption in the energy and transportation sectors, we also need to replace oil based chemicals and materials.

Luckily, the solution may already be in our hands. Humans have used microorganisms for thousands of years to make fermented drinks and foods, long before discovering what microorganisms were. Technological breakthroughs in the last 50 years have enabled us to engineer and rewire microorganisms towards producing the compounds that are currently derived from fossil fuels, through the process of microbial fermentation. The key advantage of this bio-based production process is the ability to use renewable raw materials, such as sugars, agricultural and forestry waste, or even waste gasses from heavy industry.

However, there are still many challenges to overcome before bio-based production can become the predominant production process. Building and operating the production facilities is extremely costly. The microbial strains and production processes have to therefore be highly efficient. This thesis focuses on developing strains and tools towards improved two-stage fermentations in the bacterial platform organisms *Escherichia coli* and *Bacillus subtilis.* Two-stage fermentations divide the fermentation into dedicated growth and production phases, which allows for more efficient use of the feedstock, which is a major contributor. This division of phases can be mediated by synthetic biology tools that are used to dynamically control the metabolism of the cells. Lastly, this thesis touches on heterologous protein expression, which is an important part of microbial cell factory engineering.

Dansk resumé

For hver dag som går, bliver behovet for menneskeheden at ændre måden vores økonomi opererer på tydeligere og tydeligere. En af de vigtigste ændringer er at erstatte brugen af ikke-fornyelige fossile brændstoffer med fornyelige og bæredygtige teknologier. Selvom vindmøller og elbiler kan nå langt med hensyn til at erstatte fossile brændstoffer i energi- og transportsektoren, har vi også brug for at erstatte kemikalier og materialer som produceres fra olie.

Heldigvis findes løsningen muligvis allerede i dag. Mennesker har brugt mikroorganismer til at fremstille fermenterede mad og drikke i mange tusinder af år, lang før mikroorganismerne blev opdaget. Teknologiske gennembrud i de seneste 50 år muliggjort konstrueringen af mikroorganismer, som gennem mikrobiel fermentering kan producere de stoffer som i dag produceres fra olie. Den store fordel ved biologisk produktion er evnen til at bruge fornyelige råmaterialer, såsom sukker, affald fra land- og skovbrug, og endda gasser udledt af tung industri.

Der er desværre mange udfordringer tilbage som skal overkommes, før biologisk produktion kan blive den dominerende produktionsproces. Oprettelse og drift af produktionsfaciliteter er ekstremt kostbart. Derfor skal de mikrobielle stammer og processer være meget effektive. Denne afhandling fokuserer på udviklingen af stammer og værktøjer til at forbedre to-stadie fermenteringer for bakterielle platform organismer *Escherichia coli* og *Bacillus subtilis.* To-stadie fermenteringer deler fermenteringen i dedikerede vækst- og produktionsfaser, hvilket gør mere effektiv brug af råmateriale, som er meget kostbart. Denne deling kan opnås ved hjælp af værktøjer fra syntetisk biologi, som dynamisk kan kontrollere den cellulære metabolisme. Endeligt berører afhandlingen også om heterolog ekspression af proteiner, hvilket er en vigtig del i konstruering af mikrobielle cellefabrikker.

Acknowledgements

PhD projects rarely go as planned, and my PhD project was definitely no exception. There were many twists and turns along the way, but I made it through in the end, with many people who helped me along the way, who I would like to extend my gratitude towards.

First and foremost I would like to thank all of my supervisors, Alex Toftgaard Nielsen, Sheila Ingemann Jensen and John Woodley. I really felt that my supervisors had in mind what was best for me and my project, which isn't always true for supervisors. Thank you Alex for all of your advice and guidance, and most of all your always positive attitude, something which I at times had a hard time maintaining. Thank you Sheila for all the help and guidance throughout my project, for helping me troubleshoot all the failed experiments, and for helping me becoming a better scientist. Thank you John for our perhaps too few, but always insightful discussions. I wish we had more opportunity to collaborate than what the circumstances ended up permitting.

I would of course also like to thank all the current and former members of the BCF(O) group for creating a nice work environment. Thanks to Jenny, who helped me around the lab in the beginning of my PhD. Thanks to the Dutch army of Jasper, Viviënne and Philip, my knowledge about the Netherlands has definitely increased during my PhD. Thanks to Arrate with whom I could always share the quirks of working with Sheila. Thanks to Torbjørn, Amalie, Marie, Ivan, Regiane, Stephanie, Adrian, Ácil, Lucas and Gustavo as well.

Thanks to Professor Kresten Lindorff-Larsen and for hosting me in your group at the University of Copenhagen. The pandemic and other circumstances made it difficult to plan external stays, but I'm glad to have had the chance to learn about your research and collaborate on the paper. Thanks also to the KLL group for welcoming me in, it was a nice albeit quite short of a stay.

Thanks to the FBM and the FBM PhD community, it has been quite the journey from 6 PhD's sitting in a meeting room to where the program is today. It was nice having a group of people that were in the same boat as you. Also thanks to the CfB PhD community, both were a nice community to be a part of. Thanks to CfB staff who have helped me, such as analytics, support staff and administrative staff.

Last but not least, I want to thank friends and family. Very special thanks go to my parents of course, who I can always count on for their unwavering support. It is quite special to have both your parents also work in academia, and although at times I felt like I had a fourth and fifth supervisor, your support was invaluable during these three and a half years of my PhD.

List of publications

Y. Rong, S. I. Jensen, J. Woodley, A. T. Nielsen*, “Modulating metabolism through synthetic biology – opportunities for two-stage fermentation”. Manuscript ready for submission.

Y.Rong, E. Özdemir, S. Li, A. Sainz de la Maza Larrea, A. T. Nielsen, S. I. Jensen*, “CRISPRi-mediated metabolic switch enables concomitant “anaerobic” and aerobic fermentations”. Manuscript in preparation.

Y. Rong, R. Alves de Oliveira, K. B. Falkenberg, Á. Will, A. T. Nielsen, S. I. Jensen*, ” Expanding the CRISPRi toolbox in Bacillus subtilis for application towards metabolic switches”. Manuscript in preparation

Y. Rong, S. I. Jensen, K. Lindorff-Larsen*, A. T. Nielsen*, ” Folding and expression of heterologous proteins in bacterial cell factories”. Manuscript ready for submission.

*Corresponding author

Thesis outline

Humankind has harnessed the power of microorganisms, long before even discovering what these organisms were. Fast forward a few millennia, the advent of recombinant DNA technologies in the latter half of the 20th century sparked rapid developments in the field of industrial biotechnology, enabling the production of various products through microbial fermentations. At the same time, it has become clearer than ever that the need to transition away from fossil fuels and towards a renewable bio-economy is urgent. Bio-based manufacturing plays a key role in this transition, serving as a sustainable alternative to producing chemicals, fuels and materials. Unfortunately, broad industrial implementation of bio-based manufacturing is still a very challenging endeavor, with one of the main challenges being the high costs of both establishing and operating bio-based production processes.

The microbial cell factories at the heart of bio-based manufacturing need to be engineered in order to reach titers, rates and yields (TRY) that make the manufacturing process economically viable. To achieve these highly efficient cell factories, both traditional “static”, and dynamic metabolic engineering need to be explored. Developments in synthetic biology enables controlling cellular metabolism during the fermentation, in order to favor product formation over growth once sufficient biomass concentrations are achieved.

Another challenge in cell factory engineering is the expression of heterologous proteins, which is used to introduce new proteins or biosynthetic pathways, or to enhance the native metabolism. However, due to various factors, not all proteins can be expressed in industrial platform organisms.

The focus of this PhD thesis is engineering strains and using synthetic biology tools that can contribute to decreasing the costs of fermentation, mainly CRISPR interference, to knockdown expression of genes that result in changes in phenotype that can improve the economics through increasing yield. This work is done in the bacterial platform organisms *Escherichia coli* and *Bacillus subtilis*. Furthermore, the thesis will address the challenges of heterologous protein expression with a thorough literary review of the cellular mechanisms that are important for protein expression, and how they relate to experimental strategies that are applied to improve heterologous protein expression.

Chapter 1 serves as a general introduction to industrial biotechnology, bio-based manufacturing and microbial cell factories. The chapter briefly describes cell factory/metabolic engineering, CRISPR interference, and other topics relevant for the following chapters.

Chapter 2 is a literary review on using synthetic biology tools to mediate two-stage fermentations. The chapter discusses how synthetic biology based methods compare to other approaches, and the status in regards to large scale application.

Chapter 3 explored using CRISPRi as a metabolic switch to change *E. coli* metabolism to anaerobic metabolism under aerobic conditions; this enabled co-cultivation with another strain to consume the byproduct and produce a secondary product. This unique setup could be beneficial compared to conventional monocultures.

Chapter 4 tuned the protein levels of dCas9 in *B. subtilis* for improved compatibility with various promoters when used for CRISPRi metabolic switches.

Chapter 5 reviewed cellular mechanisms that could affect heterologous protein folding, and how they differ between prokaryotic and eukaryotic organisms. Furthermore, strategies to improve heterologous protein expression in bacterial cell factories were reviewed and related to the cellular mechanisms.

Table of Contents

Chapter 1

Introduction

Long before knowing what microbial organisms were, humans learned how to harness their abilities for the benefit of themselves by making food and drink preserve for longer, in the form of fermented drinks such as beer and wine, and various kinds of fermented foods. The oldest known instances of fermented drinks and foods date back thousands of years; the first larger scale beer brewing operations date back all the way to 3600 BC in ancient Egypt [1]. Moreover, fermented foods and drink are a part of cultures all over the world, and still play an important role of what is being consumed today. It would not be until thousands of years after the onset of food and drink fermentation that the first microorganisms would be observed in the 17th century by Robert Hooke and Antony van Leeuwenhoek using primitive microscopes [2][3]. In the 19th century, scientists such as Louis Pasteur, Joseph Lister and Robert Koch established the theories of microscopic germs and bacteria being responsible for food spoiling, fermentation, and infections/diseases [4]. The earliest non-food or drink bio-based chemical production also started in the 19th century in the form of ethanol, which was used in e.g. oil lamps and as an alternative fuel to petroleum for combustion engines [5]. Spurred on by demands from the First World War, industrial ethanol production was followed in the first half of the 20th century by the acetone-butanol-ethanol (ABE) fermentation using the bacterium *Clostridium acetobutylicum*, also known as the Weizmann process [6]. Likewise, demand and interest deriving from the Second World War lead to the development of large-scale production processes for the antibiotic penicillin, discovered in 1928 by Alexander Fleming as a compound produced by filamentous fungi of the *Penicillium* genus [7]. Both of these products were among the first to be produced in large scale using bio-based chemical manufacturing processes.

All early biobased chemical manufacturing relied on wild-type microbes; although the inheritance of traits was described by Mendel and others, the mechanism of which this information was passed on to subsequent offspring was still unknown. This changed in 1953 when James Watson, Francis Crick, and Rosalind Franklin discovered the structure of DNA: two complementary chains of nucleotides wound in a double helix [8]. The first major developments towards modern biotechnology happened soon after in the 1970's, with the discovery of type-II restriction endonucleases and their application in cloning DNA from different sources to create recombinant DNA, and Sanger sequencing of DNA [9][10]. One of the first

recombinant products available for purchase was recombinant human insulin produced in engineered strains of the bacterium *Escherichia coli*, which was approved in 1982 [11]. Today, the plethora of products and technologies produced by engineered microbes are used to address many different problems and demands in modern society, ranging from fuels, chemicals, and materials to pharmaceuticals and even microbial foods and animal feed.

1.1 Industrial biotechnology and bio-based manufacturing

Industrial biotechnology can broadly be defined as the industrial scale manufacturing of compounds, like chemicals and proteins, using biological catalysts such as microbes or enzymes. Nature is the greatest engineer, and through billions of years of evolution, solutions have emerged that allow life to survive in various different environments. Progress in science now enables us to harness these solutions, both to produce new compounds, but also to have an alternative to current production processes. One of the main driving forces behind the growth of industrial biotechnology is to reduce dependence on non-renewable fossil resources such as oil.

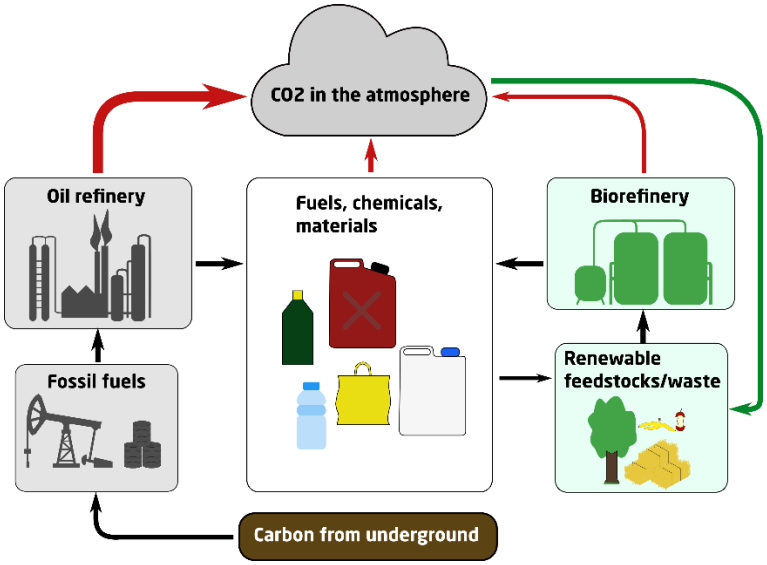


Figure 1 *Production of fuels and chemicals through petrochemical vs bio-based manufacturing. With petrochemicals, the flow of carbon is linear from oil extraction to the release of carbon in the production and consumption of the product. With bio-based manufacturing, less GHG is emitted during production, and the renewable feedstock are able to fix some of the carbon back, which then can go through another cycle.*

The key advantage when comparing bio-based and petrochemistry based manufacturing lies in the ability of bio-based manufacturing to operate with renewable feedstocks and carbon circularity, as illustrated in a simplified way in Figure 1. It has been evident that industrial biotechnology has the potential to replace oil since its beginnings in ethanol production as fuel for combustion engines. Today, the sentiment for changing from petrochemical manufacturing to bio-based manufacturing is stronger than ever, as industrial biotechnology plays a key role in achieving the UN Sustainable Development Goal of responsible production and consumption and transitioning towards a circular bio-economy [12][13].

Other advantages of bio-based production of chemicals are the stereoselectivity of enzymes compared to chemical catalysts, the often milder and/or less hazardous conditions and reagents used compared to synthetic chemistry and petro-chemistry, as well as potentially reducing the CO_2 emissions resulting from production [14]. It is not unusual for conventional chemical processes to have operation conditions with high temperatures from 200 °C to ≥600 °C, and high pressures from 50 atmospheric pressure (atm) to 300 atm [15][16]. For example, chemical synthesis of higher alcohols (C3 and higher) from ethanol, hydrogen, and carbon monoxide gas using metal catalysts operates at 320 °C and 79 atm [17]. Some chemical synthesis processes also utilize highly toxic pre-cursors; one example would be the production of methyl methacrylate (MMA), which polymer form (PMMA) is used for clear plastics [18]. Here the most prevalent synthesis route requires hydrogen cyanide (HCN) and produces large amounts of ammonium bisulfate as byproduct. Bio-based manufacturing processes on the other hand mainly operate at temperatures between 20 and 50 °C, and at, or moderately higher, than atmospheric pressure (up to 15 atm) [19]. They also inherently cannot use highly toxic compounds that kill the cells or denature the enzymes used as catalysts. In terms of CO_2 emissions, in the vast majority of cases, the bio-based production processes have a lower contribution to greenhouse gas emissions compared petrochemical processes of the same product [20]–[24].

Industrial biotechnology is a growing industry; market research has indicated that the global biorenewables sales increased from 207.2 billion US dollars (USD) in 2015 to 475.9 billion USD in 2020, with an annual growth rate of 18% in these five years [25]. The market share also increased from 10.5% in 2015 to 19.5% in 2020. Dividing the industry into categories, the largest category is polymers and fibers at 38.8%,

followed by consumer chemicals, specialty chemicals, and base chemicals, at 24%, 20.7% and 16.5% respectively. These numbers can differ quite substantially between studies; a study looking at 50 different commercialized bio-based chemicals out of 208 total in Europe found that their average market share compared to the total market was only 3% [26]. Although the numbers are quite different, it is clear that bio-based production is still in the minority when compared to production using non-renewable feedstocks, despite sustainability having been high on the global agenda for a number of years. This raises the question: What are some of the key challenges, technological, economic, and political, that need to be overcome for bio-based manufacturing to become the predominant way of production?

Policy can play a major role in broader adoption of bio-based manufacturing. Continued research and innovation are necessary to develop new and improved strains, processes, and products to drive the industry forward [27]. Research and innovation requires education of skilled scientists and engineers, where inter-disciplinarity has been emphasized as an important qualification [28]. Policies that support funding research and education are therefore crucial. Furthermore, policies that favor the fossil fuel industry, such as post-tax subsidies, should be removed in order to make the competition fair; petrochemicals already have many advantages in mature processes, supply-chains, and infrastructure [29]. Additionally, a carbon-tax could further increase the competitiveness of bio-based manufacturing. Lastly, policies that make it easier for bio-based production companies to acquire investment funding, such as government backed low-interest loans, could help to commercialize more bio-based manufacturing processes [30].

Sourcing and supply of renewable feedstocks are important for the economic viability and minimizing possible negative impacts, such as land use and greenhouse gas emissions. In order to meet the total demand for chemicals, the supply of feedstock has to be able to keep up; this can in turn compete with food production, which in itself is a challenge with the growing world population and lower crop yields due to more droughts and floods as well as soil depletion [28][31]. Alternatively, using more land for biomanufacturing feedstocks can lead to deforestation and desertification. It is therefore important to establish robust supply chains for second generation lignocellulosic biomass, such as agricultural waste and waste from

forests (paper and wood production), and explore additional feedstock sources such as food waste and industrial off-gasses [32], [33].

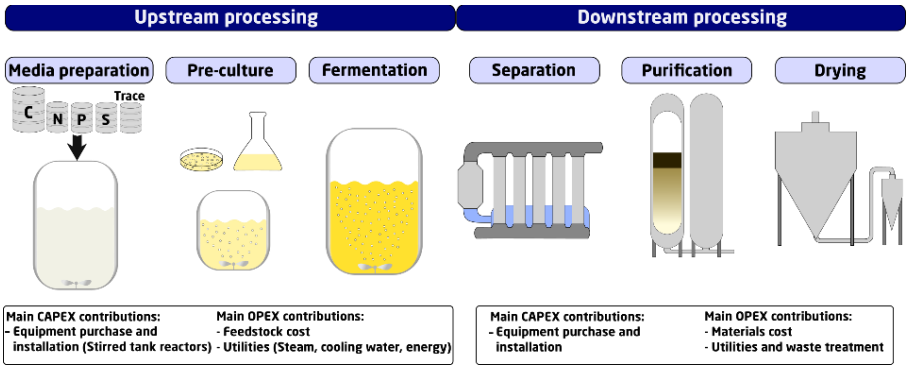


Figure 2 *A bioprocess consists of upstream processing and downstream processing. In upstream processing, the fermentation media gets prepared and the production organism is propagated from cold storage in smaller vessels first. This pre-culture is then used to seed the main fermentation. The broth from fermentation goes through downstream processing, where the biomass is separated, and the remaining liquid goes through purification and water removal steps before the finished product is formulated, e.g. by a drying step.*

Both the development and deployment of a bio-based manufacturing process are slow and expensive. Advances in technology have accelerated the process of developing production strains; however, the costs of building and operating a large-scale biorefinery remain prohibitive [34]. Figure 2 shows the different steps in a biomanufacturing process and equipment and costs related to them. A bioprocess starts with the preparation of the fermentation media and the inoculation of a pre-culture, which will be used to seed the main fermentation. In the main fermentation, the feedstock is converted into biomass and product under controlled conditions (temperature, pH, oxygen etc.), where the majority of industrial fermentations operate in fed-batch mode. Once the fermentation is complete, the biomass is separated and the product goes through different up-concentration and purification steps.

The costs related to biorefineries can be divided to capital expenditure or CAPEX, which are the initial costs of building the facilities, and operational expenditure or OPEX, which are the costs incurring from running the production [35]. In terms of CAPEX, the majority of the cost comes from purchase of equipment and associated installation [36][37]. Within the different pieces of equipment, the stainless steel

stirred tank reactors (STR) and agitators, as well as downstream processing (DSP) equipment make up a large part of the total cost [38][39]. CAPEX is influenced most by the production rate and product titer from the fermentation, as lower productivity and titer would require a larger capacity to reach the same production goals. The price of the product is also a deciding factor for the production goal; lower value products need to be produced in higher amounts to be economically viable. For OPEX, the largest contributors are the feedstock and other chemicals (e.g. acids and bases), followed by utilities, such as steam, cooling water and aeration, and DSP [37][38]. One of the fermentation parameter that influences OPEX the most is the yield of product per substrate, especially for lower value products, as well as the downstream process. Decreasing these costs is crucial in making bio-based manufacturing more economically viable.

1.2 Microbial cell factories – the workhorses of industrial biotech

Microbial cell factories are the catalysts in industrial biotechnology that are responsible for converting the renewable feedstocks into products like chemicals and proteins. The term "cell factory" applies to any microbial organisms, both prokaryotic and eukaryotic, that are used for production in industrial biotechnology [40]. Cell factories can be classified as model/platform organisms vs. non-model or novel production organisms. Alternatively, they can also be categorized as generalist vs specialist organisms. Platform organisms are generally very extensively studied and characterized, meaning they have many tools available for genetic modifications, databases of genes and pathways, as well as models such as genome scale metabolic models or GEM's [41]. There is also more knowledge and expertise on scale up and media optimization for platform organisms on the upstream processing side. On the other hand, platform organisms are often not specialized or pre-disposed to produce a specific compound or family of compounds, nor are they suitable for growth in "extreme" environments. Non-model organisms have fewer tools available, but can have other interesting properties for industrial biotechnology. For example, fermentations with halophilic and thermophilic bacteria are less likely to get contaminated and thus may not require steam sterilization of the reactors, meaning less expensive reactors could be used [42], [43]. Acetogenic bacteria could be used to fix CO_2 of industrial off-gasses via gas fermentations [33]. For the tools and methods focus of this thesis, it made sense to

work with the platform organisms *E. coli* and *B. subtilis*, so the tools and methods could potentially be applied towards many different products.

1.2.1 Escherichia coli

E. coli, the organism of choice in Chapter 3, is perhaps the best characterized model organism of all, or at least of prokaryotic organisms. It is a rod-shaped, Gram-negative bacterium, which is naturally found in the gut microbiota. *E. coli* is a facultative anaerobe, meaning it can grow under both aerobic and anaerobic conditions [44]. It is a mesophilic bacterium, meaning it has a moderate optimal growth temperature of 37 °C. The genome of *E. coli* was first sequenced in 1997 [45]. Many tools have since then been developed that allow for efficient editing of the *E. coli* genome, both gene insertions, knock-outs and more precise edits like point mutations [46]–[48]. Its metabolic pathways, regulation, and the genes and enzymes responsible are well studied, and the information is easily accessible via databases such as EcoCyc [49]. Cell factory engineering of *E. coli* can further be aided by computational tools, such as GEM's and gene expression tuning tools such as EMOPEC [50], [51].

Due to its fast growth rates and otherwise ease of cultivation, *E. coli* is a prevalent host organism for industrial biomanufacturing. As mentioned in the introduction, *E. coli* was originally known as a producer of therapeutic proteins, such as recombinant human insulin. To this day, *E. coli* is still a commonly used production organism for non-glycosylated therapeutic proteins [52]. Nevertheless, due to the tools and knowledge available, *E. coli* has also been engineered to be a capable producer of different chemicals. Several industrial production processes are currently operational with *E. coli* as the production organism, with the most successful example being 1,3-propanediol production by DuPont, where engineered strains can reach titers as high as 130 g/L [53]. Other chemicals include 1,4-butanediol produced by BASF and Genomatica as well as various different amino acids for nutrition and animal feed, where the production of serine and cysteine are currently being commercialized by a start-up company Cysbio [44][54]. Amino acids for human or animal consumption require approval of the production process as "Generally Regarded as Safe" or GRAS, which has previously been granted to L-methionine for animal feed produced in *E. coli K12* [44]. Many other compounds have also been successfully produced in *E.* coli, but never commercialized due to them not being economically viable. One drawback for industrial production using

E. coli (or other bacteria) include the potential susceptibility to phage contamination, which may require engineering phage resistant strains and designing process steps to minimize the risk of contamination [55].

1.2.2 Bacillus subtilis

B. subtilis, the organism of choice for chapter 4, is also a very well characterized prokaryotic model organism, and is often considered the Gram-positive equivalent of *E. coli* [56]. It is a rod-shaped bacterium and was originally isolated from plant roots and surrounding soil [57]. *B. subtilis* was initially thought to be an obligate aerobe, meaning it can only grow in the presence of oxygen, but studies since then have shown that it can use nitrate as an alternative external electron acceptor [58]. However, fermentation media does not commonly use nitrate as the nitrogen source, meaning they have to be fermented aerobically. *B. subtilis* is a mesophilic bacterium with a similar optimal growth temperature as *E. coli*, but interestingly can grow at higher temperatures up to 52 °C [59]. Its genome was also first sequenced in 1997, shortly after the first *E. coli* genome [56]. For databases, there is a *B. subtilis* version of EcoCyc called BsubCyc [60]. Unlike *E. coli*, *B. subtilis* has a broader GRAS status, enabling its use in food applications [61].

Due to it being an environmental bacterium, *B. subtilis* has evolved different survival mechanics that can be both beneficial and problematic for application in industrial biotechnology. One of these mechanics is natural competence, where the bacterial cells are able to receive extracellular DNA, a quorum-sensing mechanism regulated by ComK and ComS [62]. This mechanism can be induced by growing the cells in a starvation media [63]. Alternatively, ComK and ComS can be expressed with an inducible promoter [64]. One can then provide a selection cassette for a knockout or an expression cassette for insertion of genes, flanked by ~500 bp long homology arms targeting an insertion site to facilitate a double crossover homologous recombination event. Commonly targeted sites include *amyE, sacB, bpr, epr* and *vpr,* with additional sites identified and characterized in a recent study on a standardized vector toolbox for *B. subtilis* called proUSER 2.0 [65]. A survival mechanism that is detrimental for industrial production is its ability to form endospores when experiencing starvation. These spores have a thick coating that makes them resilient to external stresses, but are metabolically inactive and the mother cell is lysed in the sporulation process [66]. Not only is sporulation detrimental for production, the spores can also complicate waste biomass disposal and steam sterilization of tanks,

as they require higher temperatures or longer durations of heat for inactivation [67]. Sporulation can be reduced by deletion of regulatory genes Spo0A, SigF, SigE or SigG [56][68].

Being a Gram-positive environmental bacterium, *B. subtilis* is capable of secreting large amounts of enzymes, like amylases and proteases, through its singular cell membrane in order to break down complex substrates and use them as nutrients [69]. The secretion is mainly facilitated by the Sec and twin arginine translocation (Tat) pathway, where the main difference is the majority of proteins secreted through the Sec-pathway fold outside the cell, while proteins secreted through the Tat-pathway fold prior to secretion [70]. Industrial enzyme production is pivoting towards other *Bacilli* such as *B. licheniformis*, but *B. subtilis* remains as an excellent thoroughly characterized model organism [56][69]. For chemical production, *B. subtilis* is mainly used for production of nutritional supplements, such as DSM's production of riboflavin (vitamin B2) and Omnigene/BASF's production of (R)-pantothenate [56]. *B. subtilis* has also been explored as a host for other chemicals, such as surfactants, N-acetylglucosamine (GlcNAc), and 3-hydroxypropionate, but these processes were not commercialized [61], [71]–[74].

1.3 Metabolic engineering

Although some organisms are naturally able to produce certain products at moderately high amounts, such as *B. subtilis* and secreted enzymes, they tend not to overproduce said products to the extent that is needed for commercial viability. This is due to the organism not gaining any evolutionary advantage from this production beyond a certain point, where the production simply becomes a metabolic burden. It is therefore necessary to engineer the cellular metabolism to overproduce the desired product. As described above for the organisms used in this thesis, metabolic engineering has resulted in commercial production of several different compounds, with many other examples of commercialized processes using other metabolically engineered organisms.

The process of metabolic engineering is often described as a four-step cyclical model of Design, Build, Test and Learn or DBTL for short (Figure 3):

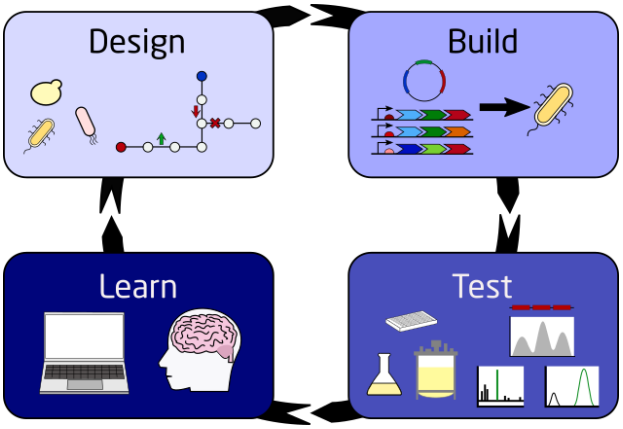


Figure 3 *The Design-Build-Test-Learn (DBTL) cycle used in metabolic engineering*

Design: In the first phase of the cycle the design choices are made. For the first iterations of the cycle, the design is mainly about the fundamental aspects of the end target of the engineering process. This includes the choice of which organism to use, depending on many factors, such as its natural ability to produce the product or pre-cursors to the product, availability of tools for genetic modification, potential product toxicity, and ability to utilize different feedstocks. If the product is not naturally produced, the enzymes catalyzing the missing steps will need to be expressed heterologously. These enzymes can be found from prior art, or databases such as KEGG and Brenda [75][76]. Alternatively, if the pathway or parts of the pathway are unknown, there are bioinformatics tools that can help predict possible pathways to the product [77][78]. For engineering the native metabolism, common rational engineering strategies include deletion of reactions that compete for pre-cursors, increasing pre-cursor supply through overexpression, increasing co-factor supply and removing negative regulation such as feedback inhibition. In recent years, various modelling based approaches have increasingly been applied as a method to identify targets for metabolic engineering [79][80]. These computational methods can help identify targets that are less apparent for rational engineering. The design phase also includes strategies for expression, such as inducible or constitutive promoters of different strength, tuning of the ribosomal binding site (RBS), expression from a plasmid or genome integrated etc.

Build: The build phase is where the strain modifications from the design phase such as deletions, insertion of new genes or additional copies of native genes, as well as up- and downregulations are generated. Vectors for plasmid based expression or genome integration and other expression elements can be obtained through repositories such as AddGene. As the cost of synthesizing DNA has decreased, it is now possible to order entire genes or even pathways [81]. Using synthetized DNA also allows the opportunity for codon optimization which can sometimes improve heterologous expression [82]. Due to efficient cloning and transformation methods, it is possible to construct combinatorial libraries, such as using different promoters and RBS of different strength and homologs of a gene from different organisms [83]. Library-based methods are limited in scale by the availability of high-throughput methods to quantify strain performance [84]. Another strategy to building strains is using evolution-based methods such as adaptive laboratory evolution (ALE), which can be used to improve cellular tolerance to product, intermediate or feedstock toxicity [85]. ALE is based on serial passaging of cells into fresh growth media, while for example gradually increasing inhibitor concentration; mutations that allow the cells to grow faster will naturally become dominant in the population.

Test: In the test phase, the strains built in the build phase are tested for their performance, such as growth and product formation in terms of TRY (titer, rate, and yield). For the first iterations of the cycle, the testing is done on smaller scale, such as deep well plates, shake flasks or small reactors. The product is then quantified by analytics methods like High Performance Liquid Chromatography (HPLC) or Liquid/Gas Chromatography-Mass Spectrometry (LC-MS or GC-MS). Different omics data, such as transcriptomics, metabolomics or proteomics, can be collected to obtain more systems level information on how the modifications are affecting the organism [86]. Alternatively, if the production can be coupled to a fluorescent output with a biosensor or to antibiotic resistance, it is possible to test a large number of strains by screening with fluorescence activated cell sorting (FACS) or by selection based on antibiotic resistance [87][88].

Learn: Learn is where the data from the test phase is analyzed for potential further improvements going into the next iteration of the cycle. Unfortunately, while this phase provides important input for the next design phase, it is also the least standardized. There are some frameworks on how the data from the test phase can be integrated. Omics data can be integrated into GEM's to improve their predictions

[89]. Statistical analyses like principal or independent component analysis (PCA or ICA) have been applied on omics data to help guide further engineering, although other analysis methods could be more suitable for biological systems [90]–[93]. With strain libraries, the focus is often on the top performing variants, which are chosen for further rounds of engineering. Machine learning could be applied to use the data from the entire library and guide further engineering [94].

1.4 Dynamic control of cellular metabolism

As discussed earlier, metabolic engineering uses genetic modifications to enhance the production of a chosen compound. However, modifications such as deletions or insertions with constitutive promoters act in a "static" manner, whereas the conditions both inside and outside the cell can vary drastically during a fermentation. It can therefore be beneficial to employ methods for dynamic control of the cellular metabolism during a fermentation. This form of dynamic metabolic engineering can be applied to control a pathway based on substrate or intermediate concentrations in order to balance the pathway and avoid potential accumulation of toxic intermediates. For example, a study by Dahl and co. alleviated the toxicity from isoprenoid pre-cursor farnesyl pyrophosphate (FPP) by identifying FPP responsive promoters through transcriptomics, and placing the expression of the FPP consuming enzyme under said promoter [95]. This lead to increased production of amorphadiene compared to using inducible promoters for the entire pathway. Another method of dynamic metabolic engineering is using metabolic switches. Genes that are essential or otherwise significantly reduce growth cannot be permanently deleted or downregulated. This could be the case if the production pathway uses pre-cursors from central metabolism that are required for growth. It can therefore be beneficial to alter the expression of these genes during a fermentation.

1.4.1 Signals and inducible promoters

In order to dynamically control metabolism, promoters are needed that respond to a specific signal that can be given at a certain time or under certain conditions. Figure 4 illustrates the different kind of signals that can be utilized to control gene expression. The most commonly used inducible promoters respond to signals conveyed by small molecules either supplied externally to the media or produced by the cell. These small molecules can be synthetic chemicals, or native metabolites. The transcriptional control is often mediated by a protein that alters its DNA binding

capability based on binding to said small molecule. One benefit of using added inducers is the ability to control the expression strength by changing the inducer concentration [96]–[98].

Examples of chemically induced promoters are the lactose promoter P_{Lac}/LacI and the tetracycline promoter P_{Tet}/tetR. LacI is a repressor that binds P_{Lac}, where the repression can be lifted by adding a chemical analogue isopropyl-beta-D-thiogalactoside (IPTG) or lactose [99]. Likewise, TetR is a repressor that binds to the P_{Tet} promoter, where the repression can be lifted by adding tetracycline or an analogue such as anhydrotetracycline [100]. The Tet promoter is more tightly regulated in *E. coli*, meaning the expression level is lower when inducers are absent; this makes it more suitable for expression of genes for toxic proteins or toxic intermediates in a pathway that can inhibit growth [100][101]. The tightness of the regulation can depend on the organism in question. One potential consideration, especially for bulk/commodity products, is the cost of using an inducer in a production process. IPTG for example is considered to be an expensive inducer [102]. Lactose could be used as a cheaper inducer, or other inducible promoters with cheaper inducers, such as xylose, rhamnose and arabinose could be used [98][103][104]. With these inducers it is important to consider that many host organisms can consume them as a carbon source (depending on catabolite repression and the presence of glucose in the media), and that feedstocks including hemicellulosic biomass contain these sugars [105][106]. P_{tet} was the promoter chosen for chapters 3 and 4, due to its tight regulation in *E. coli*, compatibility with both *E. coli* and *B. subtilis*, and the lack of interference by sugars found in biomass hydrolysates.

Alternatively, promoters can also be induced by the absence of a small molecule rather than the presence of one, typically based on nutrient depletion. In this case, the media has to be optimized in regards to the starting concentration of the nutrient that serves as the depletion signal in order to control the time of induction. Some examples of nutrient depletion based promoters include the tryptophan promoter P_{trp}/trpR and the alkaline phosphatase promoter P_{phoA}/phoB. P_{trp} has two repression mechanisms: TrpR binds as a repressor to P_{trp} under the presence of tryptophan, and the leader peptide trpL contains a stem loop that terminates translation unless the ribosome stalls from low tryptophan concentration [107], [108]. P_{phoA} is repressed by phoB, which binds to the promoter under the presence

of free phosphate [109][110]. Potential advantages of depletion-based induction are cost savings from not needing an inducer, and not risking contamination by opening the reactor during fermentation.

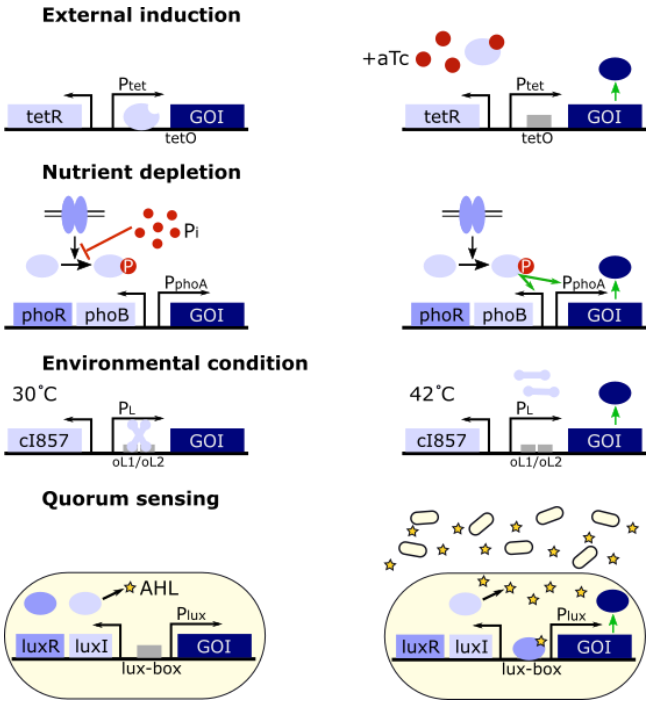


Figure 4 *Inducible promoters and how different types of signals can be used to induce expression*

Inducible promoters can also be induced by conditions of the external environment in the bioreactor, such as temperature, pH, or even light. The P_L/P_R promoters from the lambda phage can be induced by increasing temperature to above 37 °C, which causes a thermolabile mutant the repressor cI857 to become unstable [111]. The P_{cadBA} promoter is activated by a membrane bound regulator cadC, which activates expression when pH decreases below pH 5.5 and lysine is present in the media [112][113]. The P_{FixK2} promoter gets activated by FixJ, when FixJ gets phosphorylated by light sensitive kinase YF1; the promoter can control the gene(s) of interest (GOI) directly for light induction, or it can control a repressor for GOI for light repression [114]. Although these promoters are also inducer-free, the heterogeneous conditions of large reactors could pose a challenge for temperature and pH based

induction [115][116]. For light based induction, evenly illuminating a large reactor could prove to be challenging as well.

Quorum sensing systems are a mechanism whereby a population of bacterial cells can communicate, but have also been applied as autoinducible promoters for dynamic metabolic engineering. One of the most characterized quorum sensing systems is the *lux* mechanism, originally discovered in *Vibrio fischeri*, in which LuxI synthesizes acyl-homoserine lactone (AHL) and LuxR acts as an AHL concentration dependent regulator of luminescence; because AHL can diffuse through membranes, the concentration correlates with cell density, leading to cell density based activation of luminescence [117]. The *lux* mechanism has been used in *E. coli* as quorum sensing based induction of metabolic engineering [118][119].

1.4.2 Mechanisms for controlling expression

Besides expressing the GOI directly under an inducible promoter, other mechanisms can also be used to control the expression. This control can act at the transcriptional level, post-transcriptional level or post-translational level. Post-transcriptional control is based on RNA interference or RNAi, where an antisense RNA or asRNA is expressed, which is complementary to and binds the mRNA transcript of the GOI and results in its degradation by RNAses [120][121]. Post-translational control is often exerted through addition of a degradation tag onto the GOI in conjunction with a protease cleavage site, where the induction of said protease either removes the tag, leading to a stable protein, or exposes the tag, leading to the protein being targeted for degradation [122], [123]. In order to reduce waste of cellular resources, it can be beneficial to control expression at the earliest stage, which is transcription.

1.4.3 Gene silencing with CRISPR interference (CRISPRi)

Clustered Regularly Interspaced Palindromic Repeat or CRISPR sequences were originally discovered in 1987 as a previously unseen repetitive region in the genome of *E. coli*; subsequent observations of similar sequences in many other bacterial and archaeal organisms indicated of its evolutionary importance [124]. It was later discovered that CRISPR spacer sequences were also found in phages; subsequent experiments confirmed it acted as a defense mechanism against phages together with CRISPR-associated genes or Cas-genes [125]. The breakthrough towards CRISPR-Cas9 gene editing came when it was discovered that Cas9 could be guided using two short RNA sequences CRISPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA), and that this could be used to generate double-stranded breaks

(DSB) for gene editing [126][127]. Both RNA's were since combined in a single chimeric guide RNA, coined as sgRNA [127]. CRISPR-Cas9 was a huge boon for gene editing of previously hard to engineer organisms, and has also been used in combination with previous methods, such as lambda-Red recombineering and MAGE, to improve their efficiency [47], [128]–[130].

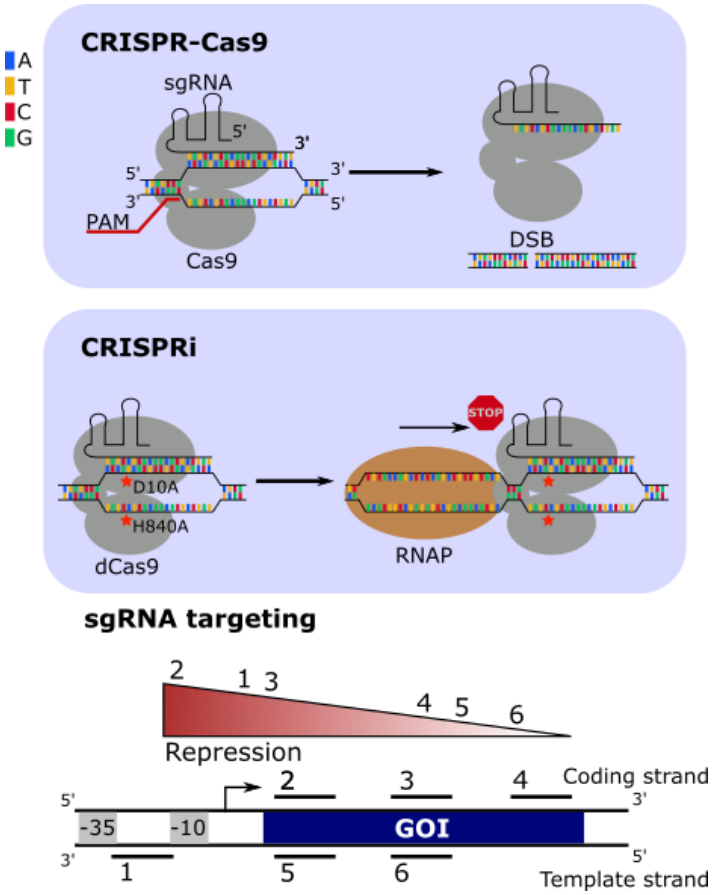


Figure 5 *Cas9 can be targeted with a sgRNA to make a DSB 4 basepairs after the PAM sequence (NGG). dCas9 can bind the same way, but is inactive due to the mutations D10A and H840A. Instead it blocks RNA-polymerase elongation through collision. sgRNA's targeting the coding strand at the start of the GOI is the most effective, while targeting the template strand has little or no effect* [131].

The programmability of CRISPR-Cas9 was unique and naturally made it interesting to be able to leverage this property for other purposes. In the original article from Jinek and co, it was discovered that the HNH domain was responsible for cleaving

the template strand, and the RuvC-like domain was responsible for cleaving the coding strand [126]. By introducing two point mutations D10A and H840A, one to each domain, it was possible to create a catalytically inactive version of Cas9 called dead Cas9 or dCas9 for short, which can still be guided by a sgRNA to bind a specific sequence, but then remain bound instead of creating a DSB. The only requirement for a target sequence is being adjacent to a short (often three-base pair) protospacer motif (PAM), which for the *Streptococcus pyogenes* derived Cas9/dCas9 is NGG [126]. The catalytically inactive dCas9 was shown to be an effective tool for gene silencing, called CRISPR interference or CRISPRi for short [131]. Figure 5 illustrates the principals and proposed mechanisms for CRISPRi.

The main mechanisms of gene silencing are the inhibition of transcription initiation or the inhibition of transcription elongation. The efficiency of the repression is highly affected by the positioning of the sgRNA, where the most effective positions are targeting the coding strand right after the start codon for elongation inhibition, or between the -35 and -10 elements in the promoter for initiation inhibition [131]. The efficiency of repression decreases the further away the sgRNA target sequence is from the start of the gene. Interestingly, whereas Cas9 can be targeted to either strand due to it generating a break on both strands, CRISPRi with dCas9 shows a stringent requirement towards targeting the coding strand, while targeting the template strand shows little or no repression of transcription. Other mechanisms such as DNA accessibility have also been proposed to play a role [132]. Because CRISPRi is programmable and can be controlled using an inducible promoter, it has enabled the characterization of how knocking down essential genes affects cellular metabolism and physiology, some of which could potentially be beneficial for the production of chemicals and proteins [133].

1.5 Growth decoupling and two-stage fermentations

One of the fundamental challenges of metabolic engineering is the competition for resources between growth and biomass formation and production of a desired compound. On one hand, conventional principles favor higher biomass formation, as higher biomass concentrations lead to higher volumetric productivities, assuming the specific productivity remains the same. On the other hand, the feedstock substrate is the most costly component of operating a bioprocess, meaning high yields of product from substrate is important. Converting excessive amounts of substrate into biomass negatively affects this yield. It is therefore desirable to be

able to control when the cells prioritize biomass formation and when they prioritize product formation: this concept is known as growth decoupling, as it leads to production no longer being linked to growth. Growth decoupled fermentations are operated in two stages: the first stage prioritizes fast accumulation of biomass, and the second stage downregulates biomass formation to re-channel flux towards product formation [134]. This general principal is illustrated in Figure 6. There are three main types of methods to limit biomass formation; nutrient limitation, addition of toxins or expression of toxin-antitoxin proteins, and applying synthetic biology tools as metabolic switches. These methods are compared in more detail in Chapter 2, so here is a brief description on the principals behind these methods.

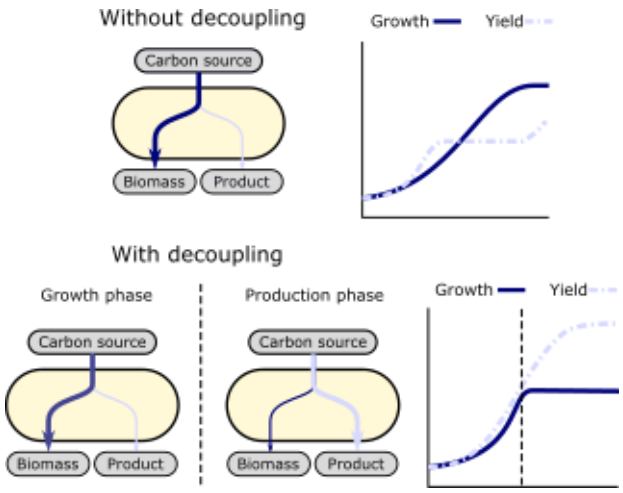


Figure 6 *Growth decoupling as a two-stage fermentation. Without growth decoupling, the cells are allowed to grow to the maximum cell density supported by the fermentation conditions, where significant amounts of the carbon source is used for growth and biomass. With growth decoupling, carbon is redirected to product by limiting growth, leading to an increased yield of product from substrate in the production phase*

One important factor when attempting to inhibit biomass formation is to avoid native cellular stress responses to the given conditions. A broadly conserved cellular response to nutrient starvation is the stringent response. The response is controlled by the signaling molecules guanosine 5',3' bispyrophosphate and guanosine pentaphosphate, abbreviated as (p)ppGpp [135]. In *E. coli,* (p)ppGpp is mainly synthesized by RelA, which detects uncharged tRNA in the A position of ribosomes, and also by SpoT, which responds to iron, carbon and fatty acid starvation [135]. (p)ppGpp leads to a strong downregulation of rRNA transcription, which results in

an overall significant decrease in cellular protein synthesis levels [136]. Therefore, stringent response should be avoided during the fermentation process. As mentioned above, *B. subtilis* can form endospores under nutrient starvation. Sporulation is potentially linked to stringent response, as stringent response leads to increased transcription of kinases *kinA* and *kinB* [137].

With nutrient limitation, the fermentation is first operated in a batch phase with nutrient surplus in order to facilitate fast growth. The fermentation is then switched to the fed-batch phase, where a nutrient limited feed media is added to the reactor [138], [139]. The limited nutrient can be a macronutrient, such as the carbon, nitrogen, phosphorus or sulfur source, or a micronutrient such as vitamins or trace elements. Which nutrient limitation is applied depends on the product and the host organism, as the goal of nutrient limitation is reducing metabolic flux towards components needed for biomass, without affecting metabolic activity such as glucose uptake or pre-cursor supply, or leading to starvation responses in the host. For example, limiting tryptophan or leucine in *E. coli* leads to the same decrease in glucose uptake rate as limiting nitrogen, but *B. subtilis* maintains a much higher glucose uptake rate with the specific amino acid limitation compared to overall nitrogen limitation [140].

Another method of inhibiting biomass formation is by adding chemical inhibitors or expressing toxin proteins. The goal is to inhibit cellular processes involved in growth, without causing cell death or significantly affecting metabolic activity. Studies have explored using antibiotics such as kanamycin, other chemicals such as indole, or toxic proteins such as HipA to achieve non-growing but metabolically active cells [141]–[143].

Programmable technologies to repress specific genes such as CRISPRi combined with inducible promoters have opened opportunities for applications towards metabolic switches in two stage fermentations. For example, a study by Kim and co used CRISPRi to repress the four genes *adhE, ldhA, pta* and *frdA* in *E. coli* to avoid byproduct formation and increase NADH supply under micro-aerobic fermentation for n-butanol production, where they obtained a several fold improvement to yield and productivity [144]. This shares some similarity to Chapter 3, where the goal was also to use a CRISPRi metabolic switch increase NADH availability and reduce byproduct formation that occurs from anaerobic fermentations, while also acting as

a growth decoupling mechanism. CRISPRi has potential to act as a more product independent method of growth decoupling in *E. coli* by targeting pyrimidine biosynthesis or DNA replication genes (*pyrF, thyA, dnaA* and *oriC)*, which can improve both chemical and protein production [145][146]. These targets were chosen based on rational design, in terms of being able to inhibit growth without being known triggers of stringent response. Chapter 4 looks into the application of these CRISPRi targets and optimizing dCas9 expression levels in *B. subtilis*. Figure 7 compares the de novo pyrimidine biosynthesis pathways of *E. coli* and *B. subtilis*, which is largely similar between both organisms.

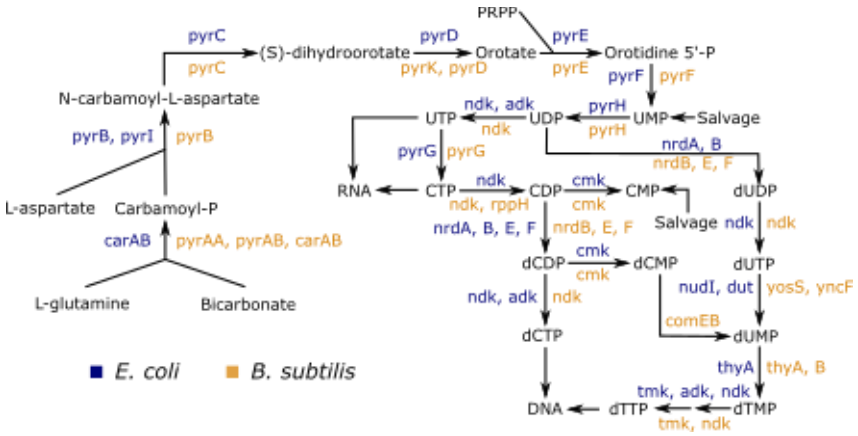


Figure 7 *De novo pyrimidine nucleotide biosynthesis pathway for both E. coli (genes in blue text) and B. subtilis (genes in orange text). "Salvage" indicates the entry point for the salvage pathway. Abbreviations: *MP = monophosphate, *DP = diphosphate, *TP = triphosphate, d*** = deoxy nucleotide, U** = uridine, C** = cytidine, T** = thymidine, PRPP = 5-phospho-α-D-ribose 1-diphosphate. Based on the pathways from KEGG database and BioCyc database* [60], [76].

1.6 Aerobic vs anaerobic metabolism for chemical production

Different host organisms have different requirements for oxygen presence. *B. subtilis* is for the most part an obligate aerobe as it lacks the metabolic reactions for fermentative growth, while organisms such as *C. acetobutylicum* are obligate anaerobes that are sensitive to oxygen [147]. *E. coli* on the other hand is a facultative anaerobe that can perform both respiratory and fermentative growth. What are the main differences in metabolism resulting from aerobic vs anaerobic conditions, and how do they relate to production of chemicals? Under aerobic conditions, the reducing equivalents NAD(P)H and $FADH_2$ are used by the electron transport chain

(ETC) to generate ATP, yielding a maximum of 18.7 ATP per glucose based on stoichiometry [148]. For anaerobic conditions, only substrate level phosphorylation can happen, leading to a net 2 ATP produced per glucose (accounting for one PEP consumed per glucose imported via the PTS system), with the ability of producing one extra ATP via acetate [149]. Furthermore, the NAD(P)H/NAD(P)$^+$ ratio is increased due to NAD(P)H not being consumed by ETC [150].

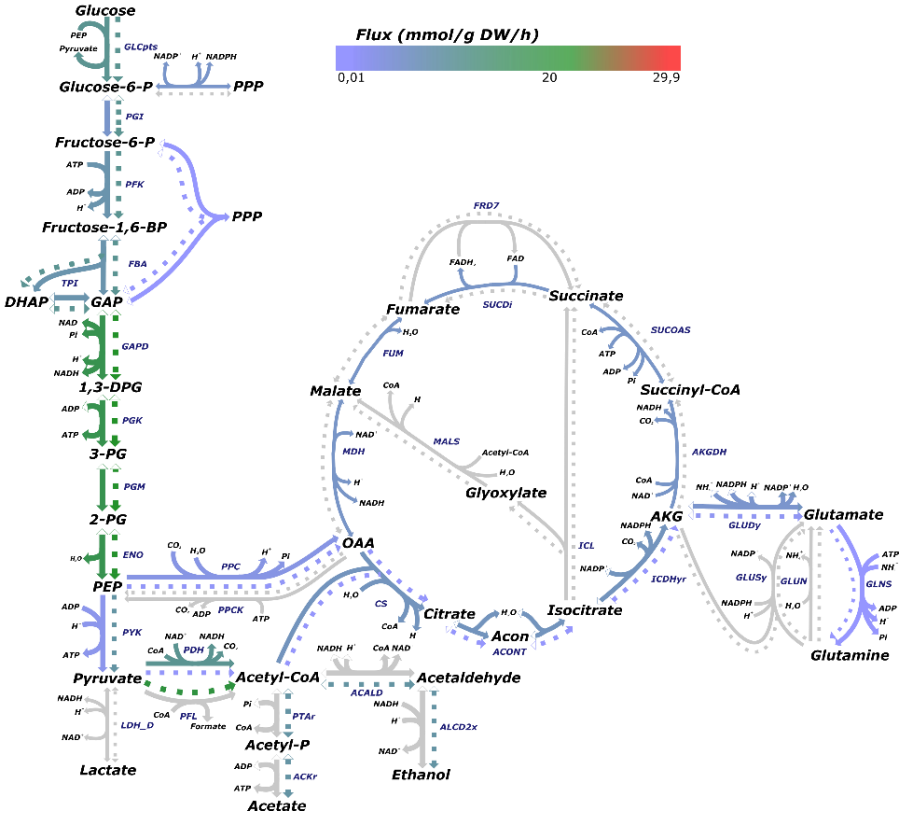


Figure 8 *Flux balance analysis (FBA) of E. coli GEM iJO1366 under aerobic conditions (solid lines) and under anaerobic conditions with oxygen exchange flux set to 0 (dashed lines), with the objective set to the biomass reaction. Modified from pathway maps were generated using Escher-FBA* [154]. *Abbreviations: PPP = pentose phosphate pathway, DHAP = dihydroxyacetone-phosphate, GAP = glyceraldehyde-3-phosphate, 1,3-DPG = 1,3-diphosphoglycerate, 3-PG = 3-phosphoglycerate, 2-PG = 2-phosphoglycerate, PEP = phosphoenolpyruvate, OAA = oxaloacetate, Acon = aconitate, AKG = α-ketogluterate*

Figure 8 shows predicted changes in central carbon metabolism between aerobic and anaerobic conditions using flux balance analysis (FBA). There is increased flux through glycolysis and preference for using pyruvate formate lyase (PFL) instead of pyruvate dehydrogenase (PDH) to have less NADH to regenerate. Only the lower part of TCA-cycle is active to produce the amino acids glutamate and glutamine. Lastly, byproducts are formed; acetate for additional ATP, and ethanol to regenerate NADH to NAD^+. The predicted growth rate decreases from 0.874 h^{-1} to 0.211 h^{-1} (not shown on figure). The increased NAD(P)H/$NAD(P)^+$ ratio can be leveraged to produce cofactor-limited products, such as succinate through the anaplerotic route [151]. Byproduct accumulation would still occur from acetate formation to generate ATP and potentially ethanol/lactate formation if the flux to the product is not enough to maintain redox balance, which can lead to byproduct toxicity [152][153]. Chapter 3 shows a novel approach to leveraging the higher NAD(P)H/$NAD(P)^+$ ratio from anaerobic metabolism while also dealing with byproduct accumulation.

1.7 Microbial consortia-based production

Traditionally, almost all industrial biotechnology processes are based on pure cultures, cultivating a singular strain. However, many microbes come from environments where they exist in communities with numerous other microbial organisms. Here they both compete for resources but also benefit from metabolites secreted by other organisms in syntrophic interactions, e.g. fermentative bacteria and methanogenic archaea working together to transform organic compounds to methane [155]. The bacteria used in this thesis are no exception; *E. coli* comes from the gut, which has a complex microbiome, and *B. subtilis* comes from the soil environment of plant roots, where there are also many other species. Therefore it is interesting to explore mimicking these natural mechanisms for optimizing metabolism by engineering artificial syntrophic consortia, consisting of multiple strains or even multiple different organisms, which could be applied towards industrial biotechnology.

With conventional monocultures, the strain can be heavily engineered, expressing long biosynthetic pathways and having modifications of the native metabolism to achieve the highest possible production. This however places a metabolic burden upon the cells, which can negatively affect both growth and production [163]. Consortia can be used to relieve this burden, by dividing the pathway between different strains or organisms where one uptakes the substrate and secretes an

intermediate, and another uptakes the intermediate and finishes the conversion to the product [157]. Consortia with different organisms can be engineered to take advantage of what each organism is naturally suited for, or different strains can be engineered to be specialized towards certain aspects of metabolism. One example of an application that could benefit from consortia based processes is the utilization of lignocellulosic biomass. Hydrolysates from lignocellulosic biomass contain many different sugars and toxic compounds; consortia can be engineered to have strains that can detoxify the hydrolysate or to be able to utilize the different sugars simultaneously [164][165]. A consortium was engineered for direct use of cellulosic biomass using industrial enzyme producing fungus *Trichoderma reesei* to produce cellulases and *E. coli* to convert the released glucose to other products [166]. Table X lists some other examples of improved chemical production using microbial consortia. In Chapter 3, a consortium of two *E. coli* strains were engineered, where one *E. coli* strain is engineered to leverage the increased NAD(P)H/NAD(P)$^+$ ratio of anaerobic metabolism, and another *E. coli* strain is engineered to leverage aerobic metabolism in order to consume byproducts from the first strain.

Table 1 *Examples of improved chemical production using microbial consortia. Adapted from* [156] and [157].

Type	Organisms	Product	Improvement	Reference
Single organism	*E.coli – E. coli*	3-hydroxybenzoate	5.3-fold	[158]
Single organism	*E.coli – E. coli*	3-amino-benzoate	15-fold	[159]
Single organism	*E.coli – E. coli*	n-butanol	2-fold	[160]
Bacteria - Eukaryote	*E. coli - S.cerevisiae*	Oxygenated taxenes	33 mg/L vs none in monoculture	[161]
Bacteria - Eukaryote	*E. coli - S.cerevisiae*	Naringenin	~8-fold	[162]

There are some challenges associated with engineering microbial consortia. If a consortium uses different organisms, they have to be able to grow under the same conditions, such as media composition, temperature, pH and oxygenation. Furthermore, fungi can produce antibacterial compounds, which would limit co-cultivation with bacteria [167]. It can therefore be beneficial to use different strains

of the same organism instead. With the division of biosynthesis pathways, the intermediate at the division point must be stable and efficiently transported in and out of the cell, setting limitations for how pathways can be divided. Another important aspect of cultivating consortia is being able to control the population of each strain. If the strains are neutral in terms of interaction and utilize the same substrate, the strain with the faster growth rate will always end up taking over as the fermentation progresses. Strains can be engineered to use different substrates to assist in compensating for differences in growth rates by adjusting each substrates concentration [157]. Alternatively, metabolite dependencies and cross-feeding can be used as a mechanism for population control; a study by Losoi and co engineered a consortium where *Acinetobacter baylyi* converts glucose to gluconate which the co-cultivated *E. coli* needs to grow, and the *E. coli* then produces acetate which *A. baylyi* needs to grow [168]. Synthetic biology tools have also been explored for population control in a mixed culture, such as expressing CRISPRi under the P_{lux} quorum-sensing promoter that targets the DNA-replication of the faster growing strain once a certain cell density is reached [169]. The consortium in Chapter 3 is population controlled by a combination of synthetic biology tools and feeding, as one strain has a CRISPRi metabolic switch that inhibits its growth, while the other strain relies on feeding of acetate from the first strain.

1.8 Heterologous protein expression in industrial biotechnology

Heterologous protein expression is very important in industrial biotechnology. The microbial production strains need to be engineered to produce at or near theoretical maxima to achieve economic viability, especially for low value products. This strain engineering process often involves taking proteins from other organisms and expressing them heterologously in a platform organism, whether it is a part of or an entire biosynthetic pathway for chemical production, or if the protein itself is the product, such as industrial enzymes or pharmaceutical proteins [52]. Furthermore, the native metabolism can be engineered by introducing higher activity or not feedback regulated variants of native enzymes.

However, these heterologous proteins do not always reach their functional state, as the cellular mechanisms assisting the protein in correct folding and expression change when the protein is expressed in a new organism. This can result in poor expression levels or protein aggregation into inclusion bodies, which is a known challenge with some bacterial platform organisms such as *E. coli* [170].

Furthermore, it can be difficult to evaluate how many proteins cannot be expressed, as that would not get reported in literature. Chapter 5 discusses this topic by providing an overview of the numerous mechanisms that affect heterologous protein expression, what experimental strategies are being applied to improve heterologous protein expression, and how the theory and experimental strategies relate. Here is a brief overview of some of the most important mechanisms.

Codons that code for amino acids are degenerate, meaning that there are 64 codons that code for only 20 amino acids. This means most amino acids are coded for by multiple codons. Organisms do not have the same level of tRNA's for all the codons of the same amino acid, rather they prefer some codons over others, which reflects in tRNA level and codon usage in their genes [171]. This phenomenon is known as codon bias, and affects heterologous protein expression, as the new host might have a different codon bias compared to the native organism of the protein. Codon bias affects protein expression through changing the speed of translation: rare codons cause the ribosome to slow down or pause [171]. This pausing influences mechanisms such as co-translational protein folding and protein complex assembly, which can be important for functional expression of the protein [172][173].

Proteins need to be folded correctly in order to function. Protein folding is a very complex process, and the inside of a cell can be a crowded folding environment. Some proteins therefore need assistance to both reach and maintain their correct structure; this function is mediated by proteins called chaperones. Chaperones are involved in both the synthesis of new proteins and maintenance of existing proteins, and core chaperones are often important or even essential for cell growth, even in non-stressed conditions [174][175]. With heterologous expression, proteins are now interacting with different chaperones of the host organism. Although the core chaperone families such as Hsp (heat shock protein) 40, 70 and 90 are present throughout the tree of life, some organisms have evolved to have different chaperones and increased interaction between them [176]. These differences between the native organism and the new host organism can affect whether the heterologous protein gets recognized by the chaperones that it needs to fold correctly. If the protein does not fold correctly, it can get recognized by the protein quality control (PQC) mechanisms. Maintaining proper protein folding and homeostasis is important for cell health and survival [177]. Misfolded proteins are recognized by various proteases and get degraded. These PQC mechanisms may also

degrade heterologous proteins, and deletion of non-essential proteases is a common method used to generate strains for protein expression, such as *E. coli* BL21 or *B. subtilis* KO7 [178][179].

Post-translational modifications (PTM) can affect protein activity and stability. PTMs can differ substantially from organism to organism, especially going from eukaryotes to prokaryotes. Two of the most important PTMs are disulfide bond formation and N-glycosylation. Disulfide bonds form between cysteine residues and are important for protein stability; mutating cysteines that form disulfide bonds almost always results in instability or even loss of function [180][181]. Most gram-positive bacteria lack the ability to form disulfide bonds, while gram-negative bacteria can only form disulfide bonds in the periplasm [182][183]. N-glycosylation can affect protein folding and stability, and both *E. coli* and *B. subtilis* lack the ability to N-glycosylate proteins [184]–[186]. PTMs therefore affect which proteins can be heterologously expressed.

Conclusions and future perspectives

As the need for humanity to transition away from using fossil fuels is becoming more and more evident, bio-based manufacturing using microbial cell factories has emerged as a technologically feasible and more sustainable alternative way of producing fuels, chemicals and materials using renewable feedstocks and waste streams. However, there are still many challenges that need to be addressed before bio-based manufacturing can become the predominant route of production for the variety of compounds that are needed in a modern society.

In this thesis, the goal was to study different tools and methods that could potentially help address some of these challenges. The main focus of the thesis was using CRISPRi to dynamically control metabolism in order to shift from biomass to product formation and thus improving product yield, which can potentially help reduce the production costs. In chapters 2-4, recent progress on using synthetic biology tools to achieve two-stage fermentations with growth decoupling was reviewed and compared with other methods. CRISPRi was applied in *E.coli*, in order to enable a unique consortium with potential to increase efficiency of resource and reactor capacity usage. Furthermore, the CRISPRi toolbox in *B. subtilis* was expanded by simplifying gRNA cloning, and by tuning down dCas9 protein levels to improve compatibility with different promoters. Lastly, a thorough literary review was made on a more scientifically fundamental challenge in cell factory engineering; namely folding and expression of heterologous proteins.

Further work is needed for these technologies to reach the level of maturity needed for industrial applications. It would be interesting to test the CRISPRi systems deployed in this thesis using different products to better understand how widely applicable they are. Furthermore, it could also be interesting to explore setups that are more closely relatable to industrial ones, such as fed-batch fermentations in lab or pre-pilot scale bioreactors, to see how the results obtained in this thesis perform at larger scale. The data obtained from a more industrially relevant setup could also be used to perform modelling, such as a techno-economical assessment (TEA) for more detailed and quantified measurement on the potential economic benefits. One of the main aspects of CRISPRi metabolic switches that needs to be optimized, especially when used to achieve growth decoupling, is stability for longer durations of time. The stability could be influenced by the tightness of regulation resulting in

instability over time (mutations) and/or changes in the physiology where genes that to some extent can compensate for the “loss” of function are upregulated once CRISPRi is induced. For the future, developing tightly regulated, possibly also inducer-free, promoter systems for as many platform organisms as possible, could be an additional step towards industrial application.

It is clear that establishing large scale bio-based manufacturing is a difficult feat, facing a variety of different challenges. For the field of metabolic and cell factory engineering, this entails continuing to engineer the most efficient and highest producing strains possible, while simultaneously lowering the cost and time needed to achieve these strains. There has been efforts to build biofoundries that can fully automate the DBTL cycle, thus reducing the time and labor costs of building cell factories [187]. Dynamic metabolic engineering and growth decoupling are likely to play a role, as they address a very fundamental concept of competition for resources between production and growth. Lastly, increased collaboration and interdisciplinarity between different areas of industrial biotechnology can help guide research in a more applicable direction.

1.10 References for Chapter 1

[1] M. A. Farag, M. M. Elmassry, M. Baba, and R. Friedman, "Revealing the constituents of Egypt's oldest beer using infrared and mass spectrometry," *Scientific Reports*, vol. 9, no. 1, p. 16199, Dec. 2019, doi: 10.1038/s41598-019-52877-0.

[2] M. C. Hunter, "Experiment, Theory, Representation: Robert Hooke's Material Models," 2010, pp. 193–219. doi: 10.1007/978-90-481-3851-7_9.

[3] P. Smit and J. Heniger, "Antoni van Leeuwenhoek (1632–1723) and the discovery of bacteria," *Antonie Van Leeuwenhoek*, vol. 41, no. 1, pp. 217–228, Dec. 1975, doi: 10.1007/BF02565057.

[4] J. Steere-Williams, "The Germ Theory," in *A Companion to the History of American Science*, Chichester, UK: John Wiley & Sons, Ltd, 2015, pp. 397–407. doi: 10.1002/9781119072218.ch31.

[5] D. D. Songstad, P. Lakshmanan, J. Chen, W. Gibbons, S. Hughes, and R. Nelson, "Historical perspective of biofuels: learning from the past to rediscover the future," *In Vitro Cellular & Developmental Biology - Plant*, vol. 45, no. 3, pp. 189–192, Jun. 2009, doi: 10.1007/s11627-009-9218-6.

[6] N.-P.-T. Nguyen, C. Raynaud, I. Meynial-Salles, and P. Soucaille, "Reviving the Weizmann process for commercial n-butanol production," *Nature Communications*, vol. 9, no. 1, p. 3682, Dec. 2018, doi: 10.1038/s41467-018-05661-z.

[7] R. Gaynes, "The Discovery of Penicillin—New Insights After More Than 75 Years of Clinical Use," *Emerging Infectious Diseases*, vol. 23, no. 5, pp. 849–853, May 2017, doi: 10.3201/eid2305.161556.

[8] T. v. Danylova and S. v. Komisarenko, "Standing on the shoulders of giants: James Watson, Francis Crick, Maurice Wilkins, Rosalind Franklin and the birth of molecular biology," *The Ukrainian Biochemical Journal*, vol. 92, no. 4, pp. 154–165, Sep. 2020, doi: 10.15407/ubj92.04.154.

[9] K. E. van Holde and J. Zlatanova, "Recombinant DNA: The Next Revolution," in *The Evolution of Molecular Biology*, Elsevier, 2018, pp. 149–163. doi: 10.1016/B978-0-12-812917-3.00014-0.

[10] J. M. Heather and B. Chain, "The sequence of sequencers: The history of sequencing DNA," *Genomics*, vol. 107, no. 1, pp. 1–8, Jan. 2016, doi: 10.1016/j.ygeno.2015.11.003.

[11] J. Sandow, W. Landgraf, R. Becker, and G. Seipke, "Equivalent Recombinant Human Insulin Preparations and their Place in Therapy," *European Endocrinology*, vol. 11, no. 1, p. 10, 2015, doi: 10.17925/EE.2015.11.01.10.

[12] UN General Assembly, "Transforming our world: the 2030 Agenda for Sustainable Development," 2015. [Online]. Available: https://sustainabledevelopment.un.org/post2015/transformingourworld

[13] E. C. D. Tan and P. Lamers, "Circular Bioeconomy Concepts—A Perspective," *Frontiers in Sustainability*, vol. 2, Jul. 2021, doi: 10.3389/frsus.2021.701509.

[14] B. Hauer, "Embracing Nature's Catalysts: A Viewpoint on the Future of Biocatalysis," *ACS Catalysis*, vol. 10, no. 15, pp. 8418–8427, Aug. 2020, doi: 10.1021/acscatal.0c01708.

[15] C. C. Ibeh, "Phenol-Formaldehyde Resins," in *Handbook of Thermoset Plastics*, Elsevier, 1998, pp. 23–71. doi: 10.1016/B978-081551421-3.50005-9.

[16] G. Busca, "Catalysts for Hydrogenations, Dehydrogenations and Metathesis," in *Heterogeneous Catalytic Materials*, Elsevier, 2014, pp. 345–374. doi: 10.1016/B978-0-444-59524-9.00010-9.

[17] K. M. Walter, M.-A. Serrer, W. Kleist, and J.-D. Grunwaldt, "Continuous production of higher alcohols from synthesis gas and ethanol using Cs-modified CuO/ZnO/Al2O3 catalysts," *Applied Catalysis A: General*, vol. 585, p. 117150, Sep. 2019, doi: 10.1016/j.apcata.2019.117150.

[18] M. J. Darabi Mahboub, J.-L. Dubois, F. Cavani, M. Rostamizadeh, and G. S. Patience, "Catalysis for the synthesis of methacrylic acid and methyl methacrylate," *Chemical Society Reviews*, vol. 47, no. 20, pp. 7703–7738, 2018, doi: 10.1039/C8CS00117K.

[19] M. Lopes, I. Belo, and M. Mota, "Over-pressurized bioreactors: Application to microbial cell cultures," *Biotechnology Progress*, vol. 30, no. 4, pp. 767–775, Jul. 2014, doi: 10.1002/btpr.1917.

[20] H. K. Jeswani, A. Chilvers, and A. Azapagic, "Environmental sustainability of biofuels: a review," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 476, no. 2243, p. 20200351, Nov. 2020, doi: 10.1098/rspa.2020.0351.

[21] J. L. Field *et al.*, "Robust paths to net greenhouse gas mitigation and negative emissions via advanced biofuels," *Proceedings of the National Academy of*

Sciences, vol. 117, no. 36, pp. 21968–21977, Sep. 2020, doi: 10.1073/pnas.1920877117.

[22] Ó. Ögmundarson, M. J. Herrgård, J. Forster, M. Z. Hauschild, and P. Fantke, "Addressing environmental sustainability of biochemicals," *Nature Sustainability*, vol. 3, no. 3, pp. 167–174, Mar. 2020, doi: 10.1038/s41893-019-0442-8.

[23] E. Budsberg, R. Morales-Vera, J. T. Crawford, R. Bura, and R. Gustafson, "Production routes to bio-acetic acid: life cycle assessment," *Biotechnology for Biofuels*, vol. 13, no. 1, p. 154, Dec. 2020, doi: 10.1186/s13068-020-01784-y.

[24] F. Adom, J. B. Dunn, J. Han, and N. Sather, "Life-Cycle Fossil Energy Consumption and Greenhouse Gas Emissions of Bioderived Chemicals and Their Conventional Counterparts," *Environmental Science & Technology*, vol. 48, no. 24, pp. 14624–14631, Dec. 2014, doi: 10.1021/es503766e.

[25] G. Festel, "Economic Aspects of Industrial Biotechnology," 2018, pp. 53–74. doi: 10.1007/10_2018_70.

[26] J. Spekreijse, T. Lammens, C. Parisi, T. Ronzon, and M. Vis, *Insights into the European market for bio-based chemicals*, vol. 19, no. October. 2019. doi: 10.2760/673071.

[27] G. Schütte, "What kind of innovation policy does the bioeconomy need?," *New Biotechnology*, vol. 40, pp. 82–86, 2018, doi: https://doi.org/10.1016/j.nbt.2017.04.003.

[28] OECD, "Meeting Policy Challenges for a Sustainable Bioeconomy," Paris, 2018.

[29] J. Philp, "The bioeconomy, the challenge of the century for policy makers," *New Biotechnology*, vol. 40, pp. 11–19, Jan. 2018, doi: 10.1016/J.NBT.2017.04.004.

[30] L. et al. Pelkmans, "D6.3 Policy options to mobilize sustainable non-food biomass resources for the biobased economy ," 2016.

[31] A. Muscat, E. M. de Olde, I. J. M. de Boer, and R. Ripoll-Bosch, "The battle for biomass: A systematic review of food-feed-fuel competition," *Global Food Security*, vol. 25, p. 100330, Jun. 2020, doi: 10.1016/j.gfs.2019.100330.

[32] J. Popp, S. Kovács, J. Oláh, Z. Divéki, and E. Balázs, "Bioeconomy: Biomass and biomass-based energy supply and demand," *New Biotechnology*, vol. 60, pp. 76–84, Jan. 2021, doi: 10.1016/j.nbt.2020.10.004.

[33] S. Redl, M. Diender, T. Ø. Jensen, D. Z. Sousa, and A. T. Nielsen, "Exploiting the potential of gas fermentation," *Industrial Crops and Products*, vol. 106, pp. 21–30, Nov. 2017, doi: 10.1016/j.indcrop.2016.11.015.

[34] L. Nattrass, C. Biggs, A. Bauen, C. Parisi, E. Rodríguez-Cerezo, and M. Gómez-Barbero, *The EU bio-based industry : Results from a survey*. 2016. doi: 10.2791/806858.

[35] J. Zetterholm, E. Bryngemark, J. Ahlström, P. Söderholm, S. Harvey, and E. Wetterlund, "Economic evaluation of large-scale biorefinery deployment: A framework integrating dynamic biomass market and techno-economic models," *Sustainability (Switzerland)*, vol. 12, no. 17, Sep. 2020, doi: 10.3390/su12177126.

[36] S. Banerjee, M. Munagala, Y. Shastri, R. Vijayaraghavan, A. F. Patti, and A. Arora, "Process Design and Techno-Economic Feasibility Analysis of an Integrated Pineapple Processing Waste Biorefinery," *ACS Engineering Au*, p. acsengineeringau.1c00028, Feb. 2022, doi: 10.1021/acsengineeringau.1c00028.

[37] X. Li and E. Mupondwa, "Empirical analysis of large-scale bio-succinic acid commercialization from a technoeconomic and innovation value chain perspective: BioAmber biorefinery case study in Canada," *Renewable and Sustainable Energy Reviews*, vol. 137, p. 110587, Mar. 2021, doi: 10.1016/j.rser.2020.110587.

[38] T. Gytkjær, "Commercial Development of Fermentation Processes," in *Fundamental Bioengineering*, J. Villadsen, Ed. Wiley-VCH, 2016, pp. 499–543.

[39] S. M. Ioannidou, K. Filippi, I. K. Kookos, A. Koutinas, and D. Ladakis, "Techno-economic evaluation and life cycle assessment of a biorefinery using winery waste streams for the production of succinic acid and value-added co-products," *Bioresource Technology*, vol. 348, p. 126295, Mar. 2022, doi: 10.1016/j.biortech.2021.126295.

[40] P. I. Nikel and D. Mattanovich, "Microbial cell factories: a biotechnology journey across species," *Essays in Biochemistry*, vol. 65, no. 2, pp. 143–145, Jul. 2021, doi: 10.1042/EBC20210037.

[41] J. Nielsen and J. D. Keasling, "Engineering Cellular Metabolism," *Cell*, vol. 164, no. 6, pp. 1185–1197, 2016, doi: 10.1016/j.cell.2016.02.004.

[42] R. Kananavičiūtė and D. Čitavičius, "Genetic engineering of Geobacillus spp.," *Journal of Microbiological Methods*, vol. 111, pp. 31–39, Apr. 2015, doi: 10.1016/j.mimet.2015.02.002.

[43] G.-Q. Chen and X.-R. Jiang, "Next generation industrial biotechnology based on extremophilic bacteria," *Current Opinion in Biotechnology*, vol. 50, pp. 94–100, Apr. 2018, doi: 10.1016/j.copbio.2017.11.016.

[44] M. Theisen and J. C. Liao, "Industrial Biotechnology: *Escherichia coli* as a Host," in *Industrial Biotechnology*, Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2016, pp. 149–181. doi: 10.1002/9783527807796.ch5.

[45] F. R. Blattner *et al.*, "The Complete Genome Sequence of *Escherichia coli* K-12," *Science (1979)*, vol. 277, no. 5331, pp. 1453–1462, Sep. 1997, doi: 10.1126/science.277.5331.1453.

[46] S. I. Jensen, R. M. Lennen, M. J. Herrgård, and A. T. Nielsen, "Seven gene deletions in seven days: Fast generation of Escherichia coli strains tolerant to acetate and osmotic stress," *Scientific Reports*, vol. 5, no. 1, p. 17874, Nov. 2016, doi: 10.1038/srep17874.

[47] C. Ronda, L. E. Pedersen, M. O. A. Sommer, and A. T. Nielsen, "CRMAGE: CRISPR Optimized MAGE Recombineering," *Scientific Reports*, vol. 6, no. 1, p. 19452, May 2016, doi: 10.1038/srep19452.

[48] X. Ao *et al.*, "A Multiplex Genome Editing Method for Escherichia coli Based on CRISPR-Cas12a," *Frontiers in Microbiology*, vol. 9, Oct. 2018, doi: 10.3389/fmicb.2018.02307.

[49] I. M. Keseler *et al.*, "The EcoCyc Database in 2021," *Frontiers in Microbiology*, vol. 12, Jul. 2021, doi: 10.3389/fmicb.2021.711077.

[50] J. M. Monk *et al.*, "iML1515, a knowledgebase that computes Escherichia coli traits," *Nature Biotechnology*, vol. 35, no. 10, pp. 904–908, Oct. 2017, doi: 10.1038/nbt.3956.

[51] M. T. Bonde *et al.*, "Predictable tuning of protein expression in bacteria," *Nature Methods*, vol. 13, no. 3, pp. 233–236, Mar. 2016, doi: 10.1038/nmeth.3727.

[52] G. Walsh, "Biopharmaceutical benchmarks 2018," *Nature Biotechnology*, vol. 36, no. 12, pp. 1136–1145, 2018, doi: 10.1038/nbt.4305.

[53] V. E. T. Maervoet, M. de Mey, J. Beauprez, S. de Maeseneire, and W. K. Soetaert, "Enhancing the Microbial Conversion of Glycerol to 1,3-Propanediol Using Metabolic Engineering," *Organic Process Research & Development*, vol. 15, no. 1, pp. 189–202, Jan. 2011, doi: 10.1021/op1001929.

[54] "Disruptive platform for sustainable and low-cost amino acid production." https://cordis.europa.eu/project/id/101009882 (accessed Apr. 04, 2022).

[55] P. Li, H. Lin, Z. Mi, S. Xing, Y. Tong, and J. Wang, "Screening of Polyvalent Phage-Resistant Escherichia coli Strains Based on Phage Receptor Analysis," *Frontiers in Microbiology*, vol. 10, Apr. 2019, doi: 10.3389/fmicb.2019.00850.

[56] H.-P. Hohmann, J. M. van Dijl, L. Krishnappa, and Z. Prágai, "Host Organisms: *Bacillus subtilis*," in *Industrial Biotechnology*, Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2016, pp. 221–297. doi: 10.1002/9783527807796.ch7.

[57] A. M. Earl, R. Losick, and R. Kolter, "Ecology and genomics of Bacillus subtilis," *Trends in Microbiology*, vol. 16, no. 6, pp. 269–275, Jun. 2008, doi: 10.1016/j.tim.2008.03.004.

[58] E. Härtig and D. Jahn, "Regulation of the Anaerobic Metabolism in Bacillus subtilis," 2012, pp. 195–216. doi: 10.1016/B978-0-12-394423-8.00005-6.

[59] S. Condon, M. Bayarte, and F. J. Sala, "Influence of the sporulation temperature upon the heat resistance of Bacillus subtilis," *Journal of Applied Bacteriology*, vol. 73, no. 3, pp. 251–256, Sep. 1992, doi: 10.1111/j.1365-2672.1992.tb02985.x.

[60] P. D. Karp *et al.*, "The BioCyc collection of microbial genomes and metabolic pathways," *Briefings in Bioinformatics*, vol. 20, no. 4, pp. 1085–1093, Jul. 2019, doi: 10.1093/bib/bbx085.

[61] Y. Su, C. Liu, H. Fang, and D. Zhang, "Bacillus subtilis: a universal cell factory for industry, agriculture, biomaterials and medicine," *Microbial Cell Factories*, vol. 19, no. 1, p. 173, Dec. 2020, doi: 10.1186/s12934-020-01436-8.

[62] N. Mirouze and D. Dubnau, "Chance and Necessity in *Bacillus subtilis* Development," *Microbiology Spectrum*, vol. 1, no. 1, Oct. 2013, doi: 10.1128/microbiolspectrum.TBS-0004-2012.

[63] L. Vojcic, D. Despotovic, R. Martinez, K.-H. Maurer, and U. Schwaneberg, “An efficient transformation method for Bacillus subtilis DB104,” *Applied Microbiology and Biotechnology*, vol. 94, no. 2, pp. 487–493, Apr. 2012, doi: 10.1007/s00253-012-3987-2.

[64] R. Rahmer, K. Morabbi Heravi, and J. Altenbuchner, “Construction of a Super-Competent Bacillus subtilis 168 Using the PmtlA-comKS Inducible Cassette,” *Frontiers in Microbiology*, vol. 6, Dec. 2015, doi: 10.3389/fmicb.2015.01431.

[65] K. B. Falkenberg *et al.*, “The ProUSER2.0 Toolbox: Genetic Parts and Highly Customizable Plasmids for Synthetic Biology in *Bacillus subtilis*,” *ACS Synthetic Biology*, vol. 10, no. 12, pp. 3278–3289, Dec. 2021, doi: 10.1021/acssynbio.1c00130.

[66] E. P. Riley, J. Lopez-Garrido, J. Sugie, R. B. Liu, and K. Pogliano, “Metabolic differentiation and intercellular nurturing underpin bacterial endospore formation,” *Science Advances*, vol. 7, no. 4, Jan. 2021, doi: 10.1126/sciadv.abd6385.

[67] P. Setlow, “Spores of Bacillus subtilis: their resistance to and killing by radiation, heat and chemicals,” *Journal of Applied Microbiology*, vol. 101, no. 3, pp. 514–525, Sep. 2006, doi: 10.1111/j.1365-2672.2005.02736.x.

[68] W. Overkamp and O. P. Kuipers, “Transcriptional Profile of Bacillus subtilis sigF-Mutant during Vegetative Growth,” *PLOS ONE*, vol. 10, no. 10, p. e0141553, Oct. 2015, doi: 10.1371/journal.pone.0141553.

[69] J. M. van Dijl and M. Hecker, “Bacillus subtilis: from soil bacterium to super-secreting cell factory,” *Microbial Cell Factories*, vol. 12, no. 1, p. 3, 2013, doi: 10.1186/1475-2859-12-3.

[70] E. R. Green and J. Mecsas, “Bacterial Secretion Systems: An Overview,” *Virulence Mechanisms of Bacterial Pathogens*, vol. 4, no. 1, pp. 213–239, 2016, doi: 10.1128/9781555819286.ch8.

[71] Q. Zhang *et al.*, “Production of proteins and commodity chemicals using engineered *Bacillus subtilis* platform strain,” *Essays in Biochemistry*, vol. 65, no. 2, pp. 173–185, Jul. 2021, doi: 10.1042/EBC20210011.

[72] E. J. GudiÃ±a, E. C. Fernandes, A. I. Rodrigues, J. A. Teixeira, and L. R. Rodrigues, “Biosurfactant production by Bacillus subtilis using corn steep liquor as culture medium,” *Frontiers in Microbiology*, vol. 6, Feb. 2015, doi: 10.3389/fmicb.2015.00059.

[73] W. Ma, Y. Liu, X. Lv, J. Li, G. Du, and L. Liu, "Combinatorial pathway enzyme engineering and host engineering overcomes pyruvate overflow and enhances overproduction of N-acetylglucosamine in Bacillus subtilis," *Microbial Cell Factories*, vol. 18, no. 1, p. 1, Dec. 2019, doi: 10.1186/s12934-018-1049-x.

[74] A. Kalantari *et al.*, "Conversion of Glycerol to 3-Hydroxypropanoic Acid by Genetically Engineered Bacillus subtilis," *Frontiers in Microbiology*, vol. 8, Apr. 2017, doi: 10.3389/fmicb.2017.00638.

[75] A. Chang *et al.*, "BRENDA, the ELIXIR core data resource in 2021: new developments and updates," *Nucleic Acids Research*, vol. 49, no. D1, pp. D498–D508, Jan. 2021, doi: 10.1093/nar/gkaa1025.

[76] M. Kanehisa, "KEGG: Kyoto Encyclopedia of Genes and Genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, Jan. 2000, doi: 10.1093/nar/28.1.27.

[77] H. Ren, C. Shi, and H. Zhao, "Computational Tools for Discovering and Engineering Natural Product Biosynthetic Pathways," *iScience*, vol. 23, no. 1, p. 100795, Jan. 2020, doi: 10.1016/j.isci.2019.100795.

[78] Y. Moriya *et al.*, "PathPred: an enzyme-catalyzed metabolic pathway prediction server," *Nucleic Acids Research*, vol. 38, no. Web Server, pp. W138–W143, Jul. 2010, doi: 10.1093/nar/gkq318.

[79] M. R. Long, W. K. Ong, and J. L. Reed, "Computational methods in metabolic engineering for strain design," *Current Opinion in Biotechnology*, vol. 34, pp. 135–141, Aug. 2015, doi: 10.1016/j.copbio.2014.12.019.

[80] E. J. Kerkhoven, P.-J. Lahtvee, and J. Nielsen, "Applications of computational modeling in metabolic engineering of yeast," *FEMS Yeast Research*, p. n/a-n/a, Sep. 2014, doi: 10.1111/1567-1364.12199.

[81] C. Payen and D. Thompson, "The renaissance of yeasts as microbial factories in the modern age of biomanufacturing," *Yeast*, vol. 36, no. 12, pp. 685–700, Dec. 2019, doi: 10.1002/yea.3439.

[82] H. G. Menzella, "Comparison of two codon optimization strategies to enhance recombinant protein production in Escherichia coli," *Microbial Cell Factories*, vol. 10, no. 1, p. 15, 2011, doi: 10.1186/1475-2859-10-15.

[83] G. Naseri and M. A. G. Koffas, "Application of combinatorial optimization strategies in synthetic biology," *Nature Communications*, vol. 11, no. 1, p. 2446, Dec. 2020, doi: 10.1038/s41467-020-16175-y.

[84] V. D'Ambrosio and M. K. Jensen, "Lighting up yeast cell factories by transcription factor-based biosensors," *FEMS Yeast Research*, vol. 17, no. 7, Nov. 2017, doi: 10.1093/femsyr/fox076.

[85] M. M. Matson *et al.*, "Adaptive laboratory evolution for improved tolerance of isobutyl acetate in Escherichia coli," *Metabolic Engineering*, vol. 69, pp. 50–58, Jan. 2022, doi: 10.1016/j.ymben.2021.11.002.

[86] Y.-M. Kim, C. J. Petzold, E. J. Kerkhoven, and S. E. Baker, "Editorial: Multi-Omics Technologies for Optimizing Synthetic Biomanufacturing," *Frontiers in Bioengineering and Biotechnology*, vol. 9, Dec. 2021, doi: 10.3389/fbioe.2021.818010.

[87] L. Li *et al.*, "Development of a Synthetic 3-Dehydroshikimate Biosensor in *Escherichia coli* for Metabolite Monitoring and Genetic Screening," *ACS Synthetic Biology*, vol. 8, no. 2, pp. 297–306, Feb. 2019, doi: 10.1021/acssynbio.8b00317.

[88] M. Rennig *et al.*, "Industrializing a Bacterial Strain for <scp>l</scp> -Serine Production through Translation Initiation Optimization," *ACS Synthetic Biology*, vol. 8, no. 10, pp. 2347–2358, Oct. 2019, doi: 10.1021/acssynbio.9b00169.

[89] C. Ramon, M. G. Gollub, and J. Stelling, "Integrating –omics data into genome-scale metabolic network models: principles and challenges," *Essays in Biochemistry*, vol. 62, no. 4, pp. 563–574, Oct. 2018, doi: 10.1042/EBC20180011.

[90] J. Alonso-Gutierrez *et al.*, "Principal component analysis of proteomics (PCAP) as a tool to direct metabolic engineering," *Metabolic Engineering*, vol. 28, pp. 123–133, Mar. 2015, doi: 10.1016/j.ymben.2014.11.011.

[91] K. Rychel, K. Decker, A. v Sastry, P. v Phaneuf, S. Poudel, and B. O. Palsson, "iModulonDB: a knowledgebase of microbial transcriptional regulation derived from machine learning," *Nucleic Acids Research*, vol. 49, no. D1, pp. D112–D120, Jan. 2021, doi: 10.1093/nar/gkaa810.

[92] M. Park, D. Kim, K. Moon, and T. Park, "Integrative Analysis of Multi-Omics Data Based on Blockwise Sparse Principal Components," *International*

Journal of Molecular Sciences, vol. 21, no. 21, p. 8202, Nov. 2020, doi: 10.3390/ijms21218202.

[93] J. C. Liao, R. Boscolo, Y.-L. Yang, L. M. Tran, C. Sabatti, and V. P. Roychowdhury, "Network component analysis: Reconstruction of regulatory signals in biological systems," *Proceedings of the National Academy of Sciences*, vol. 100, no. 26, pp. 15522–15527, Dec. 2003, doi: 10.1073/pnas.2136632100.

[94] J. Zhang *et al.*, "Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism," *Nature Communications*, vol. 11, no. 1, p. 4880, Dec. 2020, doi: 10.1038/s41467-020-17910-1.

[95] R. H. Dahl *et al.*, "Engineering dynamic pathway regulation using stress-response promoters," *Nature Biotechnology*, vol. 31, no. 11, pp. 1039–1046, Nov. 2013, doi: 10.1038/nbt.2689.

[96] J. P. N. Silva, S. V. Lopes, D. J. Grilo, and Z. Hensel, "Plasmids for Independently Tunable, Low-Noise Expression of Two Genes," *mSphere*, vol. 4, no. 3, Jun. 2019, doi: 10.1128/mSphere.00340-19.

[97] C. L. Kelly *et al.*, "Synthetic Chemical Inducers and Genetic Decoupling Enable Orthogonal Control of the *rhaBAD* Promoter," *ACS Synthetic Biology*, vol. 5, no. 10, pp. 1136–1145, Oct. 2016, doi: 10.1021/acssynbio.6b00030.

[98] L. Kim, A. Mogk, and W. Schumann, "A xylose-inducible Bacillus subtilis integration vector and its application," *Gene*, vol. 181, no. 1–2, pp. 71–76, Nov. 1996, doi: 10.1016/S0378-1119(96)00466-0.

[99] R. S. Donovan, C. W. Robinson, and B. R. Glick, "Review: Optimizing inducer and culture conditions for expression of foreign proteins under the control of thelac promoter," *Journal of Industrial Microbiology*, vol. 16, no. 3, pp. 145–154, Mar. 1996, doi: 10.1007/BF01569997.

[100] R. Bertram and W. Hillen, "The application of Tet repressor in prokaryotic gene regulation and expression," *Microbial Biotechnology*, vol. 1, no. 1, pp. 2–16, Jan. 2008, doi: 10.1111/j.1751-7915.2007.00001.x.

[101] Y. Kato, "Extremely Low Leakage Expression Systems Using Dual Transcriptional-Translational Control for Toxic Protein Production," *International Journal of Molecular Sciences*, vol. 21, no. 3, p. 705, Jan. 2020, doi: 10.3390/ijms21030705.

[102] V. M. Cardoso *et al.*, "Cost analysis based on bioreactor cultivation conditions: Production of a soluble recombinant protein using Escherichia coli BL21(DE3)," *Biotechnology Reports*, vol. 26, p. e00441, Jun. 2020, doi: 10.1016/j.btre.2020.e00441.

[103] M. J. Giacalone, A. M. Gentile, B. T. Lovitt, N. L. Berkley, C. W. Gunderson, and M. W. Surber, "Toxic protein expression in *Escherichia coli* using a rhamnose-based tightly regulated and tunable promoter system," *Biotechniques*, vol. 40, no. 3, pp. 355–364, Mar. 2006, doi: 10.2144/000112112.

[104] J. R. Newman and C. Fuqua, "Broad-host-range expression vectors that carry the l-arabinose-inducible Escherichia coli araBAD promoter and the araC regulator," *Gene*, vol. 227, no. 2, pp. 197–203, Feb. 1999, doi: 10.1016/S0378-1119(98)00601-5.

[105] J. Stülke and W. Hillen, "Carbon catabolite repression in bacteria," *Current Opinion in Microbiology*, vol. 2, no. 2, pp. 195–201, Apr. 1999, doi: 10.1016/S1369-5274(99)80034-4.

[106] A. Kuenz, M. Jäger, H. Niemi, M. Kallioinen, M. Mänttäri, and U. Prüße, "Conversion of Xylose from Birch Hemicellulose Hydrolysate to 2,3-Butanediol with Bacillus vallismortis," *Fermentation*, vol. 6, no. 3, p. 86, Sep. 2020, doi: 10.3390/fermentation6030086.

[107] E. Merino, R. A. Jensen, and C. Yanofsky, "Evolution of bacterial trp operons and their regulation," *Current Opinion in Microbiology*, vol. 11, no. 2, pp. 78–86, Apr. 2008, doi: 10.1016/j.mib.2008.02.005.

[108] J. Landberg, H. Mundhada, and A. T. Nielsen, "An autoinducible trp-T7 expression system for production of proteins and biochemicals in Escherichia coli," *Biotechnology and Bioengineering*, vol. 117, no. 5, pp. 1513–1524, May 2020, doi: 10.1002/bit.27297.

[109] P. Bachhawat, G. V. T. Swapna, G. T. Montelione, and A. M. Stock, "Mechanism of Activation for Transcription Factor PhoB Suggested by Different Modes of Dimerization in the Inactive and Active States," *Structure*, vol. 13, no. 9, pp. 1353–1363, Sep. 2005, doi: 10.1016/j.str.2005.06.006.

[110] T. Gundinger, S. Kittler, S. Kubicek, J. Kopp, and O. Spadiut, "Recombinant Protein Production in E. coli Using the phoA Expression System,"

Fermentation, vol. 8, no. 4, p. 181, Apr. 2022, doi: 10.3390/fermentation8040181.

[111] N. A. Valdez-Cruz, L. Caspeta, N. O. Pérez, O. T. Ramírez, and M. A. Trujillo-Roldán, "Production of recombinant proteins in E. coli by the heat inducible expression system based on the phage lambda pL and/or pR promoters," *Microbial Cell Factories*, vol. 9, no. 1, p. 18, Dec. 2010, doi: 10.1186/1475-2859-9-18.

[112] C.-H. Chou, A. A. Aristidou, S.-Y. Meng, G. N. Bennett, and K.-Y. San, "Characterization of a pH-inducible promoter system for high-level expression of recombinant proteins inEscherichia coli," *Biotechnology and Bioengineering*, vol. 47, no. 2, pp. 186–192, Jul. 1995, doi: 10.1002/bit.260470210.

[113] S. Brameyer *et al.*, "DNA-binding directs the localization of a membrane-integrated receptor of the ToxR family," *Communications Biology*, vol. 2, no. 1, p. 4, Dec. 2019, doi: 10.1038/s42003-018-0248-7.

[114] M. A. Lalwani *et al.*, "Optogenetic control of the lac operon for bacterial chemical and protein production," *Nature Chemical Biology*, vol. 17, no. 1, pp. 71–79, Jan. 2021, doi: 10.1038/s41589-020-0639-1.

[115] R. Spann *et al.*, "CFD predicted pH gradients in lactic acid bacteria cultivations," *Biotechnology and Bioengineering*, vol. 116, no. 4, pp. 769–780, Apr. 2019, doi: 10.1002/bit.26868.

[116] X. Luo, J. Yu, B. Wang, and J. Wang, "Heat Transfer and Hydrodynamics in Stirred Tanks with Liquid-Solid Flow Studied by CFD–DEM Method," *Processes*, vol. 9, no. 5, p. 849, May 2021, doi: 10.3390/pr9050849.

[117] A. L. Schaefer *et al.*, "LuxR- and LuxI-Type Quorum-Sensing Circuits Are Prevalent in Members of the Populus deltoides Microbiome," *Applied and Environmental Microbiology*, vol. 79, no. 18, pp. 5745–5752, Sep. 2013, doi: 10.1128/AEM.01417-13.

[118] Y. Soma *et al.*, "Design of Synthetic Quorum Sensing Achieving Induction Timing-Independent Signal Stabilization for Dynamic Metabolic Engineering of *E. coli*," *ACS Synthetic Biology*, vol. 10, no. 6, pp. 1384–1393, Jun. 2021, doi: 10.1021/acssynbio.1c00008.

[119] C. v. Dinh and K. L. J. Prather, "Development of an autonomous and bifunctional quorum-sensing circuit for metabolic flux control in engineered

Escherichia coli," *Proceedings of the National Academy of Sciences*, vol. 116, no. 51, pp. 25562–25568, Dec. 2019, doi: 10.1073/pnas.1911144116.

[120] N. Nakashima, T. Tamura, and L. Good, "Paired termini stabilize antisense RNAs and enhance conditional gene silencing in Escherichia coli," *Nucleic Acids Research*, vol. 34, no. 20, pp. e138–e138, Nov. 2006, doi: 10.1093/nar/gkl697.

[121] N. Nakashima, S. Ohno, K. Yoshikawa, H. Shimizu, and T. Tamura, "A Vector Library for Silencing Central Carbon Metabolism Genes with Antisense RNAs in Escherichia coli," *Applied and Environmental Microbiology*, vol. 80, no. 2, pp. 564–573, Jan. 2014, doi: 10.1128/AEM.02376-13.

[122] G. Durante-Rodríguez, V. de Lorenzo, and P. I. Nikel, "A Post-translational Metabolic Switch Enables Complete Decoupling of Bacterial Growth from Biopolymer Production in Engineered Escherichia coli," *ACS Synthetic Biology*, vol. 7, no. 11, pp. 2686–2697, Nov. 2018, doi: 10.1021/acssynbio.8b00345.

[123] C. Gao *et al.*, "Programmable biomolecular switches for rewiring flux in Escherichia coli," *Nature Communications*, vol. 10, no. 1, p. 3751, Dec. 2019, doi: 10.1038/s41467-019-11793-7.

[124] Y. Ishino, M. Krupovic, and P. Forterre, "History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology," *Journal of Bacteriology*, vol. 200, no. 7, Apr. 2018, doi: 10.1128/JB.00580-17.

[125] S. J. J. Brouns *et al.*, "Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes," *Science (1979)*, vol. 321, no. 5891, pp. 960–964, Aug. 2008, doi: 10.1126/science.1159689.

[126] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, "A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity," *Science (1979)*, vol. 337, no. 6096, pp. 816–821, Aug. 2012, doi: 10.1126/science.1225829.

[127] F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, and F. Zhang, "Genome engineering using the CRISPR-Cas9 system," *Nature Protocols*, vol. 8, no. 11, pp. 2281–2308, Nov. 2013, doi: 10.1038/nprot.2013.143.

[128] Y. Tong, T. Weber, and S. Y. Lee, "CRISPR/Cas-based genome engineering in natural product discovery," *Natural Product Reports*, vol. 36, no. 9, pp. 1262–1280, 2019, doi: 10.1039/C8NP00089A.

[129] V. T. Chu *et al.*, "Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells," *Nature Biotechnology*, vol. 33, no. 5, pp. 543–548, May 2015, doi: 10.1038/nbt.3198.

[130] C. R. Reisch and K. L. J. Prather, "The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in Escherichia coli," *Scientific Reports*, vol. 5, no. 1, p. 15096, Dec. 2015, doi: 10.1038/srep15096.

[131] L. S. Qi *et al.*, "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression," *Cell*, vol. 152, no. 5, pp. 1173–1183, Feb. 2013, doi: 10.1016/j.cell.2013.02.022.

[132] J. Zhang *et al.*, "Gene repression via multiplex gRNA strategy in Y. lipolytica," *Microbial Cell Factories*, vol. 17, no. 1, p. 62, Dec. 2018, doi: 10.1186/s12934-018-0909-8.

[133] J. M. Peters *et al.*, "A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria," *Cell*, vol. 165, no. 6, pp. 1493–1506, Jun. 2016, doi: 10.1016/j.cell.2016.05.003.

[134] S. Klamt, R. Mahadevan, and O. Hädicke, "When Do Two-Stage Processes Outperform One-Stage Processes?," *Biotechnology Journal*, vol. 13, no. 2, p. 1700539, Feb. 2018, doi: 10.1002/biot.201700539.

[135] M. F. Traxler *et al.*, "The global, ppGpp-mediated stringent response to amino acid starvation in Escherichia coli," *Molecular Microbiology*, vol. 68, no. 5, pp. 1128–1148, Jun. 2008, doi: 10.1111/j.1365-2958.2008.06229.x.

[136] S. E. Irving, N. R. Choudhury, and R. M. Corrigan, "The stringent response and physiological roles of (pp)pGpp in bacteria," *Nature Reviews Microbiology*, vol. 19, no. 4, pp. 256–271, Apr. 2021, doi: 10.1038/s41579-020-00470-y.

[137] S. Tojo, K. Hirooka, and Y. Fujita, "Expression of *kinA* and *kinB* of Bacillus subtilis, Necessary for Sporulation Initiation, Is under Positive Stringent Transcription Control," *Journal of Bacteriology*, vol. 195, no. 8, pp. 1656–1665, Apr. 2013, doi: 10.1128/JB.02131-12.

[138] S. Macauley-Patrick and B. Finn, "Modes of Fermenter Operation," in *Practical Fermentation Technology*, Chichester, UK: John Wiley & Sons, Ltd, pp. 69–95. doi: 10.1002/9780470725306.ch4.

[139] F. Shang, S. Wen, X. Wang, and T. Tan, "Effect of nitrogen limitation on the ergosterol production by fed-batch culture of Saccharomyces cerevisiae," *Journal of Biotechnology*, vol. 122, no. 3, pp. 285–292, Apr. 2006, doi: 10.1016/j.jbiotec.2005.11.020.

[140] V. Chubukov and U. Sauer, "Environmental Dependence of Stationary-Phase Metabolism in Bacillus subtilis and Escherichia coli," *Applied and Environmental Microbiology*, vol. 80, no. 9, pp. 2901–2909, May 2014, doi: 10.1128/AEM.00061-14.

[141] S. Li, C. B. Jendresen, and A. T. Nielsen, "Increasing production yield of tyrosine and mevalonate through inhibition of biomass formation," *Process Biochemistry*, vol. 51, no. 12, pp. 1992–2000, Dec. 2016, doi: 10.1016/j.procbio.2016.09.007.

[142] G. Bokinsky *et al.*, "HipA-Triggered Growth Arrest and -Lactam Tolerance in Escherichia coli Are Mediated by RelA-Dependent ppGpp Synthesis," *Journal of Bacteriology*, vol. 195, no. 14, pp. 3173–3182, Jul. 2013, doi: 10.1128/JB.02210-12.

[143] C.-C. Chen, R. Walia, K. J. Mukherjee, S. Mahalik, and D. K. Summers, "Indole generates quiescent and metabolically active Escherichia coli cultures," *Biotechnology Journal*, vol. 10, no. 4, pp. 636–646, Apr. 2015, doi: 10.1002/biot.201400381.

[144] S. K. Kim, W. Seong, G. H. Han, D.-H. Lee, and S.-G. Lee, "CRISPR interference-guided multiplex repression of endogenous competing pathway genes for redirecting metabolic flux in Escherichia coli," *Microbial Cell Factories*, vol. 16, no. 1, p. 188, Dec. 2017, doi: 10.1186/s12934-017-0802-x.

[145] S. Li *et al.*, "Enhanced protein and biochemical production using CRISPRi-based growth switches," *Metabolic Engineering*, vol. 38, pp. 274–284, Nov. 2016, doi: 10.1016/j.ymben.2016.09.003.

[146] J. Landberg, N. R. Wright, T. Wulff, M. J. Herrgård, and A. T. Nielsen, "CRISPR interference of nucleotide biosynthesis improves production of a single-domain antibody in *Escherichia coli*," *Biotechnology and Bioengineering*, vol. 117, no. 12, pp. 3835–3848, Dec. 2020, doi: 10.1002/bit.27536.

[147] F. Hillmann, R.-J. Fischer, F. Saint-Prix, L. Girbal, and H. Bahl, "PerR acts as a switch for oxygen tolerance in the strict anaerobe Clostridium acetobutylicum," *Molecular Microbiology*, vol. 68, no. 4, pp. 848–860, May 2008, doi: 10.1111/j.1365-2958.2008.06192.x.

[148] A. Varma and B. O. Palsson, "Metabolic Capabilities of Escherichia coli: I. Synthesis of Biosynthetic Precursors and Cofactors," *Journal of Theoretical Biology*, vol. 165, no. 4, pp. 477–502, Dec. 1993, doi: 10.1006/jtbi.1993.1202.

[149] J. Deutscher, C. Francke, and P. W. Postma, "How Phosphotransferase System-Related Protein Phosphorylation Regulates Carbohydrate Metabolism in Bacteria," *Microbiology and Molecular Biology Reviews*, vol. 70, no. 4, pp. 939–1031, Dec. 2006, doi: 10.1128/MMBR.00024-06.

[150] M. R. de Graef, S. Alexeeva, J. L. Snoep, and M. J. Teixeira de Mattos, "The Steady-State Internal Redox State (NADH/NAD) Reflects the External Redox State and Is Correlated with Catabolic Adaptation in *Escherichia coli*," *Journal of Bacteriology*, vol. 181, no. 8, pp. 2351–2357, Apr. 1999, doi: 10.1128/JB.181.8.2351-2357.1999.

[151] H. Wu, Z. Li, L. Zhou, and Q. Ye, "Improved Succinic Acid Production in the Anaerobic Culture of an *Escherichia coli pflB ldhA* Double Mutant as a Result of Enhanced Anaplerotic Activities in the Preceding Aerobic Culture," *Applied and Environmental Microbiology*, vol. 73, no. 24, pp. 7837–7843, Dec. 2007, doi: 10.1128/AEM.01546-07.

[152] C. T. Trinh, S. Huffer, M. E. Clark, H. W. Blanch, and D. S. Clark, "Elucidating mechanisms of solvent toxicity in ethanologenic Escherichia coli," *Biotechnology and Bioengineering*, vol. 106, no. 5, pp. 721–730, Mar. 2010, doi: 10.1002/bit.22743.

[153] A. J. Roe, C. O'Byrne, D. McLaggan, and I. R. Booth, "Inhibition of Escherichia coli growth by acetic acid: a problem with methionine biosynthesis and homocysteine toxicity," *Microbiology (N Y)*, vol. 148, no. 7, pp. 2215–2222, Jul. 2002, doi: 10.1099/00221287-148-7-2215.

[154] E. Rowe, B. O. Palsson, and Z. A. King, "Escher-FBA: a web application for interactive flux balance analysis," *BMC Systems Biology*, vol. 12, no. 1, p. 84, Dec. 2018, doi: 10.1186/s12918-018-0607-5.

[155] A. Kouzuma, S. Kato, and K. Watanabe, "Microbial interspecies interactions: recent findings in syntrophic consortia," *Frontiers in Microbiology*, vol. 6, May 2015, doi: 10.3389/fmicb.2015.00477.

[156] X. Qian *et al.*, "Biotechnological potential and applications of microbial consortia," *Biotechnology Advances*, vol. 40, p. 107500, May 2020, doi: 10.1016/j.biotechadv.2019.107500.

[157] K. Jawed, S. S. Yazdani, and M. AG. Koffas, "Advances in the development and application of microbial consortia for metabolic engineering," *Metabolic Engineering Communications*, vol. 9, p. e00095, Dec. 2019, doi: 10.1016/j.mec.2019.e00095.

[158] Y. Zhou, Z. Li, X. Wang, and H. Zhang, "Establishing microbial co-cultures for 3-hydroxybenzoic acid biosynthesis on glycerol," *Engineering in Life Sciences*, vol. 19, no. 5, pp. 389–395, May 2019, doi: 10.1002/elsc.201800195.

[159] H. Zhang and G. Stephanopoulos, "Co-culture engineering for microbial biosynthesis of 3-amino-benzoic acid in *Escherichia coli*," *Biotechnology Journal*, vol. 11, no. 7, pp. 981–987, Jul. 2016, doi: 10.1002/biot.201600013.

[160] M. Saini, M. Hong Chen, C.-J. Chiang, and Y.-P. Chao, "Potential production platform of n-butanol in Escherichia coli," *Metabolic Engineering*, vol. 27, pp. 76–82, Jan. 2015, doi: 10.1016/j.ymben.2014.11.001.

[161] K. Zhou, K. Qiao, S. Edgar, and G. Stephanopoulos, "Distributing a metabolic pathway among a microbial consortium enhances production of natural products," *Nature Biotechnology*, vol. 33, no. 4, pp. 377–383, Apr. 2015, doi: 10.1038/nbt.3095.

[162] W. Zhang *et al.*, "Production of naringenin from D-xylose with co-culture of *E. coli* and *S. cerevisiae*," *Engineering in Life Sciences*, vol. 17, no. 9, pp. 1021–1029, Sep. 2017, doi: 10.1002/elsc.201700039.

[163] H. Zhang and X. Wang, "Modular co-culture engineering, a new approach for metabolic engineering," *Metabolic Engineering*, vol. 37, pp. 114–121, Sep. 2016, doi: 10.1016/j.ymben.2016.05.007.

[164] Y. Liu, S. Yang, and X. Jia, "Construction of a 'nutrition supply–detoxification' coculture consortium for medium-chain-length polyhydroxyalkanoate production with a glucose–xylose mixture," *Journal of Industrial Microbiology and Biotechnology*, vol. 47, no. 3, pp. 343–354, Mar. 2020, doi: 10.1007/s10295-020-02267-7.

[165] T. Xia, M. A. Eiteman, and E. Altman, "Simultaneous utilization of glucose, xylose and arabinose in the presence of acetate by a consortium of Escherichia coli strains," *Microbial Cell Factories*, vol. 11, no. 1, p. 77, Dec. 2012, doi: 10.1186/1475-2859-11-77.

[166] J. J. Minty *et al.*, "Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass,"

Proceedings of the National Academy of Sciences, vol. 110, no. 36, pp. 14592–14597, Sep. 2013, doi: 10.1073/pnas.1218447110.

[167] S. K. Deshmukh, S. A. Verekar, and S. v. Bhave, "Endophytic fungi: a reservoir of antibacterials," *Frontiers in Microbiology*, vol. 5, Jan. 2015, doi: 10.3389/fmicb.2014.00715.

[168] P. S. Losoi, V. P. Santala, and S. M. Santala, "Enhanced Population Control in a Synthetic Bacterial Consortium by Interconnected Carbon Cross-Feeding," *ACS Synthetic Biology*, vol. 8, no. 12, pp. 2642–2650, Dec. 2019, doi: 10.1021/acssynbio.9b00316.

[169] Y. Liu, J. Chen, D. Crisante, J. M. Jaramillo Lopez, and R. Mahadevan, "Dynamic Cell Programming with Quorum Sensing-Controlled CRISPRi Circuit," *ACS Synthetic Biology*, vol. 9, no. 6, pp. 1284–1291, Jun. 2020, doi: 10.1021/acssynbio.0c00148.

[170] A. Bhatwa, W. Wang, Y. I. Hassan, N. Abraham, X.-Z. Li, and T. Zhou, "Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications," *Frontiers in Bioengineering and Biotechnology*, vol. 9, Feb. 2021, doi: 10.3389/fbioe.2021.630551.

[171] E. Samatova, J. Daberger, M. Liutkute, and M. v. Rodnina, "Translational Control by Ribosome Pausing in Bacteria: How a Non-uniform Pace of Translation Affects Protein Production and Folding," *Frontiers in Microbiology*, vol. 11, Jan. 2021, doi: 10.3389/fmicb.2020.619430.

[172] G. Zhang, M. Hubalewska, and Z. Ignatova, "Transient ribosomal attenuation coordinates protein synthesis and co-translational folding," *Nature Structural & Molecular Biology*, vol. 16, no. 3, pp. 274–280, Mar. 2009, doi: 10.1038/nsmb.1554.

[173] A. Shiber *et al.*, "Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling," *Nature*, vol. 561, no. 7722, pp. 268–272, Sep. 2018, doi: 10.1038/s41586-018-0462-y.

[174] O. Fayet, T. Ziegelhoffer, and C. Georgopoulos, "The groES and groEL heat shock gene products of Escherichia coli are essential for bacterial growth at all temperatures," *Journal of Bacteriology*, vol. 171, no. 3, pp. 1379–1385, Mar. 1989, doi: 10.1128/jb.171.3.1379-1385.1989.

[175] B. Bukau and G. C. Walker, "Cellular defects caused by deletion of the Escherichia coli dnaK gene indicate roles for heat shock protein in normal

metabolism," *Journal of Bacteriology*, vol. 171, no. 5, pp. 2337–2346, May 1989, doi: 10.1128/jb.171.5.2337-2346.1989.

[176] M. E. Rebeaud, S. Mallik, P. Goloubinoff, and D. S. Tawfik, "On the evolution of chaperones and cochaperones and the expansion of proteomes across the Tree of Life," *Proceedings of the National Academy of Sciences*, vol. 118, no. 21, p. e2020885118, May 2021, doi: 10.1073/pnas.2020885118.

[177] E. M. Sontag, R. S. Samant, and J. Frydman, "Mechanisms and Functions of Spatial Protein Quality Control," *Annual Review of Biochemistry*, vol. 86, no. 1, pp. 97–122, 2017, doi: 10.1146/annurev-biochem-060815-014616.

[178] H. Jeong, H. J. Kim, and S. J. Lee, "Complete Genome Sequence of Escherichia coli Strain BL21," *Genome Announcements*, vol. 3, no. 2, Apr. 2015, doi: 10.1128/genomeA.00134-15.

[179] D. Zeigler, "Bacillus Genetic Stock Center Strain ID 1A1133." https://bgsc.org/getdetail.php?bgscid=1A1133 (accessed Apr. 22, 2022).

[180] H. Qiu *et al.*, "Impact of cysteine variants on the structure, activity, and stability of recombinant human α-galactosidase A," *Protein Science*, vol. 24, no. 9, pp. 1401–1411, Sep. 2015, doi: 10.1002/pro.2719.

[181] P. Pecher and U. Arnold, "The effect of additional disulfide bonds on the stability and folding of ribonuclease A," *Biophysical Chemistry*, vol. 141, no. 1, pp. 21–28, Apr. 2009, doi: 10.1016/j.bpc.2008.12.005.

[182] F. Hatahet, D. Boyd, and J. Beckwith, "Disulfide bond formation in prokaryotes: history, diversity and design.," *Biochim Biophys Acta*, vol. 1844, no. 8, pp. 1402–14, Aug. 2014, doi: 10.1016/j.bbapap.2014.02.014.

[183] M. E. Reardon-Robinson and H. Ton-That, "Disulfide-Bond-Forming Pathways in Gram-Positive Bacteria," *Journal of Bacteriology*, vol. 198, no. 5, pp. 746–754, Mar. 2016, doi: 10.1128/JB.00769-15.

[184] J. L. Price *et al.*, "N-glycosylation of enhanced aromatic sequons to increase glycoprotein stability," *Biopolymers*, vol. 98, no. 3, pp. 195–211, 2012, doi: 10.1002/bip.22030.

[185] P. Goettig, "Effects of Glycosylation on the Enzymatic Activity and Mechanisms of Proteases," *International Journal of Molecular Sciences*, vol. 17, no. 12, p. 1969, Nov. 2016, doi: 10.3390/ijms17121969.

[186] G. Walsh and R. Jefferis, "Post-translational modifications in the context of therapeutic proteins," *Nature Biotechnology*, vol. 24, no. 10, pp. 1241–1252, Oct. 2006, doi: 10.1038/nbt1252.

[187] M. B. Holowko, E. K. Frow, J. C. Reid, M. Rourke, and C. E. Vickers, "Building a biofoundry," *Synthetic Biology*, vol. 6, no. 1, Feb. 2021, doi: 10.1093/synbio/ysaa026.

Chapter 2

Modulating metabolism through synthetic biology – opportunities for two-stage fermentation

Yixin Rong[1], Sheila Ingemann Jensen[1], John M Woodley[2], Alex Toftgaard Nielsen[1*]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, 2800, Denmark

[2]Department of Chemical and Biochemical Engineering, Technical University of Denmark (DTU), 2800 Kgs. Lyngby, 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

Bio-based production of fuels, chemicals and materials is needed to replace current fossil fuel based production. However, bio-based production processes are very costly, so the process needs to be as efficient as possible. Developments in synthetic biology tools has made it possible to dynamically modulate cellular metabolism during a fermentation. This can be used towards two-stage fermentations, where the process is separated into a growth and a production phase, leading to more efficient feedstock utilization and thus potentially lower costs. This article reviews the current status and some recent results in application of synthetic biology tools towards two-stage fermentations, and compares this approach to pre-existing ones, such as nutrient limitation and addition of toxins/inhibitors.

2.1 Introduction

The vast majority of chemicals produced today, ranging from commodity chemicals to pharmaceuticals, are produced using conventional petrochemical-based processes. Both in the United States and in the European Union, these processes account for 97% of the total chemicals produced [1][2]. In the past few decades, bio-based manufacturing processes have emerged as a more sustainable alternative to such processes, as they can make use of renewable feedstocks [3]. Nevertheless, broader industrial implementation of new bio-based processes has in the past often been hindered by having higher manufacturing costs compared to existing processes, which use low-cost raw materials (derived from crude oil) and have been well optimized during the many decades of deployment [4].

The cost of substrate makes up a substantial part of the operational costs in a bio-based process, typically around one third. Depending on the specific product and process, it can be up to 71% of operational cost in particular cases [5][6]. A main objective in optimizing a bio-based process is therefore to increase the yield of product per substrate; techno-economic analyses (TEA) of biorefineries show yield as the most important aspect of the fermentation step in reaching a competitive minimum sales price for the product [6]–[8]. One method to achieve a higher yield is to run the fermentation in two stages (i.e. a two-stage fermentation or TSF); generally defined as having a growth phase that prioritizes biomass formation, and a subsequent production phase that prioritizes product formation [2][9]. Figure 1 illustrates TSF as a process schematic, with a fermentation (growth) and a conversion (production) phase, possibly with separate substrates and/or cell recycling. A major advantage of having a fermentation process run primarily in the production phase is the reduction in substrate usage towards generating surplus biomass [10][11]. Being able to limit excess biomass formation can also be advantageous in regards to other cost contributors. High cell density also increases the viscosity and thus lowers the oxygen transfer rate, which can lead to the added cost factor of having to supply pure oxygen to the reactor [9]–[11].

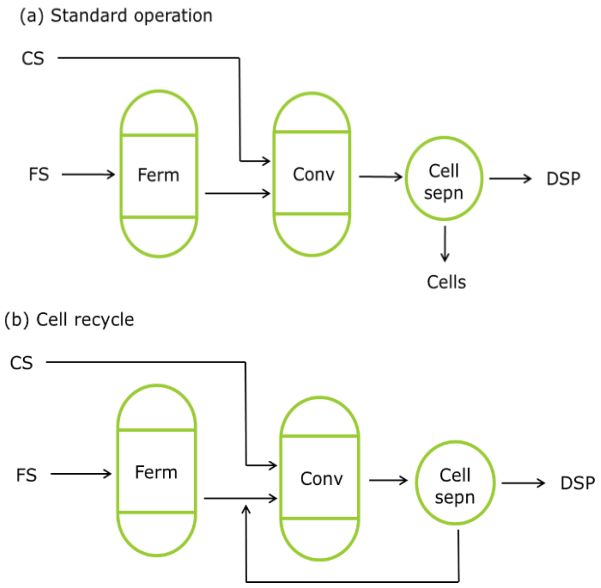


Figure 1 *Schematic representation of two-stage fermentation configurations with decoupled fermenetation (ferm) and conversion (conv) in (a) standard operation and (b) with cell recycle. FS: Fermentation substrate; CS: Conversion substrate; DSP: Downstream processing; Sepn: Separation.*

2.2 Methods to limiting biomass formation

One method to achieve reduction of cellular growth in favor of production is to limit one of the nutrients that the cell needs to grow. The limiting nutrients may for example be primary nutrients such as the carbon source, nitrogen, phosphate or sulfate, or secondary nutrients such as trace elements or vitamins [12]–[15]. Limiting a nutrient leads to changed gene expression levels in the cells, where the desired outcome is the downregulation of parts of metabolism not involved in product formation. The choice of which limitation is best can be governed by the product; carbon limitation may for example be used for secondary metabolite production, as it downregulates primary metabolism [16]. Phosphate limitation has been applied to improve yields in both primary and secondary metabolite production; e.g. in the production of L-tryptophan, as amino acid biosynthesis doesn't require phosphate [14][17][10]. Efforts have also been made in recent years to apply 'omics methods and in-silico modelling to better understand the effects of nutrient limitation on cellular metabolism [18][19]. However, in most cases it is still difficult to predict the most effective nutrient limitation strategy, meaning that preliminary fermentation

experiments have to be made for each new product. Furthermore, nutrient limitation can only affect the cellular metabolism through native regulatory mechanisms, leading to a limited set of possibilities on how the metabolism can be controlled. This can also lead to activation of cellular responses to starvation, such as stringent response, which decreases the metabolic activity of the cell [20]. For some organisms, such as *Escherichia coli*, research has been done to create strains that lack or have a reduced stringent response towards nutrient limitation like nitrogen [21].

Consequently, instead of limiting what the cells require to grow, the addition of inhibitory compounds or expression of toxins has also been used as a way of limiting biomass formation. These toxins/inhibitors often target different cellular proteins involved in processes related to growth, while also maintaining a level of metabolic activity. A study by Li and co-workers looked into using different toxic chemical inhibitors, such as kanamycin, tetracycline and 5-fluorouracil, targeting different cellular and metabolic processes in *E. coli* that are required for growth, as a way to limit biomass formation and thereby improve the production of tyrosine and mevalonate [10]. Their findings suggest that for the two products tested in the study, sulphate limitation was the most efficient way of increasing yield through growth inhibition. Another study by Chen and co-workers used indole to achieve quiescent but metabolically active cells that maintain GFP expression after growth arrest [22]. Growth inhibition can also be achieved by inducible expression of toxin proteins. Bokinsky and co-workers showed that HipA could be overexpressed to stop growth, while maintaining mevalonate productivity at a level equivalent to non-arrested cells [23]. These studies show that while it is possible to suppress growth using inhibitors and toxins while maintaining a level of metabolic activity, the benefits to product formation may be limited, possibly due to uncharacterized off-target effects.

2.3 Synthetic biology tools to control cellular growth and metabolism

Metabolic engineering of cell factories has long been applied to increase production, where the metabolism can be changed more precisely [24]. However, some genes can not be targeted for knock-outs if they have a highly deleterious effect on cell growth. In recent years, tools have emerged that enable the inducible repression of targeted genes without the need for deleting genes from the organism's genome.

In this way, it is possible to target genes that limit cell growth in order to facilitate a TSF process. The deletion/repression of a target gene should also not negatively impact overall metabolic activity, as inhibition of metabolic genes can lead to stringent response, and downregulation of the translational machinery can lead to lower levels of metabolic enzymes and thus reduce productivity [25][26].

Methods for inducible repression of gene expression can generally be divided into two overall elements: a synthetic biology tool that can repress gene expression and an inducible promoter. Important properties for the tool include stability, fold of repression, and host organism compatibility. Additionally, the ability to switch the repression on and off multiple times can be necessary if the process is run with cell recycling, such as shown in Figure 1B. In recent years, CRISPR interference (CRISPRi) has emerged as a method for dynamic modulation of gene expression across various production hosts [27]–[30]. Table 1 lists some studies that have used CRISPRi to dynamically switch metabolism to favor product formation. CRISPRi can repress gene expression very effectively, up to 300-fold in *E. coli* and >100-fold in *Bacillus subtilis* [27][28]. Other methods of controlling gene expression include the pLac + pTet system used in a study reported by Soma and co-workers, where the target gene is controlled by the pTet promoter, and the TetR repressor is controlled by the pLac promoter [31]. Another tool that has been applied for gene silencing is RNA interference (RNAi). Studies have been for example been made in the industrial relevant host organisms *E. coli* and *S. cerevisiae*; in *E. coli*, RNAi has been shown to be able to repress expression to similar levels compared to a knock-out, and in S. cerevisiae, RNAi has been shown to repress expression at ~80-95% efficiency [32] [33]. RNAi has yet to be studied in the context of TSF's. Whereas both inducible promoters, CRISPRi and RNAi assert their regulation on a transcriptional/post-transcriptional level, a study by Durante-Rodríguez and co-workers demonstrated a method to control the protein levels through targeting to proteosomal degradation [34]. Here a degradation signal is attached to the target protein via a linker that includes a site for a specific protease. When the specific protease is induced, the degradation signal is removed and the protein is stable, and vice versa.

Comparing the methods, CRISPRi is the most studied. It has been shown to function in many industrial chassis organisms, and some non-conventional host organisms, such as halophilic bacteria and obligate anaerobes like *Clostridium* species [35]–[37]. In terms of stability, a study by Li and co-workers deployed CRISPRi based growth

decoupling for up to 48 hours [38]. Likewise, the two first studies listed in Table 1 also ran their fermentations for 48 hours. CRISPRi has been deployed in fermentations up to 144 hours, although it is not clear from the results if the repression remains effective throughout [39]. It has also been demonstrated that it is possible to deactivate the repression from CRISPRi with a pTet-promoter by washing the cells in fresh media without an inducer [27]. For direct control by inducible promoters (Soma and coworkers), the fermentation was run for 72 hours, and the protein degradation based approach by Durante-Rodriguez and coworkers was tested for 24 hours [31][34]. Although CRISPRi can potentially be stable enough for typical durations of industrial fermentations, growth decoupling can also be susceptible to mutations, e.g. in the dCas9 gene in the case of CRISPRi, that enable the cells to escape the growth decoupling and take over the population. This is especially challenging if one wishes to utilize biomass recycling in the process. One aspect worth considering is the added cellular burden of expressing and then degrading the target gene when control is exerted at the protein level. In this regard, RNAi or controlling the target gene directly with an inducible promoter could potentially be the least burdensome, as CRISPRi also involves expressing dCas9, which is a large protein.

Inducible promoters are needed in order to control when in the fermentation the target gene(s) are repressed. The most important properties of inducible promoters are homogeneity of expression, “leakiness”, i.e. level of expression without the presence of the inducer, expression strength when induced, stability and host compatibility. Table 2 lists inducible promoters that have been used in conjunction with CRISPRi as well as a few other commonly used inducible promoters. Leaky expression can increase the likelihood of escapee mutations, as there would be a selective pressure favoring escapees already during the pre-culture stage. The inducer prices vary greatly, and can be prohibitory, especially with lower value specialty and commodity chemicals. Furthermore, promoters that get induced by sugars can limit what feedstocks can be used. A few studies have looked at using inducer-free promoters using optogenic (blue light), quorum sensing, (non-essential) nutrient depletion, or temperature control to induce expression [40-[43]. Some of these have not been studied at large scale, which could pose problems for some inducer-free promoters; it might e.g. be challenging to apply blue light to the entire reactor. Temperature gradients could affect temperature-based induction,

depending on how sensitive the promoter is. Out of these mechanisms, nutrient depletion based promoters, such as the tryptophan system developed by Landberg and coworkers, has the most potential as an alternative to inducer dependent promoters [42].

Besides the promoter and tool applied for inducible repression of a gene, equally important is determining the target gene(s). Earlier approaches to identify target genes have often been based on rational engineering [31][38]. In the study by Soma and co-workers, the gene citrate synthase *gltA* was expressed with a tetR repressible promoter. Repressing the expression of the gene prevents the pre-cursor acetyl-CoA from going to the TCA cycle instead of the product pathway, while inhibiting the TCA cycle also reduces biomass formation [31]. This resulted in an increase in titer and yield up to 3.7 and 3.1 fold respectively compared to a wild-type strain. However, this TSF process would only work for products derived from acetyl-CoA. Similarly, the studies listed in Table 1 also use target genes often derived from rational design, and aimed at improving the production of a specific compound. A study by Li and co-workers attempted to find more broadly applicable target genes. They reasoned based on existing literature that DNA origin of replication and pyrimidine biosynthesis would limit growth upon inhibition while not being associated with stringent response [38]. They found that using an inducible CRISPRi system to repress the pyrimidine biosynthesis gene *pyrF* gave an attenuation of growth, and a concomitant 2.16-fold increase in GFP titer, as well as a 2.9-fold increase in specific productivity and a 41% increase in yield when producing mevalonate as a proof of concept [38]. The group later published another study on a more systematic method of identifying potential targets for growth decoupling using CRISPRi, by screening a library of 12238 sgRNA's targeting all coding and also non-coding genomic locations in *E. coli* [44].

Table 1 *Examples on the use of CRISPRi as a metabolic switch to improve production of chemicals*

Organism	Product	Target(s)	Duration of fermentation	Titer	Rate	Yield	Ref.
E. coli	n-butanol	*pta, frdA, ldhA, and adhE*	48 h	-	3.2 fold increase	5.4 fold increase	[48]
E. coli	Butanoic acid	*fabI*	48 h	-	6.1 fold increase	5.9 fold increase	[47]
K. marxianus	Ethyl acetate	*ACO2b, SDH2, RIP1, and MSS51*	14 h	-	3.8 fold increase	-	[46]
S. cerevisiae	Beta-amyrin	*ADH1, ADH4, ADH5, ADH6, CIT2, MLS2 and ERG7*	144 h	43% increase	-	-	[39]
B. subtilis	GlcNAc	*zwf, pfkA, glmM*	90 h	13.2% increase	-	84.1% increase	[45]

Table 2 *Examples of conventional inducible expression systems. * Estimated based on the lowest price from Sigma-Aldrich. Prices may vary with bulk orders and other suppliers ** Used in studies to control CRISPRi expression *** Lactose can be used as cheaper inducer*

Organism	Promoter	Inducer	Price*	Stability of inducer	Leakiness
E. coli	pRha ** [48]	Rhamnose	0.14 USD/L	Depends on consumption	Low [55]
E. coli	pNEW ** [47]	Cumate	1.21 USD/L	-	Low [54]
Various	pTet **[27][38]	aTc	3.06 USD/L	Half-life ~20h [49]	Low [53]
B. subtilis	pXyl ** [45]	Xylose	0.38 USD/L	Depends on consumption	Low [52]
Various	pLac	IPTG ***	1.81 USD/L	No degradation at 32 h [49]	Medium [51]
E. coli, C. glutamicum	pAra	Arabinose	0.11 USD/L	Depends on consumption	Low [50]

2.4 Future perspectives

While some TSF processes are used in industrial production, using switches such as nutrient limitation or aerobic growth phase/anaerobic production phase, synthetic biology tool based methods have yet to be adopted [56][57]. Figure 2 lists some of the advantages and disadvantages of the different methods currently available.

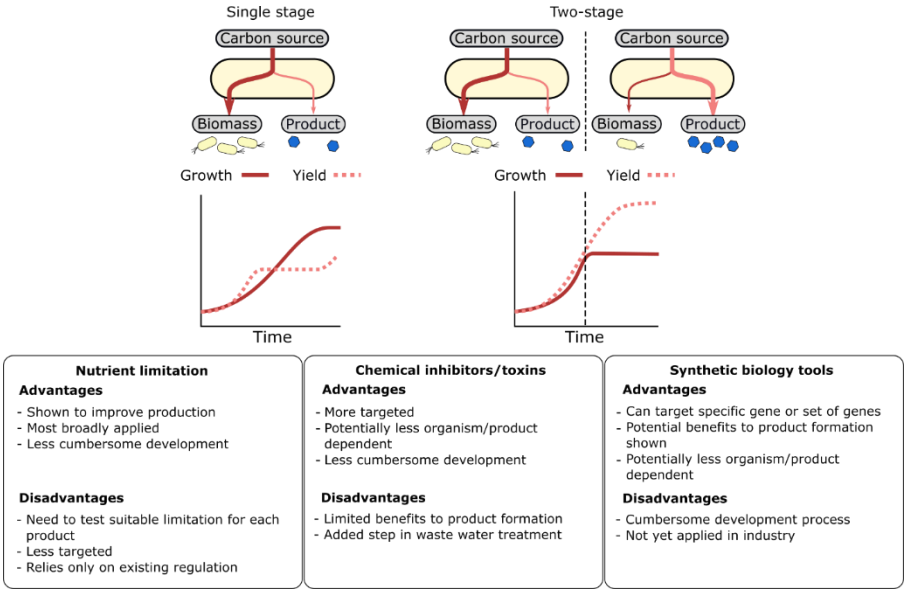


Figure 2 *Summary of advantages and disadvantages of different ways of achieving a two-stage fermentation*

Reasons for the limited uptake of synthetic biology based switches could be due to both technological barriers and incomplete knowledge. The main technological challenge is to improve the stability and robustness of these methods, which is crucial in industrial settings. Most studies are done in small lab-scale experiments without sufficient information to evaluate the feasibility of scale-up, which can be cumbersome and expensive [58]. Furthermore, it can be resource heavy to develop systems for novel host organisms, even when following previously described methods. Lastly, it can be difficult to make a comprehensive evaluation of the potential benefits of TSF for a particular process. Limiting biomass concentration can affect the volumetric productivity, if an increase in specific productivity is not enough to compensate; the potential trade off in productivity in favor of increased yield should be considered, and the optimal balance of rate and yield likely depends on product value. Other questions affecting the potential cost benefit at the end include: Does the organism grow fast and to a high enough cell concentration so it is relevant to limit growth? What is the value of the product compared to the cost of the substrate? Is aeration a crucial factor? In summary, synthetic biology tools are diverse and can enable finer control of cellular metabolism during a two-stage fermentation process, but further studies into industrial application and cost-

benefit analyses, such as scale-down of industrial fermentations and detailed techno-economic analyses, are likely needed for broader adaptation and application in industrial settings.

2.5 Acknowledgements

This work was funded by the Novo Nordisk Foundation within the framework of the Fermentation-based Biomanufacturing Initiative (FBM), grant number: NNF17SA0031362.

2.5 References for Chapter 2

[1] J. Spekreijse, T. Lammens, C. Parisi, T. Ronzon, and M. Vis, Insights into the European market for bio-based chemicals, vol. 19, no. October. 2019. doi: 10.2760/673071.

[2] J. M. Burg, C. B. Cooper, Z. Ye, B. R. Reed, E. A. Moreb, and M. D. Lynch, "Large-scale bioprocess competitiveness: the potential of dynamic metabolic control in two-stage fermentations," Current Opinion in Chemical Engineering, vol. 14, pp. 121–136, Nov. 2016, doi: 10.1016/j.coche.2016.09.008.

[3] R. A. Sheldon and J. M. Woodley, "Role of Biocatalysis in Sustainable Chemistry," 2018, doi: 10.1021/acs.chemrev.7b00203.

[4] L. Nattrass, C. Biggs, A. Bauen, C. Parisi, E. Rodríguez-Cerezo, and M. Gómez-Barbero, The EU bio-based industry : Results from a survey. 2016. doi: 10.2791/806858.

[5] A. J. J. Straathof, The Proportion of Downstream Costs in Fermentative Production Processes, Second Edi., vol. 2. Elsevier B.V., 2011. doi: 10.1016/B978-0-08-088504-9.00492-X.

[6] D. J. McClelland et al., "Renewable linear alpha-olefins by base-catalyzed dehydration of biologically-derived fatty alcohols," Green Chemistry, vol. 23, no. 12, pp. 4338–4354, 2021, doi: 10.1039/D1GC00243K.

[7] Y. Li et al., "Sustainable Lactic Acid Production from Lignocellulosic Biomass," ACS Sustainable Chemistry & Engineering, vol. 9, no. 3, pp. 1341–1351, Jan. 2021, doi: 10.1021/acssuschemeng.0c08055.

[8] S. S. Bhagwat et al., "Sustainable Production of Acrylic Acid via 3-Hydroxypropionic Acid from Lignocellulosic Biomass," ACS Sustainable Chemistry & Engineering, vol. 9, no. 49, pp. 16659–16669, Dec. 2021, doi: 10.1021/acssuschemeng.1c05441.

[9] S. Klamt, R. Mahadevan, and O. Hädicke, "When Do Two-Stage Processes Outperform One-Stage Processes?," Biotechnology Journal, vol. 13, no. 2, p. 1700539, Feb. 2018, doi: 10.1002/biot.201700539.

[10] S. Li, C. B. Jendresen, and A. T. Nielsen, "Increasing production yield of tyrosine and mevalonate through inhibition of biomass formation," Process Biochemistry, vol. 51, no. 12, pp. 1992–2000, Dec. 2016, doi: 10.1016/j.procbio.2016.09.007.

[11] N. Venayak, K. Raj, R. Jaydeep, and R. Mahadevan, "An Optimized Bistable Metabolic Switch To Decouple Phenotypic States during Anaerobic Fermentation," ACS Synthetic Biology, vol. 7, no. 12, pp. 2854–2866, Dec. 2018, doi: 10.1021/acssynbio.8b00284.

[12] L. M. Halka, C. Nowacki, A. Kleinschmidt, K. Koenen, and R. Wichmann, "Glucose limited feed strategy leads to increased production of fusicocca-2,10(14)-diene by Saccharomyces cerevisiae," AMB Express, vol. 8, no. 1, p. 132, Dec. 2018, doi: 10.1186/s13568-018-0662-8.

[13] M. Perez-Zabaleta, M. Guevara-Martínez, M. Gustavsson, J. Quillaguamán, G. Larsson, and A. J. A. van Maris, "Comparison of engineered Escherichia coli AF1000 and BL21 strains for (R)-3-hydroxybutyrate production in fed-batch cultivation," Applied Microbiology and Biotechnology, vol. 103, no. 14, pp. 5627–5639, Jul. 2019, doi: 10.1007/s00253-019-09876-y.

[14] T. Schuhmacher, M. Löffler, T. Hurler, and R. Takors, "Phosphate limited fed-batch processes: Impact on carbon usage and energy metabolism in Escherichia coli," Journal of Biotechnology, vol. 190, pp. 96–104, Nov. 2014, doi: 10.1016/j.jbiotec.2014.04.025.

[15] D. K. Kundiyana, R. L. Huhnke, and M. R. Wilkins, "Effect of nutrient limitation and two-stage continuous fermentor design on productivities during 'Clostridium ragsdalei' syngas fermentation," Bioresource Technology, vol. 102, no. 10, pp. 6058–6064, May 2011, doi: 10.1016/j.biortech.2011.03.020.

[16] Z. Sun, J. A. Ramsay, M. Guay, and B. A. Ramsay, "Carbon-limited fed-batch production of medium-chain-length polyhydroxyalkanoates from nonanoic acid by Pseudomonas putida KT2440," Applied Microbiology and Biotechnology, vol. 74, no. 1, pp. 69–77, Feb. 2007, doi: 10.1007/s00253-006-0655-4.

[17] P. Bruheim, H. Sletta, M. Bibb, J. White, and D. Levine, "High-yield actinorhodin production in fed-batch culture by a Streptomyces lividans strain overexpressing the pathway-specific activator gene actII-ORF4," Journal of Industrial Microbiology and Biotechnology, vol. 28, no. 2, pp. 103–111, 2002, doi: 10.1038/sj/jim/7000219.

[18] M. Lempp, N. Farke, M. Kuntz, S. A. Freibert, R. Lill, and H. Link, "Systematic identification of metabolites controlling gene expression in E. coli," Nature Communications, vol. 10, no. 1, p. 4463, Dec. 2019, doi: 10.1038/s41467-019-12474-1.

[19] Z. A. Irani, A. Maghsoudi, S. A. Shojaosadati, and E. Motamedian, "Development and in silico analysis of a new nitrogen-limited feeding strategy for fed-batch cultures of Pichia pastoris based on a simple pH-control system," Biochemical Engineering Journal, vol. 98, pp. 1–9, Jun. 2015, doi: 10.1016/j.bej.2015.02.016.

[20] C. C. Boutte and S. Crosson, "Bacterial lifestyle shapes stringent response activation," Trends in Microbiology, vol. 21, no. 4, pp. 174–180, Apr. 2013, doi: 10.1016/j.tim.2013.01.002.

[21] M. Ziegler, J. Zieringer, and R. Takors, "Transcriptional profiling of the stringent response mutant strain E. coli SR reveals enhanced robustness to large-scale conditions," Microbial Biotechnology, vol. 14, no. 3, pp. 993–1010, May 2021, doi: 10.1111/1751-7915.13738.

[22] C.-C. Chen, R. Walia, K. J. Mukherjee, S. Mahalik, and D. K. Summers, "Indole generates quiescent and metabolically active Escherichia coli cultures," Biotechnology Journal, vol. 10, no. 4, pp. 636–646, Apr. 2015, doi: 10.1002/biot.201400381.

[23] G. Bokinsky et al., "HipA-Triggered Growth Arrest and -Lactam Tolerance in Escherichia coli Are Mediated by RelA-Dependent ppGpp Synthesis," Journal of Bacteriology, vol. 195, no. 14, pp. 3173–3182, Jul. 2013, doi: 10.1128/JB.02210-12.

[24] J. Nielsen and J. D. Keasling, "Engineering Cellular Metabolism," Cell, vol. 164, no. 6, pp. 1185–1197, 2016, doi: 10.1016/j.cell.2016.02.004.

[25] S. E. Irving, N. R. Choudhury, and R. M. Corrigan, "The stringent response and physiological roles of (pp)pGpp in bacteria," Nature Reviews Microbiology, vol. 19, no. 4, pp. 256–271, Apr. 2021, doi: 10.1038/s41579-020-00470-y.

[26] M. F. Traxler et al., "The global, ppGpp-mediated stringent response to amino acid starvation in Escherichia coli," Molecular Microbiology, vol. 68, no. 5, pp. 1128–1148, Jun. 2008, doi: 10.1111/j.1365-2958.2008.06229.x.

[27] L. S. Qi et al., "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression," Cell, vol. 152, no. 5, pp. 1173–1183, Feb. 2013, doi: 10.1016/j.cell.2013.02.022.

[28] Y. Wu et al., "Design of a programmable biosensor-CRISPRi genetic circuits for dynamic and autonomous dual-control of metabolic flux in Bacillus subtilis," Nucleic Acids Research, vol. 48, no. 2, pp. 996–1009, Jan. 2020, doi: 10.1093/nar/gkz1123.

[29] S. Cleto, J. V. Jensen, V. F. Wendisch, and T. K. Lu, "Corynebacterium glutamicum Metabolic Engineering with CRISPR Interference (CRISPRi)," ACS Synthetic Biology, vol. 5, no. 5, pp. 375–385, May 2016, doi: 10.1021/acssynbio.5b00216.

[30] A. Momen-Roknabadi, P. Oikonomou, M. Zegans, and S. Tavazoie, "An inducible CRISPR interference library for genetic interrogation of Saccharomyces cerevisiae biology," Communications Biology, vol. 3, no. 1, p. 723, Dec. 2020, doi: 10.1038/s42003-020-01452-9.

[31] Y. Soma, K. Tsuruno, M. Wada, A. Yokota, and T. Hanai, "Metabolic flux redirection from a central metabolic pathway toward a synthetic pathway using a metabolic toggle switch," Metabolic Engineering, vol. 23, pp. 175–184, May 2014, doi: 10.1016/j.ymben.2014.02.008.

[32] G. Magistro, C. Magistro, C. G. Stief, and S. Schubert, "A simple and highly efficient method for gene silencing in Escherichia coli," Journal of Microbiological Methods, vol. 154, pp. 25–32, Nov. 2018, doi: 10.1016/j.mimet.2018.10.003.

[33] Y. Chen, E. Guo, J. Zhang, and T. Si, "Advances in RNAi-Assisted Strain Engineering in Saccharomyces cerevisiae," Frontiers in Bioengineering and Biotechnology, vol. 8, Jul. 2020, doi: 10.3389/fbioe.2020.00731.

[34] G. Durante-Rodríguez, V. de Lorenzo, and P. I. Nikel, "A Post-translational Metabolic Switch Enables Complete Decoupling of Bacterial Growth from Biopolymer Production in Engineered Escherichia coli," ACS Synthetic Biology, vol. 7, no. 11, pp. 2686–2697, Nov. 2018, doi: 10.1021/acssynbio.8b00345.

[35] W. Tao, L. Lv, and G.-Q. Chen, "Engineering Halomonas species TD01 for enhanced polyhydroxyalkanoates synthesis via CRISPRi," Microbial Cell Factories, vol. 16, no. 1, p. 48, Dec. 2017, doi: 10.1186/s12934-017-0655-3.

[36] L. Lin et al., "Optimising PHBV biopolymer production in haloarchaea via CRISPRi-mediated redirection of carbon flux," Communications Biology, vol. 4, no. 1, p. 1007, Dec. 2021, doi: 10.1038/s42003-021-02541-z.

[37] B. M. Woolston, D. F. Emerson, D. H. Currie, and G. Stephanopoulos, "Rediverting carbon flux in Clostridium ljungdahlii using CRISPR interference (CRISPRi)," Metabolic Engineering, vol. 48, pp. 243–253, Jul. 2018, doi: 10.1016/j.ymben.2018.06.006.

[38] S. Li et al., "Enhanced protein and biochemical production using CRISPRi-based growth switches," Metabolic Engineering, vol. 38, pp. 274–284, Nov. 2016, doi: 10.1016/j.ymben.2016.09.003.

[39] J. Ni, G. Zhang, L. Qin, J. Li, and C. Li, "Simultaneously down-regulation of multiplex branch pathways using CRISPRi and fermentation optimization for enhancing β-amyrin production in Saccharomyces cerevisiae," Synthetic and Systems Biotechnology, vol. 4, no. 2, pp. 79–85, Jun. 2019, doi: 10.1016/j.synbio.2019.02.002.

[40] P. Wu, Y. Chen, M. Liu, G. Xiao, and J. Yuan, "Engineering an Optogenetic CRISPRi Platform for Improved Chemical Production," ACS Synthetic Biology, vol. 10, no. 1, pp. 125–131, Jan. 2021, doi: 10.1021/acssynbio.0c00488.

[41] J. Tian, G. Yang, Y. Gu, X. Sun, Y. Lu, and W. Jiang, "Developing an endogenous quorum-sensing based CRISPRi circuit for autonomous and tunable dynamic regulation of multiple targets in Streptomyces," Nucleic Acids Research, vol. 48, no. 14, pp. 8188–8202, Aug. 2020, doi: 10.1093/nar/gkaa602.

[42] J. Landberg, H. Mundhada, and A. T. Nielsen, "An autoinducible trp-T7 expression system for production of proteins and biochemicals in Escherichia coli," Biotechnology and Bioengineering, vol. 117, no. 5, pp. 1513–1524, May 2020, doi: 10.1002/bit.27297.

[43] S. Restrepo-Pineda, N. O. Pérez, N. A. Valdez-Cruz, and M. A. Trujillo-Roldán, "Thermoinducible expression system for producing recombinant proteins in Escherichia coli : advances and insights," FEMS Microbiology Reviews, Apr. 2021, doi: 10.1093/femsre/fuab023.

[44] S. Li et al., "Genome-Wide CRISPRi-Based Identification of Targets for Decoupling Growth from Production," ACS Synthetic Biology, vol. 9, no. 5, pp. 1030–1040, May 2020, doi: 10.1021/acssynbio.9b00143.

[45] Y. Wu et al., "CRISPRi allows optimal temporal control of N-acetylglucosamine bioproduction by a dynamic coordination of glucose and xylose metabolism in Bacillus subtilis," Metabolic Engineering, vol. 49, pp. 232–241, Sep. 2018, doi: 10.1016/j.ymben.2018.08.012.

[46] A.-K. Löbs, C. Schwartz, S. Thorwall, and I. Wheeldon, "Highly Multiplexed CRISPRi Repression of Respiratory Functions Enhances Mitochondrial Localized Ethyl Acetate Biosynthesis in Kluyveromyces marxianus," ACS Synthetic Biology, vol. 7, no. 11, pp. 2647–2655, Nov. 2018, doi: 10.1021/acssynbio.8b00331.

[47] X. Ji, H. Zhao, H. Zhu, K. Zhu, S.-Y. Tang, and C. Lou, "CRISPRi/dCpf1-mediated dynamic metabolic switch to enhance butenoic acid production in Escherichia coli," Applied Microbiology and Biotechnology, vol. 104, no. 12, pp. 5385–5393, Jun. 2020, doi: 10.1007/s00253-020-10610-2.

[48] S. K. Kim, W. Seong, G. H. Han, D.-H. Lee, and S.-G. Lee, "CRISPR interference-guided multiplex repression of endogenous competing pathway genes for redirecting metabolic flux in Escherichia coli," Microbial Cell Factories, vol. 16, no. 1, p. 188, Dec. 2017, doi: 10.1186/s12934-017-0802-x.

[49] N. Politi et al., "Half-life measurements of chemical inducers for recombinant gene expression," Journal of Biological Engineering, vol. 8, no. 1, p. 5, 2014, doi: 10.1186/1754-1611-8-5.

[50] L. M. Guzman, D. Belin, M. J. Carson, and J. Beckwith, "Tight regulation, modulation, and high-level expression by vectors containing the arabinose P(BAD) promoter," Journal of Bacteriology, vol. 177, no. 14, pp. 4121–4130, 1995, doi: 10.1128/jb.177.14.4121-4130.1995.

[51] Y. Kato, "Extremely Low Leakage Expression Systems Using Dual Transcriptional-Translational Control for Toxic Protein Production," International Journal of Molecular Sciences, vol. 21, no. 3, p. 705, Jan. 2020, doi: 10.3390/ijms21030705.

[52] A. P. Bhavsar, X. Zhao, and E. D. Brown, "Development and Characterization of a Xylose-Dependent System for Expression of Cloned Genes inBacillus subtilis: Conditional Complementation of a Teichoic Acid Mutant," Applied and Environmental Microbiology, vol. 67, no. 1, pp. 403–410, Jan. 2001, doi: 10.1128/AEM.67.1.403-410.2001.

[53] F. Lausberg, A. R. Chattopadhyay, A. Heyer, L. Eggeling, and R. Freudl, "A tetracycline inducible expression vector for Corynebacterium glutamicum allowing tightly regulable gene expression," Plasmid, vol. 68, no. 2, pp. 142–147, Sep. 2012, doi: 10.1016/j.plasmid.2012.05.001.

[54] Y. J. Choi et al., "Novel, Versatile, and Tightly Regulated Expression System for E scherichia coli Strains," Applied and Environmental Microbiology, vol. 76, no. 15, pp. 5058–5066, Aug. 2010, doi: 10.1128/AEM.00413-10.

[55] A. Haldimann, L. L. Daniels, and B. L. Wanner, "Use of new methods for construction of tightly regulated arabinose and rhamnose promoter fusions in studies of the Escherichia coli phosphate regulon.," J Bacteriol, vol. 180, no. 5, pp. 1277–86, Mar. 1998, doi: 10.1128/JB.180.5.1277-1286.1998.

[56] Y. Cao, R. Zhang, C. Sun, T. Cheng, Y. Liu, and M. Xian, "Fermentative Succinate Production: An Emerging Technology to Replace the Traditional Petrochemical Processes," BioMed Research International, vol. 2013, pp. 1–12, 2013, doi: 10.1155/2013/723412.

[57] X.-B. Li, G.-R. Zhao, H. Zheng, and Y.-J. Yuan, "Improved industrial fermentation of lincomycin by phosphorus feeding," Process Biochemistry, vol. 42, no. 4, pp. 662–668, Apr. 2007, doi: 10.1016/j.procbio.2006.12.007.

[58] F. Delvigne, R. Takors, R. Mudde, W. van Gulik, and H. Noorman, "Bioprocess scale-up/down as integrative enabling technology: from fluid mechanics to systems biology and beyond," Microbial Biotechnology, vol. 10, no. 5, pp. 1267–1274, Sep. 2017, doi: 10.1111/1751-7915.12803.

Chapter 3

CRISPRi-mediated metabolic switch enables concomitant "anaerobic" and aerobic fermentations

Yixin Rong[1], Emre Özdemir[1], Songyuan Li[1], Arrate Sainz de la Maza Larrea[1], Alex Toftgaard Nielsen[1], Sheila Ingemann Jensen[1*]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet Building 220, 2800 Kongens Lyngby, Denmark

*Corresponding author

Sheila Ingemann Jensen

The Novo Nordisk Foundation Center for Biosustainability

Technical University of Denmark

Kemitorvet, DK-2800 Kongens Lyngby, Denmark

Email: shin@biosustain.dtu.dk

Abstract

Replacing petrochemicals with compounds from bio-based manufacturing processes remains an important part of the global effort to move towards a sustainable future. However, achieving economic viability requires both optimized cell factories and innovative processes. In this study, we have addressed this challenge by developing a novel fermentation platform, which enables two concurrent fermentations in one bioreactor. We first constructed a xylitol producing *Escherichia coli* strain that enabled us to use CRISPRi-mediated gene silencing to switch the metabolism from aerobic to anaerobic, even when the bacteria are under aerobic conditions. The switch also decouples growth from production, which further increases the yield. The strain produces acetate as an unwanted byproduct, which is metabolized under aerobic conditions by a secondary *E. coli* strain that through constraint-based metabolic modelling was designed to co-valorize glucose and the excreted acetate to a secondary product. This unique syntrophic consortium concept facilitates the implementation of “two-fermentation in one go”, where the concurrent fermentation displays similar titers as compared to two separate single strain fermentations.

3.1 Introduction

Responsible production of food and chemicals are part of the sustainable development goals set up by the United Nations [1]. Bio-based manufacturing is being pursued as a key technology to enable this transition, not only for food and feed applications, but also as a sustainable alternative to fossil fuel based production of chemicals, fuels, and materials. Traditionally the development and optimization of synthetic bioprocesses focus on single-strain fermentations, however, microbial consortia based processes have been pursued as an attractive alternative in particular for more complex processes. Synthetic consortia that facilitate the production of complex molecules [2][3] or the breakdown and further valorization of cellulose [4]-[6] have successfully been engineered by dividing functional traits between the different members, and niche partitioning has been used to consume and valorize a mixture of carbon sources [7]-[9].

One of the main challenges of using microbial consortia in synthetic biology is stable co-culturing. However, the development of designed niche partitioning and mutualistic syntrophic interactions have been successfully shown to enable consortia based bioprocesses.

In this study, we present a unique consortia based fermentation process, where an engineered strain is able to switch to anaerobic metabolism through CRISPR interference (CRISPRi) mediated gene silencing [10] instead of changing the physical conditions inside the bioreactor. Because the bioreactor remains aerobic, additional strain(s) can still respire and therefore use energy-requiring pathways. To demonstrate the concept, we engineered a xylitol producing strain with inducible anaerobic physiology and an aerobic co-metabolizing isobutyric acid producing strain. We stabilized the consortia based process through inducible growth decoupled production [11] combined with mutualistic syntrophic dependency. The overall results demonstrate the successful engineering of a syntrophic microbial consortium, which enables two concurrent fermentations in bioreactor that displays comparable titers and productivity as compared to two single strain fermentations.

3.2 Results

In this study, we aimed at developing a novel fermentation platform, which enables concurrent aerobic and anaerobic metabolism in one bioreactor. To demonstrate the concept, we engineered an “anaerobic” xylitol producing strain and an aerobic

isobutyric acid producing strain that operate in a syntrophic consortia process as illustrated in Figure 1A.

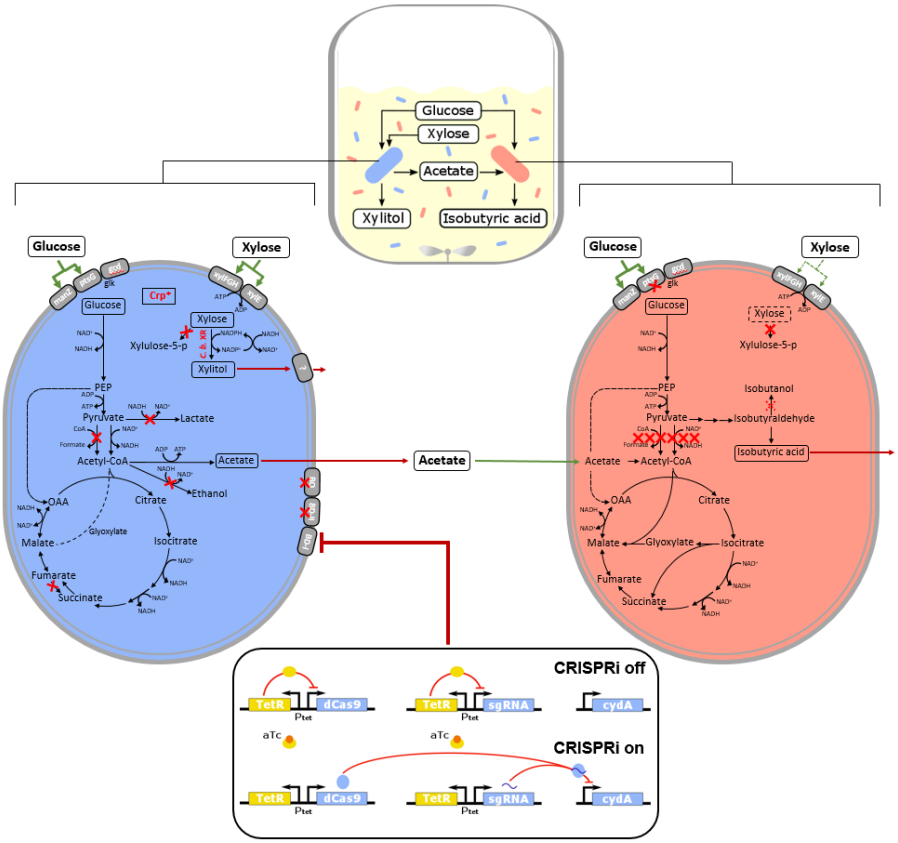


Figure 1 *Overview of the anaerobic-aerobic consortia. A) Overall illustration of the process; one strain consumes xylose and glucose and produces xylitol and acetic acid, while the other strain consumes glucose and acetic acid and produces isobutyric acid. B) Overview of knockouts in the "anaerobic" xylitol producing strain. C) Overview of knockouts in the aerobic isobutyric acid producing strain. D) Illustration of the anhydrotetracycline inducible CRISPRi-enabled switch to repress the expression of cytochrome BD-I.*

3.2.1 Construction of a xylitol producing strain with inducible anaerobic physiology under aerobic conditions

To construct a xylitol producing strain with inducible anaerobic physiology under aerobic conditions, we first created a strain devoid of native fermentation pathways by knocking out *focA::pflB*; *ldhA*; *adhE*; and *frdA* in order to limit potential by-product formation. To prevent xylose catabolism we deleted *xylAB*. We furthermore

replaced the native cAMP receptor protein (CRP) with a mutated version (CRP*), to enable simultaneous uptake of different sugars [12][13], and inserted a constitutively expressed xylose reductase into the genome. The generated strain can only oxidize glucose generated NAD(P)H by converting xylose to xylitol under anaerobic conditions. Because xylose uptake is energy requiring, the strain furthermore produces acetate to maintain redox balance under anaerobic conditions. To enable fermentative physiology under aerobic conditions [14] we first deleted *cyoB* and *appB*, which are part of cytochromes BD-o and BD-II, respectively. We subsequently integrated dCas9 into the genome under the control of an anhydrotetracycline inducible promoter, and constructed a plasmid with a guide RNA (gRNA) against *cydA* to facilitate inducible repression of cytochrome BD-I.

In order to evaluate if the inducible CRISPRi-mediated metabolic switch worked as intended, we performed a comparative growth and production experiment with CRISPRi induced or uninduced in both minimal and richer media (minimal media with 0.5% yeast extract). The repression of *cydA* was surprisingly effective and resulted in growth arrest after about only one doubling in cell density in minimal media and two doublings in cell density in richer media. The growth attenuation further remained stable for >48 hours (Figure 1A; Supplementary Fig. 1). The theoretical maximum molar yield of xylitol from glucose (oxidation of glucose generated NADP(H)) is ~4 xylitol per glucose in analogous growth arrested cells under anaerobic conditions [15]. Under aerobic conditions with anaerobic physiology we achieved a molar yield of xylitol from glucose of 3.5 (± 0.76) when CRISPRi was active in minimal media (Supplementary Fig. 2). The molar yield of xylitol from glucose in richer media decreased to ~2.3 xylitol per glucose (Figure 2B), which is still significantly higher than the molar yield for uninduced and thus respiring cells which were ~1.8 and ~1.7 in minimal and richer media, respectively.

3.2.2 Construction of an acetic acid auxotroph through constraint-based metabolic modelling

Acetate can act as a strong inhibitor of growth even at relatively low concentrations [16][17] and is typically not utilized by microorganisms, when sugars are available. To construct a compatible aerobic partner for the consortia we therefore focused our efforts on the construction of a strain that would co-utilize glucose and the secreted secondary product acetic acid. We first made three knockouts (*aceEF*, *focA-pflB*, and *poxB*), and additionally used constraint-based metabolic modelling to

identify other possible sources of acetate and acetyl-coA during growth on glucose. Three oxygen sensitive genes: *tdcE* (an oxygen sensitive pyruvate formate lyase), *pflDC* (pyruvate formate-lyase II), and pfo (an oxygen sensitive pyruvate:flavodoxin oxidoreductase), as well as *deoC* (a deoxyribose-phosphate aldolase; involved in purine and pyrimidine degradation) were identified as genes coding for proteins involved in pathways that, in addition to the deletions in the initial strain, could lead to sugar derived acetyl-coA formation. To prevent the predicted pathways from potentially evolving to be able to supply the cell with acetyl-CoA during long term experiments, these genes were therefore also deleted. A possible flux through serine or threonine that could theoretically lead to acetyl-CoA formation from the different sugars was also identified by constraint-based metabolic modelling. These two routes, shown in detail in Supplementary Figure 2, were deemed highly unlikely to provide the cells with sufficient acetyl-coA to sustain growth, and deletions to prevent these two theoretical routes were therefore not performed. To construct a glucose utilizing acetic acid auxotroph incapable of catabolizing C5-sugars, we further deleted *xylAB* and *araBA*, which encodes for proteins involved in xylose and arabinose catabolism, respectively. The resulting strain did not grow in minimal M9 media when supplemented solely with glucose. It displayed some growth when supplemented with acetic acid; however, it grew rapidly when supplemented with both glucose and acetic acid (Supplementary Figure 3). To evaluate the performance of our acetic acid auxotroph in regards to isobutyric acid production, we first did a comparison of the original strain and a strain with *ptsG* deleted. The performance of the two strains were not significantly different, but as there was a slight tendency of the Δ*ptsG* to perform better (lower total OD, higher IBA titers and yield) (Supplementary Figure 4A), our subsequent experiments were done using this strain. During our initial experiments we found that acetate was depleted after 13 h of growth (Figure 2C). We therefore tested whether an increase in acetic acid concentration from 17 mM (1 g L^{-1}) to 34 mM (2 g L^{-1}) had an impact on growth and IBA production (Supplementary Figure 4B). No differences were observed in growth and IBA production and subsequent experiments were therefore conducted using 17 mM (1 g L^{-1}) acetate.

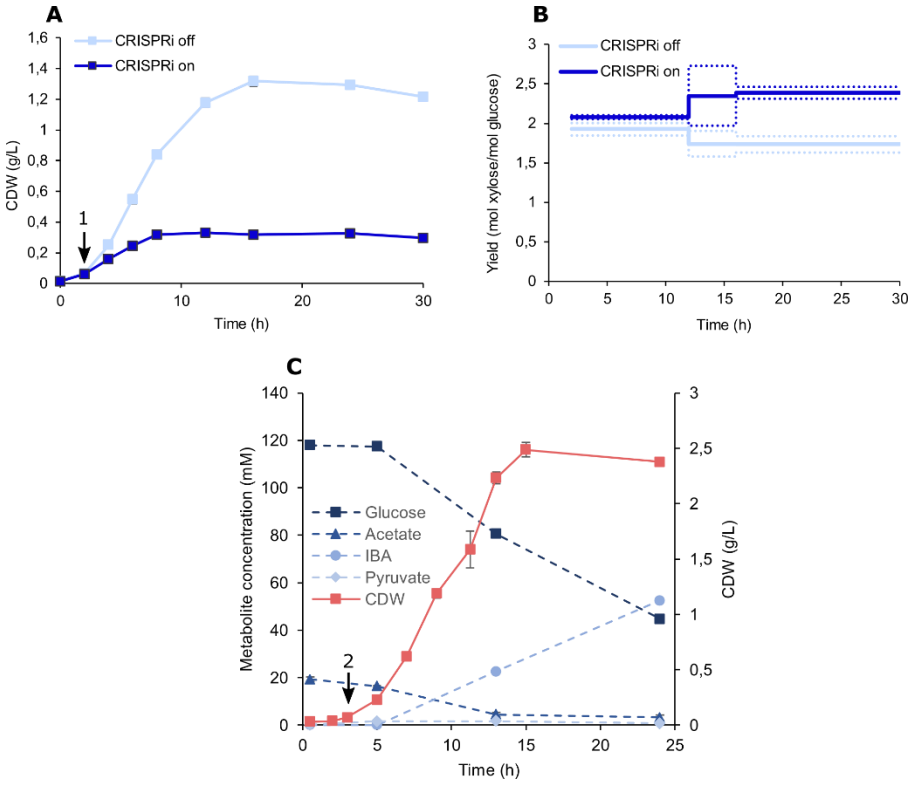


Figure 2 *Characterization of the individual strains. A) Growth curve of the xylitol strain with (dark blue) or without (light blue) the CRISPRi switch induced, standard deviation plotted as dashed lines. B) Molar yield of xylitol per glucose with and without the CRISPRi switch. C) Growth and metabolite levels of the IBA strain. 1: Induction of CRISPRi switch, 2: induction of IBA production. Error bars indicate mean ± s.d. (n=4 biological replicates for A and B, n=3 biological replicates for C).*

3.2.3 Compatibility of consortia partners

In order to determine if the presence of products from the other strain would have an effect on the growth and productivity in a consortia based process we performed single strain fermentation experiments with addition of the possibly inhibitory compounds. Xylitol productivity was not affected by the presence of isobutyric acid, however, the presence of 2.5 g L^{-1} of IBA decreased the growth rate of the xylitol strain (Supplementary Fig 4). To mitigate this effect during our consortia based experiments (see below), we therefore timed the induction of the CRISPRi-enabled switch so that we could obtain cell-arrest before the IBA concentration reached inhibitory levels (Supplementary Fig 4). The presence of anhydrous tetracycline (1 µg mL^{-1}) had no effect on the IBA producing strain, whereas the presence of xylose

(20 g L^{-1}) or xylitol (10 g L^{-1}) resulted in a slightly lower final cell density. IBA production was not affected by any of the three compounds (Supplementary Fig. 5).

3.2.4 Consortia based production

In order to enable tracking of the distribution of the two cell lines when grown together, we first integrated constitutively expressed mCherry and Yellow fluorescent protein (YFP) into the genomes of the xylitol and isobutyric acid producing strains, respectively. We subsequently performed some initial production experiments and found that the time of induction of the IBA strain had a significant impact on the overall performance of the consortia based production. During our initial experiments with the IBA strain (Figure 2C), production was induced after 3 hours of growth. However, during our consortia based optimization experiments we found that IBA titers were significantly higher when inducing after 5 hours as compared to induction after 3 and 4 hours (Supplementary Figure 7B). Induction after 5 hours of growth was therefore chosen for our final production experiment. Delaying induction of the CRISPRi system in the xylitol strain for 0.5 hours and inducing at 2.5 h instead of 2 h as was done in previous experiments did result in a slightly higher OD/proportion of the strain and consequently slightly higher xylitol titer (Supp. Figure 7A). However, after normalizing for the increase in OD, xylitol production did not differ significantly from the induction after 2-hours. We therefore continued to use the 2-hour induction time to enable comparisons with previous experiments.

Finally, both strains were cultivated together using these optimized induction conditions. The xylitol strain (initially inoculated in even proportions to the isobutyric acid strain) grew faster than the isobutyric acid strain, until growth was attenuated by the CRISPRi induced metabolic switch. The continuously growing isobutyric acid strain subsequently increased in proportion and after the growth phase, the distribution was about 80% IBA strain and 20% xylitol strain, after which it remained stable to the end of the experiment (Figure 3A). Acetate concentration initially increased in the beginning of the experiment, reaching a maximum concentration of 21.8 mM after 6 h of growth, after which it was fully consumed by the IBA strain. A small accumulation of pyruvate (5.5 mM) was further detected after 4 h of growth, but was not detectable after 8 h (Figure 3B). The final titer of IBA and xylitol was 47 mM and 23 mM, respectively, which was comparable to titers achieved using single strain inoculations, with IBA and xylitol titers of 52 mM and 22

mM, respectively. When comparing the specific productivity of the strains in the consortia-process to the respective single strain-processes (Figure 3C), we observed an increase in xylitol productivity, going from 2.16 to 3.1 mM xylitol / g L^{-1} CDW / h at 24 hours, whereas IBA productivity only decreased from 0.98 to 0.89 (± 0.08) mM IBA / g L^{-1} CDW / h at 24 hours.

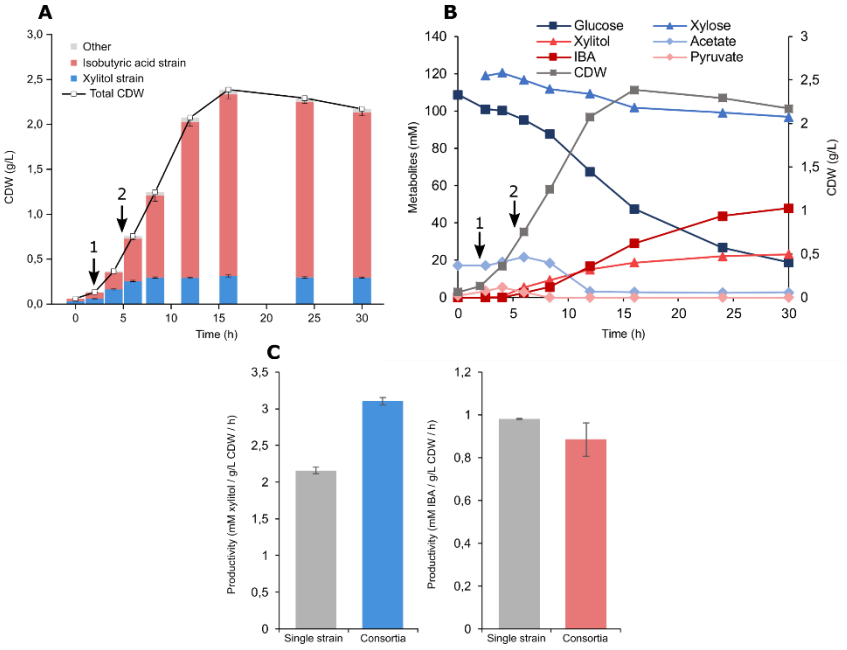


Figure 3 *Characterization of the anaerobic-aerobic consortia process. A) Distribution between the strains during the cultivation. 1: Induction of CRISPRi switch, 2: induction of IBA production B) Substrate, product and byproduct concentrations during the experiment. C) The specific productivity of the consortia-process compared to the single strain cultivations. Error bars indicate mean ± s.d. (n=4 biological replicates).*

3.3 Discussion

Microbial cell factories that can produce value added fuels and chemical building blocks have received extensive attention in order to enable more sustainable manufacturing processes [18]. Most microbial manufacturing processes are being pursued as single strain fermentations, however, in the last decade a number of consortia based approaches have been developed [2-6][8][9]. An important parameter for synthetic multi-strain bioprocess developments is population control and stability. Here we used a CRISPRi mediated metabolic switch to enable a growth

decoupled “anaerobic” production process and coupled it with a mutualistic syntrophic consortia partner that were designed to co-consume acetic acid and glucose. Bioprocesses (both single and multi-strain) are traditionally run with either aerobic, microaerobic or anaerobic conditions, pending on the organism and the product and biosynthetic pathway in question. In this study we designed an innovative consortia based process that enables both anaerobic and aerobic physiology in the same bioreactor. By reducing the oxygen requirements of one of the strains to the growth phase, we were able to concomitantly run an aerobic fermentation process in the same reactor without significant loss in titers and yields as compared to when the fermentations were run with single strains in two reactors. This highlights the potential of this process to increase efficiency of resource utilization and to lower bioreactor capacity requirements.

3.4 Acknowledgements

This study was funded by the Novo Nordisk Foundation and the Independent Research Foundation Denmark (Grant no. 7017-00321B). This work was also funded by the Novo Nordisk Foundation within the framework of the Fermentation-based Biomanufacturing Initiative (FBM), grant number: NNF17SA0031362.

3.5 Materials and methods

3.5.1 Strains, plasmids and media

All strains were routinely grown in LB broth or on LB agar plates supplemented, when needed, with appropriate antibiotics (ampicillin 100 µg mL^{-1} (ap), kanamycin 30-50 µg mL^{-1} (km), chloramphenicol 10-25 µg mL^{-1} (cm)). Cells were grown at either 30 °C or 37 °C. 100 µL liquid LB-amp supplemented with 1 g L^{-1} sodium acetate (NaOAc) was spread on LB-km or LB-cm plates for plasmid maintenance during the construction of the deletion strains, when using pSIJ8 [19]. Strains and plasmids used in this study are listed in Supplementary Table 1 and 2, and oligoes are listed in Supplementary Table S3.

3.5.2 Deletion and insertion of genes

PCR reactions were performed using standard PCR conditions. PCR templates for gene deletions were generated by amplifying gDNA extracted from previously generated single deletion strains or by using extended oligoes with 50 bp homology arms. Amplified FRT flanked antibiotic cassettes were used to delete native *E. coli* genes by the combined action of lambda Red recombineering and flippase

recombinase using previously described methods for the combined procedure for using pkd46 [20] and pcp20 [21] or for the temperature sensitive plasmid pSIJ8 [19]. Constitutively expressed fluorescent proteins (YFP and mCherry) were PCR amplified from plasmids and ligated together with an FRT-flanked kanamycin cassette by USER-ligation-PCR. The spliced cassettes were subsequently integrated into the genomes of the specific strains (9 base pairs downstream of the glmS gene) after which the integrated FRT-flanked kanamycin cassettes was removed.

3.5.3 Constraint based metabolic modeling

Cameo (Computer Aided Metabolic Engineering and Optimization) and Escher were used for simulation and visualization of metabolic models [22][23]. Gene deletions necessary for complete acetic acid auxotrophy were determined through testing of the designs with the *E. coli* metabolic model, iJO136649 [24], with a minor modification: Tryptophanase reaction catalyzed by *tnaA* gene was considered irreversible under physiological conditions. The knockouts of the initial design were projected to iJO1366 reactions as follows: aceEF/PDH; focA/FORt2pp; pflB/PFL and poxB/POX. Further deletions necessary to ensure acetic acid auxotrophy were determined by examining the flux profiles generated for utilized pathways leading to acetate/acetyl-CoA and iteratively eliminating them.

3.5.4 Shake flask cultivations of engineered *E. coli* strains

Cultivation experiments was performed with modified M9 minimal media called M9extra (6.8 g L^{-1} of Na_2HPO_4, 3 g L^{-1} of KH_2PO_4, 0.5 g L^{-1} of NaCl, 1 g L^{-1} NH_4Cl, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$) supplemented with 60 µM $FeCl_3$, a trace element solution (1.25 µM $MnCl_2{\cdot}4H_2O$, 0.21 µM $CoCl_2{\cdot}6H_2O$, 0.85 µM $ZnSO_4{\cdot}7H_2O$, 0.05 µM $CuCl_2{\cdot}2H_2O$, 0,08 µM H_3Bo_3, 0.105 µM $NiCl_2{\cdot}6H_2O$, 0.125 µM $NaMoO_4{\cdot}2H_2O$). For cultivations that included yeast extract, 5 g L^{-1} was added. For cultivations that included the IBA strain, 1 g L^{-1} of sodium acetate was added. A colony was picked off an agar-plate and cultivated in 50 mL of the M9 media in a 250 mL shake flask with 111 mM (20 g L^{-1}) glucose added. The cells were spun down at 5000 G for 5 minutes, and resuspended in M9 media. 250 mL shake flasks with 50 mL of media were inoculated to OD 0.1 (or 0.05 in experiments with the xylitol strain with yeast extract). Experiments with both strains were inoculated to OD 0.1 of each strain. Glucose was added in the following amounts: xylitol strain without yeast extract, 30 mM (5.4 g L^{-1}); xylitol strain with yeast extract, 60 mM (10.8 g L^{-1}); IBA strain, 111 mM (20 g L^{-1}); consortia, 111 mM (20 g L^{-1}). 133 mM (20 g L^{-1}) xylose was added to

cultivations with the xylitol strain present. The cultures were grown at 37 °C and 250 RPM shaking until induction, after which the temperature was lowered to 30 °C. The CRISPRi switch was induced after 2 hours with 1 µg mL^{-1} anhydrotetracycline, and IBA production was induced after 5 hours with 0.1 M IPTG. If both were induced, the temperature was lowered after induction of IBA production. Samples were taken for OD600, extracellular metabolites and measuring the distribution of the strains in 2 to 4 hour intervals during growth phase and 6 to 12 hour intervals after the growth phase.

3.5.5 Quantification of cell density

The cell densities were measured during the cultivations using optical density at a wavelength of 600 nm. Conversion of the optical cell density measurements at 600nm to equivalent cell dry weight in g L^{-1}, a conversion factor of 1 OD600 = 0.31 g L^{-1} CDW was used, which is well established in literature [25]-[27].

3.5.6 Flow cytometry

900 µL sample was spun down at 6500g for 5 min. After collection of the supernatant for extracellular metabolite quantification (see below), the pellet was fixed by re-suspending in 2% paraformaldehyde in 1x phosphate-buffered saline (PBS) and incubated for a minimum of 30 minutes. After fixation, the cells were spun down at 6500g for 5 minutes, re-suspended in 1x PBS and stored at 4 °C until further analysis, which was performed within two weeks. The relative abundance of the cells in the consortia was determined by flow cytometry using a MacsQUANT VYB cytometer (Miltenyi Biotec, Germany). Gating in the forward scatter and side scatter channels was fine-tuned to ensure that single cells were obtained, and that all cells in a given sample were represented in the measurement. mCherry fluorescence was measured with a 561 nm laser; 615/20 nm band-pass filter, whereas YFP fluorescence was measured with a 488 nm laser; 525/50 band-pass filter. Gating of the channels was performed using single strain controls, and pre-determined mixtures of the strains were used to ensure that cells were not present in both of the gatings, and that the number of cells in the two gates corresponded to the total number of cells counted.

3.5.7 Quantification of extracellular metabolites

900 µL of culture was spun down at 6500 G for 5 minutes, and the supernatant was collected and stored at -20 °C until further analysis. The thawed supernatant was then analyzed by high performance liquid chromatography (HPLC) using an Ultimate

3000 HPLC equipped with a Biorad Aminex HPX-87H column. Samples (run for 33 min.) were analyzed using 5 mM sulfuric acid as mobile phase at a flow rate of 0.6 mL min^{-1}, and a column temperature of 30 °C. Glucose, xylose, xylitol, acetic acid, pyruvic acid, and isobutyric acid were estimated using a refractive index detector (RI) and/or a tunable absorbance detector set at 210 nm.

3.6 References for Chapter 3

[1] "UN General Assembly, “Transforming our world: the 2030 Agenda for Sustainable Development,”" 2015.

[2] K. Zhou, K. Qiao, S. Edgar, and G. Stephanopoulos, "Distributing a metabolic pathway among a microbial consortium enhances production of natural products," *Nat Biotechnol,* vol. 33, pp. 377-83, Apr 2015.

[3] H. Zhang, B. Pereira, Z. Li, and G. Stephanopoulos, "Engineering Escherichia coli coculture systems for the production of biochemical products," *Proc Natl Acad Sci U S A,* vol. 112, pp. 8266-71, Jul 7 2015.

[4] R. L. Shahab, J. S. Luterbacher, S. Brethauer, and M. H. Studer, "Consolidated bioprocessing of lignocellulosic biomass to lactic acid by a synthetic fungal-bacterial consortium," *Biotechnol Bioeng,* vol. 115, pp. 1207-1215, May 2018.

[5] J. J. Minty, M. E. Singer, S. A. Scholz, C. H. Bae, J. H. Ahn, C. E. Foster*, et al.*, "Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass," *Proc Natl Acad Sci U S A,* vol. 110, pp. 14592-7, Sep 3 2013.

[6] T. R. Zuroff, S. B. Xiques, and W. R. Curtis, "Consortia-mediated bioprocessing of cellulose to ethanol with a symbiotic Clostridium phytofermentans/yeast co-culture," *Biotechnol Biofuels,* vol. 6, p. 59, Apr 29 2013.

[7] M. A. Eiteman, S. A. Lee, and E. Altman, "A co-fermentation strategy to consume sugar mixtures effectively," *J Biol Eng,* vol. 2, p. 3, Feb 27 2008.

[8] H. C. Bernstein, S. D. Paulson, and R. P. Carlson, "Synthetic Escherichia coli consortia engineered for syntrophy demonstrate enhanced biomass productivity," *J Biotechnol,* vol. 157, pp. 159-66, Jan 2012.

[9] T. Xia, M. A. Eiteman, and E. Altman, "Simultaneous utilization of glucose, xylose and arabinose in the presence of acetate by a consortium of Escherichia coli strains," *Microb Cell Fact,* vol. 11, p. 77, Jun 12 2012.

[10] M. Chen and L. S. Qi, "Repurposing CRISPR System for Transcriptional Activation," *Adv Exp Med Biol,* vol. 983, pp. 147-157, 2017.

[11] S. Li, C. B. Jendresen, A. Grunberger, C. Ronda, S. I. Jensen, S. Noack*, et al.*, "Enhanced protein and biochemical production using CRISPRi-based growth switches," *Metab Eng,* vol. 38, pp. 274-284, Nov 2016.

[12] R. Khankal, J. W. Chin, D. Ghosh, and P. C. Cirino, "Transcriptional effects of CRP* expression in Escherichia coli," *J Biol Eng,* vol. 3, p. 13, Aug 24 2009.

[13] J. G. M. Harman, K.; Peterkofsky, A., "Structure-function analysis of three cAMP-independent forms of the cAMP receptor protein," *JBC,* vol. 261, pp. 16332-16339, 1986.

[14] V. A. Portnoy, M. J. Herrgard, and B. O. Palsson, "Aerobic fermentation of D-glucose by an evolved cytochrome oxidase-deficient Escherichia coli strain," *Appl Environ Microbiol,* vol. 74, pp. 7561-9, Dec 2008.

[15] O. Akinterinwa and P. C. Cirino, "Anaerobic obligatory xylitol production in Escherichia coli strains devoid of native fermentation pathways," *Appl Environ Microbiol,* vol. 77, pp. 706-9, Jan 2011.

[16] G. W. Luli and W. R. Strohl, "Comparison of growth, acetate production, and acetate inhibition of Escherichia coli strains in batch and fed-batch fermentations," *Appl Environ Microbiol,* vol. 56, pp. 1004-11, Apr 1990.

[17] R. M. Lennen and M. J. Herrgard, "Combinatorial strategies for improving multiple-stress resistance in industrially relevant Escherichia coli strains," *Appl Environ Microbiol,* vol. 80, pp. 6223-42, Oct 2014.

[18] R. A. Sheldon and J. M. Woodley, "Role of Biocatalysis in Sustainable Chemistry," *Chem Rev,* vol. 118, pp. 801-838, Jan 24 2018.

[19] S. I. Jensen, R. M. Lennen, M. J. Herrgard, and A. T. Nielsen, "Seven gene deletions in seven days: Fast generation of Escherichia coli strains tolerant to acetate and osmotic stress," *Sci Rep,* vol. 5, p. 17874, Dec 8 2015.

[20] K. A. Datsenko and B. L. Wanner, "One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products," *Proc Natl Acad Sci U S A,* vol. 97, pp. 6640-5, Jun 6 2000.

[21] P. P. W. Cherepanov, W, "Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant," *Gene,* vol. 158, pp. 9-14, 1995.

[22] N. C. Sonnenschein, J. G. R.; Özdemir, E., "cameo: Improved cameo.api.," *doi:105281/zenodo19827*.

[23] Z. A. King, A. Drager, A. Ebrahim, N. Sonnenschein, N. E. Lewis, and B. O. Palsson, "Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways," *PLoS Comput Biol,* vol. 11, p. e1004321, Aug 2015.

[24] J. D. Orth, T. M. Conrad, J. Na, J. A. Lerman, H. Nam, A. M. Feist*, et al.*, "A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011," *Mol Syst Biol,* vol. 7, p. 535, Oct 11 2011.

[25] J. Soini, K. Ukkonen, and P. Neubauer, "High cell density media for Escherichia coli are generally designed for aerobic cultivations - consequences for large-scale bioprocesses and shake flask cultures," *Microb Cell Fact,* vol. 7, p. 26, Aug 7 2008.

[26] Y. Wu, P. Yan, Y. Li, X. Liu, Z. Wang, T. Chen*, et al.*, "Enhancing beta-Carotene Production in Escherichia coli by Perturbing Central Carbon Metabolism and Improving the NADPH Supply," *Front Bioeng Biotechnol,* vol. 8, p. 585, 2020.

[27] C. P. Long, J. E. Gonzalez, N. R. Sandoval, and M. R. Antoniewicz, "Characterization of physiological responses to 22 gene knockouts in Escherichia coli central carbon metabolism," *Metab Eng,* vol. 37, pp. 102-113, Sep 2016.

3.7 Supplementary materials

3.7.1 Supplementary Table S1 - Strains used in this study

Strain	Genotype	Reference
E. coli K-12 MG1655	F- λ - ilvG- rfb-50 rph-1	Strain collection
SIJ-216-dCas9	*E. coli* K-12 MG1655 Δ*focA-pflB*; Δ*ldhA*; Δ*adhE::*$P_{BBJ23100}$*-CBXR66*; Δ*frdA;* Δ*xylAB;* Δ*cyoB;* Δ*appB;* Δ*fhuA;* CRP*; attB186(O)::P_{TetA}*-dCas9*; Tn7::$P_{BBJ23100}$*-mCherry*	This study
SIJ-1247-3G	*E. coli* K-12 MG1655 Δ*focA-pflB*; Δ*aceEF*; Δ*poxB*; Δ*tdcE*; Δ*pflDC*; Δ*deoC*; Δ*ydbK*; Δ*yqhD*; Δ*araBA*; Δ*xylAB*; Tn7::$P_{BBJ23100}$*-yfp*	This study
SIJ-1247-3G-ptsG	1247-3G Δ*ptsG*	This study

3.7.2 Supplementary Table S2 - Plasmids used in this study

Plasmid	Features	Reference
pKD46	P_{BAD} λ Red, ori pSC101, Amp^r, Rep^{ts}	[1]
pCP20	FLP recombinase, ori pSC101, λ $cI857^+$ λ P_r Rep^{ts}, Amp^r, Cm^r	[2]
pKD3	Amp^r, FRT-Cm^r-FRT, oriR6K	[1]
pKD4	Amp^r, FRT-Km^r-FRT, oriR6K	[1]
pSIJ8	pKD46, rhaRS-P_{rha}-FLP, Amp^r	[3]
pOSIP-KO	Clonetegration plasmid;.attB::186; Km^r	[4]
pdCas9	ori p15A; P_{tetA}-dCas9, Cm^r	[5]
pSLQ1236	ori ColE1, P_{tetA}-sgRNA-RFP; Amp^r	[6]
pSIJ11	pSLQ1236-sgRNA-cydA1	This study
pIBA1	ori p15A, lacI-P_{Lac}-ilvD-alsS, Km^r	[7]
pIBA7	ori colE1, lacI-P_{Lac}-kivd-ydcW, Amp^r	[7]

3.7.3 Supplementary Table S3 - Primers used in this study

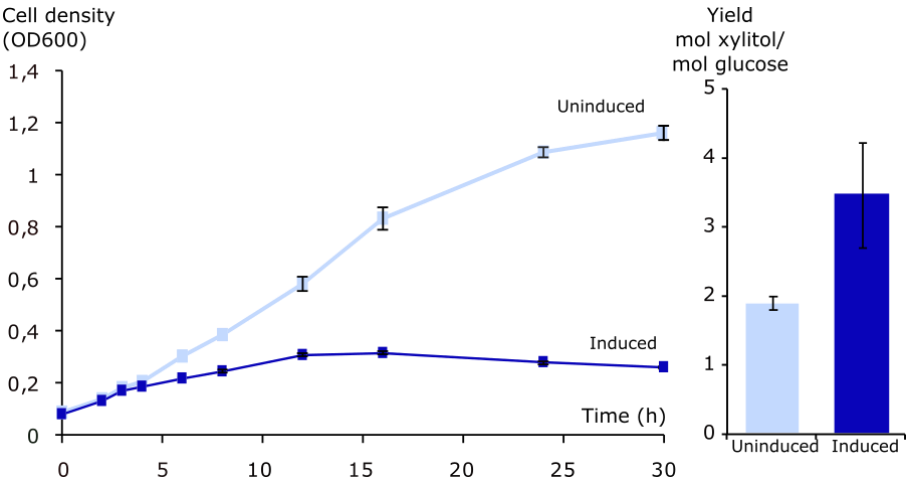
Primer name	Sequence
ldhA_knock_fw	TATTTTTAGTAGCTTAAATGTGATTCAACATCACTGGAGAAAGT CTTATGGTGTAGGCTGGAGCTGCTTC
ldhA_knock_rv	CTCCCCTGGAATGCAGGGGAGCGGCAAGATTAAACCAGTTCGT TCGGGCACATATGAATATCCTCCTTAG
adhE_knock_fw	CGAGCAGATGATTTACTAAAAAAGTTTAACATTATCAGGAGAG CATTATGGTGTAGGCTGGAGCTGCTTC
adhE_knock_rv	CCGTTTATGTTGCCAGACAGCGCTACTGATTAAGCGGATTTTTT CGCTTTCATATGAATATCCTCCTTAG

frdA_knock_fw	ACCCTGAAGTACGTGGCTGTGGGATAAAAACAATCTGGAGGAA TGTCGTGGTGTAGGCTGGAGCTGCTTC
frdA_knock_rv	GCACCACCTCAATTTTCAGGTTTTTCATCTCAGCCATTCGCCTTC TCCTTCATATGAATATCCTCCTTAG
cyoB_knock_fw	TGAGCCACGCGGAATCCGCCCATTAAAGGGGTTGAGGAAGAA TAAAGATGGTGTAGGCTGGAGCTGCTTC
cyoB_knock_rv	GCGCGTGGGCAGTCGCGTGCGTCAAAGTATCAGTTGCCATTTT TCAGCCCCATATGAATATCCTCCTTAG
appB_knock_fw	ATGCAGAGTGAACAACCGACGCAGCAACAGGGGTAAAGGAGA AAATCATGGTGTAGGCTGGAGCTGCTTC
appB_knock_rv	CAAAGTAAATACCACATTGTTTCTGCTCCTTAGTACAACTCGTTT TCGTTCATATGAATATCCTCCTTAG
xylAB_knock_fw	ACGACATCATCCATCACCCGCGGCATTACCTGATTATGGAGTTC AATATGGTGTAGGCTGGAGCTGCTTC
xylAB_knock_rv	CCCCCACCCGGTCAGGCAGGGGATAACGTTTACGCCATTAATG GCAGAAGCATATGAATATCCTCCTTAG
focA-pflB_knock_fw	ATGCTTTGTTAGTATCTCGTCGCCGACTTAATAAAGAGAGAGTT AGTGTGGTGTAGGCTGGAGCTGCTTC
focA-pflB_knock_rv	TTTTACTGTACGATTTCAGTCAAATCTAATTACATAGATTGAGTG AAGGTCATATGAATATCCTCCTTAG
aceEF_knock_fw	ACAGGTTCCAGAAAACTCAACGTTATTAGATAGATAAGGAATA ACCCATGGTGTAGGCTGGAGCTGCTTC
aceEF_knock_rv	AAAGCCGGCCGTTGGGCCGGCTCTTTTACTTACATCACCAGACG GCGAATCATATGAATATCCTCCTTAGTTCC
poxB_ knock_fw	GATGAACTAAACTTGTTACCGTTATCACATTCAGGAGATGGAG AACCATGGTGTAGGCTGGAGCTGCTTC
poxB_ knock_rv	CCTTATTATGACGGGAAATGCCACCCTTTTTACCTTAGCCAGTTT GTTTTCATATGAATATCCTCCTTAGTTCC
deoC_knock_fw	ATCGAAGTGTGTTGCGGAGT
deoC_knock_rv	GTCTCGTAGGCCTGATAAGC
tdcE_knock_fw	TTAGGCAAAGTTAACGCGCC
tdcE_knock_rv	CCTGAACATAAGGGCCGATTG
pflDC_knock_fw	CTTCTCCCGCTCGCAAGGGCGGGTTCGCTTTCCCACAGGAGTTC CTCATG TGTAGGCTGGAGCTGCTTC
pflDC_knock_rv	AGGCGGTTACTGCCACCAGGTATGCCATTTTAACCTCCCACGGT AACCTGGCTTGCATATGAATATCCTCCTTAG
ydbK_knock_fw	ATTACTACTGGGCGCGCA
ydbK_knock_rv	GAACAGTCGTCCGGGCTT
yqhD_knock_fw	ACATCAGGCAGATCGTTCTCT
yqhD_knock_rv	GTTGTGAACTTAAGTCTGGACGA
ptsG_knock-fw	CTTGCCACGCGTGAGAAC
ptsG_knock-rv	CGTATCAATTCTGAATAACACCTGTA

araBA-knock-fw	ACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGATGGAGTGAAA CGATGGTGTAGGCTGGAGCTGCTTC
araBA-knock-rv	ATCAGGCGTTACATACCGGATGCGGCTACTTAGCGACGAAACC CGTAATAGCTTGCATATGAATATCCTCCTTAG
crp*_flank_fw	TCGAATGGTTCTTGTCTCATTGCCACATTCATAAGTACCCATCCA AGAGCAAGCTTATTCACCAGGGTGAAAAAGCG
crp*_flank_rv	TCCGGGTGAGTCATAGCGTCTGGTTGTTTTGCCAGATTCAGCAG AGTCTGTGTAATGCGGCCCGTCACG
crp*_fw	AAGCTTATTCACCAGGGTGAAAAAGCG
crp*_rv1	GTCCGGGTTTACCTGAATCAATTGG
crp*_rv2	TGTAATGCGGCCCGTCACGTCGAGGAACGCCAGGTTGCCCACT TTCTCTGAAATGACTTGCAGACGACGCGCCATCTGTGCAGACAA ACGCATCAGCAGGTCCGGGTTTACCTGAATCAATTGG
Tn7-ins-fw	CTTACCATGTCGCGCTGATC
Tn7-ins-rv	ATGACGGTTTGTCACATGGAGT
BBa_J23100_XR-fw	TACTAGAGTTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGC TACTAGAGATTAAAGAGGAGAAATACTAGATGTCAAGCCCACT TTTAAC
XR_rv	TTAAATAAATGTTGGAATATTGTAACC
adhE_XR-ins_fw	AGCGATGCTGAAAGGTGTCAGCTTTGCAAAAATTTGATTTGGA TCACGTATACTAGAGTTGACGGCTAGCTCAGT
adhE_XR_ins_rv	ATCGGCATTGCCCAGAAGGGGCCGTTTATGTTGCCAGACAGCG CTACTGATTAAATAAATGTTGGAATATTGTAACCC
ldhA_check_fw	AGTAATAACAGCGCGAGAACG
ldhA_check_rv	TTTCTGGCGGATTTTTATCG
adhE_check_fw	AATCTTGCTTACGCCACCTG
adhE_check_rv	AGAAAGCGTCAGGCAGTGTT
frdA_check_fw	ACGGCGAGACAAATTTTACG
frdA_check_rv	ACGCTTCAACCTTCATACCG
ptsG_check_fw	CGTCAAACAAATTGGCACTG
ptsG_check_rv	TCTCGTTACAGGGGAACGTC
cyoB_check_fw	CGCCTAGCGAATACAACCAG
cyoB_check_rv	TGATGGAGCTGAACAACAGC
appB_check_fw	GTGGGCGATACAGGACATCT
appB_check_rv	CTCACTCTGAGCGAATGCAG
xylAB_check_fw	GCAACTAAACAGGGGAAAACA
xylAB_check_rv	GTTGCCAAAAGTTGCTGTCA
focA-pflB_check_fw	TTAATGCCCCGCTTTACATA
focA-pflB_check_rv	GCGATAGGTCACCACTTCCT
aceEF_check_fw	TGTGGTTCTGCTTCATCTGC

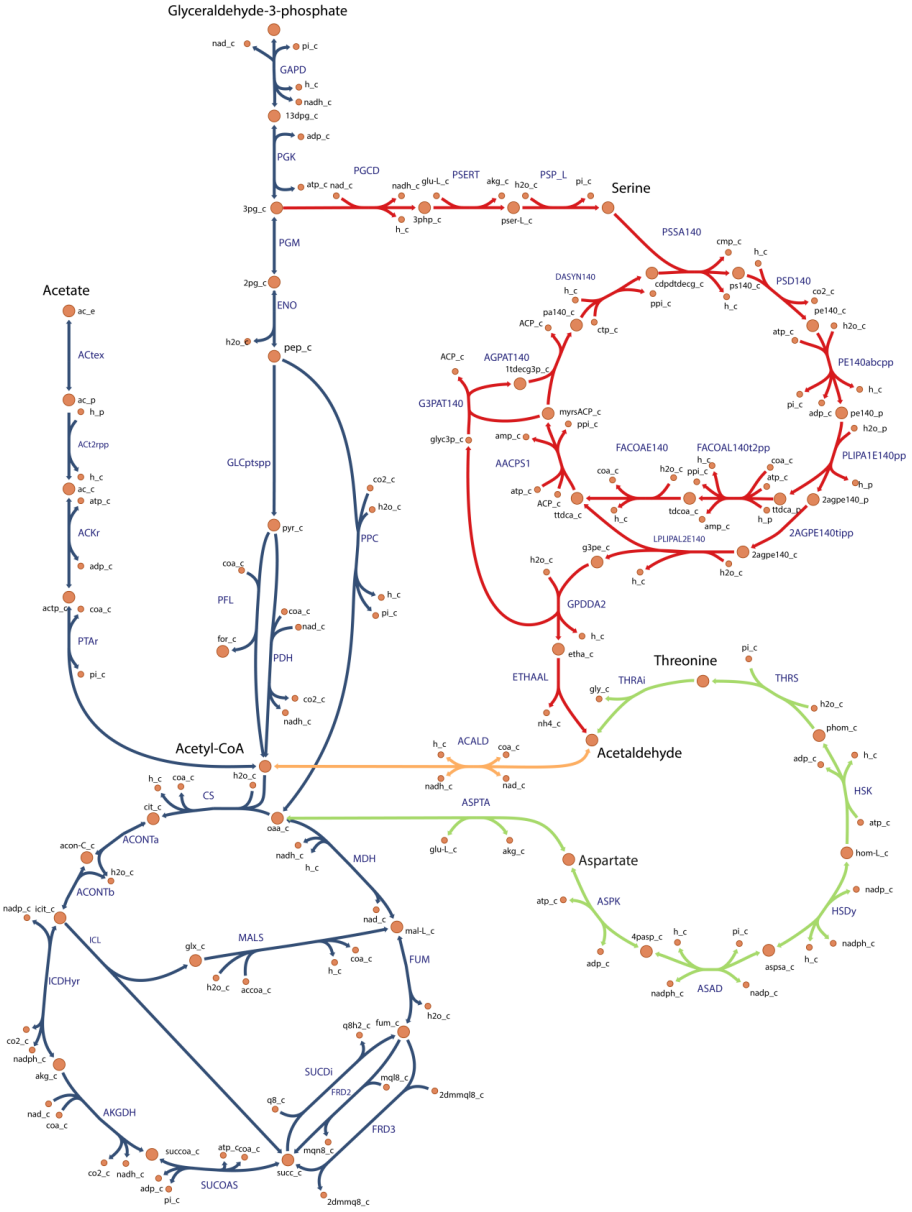
aceEF_check_rv	AGACCTAAATCAGCGCAACG
poxB-check-fw	ATGGATTGGGTAGAGCAGGA
poxB-check-rv	TGGATATCGTCGGGTTTGAT
deoC_check_fw	TGATGCGCCATTGTAAGAAGT
deoC_check_rv	GGAAGAAAATGGTCATCGCGA
tdcE_check_fw	ATTCGTCGTCTGGTCATGGA
tdcE_check_rv	AGCAACCACGATCGCTTTG
pflDC_check_fw	AGGCTGGGAAGTGAAAGTAGA
pflDC_check_rv	GCCCTGAGTTTCAATGCTGA
ydbK_check_fw	CGTCGAGAGCTGTCCTGC
ydbK_check_rv	ATTCTGCGCGGGTTATATGC
yqhD_check_fw	AACTTTATCCGCCAGCAAGC
yqhD_check_rv	TCGAGCGATAACCCACTTCT
araBA-check-fw	GCTCTTCTCGCTAACCCAAC
araBA-check-rv	CGGTCATGACGCTGTAATCG
Tn7_check_fw	GGCGGTCAGTTGTATGTCTTC
Tn7_check_rv	AGGACAAACAGGTGACAGTTATATG
adhE-ins_check_fw	AAAGACGCGCTGACAATACG
attB::186-check1-fw	AGGTGAATGGGAACGCATAA
attB::186-check1-rv	GTGCCACGTACCAAAACACC
attB::186-check2-fw	TGACGAACTACGCCATGAAC
attB::186-check2-rv	CGTTGGACGGTCTACCAGTT
dCas9_seq1	TTCTTCTTGGCTAGCTCCCC
dCas9_seq2	TTGGCGCAAATTGGAGATCA
dCas9_seq3	AACTATCATCATGGATCAGCTGC
dCas9_seq4	CGTCGCCGTTATACTGGTTG
dCas9_seq5	TCTCGCCCTTTATCCCAGAC
dCas9_seq6	TGAGCAAGAAATAGGCAAAGCA

3.7.4 Supplementary Figure 1



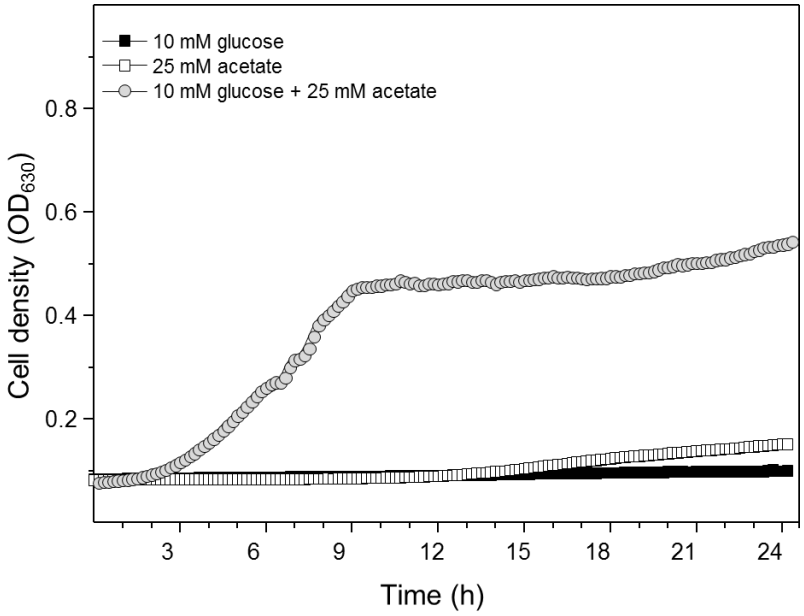
Supplementary Figure 1 *Growth curve and product molar yield per substrate for the CRISPRi anaerobic switchable strain producing xylitol (SIJ216), using M9 Extra media supplemented with 28 mM (5 g L^{-1}) glucose. CRISPRi was induced and 107 mM (16 g L^{-1}) xylose were added after 3 hours of growth, indicated by the arrow. Yield was calculated for the interval between 16 and 24 hours. Error bars indicate mean ± s.d. (n=4 biological replicates).*

3.7.5 Supplementary Figure 2



Supplementary Figure 2 *Metabolic map displaying model based alternative routes to acetyl-CoA through the formation of acetaldehyde (orange), derived either via serine (red) or via threonine (green). The metabolic map is based on reactions and metabolites of iJO1366 E. coli metabolic model and was generated using Escher.*

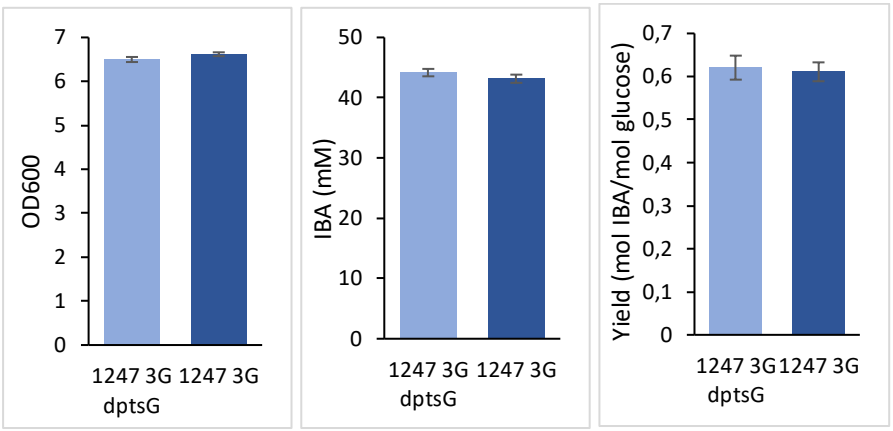
3.7.6 Supplementary Figure 3



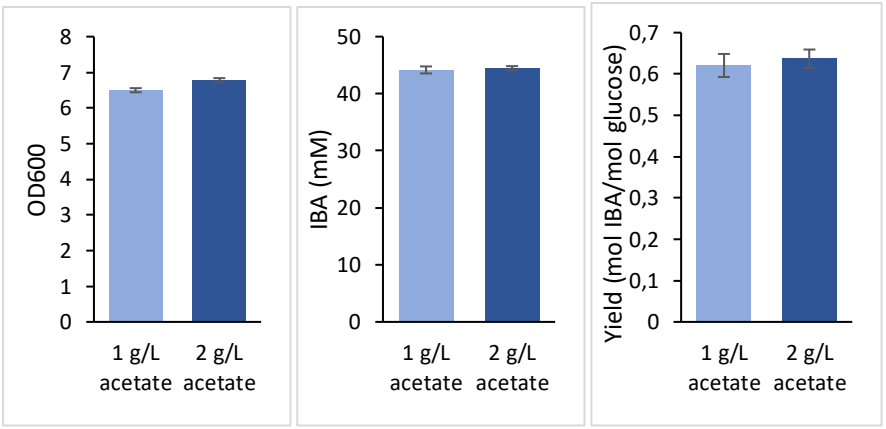
Supplementary Figure 3. *Growth curves of the acetic acid auxotroph E. coli K-12 MG1655 ΔfocA-pflB, ΔaceEF, ΔpoxB, ΔtdcE, ΔpflDC, Δpfo, ΔdeoC in minimal M9 media supplemented with 10 mM glucose (black symbols), 25 mM acetate (open squares) or 10 mM glucose and 25 mM acetate (grey circles).*

3.7.7 Supplementary Figure 4

A.

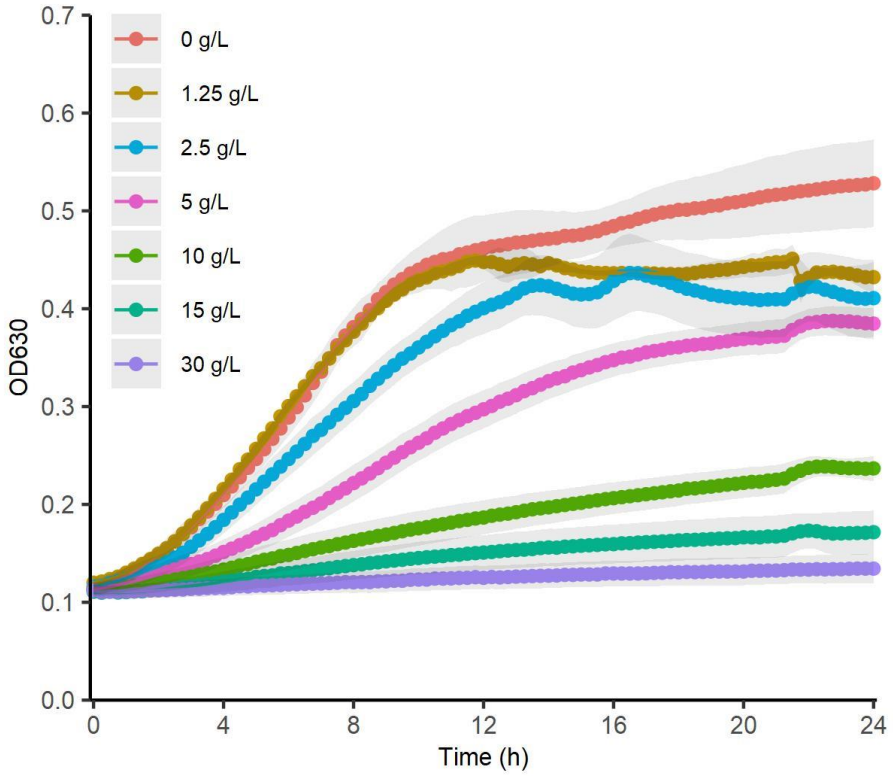


B.



Supplementary Figure 4 A) *Cell density (OD_{600}), IBA titer, and IBA yield from glucose after 24 h for strains 1247 3G ΔptsG and 1247 3G. Cells were grown in M9 Extra media supplemented with 111 mM (20 g L^{-1}) of glucose, 17 mM (1 g L^{-1}) NaOAc, and 0.5% yeast extract. Cells were induced with 0.1 mM IPTG after 5 h.* ***B)*** *Cell density (OD_{600}), IBA titer, and IBA yield from glucose after 24 h for strain 1247 3G ΔptsG, grown in M9 Extra media supplemented with 111 mM (20 g L^{-1}) of glucose, either 17 or 34 mM (1 or 2 g L^{-1}) NaOAc and 0.5% yeast extract. Cells were induced with 0.1 mM IPTG after 5 hours. Error bars indicate mean ± s.d. (n=4 biological replicates).*

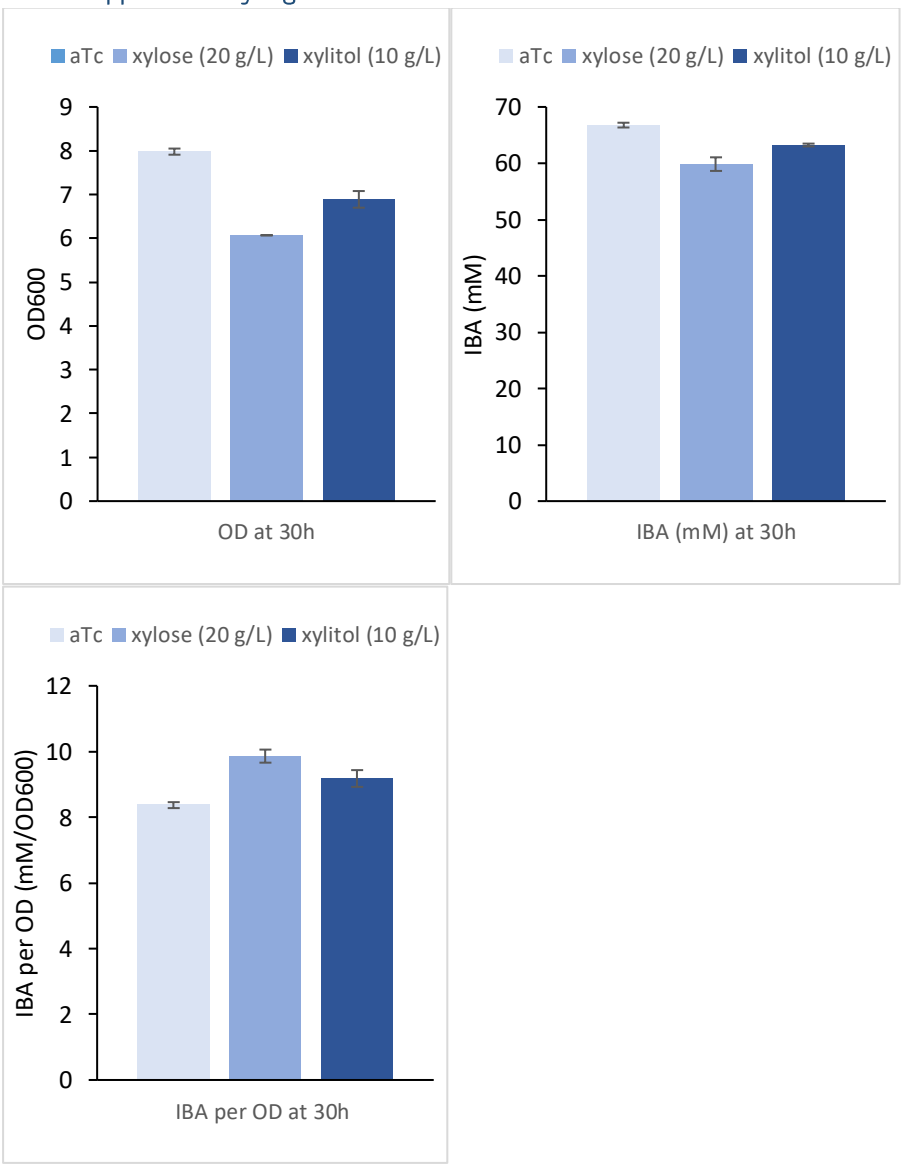
3.7.8 Supplementary Figure 5



Supplementary Figure 5 *Growth of strain SIJ216 with different concentrations of IBA in the media, measured using a microtiter plate reader. Cells were grown in M9 Extra media supplemented with 56 mM (10 g L^{-1}) of glucose, and between 1.25 to 30 g/L of IBA. Grey shading indicate mean ± s.d. (n=6 biological replicates).*

Based on these results, a production experiment with duplicates was setup, using the same parameters as in Supp. Fig 1, but with 2.5 g L^{-1} of IBA added to the media. While the OD after 24 hours was lower (0.696±0.004 vs. 1.0875±0.018), the xylitol production was not affected (6.818±0.03 vs. 5.69±0.055 g L^{-1}). Higher titer is likely due to inhibited growth leaving more glucose for xylitol production.

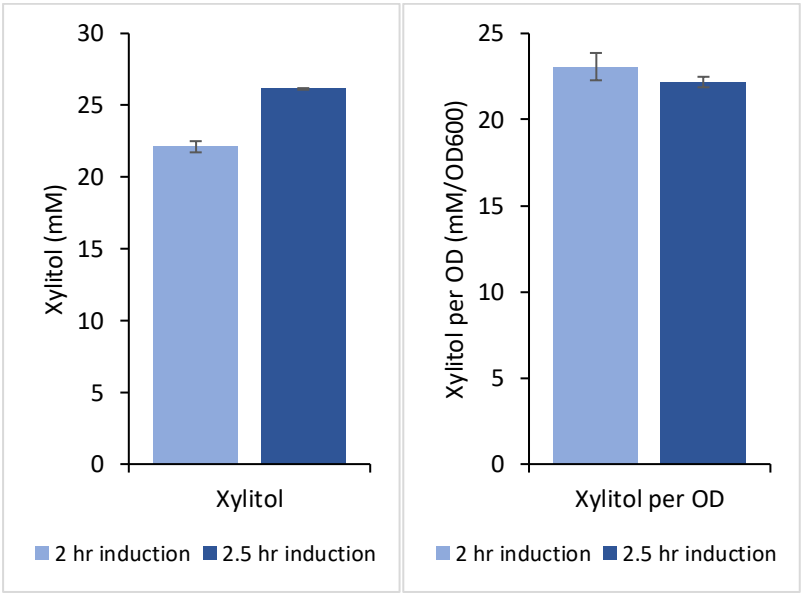
3.7.9 Supplementary Figure 6



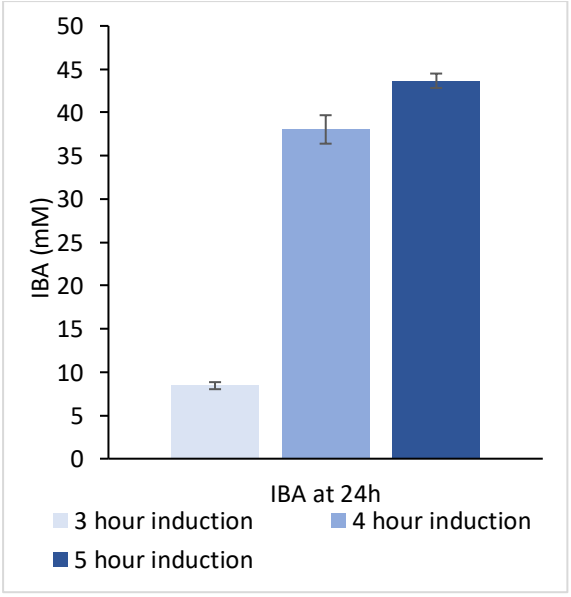
Supplementary Figure 6 *Test of the effects of substrate, product, and inducer of strain SIJ216 on strain 1247-3G-ptsG. Three sets of duplicates, with 1 ug mL^{-1} of aTc, 133 mM (20 g/L) xylose or 66 mM (10 g/L) xylitol added (aTc or xylose at 2 hours mimicking co-cultivation induction, xylitol at 6 hours). Figures show OD, IBA titer (mM) ,and IBA per OD (mM/OD600) after 30 hours.*

3.7.10 Supplementary Figure 7

A.



B.



Supplementary Figure 7 *Optimizing induction times for co-cultivation. A) Xylitol and OD-normalized xylitol at 24 hours of a co-cultivation with different induction times for xylitol strain. M9 Extra media supplemented with 111 mM (20 g L^{-1}) of glucose, 17 mM (1 g L^{-1}) acetate and 0.5% yeast extract. Induced at 2 or 2.5 hours with 120 mM (18 g L^{-1}) xylose and 1 ug mL^{-1} aTc. N=4 replicates were used. B) IBA production at 24 hours of a co-cultivation with different induction times for the IBA strain. M9 Extra media supplemented with 111 mM (20 g L^{-1}) of glucose, 17 mM (1 g L^{-1}) acetate and 0.5% yeast extract. Induced at 3, 4 or 5 hours with 0.1 mM IPTG. Error bars indicate mean ± s.d. (n=4 biological replicates for induction after 4 and 5 h of growth; n=3 biological replicates for induction after 3 h of growth).*

3.7.11 References for supplementary materials

[1] Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 6640-6645 (2000).

[2] Cherepanov, P.P. & Wackernagel, W. Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158**, 9-14 (1995).

[3] Jensen, S.I., Lennen, R.M., Herrgard, M.J. & Nielsen, A.T. Seven gene deletions in seven days: Fast generation of Escherichia coli strains tolerant to acetate and osmotic stress. *Scientific reports* **5**, 17874 (2015).

[4] St-Pierre, F. et al. One-step cloning and chromosomal integration of DNA. *ACS synthetic biology* **2**, 537-541 (2013).

[5] Qi, L.S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183 (2013).

[6] Larson, M.H. et al. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature protocols* **8**, 2180-2196 (2013).

[7] Zhang, K., Woodruff, A.P., Xiong, M., Zhou, J. & Dhande, Y.K. A synthetic metabolic pathway for production of the platform chemical isobutyric acid. *ChemSusChem* **4**, 1068-1070 (2011).

Chapter 4

Expanding the CRISPRi toolbox in *Bacillus subtilis* for application towards metabolic switches

Yixin Rong[1], Regiane Alves de Oliveira[1], Kristoffer Bach Falkenberg[1], Ácil de Almeida Will[1], Alex Toftgaard Nielsen[1], Sheila Ingemann Jensen[1*]

[1] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet Building 220, 2800 Kongens Lyngby, Denmark

*Corresponding author

Sheila Ingemann Jensen

The Novo Nordisk Foundation Center for Biosustainability

Technical University of Denmark

Kemitorvet, DK-2800 Kongens Lyngby, Denmark

Email: shin@biosustain.dtu.dk

Abstract

Synthetic biology tools enable dynamic metabolic engineering, such as using CRISPR interference (CRISPRi) as a metabolic switch. One interesting way of implementing CRISPRi metabolic switches is trying to achieve growth decoupling, where cell growth is inhibited but production is maintained. This can be beneficial for the production of proteins and biochemicals. CRISPRi growth decoupling requires tightly regulated promoters, which greatly limits what promoters can be used. In this study, we looked into expanding the CRISPRi toolbox for the industrially relevant model organism *Bacillus subtilis.* In order to improve promoter compatibility, alternate start codons and protein degradation tags were added onto dCas9 to tune protein levels and compensate for expression prior to induction. Furthermore, a one-pot cloning method was developed for cloning of sgRNA plasmids. We also showed that pyrimidine biosynthesis, specifically *pyrG,* could potentially serve as a target gene towards achieving growth decoupling.

4.1 Introduction

Bio-based production is becoming increasingly important to decrease global dependence on fossil fuel and to transition towards a circular bio-economy [1] Advances in metabolic engineering and synthetic biology have enabled the development of microbial cell factories for the production of numerous different chemicals and proteins. In the past, metabolic engineering was in most cases static, meaning that modifications to the genome would be permanent. This approach to metabolic engineering resulted in certain limitations as to what parts of metabolism could be knocked out or downregulated, as targeting essential metabolism would result in deficiencies in growth. Recently, development in synthetic biology tools have enabled a new approach of dynamic metabolic engineering, where genes can be repressed in a controllable manner, either using inducible promoters or auto regulated using biosensors or quorum sensing [2][3].

One of the most studied and broadly applied methods of controllable gene repression is CRISPR (Clustered Regularly InterSpaced Palindromic Repeats) interference or CRISPRi for short. Using a catalytically inactive version of the Cas9 nuclease, dCas9, CRISPRi enables programmable, orthogonal and highly effective gene repression, up to 300 fold in *Escherichia coli* and up to >100 fold in *Bacillus subtilis* [3][4]. CRISPRi has been shown to have potential benefits towards the production of different biochemicals by dynamically rewiring the metabolism during the fermentation [5]–[7]. *B. subtilis* is an industrially relevant, Gram-positive model organism, which originally was of interest due to its ability to secrete proteins, but has also been explored as a possible platform organism for chemical production [8][9]. CRISPRi in B. subtilis has been developed using a xylose inducible promoter controlling the dCas9 protein combined with a constitutively expressed sgRNA. However, the system is somewhat leaky, which limits its applicability, when targeting essential genes for e.g. growth decoupled production processes. The use of a sugar dependent promoter such as P_{xylA} further limits the use of biomass hydrolysates [6][10][11]. The tightly regulated anhydrotetracycline inducible promoter P_{tet} [12] could serve as an alternative that does not get induced by sugars found in common feedstocks.

One interesting application of CRISPRi is growth decoupled production. The overall goal of growth decoupled production is to achieve a cellular state, where the cells stop growing but maintain a high productivity, thereby increasing the efficiency of

feedstock utilization, which is a major production cost contributor [13][14]. Growth decoupling requires a tightly regulated promoter, as basal expression without induction leads to growth inhibition at unintended phases of the fermentation. One possible way to compensate for this leakiness is to tune the expression level of the dCas9 protein, so that the basal expression has no effect on growth, while fully induced dCas9 levels still repress sufficiently to achieve the desired phenotype.

In this study, the goal was to expand the CRISPRi toolbox in *B. subtilis* by developing a streamlined protocol for cloning guide RNA's (sgRNA) and by using two different methods to optimize dCas9 protein basal expression levels. The first approach is by substituting the default AUG start codon with UUG or GUG, which were shown in literature to reduce β-galactosidase activity by ~2.3 fold and ~3.2 fold respectively [15]. The second approach was to create unstable variants of dCas9 by attaching a SsrA degradation tag to the C-terminal of the protein, which targets dCas9 to the ClpXP protease [16]. This has previously been shown in literature to lower basal protein levels in *B. subtilis* [17].

4.2 Materials and methods

4.2.1 Media and culture conditions

Cultivation of bacteria was done at 37 °C and 250 RPM shaking unless specified otherwise. For cloning and plasmid propagation, Luria-Bertani (LB) broth was used (LB Broth (Lennox), Sigma-Aldrich). LB agar was used as solid media. For selection of Escherichia coli during DNA manipulation, 100 mg L^{-1} of ampicillin was used, and for selection of B. subtilis strains during transformation and other experiments, the following antibiotic concentrations were used: spectinomycin at 100 mg L^{-1}, erythromycin at 7.5 mg L^{-1}, kanamycin at 7.5 mg L^{-1} or chloramphenicol at 5 mg L^{-1}. For experiments to characterize the CRISPRi performance, a modified M9 medium called M9extra was used (12.8 g L^{-1} $Na_2HPO_4 \cdot 7\ H_2O$, 3 g L^{-1} KH_2PO_4, 0.5 g L^{-1} NaCl, 1 g L^{-1} NH_4Cl, 2 mM $MgSO_4$, and 0.1 mM $CaCl_2$) supplemented with 60 µM $FeCl_3$, a trace element solution (1.25 µM $MnCl_2 \cdot 4\ H_2O$, 0.21 µM $CoCl_2 \cdot 6\ H_2O$, 0.85 µM $ZnSO_4 \cdot 7\ H_2O$, 0.05 µM $CuCl_2 \cdot 2\ H_2O$, 0,08 µM H_3Bo_3, 0.105 µM $NiCl_2 \cdot 6\ H_2O$, 0.125 µM $NaMoO_4 \cdot 2\ H_2O$), and with 20 g L-1 glucose added as carbon source. For overnight pre-cultures, the M9extra medium was also supplemented with 0.2 g L^{-1} of yeast extract.

4.2.2 Plasmid and strain construction

Cloning and plasmid propagation were performed using the strain E. coli DH5λ pir. Integrative plasmids were constructed using backbones from the ProUSER2.0 toolbox, following the cloning protocol described in the study [12]. PCR fragments for cloning were generated using Phusion U polymerase (Thermo Fisher Scientific). Purification of plasmids and PCR fragments were performed using NucleoSpin Plasmid or PCR kits respectively (Macherey-Nagel). Transformations were performed as described in previous studies [18]. Sequencing was performed using Mix2Seq kits (Eurofins Genomics).

The parental strain used to characterize CRISPRi performance was *B. subtilis* 168 KO7S (*ΔnprE ΔaprE Δepr Δmpr ΔnprB Δvpr Δbpr trp*$^{+}$ *ΔsigF glmS*::P_{MtlA}-comKS)(unpublished work). Transformations were performed using the inducible comKS system described in a previous study [19]. The protocol was altered to 2 hours of recovery at 100-150 RPM shaking at 37 °C after mixing with DNA. Transformants were verified using colony PCR with OneTaq polymerase (New England Biolabs). Colonies were picked and transferred to 20 µL LB media. 10 µL cell mixture was added to 10 µL MilliQ water, put on ice for 5 minutes then heated and cooled for three cycles (1 minute in 800W microwave, 30 seconds on ice), placed on ice for 5 minutes before using 1 µL as template for the PCR reaction.

The list of primers, plasmids and strains generated and used for the study can be found in Supplementary Tables S1-S3.

4.2.3 One-pot sgRNA cloning

Oligos were diluted to 10 µM and 25 µL of each complementary oligo were mixed in a PCR tube and annealed in a PCR cycler (95 °C for 5 minutes, then cooled 1 °C per minute until 10 °C). The annealed oligos were diluted 100x and 1 µL was mixed with 200-250 ng of backbone, 1 µL of SapI restriction enzyme (New England Biolabs), 1 µL CutSmart buffer (New England Biolabs), 1 µL of T4 ligase (Thermo Fisher Scientific), 1 µL of T4 ligase buffer (Thermo Fisher Scientific) and filled to 10 µL with MilliQ water. The reaction mixture was cycled between restriction and ligation in a PCR cycler: 37 °C for 5 minutes, 22 °C for 10 minutes, repeat for 10 cycles, then restrict at 37 °C for 30 minutes and inactivate enzymes at 75 °C for 15 minutes. The mixture was diluted 2x using MilliQ water and 5 µL of the mixture was used for transformation.

4.2.4 Growth experiments with microtiter plate reader

Strains were streaked from cryostocks onto LB agar plates the day before, and one colony was used to inoculate 1 mL of LB media in a 24 deep well plate (DWP) and grown during the day at 37 °C and 250 RPM shaking for 7-9 hours. The day culture was diluted 5000 times in 2.5 mL M9extra media with 20 g L^{-1} glucose and 0.2 g L^{-1} yeast extract in a 24 DWP and grown overnight at 37 °C and 250 RPM shaking for no more than 16 hours. The overnight culture was diluted to an optical density at 600nm (OD600) of 0.05 in 135 µL of M9extra media with 20 g L^{-1} glucose added in microtiter plate wells. The plate was sealed with a gas permeable film, and measured using a plate reader with continuous shaking at medium intensity at 37 °C, with readings every 15 minutes. For experiments only measuring growth, EL808x plate readers were used and absorbance measured at 630nm. For experiments measuring fluorescence, Synergy HM1 plate readers were used, growth was measured with absorbance at 600nm, mRuby fluorescence was measured at excitation/emission wavelengths of 558/605nm and sfGFP fluorescence was measured at 485/510nm. CRISPRi was induced after one doubling or approx. 8 hours with 0.1 µg mL^{-1} anhydrotetracycline (aTc).

4.2.5 Growth experiments with shake flasks

Strains were streaked from cryostocks onto LB agar plates the day before, and one colony was used to inoculate 1 mL of LB media in a 24 DWP and grown during the day at 37 °C and 250 RPM shaking for 7-9 hours. The day culture was diluted 5000 times in 50 mL M9extra media with 20 g L^{-1} glucose and 0.2 g L^{-1} yeast extract in a 250 mL bottom baffled Erlenmeyer flask and grown overnight at 37 °C and 250 RPM shaking for no more than 16 hours. The overnight culture was diluted to an optical density at 600nm (OD600) of 0.05 in 50 mL of M9extra media with 20 g L^{-1} glucose added in 250 mL flasks. The cultures were grown at 37 °C and 250 RPM shaking. CRISPRi was induced after one doubling or approx. 4 hours with 0.1 µg mL^{-1} aTc, and OD600 was measured every 2 hours using a spectrophotometer until 16 hours from inoculation.

For fluorescence quantification using flow cytometry, 1 mL of culture was spun down at 6500G for 5 minutes in an Eppendorf tube. The supernatant was discarded and the pellet was resuspended in 500 µL 2% w/v paraformaldehyde in phosphate buffer solution (PBS). The cells were incubated for 30-60 minutes, then spun down at 6500G for 5 minutes. The supernatant was discarded and the pellet was

resuspended in 500 µL of PBS and stored at 4 °C (for no more than 7 days) prior to analysis.

4.2.6 Measuring fluorescence with flow cytometry

Flow cytometry, with a Macsquant VYB (Miltenyi Biotech), was used to quantify fluorescence of constitutively expressed mRuby when *pyrG* was inhibited, and constitutively expressed sfGFP when determining repression fold of the TTG dCas9 variant. Samples were diluted to a volume of 200 µL in microtiter plate wells with PBS according to the OD600 at the time of sampling, to what would correspond to approx. OD 0.1. mRuby2 fluorescence was measured using an excitation wavelength of 561 nm with a 615/20 nm bandpass filter for emission, and sfGFP fluorescence was measured by using an excitation wavelength of 488 nm and a 525/50 nm bandpass filter for emission. The gain settings used were as follows: 450 V for forward scatter, 425 V for side scatter, 400 V for Y2 channel and 450 V for B1 channel. The events were gated for single cells, and 50000 or 100000 events were recorded for sfGFP and mRuby2 measurements, respectively.

4.3 Results and discussion

4.3.1 One-pot sgRNA cloning allows for fast and simple construction of sgRNA plasmids

As a part of expanding the CRISPRi toolbox for *B. subtilis*, we developed a “one-pot” method for cloning new sgRNA’s into vector backbones, as illustrated in Figure 1. This method uses a modified backbone from the ProUSER 2.0 toolbox, which makes it possible to change parts of the backbone, such as the integration site, antibiotic marker, or changing to plasmid-based expression, without PCR amplification [12]. Changing promoters would first require a USER-ligation PCR.

The principle is similar to a recently published study by Bradley, describing a method for sgRNA cloning for plasmid based expression in *E. coli*, which also used type IIS restriction enzymes [20]. The approach allows for simultaneous restriction and ligation, due to successfully ligated plasmids no longer having the restriction sites present. Furthermore, the cloning is fully scarless as the overhangs on the oligos can be changed to match the promoter which enables the sgRNA to start at the +1 site. The method as described in this study can be completed in a single day and has an efficiency of >95% (see Supplementary Table 4). For cloning libraries, the efficiency could potentially be further improved by increasing the number of cycles to 20 or 30.

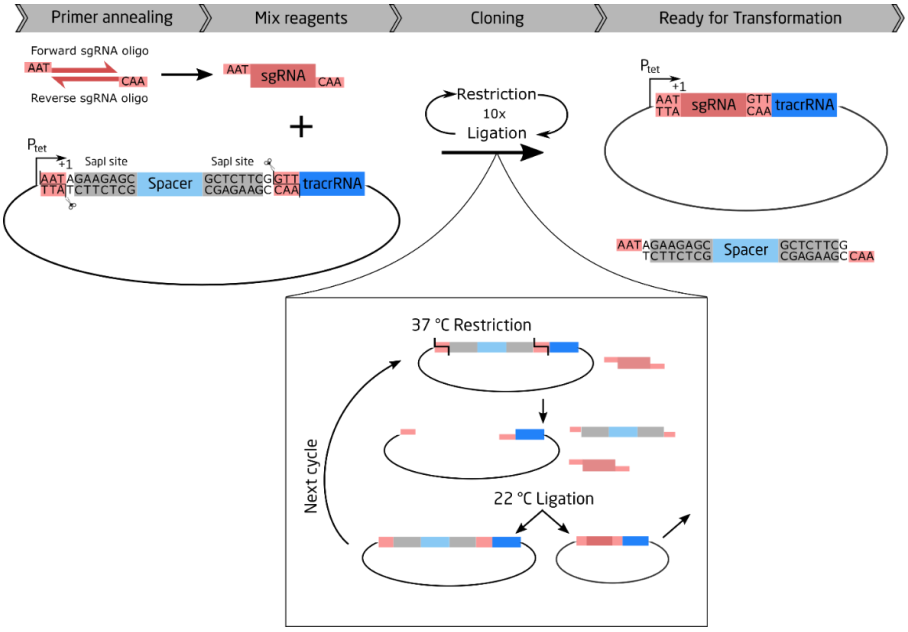


Figure 1 *One-pot cloning method for sgRNA plasmids for B. subtilis. The desired sgRNA sequence(s) can be ordered as a pair of complementary oligos, with overhangs compatible with the promoter of choice and tracrRNA. Annealed oligos are mixed with backbone, SapI restriction enzyme and T4 ligase and goes through 10 cycles of restriction and annealing; upon ligation of the sgRNA brick, the plasmid will not get digested during the remaining cycles.*

4.3.2 Alternate start codons can be used to tune dCas9 protein levels

To evaluate how CRISPRi functions when the dCas9 protein is expressed when using the alternate start codons TTG and GTG, growth experiments were performed in Erlenmeyer flasks with a gRNA targeting the *pyrG* gene in the pyrimidine biosynthesis pathway. The pyrimidine pathway has in previous studies been shown to be an interesting target for growth decoupling for the purpose of improving the production of proteins and biochemicals both in *Escherichia coli* [14][21], and in *B. subtilis* [22]. Growth inhibiting targets are also suitable for evaluating basal promoter expression, as expression while uninduced would lead to either impaired growth when compared to the parental strain without CRISPRi or loss of function due to e.g. mutations in the dCas9 protein.

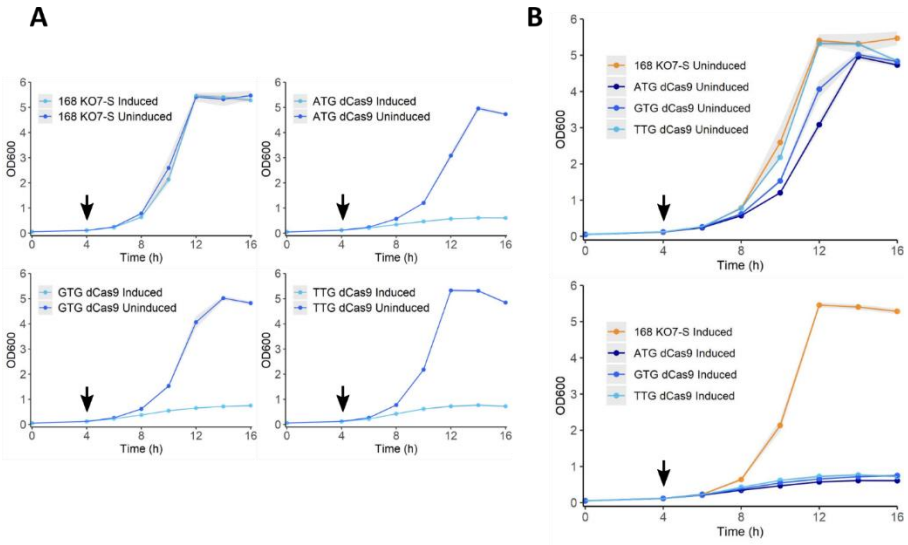


Figure 2 *Growth experiment performed in Erlenmeyer flasks to evaluate CRISPRi performance with dCas9 expressed using alternate start codons. CRISPRi was induced at 4 hours with 0.1 µg mL^{-1} aTc, indicated by the arrow. A) Each start codon plotted separately, with the parental strain 168 KO7-S as a control. Light blue represents the induced sets, and dark blue the uninduced sets. B) All start codons and parental strain plotted together, when uninduced and induced, respectively. Standard deviation (n=3 biological replicates) is depicted as gray shading around the line.*

Each individual start codon exhibit a strong attenuation of growth when CRISPRi is induced, with growth stopping around 6-8 hours post induction (Figure 1A). However, when comparing the growth of uninduced cells (Figure 2B), it is clear that the P_{tet} promoter exhibits enough expression with the consensus start codon ATG dCas9 to negatively impact growth, even without the presence of inducer. This is not caused from expression of dCas9 alone (see Supplementary Figure 1). Interestingly, changing the start codon to either GTG or TTG shows growth that is closer to the parental strain, especially with the TTG start codon. Literature suggests that TTG is preferred over GTG, both in prevalence in the genome and when used instead of ATG [15][23]. The results here suggest that in the case of dCas9, the preference is reversed. This could be due to overall translation initiation efficiency, which is influenced by the sequence of the translation initiation region (TIR) stretching from the Shine-Delgarno (SD) sequence to the first nucleotides of the open reading frame (ORF) [24][25]. Nonetheless, the TTG variant of dCas9 shows clear improvement on growth when uninduced, while simultaneously showing very

similar levels of growth inhibition compared to the consensus ATG start codon when induced. This makes it an interesting variant for further experiments. Overall the results show that using alternate codons is a simple and viable method to tune down dCas9 protein levels to compensate for P_{tet} expression prior to induction.

In order to achieve a more quantifiable measure of repression efficiency, a gRNA was designed to target superfolder GFP (sfGFP) expressed under the constitutive P_{veg} promoter. Here it was estimated that the dCas9 using the TTG start codon had a repression fold of 13.9±1 (see Supplementary Figure 2). Due to time constraints, it was not possible to compare this repression fold to the consensus ATG start codon in order to quantify the effect on the repression strength from changing the start codon.

4.3.3 Degradation tags targeting ClpXP can be used to create unstable dCas9 variants

Another potential approach to tune dCas9 protein levels is by creating “unstable” variants through addition of a C-terminal tag to target the protein for degradation. By targeting dCas9 for degradation, the goal is to compensate for accumulation of dCas9 prior to induction. This was accomplished by targeting dCas9 to the ClpXP protease through addition of a 15 AA long SsrA-tag AGKTNSFNQNVA(LAA), where the last 3 amino acids can be changed to alter the rate of degradation [17][26]. AAV and LVA were chosen initially as the last 3 amino acids. The LVA tag sequence showed protein levels during exponential phase that were similar to the wildtype LAA sequence, while AAV tag sequence is degraded slower during stationary phase.

The individual plots in Figure 3A show that the unstable dCas9 variants are able to attenuate growth upon induction of CRISPRi. When comparing the uninduced growth to the stable untagged dCas9 in Figure 3B, it can be seen that both versions of the degradation tags behave similarly and display improved growth, much like the alternate start codons in Figure 2. However, with CRISPRi induced, it can be seen that the unstable dCas9 variants has a lag-phase prior to growth attenuation compared to the stable untagged variant. This indicates that gene repression is weaker, likely due to dCas9 being less stable and a target for degradation. Interestingly, the unstable variants show a slight decrease in OD from 12 to 16 hours. This could be related to dCas9 being targeted for the ClpXP protease, and thus affecting its ability to perform its usual housekeeping tasks [27]. Outside of time constraints, it would be interesting to change the last 3 AAV to ASV, which has been

shown to be more stable [17]. Furthermore, specifically deploying with pyrimidine inhibition could impact long-term stability of the CRISPRi, as lack of pyrimidine inhibits synthesis of new proteins, including dCas9, to replace the dCas9 that is degraded. Overall, these results show that it is possible to create unstable dCas9 variants via C-terminal addition of a SsrA degradation tag. However, for growth decoupling applications via inhibition of pyrimidine biosynthesis, it does not appear to be the optimal approach. Unstable dCas9 variants could be useful in other scenarios where uninduced CRISPRi gene repression needs to be tight, or where the repression needs to be lifted.

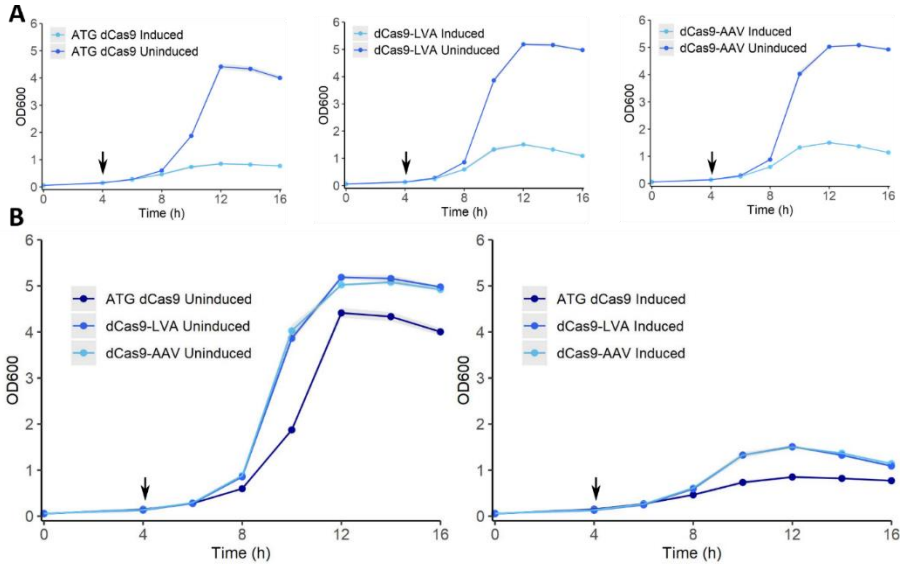


Figure 3 *Growth experiment performed in Erlenmeyer flasks to evaluate CRISPRi performance with unstable variants of dCas9. CRISPRi was induced at 4 hours with 0.1 µg mL^{-1} aTc, indicated by the arrow. A) Each start codon plotted separately. Light blue represents the induced sets, and dark blue the uninduced sets. B) The stable dCas9 and two unstable variants plotted together, when uninduced and when induced, respectively. Standard deviation (n=3 biological replicates) is depicted as gray shading around the line.*

4.3.4 *pyrG* as a target provides strong growth inhibition

To demonstrate the simplicity of the one-pot sgRNA cloning method, we created plasmids of sgRNA's that target 5 other genes within either pyrimidine biosynthesis or DNA replication, which were shown in the study by Li et al to be viable mechanisms to target for achieving growth decoupling in E. coli [14]. Of the other target genes, *gmk* and *cmk* are also a part of pyrimidine biosynthesis, while *LigA*,

dnaB and *dnaI* are involved in DNA replication [28]–[31]. To test their suitability for growth decoupling, growth with CRISPRi was measured in a plate reader (Figure 4).

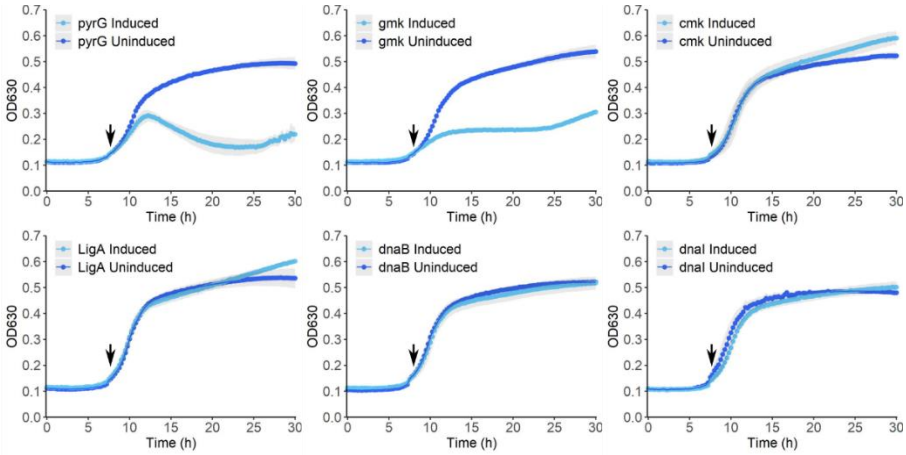


Figure 4 *Growth experiment comparing different gRNA targets. CRISPRi was induced at 7.5 hours with 0.1 µg mL^{-1} aTc, indicated by the arrow. Standard deviation (n=4 biological replicates) is depicted as gray shading around the line.*

Results in Figure 4 shows that among the target genes tested, *pyrG* results in the strongest growth inhibition, with *gmk* also showing potential for growth decoupling. However, purine metabolism is less preferable as a target due to possibly affecting adenosine nucleotide supply and thus the energy metabolism. Other targets did not show growth inhibition, possibly due to other genes being able to complement their function.

4.3.5 Inhibiting *pyrG* results in growth decoupled but metabolically active cells

In order to study the metabolic activity of the cells under growth decoupling induced by pyrG repression, red fluorescence from mRuby2 expressed under the strong constitutive promoter P3P was measured using flow cytometry. TTG dCas9 was used as this was the most promising variant from earlier experiments.

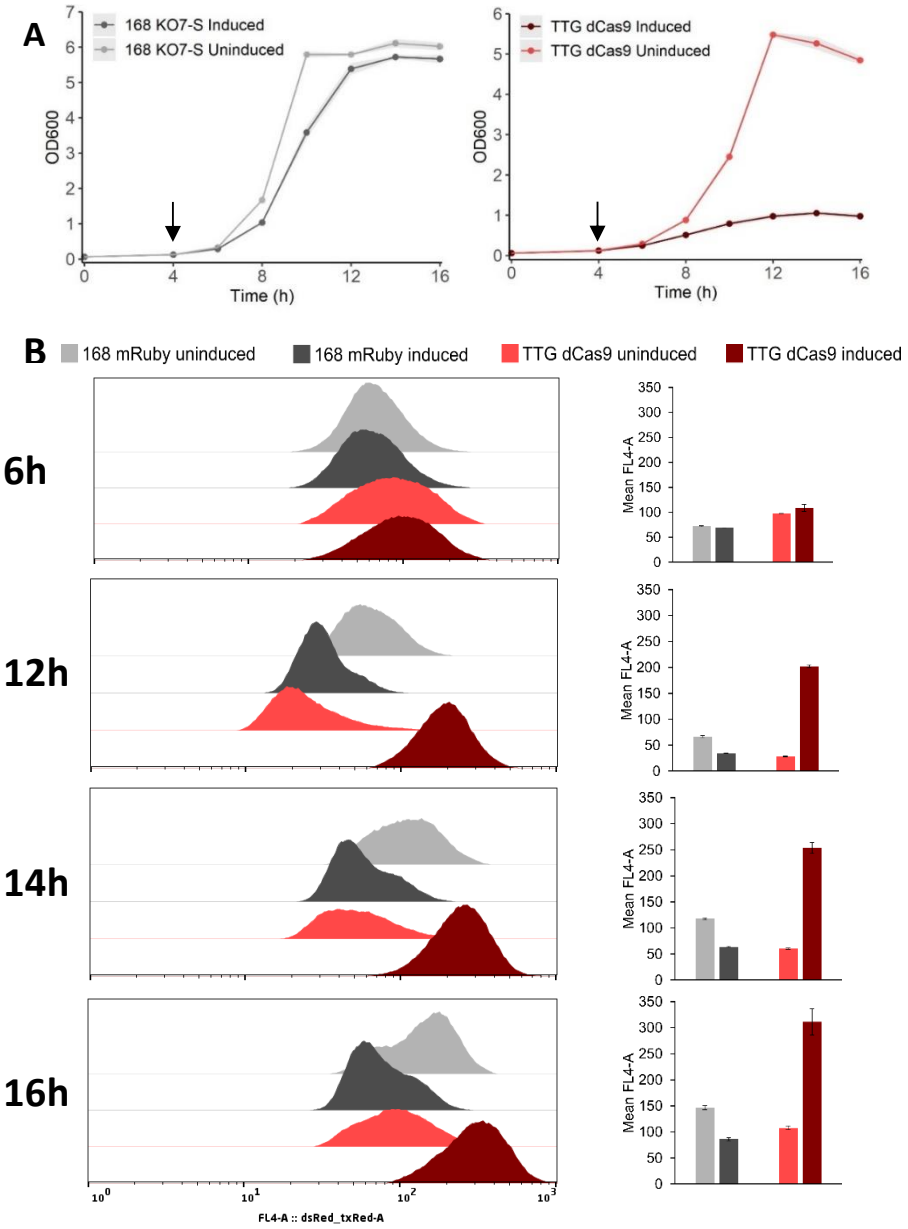


Figure 5 *Growth and constitutive mRuby expression under pyrG inhibition using the TTG start codon dCas9, compared to parental strain. CRISPRi was induced at 4 hours with 0.1 µg mL^{-1} aTc.. A) Growth of parental strain and strain with TTG dCas9 with pyrG gRNA. B) Red fluorescence measurement using flow cytometry at selected time points. Standard deviation (n=4 biological replicates) is depicted as shading around the lines or error bars. For additional time points, see Supp. Figure 3.*

The growth in Figure 5A resembles previous results in Figure 2. After 6 hours, or 2 hours after induction, the red fluorescence values are similar. However, when the growth attenuation is in full effect at 12 hours, it can be seen that cells with CRISPRi induced has a much higher red fluorescence value. This is also the case at 14 hours and 16 hours. This is a clear indication that the cells stay metabolically active and express heterologous protein while growth is inhibited. As discussed above, for the purpose of protein production (intracellular or secreted), depleting pyrimidine would in theory be inhibitory towards protein synthesis, although some results in *E. coli* suggest this is not necessarily the case [21]. However, in the case of biochemical production, enzymes in the biosynthesis pathway have time to accumulate prior to full repression from CRISPRi, making it an interesting target for such applications.

4.4 Conclusions

CRISPRi enables deployment of metabolic switches, such as for growth decoupling, in order to improve the production of biochemicals and proteins. This requires tightly regulated promoters. In this study, we explored strategies to tune dCas9 protein levels for better compatibility with different promoters. We showed that replacing the consensus start codon ATG with TTG leads to improved growth characteristics when repressing *pyrG*, a potential target for growth decoupled production. We additionally developed unstable variants of dCas9, which while not the optimal approach for this application, could be of use for other applications. Repressing *pyrG* with TTG dCas9 shows maintained expression of heterologous proteins during growth inhibition. Furthermore, we developed a fast and simple method for cloning sgRNA plasmids. Altogether, we have expanded the toolbox for deploying CRISPRi towards metabolic switches in *B. subtilis*.

4.5 Acknowledgements

This work was funded by the Novo Nordisk Foundation within the framework of the Fermentation-based Biomanufacturing Initiative (FBM), grant number: NNF17SA0031362.

4.5 References for Chapter 4

[1] E. C. D. Tan and P. Lamers, "Circular Bioeconomy Concepts—A Perspective," Frontiers in Sustainability, vol. 2, Jul. 2021, doi: 10.3389/frsus.2021.701509.

[2] Y. Liu, J. Chen, D. Crisante, J. M. Jaramillo Lopez, and R. Mahadevan, "Dynamic Cell Programming with Quorum Sensing-Controlled CRISPRi Circuit," ACS Synthetic Biology, vol. 9, no. 6, pp. 1284–1291, Jun. 2020, doi: 10.1021/acssynbio.0c00148.

[3] Y. Wu et al., "Design of a programmable biosensor-CRISPRi genetic circuits for dynamic and autonomous dual-control of metabolic flux in Bacillus subtilis," Nucleic Acids Research, vol. 48, no. 2, pp. 996–1009, Jan. 2020, doi: 10.1093/nar/gkz1123.

[4] L. S. Qi et al., "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression," Cell, vol. 152, no. 5, pp. 1173–1183, Feb. 2013, doi: 10.1016/j.cell.2013.02.022.

[5] X. Ji, H. Zhao, H. Zhu, K. Zhu, S.-Y. Tang, and C. Lou, "CRISPRi/dCpf1-mediated dynamic metabolic switch to enhance butenoic acid production in Escherichia coli," Applied Microbiology and Biotechnology, vol. 104, no. 12, pp. 5385–5393, Jun. 2020, doi: 10.1007/s00253-020-10610-2.

[6] Y. Wu et al., "CRISPRi allows optimal temporal control of N-acetylglucosamine bioproduction by a dynamic coordination of glucose and xylose metabolism in Bacillus subtilis," Metabolic Engineering, vol. 49, pp. 232–241, Sep. 2018, doi: 10.1016/j.ymben.2018.08.012.

[7] A.-K. Löbs, C. Schwartz, S. Thorwall, and I. Wheeldon, "Highly Multiplexed CRISPRi Repression of Respiratory Functions Enhances Mitochondrial Localized Ethyl Acetate Biosynthesis in Kluyveromyces marxianus," ACS Synthetic Biology, vol. 7, no. 11, pp. 2647–2655, Nov. 2018, doi: 10.1021/acssynbio.8b00331.

[8] H.-P. Hohmann, J. M. van Dijl, L. Krishnappa, and Z. Prágai, "Host Organisms: Bacillus subtilis," in Industrial Biotechnology, Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2016, pp. 221–297. doi: 10.1002/9783527807796.ch7.

[9] Q. Zhang et al., "Production of proteins and commodity chemicals using engineered Bacillus subtilis platform strain," Essays in Biochemistry, vol. 65, no. 2, pp. 173–185, Jul. 2021, doi: 10.1042/EBC20210011.

[10] C. Wang, Y. Cao, Y. Wang, L. Sun, and H. Song, "Enhancing surfactin production by using systematic CRISPRi repression to screen amino acid biosynthesis genes in Bacillus subtilis," Microbial Cell Factories, vol. 18, no. 1, p. 90, Dec. 2019, doi: 10.1186/s12934-019-1139-4.

[11] X. Dong et al., "CRISPRi-Guided Multiplexed Fine-Tuning of Metabolic Flux for Enhanced Lacto- N -neotetraose Production in Bacillus subtilis," Journal of Agricultural and Food Chemistry, vol. 68, no. 8, pp. 2477–2484, Feb. 2020, doi: 10.1021/acs.jafc.9b07642.

[12] K. B. Falkenberg et al., "The ProUSER2.0 Toolbox: Genetic Parts and Highly Customizable Plasmids for Synthetic Biology in Bacillus subtilis," ACS Synthetic Biology, vol. 10, no. 12, pp. 3278–3289, Dec. 2021, doi: 10.1021/acssynbio.1c00130.

[13] Y. Li et al., "Sustainable Lactic Acid Production from Lignocellulosic Biomass," ACS Sustainable Chemistry & Engineering, vol. 9, no. 3, pp. 1341–1351, Jan. 2021, doi: 10.1021/acssuschemeng.0c08055.

[14] S. Li et al., "Enhanced protein and biochemical production using CRISPRi-based growth switches," Metabolic Engineering, vol. 38, pp. 274–284, Nov. 2016, doi: 10.1016/j.ymben.2016.09.003.

[15] R. L. Vellanoweth and J. C. Rabinowitz, "The influence of ribosome-binding-site elements on translational efficiency in Bacillus subtilis and Escherichia coli in vivo," Molecular Microbiology, vol. 6, no. 9, pp. 1105–1114, May 1992, doi: 10.1111/j.1365-2958.1992.tb01548.x.

[16] S. Gottesman, E. Roche, Y. N. Zhou, and R. T. Sauer, "The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system," Genes and Development, vol. 12, no. 9, pp. 1338–1347, 1998, doi: 10.1101/gad.12.9.1338.

[17] S. Guiziou et al., "A part toolbox to tune genetic expression in Bacillus subtilis," Nucleic Acids Research, p. gkw624, Jul. 2016, doi: 10.1093/nar/gkw624.

[18] H. Inoue, H. Nojima, and H. Okayama, "High efficiency transformation of Escherichia coli with plasmids," Gene, vol. 96, no. 1, pp. 23–28, Jan. 1990, doi: 10.1016/0378-1119(90)90336-P.

[19] R. Rahmer, K. Morabbi Heravi, and J. Altenbuchner, "Construction of a Super-Competent Bacillus subtilis 168 Using the PmtlA-comKS Inducible Cassette," Frontiers in Microbiology, vol. 6, Dec. 2015, doi: 10.3389/fmicb.2015.01431.

[20] R. W. Bradley, "An easy-to-use CRISPRi plasmid tool for inducible knockdown in E. coli," Biotechnology Reports, vol. 32, p. e00680, Dec. 2021, doi: 10.1016/j.btre.2021.e00680.

[21] J. Landberg, N. R. Wright, T. Wulff, M. J. Herrgård, and A. T. Nielsen, "CRISPR interference of nucleotide biosynthesis improves production of a single-domain antibody in Escherichia coli," Biotechnology and Bioengineering, vol. 117, no. 12, pp. 3835–3848, Dec. 2020, doi: 10.1002/bit.27536.

[22] K. Bach Falkenberg, "Optimizing production of industrial enzymes in Bacillus subtilis using synthetic biology," Master's Thesis, Technical University of Denmark, Kongens Lyngby, 2017.

[23] E. P. C. Rocha, A. Danchin, and A. Viari, "Translation in Bacillus subtilis: roles and trends of initiation and termination, insights from a genome analysis," Nucleic Acids Research, vol. 27, no. 17, pp. 3567–3576, Sep. 1999, doi: 10.1093/nar/27.17.3567.

[24] K. Bentele, P. Saffert, R. Rauscher, Z. Ignatova, and N. Blüthgen, "Efficient translation initiation dictates codon usage at gene start," Molecular Systems Biology, vol. 9, no. 1, p. 675, Jan. 2013, doi: 10.1038/msb.2013.32.

[25] D. B. Goodman, G. M. Church, and S. Kosuri, "Causes and Effects of N-Terminal Codon Bias in Bacterial Genes," Science (1979), vol. 342, no. 6157, pp. 475–479, Oct. 2013, doi: 10.1126/science.1241934.

[26] K. L. Griffith and A. D. Grossman, "Inducible protein degradation in Bacillus subtilis using heterologous peptide tags and adaptor proteins to target substrates to the protease ClpXP," Molecular Microbiology, Nov. 2008, doi: 10.1111/j.1365-2958.2008.06467.x.

[27] J.-H. Shin and C. W. Price, "The SsrA-SmpB Ribosome Rescue System Is Important for Growth of Bacillus subtilis at Low and High Temperatures," Journal of Bacteriology, vol. 189, no. 10, pp. 3729–3737, May 2007, doi: 10.1128/JB.00062-07.

[28] A. Kriel et al., "Direct Regulation of GTP Homeostasis by (p)ppGpp: A Critical Component of Viability and Stress Resistance," Molecular Cell, vol. 48, no. 2, pp. 231–241, Oct. 2012, doi: 10.1016/j.molcel.2012.08.009.

[29] A. Danchin, "Comparison Between the Escherichia coli and Bacillus subtilis Genomes Suggests That a Major Function of Polynucleotide Phosphorylase is to Synthesize CDP," DNA Research, vol. 4, no. 1, pp. 9–18, Jan. 1997, doi: 10.1093/dnares/4.1.9.

[30] M.-A. Petit, "The NAD-dependent ligase encoded by yerG is an essential gene of Bacillus subtilis," Nucleic Acids Research, vol. 28, no. 23, pp. 4642–4648, Dec. 2000, doi: 10.1093/nar/28.23.4642.

[31] C. Bruand, M. Farache, S. McGovern, S. D. Ehrlich, and P. Polard, "DnaB, DnaD and DnaI proteins are components of the Bacillus subtilis replication restart primosome," Molecular Microbiology, vol. 42, no. 1, pp. 245–256, Jul. 2008, doi: 10.1046/j.1365-2958.2001.02631.x.

4.6 Supplementary materials

4.6.1 Supplementary Table 1 – Oligos

Number	Name	Sequence
553	553_pSEVA_seq_rv	ccgagcgttctgaacaaatc
1402	1402_dCAS-nick-U-rv	ggtgcgaU ttagtcacctcctagctgactca
1431	1431_Seva-seq-fw	cggcggatttgtcctactca
1437	1437_dCas-seq1-rv	ttcttcttggctagctcccc
1438	1438_dCas-seq2-fw	ttggcgcaaattggagatca
1439	1439_dCas-seq3-rv	aactatcatcatggatcagctgc
1440	1440_dCas-seq4-fw	cgtcgccgttatactggttg
1441	1441_dCas-seq5-rv	tctcgccctttatcccagac
1442	1442_dCas-seq6-fw	tgagcaagaaataggcaaagca
1537	1537_pMazCr2-2.0-fw	gtctccccatgcgagagtag
1572	1572_tetR-seq-ekstra-fw	tcgcgatgacttagtaaagcac
1573	1573_tetR-seq-fw	aagcagctctaatgcgctgt
1823	1823_CB-22_seq_tetR_rv	cgcccagaagctaggtgtag
1824	1824_CB-23_seq_tetR_fw	tgcagagccagccttcttat
1845	1845_sfGFP-U-fw	ggcgaU cgc aggaggaatacat atgcgtaaaggcgaagagctgt
2074	2074_glmS-check-fw	TGAGCAGCCTGTTGTTATGC
2075	2075_glmS-check-rv	TGGAAACGGGAAGAGATTGT
2076	2076_yhgE-check-fw	GCTCAGAGCAATTGGTGACA
2077	2077_yhgE-check-rv	TGCATAGAAGCGGTGAAGAA
2078	2078_rsbP-check-fw	TGAGCTGCTGATAGCCAGAA
2079	2079_rsbP-check-rv	ACAGGCTGGGGTCAACTATG
2714	2714_dCas9-direct-U-fw	aggcgaU aggaggaatatac atggataagaaatactcaataggct
2797	2797_dCas-ttg-U-fw	ggcgaU aggaggaatatacc ttggataagaaatactcaataggcttagc
2798	2798_dCas-gtg-U-fw	ggcgaU aggaggaatatacc gtggataagaaatactcaataggcttagc
2811	2811_pyrG-direct-fw	aat aagtgaggatacaactcccc
2812	2812_pyrG-direct-rv	aac ggggagttgtatcctcactt
2813	2813_dnaI_direct-fw	aat ccctgcagggaacggccgat
2814	2814_dnaI_direct-rv	aac atcggccgttccctgcaggg
2815	2815_dnaB_direct-fw	aat ttaaccacatacgggtctac
2816	2816_dnaB_direct-rv	aac gtagacccgtatgtggttaa
2817	2817_cmk_direct-fw	aat ttttttctcagccacaattt
2818	2818_cmk_direct-rv	aac aaattgtggctgagaaaaaa
2821	2821_gmk_direct-fw	aat gttcctttaccaactcctga
2822	2822_gmk_direct-rv	aac tcaggagttggtaaaggaac

2823	2823_ligA_direct-fw	aat tattcggcatctgggacgct
2824	2824_ligA_direct-rv	aac agcgtcccagatgccgaata
3057	3057_sfGFP-CC-U-fw	ggcgaU aggaggaatatac atgcgtaaaggcgaagagctgttcacc GGTTTCGTCACTATTCTGGT
3059	3059_sfGFP-gRNA1-fw	aat caccagaatagtgacgaaac
3060	3060_sfGFP-gRNA1-rv	aac gtttcgtcactattctggtg
3072	3072_dCas-AAV-BS-U-rv	ggtgcgaU tta GAC TGC CGC agcaacattttgattaaatgaatttgttttgcctgc gtcacctcctagctgactca
3073	3073_dCas-LVA-BS-U-rv	ggtgcgaU tta CGC GAC AAG agcaacattttgattaaatgaatttgttttgcctgc gtcacctcctagctgactca
oYR76	oYR67_dCas9-seq-fw	tgcttcattaggtacctacca

4.6.2 Supplementary Table 2 – Plasmids

Name	Description	Source/Ref.
pProUSER13E3F	proUSER2.0 vector, *yhgE* homology, P_{tet}-spR	[12]
pProUSER13C1B	proUSER2.0 vector, *glmS* homology, P_{3P}-cmR	[12]
proUSER v8	Modified version of pProUSER13G2F, with cargo including SapI sites and tracrRNA, *rsbP* homology, P_{tet}-eryR	This work
pProUSER13F4A	proUSER2.0 vector, *sigF* homology, P_{veg}-kmR	This work
proUSER v8 pyrG	proUSER v8 with *pyrG* gRNA	This work
pSIJ388	pSEVA 441 tetR-P_{tet}-dCas9	Unpublished
pSIJ868	pProUSER13E3F + ATG-dCas9	This work
pSIJ874	pProUSER13E3F + TTG-dCas9	This work
pSIJ875	pProUSER13E3F + GTG-dCas9	This work
pSIJ599	pProUSER13C1B + mRuby	[12]
proUSER v8 gmk	proUSER v8 + *gmk* gRNA	This work
proUSER v8 LigA	proUSER v8 + *LigA* gRNA	This work
proUSER v8 cmk	proUSER v8 + *cmk* gRNA	This work
proUSER v8 dnaB	proUSER v8 + *dnaB* gRNA	This work
proUSER v8 dnaI	proUSER v8 + *dnaI* gRNA	This work
pSIJ774	pProUSER13F4A + sfGFP	This work
pSIJ768	proUSER v8 + sfGFP_gRNA1	This work

pSIJ890	pProUSER13E3F + dCas9-AAV	This work
pSIJ891	pProUSER13E3F + dCas9-LVA	This work

4.6.3 Supplementary Table 3 – Strains

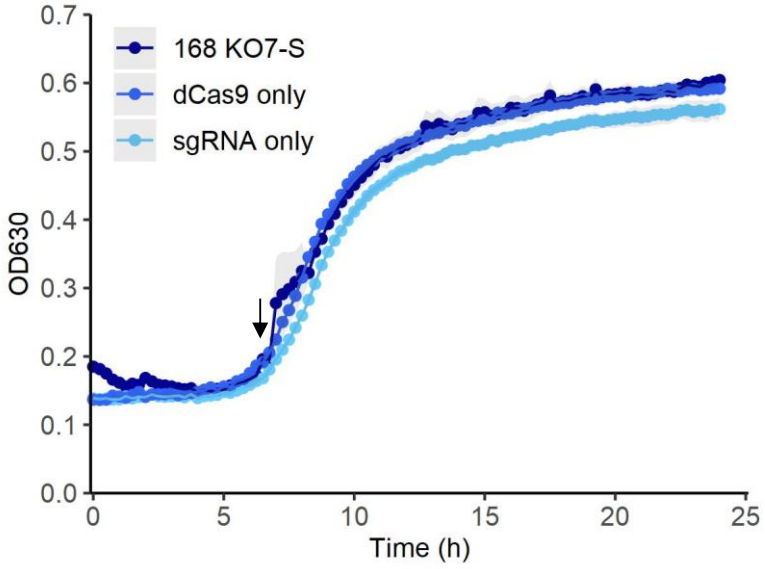
Strain	Genotype	Source/Ref.
168 KO7-S	*ΔnprE ΔaprE Δepr Δmpr ΔnprB Δvpr Δbpr trp*$^{+}$*ΔsigF glmS::pMtlA-comKS*	Unpublished data
sYR96	168 KO7-S *rsbP::*P_{tet}-pyrG_gRNA::eryR	This work
sYR97	168 KO7-S *yhgE*::P_{tet}-ATG_dCas9::spR	This work
sYR98	168 KO7-S *yhgE*::P_{tet}-TTG_dCas9::spR	This work
sYR99	168 KO7-S *yhgE*::P_{tet}-GTG_dCas9::spR	This work
sYR107	168 KO7-S *yhgE*::P_{tet}-ATG_dCas9::spR + *rsbP::*P_{tet}-pyrG_gRNA::eryR	This work
sYR108	168 KO7-S *yhgE*::P_{tet}-TTG_dCas9::spR + *rsbP::*P_{tet}-pyrG_gRNA::eryR	This work
sYR109	168 KO7-S *yhgE*::P_{tet}-GTG_dCas9::spR + *rsbP::*P_{tet}-pyrG_gRNA::eryR	This work
sYR122	168 KO7-S *glmS*::P_{3P}-mRuby::cmR	This work
sYR124	168 KO7-S *yhgE*::P_{tet}-TTG_dCas9::spR + *rsbP::*P_{tet}-pyrG_gRNA::eryR + *glmS*::P_{3P}-mRuby::cmR	This work
sYR138	168 KO7-S *yhgE*::P_{tet}-TTG_dCas9::spR + *rsbP::*P_{tet}-gmk_gRNA::eryR	This work
sYR139	168 KO7-S *yhgE*::P_{tet}-TTG_dCas9::spR + *rsbP::*P_{tet}-LigA_gRNA::eryR	This work
sYR140	168 KO7-S *yhgE*::P_{tet}-TTG_dCas9::spR + *rsbP::*P_{tet}-cmk_gRNA::eryR	This work
sYR142	168 KO7-S *yhgE*::P_{tet}-TTG_dCas9::spR + *rsbP::*P_{tet}-dnaB_gRNA::eryR	This work

sYR143	168 KO7-S *yhgE*::P_{tet}-TTG_dCas9::spR + *rsbP::*P_{tet}-dnaI_gRNA::eryR	This work
sYR147	168 KO7-S *sigF*::P_{veg}-sfGFP::kmR	This work
sYR154	168 KO7-S *sigF*::P_{veg}-sfGFP::kmR + *rsbP*::P_{tet}-sfGFP_gRNA1::eryR	This work
sYR157	168 KO7-S *yhgE*::P_{tet}-dCas9_AAV::spR	This work
sYR158	168 KO7-S *yhgE*::P_{tet}-dCas9_LVA::spR	This work
sYR163	168 KO7-S *yhgE*::P_{tet}-dCas9_AAV::spR + *rsbP::*P_{tet}-pyrG_gRNA::eryR	This work
sYR164	168 KO7-S *yhgE*::P_{tet}-dCas9_LVA::spR + *rsbP::*P_{tet}-pyrG_gRNA::eryR	This work

4.6.4 Supplementary Table 4 – Cloning efficiency

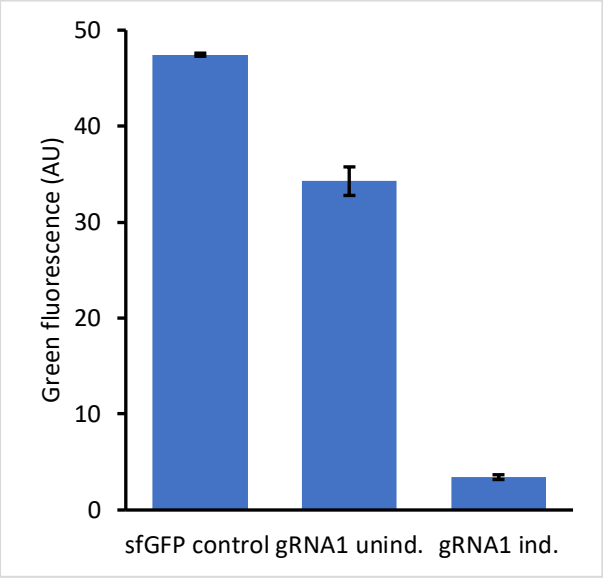
sgRNA	**Correct colonies**	**Total colonies**
pyrG	8	8
gmk	7	8
cmk	8	8
LigA	8	8
dnaB	7	8
dnaI	8	8
Total	46	48
Percent of correct colonies	95,8%	

4.6.5 Supplementary Figure 1 – Growth with dCas9 only and sgRNA only



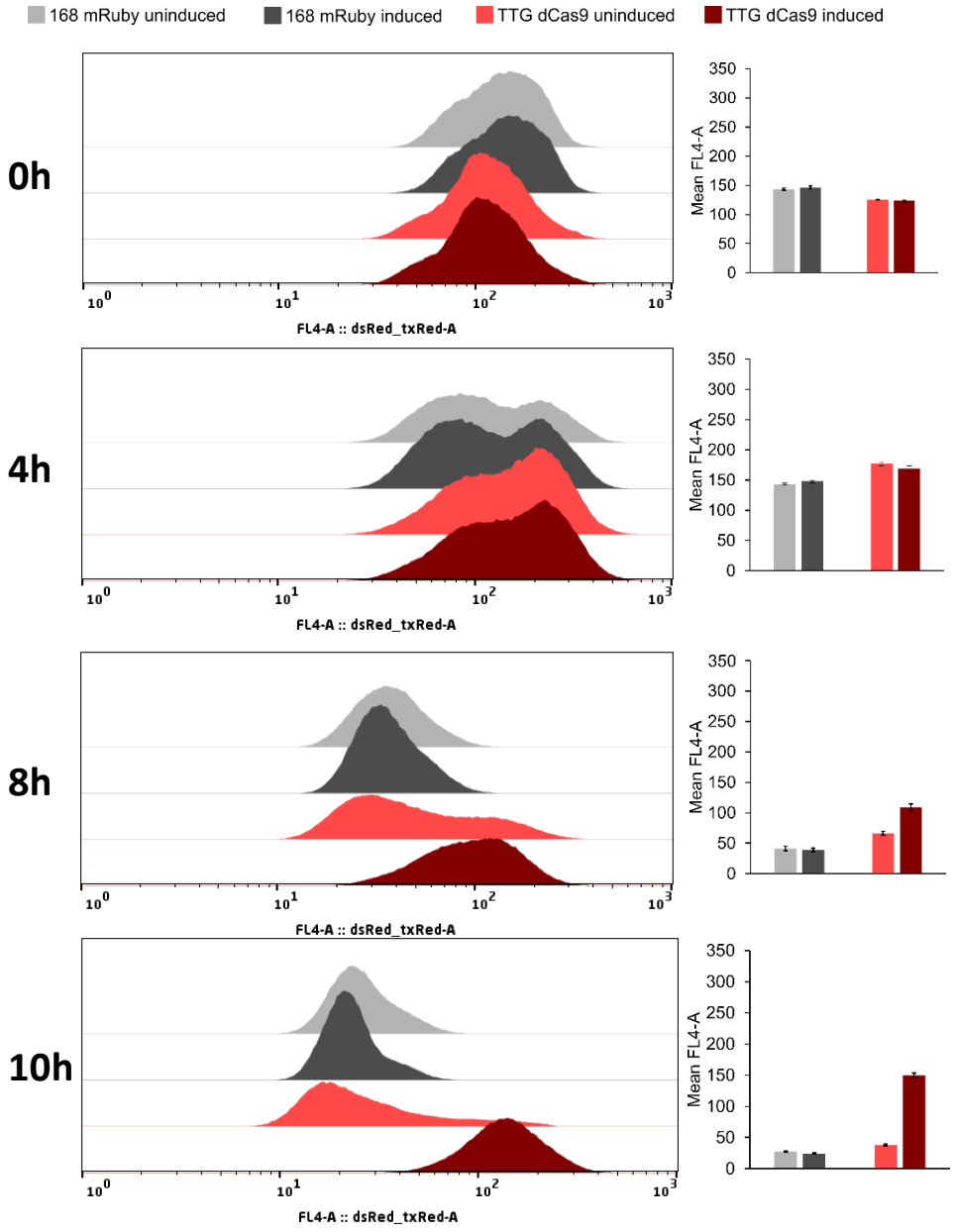
Supplementary Figure 1 *– Growth with parental strain 168 KO7-S, dCas9 only or sgRNA only. Measured in a plate reader with gray shading indicating mean ± s.d. (n=4 biological replicates). CRISPRi was induced at 7 hours with 0.1 µg mL^{-1} aTc, indicated by the arrow. Growth is not affected by expressing the wt dCas9 protein or the pyrG gRNA by themselves.*

4.6.6 Supplementary Figure 2 – Repression fold of TTG dCas9



Supplementary Figure 2 *– Measuring repression fold of TTG dCas9 on sfGFP expressed constitutively under the P_{veg} promoter, using flow cytometry. Sample was taken 8 hours after induction of CRISPRi with 0.1 µg mL^{-1} of aTc. gRNA is targeted at the non-template strand, 32 bp from the start of the ORF. Autofluorescence was measured based on parental strain with no fluorophore and subtracted from other fluorescence values. Error bars indicate mean ± s.d. (n=3 biological replicates for gRNA strain, n=2 biological replicates for controls).*

4.6.7 Supplementary Figure 3 – Additional time points from Figure 5B



Chapter 5

Folding and expression of heterologous proteins in bacterial cell factories

Yixin Rong[1], Sheila Ingemann Jensen[1], Kresten Lindorff-Larsen[2*], Alex Toftgaard Nielsen[1*]

[1] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, 2800 Kgs. Lyngby, Denmark.

[2] Structural Biology and NMR Laboratory, Department of Biology, University of Copenhagen, Ole Maaloes Vej 5, 2200 Copenhagen N, 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

Kresten Lindorff-Larsen

Structural Biology and NMR Laboratory, Department of Biology

University of Copenhagen

Ole Maaloes Vej 5, DK-2200 Copenhagen N, Denmark

Email: lindorff@bio.ku.dk

Abstract

The expression of correctly folded and functional heterologous proteins is important in many biotechnological production processes, whether it is enzymes, biopharmaceuticals or biosynthetic pathways for production of sustainable chemicals. For industrial applications, bacterial platform organisms, such as *E. coli*, are still broadly used due to the availability of tools and proven suitability at industrial scale. However, expression of heterologous proteins in these organisms can result in protein aggregation and otherwise low amounts of functional protein. This review provides an overview of the cellular mechanisms that can influence protein folding and expression, such as co-translational folding and assembly, chaperone binding, and protein quality control, across different model organisms. The knowledge of these mechanisms is then related to different experimental methods that have been applied in order to improve functional heterologous protein expression, such as codon optimization, fusion tagging, chaperone co-expression, as well as strain and protein engineering strategies.

5.1 Introduction

Biological production of various compounds has gained ground in the past few decades. For advances in medicine, therapeutic proteins have increasingly become the source of new bioactive compounds. Between 2014 and 2018 there were 72 new bioactive compounds approved by the FDA; out of those, 61 were therapeutic proteins [1]. The total market value for therapeutic proteins and peptides was predicted to be 222.7 billion USD in 2019 [2]. Although mammalian expression systems are used to produce the majority of therapeutic proteins, bacterial hosts such as Escherichia coli are also commonly used for proteins that do not require complex post-translational modifications. Out of the 316 therapeutic proteins currently produced, 88 are produced in *E. coli* due to the much simpler requirements for cultivation and the availability of tools for easy genetic modification [1][3]. In particular, therapeutic peptides between 30 and 100 amino acids in length, which make up 25.4 billion USD of the market, are preferentially produced in *E. coli* [2][3]. Outside of therapeutic proteins, about half of the industrial enzymes are produced in bacterial expression systems such as *Bacillus licheniformis* [4]. The industrial enzyme market is estimated to be worth 6.2 billion USD in 2020 [5]. Heterologous protein expression can also play a role in natural products and biochemical production, as biosynthesis pathways and enzymes from various organisms are expressed in the preferred platform organisms, such as *E. coli* and *Corynebacterium glutamicum* [6].

Although heterologous protein expression plays an important role in the production of everything from therapeutic proteins to biochemicals, it is often not a straightforward process to achieve the expression of a functional protein at the desired levels [7]. One of the key aspects to achieving a functional protein is the folding of the protein into its correct three-dimensional structure. This native structure is ultimately determined by the sequence of amino acids, and is generally the state where the protein has the lowest free energy [8]. The folding process itself is understood as a self-assembly process, wherein the protein goes from an unfolded state to the native state. The process of moving between the two states is often described as an energy landscape, consisting of all the possible intermediate states the protein can assume on its way to the native folded state [9]. This landscape contains an immense number of possible solutions for what intermediate states the protein can assume before reaching its native state. Constraints based on

molecular force fields can be applied to reduce the number of possible solutions, making it possible to simulate changes in protein structure. One such approach is called molecular dynamic (MD) simulations, and have been used to study mechanisms of protein folding/unfolding, ligand binding, and protein-protein interactions [10]. However, these simulations are limited by the requirements for computing power and the timescale for the simulation [10].

While MD-simulations can provide insight into the molecular mechanics and the dynamisms of protein structure, it is difficult to simulate folding in a cellular environment. It is known that the environment a protein folds in affects folding thermodynamics, kinetics, and fidelity, through factors ranging from temperature and pH to protein-protein interactions, and can determine whether the protein reaches and maintains the native structure [8]. Many of these environmental factors are changed when expressing heterologous proteins, as the protein experiences a different cellular environment than its native host organism. These changes in the environment can cause heterologous proteins to misfold or otherwise not to be expressed in a functional state. One example of such misfolding is the formation of protein aggregates known as inclusion bodies (IB) when expressing heterologous proteins in *E. coli* [11]. Various studies have been made to elucidate different factors and mechanisms behind protein folding, though often with the perspective of misfolding related diseases. Similarly, many studies have explored methods of improving heterologous protein expression, but the mechanisms leading to said improvements are not always clear. This review is structured to first provide an overview of the different cellular mechanisms that affect heterologous protein folding and how they differ between prokaryotic and eukaryotic organisms, and then to look at prior and state-of-the-art of methods that have been applied to obtain correctly folded heterologous proteins. By combining overviews of both the theory and applications, the goal is to help connect them and to inspire future studies towards improved heterologous protein expression.

5.2 Not all aggregates are made equal

The general goal when expressing heterologous proteins is to avoid or minimize aggregation/inclusion body formation. However, it is worth considering the advantages and disadvantages of having inclusion bodies, as they can have varying properties, such as size, purity, activity, and resistance to denaturing agents, and can even have direct applications themselves. IB formation may even be protective

in some protein misfolding diseases such as Huntington's; which could potentially be useful to sequester toxic proteins during production [12]. However, IB formation is unwanted for enzymes in biosynthetic pathways, as it is not possible to denature and refold the IBs.

In terms of factors that affect IB properties, it has been shown that lower temperature can increase the amount of functional proteins within aggregates [13][14]. Conversely, higher temperatures increase not only the fraction of protein found in inclusion bodies, but also the heterogeneity of protein structure within the inclusion body [15]. Other methods of modulating inclusion body properties include pH and feeding during fermentation [16][17]. Inclusion body properties can also depend on the protein itself; asparaginase inclusion bodies were shown to require a much lower concentration of urea to denature than inclusion bodies of human growth hormone [18]. Interestingly, inclusion bodies can generally be quite pure, with 80-95% of it being the heterologous protein, and the rest being other proteins such as chaperones DnaK, ClpB and IbpA/B [19][20]. Furthermore, up to 50% of the protein can be found in an active formation [21]. This has sparked interest in using them as a method of drug delivery; a study on soluble encapsulated versus inclusion body form of matrix metalloproteinase-9 showed higher in vivo activity in mice for the inclusion body [22]. With direct use of the inclusion bodies, it is not necessary to denature and refold, thus potentially streamlining the downstream processing steps. Another potential direct use is as carrier-free immobilized enzymes for biocatalysis, where fusion tags have been applied to induce catalytically active inclusion bodies, which could be stabilized by magnetization to prolong activity [23][24]. Lastly, if the product is of high value, it could be more economically advantageous to denature and refold the IBs, if obtaining soluble expression negatively affects the total protein produced and thus outweighs the increased downstream processing costs. While the understanding of inclusion bodies has shifted from the original concept of being essentially waste, due to denaturing and refolding still being time and resource intensive and unpredictable, obtaining correctly folded soluble proteins will likely remain the preferred outcome in most cases [17][25].

5.3 Physical and chemical factors' effect on protein folding

Protein folding can be disrupted by changes in physical and chemical conditions, such as temperature, pH, and the redox conditions [26]. Higher temperatures have been shown to increase misfolding and aggregation with heterologous protein production in E. coli, resulting in the formation of aggregates in the form of inclusion bodies [27]. Temperature can affect both the host organism and the protein itself. Energetically, heat induced destabilization of the native protein structure is caused by the increased thermal fluctuations, resulting in particular changes of the hydrophobic effect and the native structure no longer being the most energetically favorable [28]. Cold denaturation from lowering of temperature also depends on the strength of the hydrophobic interactions in a protein being weaker at lower temperatures [29]. Cold denaturation does not usually occur in production processes, as most proteins cold denature at temperatures below 0 °C [30]. Heat denaturing can pose problems to production of proteins such as enzymes derived from psychrophilic microorganisms, which are evolved to be active under very low temperatures by having weaker intramolecular forces and thus greater flexibility at low temperatures. These enzymes can unfold at temperatures of 40 °C or lower [31]. Temperature can also affect certain metastable human proteins such as serine protease inhibitors (serpins) [32]. Indirect effects of temperature include the general protein synthesis and macromolecular levels, where a less crowded cellular environment is favorable for protein folding [27]. Changes in temperature also affect the expression of proteins involved in protein folding modulation and degradation of misfolded proteins, this will be described in more detail in a later chapter [33].

Other changes in the chemical properties of the folding environment that affect protein folding include pH and redox environment. Changes in pH mainly affects protein folding and structural stability by protonation/deprotonation of charged amino acid residues, which can affect charged interactions such as salt bridges [34]. Out of the charged amino acids, histidine is the most prone to a change of protonation with a pKa value of 6, whereas the other positively charged residues of lysine and arginine have pKa values of 10.5 and 12.5 respectively. The negatively charged residues of aspartate and glutamate have pKa values of 3.7 and 4.3 [35]. These pKa values can vary, especially for buried residues [36]. Oxidative stress may be caused by the presence of reactive oxygen species (ROS), which can be formed during e.g. cellular respiration or disulfide bond formation [37][38]. ROS can affect

protein folding by oxidizing cysteine residues and creating non-native disulfide bonds, which can lead to misfolding and aggregation [39].

5.4 Prokaryotic and eukaryotic chaperones and protein quality control

Maintaining correctly folded proteins is crucial for cellular health, as misfolding and aggregation of proteins disrupt cellular functions and can ultimately lead to cell death [40][41]. Therefore, all organisms have mechanisms to promote the correct folding of proteins and to degrade misfolded proteins, mediated by a network of different proteins including chaperones and proteases [42]. These protein quality control (PQC) mechanisms help to maintain a state of balanced and functional proteome in the cell, also referred to as proteostasis [43]. Although many aspects of these systems are conserved, there are differences between organisms, especially between prokaryotic and eukaryotic organisms. This chapter will provide a broad overview of chaperones and PQC of both prokaryotic and eukaryotic organisms, and how the differences may affect heterologous protein folding.

5.4.1 Bacterial chaperones

Bacterial cells are less complex than eukaryotic cells, as they essentially have the cytosol as the only main compartment. Apart from the cytosol, proteins can also be targeted to the membrane or be secreted to the extracellular environment. Within Gram-negative bacteria, proteins can additionally be targeted to the periplasm and the outer membrane. Regardless of the destination and trafficking pathway, newly synthesized proteins interact with various chaperones that can assist in folding, unfolding and refolding of any misfolded proteins or help target irreversibly misfolded proteins for proteolytic degradation [44]. Many molecular chaperones belong to a type of proteins called heat shock proteins (Hsp), due to their initial discovery as proteins that make up a major part of the cellular response to heat shock. These Hsp's are further divided into different families based on their approximate molecular weight, such as Hsp60, Hsp70, Hsp90 etc. [45]. An overview of bacterial chaperones, their families and functions, as well as proteases involved in clearing misfolded proteins can be seen in Table 1.

Cytosolic chaperones are the most well studied ones in literature, likely due to them being much easier to purify compared to membrane proteins, which makes it possible to determine the structure of the protein as well as to perform various in-

vitro studies. The main cytosolic chaperones in bacteria are the GroEL-GroES system and DnaK-DnaJ-GrpE system [43]. Other chaperones include trigger factor (TF), ClpB, HtpG and IbpA/IbpB [46]-[49].

GroEL belongs to the Hsp60 family of chaperones, and forms a protein complex consisting of two heptameric rings stacked on top of each other, forming a chamber in the middle [50]. This chamber can be closed off by its co-chaperone GroES, which also forms a heptameric ring. The GroEL-ring that is bound to the GroES-ring is referred to as the cis-ring, with the other GroEL-ring being the trans-ring [50]. The substrate protein binds to residues in the GroES-ring, though the mechanism of it initially entering the GroEL-ring is unclear [51]. GroEL can bind proteins and peptides of various sizes; while there are no known consensus sequences or structural motifs, GroEL preferentially binds exposed hydrophobic regions [52][53]. The main function of the chaperone system is thought to be to capture folding intermediates or misfolded proteins, help unfold them and then provide a folding environment isolated from the bulk solvent of the cell. Each GroEL-subunit can bind ATP; the hydrolysis of ATP to ADP likely provides the necessary energy to unfold and refold the protein. ATP hydrolysis also prepares the trans-ring to be bound to GroES, and ultimately leads to the release of the GroES, ADP and substrate protein alike [50][51].

DnaK is a chaperone belonging to the Hsp70 family, and prevents aggregation and assists in the folding of misfolded or partially misfolded proteins by an ATP-dependent bind and release cycle [54]. More specifically, DnaK is thought to bind exposed hydrophobic surfaces of proteins [55]. DnaK also interacts with nascent chains by binding to them and preventing the formation of non-native intermolecular interactions [56]. Like GroEL, DnaK also acts in coordination with accessory proteins; namely DnaJ and GrpE, which are co-chaperones and nucleotide exchange factors (NEF) respectively. DnaJ increases the ATPase-activity of DnaK when bound, while GrpE promotes the disassociation of ADP from DnaK [57][58].

DnaK is also known to act on protein aggregates together with another chaperone, ClpB. ClpB belongs to the Hsp100 family of chaperones, and has been shown to act as a disaggregase on stable protein aggregates [45]. The bi-chaperone system works by DnaK first extracting polypeptides from an aggregate, after which ClpB

disaggregates the polypeptide; the resulting disaggregated chains are then refolded by DnaK [59].

HtpG is a chaperone belonging to the Hsp90 family; the exact function of this chaperone is not entirely clear, although it is also known to act through an ATP and ATPase mediated binding and release cycle like DnaK, and HtpG deletions lead to slowed growth at temperatures higher than 42 °C [47][60]. HscA and HscB are specialized chaperones in the Hsp70 family that are involved in the synthesis of iron-sulfur cluster containing proteins by transferring the iron-sulfur clusters to acceptor proteins [61]. IbpA and IbpB are small Hsp's which can help recruit the major chaperones and prevent aggregate formation; the absence of them lead to larger aggregates being formed [62]. They have also been shown to delay the degradation of inclusion bodies [63]. Lastly there is trigger factor, which is classified as a protein prolyl isomerase (PPIase). Trigger factor has been shown to bind to nascent chains like DnaK, although trigger factor binds to shorter chains than DnaK [46].

Apart from their role in assisting in the proper folding of cytosolic proteins, chaperones also interact with proteins in other targeting pathways. Some pathways have dedicated chaperones; e.g. SecB in the Sec-pathway for secreting proteins is a chaperone, which binds proteins and prevents them from folding before being secreted [64]. GroEL, trigger factor and DnaK have all been shown to act on proteins secreted via the Sec-pathway, and possibly also on proteins secreted via the Tat (twin-arginine translocation) pathway, although the mechanisms of the interaction with the Tat-pathway remains unclear [65]. The Sec-pathway likely involves more chaperones due to it secreting unfolded proteins, whereas the Tat-pathway secretes folded proteins [66]. DnaK can also be recruited to the membrane by an alternative co-chaperone DjlA, where it is involved in the insertion of membrane proteins, although it is not known if the membrane localization is needed for it to act on membrane protein insertion [65][67]. Gram-negative bacteria also have chaperones in the periplasm; these include the PPIases SurA and FkpA, as well as the chaperones Skp and Spy [68]. Both Skp and Spy are ATP-independent chaperones, and have been shown to unfold and refold partially misfolded proteins, as well as preventing outer membrane proteins from folding prior to membrane insertion [69][70].

Table 1 *Overview of bacterial chaperones and proteases*

Protein name	Family	Localization	Function
GroEL	Hsp60	Cytoplasm	ATP-dependent chaperone, assists in protein folding, involved in secretion
GroES	Hsp10	Cytoplasm	Co-chaperone for GroEL
DnaK	Hsp70	Cytoplasm	ATP-dependent chaperone, assists in protein folding, binds nascent chains, involved in secretion and membrane protein insertion
DnaJ	Hsp40	Cytoplasm	Co-chaperone for DnaK
GrpE	-	Cytoplasm	Co-chaperone for DnaK
ClpB	Hsp100/AAA+	Cytoplasm	ATP-dependent chaperone, disaggregase, acts with DnaK
HtpG	Hsp90	Cytoplasm	ATP-dependent chaperone, involved in responding to heat shock induced misfolding
HscA/B	Hsp70	Cytoplasm	ATP-dependent chaperone, involved in synthesis of FE/S cluster containing proteins
IbpA/IbpB	sHsp	Cytoplasm	ATP-independent chaperone, prevents protein aggregation
Trigger factor	PPIase	Cytoplasm	Binds to short nascent chains, involved in secretion
SecB	-	Cytoplasm	Prevents proteins designated for the Sec-pathway
SurA	PPIase	Periplasm	Lower outer-membrane protein levels when absent
FkpA	PPIase	Periplasm	Chaperone-like activity with some recombinant proteins
Skp	-	Periplasm	Chaperone, shields outer-membrane proteins prior to membrane insertion
Spy	-	Periplasm	Chaperone, shields outer-membrane proteins prior to membrane insertion
ClpP	Protease	Cytoplasm	Degrades unfolded proteins with ClpA/ClpX

ClpA/ClpX	AAA+	Cytoplasm	ATP-dependent chaperone-like functions, binds unfolded protein to release or degrade by ClpP, degradation of regulators (housekeeping)
HslVU (ClpQY)	Protease	Cytoplasm	Degradation of regulators (housekeeping) and misfolded proteins
Lon	Protease	Cytoplasm	Degradation of regulators (housekeeping), naturally unstable proteins and misfolded proteins
FtsH	Protease	Cytoplasmic membrane	Quality control of membrane proteins, degradation of regulatory proteins
RseP	Protease	Cytoplasmic membrane	Regulation of cellular responses by degrading membrane bound regulators
DegP	Protease	Periplasm	Bind and release or degrade misfolded proteins, shields outer-membrane proteins prior to membrane insertion

5.4.2 Bacterial proteases and protein quality control

Not all misfolded or aggregated proteins can be rescued by chaperones; these need to be degraded by various proteases in the cell. The main proteases in the cytosol are the ClpAP and ClpXP complexes, alongside with Lon [45][71][72]. ClpP is a protease that requires the ATP-dependent chaperones ClpA and/or ClpX in order to degrade proteins that have been tagged for proteolysis [73]. The active site of the ClpP complex is hidden away inside a barrel-shape formed by two heptameric rings, where only small peptides can enter the pore from the ends [73]. ClpA and ClpX both form hexameric rings that can bind either end of the ClpP-barrel, and can recognize certain N- and C-terminal tags, leading to unfolding of the tagged protein [73][74]. The target protein can then either be fed into ClpP to be degraded, or released from the chaperone in an unfolded state, ready to be refolded [75]. ClpA and ClpX target regulators related to functions such as cell cycle and stress response, such as FtsZ, RepA, RseA and RpoS [75]-[78]. Another protease with a similar mechanism, HslVU/ClpQY, has also been shown to degrade regulatory proteins, such as FtsZ inhibitor SulA, as well as heterologous proteins [79]. In addition, proteins can be

targeted to ClpA via N-degrons or to ClpX via tagging of a short amino acid sequence by ssrA [77][80]. SsrA is an RNA structure that can recognize stalled ribosomes, and attach an 11 amino acid long sequence to the nascent chain; this tag gets recognized by ClpX, in some cases with the help of adaptor protein SspB, and the incomplete protein gets degraded [77].

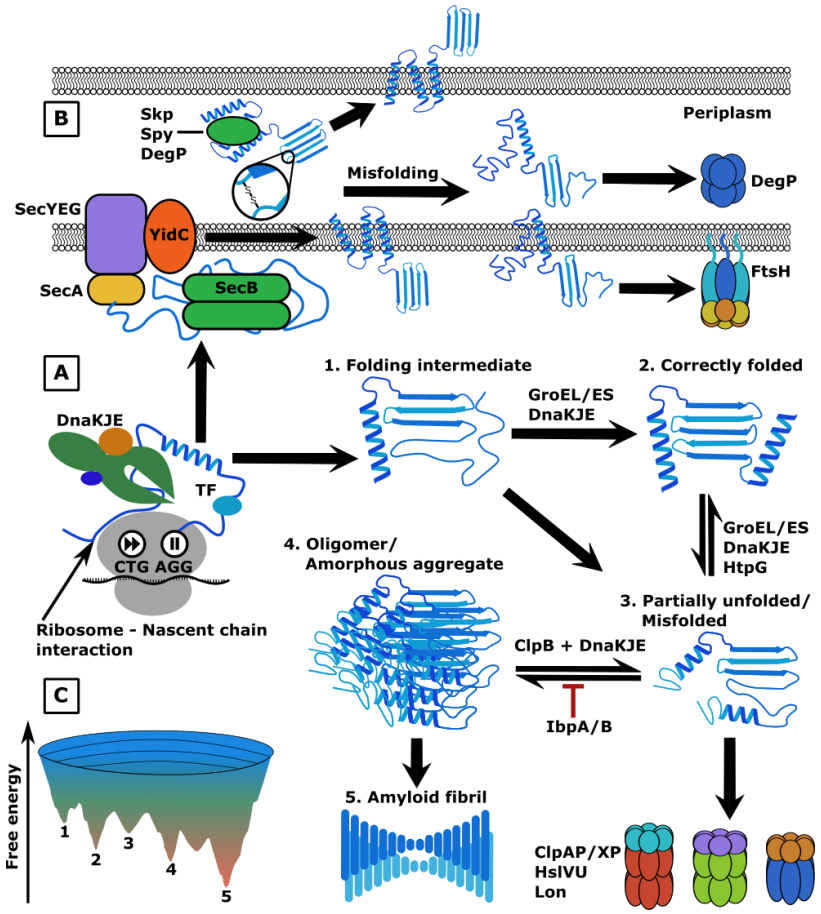


Figure 1 *Protein folding in bacteria. A) Folding starts during translation, where chaperones DnaKJE and TF and the ribosomal surface can act as holdases. Folding is also affected by codon usage. Subsequently, the protein can assume different folding states, where chaperones assist it in reaching and maintaining the correct fold. Misfolded proteins get degraded by proteases ClpAP/XP, HslVU and Lon. B) Alternatively, the protein can get secreted to the periplasm or inserted to the inner membrane (IM) via the Sec pathway. Periplasmic chaperones protect transmembrane regions of outer membrane proteins. Disulfide bonds also form in the periplasm. Misfolded proteins get degraded by FtsH and DegP. C) Free energy of the different folding states.*

Degrons act as signals for degradation, usually consisting of a single AA residue or a short sequence of AA residues on either terminal of a protein [81]. ClpS is an adaptor for ClpA that recognizes N-terminal hydrophobic residues L, F, W and Y that can be exposed through cleavage by other proteases or the removal of the initiating M, or an L that is added to N-terminal charged residues by Aat or Bpt [80]. ClpS can also target proteins with the initiating M intact and a hydrophobic residue at the 4th position [80]. Additionally, failure of PDF deformylase to deformylate formyl-methionine during translation has been indicated to act as a N-degron as well [82]. There are some differences depending on the organism in question. The Gram positive bacteria Bacillus subtilis has a different ssrA tag sequence and no SspB orthologue. Its ClpX also recognizes C-terminal poly-alanine tags, which get added onto dissociated 50S ribosomal subunit complexes with the nascent chain still attached by the quality control proteins RqcH and RqcP [83][84]. The ATPase ClpA and its adaptor protein ClpS are replaced by ClpC and MecA instead [85].

Unlike the Clp-protease complex, other proteases such as Lon have both the substrate recognition site and protease site on the same polypeptide chain [72]. Lon is also involved in degradation of many naturally unstable regulatory proteins, which are involved in various physiological processes in bacterial cells. Lon also degrades aggregation prone, misfolded proteins and has overlap in substrate recognition with other proteases such as Clp, and although Lon was shown to bind sequences rich in aromatic amino acids, no consensus sequence or motif for Lon-binding has been discovered [72][86]. Outside of cytosolic proteases, bacteria also have inner membrane anchored proteases such as FtsH and RseP (previously known as YaeL or EcfE) [87][88]. FtsH is a cytoplasmic membrane-anchored protease that is structured as a homo-hexameric pore, and mainly functions as quality control for membrane proteins [87]. It has been shown to degrade membrane proteins such as dissociated ATP-synthase subunits and SecY, and it is also involved in degradation of certain regulatory proteins, such as the sigma-factor rpoH, which activates genes involved in heat shock response (HSR) [87]. RseP is a protease that has mainly been implicated in the regulation of certain extracellular responses by cleaving the transmembrane regions of membrane-anchored anti sigma-factors, such as with rpoE in *E. coli* and the lut system in *Pseudomonas putida*, although it is unknown if RseP acts on other targets as well [88][89]. Lastly, DegP is a protease and chaperone located in the periplasm of Gram negative bacteria. DegP is multimeric, being in a

hexamer structure while inactive, and in a 12mer or 24mer structure when active [90]. The hexamer form is able to bind misfolded proteins and reassemble into the 12- or 24mer form, which is able to degrade terminally misfolded proteins, or release proteins that can refold [90]. Like Skp and Spy, DegP can also shield the hydrophobic membrane anchors of outer membrane proteins while in the periplasm [90]. Figure 1 shows different routes a heterologous protein can take and where chaperones and proteases would act on it.

5.4.3 Eukaryotic chaperones

Eukaryotic organisms also possess many chaperones belonging to the same families as those in bacteria. However, the presence of organelles necessitates the ability of the cells to detect and facilitate a response to misfolded proteins in many different compartments. Unlike bacterial cells, eukaryotes can synthesize new proteins either in the cytosol or directly translocate the nascent chain into the endoplasmic reticulum (ER). On the other hand, many eukaryotes have been shown to have several homologs of Hsp70 family chaperones in the cytosol: *Saccharomyces cerevisiae* is known to have four Hsp70 homologs, where Ssa1 and Ssa2 are constitutively expressed, and Ssa3 and Ssa4 are expressed under heat shock [91]. Deletion of Ssa1 and 2 leads to slow growth and sensitivity to heat, indicating that they have non-redundant functions compared to Ssa3 and 4 [92]. Mammals also have genes for several functionally different Hsp70 chaperones in their genome; overexpression of some Hsp70 chaperones gives increased thermotolerance, while others do not [93]. The overall binding and release mechanism is conserved, where ATP hydrolysis induced by a J-protein/Hsp40 family co-chaperone causes Hsp70 to change to a closed conformation, binding to an unfolded protein [94][95]. In S. cerevisiae, Sis1 and Ydj1 are the primary J-proteins, and in humans as many as 49 genes coding for J-proteins have been identified [91][96]. Many protein aggregation diseases in humans have been linked to mutated J-proteins, although whether the link between specific J-proteins and diseases is due to tissue or target specificity of the J-proteins is not currently known [96]. The exchange of ADP with ATP then promotes opening of Hsp70 and the release of the substrate, which is mediated by Bcl2-associated athanogen or BAG-proteins in eukaryotes [97]. A study has shown differences in the allosteric state of the chaperone in bacteria vs. eukaryotes; while DnaK from E. coli shifts from open to partially closed upon substrate binding, the

human Hsp70, HspA1, requires both substrate binding and ATP-hydrolysis to shift to a closed state [94].

Another major cytosolic chaperone is Hsp90. The eukaryotic Hsp90 is an ATP-dependent chaperone, structured as a homodimer [98]. It has a different open and close allosteric mechanism than Hsp70; the nucleotide binding domains of each subunit interact and form the closed conformation upon the binding of ATP, and returns to the open conformation upon ATP-hydrolysis and the dissociation of ADP [99]. Like Hsp70, Hsp90 is reliant on co-chaperones to induce ATP-hydrolysis and nucleotide exchange; it has however been shown that these co-chaperones act in a target specific manner [98]. Hsp90 can also act in coordination with Hsp70 through an adaptor protein Sti1/Hop, where Hsp90 can be directed towards a protein that is first be processed by Hsp70 [99][100]. Interestingly, metazoans have one major difference compared to bacteria, fungi and plants in that they do not have a homolog of the Hsp104 chaperone/AAA+ ATPase (ClpB in *E. coli*), which is tasked with disaggregation of protein aggregates in those organisms [101]. A study has shown that the mammalian Hsp110 (Apg-2, or Sse1 in yeast) can partially complement the disaggregase function together with Hsp70 chaperones and Hsp40 co-chaperones; it was however unable to dissolve certain prion-based aggregates, where Hsp104 was able to act on the aggregates [101].

Eukaryotes also have ATP-independent chaperones in the form of sHsp's; S. cerevisiae has two sHsp's in the cytosol in the form of Hsp42 and Hsp26, whereas the human genome contains ten sHsp's named HspB1 to HspB10 [102][103]. The two yeast sHsp's both show ability to prevent aggregate formation in vitro, much like their bacterial equivalents. Deleting either one or the other does not result in altered growth phenotypes, suggesting a degree of complementarity; however the in vivo activity differs in that Hsp42 can also promote aggregate formation, while Hsp26 only shows the typical holdase activity and prevents aggregation [103]. The aggregase activity of Hsp42 could at as a mitigation mechanism against toxic protein aggregates, as larger aggregates have a comparatively smaller exposed surface area [104]. On the other hand, Hsp26 deletion mutants show slower refolding mediated by Hsp70/Hsp100 whereas Hsp42 deletion mutants do not, indicating that Hsp26 is better at facilitating refolding [103]. Out of the human sHsp's, HspB1, B5, B6 and B8 are ubiquitous, and the others are tissue specific [102]. Like the yeast Hsp42, HspB1 has been shown to be able to mitigate the toxicity of protein aggregates by

sequestering them to larger aggregates [105]. HspB1 as well as HspB5 are also known to be able to prevent the formation of aggregates, where overexpression reduces intracellular α-synuclein [106]. HspB6 has also been indicated to display the classical sHsp function of preventing aggregate formation [107]. HspB8 can form a complex with Bag3 and Hsp70, where this complex was shown to disassemble stress granules, which are stress-induced ribonucleoprotein complexes [108]. Interestingly, these ten sHsp's have been shown to form both homodimers, homo-oligomers and different hetero-oligomers with each other. They can all form homodimers and small homo-oligomers with the exception of HspB1 and HspB5, which can form larger multimers [109]. Some of the sHsp's can also form hetero-oligomers, and the known combinations are shown in Figure 2. The composition of the hetero-oligomers could be affected by the expression levels of the different sHsp's [110]. The presence of both homo- and hetero-oligomers and of different sizes results in a large number of combinations with potentially different target specificities.

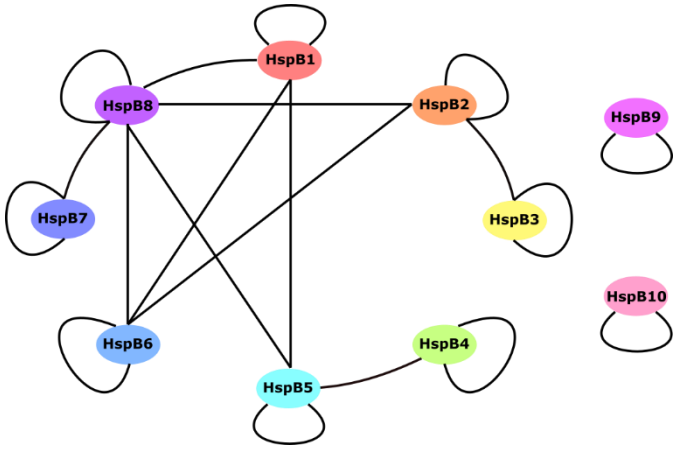


Figure 2 All the known combinations of human/mammalian small heat shock protein multimers [82]. All 10 can form homo-dimers, as indicated by the looped lines, as well as certain hetero-dimers, indicated by straight lines

In eukaryotes, proteins targeted for secretion or the cell membrane get trafficked through the ER. The main chaperones in the ER include BiP (Immunoglobulin Binding Protein), GRP94, calnexin (CNX) and calreticulin (CRT) [111]-[113]. BiP is a Hsp70-family chaperone located in the lumen of the ER, and like other Hsp70 chaperones is ATP-dependent and requires co-chaperones and nucleotide exchange factors to perform its function [114]. For mammalian cells, 7 ER-localized J-proteins named

ERdj1-7 have been identified, as well as the two NEF's Grp170 and Sil1 [111][114]. ERdj1, ERdj2 (also referred to as Sec63) and ERdj3 are involved in recruiting BiP to the Sec61 translocon, through which the nascent peptide gets synthesized [114][115]. Here BiP can bind to hydrophobic patches on the nascent chains to prevent them from prematurely folding, and has also been proposed to act as a molecular motor, promoting the one-directional translocation of the nascent chain [116][117]. ERdj4-7 are thought to act as adaptor proteins, enabling BiP to interact with specific target proteins [114]. Interestingly, the NEF Grp170 has been shown to exhibit chaperone activity itself, with ERdj5 as a co-chaperone [115]. ERdj5 also has disulfide reductase activity, and promotes degradation of misfolded proteins [118]. GRP94 is an ER localized Hsp90 chaperone, where it assists in protein folding, albeit acting in a more substrate specific manner compared to BiP[113]. Outside of classical chaperones, the ER also has the lectins calnexin and calreticulin that have chaperone activity. They bind to the N-glycosylation moiety Glc1Man9GlcNAc2 of glycoproteins and assist in folding and recognizing misfolded proteins [119].

Besides the ER, chaperones are also present in metabolic organelles such as the mitochondrion and the chloroplast of plants. Due to mitochondria likely originating as a symbiotic prokaryote, they have homologs of the core chaperones from bacteria. Hsp60-Hsp10 is a GroEL-GroES homolog, Ssc1-Mdj1-Mge1 is a DnaK-DnaJ-GrpE homolog, TRAP-1 is a HtpG homolog and Hsp78 is a ClpB homolog [120]–[123]. Although these chaperones' function is mostly conserved, Ssc1 has acquired a new function in eukaryotic mitochondria via eukaryote only co-chaperones Pam16, Pam18 and Tim44, which enable it to act as a motor for translocation of membranes into the mitochondrial matrix [121]. Other proteins unique to mitochondria include the Tim9-Tim10 chaperone-like complex, which also acts on protein translocation [124][125]. Chloroplast enzymes can be interesting for heterologous expression, as the biosynthesis of plant compounds such as monoterpenes, diterpenes and carotenoids takes place in the chloroplast [126]. Cpn60 is a GroEL homolog known to be involved in the folding of Rubisco, although unlike GroEL, the chloroplast Cpn60 assembles as a hetero- oligomeric complex of both α and β subunits [127]. Hsp70B-CDJ2-CGE1 is a homolog of DnaK-DnaJ-GrpE, Hsp90C is a HtpG homolog, and Hsp93 is a ClpB homolog [128]–[130]. Like the mitochondria, the chloroplast has lost much of its genome during evolution, and has to import the majority of its proteins from the cytosol. But unlike in mitochondria, it is suggested that Hsp70B,

Hsp90C and Hsp93 are all involved in protein import, although consensus has not been reached regarding the exact mechanism and the co-chaperones required [131].

5.4.4 Protein quality control in eukaryotes

Eukaryotes have many PQC pathways, and the majority of these lead to degradation by the ubiquitin proteasome system (UPS). Ubiquitylation is performed by the E1-E2-E3 enzyme cascade, where E1 activates and transfers ubiquitin to E2, E3 ligase recognizes and binds the substrate and recruits the E2 conjugating enzyme, which ubiquitylates the substrate [132]. The specificity for proteasomal targeting is in part mediated by the large number of different ubiquitin E3 ligases, which transfer ubiquitin from the E2 carrier to the substrate. The number of different E3 ligases differs based on the organism; yeast has between 60 and 100 E3 ligases, while humans have more than 600 putative E3 ligases [133][134]. The specificity comes from the substrate specificity of E3 ligases themselves, but also from co-localization via recruitment by adaptor proteins.

As with prokaryotes, PQC starts already during translation; stalled ribosomes can get detected and rescued, and the incomplete nascent chain gets targeted for degradation. This quality control process is mediated by the ribosome-associated protein quality control complex RQC, where Rqc2 binds to the large ribosomal subunit and a peptidyl-tRNA in the P site; a conformation that is only present upon stalling induced dissociation of the two ribosome subunits, but does not bind intact ribosomes [135]. The complex can act on both cytosolic and ER/mitochondrial membrane associated ribosomes. Rqc2 recruits the ubiquitin E3 ligase Ltn1/listerin, leading to ubiquitylation of the nascent chain. Cdc48 can then free the ribosome from the nascent chain, which gets degraded by the proteasome [135]. If no lysine residue is available, Rqc2 can add alanine and threonine residues in a process called CATylation, until lysines potentially in the exit site of the ribosome get exposed; if ubiquitylation fails all together, the CATylated nascent chains get released and can potentially aggregate [135].

Cytosolic protein quality control in eukaryotes starts by binding of cytosolic chaperones. Hsp70 chaperones bind to short sequences of exposed hydrophobic residues and/or positively charged residues in misfolded or unfolded proteins, in some cases assisted by Hsp40/J-domain proteins with regards to substrate

specificity and promoting ATPase activity [136][137]. The Hsp70-Hsp40-substrate complex can then be transferred to Hsp90 by the adaptor protein HOP/Sti1 to further facilitate refolding [138]. If the substrate protein remains misfolded, it can either be trafficked to a subcellular compartment, or get an ubiquitin moiety attached by an ubiquitin E3 ligase and get targeted for degradation by the proteasome. Aggregation prone proteins such as amyloids get trafficked to the insoluble protein deposit IPOD, located at the vacuole [139]. The other compartments for misfolded proteins are the juxtanuclear quality control compartment JUNQ and the intranuclear quality control compartment INQ, where the targeting involves sHsp's and co-chaperones to the main Hsp70/90 chaperones [140]. Proteasomal targeting involves the co-chaperone CHIP and the NEF BAG1. CHIP can associate with Hsp70 and Hsp90, blocking ATP-hydrolysis and protein refolding, while also ubiquitylating the substrate. BAG1 can then bind Hsp70, associate with the proteasome via its ubiquitin-like domain and promote substrate release by nucleotide exchange [141]. Hsp70 has also been implicated in lysosomal degradation via targeting to macroautophagy, where it can form a complex with BAG3 and HspB8 [141].

Secreted proteins and membrane proteins go through the ER, where the ER chaperones are responsible for quality control. Calnexin and calreticulin act on glycoproteins in a cyclical manner. They bind the N-glycan moiety and dissociate when glucosidase GII cleaves the remaining glucose; if the protein is natively folded, it can exit the ER towards the Golgi apparatus, whereas misfolded proteins can get a glucose re-added by a glucosyltransferase, promoting re-binding of CNX/CRT [112]. Studies have also suggested calnexin playing a role in glycan-independent protein quality control of transmembrane domains (TMD) of membrane proteins [142][143]. If the protein resides in the ER for a prolonged period, it can get a mannose removed by a mannosidase; the $Man_8GlcNAc_2$ moiety is recognized by the mannosidase EDEM, which can target the protein to ER-associated degradation (ERAD) [119]. Here the protein gets retrotranslocated by the retrotranslocon and ubiquitylated and thus targeted to the proteasome [144]. GRP94 has also been indicated to be able to target ERAD, although the mechanism is currently unknown [113]. The human BAG6 chaperone can bind exposed hydrophobic regions of tail-anchored membrane proteins that fail to insert, recruit the E3 ligase RNF126, and bind the proteasome to target the substrate for degradation [145]. Once integrated

into the target membrane, membrane proteins can get targeted for degradation when they undergo conformational changes, e.g. from heat stress, and binding motifs for E3 ligases and their adaptor proteins get exposed [146]. In yeast, the E3 ligase Rsp5 has been discovered to be present at both the plasma membrane and intracellular membranes, where it helps ubiquitylate membrane proteins [146].

5.4.5 Eukaryotic degron systems

N-degron targeting is generally divided in three main pathways: the Arg/N-degron, the Ac/N-degron and the Pro/N-degron [81]. The Arg/N-degron pathway recognizes N-terminal unacetylated amino acids M*, L, F, I, Y, W, D, E, R, K and H, as well as N and Q if they get converted to D or E via a N-terminal amidase (* being a large hydrophobic residue) [147]. They get recognized by the UBR family of E3 ligases; in the case of D and E, they first get arginylated by an arginine transferase [147]. The Ac/N-degron pathway recognizes N-terminal acetylated amino acids M*(*: large hydrophobic residues F, W, Y, I, L), MX (X: acidic/hydrophilic residues D, E, N, Q) and MZ (Z: small residues A, C, S, T, V) [148]. These N-terminal residues can get acetylated by N-terminal acetylases NatC, NatB and NatF respectively, and NatA can directly acetylate the small residues without the initiator M; the acetylated N-degrons are then recognized by the ER membrane bound E3 ligase Doa10/Not4, which is also involved in retrotranslocation of proteins designated for ERAD [148], [149]. The Pro/N-degron pathway involves the GID ubiquitin ligase, where the Gid4 subunit recognizes MP and MSP N-terminal sequences with some biases for the following 2-6 residues [150].

Lately also fMet/N-degron and Gly/N-degron pathways have been discovered. A study by Kim et al. found that in yeast, the mitochondrial formyltransferase Fmt1 gets retained in the cytosol under certain stress conditions, where it formylates methionine; the fMet gets recognized by the Psh1 ubiquitin ligase [151]. N-terminal glycines were found by Timms et al. to be degraded by Cul2 E3 ligase complex, where N-terminal glycines often occur natively as a result of failed post-translational modification (PTM) myristoylation [152].

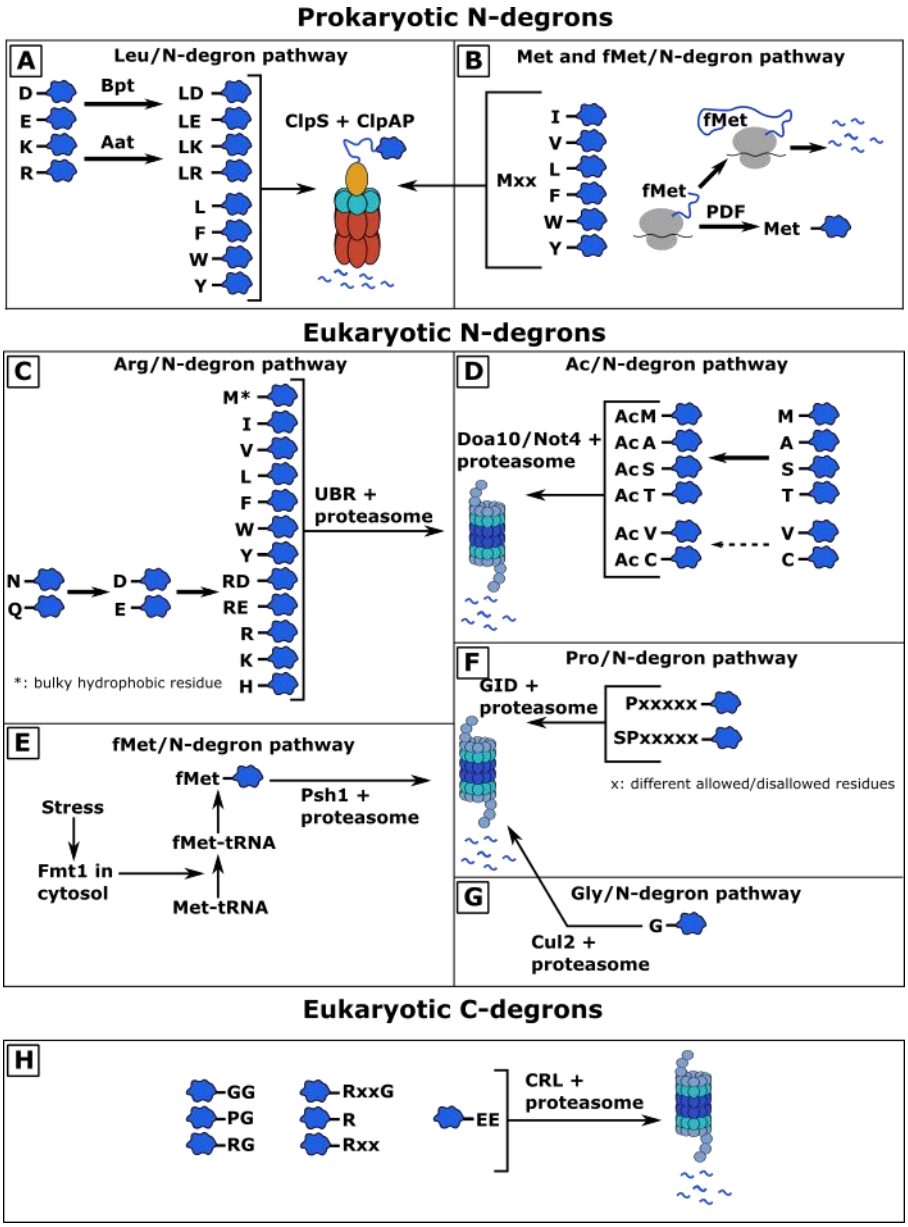


Figure 3 *Overview of N- and C-degron pathways in prokaryotes and eukaryotes*

Proteins can also be targeted for degradation through C-degrons. C-degrons are degraded by the CRL ubiquitin complexes, where adaptor subunits recognize different moieties. Most C-degrons involve either glycine and/or arginine, such as –

GG, -PG, -RG, -RxxG, -R and –Rxx, with –EE being an exception [153]. It is clear that there are mechanisms in place to degrade virtually any protein. However, between 50 and 80% of proteins are acetylated, yet N-terminal acetylation is often regarded as a stabilizing PTM despite also being able to act as a degron [81]. Proteins can avoid degradation through sequestering their degrons, either by association in a complex, where it is shielded by other subunits in the complex, or intramolecular folding [148]. Degrons can then get exposed upon dissociation from said complex due to stress like heat or imbalanced subunit stoichiometry, or the protein can get cleaved by other proteases [81][148]. A recent study constructed a predictive model for degrons based on the stability of a peptide library in yeast, and found in agreeance with other literature that hydrophobic residues get targeted for degradation [154]. Furthermore, transmembrane domains were strong signals for degradation, reinforcing the hypothesis that parts of a protein that are normally not exposed act as degradation signals.

5.4.6 Summary of major differences in chaperones and PQC

In regards to cytosolic chaperones, eukaryotes seem to lack a Hsp60 family chaperone like the GroEL-GroES system of bacteria in their cytosol. Furthermore, metazoans differentiate themselves from other eukaryotes by not having a Hsp104 homolog, instead relying on a Hsp110-family chaperone for disaggregase activity. Also worth noting is that both the human mitochondrial Hsp90 TRAP1 and bacterial Hsp90 HtpG have no co-chaperones, while eukaryotic cytosolic Hsp90 does [122]. Eukaryotes have organelles and thus also chaperones localized in them. As described above, organelles like the mitochondria and the chloroplast have conserved chaperones from bacteria due to their endosymbiotic evolutionary origin. Some of these chaperones have acquired a new function in protein import by complexing with new co-chaperones. Other organelles have chaperones not present in bacteria, such as calnexin/calreticulin in the ER. And although the ER chaperone BiP is a conserved Hsp70 chaperone, the ER of mammals have 7 JDP's (of which only ERdJ3 is a homolog of the E. coli DnaJ), which assist BiP in its functions in the ER [133].

Eukaryotes, especially more complex organisms, have a larger more complex proteome, with more multi domain proteins and unique folds, but also more aggregation prone proteins [156]. An evolutionary study on chaperones by Rebeaud et al shows that the core chaperone families, such as Hsp20, Hsp60, Hsp70, Hsp90

and Hsp100 are conserved from bacteria to higher eukaryotes, with surprisingly no novel core chaperones emerging. Although no new core chaperones emerged, the conserved core chaperones were observed to have an increased expression level, where the ratio of mRNA levels of core chaperones compared to other mRNAs is 6 fold higher in metazoans compared to mesophilic bacteria [156]. The more expansive evolution seems to have happened at the co-chaperone level: eukaryotic co-chaperones are more abundant, make up a higher percentage of the genome and exist in a larger number of distinct protein families [156]. For example, *E. coli* is only known to have 3 JDPs, while yeast has 23, humans have 46, and plants can have over 100 [133]. As described above, humans and mammals also have a more complex network of ATP-independent chaperones of the small heat shock protein class compared to bacteria. The study indicates that the increased abundance of core chaperones along with the increased diversity of co-chaperones and sHsp's enable the evolutionarily conserved core chaperones to tackle the challenges presented by the increased aggregation propensity, prevalence of multi-domain proteins and novel types of folds presented in eukaryotic proteomes. An example of this could be the mitochondrial/chloroplast chaperones, which gained a new function in protein import through interacting with novel co-chaperones.

As the eukaryotic proteome is more complex and prone to aggregation, the PQC needs to be robust and able to deal with a variety of substrate proteins. Some major differences in prokaryotic and eukaryotic PQC occur in the targeting mechanism and target recognition. Eukaryotic degradation uses the PTM ubiquitylation to target substrates for the UPS, or alternatively substrates can get degraded via autophagy into lysosomes, which is regarded as less target specific than the UPS [141]. The ubiquitin gets added by E3 ligases, where the substrate may get recognized by the E3 ligase itself, by another subunit in a complex, or the E3 ligase can get recruited to its substrate by another protein. Prokaryotic degradation involves a few different proteases, where the protease itself recognizes its substrate with no PTM's necessary, as is the case with Lon or FtsH, or in the case of ClpP through AAA ATPases and adaptor proteins.

For target recognition, both eukaryotes and prokaryotes recognize non-specific sequences rich in hydrophobic residues (eukaryotic chaperones also respond to positively charged residues) and more specific signals in N- and C-degrons. Figure 3 summarizes prokaryotic and eukaryotic degrons. Hydrophobic sequences are

recognized by chaperones GroEL-GroES or DnaK-DnaJ-GrpE in prokaryotes, and by Hsp70-Hsp40 and Hsp90 in eukaryotes. If a protein cannot be refolded, it gets targeted for degradation. Prokaryotic chaperones are not suggested to directly target substrates for degradation, rather failure to refold could lead to Lon or other proteases recognizing exposed hydrophobic regions, or to a degron being exposed. Eukaryotic cytosolic chaperones can target substrates to either the UPS by CHIP and BAG1, or to macroautophagy by HspB8 and BAG3. In terms of degrons, the eukaryotic PQC can recognize virtually any N-terminal residue through the different N-degron pathways and the E3 ligases associated with them, while prokaryotic PQC mainly recognizes a few hydrophobic residues in L, F, W and Y, as well as charged residues. Eukaryotic PQC can also recognize some C-terminal degrons, while the only known C-degron in prokaryotes is the sequence added by ssrA. Because more residues can act as degrons in eukaryotes, it is unlikely a stable eukaryotic protein would get degraded in a prokaryotic expression system due to a degron existing in prokaryotes but not eukaryotes. If a heterologous protein gets degraded, it could be due to misfolding exposing hydrophobic regions or previously sequestered degrons, or due to expressing a subunit without the rest of the complex shielding a degron.

5.5. Post-translational modifications and their effect on folding and stability

In addition to their functional and regulatory roles, PTMs also play a role in protein folding, function and stability. Glycosylation is one of the most prevalent PTMs, as more than 50% of human proteins are glycoproteins [157]. As mentioned above, N-glycosylation in eukaryotes targets the protein for the calnexin/calreticulin binding and quality control. N-glycosylation can also affect protein folding and stability outside of chaperone binding and quality control. Studies suggest that N-glycans affect the folding process itself; it was shown that unfolding and refolding a plant lectin ECorL from aggregates yielded much more functional protein for the N-glycosylated variant compared to the unglycosylated variant [158]. This may be related to N-glycosylation affecting the rate of folding, e.g. through interaction with aromatic residues [159][160]. Especially for biopharmaceuticals, glycosylation is important for immunogenicity, stability and activity. Commonly used bacterial hosts, such as industrially used strains of E. coli, do not N-glycosylate its proteins [161]. Some non-industrial bacteria, such as Campylobacter jejuni and other pathogens are able to N-glycosylate proteins, albeit with a much different glycan

structure compared to humans [162]. Thus bacterial hosts are not commonly used for the production of glycosylated biopharmaceuticals. Glycosylation can also affect enzymes in heterologous enzyme production and biosynthetic pathways for chemical production. In a study by Bonzom et al., a feruloyl esterase from the filamentous fungus *Myceliophthora* thermophile was expressed in *Pichia pastoris* (*Komagataella phaffii*) and *E. coli*, where they observed that the unglycosylated enzyme from *E. coli* had lower catalytic efficiency and melting and optimal temperature [163]. Studies looking at molecular dynamics have indicated that glycoproteins tend to be more thermodynamically stable compared to their unglycosylated counterparts [164][165]. N-glycosylation can also increase stability by acting as a steric hindrance to protease sites [166]. Some human proteins cannot be expressed in *E. coli* but can be expressed in eukaryotes like insect cells or *Leishmania*, which have different N-glycan structures than humans; this suggests that in some cases, the presence of glycosylation is more important than the glycan structure [166].

Another prevalent PTM that affects protein folding and stability is disulfide bridge formation; mutating cysteines involved in disulfide bonds often result in loss of function and/or misfolding, and creating additional disulfide bonds can in some cases increase stability [167]–[169]. Furthermore, oxidative protein folding can be the rate-limiting step in protein folding, making it a potential target for optimization in heterologous expression [170][171]. In eukaryotes, oxidative protein folding happens co-translationally in the ER, catalyzed by thiol-disulfide oxidoreductases of the PDI family [170]. In gram-negative bacteria like E. coli, oxidative folding occurs in the periplasm, catalyzed by the Dsb-family of proteins [172]. The oxidative folding pathway is less well defined for gram-positive bacteria, as they do not have a periplasmic space. In B. subtilis, BdbA/B and BdbC/D are able to catalyze disulfide bond formation, although it is suggested that they only act on specific targets instead of being a general mechanism, possibly due to originating from horizontal gene transfer [173]. Other gram-positives, such as Streptomyces and Corynebacterium, are known to secrete many disulfide bond containing proteins [173]. In Streptomyces lividans, homologs for DsbA-D were identified, and were shown to be required to obtain a functional heterologous α-amylase containing a disulfide bond, secreted through the Sec-pathway [174]. In Corynebacterium diptheriae, MdbA catalyzes disulfide bond formation and MdbB re-oxidizes MdbA,

though mechanisms for re-oxidation of MdbB or reduction of incorrect bonds are not yet characterized [173].

Other PTMs that have also been shown to influence folding and stability include lipid modifications, such as N-myristoylation or S-acylation. Myristoylation occurs mainly in eukaryotic cytosolic proteins, where the C14 fatty acid myristate gets irreversibly attached to an N-terminal glycine and can assist the protein with membrane association [175]. A molecular dynamics study looking at the protein hisactophilin observed that myristoylation lowered the free energy barrier for folding and stabilizing the protein through nonnative interactions [176]. Bacterial hosts like *E. coli* lack native N-myristoyl transferase (NMT) activity [177]. In eukaryotic S-acylation, also referred to as palmitoylation, a C16 fatty acid palmitate is reversibly attached to a cysteine by palmitoyl acyltransferases (PAL) and detached by acyl protein thioesterases (APT) [178]. Palmitoylation is involved in membrane association, membrane protein stability and can also function as a membrane anchor [178][179]. Bacterial S-acylation occurs at the plasma membrane and involves a conserved lipobox motif on a type II signal peptide, with cysteine at the +1 position; this means the palmitoylation is limited to be at the N-terminus, a limitation that is not present in eukaryotes [180]. Lysine methylation is found to assist in thermostability of thermophile archaeal enzymes, and does not occur in *E. coli* [181].

5.6 Translational pausing, co-translational folding and assembly

Protein folding is a highly dynamic process, where the speed of translation elongation and different interactions happening co-translationally play an important role in protein folding. The speed of elongation is not uniform, and is determined largely by codon usage [182]. Many amino acids can be encoded for by several codons, and the prevalence of these codons differ depending on the organism; this is often referred to as codon bias, and it can modulate elongation rate through factors such as aa-tRNA abundance, certain amino acid sequences, and mRNA structure [182][183]. It has been shown in both prokaryotes and eukaryotes that rarer codons are decoded by lower abundance tRNA's, leading to slower elongation; although the correlation is lower in *E. coli* than in yeast [182]. Codon bias has been shown to extend to pairs and clusters, and in yeast, 17 codon pairs were observed to slow elongation rates [182], [184]–[186]. Repeated proline residues can stall translation due to being a poor donor and acceptor to peptide

bond formation; both pro- and eukaryotes have an elongation factor to assist in translation of proline repeats, EF-P and eIF5A respectively [187]. Ribosomal pausing can also occur due to "stall peptides", which are 20 AA long peptides with specific sequences that interact with the polypeptide exit tunnel, or due to repeats of A longer than 9 nucleotides [182]. Lastly, elongation can be inhibited by secondary structures such as stem loops in the mRNA, despite the ribosome having helicase activity [188].

As mentioned earlier, ribosome stalling can lead to ribosomal dissociation and degradation of the incomplete protein. Despite of this, ribosomal pausing can be crucial for the correct folding of some proteins. Nascent chains can begin folding very soon after being synthetized; it has been shown that transmembrane □-helixes can form in the ribosomal exit tunnel [189]. Furthermore, nascent chains can interact with the surface of the ribosome, which acts in a holdase-like manner, affecting folding kinetics and thermodynamics [190]. Studies have shown that silent mutations can lead to altered protein conformations and folding kinetics [191][192]. Altering the translational pace can also result in increased protease susceptibility; the *E. coli* multidomain protein SufI became more protease labile upon addition of aa-tRNA of rare codons or substitution of rare codons to more common ones [193]. Also in *E. coli*, altering codons on chloramphenicol acyltransferase CAT, where the overall codon usage was similar but locations of rare codons were changed, resulted in impaired growth in chloramphenicol media despite similar protein abundance; further experiments showed that the altered codon variant was more susceptible to ClpXP degradation upon the addition of an ssrA-tag [194]. Ribosomal pausing also plays a role in TMD insertion; the largest rare codon clusters are often found in transmembrane proteins, and both computational and experimental studies have shown how ribosomal pausing can affect the topology of transmembrane helixes [186][195][196].

Besides the co-translational folding of individual proteins, protein complexes can also assemble in a co-translational manner. When the bioluminescence genes luxA and luxB are expressed from different genomic locations in *E. coli*, the luminescence is decreased by 40% compared to expressing the genes as an operon [197]. It was also observed that the LuxA protein associated with the LuxB nascent chain, but the reverse happened in a much lesser extent. This suggests that the complex assembles co-translationally with a preferred order; a theory further corroborated by a study

from Marsh et al., where it was shown that subunit fusion was significantly more likely to occur in a manner that conserved the complex assembly pathway [198]. Co-translational assembly has also been shown to be wide-spread in eukaryotes; a study in *Schizosaccharomyces pombe* showed that 12 out of 31 proteins with no inherent RNA-binding activity co-purified with mRNAs of known interaction partners [199]. Similarly, 9 out of 12 heteromeric complexes in S. cerevisiae were shown to assemble co-translationally [200]. Chaperones like DnaK and trigger factor in *E. coli* and ribosome associated Hsp70 Ssb1 in yeast shield the nascent chain until it binds with its assembly partner [200][201]. Eukaryotes have a larger proportion of complexes that are heteromers; furthermore, orphaned proteins that lack binding partners are known to be aggregation prone [202][203]. It could therefore be challenging to heterologously express proteins that are a part of a complex.

5.7 Intrinsically disordered proteins and protein regions

Most proteins fold into a stable native structure as the default state; this is however not the case for intrinsically disordered proteins (IDP) or protein regions (IDR). For simplicity, both fully and partially disordered proteins will here be referred to as IDPs. IDPs are generally more prevalent in eukaryotes than prokaryotes, and are often involved in key signaling and regulatory functions; the disorderly nature allows for increased promiscuity and a larger number of interaction partners, which is a consequence of the increased proteomic and organismal complexity [204]–[206]. Industrially, IDPs are not a major class of biopharmaceuticals, although research has been made towards disorder modulators as therapeutic molecules [207]. Enzymes were thought to be an exemption to intrinsic disorder, but a bioinformatics study looking at gaps in x-ray crystallography structures discovered that while there is overall less disorder in enzymes vs. non-enzymes, enzymes with transferase, hydrolase or multiple functional annotations had similar length disordered regions compared to non-enzymes [208]. Heterologous expression in *E. coli* is a commonly used method to obtain IDPs for studying their properties, where many IDPs can be solubly expressed due to being enriched to charged residues; they are however also susceptible to proteolytic degradation [209][210]. IDPs are also enriched in proline residues, which could be problematic for translational elongation as described above [204]. Interestingly, a nuclear E3 ligase in yeast, San1, has a disordered substrate binding region which avoids auto-ubiquitylation and subsequent degradation by not having lysine residues for ubiquitin attachment [211]. It is

unclear if this mechanism would work upon heterologous expression in a bacterial host, as bacteria do not rely on ubiquitylation for proteolytic targeting.

5.8 Strategies for improved folding and expression heterologous proteins

A large variety of methods have been applied towards improving the folding and expression of heterologous proteins, ranging from the process level to strain and protein engineering and optimization of transcription and translation. Protein expression fermentations are often operated at temperatures lower than the optimal temperature for growth, e.g. in a two-stage process with a growth phase at 37 °C and a production phase at 15-25 °C [212]. In terms of the thermodynamics, hydrophobic effect correlates with temperature, and a weaker hydrophobic effect has been shown to promote correct folding [213][214]. Temperature also affects elongation rate; as discussed earlier, proteins can benefit from slowdowns in elongation where they have more time to fold [215]. Lowering the temperature from 37 °C has in general been shown to be a beneficial strategy for heterologous protein expression in *E. coli* [216]–[218]. Figure 4 shows an overview of the different strategies, with the commonly used order of application going top to bottom.

5.8.1 Optimizing the expression cassette

Tuning of transcription and translation initiation have also proven to be useful tools for optimizing heterologous protein expression. In terms of promoter strength, it has been observed in various studies that while stronger promoters lead to an increase in the total amount of protein, much of that increase is often found in the insoluble fraction [219]–[221]. Promoter strength can be adjusted with either inducer concentration, choosing different constitutive promoters, or engineering synthetic promoters [221][222]. Expression can also be tuned by altering the sequence of the translation initiation region (TIR), which includes the Shine-Delgarno sequence (SD), the space between the SD and start codon, as well as the first 5 codons of the open reading frame (ORF) [223]. A computational tool, EMOPEC, has been developed for generating SD sequences with predicted expression levels in *E. coli*; however, to our knowledge, no predictive tools have yet been developed for the entire TIR [224]. Lastly, the coding sequence itself can be optimized by using synonymous codons. Early codon optimization strategies revolve around altering low abundance codons to reduce bottlenecks for translation elongation, applying metrics such as codon adaptation index (CAI), where a higher

CAI indicating fewer rare codons was preferable [225]. While this can be an effective strategy to improve expression, it has been shown that increased CAI does not always lead to increased expression; we now know that rare codons and ribosomal pausing can play functional roles in co-translational folding and membrane topology as described above [226][227]. Codon harmonization looks at mimicking the codon usage of the native organism, and has shown potential as an alternative to maximizing CAI, although the optimal codon usage profile can still be highly dependent on the protein in question [228]–[231]. Algorithms have been developed which can generate codon harmonized gene sequences, such as CodonWizard and CHARMING [232][233].

Another broadly used strategy to obtain soluble expressed heterologous proteins (mostly in *E. coli*) are solubility tags or fusion tags. Here another protein or short peptide is linked to the N- or C-terminal of the target protein to enhance soluble expression. Some commonly used solubility tags are listed in Table 2. The mechanisms by which these tags act has not been fully elucidated. Maltose-binding protein (MBP) is one of the most commonly and successfully used solubility tags, where it has been used to obtain soluble expression of many recombinant human proteins in E. coli [234]–[236]. It has been suggested that MBP could have inherent chaperone activity, although other studies suggest that it acts in a more holdase-like manner, where it sterically blocks aggregate formation or promotes binding of the GroEL chaperone [237][238]. Small ubiquitin-like modifier (SUMO) has also been suggested to sterically block aggregation, as well as potentially acting as a nucleation site for folding [239][240]. The synthetic SET tags on the other hand rely on the repulsive energy of their negatively charged residues to prevent aggregation [241]. Recently, studies have shown that IDPs have potential as a new source of fusion tags that have improved performance compared to commonly used tags such as MBP, GST and Fh8 [242][243]. The proposed mechanisms of action are that the IDP prevents aggregation due to thermally driven random motion, and the low hydrophobicity characteristic of IDPs promote solubility.

One aspect of applying solubility tags that is worth considering is the subsequent cleavage of the tag from the target protein. For protein production, the cleavage happens during purification, where in some cases, the solubility promoting activity is lost after cleavage [244][245]. For enzymes in biosynthetic pathways, if the solubility enhancement is lost upon cleavage, an option would be to leave the tag

on. Here it would be important to consider that improved solubility does not always correlate with improved or even maintained activity. In this case it could be advantageous to use smaller tags that are less likely to influence the activity of the target protein, and look into optimizing the linker sequence between the tag and the target [246]. The selection of the optimal tag for a given target protein is still mostly a trial-and-error process, as different tags work best for different proteins. Some studies have looked into a machine learning based approach to solubility tag selection, which could be a promising future direction for the use of solubility tags [247].

Table 2 *List of commonly used solubility tags for heterologous protein expression in E. coli*

Tag	Length (aa)	Origin	References
MBP (Maltose-binding protein)	396	*E. coli*	[234]-[238]
GST (Glutathione-S transferase)	211	*Schistosoma japonicum*	[248][249]
Trx (Thioredoxin)	109	*E. coli*	[250]
NusA (N-utilization substance)	495	*E. coli*	[244]
SUMO (Small ubiquitin modified)	~100	*Homo sapiens*	[239]
SET (Solubility enhancer peptide sequences)	<20	Synthetic	[241]
Fh8 (*Fasciola hepatica* 8-kDa antigen)	69	*Fasciola hepatica*	[240]

5.8.2 Expression strain engineering

The host strains for expression have also been an interesting target for engineering in order to improve their ability to express heterologous proteins. As discussed above, chaperones play a very important role for the folding and functional expression of proteins, and prokaryotes tend to have lower chaperone abundance. Overloading of the native chaperones of the host organism is hypothesized to be one of the contributing factors to the formation of aggregates/inclusion bodies when expressing heterologous protein. Expressing an unfolded protein can induce HSR, due to DnaK and GroEL being negative regulators of the heat shock sigma-factor σ32, which get titrated by binding the unfolded protein [251]. HSR leads to a sharp transient increase in chaperones and proteases, which could be beneficial for protein expression [252]. However, the σ32 regulon also involves many other genes, and thus HSR induction could cause additional burden on the cell. Studies have

therefore looked into overexpression or co-expression of specific cytosolic chaperones to improve the production of difficult to express proteins, such as GroEL-GroES, DnaK-DnaJ-GrpE and trigger factor [253]–[256]. Table 3 shows examples of studies that have used co- or overexpression of native chaperones to improve protein expression in E. coli.

Table 3 *Studies that co- or overexpressed native chaperones in E. coli to improve heterologous protein expression*

Protein	Organism	Co- or overexpressed chaperone	Outcome	Ref.
Interferon-γ	*H. sapiens*	GroEL-GroES	2.2-fold increase in expression	[260]
Phytase	*Aspergillus niger*	GroEL-GroES	~2-fold increase in specific activity	[261]
Cytochrome P450's 27C1, 2U1 and 2W1	*H. sapiens*	GroEL-GroES	2.1 – 22-fold increase in expression	[262]
Organophosphorus hydrolase	*Flavobacterium sp. ATCC 27551*	GroEL-GroES	5.5-fold increase in activity	[263]
FucT2 (fucosyltransferase)	*Helicobacter pylori*	GroEL-GroES + DnaK-DnaJ-GrpE	5-fold increase in activity	[264]
LZip (cAMP-responsive element binding protein 3)	*H. sapiens*	DnaK-DnaJ-GrpE	Only soluble with chaperone co-expression	[254]

Chaperones differ from organism to organism on many levels, from sequence to abundance and interactions with co-chaperones and other chaperones. Therefore it can also be beneficial to express chaperones from other organisms, including the organism the protein comes from. A study by Ahn and Yun showed a 15-fold increase in the catalytic activity of human cytochrome P450 CYP3A4 expressed in *E.coli* with co-expression of human J-protein HDJ-1 [257]. HDJ-1 is a homolog of DnaJ and in humans acts as a co-chaperone to the cytosolic Hsp70; this could indicate that it can also act as a co-chaperone for DnaK of *E. coli*. Similarly, co-expression of Hsp70 from Plasmodium falciparum enabled Stephens and co to obtain functional expression of P. falciparum GTP cyclohydrolase I in *E. coli* [258]. In an alternative approach, Shanmugasundaram et al. screened extremophile chaperones and

discovered that Cpn60 from *Pyrococcus furiosus* (homologue of GroEL) was able to improve the solubility and activity of a poorly soluble GFP variant from *Aequorea victoria* when co-expressed in E. coli [259]. It could be interesting to construct chaperone co-expression libraries to improve heterologous protein expression, as both native and heterologous chaperones have shown potential to improve folding and expression.

As described above, bacterial host organisms often lack the ability to perform certain PTM's that can be important for the folding and stability of a protein. Efforts have therefore been made towards introducing PTM pathways into bacterial hosts. E. coli has been successfully engineered to perform N-glycosylation by heterologous expression of oligosaccharyltransferase PglB from *Campylobacter jejuni* and glycosyltransferases from *S. cerevisiae*, where the glycan gets attached upon translocation of the target protein to the periplasm, resulting in a $Man_3GlcNAc_2$ N-glycosylated protein [265]. It is also possible to N-glycosylate cytosolic proteins by expressing N-glucosyltransferases from *Actinobacillus pleuropneumoniae* [266]. The primary focus of these glycoengineering studies is to produce glycoproteins with human-like N-glycans for therapeutic applications, and it is unclear if the glycosylation benefits protein folding; as bacteria do not have CNX/CRT, the glycan can only act through CNX-independent mechanisms like interacting with aromatic residues during folding and steric inhibition of degradation sites. Furthermore, titers of glycoproteins produced in bacteria can be quite low, limiting the viability of using glycoengineered strains as a strategy for improved protein folding and expression [267].

For disulfide bonds, gram-negative bacteria are able to catalyze their formation in the periplasm, but not the cytoplasm. Disulfide bond containing proteins, such as single-chain antibody fragments and human growth hormone, are therefore often targeted to the periplasm using signal peptides from native periplasmic proteins that target the Sec/SRP pathway or the Tat pathway [268]–[270]. Here the signal peptide sequence is important to optimize for efficient transport, and the optimal signal peptide differs depending on the protein [271][272]. Like with cytosolic expression, co-expression of chaperones has also been used to enhance folding and expression in the periplasm [273][274]. Alternatively, studies have also attempted to achieve a more oxidative cytosol in order to facilitate disulfide bond formation. Earlier approaches relied on knocking out thiosulfide reductase activity by deleting

gor and trxB, overexpressing disulfide isomerase DsbC and mutating AhpC to maintain viability; strains with these modifications were verified to be able to express several DSB containing proteins in the cytoplasm [275]. More recently, a study by Gąciarz et al achieved disulfide bond formation in the cytoplasm by heterologously expressing a sulfhydryl oxidase Erv1p from *S. cerevisiae* and a human disulfide bond isomerase PDI in a system called CyDisCo [276]. This strain could also grow on minimal media, likely due to not needing to knock out the native reductases, and was verified to be able to produce several DSB containing mammalian proteins in a fed-batch fermentation. The system was improved in a new plasmid system called DisCoTune, using a new plasmid backbone allowing for better balancing of expression between the GOI and CyDisCo genes [277].

Strain engineering has also been used to address codon bias, by modifying the strain to increase the levels of AA-tRNA complexes of otherwise rare codons. One of the well-known commercial *E. coli* strains with increased expression of tRNA for rare codons is Rosetta(DE3) by Novagen, which compensates for the rare codons AGG, AGA, AUA, CUA, CCC and GGA by having an extra copy of the tRNA-genes expressed on a plasmid [278]. This strain showed increased production of short 25 – 150 AA regions of human proteins compared to another protein production strain BL21(DE3). The results may look different for a set of complete proteins, as short/partial proteins could be less demanding to fold. Furthermore, a study was published that indicates the improvements in expression when using the Rosetta strain comes from increased T7 polymerase levels due to the plasmid backbone used, rather than the tRNA themselves [279]. Another study looked at recreating the strain but having the extra copies of tRNA-genes genome integrated instead. This strain, coined by the authors as “SixPack”, showed improved growth when compared to a Rosetta-derivative strain, possibly due to integration into an rRNA operon, and thus having regulated rather than constitutive expression [280]. For protein expression, 4 out of 8 proteins showed improved expression in the SixPack strain compared to BL21. As the price of de novo DNA synthesis has decreased over the years, synthesis of codon optimized genes has emerged as the more flexible method to address codon bias, as it is possible to account for the codon usage of the native organism in order to preserve functionally relevant rare codons [281].

5.8.3 Protein engineering

It is also possible to engineer the protein of interest itself to improve folding and expression, where both evolutionary and rational engineering based methods have been explored. A method based on phage infection, named phage assisted continuous evolution (PACE), has been used to select for correct folding of a protein of interest. The method is based on a split T7-polymerase and phage infection, where the small fragment of the polymerase is linked to the POI, and the full T7-polymerase is responsible for transcribing a phage proliferation gene; because the culture is continuously diluted with fresh cells, only phages with POI variants that fold correctly and allow the T7-polymerase to transcribe the phage protein will be able to persist in the culture [282]. It was also possible to select for binding when evolving an antibody, where the phage gene was split in two, and the other half's expression was controlled by an rpoZ-POI fusion, where antibody binding was required for expression. Using this method, mutations that allowed for increased expression and maintained binding activity were identified. While the folding selection could potentially be applied for other proteins, the binding activity selection may not be applicable for proteins without inherent binding partners. Another evolution based method would be directed evolution; a reporter-system for insoluble proteins was recently published for *E. coli*, where it was possible to select for more soluble variants in a library of an insoluble protein with fluorescence activated cell sorting (FACS) [283]. With this method it is not possible to simultaneously select for activity; more data on how activity is affected by selecting for improved solubility is needed to evaluate the broader applicability of this reporter system.

Computational biology has in recent years emerged to potentially provide more rational engineering based methods for protein engineering, with the most known example being AlphaFold2, which can predict protein 3D-structures very accurately [284]. While prediction of heterologous expression success is not fully possible yet due to there being a plethora of factors at play and incomplete knowledge of all the underlying mechanisms, some studies show that using certain predictable parameters, such as $\Delta\Delta G$, as a proxy for improved heterologous expression could be a viable approach to protein engineering. Johansson et al. developed a method called Global Multi-Mutation Analysis (GMMA), which combines experiments and computation to identify combinations of mutations that improve the stability of a

protein; they were able to increase the fluorescence of a GFP variant several fold by introducing the top 6 predicted mutations, many of which were also present in the known GFP variant superfolder GFP (sfGFP) [285]. Another computational method by Goldenzweig et al. combines $\Delta\Delta G$ from Rosetta modeling predictions with a position-specific substitution matrix (PSSM) generated from a sequence alignment of homologous sequences to reduce the number of false positive stabilizing mutations in a method [286]. The algorithm was verified with experimental data of single-point mutations in the enzymes fungal endoglucanase Cel5A and yeast triosephosphate isomerase (TIM), and was able to eliminate all destabilizing mutations and retain 35% and 65% of stabilizing mutations respectively. For other targets, the algorithm designed a highly active and higher expression variant of human acetylcholinesterase hAChE, and also improved the activity and expression of human histone deacylase SIRT6 and bacterial phosphotriesterase. For human DNA methyltransferase Dnmt3a, the variant generated by the algorithm showed increased activity but lower expression. Overall, computational methods have shown potential for rational engineering for improved folding and expression of heterologous proteins, and will likely keep improving for the foreseeable future. Neural network based models are also being explored for predicting stabilizing mutations, but are currently mostly limited to smaller proteins [287]. Recently, a neural network named MutCompute was used to successfully engineer a thermostabile plastic degrading enzyme [288].

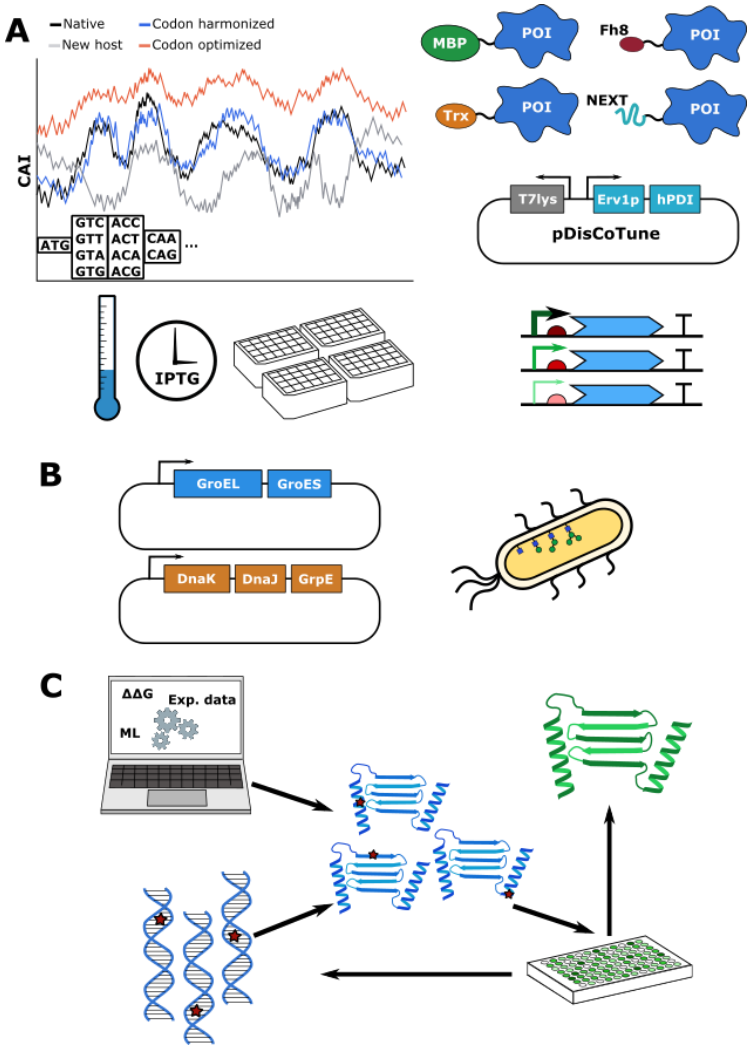


Figure 4 Strategies for improving heterologous protein folding and expression. A) Synthesizing the gene gives options to adjust for changes in codon usage, either prioritizing frequently used codons or trying to mimic the usage of the native organism. With cloning, solubility tags and different promoters and SD-sequences can be incorporated. If the protein has disulfide bonds, it can be co-expressed with the pDisCoTune plasmid. These constructs can be screened and cultivation conditions such as temperature, inducer concentration and induction time can be optimized. B) If the expression is still poor, chaperones can be co-expressed, either from the new host or native organism. If the protein has PTM's, strains that are engineered to be able to perform them can be used. C) The protein can be engineered, either using evolution-based methods (if high-throughput screens are available), or using computational approaches.

5.9 Closing remarks

Heterologous protein expression is important across many applications of biotechnology. Bacterial platform organisms such as *E. coli* play a major role, but heterologous protein expression can be challenging, leading to misfolding, formation of aggregates and low expression. Within the cell, heterologous proteins interact with the ribosome, chaperones and proteases, which affect whether the protein folds into the correct structure. Other factors such as co-translational folding, assembly and post-translational modifications can also affect the protein. While some aspects are conserved from prokaryotes to eukaryotes, there are also areas with significant differences. Many strategies have been applied towards improving heterologous protein expression, such as codon optimization, fusion tags, and strain and protein engineering. Having an overview of commonly applied strategies together with an understanding of the cellular mechanisms and how they might be related can hopefully prove to be useful in designing new studies and gaining further understanding and increased predictability in regards to heterologous protein expression.

5.10 Acknowledgements

This work was funded by the Novo Nordisk Foundation within the framework of the Fermentation-based Biomanufacturing Initiative (FBM), grant number: NNF17SA0031362.

5.10 References for Chapter 5

[1] G. Walsh, "Biopharmaceutical benchmarks 2018," Nature Biotechnology, vol. 36, no. 12, pp. 1136–1145, 2018, doi: 10.1038/nbt.4305.

[2] I. Ozer and A. Chilkoti, "Site-Specific and Stoichiometric Stealth Polymer Conjugates of Therapeutic Peptides and Proteins," Bioconjugate Chemistry, vol. 28, no. 3, pp. 713–723, 2017, doi: 10.1021/acs.bioconjchem.6b00652.

[3] M. N. Baeshen et al., "Production of Biopharmaceuticals in E. coli: Current Scenario and Future Perspectives," Journal of Microbiology and Biotechnology, vol. 25, no. 7, pp. 953–962, Jul. 2015, doi: 10.4014/jmb.1412.12079.

[4] F. F. C. Barros, A. P. R. Simiqueli, C. J. de Andrade, and G. M. Pastore, " Production of Enzymes from Agroindustrial Wastes by Biosurfactant-Producing Strains of Bacillus subtilis ," Biotechnology Research International, vol. 2013, pp. 1–9, 2013, doi: 10.1155/2013/103960.

[5] R. Singh, M. Kumar, A. Mittal, and P. K. Mehta, "Microbial enzymes: industrial progress in 21st century," 3 Biotech, vol. 6, no. 2, pp. 1–15, 2016, doi: 10.1007/s13205-016-0485-8.

[6] J. Nielsen and J. D. Keasling, "Engineering Cellular Metabolism," Cell, vol. 164, no. 6, pp. 1185–1197, 2016, doi: 10.1016/j.cell.2016.02.004.

[7] F. Baneyx and M. Mujacic, "Recombinant protein folding and misfolding in Escherichia coli," Nature Biotechnology, vol. 22, no. 11, pp. 1399–1407, 2004, doi: 10.1038/nbt1029.

[8] A. Dwevedi, Protein Folding. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-12592-3.

[9] F. Mallamace et al., "Energy landscape in protein folding and unfolding," Proc Natl Acad Sci U S A, vol. 113, no. 12, pp. 3159–3163, 2016, doi: 10.1073/pnas.1524864113.

[10] E. Lindahl et al., Molecular Modeling of Proteins. 2015. [Online]. Available: https://link-springer-com.sire.ub.edu/content/pdf/10.1007%2F978-1-4939-1465-4.pdf

[11] G. Skretas and S. Ventura, "Editorial: Protein Aggregation and Solubility in Microorganisms (Archaea, Bacteria and Unicellular Eukaryotes): Implications and Applications," Frontiers in Microbiology, vol. 11, Nov. 2020, doi: 10.3389/fmicb.2020.620239.

[12] M. Arrasate, S. Mitra, E. S. Schweitzer, M. R. Segal, and S. Finkbeiner, "Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death," Nature, vol. 431, no. 7010, pp. 805–810, Oct. 2004, doi: 10.1038/nature02998.

[13] N. S. de Groot and S. Ventura, "Effect of temperature on protein quality in bacterial inclusion bodies," FEBS Letters, vol. 580, no. 27, pp. 6471–6476, Nov. 2006, doi: 10.1016/j.febslet.2006.10.071.

[14] Š. Peternel, J. Grdadolnik, V. Gaberc-Porekar, and R. Komel, "Engineering inclusion bodies for non denaturing extraction of functional proteins," Microbial Cell Factories, vol. 7, no. 1, p. 34, 2008, doi: 10.1186/1475-2859-7-34.

[15] A. Singh, V. Upadhyay, A. Singh, and A. K. Panda, "Structure-Function Relationship of Inclusion Bodies of a Multimeric Protein," Frontiers in Microbiology, vol. 11, May 2020, doi: 10.3389/fmicb.2020.00876.

[16] D. J. Wurm et al., "Teaching an old pET new tricks: tuning of inclusion body formation and properties by a mixed feed system in E. coli," Applied Microbiology and Biotechnology, vol. 102, no. 2, pp. 667–676, Jan. 2018, doi: 10.1007/s00253-017-8641-6.

[17] C. Slouka et al., "Custom made inclusion bodies: impact of classical process parameters and physiological parameters on inclusion body quality attributes," Microbial Cell Factories, vol. 17, no. 1, p. 148, Dec. 2018, doi: 10.1186/s12934-018-0997-5.

[18] A. K. Upadhyay, A. Murmu, A. Singh, and A. K. Panda, "Kinetics of inclusion body formation and its correlation with the characteristics of protein aggregates in escherichia coli," PLoS ONE, vol. 7, no. 3, 2012, doi: 10.1371/journal.pone.0033951.

[19] B. Jürgen et al., "Quality control of inclusion bodies in Escherichia coli," Microbial Cell Factories, vol. 9, no. 1, p. 41, Dec. 2010, doi: 10.1186/1475-2859-9-41.

[20] A. Ramón, M. Señorale-Pose, and M. Marín, "Inclusion bodies: not that bad…," Frontiers in Microbiology, vol. 5, 2014, doi: 10.3389/fmicb.2014.00056.

[21] C. Slouka, J. Kopp, O. Spadiut, and C. Herwig, "Perspectives of inclusion bodies for bio-based products: curse or blessing?," Applied Microbiology and Biotechnology, vol. 103, no. 3, pp. 1143–1153, Feb. 2019, doi: 10.1007/s00253-018-9569-1.

[22] L. Gifre-Renom et al., "The Biological Potential Hidden in Inclusion Bodies," Pharmaceutics, vol. 12, no. 2, p. 157, Feb. 2020, doi: 10.3390/pharmaceutics12020157.

[23] V. D. Jäger et al., "Catalytically-active inclusion bodies for biotechnology—general concepts, optimization, and application," Applied Microbiology and Biotechnology, vol. 104, no. 17, pp. 7313–7329, Sep. 2020, doi: 10.1007/s00253-020-10760-3.

[24] R. Koszagova et al., "Magnetization of active inclusion bodies: comparison with centrifugation in repetitive biotransformations," Microbial Cell Factories, vol. 17, no. 1, p. 139, Dec. 2018, doi: 10.1186/s12934-018-0987-7.

[25] A. Singh, V. Upadhyay, A. K. Upadhyay, S. M. Singh, and A. K. Panda, "Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process," Microbial Cell Factories, vol. 14, no. 1, p. 41, Dec. 2015, doi: 10.1186/s12934-015-0222-8.

[26] D. Kishore, S. Kundu, and A. M. Kayastha, "Thermal, Chemical and pH Induced Denaturation of a Multimeric β-Galactosidase Reveals Multiple Unfolding Pathways," PLoS ONE, vol. 7, no. 11, 2012, doi: 10.1371/journal.pone.0050380.

[27] H. P. Sørensen and K. K. Mortensen, "Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli," Microbial Cell Factories, vol. 4, no. Figure 1, pp. 1–8, 2005, doi: 10.1186/1475-2859-4-1.

[28] C. Camilloni et al., "Towards a structural biology of the hydrophobic effect in protein folding," Scientific Reports, vol. 6, no. May, pp. 1–9, 2016, doi: 10.1038/srep28285.

[29] E. van Dijk, A. Hoogeveen, and S. Abeln, "The Hydrophobic Temperature Dependence of Amino Acids Directly Calculated from Protein Structures," PLoS Computational Biology, vol. 11, no. 5, pp. 1–17, 2015, doi: 10.1371/journal.pcbi.1004277.

[30] A. Pastore, S. R. Martin, A. Politou, K. C. Kondapalli, T. Stemmler, and P. A. Temussi, "Unbiased cold denaturation: Low- and high-temperature unfolding of yeast frataxin under physiological conditions," J Am Chem Soc, vol. 129, no. 17, pp. 5374–5375, 2007, doi: 10.1021/ja0714538.

[31] G. Feller, "Psychrophilic Enzymes: From Folding to Function and Biotechnology," Scientifica (Cairo), vol. 2013, pp. 1–28, 2013, doi: 10.1155/2013/512840.

[32] H. Im, E. J. Seo, and M. H. Yu, "Metastability in the inhibitory mechanism of human α1-antitrypsin," Journal of Biological Chemistry, vol. 274, no. 16, pp. 11072–11077, 1999, doi: 10.1074/jbc.274.16.11072.

[33] J. A. Vasina and F. Baneyx, "Recombinant protein expression at low temperatures under the transcriptional control of the major Eschenchia coli cold shock promoter cspA," Applied and Environmental Microbiology, vol. 62, no. 4, pp. 1444–1447, 1996, doi: 10.1128/aem.62.4.1444-1447.1996.

[34] J. Singh and J. B. Udgaonkar, "Unraveling the Molecular Mechanism of pH-Induced Misfolding and Oligomerization of the Prion Protein," Journal of Molecular Biology, vol. 428, no. 6, pp. 1345–1355, 2016, doi: 10.1016/j.jmb.2016.01.030.

[35] T. J. Sereda, C. T. Mant, A. M. Quinn, and R. S. Hodges, "Effect of the α-amino group on peptide retention behaviour in reversed-phase chromatography Determination of the pKa values of the α-amino group of 19 different N-terminal amino acid residues," Journal of Chromatography A, vol. 646, no. 1, pp. 17–30, Aug. 1993, doi: 10.1016/S0021-9673(99)87003-4.

[36] M. J. Harms et al., "The pKa Values of Acidic and Basic Residues Buried at the Same Internal Location in a Protein Are Governed by Different Factors," Journal of Molecular Biology, vol. 389, no. 1, pp. 34–47, May 2009, doi: 10.1016/j.jmb.2009.03.039.

[37] M. Schulte et al., "A mechanism to prevent production of reactive oxygen species by Escherichia coli respiratory complex I," Nature Communications, vol. 10, no. 1, 2019, doi: 10.1038/s41467-019-10429-0.

[38] J. L. Martínez, E. Meza, D. Petranovic, and J. Nielsen, "The impact of respiration and oxidative stress response on recombinant α-amylase production by Saccharomyces cerevisiae," Metabolic Engineering Communications, vol. 3, pp. 205–210, 2016, doi: 10.1016/j.meteno.2016.06.003.

[39] E. Lévy et al., "Causative links between protein aggregation and oxidative stress: A review," International Journal of Molecular Sciences, vol. 20, no. 16, 2019, doi: 10.3390/ijms20163896.

[40] H. Saibil, "Chaperone machines for protein folding, unfolding and disaggregation," Nature Reviews Molecular Cell Biology, vol. 14, no. 10, pp. 630–642, 2013, doi: 10.1038/nrm3658.

[41] A. Buchberger, B. Bukau, and T. Sommer, "Protein Quality Control in the Cytosol and the Endoplasmic Reticulum: Brothers in Arms," Molecular Cell, vol. 40, no. 2, pp. 238–252, 2010, doi: 10.1016/j.molcel.2010.10.001.

[42] J. Sawa, H. Malet, T. Krojer, F. Canellas, M. Ehrmann, and T. Clausen, "Molecular adaptation of the DegQ protease to exert protein quality control in the bacterial cell envelope," Journal of Biological Chemistry, vol. 286, no. 35, pp. 30680–30690, 2011, doi: 10.1074/jbc.M111.243832.

[43] Y. E. Kim, M. S. Hipp, A. Bracher, M. Hayer-Hartl, and F. Ulrich Hartl, Molecular Chaperone Functions in Protein Folding and Proteostasis, vol. 82, no. 1. 2013. doi: 10.1146/annurev-biochem-060208-092442.

[44] P. A. Lund, "Microbial molecular chaperones," Advances in Microbial Physiology, vol. 44, pp. 93–140, 2001, doi: 10.1016/S0065-2911(01)44012-4.

[45] D. A. Dougan, A. Mogk, and B. Bukau, "Protein folding and degradation in bacteria: To degrade or not to degrade? That is the question," Cellular and Molecular Life Sciences, vol. 59, no. 10, pp. 1607–1616, 2002, doi: 10.1007/PL00012487.

[46] G. Kramer et al., "Functional Dissection of Escherichia coli Trigger Factor: Unraveling the Function of Individual Domains," Journal of Bacteriology, vol. 186, no. 12, pp. 3777–3784, Jun. 2004, doi: 10.1128/JB.186.12.3777-3784.2004.

[47] J. G. Thomas and F. Baneyx, "ClpB and HtpG facilitate de novo protein folding in stressed Escherichia coli cells," Molecular Microbiology, vol. 36, no. 6, pp. 1360–1370, 2000, doi: 10.1046/j.1365-2958.2000.01951.x.

[48] K. Motohashi, Y. Watanabe, M. Yohda, and M. Yoshida, "Heat-inactivated proteins are rescued by the DnaK·J-GrpE set and ClpB chaperones," Proc Natl Acad Sci U S A, vol. 96, no. 13, pp. 7184–7189, 1999, doi: 10.1073/pnas.96.13.7184.

[49] D. Kuczyńska-Wiśnik et al., "The Escherichia coli small heat-shock proteins IbpA and IbpB prevent the aggregation of endogenous proteins denatured in vivo during extreme heat shock," Microbiology (N Y), vol. 148, no. 6, pp. 1757–1765, 2002, doi: 10.1099/00221287-148-6-1757.

[50] O. Keskin, I. Bahar, D. Flatow, D. G. Covell, and R. L. Jernigan, "Molecular mechanisms of chaperonin GroEL-GroES function," Biochemistry, vol. 41, no. 2, pp. 491–501, 2002, doi: 10.1021/bi011393x.

[51] T. Ueno, H. Taguchi, H. Tadakuma, M. Yoshida, and T. Funatsu, "GroEL mediates protein folding with a two successive timer mechanism," Molecular Cell, vol. 14, no. 4, pp. 423–434, 2004, doi: 10.1016/S1097-2765(04)00261-8.

[52] Y. Li, Z. Zheng, A. Ramsey, and L. Chen, "Analysis of peptides and proteins in their binding to GroEL," Journal of Peptide Science, vol. 16, no. 12, pp. 693–700, Dec. 2010, doi: 10.1002/psc.1288.

[53] Z. Lin and H. S. Rye, "GroEL-Mediated Protein Folding: Making the Impossible, Possible," Critical Reviews in Biochemistry and Molecular Biology, vol. 41, no. 4, pp. 211–239, Jan. 2006, doi: 10.1080/10409230600760382.

[54] P. Goloubinoff, A. Mogk, A. P. ben Zvi, T. Tomoyasu, and B. Bukau, "Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network," Proc Natl Acad Sci U S A, vol. 96, no. 24, pp. 13732–13737, 1999, doi: 10.1073/pnas.96.24.13732.

[55] X. Zhu et al., "Structural Analysis of Substrate Binding by the Molecular Chaperone DnaK," Science (1979), vol. 272, no. 5268, pp. 1606–1614, 1996, doi: 10.1016/j.physbeh.2017.03.040.

[56] V. R. Agashe et al., "Function of Trigger Factor and DnaK in Multidomain Protein Folding: Increase in Yield at the Expense of Folding Speed.," Cell, vol. 117, no. 2, pp. 199–209, 2004, doi: 10.1016/s0092-8674(04)00299-5.

[57] W. Han and P. Christen, "Mechanism of the targeting action of DnaJ in the DnaK molecular chaperone system," Journal of Biological Chemistry, vol. 278, no. 21, pp. 19038–19043, 2003, doi: 10.1074/jbc.M300756200.

[58] C. Harrison, "GrpE, a nucleotide exchange factor for DnaK," Cell Stress and Chaperones, vol. 8, no. 3, pp. 218–224, 2003, doi: 10.1379/1466-1268(2003)008<0218:GANEFF>2.0.CO;2.

[59] R. Rosenzweig, S. Moradi, A. Zarrine-Afsar, J. R. Glover, and L. E. Kay, "Unraveling the Mechanism of Protein Disaggregation Through a ClpB-DnaK Interaction," Science (1979), vol. 1080, no. March, pp. 1–4, 2013, doi: 10.1126/science.1233066.

[60] Y. Jin, R. S. Hoxie, and T. O. Street, "Molecular mechanism of bacterial Hsp90 pH-dependent ATPase activity," Protein Science, vol. 26, no. 6, pp. 1206–1213, 2017, doi: 10.1002/pro.3174.

[61] F. Bonomi, S. Iametti, A. Morleo, D. Ta, and L. E. Vickery, "Studies on the mechanism of catalysis of iron-sulfur cluster transfer from IscU[2Fe2S] by HscA/HscB chaperones," Biochemistry, vol. 47, no. 48, pp. 12795–12801, 2008, doi: 10.1021/bi801565j.

[62] E. Ratajczak, S. Zietkiewicz, and K. Liberek, "Distinct Activities of Escherichia coli Small Heat Shock Proteins IbpA and IbpB Promote Efficient Protein Disaggregation," Journal of Molecular Biology, vol. 386, no. 1, pp. 178–189, 2009, doi: 10.1016/j.jmb.2008.12.009.

[63] H. LeThanh, P. Neubauer, and F. Hoffmann, "The small heat-shock proteins IbpA and IpbB reduce the stress load of recombinant Escherichia coli and delay degradation of inclusion bodies," Microbial Cell Factories, vol. 4, pp. 1–12, 2005, doi: 10.1186/1475-2859-4-6.

[64] P. Bechtluft, N. Nouwen, S. J. Tans, and A. J. M. Driessen, "SecB - A chaperone dedicated to protein translocation," Molecular BioSystems, vol. 6, no. 4, pp. 620–627, 2010, doi: 10.1039/b915435c.

[65] M. P. Castanié-Cornet, N. Bruel, and P. Genevaux, "Chaperone networking facilitates protein targeting to the bacterial cytoplasmic membrane," Biochimica et Biophysica Acta - Molecular Cell Research, vol. 1843, no. 8, pp. 1442–1456, 2014, doi: 10.1016/j.bbamcr.2013.11.007.

[66] E. R. Green and J. Mecsas, "Bacterial Secretion Systems: An Overview," Virulence Mechanisms of Bacterial Pathogens, vol. 4, no. 1, pp. 213–239, 2016, doi: 10.1128/9781555819286.ch8.

[67] M. Peschke et al., "SRP, FtsY, DnaK and YidC Are Required for the Biogenesis of the E. coli Tail-Anchored Membrane Proteins DjlC and Flk," Journal of Molecular Biology, vol. 430, no. 3, pp. 389–403, 2018, doi: 10.1016/j.jmb.2017.12.004.

[68] M. Merdanovic, T. Clausen, M. Kaiser, R. Huber, and M. Ehrmann, "Protein Quality Control in the Bacterial Periplasm," Annual Review of Microbiology, vol. 65, no. 1, pp. 149–168, 2011, doi: 10.1146/annurev-micro-090110-102925.

[69] L. He, T. Sharpe, A. Mazur, and S. Hiller, "A molecular mechanism of chaperone-client recognition," Science Advances, vol. 2, no. 11, pp. 1–12, 2016, doi: 10.1126/sciadv.1601625.

[70] D. A. Holdbrook et al., "A Spring-Loaded Mechanism Governs the Clamp-like Dynamics of the Skp Chaperone," Structure, vol. 25, no. 7, pp. 1079-1088.e3, 2017, doi: 10.1016/j.str.2017.05.018.

[71] S. K. Singh, R. Grimaud, J. R. Hoskins, S. Wickner, and M. R. Maurizi, "Unfolding and internalization of proteins by the ATP-dependent proteases ClpXP and ClpAP," Proc Natl Acad Sci U S A, vol. 97, no. 16, pp. 8898–8903, 2000, doi: 10.1073/pnas.97.16.8898.

[72] V. Tsilibaris, G. Maenhaut-Michel, and L. van Melderen, “Biological roles of the Lon ATP-dependent protease,” Research in Microbiology, vol. 157, no. 8, pp. 701–713, 2006, doi: 10.1016/j.resmic.2006.05.004.

[73] T. A. Baker and R. T. Sauer, “ClpXP, an ATP-powered unfolding and protein-degradation machine,” Biochimica et Biophysica Acta - Molecular Cell Research, vol. 1823, no. 1, pp. 15–28, 2012, doi: 10.1016/j.bbamcr.2011.06.007.

[74] S. Gottesman, E. Roche, Y. N. Zhou, and R. T. Sauer, “The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system,” Genes and Development, vol. 12, no. 9, pp. 1338–1347, 1998, doi: 10.1101/gad.12.9.1338.

[75] J. R. Hoskins, S. K. Singh, M. R. Maurizi, and S. Wickner, “Protein binding and unfolding by the chaperone ClpA and degradation by the protease ClpAP,” Proc Natl Acad Sci U S A, vol. 97, no. 16, pp. 8892–8897, 2000, doi: 10.1073/pnas.97.16.8892.

[76] J. L. Camberg, J. R. Hoskins, and S. Wickner, “ClpXP protease degrades the cytoskeletal protein, FtsZ, and modulates FtsZ polymer dynamics,” Proceedings of the National Academy of Sciences, vol. 106, no. 26, pp. 10614–10619, Jun. 2009, doi: 10.1073/pnas.0904886106.

[77] J. M. Flynn, I. Levchenko, R. T. Sauer, and T. A. Baker, “Modulating substrate choice: the SspB adaptor delivers a regulator of the extracytoplasmic-stress response to the AAA+ protease ClpXP for degradation,” Genes & Development, vol. 18, no. 18, pp. 2292–2301, Sep. 2004, doi: 10.1101/gad.1240104.

[78] A. Studemann, “Sequential recognition of two distinct sites in S by the proteolytic targeting factor RssB and ClpX,” The EMBO Journal, vol. 22, no. 16, pp. 4111–4120, Aug. 2003, doi: 10.1093/emboj/cdg411.

[79] M. Kanemori, K. Nishihara, H. Yanagi, and T. Yura, “Synergistic roles of HslVU and other ATP-dependent proteases in controlling in vivo turnover of sigma32 and abnormal proteins in Escherichia coli,” Journal of Bacteriology, vol. 179, no. 23, pp. 7219–7225, Dec. 1997, doi: 10.1128/jb.179.23.7219-7225.1997.

[80] X. Gao, J. Yeom, and E. A. Groisman, “The expanded specificity and physiological role of a widespread N-degron recognin,” Proceedings of the National Academy of Sciences, vol. 116, no. 37, pp. 18629–18637, Sep. 2019, doi: 10.1073/pnas.1821060116.

[81] A. Varshavsky, "N-degron and C-degron pathways of protein degradation," Proceedings of the National Academy of Sciences, vol. 116, no. 2, pp. 358–366, Jan. 2019, doi: 10.1073/pnas.1816596116.

[82] K. Piatkov, T. Vu, C.-S. Hwang, and A. Varshavsky, "Formyl-methionine as a degradation signal at the N-termini of bacterial proteins," Microbial Cell, vol. 2, no. 10, pp. 376–393, Oct. 2015, doi: 10.15698/mic2015.10.231.

[83] K. L. Griffith and A. D. Grossman, "Inducible protein degradation in Bacillus subtilis using heterologous peptide tags and adaptor proteins to target substrates to the protease ClpXP," Molecular Microbiology, Nov. 2008, doi: 10.1111/j.1365-2958.2008.06467.x.

[84] C. Müller, C. Crowe-McAuliffe, and D. N. Wilson, "Ribosome Rescue Pathways in Bacteria," Frontiers in Microbiology, vol. 12, Mar. 2021, doi: 10.3389/fmicb.2021.652980.

[85] J. Kirstein et al., "Adaptor protein controlled oligomerization activates the AAA + protein ClpC," EMBO Journal, vol. 25, no. 7, pp. 1481–1491, 2006, doi: 10.1038/sj.emboj.7601042.

[86] E. Gur, M. Vishkautzan, and R. T. Sauer, "Protein unfolding and degradation by the AAA+ Lon protease," Protein Science, vol. 21, no. 2, pp. 268–278, Feb. 2012, doi: 10.1002/pro.2013.

[87] S. Langklotz, U. Baumann, and F. Narberhaus, "Structure and function of the bacterial AAA protease FtsH," Biochimica et Biophysica Acta - Molecular Cell Research, vol. 1823, no. 1, pp. 40–48, 2012, doi: 10.1016/j.bbamcr.2011.08.015.

[88] K. Koide, K. Ito, and Y. Akiyama, "Substrate recognition and binding by RseP, an Escherichia coli intramembrane protease," Journal of Biological Chemistry, vol. 283, no. 15, pp. 9562–9570, 2008, doi: 10.1074/jbc.M709984200.

[89] K. C. Bastiaansen, A. Ibañez, J. L. Ramos, W. Bitter, and M. A. Llamas, "The Prc and RseP proteases control bacterial cell-surface signalling activity," Environmental Microbiology, vol. 16, no. 8, pp. 2433–2443, 2014, doi: 10.1111/1462-2920.12371.

[90] J. Sawa, A. Heuck, M. Ehrmann, and T. Clausen, "Molecular transformers in the cell: lessons learned from the DegP protease-chaperone," Current Opinion in Structural Biology, vol. 20, no. 2, pp. 253–258, 2010, doi: 10.1016/j.sbi.2010.01.014.

[91] J. Verghese, J. Abrams, Y. Wang, and K. A. Morano, "Biology of the Heat Shock Response and Protein Chaperones: Budding Yeast (Saccharomyces cerevisiae)

as a Model System," Microbiology and Molecular Biology Reviews, vol. 76, no. 2, pp. 115–158, Jun. 2012, doi: 10.1128/MMBR.05018-11.

[92] D. Sharma and D. C. Masison, "Functionally redundant isoforms of a yeast Hsp70 chaperone subfamily have different antiprion effects," Genetics, vol. 179, no. 3, pp. 1301–1311, 2008, doi: 10.1534/genetics.108.089458.

[93] J. Hageman, M. A. W. H. van Waarde, A. Zylicz, D. Walerych, and H. H. Kampinga, "The diverse members of the mammalian HSP70 machine show distinct chaperone-like activities," Biochemical Journal, vol. 435, no. 1, pp. 127–142, 2011, doi: 10.1042/BJ20101247.

[94] W. Meng, E. M. Clerico, N. McArthur, and L. M. Gierasch, "Allosteric landscapes of eukaryotic cytoplasmic Hsp70s are shaped by evolutionary tuning of key interfaces," Proc Natl Acad Sci U S A, vol. 115, no. 47, pp. 11970–11975, 2018, doi: 10.1073/pnas.1811105115.

[95] R. Qi et al., "Allosteric opening of the polypeptide-binding site when an Hsp70 binds ATP," Nature Structural & Molecular Biology, vol. 20, no. 7, pp. 900–907, Jul. 2013, doi: 10.1038/nsmb.2583.

[96] C. Koutras and J. E. A. Braun, "J protein mutations and resulting proteostasis collapse," Frontiers in Cellular Neuroscience, vol. 8, no. JULY, pp. 1–7, 2014, doi: 10.3389/fncel.2014.00191.

[97] J. Radons, "The human HSP70 family of chaperones: where do we stand?," Cell Stress and Chaperones, vol. 21, no. 3, pp. 379–404, 2016, doi: 10.1007/s12192-016-0676-6.

[98] S. Verma, S. Goyal, S. Jamal, A. Singh, and A. Grover, "Hsp90: Friends, clients and natural foes," Biochimie, vol. 127, pp. 227–240, 2016, doi: 10.1016/j.biochi.2016.05.018.

[99] L. H. Pearl, "Review: The HSP90 molecular chaperone - An enigmatic ATPase," Biopolymers, vol. 105, no. 8, pp. 594–607, 2016, doi: 10.1002/bip.22835.

[100] O. O. Odunuga, V. M. Longshaw, and G. L. Blatch, "Hop: More than an Hsp70/Hsp90 adaptor protein," BioEssays, vol. 26, no. 10, pp. 1058–1068, 2004, doi: 10.1002/bies.20107.

[101] J. Shorter, "The mammalian disaggregase machinery: Hsp110 synergizes with Hsp70 and Hsp40 to catalyze protein disaggregation and reactivation in a cell-free system," PLoS ONE, vol. 6, no. 10, 2011, doi: 10.1371/journal.pone.0026319.

[102] R. Bakthisaran, R. Tangirala, and C. M. Rao, "Small heat shock proteins: Role in cellular functions and pathology," Biochimica et Biophysica Acta - Proteins and Proteomics, vol. 1854, no. 4, pp. 291–319, 2015, doi: 10.1016/j.bbapap.2014.12.019.

[103] S. Ungelenk et al., "Small heat shock proteins sequester misfolding proteins in near-native conformation for cellular protection and efficient refolding," Nature Communications, vol. 7, pp. 1–14, 2016, doi: 10.1038/ncomms13673.

[104] H. Sakahira, P. Breuer, M. K. Hayer-Hartl, and F. U. Hartl, "Molecular chaperones as modulators of polyglutamine protein aggregation and toxicity," Proc Natl Acad Sci U S A, vol. 99, no. SUPPL. 4, pp. 16412–16418, 2002, doi: 10.1073/pnas.182426899.

[105] J. Ojha, G. Masilamoni, D. Dunlap, R. A. Udoff, and A. G. Cashikar, "Sequestration of Toxic Oligomers by HspB1 as a Cytoprotective Mechanism," Molecular and Cellular Biology, vol. 31, no. 15, pp. 3146–3157, 2011, doi: 10.1128/mcb.01187-10.

[106] D. Cox and H. Ecroyd, "The small heat shock proteins αB-crystallin (HSPB5) and Hsp27 (HSPB1) inhibit the intracellular aggregation of α-synuclein," Cell Stress and Chaperones, vol. 22, no. 4, pp. 589–600, 2017, doi: 10.1007/s12192-017-0785-x.

[107] F. Li, H. Xiao, F. Zhou, Z. Hu, and B. Yang, "Study of HSPB6: Insights into the Properties of the Multifunctional Protective Agent," Cellular Physiology and Biochemistry, vol. 44, no. 1, pp. 314–332, 2017, doi: 10.1159/000484889.

[108] M. Ganassi et al., "A Surveillance Function of the HSPB8-BAG3-HSP70 Chaperone Complex Ensures Stress Granule Integrity and Dynamism," Molecular Cell, vol. 63, no. 5, pp. 796–810, 2016, doi: 10.1016/j.molcel.2016.07.021.

[109] J. M. Webster, A. L. Darling, V. N. Uversky, and L. J. Blair, "Small heat shock proteins, big impact on protein aggregation in neurodegenerative disease," Frontiers in Pharmacology, vol. 10, no. September, pp. 1–18, 2019, doi: 10.3389/fphar.2019.01047.

[110] J. A. Aquilina, S. Shrestha, A. M. Morris, and H. Ecroyd, "Structural and functional aspects of hetero-oligomers formed by the small heat shock proteins αB-crystallin and HSP27," Journal of Biological Chemistry, vol. 288, no. 19, pp. 13602–13609, 2013, doi: 10.1074/jbc.M112.443812.

[111] J. Behnke, M. J. Feige, and L. M. Hendershot, "BiP and Its Nucleotide Exchange Factors Grp170 and Sil1: Mechanisms of Action and Biological Functions,"

Journal of Molecular Biology, vol. 427, no. 7, pp. 1589–1608, Apr. 2015, doi: 10.1016/j.jmb.2015.02.011.

[112] J. J. Caramelo and A. J. Parodi, "Getting in and out from calnexin/calreticulin cycles," Journal of Biological Chemistry, vol. 283, no. 16, pp. 10221–10225, 2008, doi: 10.1074/jbc.R700048200.

[113] D. Eletto, D. Dersh, and Y. Argon, "GRP94 in ER quality control and stress responses," Seminars in Cell & Developmental Biology, vol. 21, no. 5, pp. 479–485, Jul. 2010, doi: 10.1016/j.semcdb.2010.03.004.

[114] K. F. R. Pobre, G. J. Poet, and L. M. Hendershot, "The endoplasmic reticulum (ER) chaperone BiP is a master regulator of ER functions: Getting by with a little help from ERdj friends," Journal of Biological Chemistry, vol. 294, no. 6, pp. 2098–2108, 2019, doi: 10.1074/jbc.REV118.002804.

[115] B. Coulombe, The Molecular Chaperones Interaction Networks in Protein Folding and Degradation. New York, NY: Springer New York, 2014. doi: 10.1007/978-1-4939-1130-1.

[116] P. de Los Rios, A. Ben-Zvi, O. Slutsky, A. Azem, and P. Goloubinoff, "Hsp70 chaperones accelerate protein translocation and the unfolding of stable protein aggregates by entropic pulling," Proc Natl Acad Sci U S A, vol. 103, no. 16, pp. 6166–6171, 2006, doi: 10.1073/pnas.0510496103.

[117] M. Griesemer, C. Young, A. S. Robinson, and L. Petzold, "BiP Clustering Facilitates Protein Folding in the Endoplasmic Reticulum," PLoS Computational Biology, vol. 10, no. 7, 2014, doi: 10.1371/journal.pcbi.1003675.

[118] R. Ushioda, J. Hoseki, K. Araki, G. Jansen, D. Y. Thomas, and K. Nagata, "ERdj5 Is Required as a Disulfide Reductase for Degradation of Misfolded Proteins in the ER," Science (1979), vol. 321, no. 5888, pp. 569–572, Jul. 2008, doi: 10.1126/science.1159293.

[119] L. Ellgaard and E. M. Frickel, "Calnexin, calreticulin, and ERp57: Teammates in glycoprotein folding," Cell Biochemistry and Biophysics, vol. 39, no. 3, pp. 223–247, 2003, doi: 10.1385/CBB:39:3:223.

[120] K. L. Nielsen, N. McLennan, M. Masters, and N. J. Cowan, "A Single-Ring Mitochondrial Chaperonin (Hsp60-Hsp10) Can Substitute for GroEL-GroES In Vivo," Journal of Bacteriology, vol. 181, no. 18, pp. 5871–5875, Sep. 1999, doi: 10.1128/JB.181.18.5871-5875.1999.

[121] N. Regev-Rudzki, K. Gabriel, and D. Bursać, "The Evolution and Function of Co-Chaperones in Mitochondria," 2015, pp. 201–217. doi: 10.1007/978-3-319-11731-7_10.

[122] F. A. Dekker and S. G. D. Rüdiger, "The Mitochondrial Hsp90 TRAP1 and Alzheimer's Disease," Frontiers in Molecular Biosciences, vol. 8, Jun. 2021, doi: 10.3389/fmolb.2021.697913.

[123] J. Krzewska, T. Langer, and K. Liberek, "Mitochondrial Hsp78, a member of the Clp/Hsp100 family in Saccharomyces cerevisiae , cooperates with Hsp70 in protein refolding," FEBS Letters, vol. 489, no. 1, pp. 92–96, Jan. 2001, doi: 10.1016/S0014-5793(00)02423-6.

[124] E. Ivanova, T. A. Jowitt, and H. Lu, "Assembly of the Mitochondrial Tim9–Tim10 Complex: A Multi-step Reaction with Novel Intermediates," Journal of Molecular Biology, vol. 375, no. 1, pp. 229–239, Jan. 2008, doi: 10.1016/j.jmb.2007.10.037.

[125] K. Weinhäupl et al., "Structural Basis of Membrane Protein Chaperoning through the Mitochondrial Intermembrane Space," Cell, vol. 175, no. 5, pp. 1365-1379.e25, Nov. 2018, doi: 10.1016/j.cell.2018.10.039.

[126] J. A. Bick and B. M. Lange, "Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane," Archives of Biochemistry and Biophysics, vol. 415, no. 2, pp. 146–154, Jul. 2003, doi: 10.1016/S0003-9861(03)00233-9.

[127] Q. Zhao and C. Liu, "Chloroplast Chaperonin: An Intricate Protein Folding Machine for Photosynthesis," Frontiers in Molecular Biosciences, vol. 4, Jan. 2018, doi: 10.3389/fmolb.2017.00098.

[128] C. Liu et al., "The chloroplast HSP70B-CDJ2-CGE1 chaperones catalyse assembly and disassembly of VIPP1 oligomers in Chlamydomonas," The Plant Journal, vol. 50, no. 2, pp. 265–277, Mar. 2007, doi: 10.1111/j.1365-313X.2007.03047.x.

[129] H. Inoue, M. Li, and D. J. Schnell, "An essential role for chloroplast heat shock protein 90 (Hsp90C) in protein import into chloroplasts," Proceedings of the National Academy of Sciences, vol. 110, no. 8, pp. 3173–3178, Feb. 2013, doi: 10.1073/pnas.1219229110.

[130] G. L. Rosano, E. M. Bruch, and E. A. Ceccarelli, "Insights into the CLP/HSP100 Chaperone System from Chloroplasts of Arabidopsis thaliana," Journal

of Biological Chemistry, vol. 286, no. 34, pp. 29671–29680, Aug. 2011, doi: 10.1074/jbc.M110.211946.

[131] H. Li, D. Schnell, and S. M. Theg, "Protein Import Motors in Chloroplasts: On the Role of Chaperones," The Plant Cell, vol. 32, no. 3, pp. 536–542, Mar. 2020, doi: 10.1105/tpc.19.00300.

[132] G. Kleiger and T. Mayor, "Perilous journey: a tour of the ubiquitin–proteasome system," Trends in Cell Biology, vol. 24, no. 6, pp. 352–359, Jun. 2014, doi: 10.1016/j.tcb.2013.12.003.

[133] D. Finley, H. D. Ulrich, T. Sommer, and P. Kaiser, "The Ubiquitin–Proteasome System of Saccharomyces cerevisiae," Genetics, vol. 192, no. 2, pp. 319–360, Oct. 2012, doi: 10.1534/genetics.112.140467.

[134] M. Iconomou and D. N. Saunders, "Systematic approaches to identify E3 ligase substrates," Biochemical Journal, vol. 473, no. 22, pp. 4083–4101, Nov. 2016, doi: 10.1042/BCJ20160719.

[135] C. A. P. Joazeiro, "Mechanisms and functions of ribosome-associated protein quality control," Nature Reviews Molecular Cell Biology, vol. 20, no. 6, pp. 368–383, Jun. 2019, doi: 10.1038/s41580-019-0118-2.

[136] E. M. Clerico, J. M. Tilitsky, W. Meng, and L. M. Gierasch, "How Hsp70 Molecular Machines Interact with Their Substrates to Mediate Diverse Physiological Functions," Journal of Molecular Biology, vol. 427, no. 7, pp. 1575–1588, Apr. 2015, doi: 10.1016/j.jmb.2015.02.004.

[137] O. Faust et al., "HSP40 proteins use class-specific regulation to drive HSP70 functional diversity," Nature, vol. 587, no. 7834, pp. 489–494, Nov. 2020, doi: 10.1038/s41586-020-2906-4.

[138] H. Wegele, S. K. Wandinger, A. B. Schmid, J. Reinstein, and J. Buchner, "Substrate Transfer from the Chaperone Hsp70 to Hsp90," Journal of Molecular Biology, vol. 356, no. 3, pp. 802–811, Feb. 2006, doi: 10.1016/j.jmb.2005.12.008.

[139] S. Rothe, A. Prakash, and J. Tyedmers, "The Insoluble Protein Deposit (IPOD) in Yeast," Frontiers in Molecular Neuroscience, vol. 11, Jul. 2018, doi: 10.3389/fnmol.2018.00237.

[140] E. M. Sontag, R. S. Samant, and J. Frydman, "Mechanisms and Functions of Spatial Protein Quality Control," Annual Review of Biochemistry, vol. 86, no. 1, pp. 97–122, 2017, doi: 10.1146/annurev-biochem-060815-014616.

[141] M. R. Fernández-Fernández, M. Gragera, L. Ochoa-Ibarrola, L. Quintana-Gallardo, and J. M. Valpuesta, "Hsp70 - a master regulator in protein degradation," FEBS Letters, vol. 591, no. 17, pp. 2648–2660, Sep. 2017, doi: 10.1002/1873-3468.12751.

[142] A. Fontanini, R. Chies, E. L. Snapp, M. Ferrarini, G. M. Fabrizi, and C. Brancolini, "Glycan-independent Role of Calnexin in the Intracellular Retention of Charcot-Marie-Tooth 1A Gas3/PMP22 Mutants," Journal of Biological Chemistry, vol. 280, no. 3, pp. 2378–2387, Jan. 2005, doi: 10.1074/jbc.M405104200.

[143] E. Swanton, "Role of calnexin in the glycan-independent quality control of proteolipid protein," The EMBO Journal, vol. 22, no. 12, pp. 2948–2958, Jun. 2003, doi: 10.1093/emboj/cdg300.

[144] S. S. Vembar and J. L. Brodsky, "One step at a time: endoplasmic reticulum-associated degradation," Nature Reviews Molecular Cell Biology, vol. 9, no. 12, pp. 944–957, Dec. 2008, doi: 10.1038/nrm2546.

[145] A. B. Abildgaard et al., "Co-Chaperones in Targeting and Delivery of Misfolded Proteins to the 26S Proteasome," Biomolecules, vol. 10, no. 8, p. 1141, Aug. 2020, doi: 10.3390/biom10081141.

[146] R. Sardana and S. D. Emr, "Membrane Protein Quality Control Mechanisms in the Endo-Lysosome System," Trends in Cell Biology, vol. 31, no. 4, pp. 269–283, Apr. 2021, doi: 10.1016/j.tcb.2020.11.011.

[147] T. T. M. Vu, D. C. Mitchell, S. P. Gygi, and A. Varshavsky, "The Arg/N-degron pathway targets transcription factors and regulates specific genes," Proceedings of the National Academy of Sciences, vol. 117, no. 49, pp. 31094–31104, Dec. 2020, doi: 10.1073/pnas.2020124117.

[148] K. T. Nguyen, S.-H. Mun, C.-S. Lee, and C.-S. Hwang, "Control of protein degradation by N-terminal acetylation and the N-end rule pathway," Experimental & Molecular Medicine, vol. 50, no. 7, pp. 1–8, Jul. 2018, doi: 10.1038/s12276-018-0097-y.

[149] C. C. Schmidt, V. Vasic, and A. Stein, "Doa10 is a membrane protein retrotranslocase in ER-associated protein degradation," Elife, vol. 9, Jun. 2020, doi: 10.7554/eLife.56945.

[150] S.-J. Chen, X. Wu, B. Wadas, J.-H. Oh, and A. Varshavsky, "An N-end rule pathway that recognizes proline and destroys gluconeogenic enzymes," Science (1979), vol. 355, no. 6323, Jan. 2017, doi: 10.1126/science.aal3655.

[151] J.-M. Kim et al., "Formyl-methionine as an N-degron of a eukaryotic N-end rule pathway," Science (1979), vol. 362, no. 6418, p. eaat0174, Nov. 2018, doi: 10.1126/science.aat0174.

[152] R. T. Timms, Z. Zhang, D. Y. Rhee, J. W. Harper, I. Koren, and S. J. Elledge, "A glycine-specific N-degron pathway mediates the quality control of protein N -myristoylation," Science (1979), vol. 365, no. 6448, Jul. 2019, doi: 10.1126/science.aaw4912.

[153] I. Koren, R. T. Timms, T. Kula, Q. Xu, M. Z. Li, and S. J. Elledge, "The Eukaryotic Proteome Is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons," Cell, vol. 173, no. 7, pp. 1622-1635.e14, Jun. 2018, doi: 10.1016/j.cell.2018.04.028.

[154] K. E. Johansson, B. Mashahreh, R. Hartmann-Petersen, T. Ravid, and K. Lindorff-Larsen, "Prediction of quality-control degradation signals in yeast proteins," bioRxiv, Apr. 2022.

[155] A. K. Verma et al., "The expanding world of plant J-domain proteins," Critical Reviews in Plant Sciences, vol. 38, no. 5–6, pp. 382–400, Nov. 2019, doi: 10.1080/07352689.2019.1693716.

[156] M. E. Rebeaud, S. Mallik, P. Goloubinoff, and D. S. Tawfik, "On the evolution of chaperones and cochaperones and the expansion of proteomes across the Tree of Life," Proceedings of the National Academy of Sciences, vol. 118, no. 21, p. e2020885118, May 2021, doi: 10.1073/pnas.2020885118.

[157] H. J. An, J. W. Froehlich, and C. B. Lebrilla, "Determination of glycosylation sites and site-specific heterogeneity in glycoproteins," Current Opinion in Chemical Biology, vol. 13, no. 4, pp. 421–426, Oct. 2009, doi: 10.1016/j.cbpa.2009.07.022.

[158] N. Mitra, S. Sinha, T. N. C. Ramya, and A. Surolia, "N-linked oligosaccharides as outfitters for glycoprotein folding, form and function," Trends in Biochemical Sciences, vol. 31, no. 3, pp. 156–163, Mar. 2006, doi: 10.1016/j.tibs.2006.01.003.

[159] J. L. Price et al., "N-glycosylation of enhanced aromatic sequons to increase glycoprotein stability," Biopolymers, vol. 98, no. 3, pp. 195–211, 2012, doi: 10.1002/bip.22030.

[160] H. Yamaguchi, "Chaperone-Like Functions of N-Glycans in the Formation and Stabilization of Protein Conformation.," Trends in Glycoscience and Glycotechnology, vol. 14, no. 77, pp. 139–151, 2002, doi: 10.4052/tigg.14.139.

[161] G. Walsh and R. Jefferis, “Post-translational modifications in the context of therapeutic proteins,” Nature Biotechnology, vol. 24, no. 10, pp. 1241–1252, Oct. 2006, doi: 10.1038/nbt1252.

[162] H. Nothaft and C. M. Szymanski, “Protein glycosylation in bacteria: sweeter than ever,” Nature Reviews Microbiology, vol. 8, no. 11, pp. 765–778, Nov. 2010, doi: 10.1038/nrmicro2383.

[163] C. Bonzom et al., “Glycosylation influences activity, stability and immobilization of the feruloyl esterase 1a from Myceliophthora thermophila,” AMB Express, vol. 9, no. 1, p. 126, Dec. 2019, doi: 10.1186/s13568-019-0852-z.

[164] D. Shental-Bechor and Y. Levy, “Effect of glycosylation on protein folding: A close look at thermodynamic stabilization,” Proceedings of the National Academy of Sciences, vol. 105, no. 24, pp. 8256–8261, Jun. 2008, doi: 10.1073/pnas.0801340105.

[165] H. S. Lee, Y. Qi, and W. Im, “Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study,” Scientific Reports, vol. 5, no. 1, p. 8926, Aug. 2015, doi: 10.1038/srep08926.

[166] P. Goettig, “Effects of Glycosylation on the Enzymatic Activity and Mechanisms of Proteases,” International Journal of Molecular Sciences, vol. 17, no. 12, p. 1969, Nov. 2016, doi: 10.3390/ijms17121969.

[167] H. Qiu et al., “Impact of cysteine variants on the structure, activity, and stability of recombinant human α-galactosidase A,” Protein Science, vol. 24, no. 9, pp. 1401–1411, Sep. 2015, doi: 10.1002/pro.2719.

[168] M. A. Weiss, “Diabetes mellitus due to the toxic misfolding of proinsulin variants,” FEBS Letters, vol. 587, no. 13, pp. 1942–1950, Jun. 2013, doi: 10.1016/j.febslet.2013.04.044.

[169] P. Pecher and U. Arnold, “The effect of additional disulfide bonds on the stability and folding of ribonuclease A,” Biophysical Chemistry, vol. 141, no. 1, pp. 21–28, Apr. 2009, doi: 10.1016/j.bpc.2008.12.005.

[170] L. Ellgaard, N. McCaul, A. Chatsisvili, and I. Braakman, “Co- and Post-Translational Protein Folding in the ER,” Traffic, vol. 17, no. 6, pp. 615–638, Jun. 2016, doi: 10.1111/tra.12392.

[171] P. Kosuri et al., “Protein Folding Drives Disulfide Formation,” Cell, vol. 151, no. 4, pp. 794–806, Nov. 2012, doi: 10.1016/j.cell.2012.09.036.

[172] F. Hatahet, D. Boyd, and J. Beckwith, "Disulfide bond formation in prokaryotes: history, diversity and design.," Biochim Biophys Acta, vol. 1844, no. 8, pp. 1402–14, Aug. 2014, doi: 10.1016/j.bbapap.2014.02.014.

[173] M. E. Reardon-Robinson and H. Ton-That, "Disulfide-Bond-Forming Pathways in Gram-Positive Bacteria," Journal of Bacteriology, vol. 198, no. 5, pp. 746–754, Mar. 2016, doi: 10.1128/JB.00769-15.

[174] S. Gullón, S. Marín, and R. P. Mellado, "Four thiol-oxidoreductases involved in the formation of disulphide bonds in the Streptomyces lividans TK21 secretory proteins," Microbial Cell Factories, vol. 18, no. 1, p. 126, Dec. 2019, doi: 10.1186/s12934-019-1175-0.

[175] W.-H. Xiong, M. Qin, and H. Zhong, "Myristoylation alone is sufficient for PKA catalytic subunits to associate with the plasma membrane to regulate neuronal functions," Proceedings of the National Academy of Sciences, vol. 118, no. 15, p. e2021658118, Apr. 2021, doi: 10.1073/pnas.2021658118.

[176] D. Shental-Bechor et al., "Nonnative interactions regulate folding and switching of myristoylated protein," Proceedings of the National Academy of Sciences, vol. 109, no. 44, pp. 17839–17844, Oct. 2012, doi: 10.1073/pnas.1201803109.

[177] J. M. Glück, S. Hoffmann, B. W. Koenig, and D. Willbold, "Single Vector System for Efficient N-myristoylation of Recombinant Proteins in E. coli," PLoS ONE, vol. 5, no. 4, p. e10081, Apr. 2010, doi: 10.1371/journal.pone.0010081.

[178] Y. Zhang, Z. Qin, W. Sun, F. Chu, and F. Zhou, "Function of Protein S-Palmitoylation in Immunity and Immune-Related Diseases," Frontiers in Immunology, vol. 12, Sep. 2021, doi: 10.3389/fimmu.2021.661202.

[179] J. Sobocińska, P. Roszczenko-Jasińska, A. Ciesielska, and K. Kwiatkowska, "Protein Palmitoylation and Its Role in Bacterial and Viral Infections," Frontiers in Immunology, vol. 8, Jan. 2018, doi: 10.3389/fimmu.2017.02003.

[180] H. Nakayama, K. Kurokawa, and B. L. Lee, "Lipoproteins in bacteria: structures and biosynthetic pathways," FEBS Journal, vol. 279, no. 23, pp. 4247–4268, Dec. 2012, doi: 10.1111/febs.12041.

[181] D. Stiefler-Jensen et al., "The extraordinary thermal stability of EstA from S. islandicus is independent of post translational modifications," Protein Science, vol. 26, no. 9, pp. 1819–1827, Sep. 2017, doi: 10.1002/pro.3220.

[182] E. Samatova, J. Daberger, M. Liutkute, and M. v. Rodnina, "Translational Control by Ribosome Pausing in Bacteria: How a Non-uniform Pace of Translation Affects Protein Production and Folding," Frontiers in Microbiology, vol. 11, Jan. 2021, doi: 10.3389/fmicb.2020.619430.

[183] S. K. Behura and D. W. Severson, "Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes," Biological Reviews, vol. 88, no. 1, pp. 49–61, Feb. 2013, doi: 10.1111/j.1469-185X.2012.00242.x.

[184] N. Neelagandan, I. Lamberti, H. J. F. Carvalho, C. Gobet, and F. Naef, "What determines eukaryotic translation elongation: recent molecular and quantitative analyses of protein synthesis," Open Biology, vol. 10, no. 12, p. 200292, Dec. 2020, doi: 10.1098/rsob.200292.

[185] T. F. Clarke and P. L. Clark, "Rare Codons Cluster," PLoS ONE, vol. 3, no. 10, p. e3412, Oct. 2008, doi: 10.1371/journal.pone.0003412.

[186] M. Chartier, F. Gaudreault, and R. Najmanovich, "Large-scale analysis of conserved rare codon clusters suggests an involvement in co-translational molecular recognition events," Bioinformatics, vol. 28, no. 11, pp. 1438–1445, Jun. 2012, doi: 10.1093/bioinformatics/bts149.

[187] P. Huter et al., "Structural Basis for Polyproline-Mediated Ribosome Stalling and Rescue by the Translation Elongation Factor EF-P," Molecular Cell, vol. 68, no. 3, pp. 515-527.e6, Nov. 2017, doi: 10.1016/j.molcel.2017.10.014.

[188] C. Bao, S. Loerch, C. Ling, A. A. Korostelev, N. Grigorieff, and D. N. Ermolenko, "mRNA stem-loops can pause the ribosome by hindering A-site tRNA binding," Elife, vol. 9, May 2020, doi: 10.7554/eLife.55799.

[189] M. Bañó-Polo et al., "Transmembrane but not soluble helices fold inside the ribosome tunnel," Nature Communications, vol. 9, no. 1, p. 5246, Dec. 2018, doi: 10.1038/s41467-018-07554-7.

[190] C. A. Waudby, C. Burridge, L. D. Cabrita, and J. Christodoulou, "Thermodynamics of co-translational folding and ribosome–nascent chain interactions," Current Opinion in Structural Biology, vol. 74, p. 102357, Jun. 2022, doi: 10.1016/j.sbi.2022.102357.

[191] C. Kimchi-Sarfaty et al., "A 'Silent' Polymorphism in the MDR 1 Gene Changes Substrate Specificity," Science (1979), vol. 315, no. 5811, pp. 525–528, Jan. 2007, doi: 10.1126/science.1135308.

[192] F. Buhr et al., "Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations," Molecular Cell, vol. 61, no. 3, pp. 341–351, Feb. 2016, doi: 10.1016/j.molcel.2016.01.008.

[193] G. Zhang, M. Hubalewska, and Z. Ignatova, "Transient ribosomal attenuation coordinates protein synthesis and co-translational folding," Nature Structural & Molecular Biology, vol. 16, no. 3, pp. 274–280, Mar. 2009, doi: 10.1038/nsmb.1554.

[194] I. M. Walsh, M. A. Bowman, I. F. Soto Santarriaga, A. Rodriguez, and P. L. Clark, "Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness," Proceedings of the National Academy of Sciences, vol. 117, no. 7, pp. 3528–3534, Feb. 2020, doi: 10.1073/pnas.1907126117.

[195] M. J. M. Niesen, C. Y. Wang, R. C. van Lehn, and T. F. Miller, "Structurally detailed coarse-grained model for Sec-facilitated co-translational protein translocation and membrane integration," PLOS Computational Biology, vol. 13, no. 3, p. e1005427, Mar. 2017, doi: 10.1371/journal.pcbi.1005427.

[196] E. Mercier, W. Wintermeyer, and M. v Rodnina, "Co-translational insertion and topogenesis of bacterial membrane proteins monitored in real time," The EMBO Journal, vol. 39, no. 15, Aug. 2020, doi: 10.15252/embj.2019104054.

[197] Y.-W. Shieh et al., "Operon structure and cotranslational subunit association direct protein assembly in bacteria," Science (1979), vol. 350, no. 6261, pp. 678–680, Nov. 2015, doi: 10.1126/science.aac8171.

[198] J. A. Marsh et al., "Protein Complexes Are under Evolutionary Selection to Assemble via Ordered Pathways," Cell, vol. 153, no. 2, pp. 461–470, Apr. 2013, doi: 10.1016/j.cell.2013.02.044.

[199] C. D. S. Duncan and J. Mata, "Widespread Cotranslational Formation of Protein Complexes," PLoS Genetics, vol. 7, no. 12, p. e1002398, Dec. 2011, doi: 10.1371/journal.pgen.1002398.

[200] A. Shiber et al., "Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling," Nature, vol. 561, no. 7722, pp. 268–272, Sep. 2018, doi: 10.1038/s41586-018-0462-y.

[201] J. Koubek, J. Schmitt, C. V. Galmozzi, and G. Kramer, "Mechanisms of Cotranslational Protein Maturation in Bacteria," Frontiers in Molecular Biosciences, vol. 8, May 2021, doi: 10.3389/fmolb.2021.689755.

[202] M. Lynch, "The Evolution of Multimeric Protein Assemblages," Molecular Biology and Evolution, vol. 29, no. 5, pp. 1353–1366, May 2012, doi: 10.1093/molbev/msr300.

[203] K.-Y. E. Kong, J. P. L. Coelho, M. J. Feige, and A. Khmelinskii, "Quality control of mislocalized and orphan proteins," Experimental Cell Research, vol. 403, no. 2, p. 112617, Jun. 2021, doi: 10.1016/j.yexcr.2021.112617.

[204] V. N. Uversky, "Intrinsically Disordered Proteins and Their 'Mysterious' (Meta)Physics," Frontiers in Physics, vol. 7, Feb. 2019, doi: 10.3389/fphy.2019.00010.

[205] V. N. Uversky, "Paradoxes and wonders of intrinsic disorder: Stability of instability," Intrinsically Disordered Proteins, vol. 5, no. 1, p. e1327757, Jan. 2017, doi: 10.1080/21690707.2017.1327757.

[206] F. Lermyte, "Roles, Characteristics, and Analysis of Intrinsically Disordered Proteins: A Minireview," Life, vol. 10, no. 12, p. 320, Nov. 2020, doi: 10.3390/life10120320.

[207] G. Fuertes, L. Nevola, and S. Esteban-Martín, "Perspectives on drug discovery strategies based on IDPs," in Intrinsically Disordered Proteins, Elsevier, 2019, pp. 275–327. doi: 10.1016/B978-0-12-816348-1.00009-0.

[208] S. DeForte and V. N. Uversky, "Not an exception to the rule: the functional significance of intrinsically disordered protein regions in enzymes," Molecular BioSystems, vol. 13, no. 3, pp. 463–469, 2017, doi: 10.1039/C6MB00741D.

[209] S. P. Graether, "Troubleshooting Guide to Expressing Intrinsically Disordered Proteins for Use in NMR Experiments," Frontiers in Molecular Biosciences, vol. 5, Jan. 2019, doi: 10.3389/fmolb.2018.00118.

[210] A. H. M. Ng, N. N. Fang, S. A. Comyn, J. Gsponer, and T. Mayor, "System-wide Analysis Reveals Intrinsically Disordered Proteins Are Prone to Ubiquitylation after Misfolding Stress," Molecular & Cellular Proteomics, vol. 12, no. 9, pp. 2456–2467, Sep. 2013, doi: 10.1074/mcp.M112.023416.

[211] E. K. Fredrickson, S. v. Clowes Candadai, C. H. Tam, and R. G. Gardner, "Means of self-preservation: how an intrinsically disordered ubiquitin-protein ligase averts self-destruction," Molecular Biology of the Cell, vol. 24, no. 7, pp. 1041–1052, Apr. 2013, doi: 10.1091/mbc.e12-11-0811.

[212] A. Bhatwa, W. Wang, Y. I. Hassan, N. Abraham, X.-Z. Li, and T. Zhou, "Challenges Associated With the Formation of Recombinant Protein Inclusion

Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications," Frontiers in Bioengineering and Biotechnology, vol. 9, Feb. 2021, doi: 10.3389/fbioe.2021.630551.

[213] J. A. Schellman, "Temperature, stability, and the hydrophobic interaction," Biophysical Journal, vol. 73, no. 6, pp. 2960–2964, Dec. 1997, doi: 10.1016/S0006-3495(97)78324-3.

[214] R. L. Crisman and T. W. Randolph, "Refolding of proteins from inclusion bodies is favored by a diminished hydrophobic effect at elevated pressures," Biotechnology and Bioengineering, vol. 102, no. 2, pp. 483–492, Feb. 2009, doi: 10.1002/bit.22082.

[215] A. Farewell and F. C. Neidhardt, "Effect of temperature on in vivo protein synthetic capacity in Escherichia coli.," J Bacteriol, vol. 180, no. 17, pp. 4704–10, Sep. 1998, doi: 10.1128/JB.180.17.4704-4710.1998.

[216] A. Vera, N. González-Montalbán, A. Arís, and A. Villaverde, "The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures," Biotechnology and Bioengineering, vol. 96, no. 6, pp. 1101–1106, Apr. 2007, doi: 10.1002/bit.21218.

[217] L. Niiranen et al., "Comparative expression study to increase the solubility of cold adapted Vibrio proteins in Escherichia coli," Protein Expression and Purification, vol. 52, no. 1, pp. 210–218, Mar. 2007, doi: 10.1016/j.pep.2006.09.005.

[218] M. Gutiérrez-González et al., "Optimization of culture conditions for the expression of three different insoluble proteins in Escherichia coli," Scientific Reports, vol. 9, no. 1, p. 16850, Dec. 2019, doi: 10.1038/s41598-019-53200-7.

[219] H. Tegel, J. Ottosson, and S. Hober, "Enhancing the protein production levels in Escherichia coli with a strong promoter," FEBS Journal, vol. 278, no. 5, pp. 729–739, Mar. 2011, doi: 10.1111/j.1742-4658.2010.07991.x.

[220] J. Hothersall, R. E. Godfrey, C. Fanitsios, T. W. Overton, S. J. W. Busby, and D. F. Browning, "The PAR promoter expression system: Modified lac promoters for controlled recombinant protein production in Escherichia coli," New Biotechnology, vol. 64, pp. 1–8, Sep. 2021, doi: 10.1016/j.nbt.2021.05.001.

[221] G. Lozano Terol, J. Gallego-Jara, R. A. Sola Martínez, A. Martínez Vivancos, M. Cánovas Díaz, and T. de Diego Puente, "Impact of the Expression System on Recombinant Protein Production in Escherichia coli BL21," Frontiers in Microbiology, vol. 12, Jun. 2021, doi: 10.3389/fmicb.2021.682001.

[222] M. de Mey, J. Maertens, G. J. Lequeux, W. K. Soetaert, and E. J. Vandamme, "Construction and model-based analysis of a promoter library for E. coli: an indispensable tool for metabolic engineering," BMC Biotechnology, vol. 7, no. 1, p. 34, 2007, doi: 10.1186/1472-6750-7-34.

[223] M. Rennig et al., "Industrializing a Bacterial Strain for <scp>l</scp> -Serine Production through Translation Initiation Optimization," ACS Synthetic Biology, vol. 8, no. 10, pp. 2347–2358, Oct. 2019, doi: 10.1021/acssynbio.9b00169.

[224] M. T. Bonde et al., "Predictable tuning of protein expression in bacteria," Nature Methods, vol. 13, no. 3, pp. 233–236, Mar. 2016, doi: 10.1038/nmeth.3727.

[225] P. M. Sharp and W.-H. Li, "The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications," Nucleic Acids Research, vol. 15, no. 3, pp. 1281–1295, 1987, doi: 10.1093/nar/15.3.1281.

[226] N. A. Burgess-Brown, S. Sharma, F. Sobott, C. Loenarz, U. Oppermann, and O. Gileadi, "Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study," Protein Expression and Purification, vol. 59, no. 1, pp. 94–102, May 2008, doi: 10.1016/j.pep.2008.01.008.

[227] H. G. Menzella, "Comparison of two codon optimization strategies to enhance recombinant protein production in Escherichia coli," Microbial Cell Factories, vol. 10, no. 1, p. 15, 2011, doi: 10.1186/1475-2859-10-15.

[228] E. van Aalst, M. Yekefallah, A. K. Mehta, I. Eason, and B. Wylie, "Codon Harmonization of a Kir3.1-KirBac1.3 Chimera for Structural Study Optimization," Biomolecules, vol. 10, no. 3, p. 430, Mar. 2020, doi: 10.3390/biom10030430.

[229] J. Konczal, J. Bower, and C. H. Gray, "Re-introducing non-optimal synonymous codons into codon-optimized constructs enhances soluble recovery of recombinant proteins from Escherichia coli," PLOS ONE, vol. 14, no. 4, p. e0215892, Apr. 2019, doi: 10.1371/journal.pone.0215892.

[230] E. Angov, P. M. Legler, and R. M. Mease, "Adjustment of Codon Usage Frequencies by Codon Harmonization Improves Protein Expression and Folding," 2011, pp. 1–13. doi: 10.1007/978-1-61737-967-3_1.

[231] C. Mignon et al., "Codon harmonization – going beyond the speed limit for protein expression," FEBS Letters, vol. 592, no. 9, pp. 1554–1564, May 2018, doi: 10.1002/1873-3468.13046.

[232] P. Rehbein, J. Berz, P. Kreisel, and H. Schwalbe, "'CodonWizard' – An intuitive software tool with graphical user interface for customizable codon

optimization in protein expression efforts," Protein Expression and Purification, vol. 160, pp. 84–93, Aug. 2019, doi: 10.1016/j.pep.2019.03.018.

[233] G. Wright, A. Rodriguez, J. Li, T. Milenkovic, S. J. Emrich, and P. L. Clark, "<scp>CHARMING</scp> : Harmonizing synonymous codon usage to replicate a desired codon usage pattern," Protein Science, vol. 31, no. 1, pp. 221–231, Jan. 2022, doi: 10.1002/pro.4223.

[234] K. D. Pryor and B. Leiting, "High-Level Expression of Soluble Protein inEscherichia coliUsing a His6-Tag and Maltose-Binding-Protein Double-Affinity Fusion System," Protein Expression and Purification, vol. 10, no. 3, pp. 309–319, Aug. 1997, doi: 10.1006/prep.1997.0759.

[235] T. T. T. Vu et al., "Soluble Prokaryotic Expression and Purification of Human Interferon Alpha-2b Using a Maltose-Binding Protein Tag," Journal of Molecular Microbiology and Biotechnology, vol. 26, no. 6, pp. 359–368, 2016, doi: 10.1159/000446962.

[236] M. T. Nguyen et al., "Prokaryotic Soluble Overexpression and Purification of Human VEGF165 by Fusion to a Maltose Binding Protein Tag," PLOS ONE, vol. 11, no. 5, p. e0156296, May 2016, doi: 10.1371/journal.pone.0156296.

[237] H. Bach et al., "Escherichia coli maltose-binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-chain antibodies 1 1Edited by R. Huber," Journal of Molecular Biology, vol. 312, no. 1, pp. 79–93, Sep. 2001, doi: 10.1006/jmbi.2001.4914.

[238] S. Raran-Kurussi and D. S. Waugh, "The Ability to Enhance the Solubility of Its Fusion Partners Is an Intrinsic Property of Maltose-Binding Protein but Their Folding Is Either Spontaneous or Chaperone-Mediated," PLoS ONE, vol. 7, no. 11, p. e49589, Nov. 2012, doi: 10.1371/journal.pone.0049589.

[239] D. Kuo, M. Nie, and A. J. Courey, "SUMO as a Solubility Tag and In Vivo Cleavage of SUMO Fusion Proteins with Ulp1," 2014, pp. 71–80. doi: 10.1007/978-1-4939-1034-2_6.

[240] S. Costa, A. Almeida, A. Castro, and L. Domingues, "Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system," Frontiers in Microbiology, vol. 5, 2014, doi: 10.3389/fmicb.2014.00063.

[241] Y.-B. Zhang, J. Howitt, S. McCorkle, P. Lawrence, K. Springer, and P. Freimuth, "Protein aggregation during overexpression limited by peptide extensions with large net negative charge," Protein Expression and Purification, vol. 36, no. 2, pp. 207–216, Aug. 2004, doi: 10.1016/j.pep.2004.04.020.

[242] B. H. Jo, "An Intrinsically Disordered Peptide Tag that Confers an Unusual Solubility to Aggregation-Prone Proteins," Applied and Environmental Microbiology, vol. 88, no. 7, Apr. 2022, doi: 10.1128/aem.00097-22.

[243] E. Morimoto, K. Tsuboyama, and Y. Tomari, "Stabilization of recombinant proteins by fusion with heat-resistant obscure (Hero) proteins," BioRxiv, Mar. 2022.

[244] S. Nallamsetty and D. S. Waugh, "Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners," Protein Expression and Purification, vol. 45, no. 1, pp. 175–182, Jan. 2006, doi: 10.1016/j.pep.2005.06.012.

[245] D. Esposito and D. K. Chatterjee, "Enhancement of soluble protein expression through the use of fusion tags," Current Opinion in Biotechnology, vol. 17, no. 4, pp. 353–358, Aug. 2006, doi: 10.1016/j.copbio.2006.06.003.

[246] A. Gräwe and V. Stein, "Linker Engineering in the Context of Synthetic Protein Switches and Sensors," Trends in Biotechnology, vol. 39, no. 7, pp. 731–744, Jul. 2021, doi: 10.1016/j.tibtech.2020.11.007.

[247] X. Han, W. Ning, X. Ma, X. Wang, and K. Zhou, "Improving protein solubility and activity by introducing small peptide tags designed with machine learning models," Metabolic Engineering Communications, vol. 11, p. e00138, Dec. 2020, doi: 10.1016/j.mec.2020.e00138.

[248] D. B. Smith and K. S. Johnson, "Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase," Gene, vol. 67, no. 1, pp. 31–40, Jul. 1988, doi: 10.1016/0378-1119(88)90005-4.

[249] K. Shimada, M. Nagano, M. Kawai, and H. Koga, "Influences of amino acid features of glutathioneS-transferase fusion proteins on their solubility," PROTEOMICS, vol. 5, no. 15, pp. 3859–3863, Oct. 2005, doi: 10.1002/pmic.200402085.

[250] E. R. LaVallie, E. A. DiBlasio-Smith, L. A. Collins-Racie, Z. Lu, and J. M. McCoy, "Thioredoxin and Related Proteins as Multifunctional Fusion Tags for Soluble Expression in E. coli," in E. coli Gene Expression Protocols, vol. 205, New Jersey: Humana Press, 2003, pp. 119–140. doi: 10.1385/1-59259-301-1:119.

[251] E. Guisbert, T. Yura, V. A. Rhodius, and C. A. Gross, "Convergence of Molecular, Modeling, and Systems Approaches for an Understanding of the Escherichia coli Heat Shock Response," Microbiology and Molecular Biology Reviews, vol. 72, no. 3, pp. 545–554, Sep. 2008, doi: 10.1128/MMBR.00007-08.

[252] X. Zhang, Y. Liu, J. C. Genereux, C. Nolan, M. Singh, and J. W. Kelly, "Heat-Shock Response Transcriptional Program Enables High-Yield and High-Quality Recombinant Protein Production in Escherichia coli," ACS Chemical Biology, vol. 9, no. 9, pp. 1945–1949, Sep. 2014, doi: 10.1021/cb5004477.

[253] M. Mamipour, M. Yousefi, and M. Hasanzadeh, "An overview on molecular chaperones enhancing solubility of expressed recombinant proteins with correct folding," International Journal of Biological Macromolecules, vol. 102, pp. 367–375, Sep. 2017, doi: 10.1016/j.ijbiomac.2017.04.025.

[254] A. de Marco, E. Deuerling, A. Mogk, T. Tomoyasu, and B. Bukau, "Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli," BMC Biotechnology, vol. 7, no. 1, p. 32, Dec. 2007, doi: 10.1186/1472-6750-7-32.

[255] Y. Tong et al., "Enhancement of soluble expression of codon-optimized Thermomicrobium roseum sarcosine oxidase in Escherichia coli via chaperone co-expression," Journal of Biotechnology, vol. 218, pp. 75–84, Jan. 2016, doi: 10.1016/j.jbiotec.2015.11.018.

[256] D.-C. Piao et al., "Trigger factor assisted soluble expression of recombinant spike protein of porcine epidemic diarrhea virus in Escherichia coli," BMC Biotechnology, vol. 16, no. 1, p. 39, Dec. 2016, doi: 10.1186/s12896-016-0268-7.

[257] T. Ahn and C.-H. Yun, "High-level expression of human cytochrome P450 3A4 by co-expression with Human molecular chaperone HDJ-1 (Hsp40)," Archives of Pharmacal Research, vol. 27, no. 3, pp. 319–323, Mar. 2004, doi: 10.1007/BF02980067.

[258] L. L. Stephens, A. Shonhai, and G. L. Blatch, "Co-expression of the Plasmodium falciparum molecular chaperone, PfHsp70, improves the heterologous production of the antimalarial drug target GTP cyclohydrolase I, PfGCHI," Protein Expression and Purification, vol. 77, no. 2, pp. 159–165, Jun. 2011, doi: 10.1016/j.pep.2011.01.005.

[259] M. Shanmugasundaram, N. v. Pavlova, A. R. Pavlov, I. K. Lednev, and F. T. Robb, "Improved folding of recombinant protein via co-expression of exogenous chaperones," 2021, pp. 145–170. doi: 10.1016/bs.mie.2021.09.001.

[260] X. Yan, S. Hu, Y.-X. Guan, and S.-J. Yao, "Coexpression of chaperonin GroEL/GroES markedly enhanced soluble and functional expression of recombinant human interferon-gamma in Escherichia coli," Applied Microbiology and

Biotechnology, vol. 93, no. 3, pp. 1065–1074, Feb. 2012, doi: 10.1007/s00253-011-3599-2.

[261] M. V. Ushasree, J. Vidya, and A. Pandey, "Gene cloning and soluble expression of Aspergillus niger phytase in E. coli cytosol via chaperone co-expression," Biotechnology Letters, vol. 36, no. 1, pp. 85–91, Jan. 2014, doi: 10.1007/s10529-013-1322-3.

[262] Z.-L. Wu, J. Qiao, Z.-G. Zhang, F. P. Guengerich, Y. Liu, and X.-Q. Pei, "Enhanced bacterial expression of several mammalian cytochrome P450s by codon optimization and chaperone coexpression," Biotechnology Letters, vol. 31, no. 10, pp. 1589–1593, Oct. 2009, doi: 10.1007/s10529-009-0059-5.

[263] D. G. Kang et al., "Coexpression of molecular chaperone enhances activity and export of organophosphorus hydrolase in Escherichia coli," Biotechnology Progress, vol. 28, no. 4, pp. 925–930, Jul. 2012, doi: 10.1002/btpr.1556.

[264] A. R. Lee, L. Li, S.-Y. Shin, J. S. Moon, H.-J. Eom, and N. S. Han, "Soluble Expression of the Fucosyltransferase Gene from Helicobacter pylori in Escherichia coli by Co-expression of Molecular Chaperones," Microbiology and Biotechnology Letters, vol. 43, no. 3, pp. 212–218, Sep. 2015, doi: 10.4014/mbl.1507.07004.

[265] J. D. Valderrama-Rincon et al., "An engineered eukaryotic protein glycosylation pathway in Escherichia coli," Nature Chemical Biology, vol. 8, no. 5, pp. 434–436, May 2012, doi: 10.1038/nchembio.921.

[266] A. Naegeli et al., "Molecular Analysis of an Alternative N-Glycosylation Machinery by Functional Transfer from Actinobacillus pleuropneumoniae to Escherichia coli," Journal of Biological Chemistry, vol. 289, no. 4, pp. 2170–2179, Jan. 2014, doi: 10.1074/jbc.M113.524462.

[267] C. J. Glasscock et al., "A flow cytometric approach to engineering Escherichia coli for improved eukaryotic protein glycosylation," Metabolic Engineering, vol. 47, pp. 488–495, May 2018, doi: 10.1016/j.ymben.2018.04.014.

[268] R. Rouet et al., "Expression of high-affinity human antibody fragments in bacteria," Nature Protocols, vol. 7, no. 2, pp. 364–373, Feb. 2012, doi: 10.1038/nprot.2011.448.

[269] G. Rigi et al., "Optimization of expression, purification and secretion of functional recombinant human growth hormone in Escherichia coli using modified staphylococcal protein a signal peptide," BMC Biotechnology, vol. 21, no. 1, p. 51, Dec. 2021, doi: 10.1186/s12896-021-00701-x.

[270] G. R. M. Kleiner-Grote, J. M. Risse, and K. Friehs, “Secretion of recombinant proteins from E. coli,” Engineering in Life Sciences, vol. 18, no. 8, pp. 532–550, Aug. 2018, doi: 10.1002/elsc.201700200.

[271] A. Karyolaimos et al., “Enhancing Recombinant Protein Yields in the E. coli Periplasm by Combining Signal Peptide and Production Rate Screening,” Frontiers in Microbiology, vol. 10, Jul. 2019, doi: 10.3389/fmicb.2019.01511.

[272] K. Mirzadeh et al., “Increased production of periplasmic proteins in Escherichia coli by directed evolution of the translation initiation region,” Microbial Cell Factories, vol. 19, no. 1, p. 85, Dec. 2020, doi: 10.1186/s12934-020-01339-8.

[273] D. Ow, D. Lim, P. Morin Nissom, A. Camattari, and V. Wong, “Co-expression of Skp and FkpA chaperones improves cell viability and alters the global expression of stress response genes during scFvD1.3 production,” Microbial Cell Factories, vol. 9, no. 1, p. 22, 2010, doi: 10.1186/1475-2859-9-22.

[274] L. C. Shriver-Lake, E. R. Goldman, D. Zabetakis, and G. P. Anderson, “Improved production of single domain antibodies with two disulfide bonds by co-expression of chaperone proteins in the Escherichia coli periplasm,” Journal of Immunological Methods, vol. 443, pp. 64–67, Apr. 2017, doi: 10.1016/j.jim.2017.01.007.

[275] A. de Marco, “Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli,” Microbial Cell Factories, vol. 8, no. 1, p. 26, 2009, doi: 10.1186/1475-2859-8-26.

[276] A. Gąciarz et al., “Efficient soluble expression of disulfide bonded proteins in the cytoplasm of Escherichia coli in fed-batch fermentations on chemically defined minimal media,” Microbial Cell Factories, vol. 16, no. 1, p. 108, Dec. 2017, doi: 10.1186/s12934-017-0721-x.

[277] A. B. Bertelsen et al., “DisCoTune: versatile auxiliary plasmids for the production of disulphide-containing proteins and peptides in the E. coli T7 system,” Microbial Biotechnology, vol. 14, no. 6, pp. 2566–2580, Nov. 2021, doi: 10.1111/1751-7915.13895.

[278] H. Tegel, S. Tourle, J. Ottosson, and A. Persson, “Increased levels of recombinant human proteins with the Escherichia coli strain Rosetta(DE3),” Protein Expression and Purification, vol. 69, no. 2, pp. 159–167, Feb. 2010, doi: 10.1016/j.pep.2009.08.017.

[279] K. M. Søgaard and M. H. H. Nørholm, "Side effects of extra tRNA supplied in a typical bacterial protein production scenario," Protein Science, vol. 25, no. 11, pp. 2102–2108, Nov. 2016, doi: 10.1002/pro.3011.

[280] Z. Lipinszki et al., "Enhancing the Translational Capacity of E. coli by Resolving the Codon Bias," ACS Synthetic Biology, vol. 7, no. 11, pp. 2656–2664, Nov. 2018, doi: 10.1021/acssynbio.8b00332.

[281] C. Payen and D. Thompson, "The renaissance of yeasts as microbial factories in the modern age of biomanufacturing," Yeast, vol. 36, no. 12, pp. 685–700, Dec. 2019, doi: 10.1002/yea.3439.

[282] T. Wang, A. H. Badran, T. P. Huang, and D. R. Liu, "Continuous directed evolution of proteins with improved soluble expression," Nature Chemical Biology, vol. 14, no. 10, pp. 972–980, Oct. 2018, doi: 10.1038/s41589-018-0121-5.

[283] A. Zutz et al., "A dual-reporter system for investigating and optimizing protein translation and folding in E. coli," Nature Communications, vol. 12, no. 1, p. 6093, Dec. 2021, doi: 10.1038/s41467-021-26337-1.

[284] J. Jumper et al., "Highly accurate protein structure prediction with AlphaFold," Nature, vol. 596, no. 7873, pp. 583–589, Aug. 2021, doi: 10.1038/s41586-021-03819-2.

[285] K. E. Johansson, K. Lindorff-Larsen, and J. R. Winther, "Global analysis of multi-mutants to discover stabilizing amino acid substitutions," bioRxiv, Sep. 2021.

[286] A. Goldenzweig et al., "Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability," Molecular Cell, vol. 63, no. 2, pp. 337–346, Jul. 2016, doi: 10.1016/j.molcel.2016.06.012.

[287] J. M. Singer et al., "Large-scale design and refinement of stable proteins using sequence-only models," PLOS ONE, vol. 17, no. 3, p. e0265020, Mar. 2022, doi: 10.1371/journal.pone.0265020.

[288] H. Lu et al., "Machine learning-aided engineering of hydrolases for PET depolymerization," Nature, vol. 604, no. 7907, pp. 662–667, Apr. 2022, doi: 10.1038/s41586-022-04599-z.