Ana Sofia Araújo Ferreira

Development of novel strains for the production of biofuels

UMinho | 2018

Universidade do Minho
Escola de Engenharia

Ana Sofia Araújo Ferreira

Development of novel strains for the production of biofuels

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA
MIT Portugal
COMPETE
POPH
UNIÃO EUROPEIA
Fundo Social Europeu

dezembro de 2018

Universidade do Minho
Escola de Engenharia

Ana Sofia Araújo Ferreira

Development of novel strains for the production of biofuels

Tese de Doutoramento em Bioengenharia

Trabalho efetuado sob a orientação da
Professora Isabel Cristina de Almeida Pereira da Rocha
do
Professor Aljoscha Wahl
e do
Doutor Rui Miguel Pinheiro da Silva Pereira

dezembro de 2018

STATEMENT OF INTEGRITY

I hereby declare having conducted my thesis with integrity. I confirm that I have not used plagiarism or any form of falsification of results in the process of the thesis elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

University of Minho, December 17th 2018

Ana Sofia Araújo Ferreira

AGRADECIMENTOS

Durante estes 4 anos tive oportunidade de conhecer, interagir e aprender com pessoas formidáveis, sem as quais este doutoramento não seria possível ou, pelo menos, teria sido muito mais difícil de alcançar. Um doutoramento é um caminho, por vezes, solitário, mas tive a sorte de estar rodeada por pessoas maravilhosas.

Em primeiro lugar, agradeço à pessoa que me aconselhou e viu em mim potencial para trilhar este longo e tortuoso caminho que é um doutoramento, a minha orientadora Isabel Rocha. Durante estes 4 anos tive oportunidade de crescer muito a nível académico/científico, mas também pessoal e, só por isso, já lhe estaria eternamente grata. Obrigada por tudo o que me ensinou, por nunca me deixar desanimar, por dar sempre conselhos e sugestões, por ser exigente e por ver em nós mais do que aquilo que nós próprios conseguimos ver, fazendo-nos sair da tentadora, mas monótona, zona de conforto. Ser orientada por alguém assim é um privilégio e uma inspiração.

I would like to thank Aljoscha Wahl for receiving me in his research group, for being always available to discuss, for sharing his infinite knowledge on metabolic engineering with me and, of course, for lending me one of his precious bikes. I would also like to acknowledge all the CSE group and the good times I spent there. Those happy hours at the botanical garden were the best. Em especial aos meus compatriotas Alexandre, Joana, Francisca e Leonor, ajudaram a tornar a Holanda menos cinzenta.

Ao meu co-orientador Rui Pereira, pela partilha do teu enorme conhecimento em basicamente tudo o que constitui esta tese. Quero agradecer toda a disponibilidade, paciência e todos os ensinamentos. Tiveste sempre uma palavra de apoio e conselhos para dar mesmo a muitos Km de distância.

Ao prof. Eugénio como diretor do CEB e do grupo BioSystems por estar sempre disponível a dar uma ajuda na parte burocrática e ter sempre um conselho para dar.

À SilicoLife, em especial ao Vilaça, Maia, Simão e Camilo, por todo o apoio dado na análise de dados computacionais e na elaboração da patente.

Um agradecimento especial ao Liu por, primeiro ter gerado as vias em que se baseou este trabalho e, segundo, por toda a paciência a explicar-me o algoritmo e as soluções.

Ao Prof. José Maria Oliveira por ter facultado o GC-FID para as minhas análises e ao Eduardo Coelho por todo o apoio com este equipamento. À Eng. Madalena, Aline e Maura por toda a ajuda dada na parte analítica.

A todos os membros do BisBII (atuais e antigos) por me terem acolhido tão bem, sem dúvida que o momento do almoço e todas as conversas surreais que tivemos ajudavam a passar os dias mais difíceis. Um grupo tão multidisciplinar ajudou-me a aprender muito e as pessoas fantásticas foram a cereja no topo do bolo. Não vou referir "em especial" porque vocês são todos especiais, cada um à sua maneira.

A todo o laboratório de bioprocessos, em especial à minha hobbit Joana (que também participou neste trabalho) pelo bom ambiente de trabalho. Desculpem lá as cantorias desafinadas.

À gêmea Cristiana Faria (que se insere em demasiados grupos destes agradecimentos), obrigada por tudo o que me ensinaste e, principalmente, pela amizade e ajuda. Este doutoramento não seria o mesmo sem ti. À Lima pela boa-disposição e vídeos de gatinhos...e por me teres ajudado a fazer umas coisas no Lab.

À Cristiana Fernandes, por ser minha amiga há tanto tempo que já parei de contar. Vamos crescendo e mudando, mas a nossa amizade é intemporal. Obrigada em especial pelas duas visitas à Holanda.

À Cris, Mé, Paulo, Vânia e Izzie por serem o que de melhor levei do curso e continuarem sempre presentes, mesmo que as agendas sejam cada vez mais incompatíveis, os momentos passados juntos são sempre tempo de qualidade. À Sara por 'attackarmos' juntas os stresses do doutoramento no Solinca. Ao pessoal de civil por me fazerem sempre rir e, pelo caminho, alargarem os meus conhecimentos a pré-esforços. À Camila por ser sempre uma amiga tão atenciosa e preocupada.

Aos coleguinhas do MIT Portugal, aquelas semanas foram intensas, mas a vossa companhia tornou tudo mais fácil. Em especial à MJJ e à Salomé. Salomé, estamos destinadas a entrar em cursos/programas juntas e... ainda bem!

À minha família, em especial, aos meus pais, irmãos e tios Lina e Manel por, mesmo às vezes sem exatamente perceberem o que eu fazia com as minhas 'bichinhas', me apoiarem incondicionalmente. Desculpem as ausências, o (um pouco acentuado) mau-feitio e falta de paciência dos últimos tempos. Obrigada por tudo!

Às minhas pequeninas Ana e Sofia por, mesmo sem saberem, serem o melhor tónico para as semanas menos boas. Têm o dom de fazer-me parar de pensar no doutoramento por momentos.

Inevitavelmente, ao David, por todo o amor e apoio. Desculpa todos os planos de fim-de-semana que tenham incluído inóculos, tirar pontos e HPLC. O melhor ainda está para vir :)

Dedico esta tese de doutoramento à memória da minha avó Teresa.

ABSTRACT

Butanol is a four-carbon alcohol with a wide range of applications, which include: solvent for paints, dyes and resins; precursor of several polymers; and flavoring agent in food and beverage industry. Recently, the potential use of butanol as a biofuel has gained attention due to its superior properties when compared with the mostly used biofuel – ethanol. Butanol is natively produced by strains from genus *Clostridium* via Acetone-Butanol-Ethanol fermentation, a challenging industrial process. In this regard, more robust microorganisms have been tested as butanol producers. This thesis addresses the construction of novel strains of *Escherichia coli* to produce butanol.

Firstly, a novel route to produce butanol was selected from pathways previously generated using a hyper-graph algorithm. To do so, the pathways were ranked according to diverse criteria such as butanol yield, thermodynamic feasibility and number of reactions. Then, the most promising pathways were integrated into *i*JO1366, the most comprehensive *E. coli* genome-scale metabolic model available. The application of computational tools allowed to search for sets of genetic modifications that maximized product excretion.

In this novel pathway, 2-oxoglutarate is converted into butanol in seven catalytic steps by enzymes from different microbial sources. This pathway was validated *in vivo* by expressing nine heterologous genes in *E. coli* distributed in three plasmids. The developed strains were able to excrete butanol (maximum titer of 128.96±7.74 $mg.L^{-1}$), representing the first reported butanol production using 2-oxoglutarate as precursor.

Then, to implement an *in silico* suggested engineering strategy, the main mixed-acid fermentation pathways were inactivated to force the use butanol production as a NADH recycling route. Although butanol production could not be improved by removing the competing pathways, ethanol production was abolished, and lactate/acetate accumulation was reduced.

Lastly, strains of *E. coli* able to produce butanol through clostridial pathway were engineered by testing alternative enzymes to catalyze two rate-limiting steps. For the first step of this pathway, the native thiolase was replaced by the gene *phbA* from *Cupriavidus necator*. The clostridial gene *bcd-etfAB* and *ter* from *Treponoema denticola* were both tested. Moreover, the effect of using a single or an individual promoter to express the genes constituting the BCS (butyryl-CoA synthesis) operon was explored. The maximum butanol titer was 58.41±2.88 $mg.L^{-1}$. These constructions also allowed to conclude about the differences between the clostridial pathway and the newly designed pathway in

terms of nutrient requirements and performance in different strain backgrounds. The novel pathway, as opposed to the clostridial one, has a better performance in defined medium and in *E. coli* K strains.

RESUMO

O butanol é um álcool de quatro carbonos com diversas aplicações, tais como: solvente na produção de tintas, corantes e resinas; precursor de vários polímeros e agente aromatizante na indústria alimentar e de bebidas. Mais recentemente, o potencial do butanol como biocombustível destacou-se devido às suas características superiores quando comparado com o biocombustível mais usado: etanol. O butanol é produzido nativamente por várias estirpes do género *Clostridium* pela fermentação Acetona-Butanol-Etanol, um processo difícil de implementar industrialmente. Por esse motivo, outros microrganismos mais robustos têm sido testados como produtores de butanol. Esta tese visa a construção de novas estirpes de *Escherichia coli* para produzir butanol.

Inicialmente, uma nova via alternativa à de *Clostridium* foi selecionada de entre as diversas vias geradas por um algoritmo de híper-grafos. Nesse sentido, as vias foram classificadas de acordo com diversos critérios, tais como rendimento em butanol, viabilidade termodinâmica e número de reações. Posteriormente, a via considerada como a mais promissora foi integrada num modelo metabólico à escala genómica de *E. coli*, *i*JO1366, o que permitiu a aplicação de ferramentas computacionais para procurar conjuntos de modificações genéticas que maximizam a produção de butanol.

Nesta nova via, o 2-oxoglutarato é convertido em butanol através de sete reações catalisadas por enzimas de várias fontes microbianas. Esta via foi validada *in vivo* através da inserção de nove genes heterólogos em *E. coli*, distribuídos por três plasmídeos. As estirpes construídas foram capazes de produzir butanol (concentração máxima de 128.96±7.74 $mg.L^{-1}$), provando pela primeira vez que butanol pode ser produzido usando 2-oxoglutarato como precursor.

De seguida, as principais vias fermentativas de *E. coli* foram apagadas de forma a testar um genótipo ótimo obtido *in silico*. Apesar da produção de butanol não ter aumentado, observou-se a abolição da produção de etanol e uma diminuição na acumulação de acetato e lactato.

Por último, a via de *Clostridium* foi manipulada em dois passos identificados como limitantes. No primeiro passo, a tiolase foi substituída pelo gene *phbA* de *Cupriavidus necator*, enquanto que os genes *bcd-etfAB* e *ter* de *Cupriavidus necator* foram ambos testados. Adicionalmente, o efeito de expressar os genes do operão BCS (síntese de butiril-CoA) num promotor único ou individual foi explorado. A produção máxima de butanol obtida nas estirpes de *E. coli* resultantes foi 58.41±2.88 $mg.L^{-1}$. Estas construções também permitiram aferir as diferenças entre a via de *Clostridium* (e seus

derivados) e a nova via em termos de requerimentos nutricionais e desempenho atendendo à estirpe utilizada. A nova via, em oposição à de *Clostridium* mostrou um melhor desempenho em meio definido e usando a estipe de *E. coli* K.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

acacCoA	Acetoacetyl-Coa
acCoA	Acetyl-Coa
Acet	Acetate
Acka	Acetate Kinase
Acp	Acetyl-Phosphate
ALE	Adaptive Laboratory Evolution
AmpR	Ampicillin Resistance
ATP	Adenosine Triphosphate
BCD	Butyryl-CoA Dehydrogenase
BPCY	Biomass Product Coupled Yield
ButOH	Butanol
CA	*Clostridium acetobutylicum*
CB	*Candida boidinii*
CBM	Constraint-Based Modeling
Cit	Citrate
CoA	Coenzyme A
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSOM	Computational Strain Optimization Methods
DNA	Deoxyribonucleic Acid
EC	Enzyme Commission
EColi	*Escherichia coli*
EDTA	Ethylene Diamine Tetraacetic Acid
EtOH	Ethanol
FA	Fatty Acids
FAD	Flavin adenine dinucleotide
$FADH_2$	Dihydroflavine-adenine dinucleotide
FBA	Flux Balance Analysis
FID	Flame Ionization Detector
Form	Formate
FP	FindPath
FRE	Fermentation Regulatory Elements

FRT	Flippase Recombination Sites
Fum	Fumarate
FVA	Flux Variability Analysis
Fw	Forward ´
GC	Gas Chromatography
Gluc	Glucose
GSMM	Genome Scale Metabolic Model
HDM	High Density Medium
HPLC	High-Performance Liquid Chromatography
IPTG	Isopropyl β-1-Thiogalactopyranoside
Isocit	Isocitrate
KanR	Kanamycin Resistance
KEGG	Kyoto Encyclopedia Of Genes And Genomes
Km	Enzymatic Half-Saturation Constant
KO	Knock-Out
Lact	Lactate
LB	Luria-Bertani Rich Medium
Leu	Leucine
LL	*Lactococcus lactis*
LMOMA	Linear Minimization Of Metabolic Adjustment
Mal	Malate
MCS	Multiple Cloning Site
ME	Metabolic Engineering
MFA	Metabolic Flux Analysis
MILP	Mixed Integer Linear Programming
MiMBl	Minimization Of Metabolites Balance
MOMA	Minimization Of Metabolic Adjustment
mRNA	Messenger Ribonucleic Acid
NAD+	β-Nicotinamide Adenine Dinucleotide
NADH	β-Nicotinamide Adenine Dinucleotide, Reduced Form
NADP+	β-Nicotinamide Adenine Dinucleotide 2′-Phosphate

NADPH	β-Nicotinamide Adenine Dinucleotide 2′-Phosphate, Reduced Form
NCBI	National Center For Biotechnology Information
OAA	Oxaloacetate
OD_{600}	Optical Density at λ600nm
OXG	2-Oxoglutarate
PCR	Polymerase Chain Reaction
PDH	Pyruvate Dehydrogenase
PEP	Phosphoenolpyruvic Acid
PEP	Phosphoenolpyruvate
pFBA	Parsimonious Flux Balance Analysis
PFL	Pyruvate Formate Lyase
PHB	Polyhydroxybutyrate
Pta	Phosphotransacetylase
Pyr	Pyruvate
RBS	Ribosome Binding Site
Rev	Reverse
RI	Refractive Index
RNA	Ribonucleic Acid
rpm	Rotations Per Minute
SC	*Saccharomyces cerevisiae*
SmR	Spectinomycin Resistance
SSG	Solution Generation
Succ	Succinate
succCoA:	Succinyl-Coa;
TB	Terrific Broth
TCA	Tricarboxylic Acid
TcR	Tetracycline Resistance
TD	*Treponema denticola*
Val	Valine
YE	Yeast Extract

CHAPTER 1 Introduction

A growing concern towards reducing the Human ecological footprint and lowering anthropogenic CO_2 emissions is promoting a shift from chemical-based to biological-based processes. The construction of efficient cell factories for biological-based processes poses as one of the main challenges for metabolic engineering (ME). In this regard, the advent of curated reconstructions of metabolic networks combined with computational tools can support more rational ME strategies. Butanol, considered a promising gasoline substitute, is still mostly synthesized via chemical routes because this process is more cost-effective than the biological equivalent (Ndaba et al., 2015). The creation of cost-competitive biological processes for butanol production could open the door for the adoption of this compound as a biofuel. Thus, several ME-driven approaches have been focusing on the improvement of biobutanol production in order to reach industrial attractive titers and productivities. Besides butanol production via Acetone-Butanol-Ethanol fermentation by cultivating clostridial strains, alternative microorganisms have been tested as producing hosts. Among those, *Escherichia coli* achieved the greatest butanol titers.

The main goal of this thesis is to develop novel strains of *Escherichia coli* able to produce butanol. This thesis addresses the identification and selection of a novel pathway to produce butanol in *E. coli* and respective implementation *in vivo*, engineering of competing pathways to improve butanol production and the expression and engineering of clostridial pathway derivatives.

1.1 Context and motivation

The limited amount of natural resources and the consequent endeavor for more sustainable processes remains as one of the major challenges nowadays, stimulating the interest towards Industrial Biotechnology. In this endeavor, Metabolic Engineering (ME) can support the efficient shift from chemical-based processes to biological-based ones. Among several different definitions, ME targets the design of microbial cell factories by genetic and consequent metabolic manipulation with the goal of improving the production of a desired target compound.

The growing global demand for energy combined with environmental awareness has motivated the search for alternative fuels, produced from renewable raw materials. Lately, bioethanol – ethanol produced from biomass – has been reported as a promising gasoline substitute (H. Chan et al., 2013). Nevertheless, ethanol presents some non-ideal properties as fuel, such as lower energy density relative to gasoline and corrosiveness. Therefore, it is more suitable as gasoline's additive, due to its high octane number (Atsumi et al., 2008b).

Unlike ethanol, butanol is gaining increasing attention as a potential gasoline substitute because of its superior characteristics as a liquid fuel (Atsumi et al., 2008b; Patakova et al., 2013). Namely, butanol is less hydrophilic and, consequently, more compatible with the existing oil infrastructure; it is less corrosive; has better energy density (about 30 % more energy accumulated per volume); has lower volatility and, therefore, is less explosive (Atsumi et al., 2008a; Jang et al., 2012; Nielsen et al., 2009; Rabinovitch-Deere et al., 2013)). Butanol (also known as 1-Butanol, Butyl Alcohol, n-Butanol, n-Butyl Alcohol and n-Butan-1-ol) is a four-carbon alcohol with molecular formula C_4H_9OH and molecular weight of 74.12 $g.mol^{-1}$. This compound is a colorless liquid with a characteristic odor. There are three other isomeric forms of butanol, namely, *sec*-butanol or 2-butanol, isobutanol and *tert*-butanol (Dong et al., 2016; Ndaba et al., 2015). Butanol is a precursor for polymers, paints and plastics. As a solvent, it can also be utilized in the manufacture of hormones, vitamins and amino acids. In a report by Markets and Markets, the global market for butanol is forecasted to be worth € 7.3 thousand million by 2020 with a compound annual growth rate (CAGR) of 5.1 % (*N-butanol market: Global trends and forecasts to 2018 by applications (butyl acrylate, butyl acetate, glycol ethers, and others) and geography*, 2015).

Although naturally produced by clostridial strains through Acetone-Butanol-Ethanol (ABE) fermentation, most of this compound is still produced chemically via hydroformylation of propylene (Tudor and Ashley, 2007). In spite of the long-history of ABE fermentation, there are still some challenges regarding the industrial implementation of this process. Namely, slow growth of the host microorganism; biphasic metabolism; spore-forming life cycle; low yields and high recovery costs (Ndaba et al., 2015; Sauer, 2016).

Other microorganisms have been tested as hosts for butanol production. The first attempt to produce butanol in *Saccharomyces cerevisiae* achieved a very modest titer of 2.5 $mg.L^{-1}$ (Steen et al., 2008). More recently, a higher titer of 0.86 $g.L^{-1}$ was achieved by increasing acetyl-CoA and NADH levels (Schadeweg and Boles, 2016). Other microorganisms such as the cyanobacteria *Synechococcus elongatus* PCC 7942 (Lan and Liao, 2012) and the bacteria *Pseudomonas putida* and *Bacillus subtilis* (Nielsen et al., 2009) were also engineered to produce butanol. However, the highest titer (30 $g.L^{-1}$) of heterologous butanol production was obtained with *Escherichia coli*, which makes this microorganism the most suitable non-natural host for producing butanol (Shen et al., 2011). Since the first reported successful butanol production in *E. coli* (Atsumi et al., 2008a), several metabolic-engineering driven strategies have been applied to improve the productivity. Among several approaches, the expression of alternative pathways more suitable to the metabolism of *E. coli* has also been explored (Dellomonaco et al., 2011; Shen and Liao, 2008). In this regard, pathway discovering tools can be used to prospect for new pathways by retrieving metabolic information from databases. Further computational analysis can help identifying the most suitable pathways to be implemented in the selected host (Wang et al., 2017).

1.2 Research aims

The main goal of this thesis was to develop novel strains of *E. coli* able to produce butanol, according to the specific aims as follows:

- Evaluate the feasibility of several pathways, previously generated using the hyper-graph algorithm FindPath (Liu et al., 2015), to seek for the most promising ones to apply *in vivo*.

- Find an optimal genotype using computational tools based on Computational Strain Optimization Methods to search for sets of genetic modifications that maximize product excretion and reduce by-product formation.

- Select the most suitable enzyme for each of the steps composing the novel pathway considering curated gene sequences and the experimental data available in the literature.

- Implement *in vivo* the selected butanol pathway from the *in silico* analysis, constructing new strains of *E. coli* able to produce butanol through this novel pathway.

- Apply the optimum genotype provided by the *in silico* optimizations, in order to improve the butanol titers and reduce by-product formation.

- Design new *E. coli* strains expressing clostridial butanol pathway derivatives by testing alternative genes in the reported bottlenecks and expressing systems to infer the respective effect on butanol production. Compare the results with the ones obtained for strains expressing the novel butanol pathway selected from the *in silico* analysis.

1.3 Outline of the thesis

This thesis is structured in seven chapters. This first chapter includes a general introduction and the research aims. In Chapter 2, a review on the rational strategies to increase butanol production in *E. coli* is described in detail. From Chapter 3 to Chapter 6, the research aims described above are covered. General conclusions and recommendations for future work are stated in Chapter 7. The main content has thus the following structure:

- In Chapter 2, a comprehensive description of the different metabolic engineering strategies applied in *E. coli* to improve butanol production was gathered. The main goals were to evaluate the improvements described so far in the literature and pinpoint the remaining challenges. The progress in the field of metabolic modeling and pathway generation

algorithms and their application to butanol production in *E. coli* were also summarized in this chapter.

- In Chapter 3, pathways previously generated using the hyper-graph algorithm FindPath (Liu et al., 2015) were analyzed to infer their feasibility to be applied *in vivo*. First the solutions were evaluated according to diverse criteria (maximum butanol yield, number of reactions and novelty) to seek the most promising solutions. Lastly, using computational tools based on constraint-based modelling, the *in silico* butanol production in *E. coli* was optimized by searching for sets of genetic modifications that maximize product excretion though the proposed pathway.

- The *in vivo* implementation of the pathway considered the most promising one in Chapter 3 was proceeded in Chapter 4. To do so, strains expressing the enzymes constituting this pathway were designed and constructed. Different culture conditions were tested to find the best conditions for butanol production. Controlled cultures in bioreactors were also tested for their influence on butanol production.

- In Chapter 5, the main mixed-acid fermentation pathways were deleted to test the optimum *in silico* genotype obtained in Chapter 2. The optimum genotype included the knock-out of the main mixed-acid fermentation pathways, active during anaerobicity. The goal was to improve the butanol titers, abolish ethanol and lactate production and reduce acetate accumulation.

- Lastly, in Chapter 6, new strains of *E. coli* were designed to produce butanol by engineering the clostridial pathway, particularly intervening in two identified rate-limiting steps by testing alternative enzymes. The effect of different expression systems was also evaluated. These constructions were also used for qualitatively comparing the clostridial pathway with the newly designed pathway in terms of nutrient requirements and performance in different strain backgrounds.

This thesis is based on the following publications:

Patent

Rocha, I.; Rocha, M.; Ferreira, S.; Liu, F.; Pereira, R.; Vilaça, P.; Soares, S.; **'Method for n-butanol production using heterologous expression of anaerobic pathways'**, European Patent (application number EP18164811.4)

Papers

S. Ferreira, R. Pereira, A. Wahl, I. Rocha **'A review on the rational design strategies to produce butanol in *E. coli*'** (manuscript in preparation)

S. Ferreira, F. Liu, P. Vilaça, R. Pereira, I. Rocha **'Metabolic engineering of *Escherichia coli* for butanol production through a computationally generated pathway'** (manuscript in preparation)

S. Ferreira, R. Pereira, A. Wahl, I. Rocha **'Engineering the mixed-acid fermentation pathways in *Escherichia coli* to improve anaerobic butanol production'** (manuscript in preparation)

CHAPTER 2 Strategies to produce butanol in *Escherichia coli*: a review

The global market of butanol is increasing due to its applications as solvent, flavoring agent and chemical precursor of several other compounds. More recently, the superior properties of butanol as a fuel over ethanol, have stimulated even more interest. Butanol is natively produced together with ethanol and acetone by *Clostridium* species through Acetone-Butanol-Ethanol (ABE) fermentation, usually with low titers. The heterologous expression of the clostridial butanol biosynthetic pathway (and others) in a more robust microbial host figures as a promising alternative. Therefore, several groups have tried to express butanol production pathways in *Escherichia coli*, a well-known microorganism with several genetic modification tools available. In order to make *E. coli* more suitable as butanol producer, several efforts have been made to enhance butanol levels by modifying the native metabolic pathways and improving the heterologous pathways used.

This chapter contextualizes the history of ABE fermentation and gathers the main challenges of using this process. Moreover, all rational metabolic engineering strategies tested in *E. coli* to increase butanol titers are reviewed, such as manipulation of central carbon metabolism, elimination of competing pathways, cofactor balancing, development of new pathways, expression of novel enzymes, consumption of different substrates and molecular biology strategies. The progress in the field of metabolic modeling and pathway generation algorithms and their application to butanol production in *E. coli* are also summarized here. The main goals of this review are to gather all the strategies, evaluate the respective progress obtained, and identify and exploit the remaining problems.

2.1 Historical context

The history of butanol dates back to 1861, when Louis Pasteur firstly described the production of butanol by microorganisms, under anaerobic conditions (Pasteur, 1862). The French microbiologist detected butanol in a culture of, what he called, *Vibrion butyrique*, presumably a mixed-culture containing clostridia species (Dürre, 2008). In the end of the 19th century, most of the following works focused on the isolation and characterization of some clostridia species. These strains have the ability to convert carbohydrates into solvents, through the Acetone-Butanol-Ethanol (ABE) fermentation. In the early twentieth century, the industrial interest on ABE fermentation had emerged due to the application of butanol as a building block for producing synthetic rubber precursors. However, it was the interest in another by-product, acetone, which has motivated great interest in this bioprocess. During World War I (WWI), high amounts of acetone were needed to produce cordite, the propellant of cartridges and shells. Coincidently, Chaim Weizmann had isolated a strain of, later known as, *Clostridium acetobutylicum* able to produce acetone, butanol and ethanol in a ratio of 3:6:1 from starch and sugars. When compared with other strains available at the time, *C. acetobutylicum* was able to produce greater amounts of acetone and butanol, which led Weizmann to apply for a patent in 1915 (Weizmann, 1919). The industrial scale-up of this bioprocess, based on available plants for ethanol production, provided a constant supply of acetone in Great Britain, Canada and United States. After the war, most industrial ABE fermentation facilities were successively disassembled because of the lower demand for acetone (Buehler and Mesbah, 2016; Dürre, 2007; García et al., 2011; Ndaba et al., 2015).

During WW1, butanol was considered an unwanted by-product from the ABE fermentation process. Its potential as a bulk chemical gained attention with the rapidly growing need for quick-drying lacquers for the automobile industry (Sauer, 2016). Originally, butanol was produced using starch as substrate. However, molasses became the main carbon source as their price dropped and availability increased. The change in carbon source was only possible due to the earlier isolation of Clostridia strains capable of readily fermenting sugars. Some of the plants disassembled after WW1 were recovered to produce butanol and new ones were constructed. In fact, until 1950, around two thirds of the butanol produced globally were of biological origin. In the same decade, a new chemical process emerged for the production of butanol based on propylene oxo synthesis. In this process,

the aldehydes resulting from propylene hydroformylation are hydrogenated to produce butanol (Dürre, 2008; García et al., 2011; Green, 2011).

The increasing cost of molasses progressively led to the decline of ABE fermentation, which could no longer compete with the appellative crude oil prices, favoring chemical production over biological processes (Ndaba et al., 2015). Until the 1980s, some fermentation industrial facilities remained operational, but were successively closed. More recently, the instability of crude oil prices led to the reestablishment of some butanol producing plants in China and Brazil (Pfromm et al., 2010).

Nowadays, most of butanol is still synthesized via petrochemical routes. For this reason, synthetic butanol costs are linked to the propylene market and are sensitive to the price of the crude oil (Green, 2011).

Most of the attention given to butanol comes from its application as a fuel. In order to be competitive in the fuel market, the bioprocesses available for butanol production need to be engineered to decrease the cost of the final product. This chapter provides a thorough review of the most important progress made in the improvement of butanol production in the last decades. Given the slow progress on the improvement of native butanol producing organisms, this chapter mostly focuses on the use of *Escherichia coli* as an alternative host for butanol production and all the improvements that have been made in this front.

2.2 Biofuels: Biobutanol vs Bioethanol

Biofuels, broadly defined as fuels produced from organic matter, can be the answer to replace fuels obtained from depleting petroleum resources and to comply with the emission regulations implemented by a growing number of countries. The most common biofuels are biodiesel and bioalcohols (including bioethanol and biobutanol). In the first generation of processes used for biofuel production, agricultural crops such as corn, wheat and sugar cane were used as substrates, raising some ethical problems related to food scarcity. To solve these issues the second generation biofuels are produced using inedible plants or plants parts as the main substrate (He et al., 2015).

The most widespread biofuel is ethanol, which can be produced by well-known fermentation processes using a wide range of microorganisms, including the industrial workhorse *Saccharomyces cerevisiae*. Ethanol's higher octane number when compared with gasoline, makes it a very convenient fuel additive. Furthermore, the emissions of carbon dioxide are reduced by using ethanol instead of petrol. The main drawback in using bioethanol as a transportation fuel is its hygroscopicity (i.e. the ability to absorb or adsorb water particles from the surroundings) and corrosiveness. For this reason, ethanol is easily contaminated with water and the blending with gasoline has to take place right before the use. Thus, ethanol is incompatible with the actual storage and transportation infrastructures (Clomburg and Gonzalez, 2010; Luque et al., 2008; Pfromm et al., 2010).

Butanol is another promising candidate as biofuel, offering several advantages when compared with ethanol. Firstly, it is less hygroscopic and corrosive, so the blending with gasoline can take place in the biorefinery, allowing the use of current infrastructures. It is also less explosive and consequently safer to handle due to the lower vapor pressure. Moreover, butanol possesses two more carbon atoms per molecule than ethanol, accumulating ≈ 30 % more energy per molecule. The higher boiling point leads to longer burning in the motor engine. Finally, due to the closer stoichiometric air/fuel ratio with gasoline, butanol can be blended in higher percentages than ethanol (Clomburg and Gonzalez, 2010; Luque et al., 2008; Ndaba et al., 2015; Pfromm et al., 2010).

In Table 2.1, some characteristics of ethanol, butanol and gasoline are summarized.

Table 2.1: Properties of ethanol, butanol and gasoline. Adapted from (He et al., 2015; Ndaba et al., 2015).

Property	Ethanol	Butanol	Gasoline
Formula	C_2H_5OH	C_4H_9OH	C_4-C_{12}
Boiling point (°C)	78.4	117.2	40-220
Lower Heating value (MJ Kg^{-1})	26.78	33.1	44.0
Stoichiometric air/fuel ratio	9.0	11.2	14.7
Research octane number	129	96	94
Density at 16 °C ($Kg.L^{-1}$)	0.79	0.81	0.75

In the spite of the long-history of ABE fermentation, some problems regarding butanol production still exist, including low yield and expensive substrates. Thus, several efforts have been devoted to improving ABE fermentation and finding alternative biological ways to produce butanol.

2.3 Butanol production through ABE fermentation

The ABE fermentation occurs naturally in microorganisms from *Clostridium* genus. *Clostridium* spp are rod-shaped, gram-negative, strict anaerobes, spore-forming bacteria (Ndaba et al., 2015).

Clostridial strains differ in their ability to ferment various substrates, the patterns of solvent ratios and some diversity of chemicals produced (Patakova et al., 2013). Particularly, *C. acetobutylicum* and *Clostridium beijerinckii* are the best studied and applied for industrial production of butanol. Usually, less than 13 $g.L^{-1}$ of butanol are produced during batch fermentation since butanol is highly toxic to the cells (Visioli et al., 2014).

The ABE fermentation is a biphasic process, comprising an acidogenic and a solvent phase. During the acidogenesis, the cells ferment sugar or starch into butyrate, acetate, carbon dioxide and hydrogen, which lowers the pH of the medium. In the end of the exponential phase a metabolic shift takes place in clostridial strains and the solventogenic phase starts: the previously excreted acids are re-assimilated and converted in neutral solvents such as acetone, butanol and ethanol. So, the solventogenic phase is a defense mechanism to prevent cell death due to low values of pH, increasing both external and internal pH, and allowing the cells to remain metabolically active for a longer time

(Dürre, 2008). Nevertheless, the solvents (mainly butanol) also damage the cells by inactivating membrane proteins and destroying the membrane. In parallel to the solventogenic phase, cells start to form endospores, which guarantee a long-term survival. Thus, the solventogenic and sporulation regulatory networks are correlated. Despite the long-history using ABE fermentation, some challenges still remain nowadays, hindering its industrial application. In Figure 2.1, a schematic representation of ABE fermentation is depicted with indication of the enzymes involved in butanol production (García et al., 2011; Green, 2011; Patakova et al., 2013).

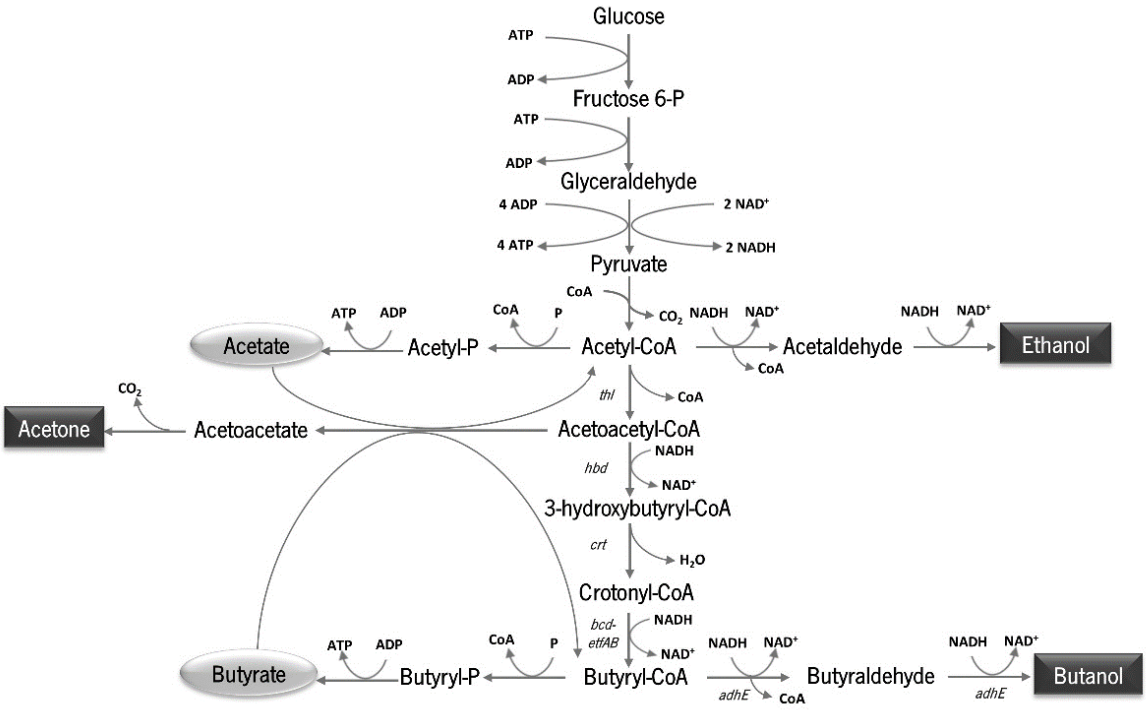


Figure 2.1: Schematic representation of ABE fermentation. Light grey circles indicate the products released during acidogenic phase and in dark grey boxes are the solvents excreted in solventogenesis. The enzymes involved in butanol pathway are indicated.

The first-generation of butanol bioprocesses was mostly based on substrates obtained from starch-rich crops, including wheat, rice and maize, yielding significant titers. Nevertheless, non-edible residues or plant parts (second generation fuels) are a much cheaper and more available feedstock (Visioli et al., 2014).

The economics of an industrial process are highly dependent on the substrate costs. Thus, the conversion of cheap lignocellulosic substrates into butanol is important to achieve an economically feasible process. These residues must be exposed to pretreatments in order to release the fermentable sugars like glucose and xylose. Compared with other microorganisms used to produce

ethanol, clostridial strains are more sensitive to the inhibitors resultant from the pretreatment processes. So, the pretreatment is an active topic of research in an effort to find an efficient process that allows butanol production from cheap raw materials (Pfromm et al., 2010; Visioli et al., 2014).

The productivity of butanol in ABE fermentation is low due to the long fermentation times (García et al., 2011). Volumetric solvent productivity has a huge impact on global process costs. Continuous cultures help to increase the productivity over batch cultures. However, the implementation of continuous cultures using clostridial strains is difficult due to the loss of the solvent production phenotype over time and the two-stage fermentation. Semi-continuous reactors and immobilized systems have been tested to improve productivity showing higher productivities when compared with batch cultures (Green, 2011; Visioli et al., 2014).

The low butanol yields due to the production of byproducts and the low volumetric solvent productivity lead to high costs in the recovery process. For that reason, there is significant research on ABE fermentation focusing in the design of an efficient downstream process. For the recovery of butanol from the fermentation broth, researchers suggest techniques such as adsorption, liquid–liquid extraction, gas stripping and pervaporation. The *in situ* application of some of these processes can smooth the butanol toxicity issues (Dürre, 2007; Visioli et al., 2014).

In order to tackle some of the problems identified above, metabolic engineering strategies have been applied to develop more robust clostridial strains. The main goals are: to increase butanol production over other solvents; to increment butanol titer and purity; to accelerate the fermentation process by omitting the acidogenic phase; to avoid the coupling between acidogenic and sporulation phase; to express other enzymes to extend the portfolio of usable substrates; or to increase butanol tolerance (García et al., 2011).

Many paths of improvement are currently being followed and a review of the main findings has been published by Sauer (Sauer, 2016). For instance, an improved strain of *C. acetobutylicum* was obtained using chemical mutagenesis, which can excrete acetone-butanol-ethanol in a 2:7:1 ratio. Another strain of *C. acetobutylicum* was engineered by inactivating the *solR* regulatory gene and overexpressing the gene encoding aldehyde/alcohol dehydrogenase, achieving a butanol titer of 17.6 $g.L^{-1}$ (Harris et al., 2001) .

The advent of new genome editing strategies, in particular of the CRISPR-Cas9 system, has facilitated strain engineering for several microorganisms and, simultaneously, enlarged the range of

applications (Tian et al., 2017). So far, this genome editing tool was efficiently developed for different strains from the genus *Clostridium*, such as *Clostridium saccharoperbutylacetonicum* (Zhang et al., 2017) and for *C. acetobutylicum* (Wasels et al., 2017). Nevertheless, the complex regulatory and metabolic mechanisms of clostridial strains hinder the development of fully efficient cell factories. The production of butanol in other well-known industrial hosts poses as a promising alternative (Clomburg and Gonzalez, 2010).

2.4 Recombinant butanol production in *Escherichia coli*

E. coli is a gram-negative bacterium, facultative anaerobe, well-characterized and with efficient genetic tools available. The large knowledge about its genetic, metabolic and physiological characteristics enables its engineering for the production of diverse target compounds. Moreover, *E. coli* possesses several industrially relevant characteristics like (i) capacity to grow on mineral media, (ii) conversion of a variety of different substrates including biomass components such as fatty acids, carbohydrates and polyols (iii) efficient growth (iv) ability to grow under aerobic and anaerobic conditions (v) robustness under industrial conditions (vi) tolerance to high concentrations of substrates and products (Clomburg and Gonzalez, 2010; Koppolu and Vasigala, 2016).

E. coli has been successfully engineered for the production of diverse products, including hormones (Rezaei and Zarkesh-Esfahani, 2012), proteins (Reyes et al., 2017) and amino acids (Lee et al., 2007; Park et al., 2007). Its potential to produce biofuels was first explored with ethanol, a native product (Ohta et al., 1991).

In 2008, Atsumi *et al*. described for the first time the production of butanol in *E. coli* (Atsumi et al., 2008a). In this work, the genes constituting the clostridial butanol biosynthetic pathway (*thl*, *crt*, *hbd*, *bcd-etfAB*, *adhE*) were expressed in *E. coli*. Alternative genes encoding enzymes catalyzing the first step (*atoB*) and the fourth step (*ccr*) of the pathway were also tested. The replacement of *thl* for *atoB* increased butanol production, while the expression of *ccr* led to lower butanol titers. The maximum butanol achieved was 552 $mg.L^{-1}$, obtained by cultivating cells semi-aerobically in Terrific Broth (TB) medium supplemented with 2 % glycerol. Considering that *Clostridium* strains are able to produce butanol titers up to 20 $g\text{-}L^{-1}$, the recombinant butanol production in *E. coli* needed further optimization to compete with the native producers.

2.5 Rational design strategies to improve recombinant butanol production in *Escherichia coli*

Since the first reported recombinant butanol production in *E. coli*, several authors have tried to increase butanol titers. To do so, various rational strategies have been designed and implemented. In Figure 2.2, the main attempts to improve butanol production in *E. coli* are summarized.

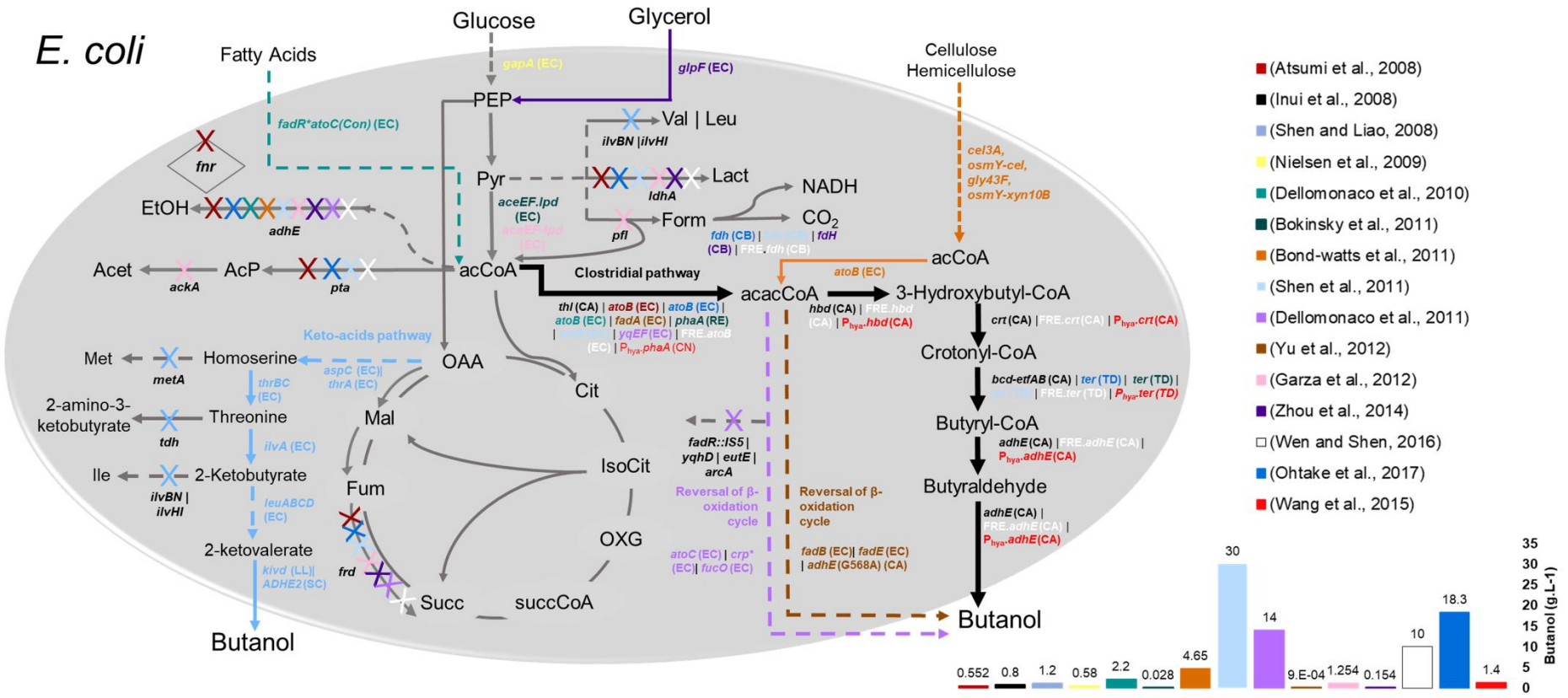


Figure 2.2: Schematic overview of the different pathways expressed into E. coli *to produce butanol and other genetic modifications. These pathways correspond to the ones described in literature in which maximum butanol production was obtained. The n-butanol synthetic pathway from* Clostridium acetobutylicum *is represented with black arrows. Alternatives described in literature to this pathway including gene knock-out, overexpression, different substrates are highlighted in the respective color. Dashed lines represent successive enzymatic reactions; X indicates gene knock-outs; diamonds correspond to transcriptional regulators. acCoA- Acetyl-CoA; acacCoA – acetoacetyl-CoA; PEP – phosphoenolpyruvic acid; OXG – 2-oxoglutarate; Cit – Citrate; Isocit – Isocitrate; succCoA: Succinyl-CoA; Succ – Succinate; Fum – Fumarate; Mal – Malate; OAA – oxaloacetate; Acp – acetyl-Phosphate; Acet – Acetate; EtOH – Ethanol Lact – Lactate; Form – Formate; Pyr – pyruvate; Val – Valine; Leu -Leucine; EColi –* Escherichia coli*; CA –* Clostridium acetobutylicum*; SC –* Saccharomyces cerevisiae*; TD –* Treponema denticola*; CB -* Candida boidinii*; LL –* Lactococcus lactis*; FRE – fermentation regulatory elements*

By analyzing Figure 2.2, it is possible to perceive a variety of strategies to produce butanol in *E. coli*, including: manipulation of the central carbon metabolism either by gene disruption or gene overexpression; using different substrates; testing different enzymes for specific steps of the clostridial pathway or expression of new biosynthetic pathways. Most of the strain designs here focused on the expression of clostridial pathway and respective derivatives. Nevertheless, two new routes were also explored: the reversed β-oxidation cycle and the ketoacids pathway.

Within the strain designs depicted in Figure 2.2 that provided the highest butanol titers, a variety of titers was obtained ranging from 0.9 $mg.L^{-1}$ to 30 $g.L^{-1}$. We also can see that the first and fourth step of clostridial pathway are the ones where more alternative enzymes have been tested. Regarding gene knock-outs, the genes responsible for the production of ethanol (*adhE*), succinate (*frdABCD*), acetate *(pta-ackA*) and lactate (*ldhA*) are the most common targets. The gene responsible for the production of ethanol, in particular, was knocked-out in nine out of the fourteen strategies shown in Figure 2.2.

The titers obtained also depend on the strain of *E. coli* used as host, culture medium used and cultivation conditions. For this reason, the details for each strategy depicted in Figure 2.2 are shown in Table 2.2, including the culture conditions and host strains used.

Table 2.2: Major features of the engineered strains of E. coli *to produce butanol and respective culture conditions.*

Strain	Genes expressed	KO	Culture conditions	Titer ($g.L^{-1}$)	Ref.
E. coli BW25113	*atoB, hbd, crt, bcd, etfA, etfB, adhE2*	*adhE, ldhA, frdBC,pta, fnr*	TB medium + 2 % glycerol. Culture in sealed 12 mL glass tube in semi-aerobically conditions- without the headspace evacuated. 37 °C for 24 h	0.552	(Atsumi et al., 2008a)
E. coli JM109	*thl, hbd, crt, bcd, etfA, etfB, adhE2*	-	M9 medium, initial OD660=20, shaken in a 100 mL bottle in an anaerobic chamber (N_2 95%+ H_2 5%) at 30 °C for 60 h. pH was kept above 6.5	1.2	(Inui et al., 2008)
E. coli BW25113	*thrA*rBC, ilvA, leuABCD, kilvd, adhE2*	*metA, tdh, ilvI, ilvB, adhE*	M9 medium with 5 $g.L^{-1}$ yeast extract and 8 $g.L^{-1}$ threonine, shaken in a 250-mL screw capped conical flask for 4 days at 30 °C	0.8	(Shen and Liao, 2008)
E. coli BL21 star (DE3)	*atoB, hbd, crt, bcd, etfA, etfB, adhE2, gapA*	-	TB medium + 5 $g.L^{-1}$ glycerol, shaken in 250-mL screw-capped flask for 48 h with initial OD_{600} = 0.05 at 37 °C	0.58	(Nielsen et al., 2009)
E. coli MG1655	*atoB, hbd, crt, bcd, etfA, etfB, adhE2, atoC(c)*	*adhE, fadR:: IS5*	MOPS minimal medium + 5 $g.L^{-1}$ palmitic acid+2 $g.L^{-1}$ Brij 58. Supplementation of 5 $g.L^{-1}$ palmitic acid every 24 h. Initial OD550 = 10, shaken in 50 mL baffled shake flask at 37 °C for 72 h	2.05	(Dellomonaco et al., 2010)

E. coli DH1	*atoB, hbd, crt, bcd, etfA, etfB, adhE2, cel3A, osmY-cel, gly43F, osmY-xyn10B*	*adhE*	EZ-rich medium (Teknova) + 3.3 % w/v IL-treated switchgrass Parafilm sealed to create microaerobic environment; 30 °C for 96 h	0.028	(Bokinsky et al., 2011)
E. coli DH1	*phaA, hbd, crt, ter, adhE2, aceEF, lpd*	-	TB medium + 1.5 % (w/v) glucose, shaken in a sealed 250-mL baffled flask selead with parafilm to prevent butanol evaporation for 3 days at 30 °C. Additional glucose (1 % (w/v)) was added after 1 day	4.65	(Bond-watts et al., 2011)
E. coli BW25113	*atoB, hbd, crt, ter, adhE2, fdh*	*ldhA adhE frdBC pta*	Cultivations using 0.35 L working volume in 0.5 L fed-batch bioreactor with TB+2 % glucose medium. Intermittent linear feeding of glucose solution to maintain a glucose concentration between 10 and 20 $g.L^{-1}$ with continuous removal of butanol; 7 days at 37 °C, pH 7 adjusted with NaOH	30	(Shen et al., 2011)
E. coli MG1655	*atoC (c), crp*, yqeF, fucO*	*fadR::IS5 yqhD, eutE, arcA, adhE, frdA, pta*	Minimal medium supplemented with 20 $g.L^{-1}$ glucose and 1 mM betaine in a SixFors multifermentation system (Infors HT) Initial OD550 of 0.05. Dissolved oxygen controlled at 5 % of saturation and pH at 7. 30 °C for 24 h	14	(Dellomonaco et al., 2011)
E. coli MG1655 $lacI^{Q}$	*atoC (c), fadB, fadE, yqhD, adhE (G568A)*	*arcA*	M9 medium+20 $g.L^{-1}$ glucose+10 g.L glycerol, shaken in 20x200-mm test tubes with ventilation plugs at 37 °C for 24 h	$9x10^{-4}$	(Yu et al., 2012)
E. coli ATCC11303	*thl, hbd, crt, bcd, etfA, etfB, adhE2, aceEF-lpd*	*adhE, ackA, frdABCD, pflB, ldhA*	LB + 50 $g.L^{-1}$ glucose, initial OD550=32, sealed serum bottle with stirring in an anaerobic jar at 30 °C for 60 h	1.254	(Garza et al., 2012)
E. coli MG1655 (DE3)	*thl, hbd, crt, bcd, etfA, etfB, adhE2, GlpF, fdh*	*adhE, ldhA, frdBC*	TB medium + 0.5 % (w/v) glycerol, initial OD_{600}=1.5. Cultures were in 50-mL sealed flask with evacuation of the headspace to create anaerobic conditions. Culture media was shaken for 56 h at 37 °C	0.154	(Zhou et al., 2014)
E. coli BW25113	*Pya:phaA, Pya:hbd, Pya:crt, Pya:ter Pya:adhE1*	*adhE, ldhA*	TB medium + 0.5 % glucose. Two stage fermentation in bioreactors. Aerobic growth was maintained with dissolved oxygen above 25 %, cultures were shifted to anaerobic conditions when OD_{600}=5, pH adjusted to 5.5 with 1M NaOH, temperature maintained at 37 °C	1.4	(Wang et al., 2015)
E. coli BW25113	FRE_{ackA}*::atoB-adhE2-crt-hbd,* FRE_{adhE}*::ter,* FRE_{adhE}*:: fdh*	*adhE, ldhA, frdBC, pta*	TB medium+60 $g.L^{-1}$ glucose. Switch to anaerobic conditions when OD_{600}=4. Linear feeding of 10 $g.L^{-1}$ glucose and pH adjusted to 7. Cells were grown for 24 h at 37 °C	10	(Wen and Shen, 2016)
E. coli BW25113	*atoB, hbd, crt, ter, adhE2, fdh*	*frdBC, ldhA, adhE*	TB medium + 2 % glucose supplemented with 2 mM cysteine, Initial OD_{600}=0.04. Cells were cultured in sealed tubes with headspace evacuated and incubated at 37 °C for 78 h at 200 rpm pH was adjusted to 7 using 10 M NaOH	18.3	(Ohtake et al., 2017)

In Table 2.2, a wide variety of strategies are described. For instance, the two first published works, (Atsumi et al., 2008a) and (Inui et al., 2008), reporting recombinant butanol production in *E. coli* achieved different titers: 552 $mg.L^{-1}$ and 1.2 $g.L^{-1}$, respectively. The value achieved by Atsumi and coworkers was obtained after disrupting the NADH-competing pathways and replacing *thl* by *atoB*, while in the work developed by Inui *et al.*, the greater butanol titer was achieved by simply expressing the clostridial butanol biosynthetic genes. So, several factors can impact butanol titer including host strain, expression system and culture conditions. Overall, within the published works, the medium most used is the complex medium Terrific Broth (TB) or mineral media supplemented with complex nutrients like yeast extract. The greater butanol titer (30 $g.L^{-1}$) was achieved by *in situ* continuous removing butanol and intermittent linear feeding of glucose. This approach to remove butanol led to an increment from 15 $g.L^{-1}$ to 30 $g.L^{-1}$.

As already mentioned, the strategies shown in Table 2.2 correspond to the optimal butanol titers obtained within each publication. Nevertheless, sub-optimal strategies were also tested within the references shown and for this reason, we will explore each one in more detail in the following sections.

2.5.1 Elimination of competing pathways

The internal metabolic fluxes in a microorganism evolved to fulfill its own requirements, which usually means that genetic interventions are required to change the fluxes to satisfy industrial goals (Burgard et al., 2003; Maia et al., 2016) The elimination of competing pathways by gene disruption is a common strategy when overproducing a certain biochemical compound. In *E. coli*, this approach was effectively applied in the production of a wide variety of compounds such as succinate (Jantama et al., 2008), ethanol (Kim et al., 2007) and L-alanine (Zhang et al., 2007).

In *E. coli* growing on glucose, under anaerobic conditions, the Tri-Carboxylic Acid (TCA) cycle is downregulated, leading to incomplete oxidation of the substrate and excretion of fermentation products. Besides the two pyruvate molecules and two ATPs produced in glycolysis per molecule of glucose, without oxygen available, two NADH molecules per glucose need to be recycled to NAD^+ using mixed-fermentation pathways. In anaerobiosis, *E. coli* metabolizes pyruvate using the pyruvate

formate lyase, which prevents the release of additional NADH, forming formate and acetyl-CoA instead. Acetyl-CoA can then be converted into acetate by the action of the enzymes encoded by *ackA-pta*, forming ATP or into ethanol by action of *adhE*, recycling four molecules of NADH per glucose. Lactate synthesis from pyruvate also allows to recycle NADH (two per glucose). In addition to these three routes, *E. coli* can also produce succinate through the reduced branch of the TCA cycle, while recycling four molecules of NADH per glucose. In sum, in anaerobiosis *E. coli* excretes a mixture of ethanol, acetate, lactate and succinate, the so-called mixed-acid fermentation products.

The redistribution of the metabolic flux from the production of native fermentation products towards butanol production, usually involves the elimination of competing NADH-recycling pathways such as ethanol, lactate and succinate. The goal of disrupting the genes involved in the production of acetate is to increase the acetyl-CoA pool, the main precursor for butanol synthesis in the clostridial pathway (Ohtake et al., 2017; Shen et al., 2011).

The disruption of *adhE* is the most common within the strategies represented in Figure 2.2. The elimination of *adhE* allows simultaneously to avoid acetyl-CoA consumption and the recycling of two molecules of NADH. Nevertheless, other enzymes with alcohol/aldehyde dehydrogenase activities such as *mhpF*, *adhP* and *yqhD* can still lead to ethanol accumulation (Yu et al., 2012).

In the two novel pathways (ketoacids and the reversal β-oxidation cycle pathways), other genes were disrupted, mostly to redirect the flux towards butanol (Dellomonaco et al., 2011; Shen and Liao, 2008).

In the work developed by Atsumi *et al.*, the disruption of several genes (*adhE*, *ldhA*, *frdBC*, *fnr* and *pta*) increased by 2.6-fold the butanol titer and reduced the accumulation of acetate, ethanol and lactate (Atsumi et al., 2008a). Liao group developed the *E. coli* strain able to achieve the maximum butanol titer so far (30 $g.L^{-1}$) and also observed a 4-fold improvement on butanol accumulation after knocking-out the genes *adhE*, *ldhA*, *frdBC* an *pta*. In this study, an enzyme converting formate into CO_2 and NADH was also expressed to provide extra NADH (Shen et al., 2011).

2.5.2 Cofactor balancing

In addition to the activity of heterologous enzymes, the metabolic flux towards the target product also depends on the concentration of precursors and cofactors. When developing an efficient cell factory

for the production of highly reduced compounds (such as alcohols), the cofactor balance is one of the major aspects to be considered. For example, producing one molecule of butanol from acetyl-CoA requires four molecules of NADH when the clostridial pathway is used. Supporting the importance of cofactor availability, it has been shown that increasing the intracellular NADH/NAD^+ ratio results in higher accumulation of fermentation by-products (de Graef et al., 1999; Zhao et al., 2017).

Therefore, given the lack of oxygen as an electron sink, it is easy to understand that anaerobic conditions favor the production of reduced compounds. During fermentation, the molecules of NADH formed in glycolysis are recycled to NAD^+ by forming reduced compounds, such as alcohols. These NADH recycling mechanisms support the maintenance of the redox balance inside the cells, a mandatory requirement for living cells sustain their growth. On the other hand, under aerobic conditions, the regeneration of NADH by oxidation is more advantageous, generating ATP. So, in order to couple cell growth with butanol synthesis, cultivations are usually oxygen-limited (Trinh et al., 2011). Several approaches were tested to increase the NADH pool under anaerobic conditions to increase butanol production.

Pyruvate can be converted into acetyl-CoA by two enzymes: Pyruvate Formate Lyase (PFL) and Pyruvate Dehydrogenase (PDH) complex. PFL is active under anaerobic conditions and catalyzes the conversion of pyruvate into acetyl-CoA and formate. On the other hand, the PDH complex releases two molecules of NADH when pyruvate is converted into acetyl-CoA and CO_2. So, if the PDH is activated in anaerobic conditions it can provide extra NADH that needs to be recycled. This approach was successfully applied to the production of butanol, achieving a 1.6-fold improvement (Bond-watts et al., 2011). Another study only achieved a 1.1-fold increment on butanol production, but a two-times higher yield on substrate (Garza et al., 2012).

The PDH complex is constituted by three subunit enzymes: pyruvate decarboxylase (encoded by *aceE*), dihydrolipoamide acetyltransferase (*aceF*), and dihydrolipoamide dehydrogenase (encoded by *lpd*). This last enzyme is inhibited when exposed to high concentrations of NADH (de Graef et al., 1999), which can explain the modest improvements on butanol observed when overexpressing PDH complex under anaerobic conditions (Kim et al., 2008; Lim et al., 2013) Another strategy followed by the Liao group consisted in inactivating the regulator *fnr* (which represses the expression of PDH complex under anaerobic conditions). By only deleting *fnr* actually resulted in decreased butanol

production. Only by coupling the previous strategy to the deletion of *pta* (involved in the production of acetate from acetyl-CoA), led to a 3-fold improvement on butanol production (Atsumi et al., 2008a).

Shen et al., developed a metabolic engineering strategy using CoA and NADH as driving forces. In particular, the NADH driving force was established by first deleting the mixed acid fermentation reactions (Δ*adhE* Δ*ldhA* Δ*frdBC*). Also, in order to couple butanol pathway with NADH driving force, the complex *bcd-etfAB*, which uses ferredoxin as reducing power, was replaced by the NADH-dependent *ter* from *Treponema denticola.* In *E. coli*, the native formate dehydrogenase catalyzes the conversion of formate into CO_2 and H_2. However, in other microorganisms, the same enzyme is able to hydrolyze formate into CO_2 and NADH. Shen *et al.*, in order to further increase the NADH pool, expressed the formate dehydrogenase from *Candida boidinii*, which increased butanol concentration when the medium was supplemented with formate (Shen et al., 2011). Nielsen *et al.* expressed *fdh* from *Saccharomyces cerevisiae*, which resulted in an improvement of 74 % on butanol concentration. The supplementation of formate led to even higher concentrations (Nielsen et al., 2009).

2.5.3 Expression of alternative enzymes

The clostridial butanol biosynthetic pathway is constituted by six catalytic steps converting two molecules of acetyl-CoA into one of butanol (Figure 2.1). The seven genes (*thl*, *hbd*, *crt*, *bcd-etfAB*, *adhE*) constituting this pathway are sufficient to support butanol production in *E. coli* (Inui et al., 2008). Nonetheless, the determination of rate-limiting steps and the expression of alternative genes more suitable to the host can lead to a more efficient cell factory.

2.5.3.1 *Thiolase*

The first step of clostridial butanol biosynthetic pathway is the thermodynamically unfavorable ($\Delta_r G'^m$ = 26.1 ± 1.7 kJ.mol^{-1}) condensation of two molecules of acetyl-CoA into one of acetoacetyl-CoA by the action of a thiolase (encoded by *thl*). Although the thermodynamics of a reaction are not dependent on the enzyme used, the rate of a reaction can be improved by an increment in enzymatic activity. For this reason, enzyme homologs from non-clostridial sources have been tested to catalyze

this step. *E. coli* expresses three enzymes with acetyl-CoA acetyltransferase activity: *atoB*, *fadA* and *yqeF*. AtoB has a higher specific activity when compared with clostridial thiolase *thl*: (AtoB, 1,078 $U.mg^{-1}$ against Thl, 216 $U.mg^{-1}$). As observed in Figure 2.2, in most of the butanol production studies in *E. coli*, *thl* was replaced by *atoB* from *E. coli*. The overexpression of this native enzyme was tested in the first reported recombinant production of butanol in *E. coli* (Atsumi et al., 2008a) resulting in a 3-fold improvement on butanol titer from 0.014 $g.L^{-1}$ to 0.040 $g.L^{-1}$. Nevertheless, in another study the increment on butanol titer by replacing *thl* for *atoB* was not so significant: from 0.20 to 0.22 $g.L^{-1}$ (Nielsen et al., 2009). This difference highlights that the activity of an enzyme is also dependent on other factors like the expression system or the host strain used. Another thiolase, *fadA,* was tested in the work where the inverted β-oxidation cycle was explored for butanol production (Yu et al., 2012). However, the final butanol titer obtained with this alternative pathway was very modest (less than 1 $mg.L^{-1}$). On the other hand, Dellomonaco *et al.* also took advantage of the native β-oxidation pathway to produce butanol in *E. coli*, but overexpressing another thiolase (*yqeF*). The results were more promising in this case and the modified strains could achieve 2.2 $g.L^{-1}$ of butanol in shake-flask and around 14 $g.L^{-1}$ in bioreactor (Dellomonaco et al., 2011).

Bond-Watts and his coworkers, inspired by the efficient production of PolyHydroxyButyrate (PHB) in *Cupriavidus necator*, have transferred part of the respective pathway (first two steps) from *C. necator* to *E. coli* to produce butanol. The single overexpression of *phaA* gene led to the greater butanol accumulation of 4.65 $g.L^{-1}$ (Bond-watts et al., 2011).

2.5.3.2 *3-Hydroxybutyryl-CoA Dehydrogenase and 3-Hydroxybutyryl-CoA Dehydratase*

In most of the published works, the protein product of the genes *hbd* and *crt* from *Clostridium acetobutylicum* are used to catalyze the reduction of acetoacetyl-CoA into 3-hydroxybutylryl-CoA and subsequent dehydration in crotonyl-CoA (Dong et al., 2016).

Nevertheless, the genes *phaB* and *phaJ* from PHB pathway could be used to replace these two clostridial enzymes, as long as they are used together (Dong et al., 2016). In the inverted aerobic fatty acid β-oxidation pathway, a single enzyme was expressed to catalyze these two steps, the protein product of the gene *fadB* from *E. coli* (Dellomonaco et al., 2010).

2.5.3.3 *Butyryl-dehydrogenase*

The reduction of crotonyl-CoA into butyryl-CoA is catalyzed in clostridial pathway by butyryl-CoA dehydrogenase (Bcd), which requires the presence of the electron-transferring flavoprotein complex (EtfAB). The expression of this enzyme in *E. coli* is challenging due to its oxygen sensitivity and the requirement of ferredoxin as the electron donor (Dong et al., 2016; Shen et al., 2011).

The Liao group have replaced *bcd-etfAB* by *ccr* from *Streptomyces coelicolor*, but lower titers of butanol were achieved (Atsumi et al., 2008a). In another study (Bond-watts et al., 2011) the authors further explored the expression of *ccr* as part of a synthetic butanol pathway, concluding that this enzyme favors ethylmalonyl-CoA formation (65 %) over butyryl-CoA (35 %), providing a route for carbon to exit the butanol pathway.

The difficulty of functionally expressing *bcd-etfAB* complex in *E. coli* was overcome by expressing another class of enzymes with the same activity: *trans*-enoyl-reductases. The reduction of crotonyl-CoA into butyryl-CoA mediated by *ter* is an irreversible reaction in contrast with the reversible reaction catalyzed by flavin-dependent Bcd-EtfAB complex. The replacement of *bcd-etfAB* by *ter* effectively increased productivity of n-butanol in *E. coli* from 150-200 $mg.L^{-1}$ to 2,950 $mg.L^{-1}$ (Bond-watts et al., 2011). Shen and coworkers have also tested the effect of expressing *ter* from different sources on butanol titer, namely from *Treponema denticola, Treponema vincentii*, *Fibrobacter succinogenes* and *Flavobacterium johnsoniae* (Shen et al., 2011). Cells expressing *ter* from *T. denticola* achieved the greater butanol titer. They also subjected the three Ter homologues to error-prone PCR mutagenesis. The *ter* mutant (Met11Lys) from *F. succinogenes* provided similar performances in terms of butanol titer when compared with cells expressing *ter* from *Treponoema denticola*. In Figure 2.2, it is possible to see that most of the published works use *ter* to catalyze this step.

2.5.3.4 *Aldehyde/Alcohol Dehydrogenase*

The two last steps of the clostridial pathway to produce butanol are the two successive reductions of butyryl-CoA into butyraldehyde and to butanol, recycling two molecules of NADH. In clostridial strains,

these two steps are catalyzed by the same enzyme, the bifunctional aldehyde/alcohol dehydrogenase. Clostridial strains express two aldehyde/alcohol dehydrogenases, one during acidogenesis (*adhE*1) and the other in the solventogenic phase (*adhE*2) (Dong et al., 2016; Nielsen et al., 2009).

Inui and coworkers have studied the effect of expressing these two enzymes on butanol accumulation in *E. coli* (Inui et al., 2008). They concluded that *adhE2* has higher specificity towards butyryl-CoA than *adhE1*, resulting in around 4-fold improvement on butanol production. On the other hand, in the work published by Nielsen *et al.*, the higher specificity towards butyryl-CoA was not reflected on butanol accumulation. Similar butanol titers were obtained independently of the bifunctional enzyme expressed (Nielsen et al., 2009).

The Dębabov group has expressed a single mutated alcohol dehydrogenase *adhE568*, so it would become active under aerobic conditions to explore the inverted β-oxidation pathway. To do so, a point mutation (G→A) was introduced in *adhE* coding sequence, leading to a Glu568Lys substitution in the protein product of the gene. Nonetheless, the maximum butanol production was less than 1 $mg.L^{-1}$ (Yu et al., 2012).

2.5.4 Conversion of alternative substrates

The production of biofuels from cheap and renewable raw-materials can lead to more sustainable processes able to compete with the petrochemical industry. Some alternative substrates that have been used for butanol production include: glycerol, cellulose, hemicellulose, switchgrass and fatty acids (Ndaba et al., 2015; Zheng et al., 2009).

Glycerol is an abundant by-product from biodiesel production and since it is a more reduced substrate when compared with glucose, it can enhance the synthesis of butanol. Zhou *et al.* explored the production of butanol from glycerol by overexpressing the transporter GlpF. This modification improved the transport of extracellular glycerol into the cell by 25 % and as a result the butanol production also improved 23 %. In this study, after disrupting NADH-competing pathways, the highest butanol titer achieved was 0.154 $g.L^{-1}$ (Zhou et al., 2014).

The Keasling group has developed a cellulolytic strain of *E. coli*, capable of growing on switchgrass. This strain was engineered to separately produce three fuels (butanol, pinene and fatty-acid ethyl

esters) without addition of exogenous glycoside hydrolases (Bokinsky et al., 2011). This strain of *E. coli* was engineered to secrete cellulases and hemicellulases, so it can grow on medium containing cellulose and hemicellulose. Switchgrass was first pretreated with ionic liquids (IL) to release cellulose and hemicellulose components. The strain containing the heterologous butanol pathway on a single plasmid was able to produce 28.5 $mg.L^{-1}$ butanol from defined rich medium containing 3.3 % w/v IL-treated switchgrass as the main carbon source. Nevertheless, in order to turn this process industrially relevant, both biomass-degrading and butanol-producing capacities should be improved.

Lignocellulosic raw-materials have been proposed as the most suitable feedstock for the biological production of fuels and chemicals. Nevertheless, the interest on using fatty acids (FA) as substrate had emerged due to the availability of (FA)-rich feedstocks and its efficient metabolism which can support high product yields on substrate. Particularly, FA metabolism to acetyl-CoA results in full carbon recovery in contrast with sugar metabolism where formate or carbon dioxide are also formed. Nevertheless, the incorporation of the highly reduced FA generates reducing-equivalents, requiring the presence of an electron acceptor. So, the conversion of FA is only possible under aerobic conditions which hinders butanol production. The conversion of FA in *E. coli* is mediated by enzymes encoded by the *fad* regulon and the *ato* operon. By expressing these genes and engineering a respiro-fermentative metabolism, Dellomonaco *et al.* have developed strains of *E. coli* converting FA into butanol. The maximum butanol titer was 2.2 $g.L^{-1}$ (Dellomonaco et al., 2010).

The lower butanol titers when compared with other studies suggests that further improvements are required to achieve significant product yields on alternative substrate. Since biobutanol production costs are highly depend on the substrate price, efforts should be devoted to developing an efficient production process from cheap and widely available raw materials derived from biomass.

2.5.5 Gene expression optimization

The maximization of the carbon flux in a pathway implies the fine-tuning of the heterologous gene expression. To do so, several techniques are available such as codon-optimization, modulation of Ribosome Binding Sites (RBS); manipulation of mRNA stability; engineering promoter strengths, or modification of gene copy number (Zhao et al., 2017).

The modification of the RBS of an mRNA transcript controls the translation efficiency, allowing to regulate enzyme production at the RNA level (Copeland et al., 2013). Ohtake and colleagues designed different RBS to control *adhE2* expression and achieve an optimal translation rate (Ohtake et al., 2017). In this study, a library of eight clones was generated using RBS calculator (Salis, 2011), optimizing the production of butanol and lowering the by-product formation.

Codon optimization allows to enhance the translation of heterologous proteins in new hosts. When expressing genes from distant organisms to produce butanol, it is common to codon-optimize the genes for translation in *E. coli* (Bond-watts et al., 2011; Nielsen et al., 2009).

The integration of butanol production genes into the genome under the control of native fermentation regulatory elements (FRE) of the major fermentative genes proved to be an efficient approach (Wen and Shen, 2016). This allowed to construct a self-regulated butanol production system in *E. coli*, a strain able to auto-induce butanol production under anaerobic conditions. Moreover, from the industrial point of view, this strategy is very advantageous, avoiding the use of IPTG which reduces the production costs. In this study, different FRE (FRE*ldhA*, FRE*frd*, FRE*adhE*, FRE*ackA*) were combined with the butanol producing genes resulting in several FRE::gene constructions. The best strain was able to excrete 10 $g.L^{-1}$ of butanol under anaerobic conditions.

2.5.6 Development of new pathways

Another possible approach to tackle the issues of expressing clostridial butanol pathway into *E. coli* is to develop novel pathways more suitable to the host metabolism. Most of the studies reported so far in this chapter have focused on engineering clostridial pathway, as depicted in Figure 2.2. Until now, only two significantly different alternative pathways for producing butanol have been tested in *E. coli*. One explores the keto-acids metabolism, the other engineering the beta oxidation pathway present in *E. coli*.

Liao's group have developed a strain of *E. coli* able to simultaneously produce butanol and propanol by exploiting the keto-acid pathway (Shen and Liao, 2008). In this work, *E. coli* was engineered to increase the pool of 2-ketobutyrate, a common keto-acid intermediate in isoleucine biosynthesis. This keto-acid can then be converted into 1-propanol, by the action of heterologous decarboxylases and dehydrogenases, or into butanol through the rare norvaline biosynthetic pathway. In this work, the

authors took advantage of the native amino acid pathway overcoming the need to involve CoA-dependent intermediates. The engineered strain was able to accumulate 0.8 $g.L^{-1}$ of butanol.

An approach to engineer the reverse β-oxidation cycle native from *E. coli* has also been demonstrated to allow n-butanol production under aerobic conditions. Dellomonaco *et al.* engineered *E. coli* to activate this pathway in the absence of the inducing substrate (fatty acids). The constitutive expression of this pathway without the respective substrate was achieved by introducing mutations in the corresponding transcriptional regulators (*fad*, *ato* and *crp*) and knock-out *arcA*. Further disruption of fermentation pathways (Δ*adhE*, Δ*frdA*, Δ*pta*) and overexpression of native thiolase (YqeF) and alcohol dehydrogenase (FucO) led to a production of n-butanol at 2.2 $g.L^{-1}$. A maximum butanol titer of 14 $g.L^{-1}$ was achieved by cultivating these cells in bioreactors (Dellomonaco et al., 2011).

In a similar approach, (Yu et al., 2012) also explored the reversed β-oxidation pathway, but only expressing enzymes from this pathway and *adhE* to convert butyryl-CoA into butanol. In this work, no competing pathways were eliminated but the maximum butanol titer achieved was quite low (less than 1 $mg.L^{-1}$) when compared with other studies.

The studies in this section show some promising results from exploring other pathways in *E. coli* for the production of butanol. In this regard, computational methods that generate new pathways from databases of enzymatic reactions are a promising tool to expand even more the portfolio of butanol production routes (Liu et al., 2015; Ranganathan and Maranas, 2010).

2.6 Computational Approaches for *in silico* Metabolic Engineering

The main goal of Metabolic Engineering (ME) is to develop efficient cell factories for the production of relevant compounds through the manipulation of biochemical pathways (McCloskey et al., 2013). The advent of the next-generation sequencing techniques combined with semiautomated annotation tools have increased the number of fully annotated genome sequences. Using the information about the genes present in a certain organism, it is possible to know the enzymes expressed and respective catalyzed reactions. The collection of reactions present in an organism can be combined with knowledge of cellular metabolism (e.g., biomass composition and energy requirements) that can

then be mathematically represented in genome-scale metabolic models (GSMMs). These models, when combined with constraint-based modeling (CBM) methods, can predict phenotypes and support rational ME-driven strategies (Baumler et al., 2011; Conrad et al., 2011; Feist et al., 2009; Maia et al., 2016).

2.6.1 Genome-scale metabolic models

The reconstruction of GSMMs starts by computing the set of genes in an organism that code for enzymes and attributing enzymatic reactions to each of them. Some reconstruction software tools such as merlin (Dias et al., 2015), RAVEN toolbox (Agren et al., 2013) and Model SEED (Devoid et al., 2013) can support this process in a (semi)automated fashion. The draft model can then be refined with gap-filing and manual curation, which is an iterative process that can also require the integration of experimental data from difference sources. Information about the biochemical reactions is usually retrieved from databases such as KEGG (Goto et al., 1997) and BRENDA (Scheer et al., 2011), while other data for model validation and refinement need to be gathered from the literature. The process of reconstructing GSMMs involves several stages in a cyclic way (as depicted in Figure 2.3) and has been reviewed in considerable detail elsewhere (Bordbar et al., 2014; Saha et al., 2014).

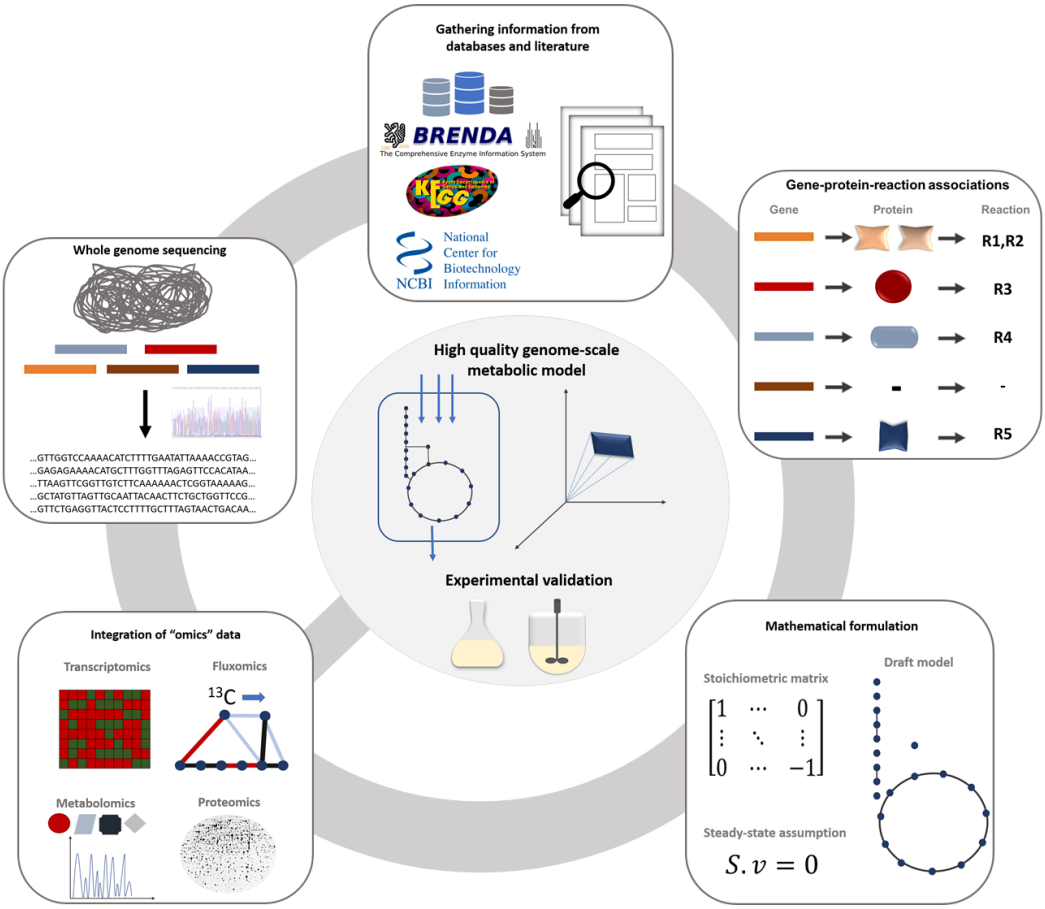


Figure 2.3: Overview of the reconstruction of a genome-scale metabolic model.

The first GSMM available for *E. coli* was published in 2000 and since then, a large number of models became available for a wide range of microorganisms. For *E. coli*, many GSMM are available, as showed in Table 2.3.

Table 2.3: List of available genome-scale models for E. coli.

Model	Strain	Genes	Reactions	Reference
*i*JE660	K-12 MG1655	660	627	(Edwards and Palsson, 2000)
*i*JR904	K12 MG1655	904	931	(Reed et al., 2003)
*i*AF1260	K12 MG1655	1260	2077	(Feist et al., 2007)
*i*CA1273	W (ATCC 9637)	1273	2477	(Archer et al., 2011)
*i*JO1366	MG1655	1366	2251	(Orth et al., 2011)

Most of the available genome-scale models available today were reconstructed for *E. coli* K12 MG1655. The only exception is the model iCA1273, which was developed for *E. coli* W. Regarding K12-based models, it is possible to observe the increasing number of genes and reactions covered by the models. The most complete *E. coli* GSMM covers 1366 genes, 2251 metabolic reactions and 1136 unique metabolites.

Examples of metabolic engineering approaches based on *E. coli* GSMM include the overproduction of lycopene (Alper et al., 2005), L-valine (Park et al., 2007) and L-threonine (Lee et al., 2007). A comprehensive list of the applications of GSMM of *E. coli* was reviewed in (McCloskey et al., 2013).

2.6.2 Phenotype simulation and optimization methods

The GSMMs include all the known metabolic reactions and the respective encoding genes. Nevertheless, in order to calculate internal flux distributions, simulation algorithms are required. FBA was the first developed method applied to the study of GSMMs (Schuetz et al., 2007). Briefly, this simulation method assumes steady state, imposes constraints on metabolite balances and computes a possible flux distribution by optimizing a certain objective function (Orth et al., 2010). A commonly used objective function for simulating microbial models relies on the assumption that biomass production should be maximized. Although the maximization of biomass works remarkably well for simulating wild-type phenotypes, when predicting growth rates for engineered strains, the results can be quite different from the experimental values (Segrè et al., 2002; Shlomi et al., 2005). It has been shown that adaptive evolution can help engineered strains to recover some of the growth deficit, converging with the values simulated using FBA (Fong et al., 2005; Ibarra et al., 2002). Nonetheless, several methods, such as MOMA and ROOM, have been proposed to tackle this issue (see Table 2.4 for a collection of simulation methods).

Table 2.4: Phenotypic simulation methods for genome-scale metabolic models.

Method	Description	Reference
FBA	Maximizes an objective function (usually biomass growth) by linear programming. Allows to predict the flux distribution and biomass growth for wild-type and knock-out strains	(Ibarra et al., 2002)

pFBA	Same as FBA but applying a two-level optimization principle: maximization of a certain objective function while minimizing the total sum of fluxes or the number of active enzymes. The underlying assumption is that there is a natural selection for the strains able to grow the most quickly and efficiently using the lowest number of enzymes possible	(Ponce De León et al., 2008)
MOMA	The underlying assumption of this simulation method is that the flux distribution in a knock-out mutant strain should remain as close as possible to the wild-type. To do so, MOMA minimizes the Euclidean distance between the reference flux distribution (for the wild-type) and the mutant set of fluxes	(Segrè et al., 2002)
LMOMA	Same as MOMA, but replaces Euclidean distance by Manhattan distance.	(Schellenberger et al., 2011)
PSEUDO	This method coincides with MOMA in the minimization of the Euclidean distance between the mutant and the wild-type set of fluxes. The difference is that PSEUDO uses a region of flux space (which allows nearly optimal growth) instead of a single flux distribution.	(Wintermute et al., 2013)
ROOM	Computes the set of fluxes in mutant strains by minimizing the number of activated or inactivated reactions when compared with the reference flux distribution (wild-type)	(Shlomi et al., 2005)
MimBI	The formulation of this method relies on the assumption that knock-out strains minimize the metabolite turnovers comparing with wild-type, i.e. the sum of producing and consuming fluxes	(Brochado et al., 2012)
Under/Over expression plugin for Optflux	This method imposes constraints to the fluxes in order to simulate under and over expressions, using as reference the wild-type flux	(Gonçalves et al., 2012) (Rocha et al., 2010)

Simulation methods are very useful to calculate flux distributions in wild-type and knock-out mutant strains. Nevertheless, manually finding combination of genetic interventions iteratively to reach a desired phenotype poses as a fastidious task. In this regard, several methods for discovering combinations of metabolic interventions required to achieve a certain engineering objective are available and have been reviewed elsewhere (Maia et al., 2016). Some examples of computation tools for target discovery using genome-scale metabolic models are presented in Table 2.5.

Table 2.5: Examples of computational tools for target discovery using genome-scale metabolic models.

Method	Description	Reference
OptKnock	Bilevel programming framework that returns a single optimum strain by setting the maximum number of knock-outs. In this method, the knock-out strategy ensures that cellular growth accompanies the biochemical overproduction.	(Burgard et al., 2003)
RobustKnock	Similar to OptKnock, but assuring that the solutions provided are robust, i.e., there is no variability on the target product flux in the strain design provided by this method.	(Tepper and Shlomi, 2010)
OptGene	Based on an evolutionary algorithm, it finds sets of knock-outs that meet a certain objective by using available simulation methods to compute the flux distributions (Table 2.4) The output is several near optimal strain designs with maximization of the target compound.	(Patil et al., 2005)
SA/SEA	Evolutionary and simulated annealing algorithms to search near optimal sets of genetic modifications that maximize the flux towards a target compound. Similar to OptGene, the main difference is the variable-sized solutions (sets of solutions with different number of knock-outs).	(Rocha et al., 2008)
OptForce	In OptForce, the set of solutions that guarantees an overproduction of the target compound is not based on biomass maximization. In this method, the set of solutions is found by minimizing the set of fluxes that must be forced to provide an optimal production of the target compound. The range of fluxes variability is calculated by iteratively maximizing and minimizing each flux.	(Ranganathan et al., 2010)

CosMos	Similar to OptForce, but fluxes are not limited to upper and lower bonds.	(Cotten and Reed, 2013)
OptSwap	This method is formulated as a MILP (mixed-integer linear programming) problem, similarly to RobustKock but including cofactor binding specificity of enzymes as possible targets for optimization.	(King and Feist, 2013)

2.6.3 Enumeration of novel pathways

Most early endeavors to design biochemical pathways were traditionally based on intuitive knowledge and by manually searching the literature and databases. Currently, with the increasing number of enzymatic reactions discovered and genome sequences available, the use of computational tools poses as an effective alternative to exploit this information. In this regard, pathway discovery tools enable to prospect for novel pathways by assembling enzymatic steps from multiple organisms (Croes et al., 2005; Kim et al., 2017). The work developed by (Yim et al., 2011) is a successful example of 1,4-Butanediol recombinant production in *E. coli* through a pathway generated by a pathway discovery tool.

The increasing number of metabolic reactions compiled in databases has broadened the spectrum of possibilities for the prospection of new pathways. In this regard, pathway discovery tools can help to assemble new pathways by identifying reactions that can fill the gap between a certain substrate and a product. To do so, information from databases must be retrieved and standardized, forming a metabolic network, the search space in which new pathways are computed by the algorithm. The application of computational tools for designing and reconstructing metabolic pathways was comprehensively reviewed by (Wang et al., 2017), gathering 139 references and by (Kim et al., 2017) referring 70 works. Briefly, the implementation of pathway design algorithms involves the following steps (Figure 2.4): (i) retrieving information from databases and the respective standardization; (ii) representation of the database; (iii) network pruning (iv) enumeration of pathways by the selected search algorithm (v) pathway ranking to select the most suitable one(s) (Wang et al., 2017).

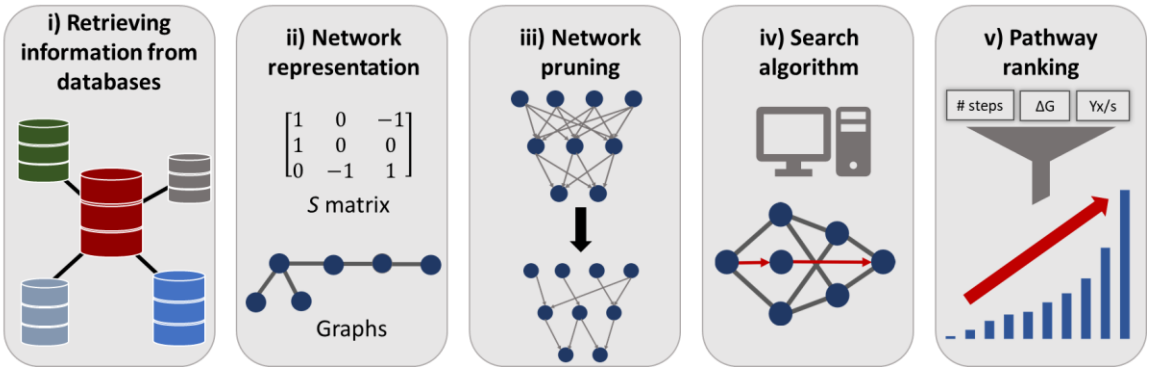


Figure 2.4: Schematic representation of the implementation of pathway algorithms workflow.

The pathway discovery tools are based on the reactions and metabolites retrieved from existing databases such as KEGG (Goto et al., 1997), BRENDA (Scheer et al., 2011) and MetaCyc (Caspi et al., 2012). Nevertheless, the information collected from these databases may include redundant names for reactions and metabolites or stoichiometric imbalances. So, a manual curation step could be necessary to standardize the gathered information. The reactions and metabolites retrieved from databases constitute a metabolic network, which is used for pathway search (Wang et al., 2017).

The metabolic network can be represented by a stoichiometric matrix (*S*) or graph-based systems. Graph-based systems are often limited to linear paths over the graph, excluding biochemical reactions with more than one reactant/product. In this regard, hypergraph representations address this issue by including multiple reactants and products. In hypergraphs, multiple reactants are connected to multiple products with a single hyperedge, the reaction. While the *S*-matrix and hypergraph representations already account for all the participating metabolites, other graph-based systems representations often require further processing steps to balance the predicted pathways (Liu et al., 2015; Wang et al., 2017).

Considering graph-based representations of metabolic networks, the shortest path between two compounds can be biologically meaningless due to the existence of cofactors and pool metabolites. The highly connected metabolites (e.g. ATP, H_2O, NADH) favor its inclusion in the shortest path as substrate or product, when usually these compounds act only as side metabolites or cofactors. This issue can be tackled by simply excluding cofactors and pool metabolites from the metabolic network. However, by doing so, solutions that are able to synthesize these compounds (e.g. ATP) will not be discovered. More systematic approaches have been suggested to carry an efficient network pruning. For instance, by applying weight to each compound, which penalizes compounds with a high degree

(i.e., highly connected metabolites), the algorithm avoids the use of currency metabolites whenever possible. The atom tracking approach, by incorporating the chemical structure of the compounds, allows to track the conservation of atoms between a substrate and product excluding currency metabolites and, simultaneously, giving insight into the efficiency of the pathway. In stoichiometric-based methods mass balances already are comprised by employing (Croes et al., 2005; Liu et al., 2015).

The selection of the search algorithm is intrinsically related with the network representation and also depends on the type of pathways desired (linear or branching) (Wang et al., 2017).

Pathway generation algorithms can generate a large number of pathways, which need to be ranked according to diverse criteria to facilitate the respective evaluation. The size of the pathway (i.e. the number of reaction steps from a substrate to the product) is the most common method to sort pathways. Shorter pathways are preferred due to the lower number of genes necessary (assuming each enzyme is encoded by a gene) and consequent lower metabolic burden imposed to the host cells. The calculation of the free Gibbs energy is also a common parameter since it allows to evaluate the thermodynamic feasibility of the pathway. These values can be estimated using eQuilibrator (Flamholz et al., 2012). The integration of the generated pathways in a GSMM for the target host and further simulation using FBA enables to infer the feasibility of the pathway in the host context. Moreover, the product yield can be predicted, which can also be used in the pathway ranking. Databases such as Tox21 (Tice et al., 2013), allow to evaluate the toxicity of intermediate compounds on the host strain, which can further help filtering the sets of generated pathways (Carbonell et al., 2012; Medema et al., 2012; Sajo Mienda and Shahir Shamsir, 2015)

The selection of the most suitable DNA sequence encoding the respective reaction is a fastidious task. The integration of an interface that give suggestions for heterologous genes based on enzymatic activity and experimental validation on the selected host is a future perspective. Currently, most of this analysis is still manually carried by consulting relevant literature and databases (Kim et al., 2017).

The application of optimization algorithms (Table 2.5) to the host GSMM including the heterologous pathway incorporated, allow to further predict the genetic interventions necessary to improve the production of the desired compound and reduce by-product formation (Maia et al., 2016).

The enumeration of novel pathways is a powerful tool, allowing to identify non-obvious combinations of enzymes to produce a certain compound. The integration of these novel pathways into the GSMM of the host cell enables a rational design of more efficient cell factories. Particularly for the production of butanol, alternative heterologous pathways have been generated (Liu et al., 2015; Ranganathan and Maranas, 2010).

2.7 Conclusions

In this chapter, we gathered the different approaches reported to produce butanol in *E. coli*, analyzing the obtained progress (Figure 2.2 and Table 2.2). Most of the metabolic engineering strategies have been applied to increase the NADH pool available either by increasing its accumulation or by disrupting competing pathways. The conversion of alternative substrates, expression of homologue enzymes and other molecular biology strategies were also covered. A wide range of culture conditions has been tested, as well as different *E. coli* strains. From the first published recombinant production of butanol in *E. coli*, butanol titer has increased more than 50-fold, from 0.552 $g.L^{-1}$ to 30 $g.L^{-1}$, the maximum value obtained so far. This titer was obtained by cultivating the cells in a continuous bioreactor with *in situ* butanol recovery. The values obtained already can compete and even exceed clostridial yields, confirming the potential of using *E. coli* as a butanol cell factory. Regarding metabolic engineering approaches, further optimization can be achieved by combining the comprehensive GSMMs with computational tools.

In parallel, the exploration of alternative pathways has showed promising results. In this regard, pathway discovery tools can generate novel routes to produce butanol in *E. coli* by retrieving metabolic reactions from databases and assembling enzymes from different sources. The integration of these novel routes into GSMM and further analysis using phenotype simulation and optimized methods can support the development of efficient cell factories.

CHAPTER 3 *In silico* analysis, simulation and optimization of alternative metabolic pathways to produce n-butanol in *Escherichia coli*

Global demand for n-butanol is increasing and it is expected to achieve a global market of USD 9.9 Billion in 2020. Although butanol can be naturally produced via ABE fermentation by *Clostridium sp.* strains, most of this compound is still produced using chemical processes. One possible approach to overcome the difficulties of industrial exploration of clostridial strains, is to produce heterologously this compound in a more robust microorganism such as *Escherichia coli* or *Saccharomyces cerevisiae*. This chapter describes the *in silico* driven design of a *E. coli* strain that is able to produce butanol through a novel pathway, generated using a hyper-graph algorithm.

The novel butanol pathway was selected from an initial set of 105,954 different pathways previously generated (Liu et al., 2015), successively downsized to a set of 24 promising pathways to express *in vivo*, by applying different filters (such as stoichiometric feasibility, size, novelty and conservation of carbon atoms). These pathways were integrated in the *E. coli* genome-scale metabolic model *i*JO1366, and using simulated annealing and evolutionary algorithms, an optimum genotype was discovered, in which butanol production is maximized and by-products formation minimized under anaerobic conditions.

Instead of the commonly used precursor acetyl-CoA, in this pathway butanol is synthesized via 2-oxoglutarate, the keto-acid formed by deamination of glutamate and an intermediate in TCA cycle.

The combination of a computational analysis, manual curation of the results and literature review, in a systematic fashion, led to a deep understanding of the genetic alterations proposed and their role in *E. coli* metabolism. Considering the exhaustive *in silico* analysis and the availability of gene sequences for the heterologous enzymes catalyzing this novel pathway, this solution has great potential to be applied *in vivo*.

3.1 Introduction

During the process of creating a new cell factory, one of the first steps involves the selection of the most appropriate pathway to produce the target chemical. In addition to known established biosynthetic pathways, the identification of alternative non-natural routes is also a potential strategy. One example reported in the literature is the synthesis of butanol and propanol by conversion of keto-acids intermediates instead of acetyl-CoA, the commonly used precursor (Atsumi et al., 2008b; Shen and Liao, 2008).

Pathway enumeration algorithms can generate improved and non-obvious pathways by retrieving information from databases of enzyme catalyzed reactions and rearrange it. Lately, several different pathway computation approaches have been developed: constraint-based modeling (Chatsurachai et al., 2012; Pharkya et al., 2004), graph-based methods (Croes et al., 2005) and knowledge-based systems of chemical rules (Arita, 2003; Heath et al., 2010). Among these alternative methods, butanol production pathways have been enumerated by Liu *et al.* (Liu et al., 2015) using two different hypergraph algorithms: Solution Structure Generation (SSG) (Friedler et al., 1995; Lee et al., 2005) and FindPath (FP) (Carbonell et al., 2012). Using modified versions of the original algorithms, Liu *et al.*, enumerated pathways to produce butanol, vanillin and curcumin in *Escherichia coli* and *Saccharomyces cerevisiae*. FP generally outperformed SSG, computing solutions with a further *radius* (minimal number of reactions between a source of substrates and the target products) and, consequently, providing larger sets of solutions.

The pathways generated with SSG and FP can be inconsistent at the stoichiometric level and further processing – such as applying Flux Balance Analysis (FBA) – is needed to validate them. FBA allows to predict flux distributions in metabolic networks at steady-state towards maximization of an objective function (Orth et al., 2010). In the butanol case-study mentioned above, out of the 60,356 solutions obtained from FP algorithm for KEGG search space, only 22,968 actually allowed butanol production when integrated in the *E. coli i*JO1366 GSMM (Liu et al., 2015). Within these compatible solutions, further manual curation is still needed. For instance, all KEGG reactions are assumed to be reversible and to determine to which extent a given reaction (or a set of reactions) is catalyzed in the direction of the product of interest, values such as change in free Gibbs energy must be evaluated. Lastly, curated gene sequences must be available for each metabolic reaction, in order to express these pathways *in vivo*.

Besides the discovery of non-obvious pathways, computational metabolic engineering tools allow to rationally design strategies to redirect metabolic fluxes to the target products to be applied *in vivo*. Thus, saving time and labor by minimizing laboratory experiments. The increasing number of total genome sequences available potentiate the reconstruction of Genome-Scale Metabolic Models (GSMMs) (H. Chan et al., 2013). These models can be used to predict phenotypes and, combined with other tools, maximize the production of a target product while minimizing by-products formation (Maia et al., 2016). Since microorganisms have evolved towards their own requirements, genetic manipulation is needed to adjust their metabolism for fulfilling industrial objectives. Computational Strain Optimization Methods (CSOMs) based on Constraint Based Modelling (CBM) can predict which genetic modifications are required to maximize the production of a certain compound. Besides gene deletions, these methods can also find heterologous gene insertions, gene over/under expressions, and cofactor specificity swapping (Maia et al., 2016; Maranas and Zomorrodi, 2016).

The main goal of this work is to find a set of novel butanol producing pathways to express in *E. coli*. For this, pathways previously generated using the hyper-graph algorithm FindPath (Liu et al., 2015) were analyzed to infer their feasibility to be applied *in vivo*. First the solutions were evaluated according to diverse criteria (maximum butanol yield, number of reactions and novelty) to seek the most promising solutions. Lastly, using computational tools based on CSOMs, the *in silico* butanol producing *E. coli* was optimized by searching for sets of genetic modifications that maximize product excretion.

3.2 Materials and methods

3.2.1 Analysis of heterologous pathways

Several alternative routes to produce n-butanol were generated by (Liu et al., 2015) using an hyper-graph algorithm based on FindPath (FP) (Carbonell et al., 2012). These solutions were analyzed to select the most promising ones to implement *in vivo*. The first step of this analytical process was, using an internal digital platform from *SilicoLife,* to test if the pathways would allow butanol production when inserted in the *i*JO1366 *E. coli* genome-scale metabolic model (GSMM) (Orth et al., 2011). This test was carried using FBA with maximization of butanol flux as the objective function, for each pathway generated by FP. Since stoichiometry is not considered by FP, the FBA simulations

without butanol production (i.e. maximum butanol flux = 0) allowed to identify the stoichiometrically infeasible pathways. Pathways without any butanol flux were discarded, and, in the same platform, diverse filters were successively applied to the remaining pathways. First, the pathways were sorted by size, i.e. the number of reactions needed to catalyze the initial precursor into butanol. Since clostridial pathway is constituted by six steps, the threshold of the pathways size was set in seven. Within the diverse groups sorted by size, the pathways were ranked accordingly to the conservation of number of carbon atoms (i.e. the difference between the number of carbons of the initial substrate and butanol). For the pathways with the greatest conservation of carbon atoms, a manual analysis was carried out.

To assure the novelty of the chosen pathways, the relevant literature was searched for any previous reports of butanol production using the same set of reactions and the ones already reported were discarded. Then, the availability of curated gene sequences encoding the enzymes that catalyze the different reactions was verified by consulting databases such as UniProt, KEGG and MetaCyc. For all the reactions constituting the most promising set of pathways, the change in Gibbs free energy ($\Delta_r G'^m$) and Equilibrium constant (K'_{eq}) were calculated using eQuilibrator 2.0 (Flamholz et al., 2012) to evaluate their reversibility. These values were estimated using Component Contribution (Noor et al., 2013) considering a pH of 7 and an ionic strength of 0.1 M.

3.2.2 Genome-scale model and software

The reactions constituting the most promising pathway obtained from the analysis described in Figure 3.2 were added to *i*JO1366 *E. coli* GSMM (Orth et al., 2011). Table 3.1 shows the stoichiometry of the reactions added to the models, as well as transport reaction to allow the excretion of butanol. The reaction R01175 was not added to the GSMM, since the model already includes this reaction (R_ACOAD1f). The reaction (S)-3-hydroxybutanoyl-CoA hydrolyase (R_ECOAH1) was deleted from the GSMM to make sure that butanol production originated from the heterologous pathway.

OptFlux (Rocha et al., 2010) was used to perform all simulations. FBA was used as the simulation algorithm using biomass growth rate as objective function.

Table 3.1: Reactions added to E. coli *Genome Scale Metabolic Model* i*JO1366 corresponding to the different catalytic steps of the most promising heterologous pathway to produce butanol, some of them have alternative cofactors, as well as drain to excrete this compound.*

Step	ID	Reaction
1	R03534	2-Hydroxyglutarate + FAD <=> 2-Oxoglutarate + $FADH_2$
	R08198	2-Hydroxyglutarate + NAD^+ <=> 2-Oxoglutarate + NADH + H^+
2	R04000	Acetyl-CoA + 2-Hydroxyglutarate <=> Acetate + 2-Hydroxyglutaryl-CoA
3	R03937	Glutaconyl-1-CoA + H_2O <=> 2-Hydroxyglutaryl-CoA
4	R03028	Glutaconyl-1-CoA <=> Crotonyl-CoA + CO_2
5	R09738	Butanoyl-CoA + $NADP^+$ <=> Crotonyl-CoA + NADPH + H^+
	R01171	Butanoyl-CoA + NAD^+ <=> Crotonyl-CoA + NADH + H^+
6	R01173	Butanal + CoA + $NADP^+$ <=> Butanoyl-CoA + NADPH + H^+
	R01172	Butanal + CoA + NAD^+ <=> Butanoyl-CoA + NADH + H^+
7	R03544	Butanal + NADH + H^+ <=> 1-Butanol + NAD^+
	R03545	Butanal + NADPH + H^+ <=> 1-Butanol + $NADP^+$
Drain	R_EX_C06142_	1-Butanol_c<=> 1-Butanol_e

3.2.3 *In silico* optimization and simulation of n-butanol production

The optimization analysis was performed in the digital platform of *SilicoLife*. Evolutionary (EA) and simulated annealing (SA) algorithms (Rocha et al., 2008) were used for strain optimization purposes. These algorithms were set up to find genetic modifications that lead to increased production of n-butanol, including co-factor swapping (King and Feist, 2013) and/or gene knock-outs (Maia et al., 2016). FBA was used as the simulation method and two different objective functions were considered for the optimization: WYIELD and WBPCY. Both objective functions promote the search for strain designs that show robust butanol production, i.e., when biomass is at its maximum value and the butanol flux is minimized (FVA_{min}) its value remains constant. WYIELD uses the maximum (FVA_{max}) and minimum (FVA_{min}) butanol fluxes to guide the optimization process using the expression: $(\alpha \cdot FVA_{max}) + ((1-\alpha) * FVA_{min}))$ with $\alpha = 0.001$. This objective function allowed to give priority to robust solutions without discarding the non-robust by giving a small weight to the FVA_{max}.

The objective function WBCPY works in the same fashion as the previous one, but also including the biomass growth: $\mu.[(\alpha \cdot FVA_{max}) + ((1-\alpha) * FVA_{min}))]$.

For both objective functions, solutions with biomass growth less than 0.01 h^{-1} were eliminated.

The number of function evaluations was set to 50,000, the maximum number of knock-outs was set to 30 and the maximum number of co-factor swaps to 4. Only the heterologous reactions with alternative co-factors belonging to the butanol production pathway were available for swapping. These reactions catalyze the conversion of the same substrate into the same product, however using different cofactors. For these reactions, the algorithm could recognize the reaction was duplicated, activating one and knocking out the alternatives.

Two different environmental conditions were tested: anaerobic (oxygen uptake flux was set to 0 $mmol.(g_{DW}.h)^{-1}$) and aerobic (oxygen: unconstrained uptake). For both, the glucose uptake rate was equal to 10 $mmol.(g_{DW}.h)^{-1}$ and ammonia, phosphate and sulfate had unconstrained uptake. For both environmental conditions, the non-targets (reactions not subjected to optimization) for deletions were: drains/transporters; critical reactions (when excluded impair biomass growth); reactions that under the respective environmental conditions are mathematically prevented from having flux, and finally, equivalent reactions, i.e. reactions from the same linear pathway whose knock-outs are mathematically equivalent with each other. A manual nontarget was included in both environmental conditions: the pseudo-reaction of ATP maintenance (R_ATPM).

In the end, the optimization results were simplified by eliminating reactions that did not contribute to the solution fitness and removing repeated solutions. The remaining solutions were aggregated by number of modifications and Biomass-Product Coupled Yield (BPCY). For both environmental conditions, first, a simulation was performed using the model without any knock-outs. Then, the simplest (i.e. with the lowest number of inactivated reactions) solution combined with the maximum BPCY was simulated using FBA (Flux Balance Analysis) with maximization of biomass growth as objective function. In order to understand the flux distribution, other genetic modifications were simulated according to the obtained results. For each mutant, biomass growth and key products were calculated.

Flux Variability Analysis (FVA) allowed to infer the robustness of the mutant that provided the maximal BCPY with the lowest number of knocked-out reactions. FVA was used to create a graph that relates

the maximum yield of butanol for diverse levels of biomass formation. The number of steps was set in 20, the environmental conditions were anaerobic with glucose uptake equal to 10 mmol.$(g_{DW}.h)^{-1}$.

3.3 Results

3.3.1 Computational analysis and manual curation of heterologous pathways

The set of heterologous pathways previously generated (Liu et al., 2015) had 105,954 alternatives. After discarding all the pathways without any flux in the target product (butanol), only 40,608 different pathways were left. Since the clostridial butanol pathway possesses 6 reactions (see Figure 2.1), a competing alternative should not be much larger. Therefore, the dataset was filtered to include only pathways with a maximum of 7 reactions. The application of this filter originated 316 pathways, which, in their turn, were ranked by size and maximum conservation of number of carbon atoms, resulting in four different batches of pathways: 4 pathways with size equal to three reactions; 4 pathways with size equal to four reactions; 120 pathways with size equal to six reactions and 188 pathways with size equal to seven reactions. Within this 4 different batches, a manual analysis was carried out for the pathways with the greatest conservation of carbon atoms. After reviewing the literature to infer their novelty, 24 pathways, all constituted by seven catalytic steps, have remained and considered the most auspicious for further analysis and implementation *in vivo*. In Figure 3.1Table 3.1, the various stages of the *in silico* analysis from the initial set of pathways are represented.

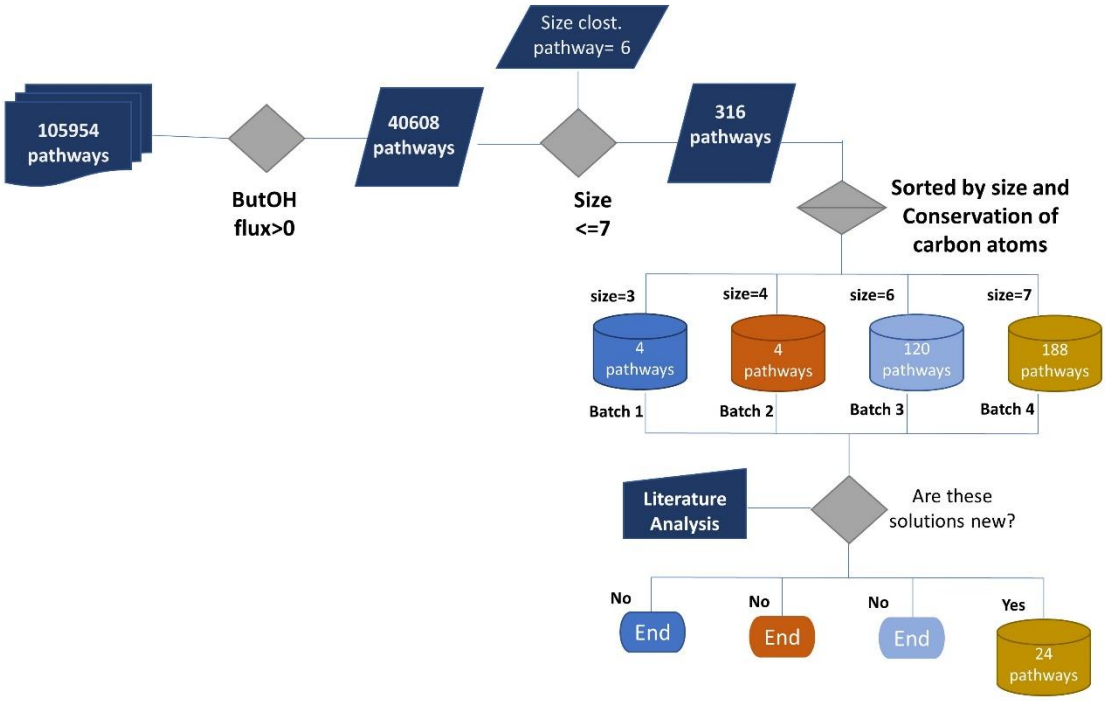


Figure 3.1: Flowchart representing in a simplified way the successive stages of the in silico *analysis of the heterologous pathways to produce butanol in* E. coli *generated by* (Liu et al., 2015)*, resulting in a set of the most promising solutions containing 24 pathways. ButOH – Butanol; clost. - Clostridial.*

The fact that pathways already described in the literature have been generated by FindPath, is a good indicator to validate the algorithm and this strategy. The pathways constituted by three steps were considered bad candidates as they all started with 3-oxohexanoyl-CoA. This metabolite is an intermediate of the fatty-acid β-oxidation pathway. In reported works using this pathway to produce butanol, more enzymes were expressed (Dellomonaco et al., 2011). Regarding batch 2 (pathways with four reactions), all pathways consisted in the conversion of keto-acids into butanol, which is something already explored (Atsumi et al., 2008a). The batch 3, constituted by pathways with six reactions included, as expected, the clostridial butanol pathway and respective alternatives, as well as pathways converting compounds from fatty-acid β-oxidation intermediates. Finally, on the set of solutions with seven catalytic steps, 24 pathways were considered as the most promising routes and were further analyzed. Besides these pathways, routes converting malonyl-CoA were discovered. This pathway was already expressed in the cyanobacteria *Synechococcus elongatus* PCC 7942 producing butanol in a photosynthetic process (Lan and Liao, 2012). In Table 3.2, we summarize, for each batch of pathways, the precursors used by the main set of solutions discarded, as well as the references to literature describing their use for butanol production.

Table 3.2: Substrates and respective references already reported for butanol production for each batch of solutions (filtered by size).

Batch	Size	Precursors	Pathway	References
1	3	3-Oxohexanoyl-CoA	Fatty acid β-oxidation	(Dellomonaco et al., 2011)
2	4	3-Methyl-2-oxobutanoic acid	Keto-acids	(Atsumi et al., 2008b)
3	6	(S)-3-Hydroxybutanoyl-CoA	Fatty acid β-oxidation	(Yu et al., 2012)
		Acetyl-CoA	Clostridial	(Inui et al., 2008)
4	7	Malonyl-CoA	Photosynthetic	(Lan and Liao, 2012)

On the other hand, the 24 pathways constituting the most promising set of solutions to apply *in vivo* catalyzed the conversion of 2-oxoglutarate, a keto-acid formed by deamination of glutamate and an intermediate in TCA cycle, into butanol. This set of pathways has not been reported in the literature and, for this reason, are a target for further analysis.

3.3.2 Set of most promising solutions

Amongst the pathways generated by (Liu et al., 2015) to produce butanol, only some with size equal or higher than seven presented novelty, as already described. Further analysis allowed the identification of a set of heterologous pathways with potential to express in *E. coli*. Instead of the commonly used precursor acetyl-CoA, in this pathway butanol is synthesized via 2-oxoglutarate. The final set of candidate pathways all consisted of similar starting and intermediate metabolites but, in some steps, their conversion could use different cofactors resulting in many variations. First, 2-oxoglutarate is reduced to 2-hydroxyglutarate via reaction R08198 or R03534. Then, in reaction R0400, Coenzyme A is transferred from acetyl-CoA (releasing acetate) to 2-hydroxyglutarate forming 2-hydroxyglutaryl-CoA, followed by its dehydration into glutaconyl-CoA (reaction R03028). Step 4 is the reaction R03028, where glutaconyl-CoA is decarboxylated into crotonyl-CoA. Three different alternatives (R01171; R01175 and R09738) have been enumerated for the subsequent reduction of crotonyl-CoA in butanoyl-CoA. The last two steps – common to all the solutions generated – are the successive reductions of butanoyl-CoA to butanal (R01173 or R01172) and of butanal to n-butanol (R03544 and R03545). Regarding the conservation of carbon atoms, from 2-oxoglutarate

($C_5H_6O_5$) to n-butanol ($C_4H_{10}O$) one single carbon atom is lost under the form of CO_2 in step 4. In Figure 3.2, the set of the most promising pathways is represented.

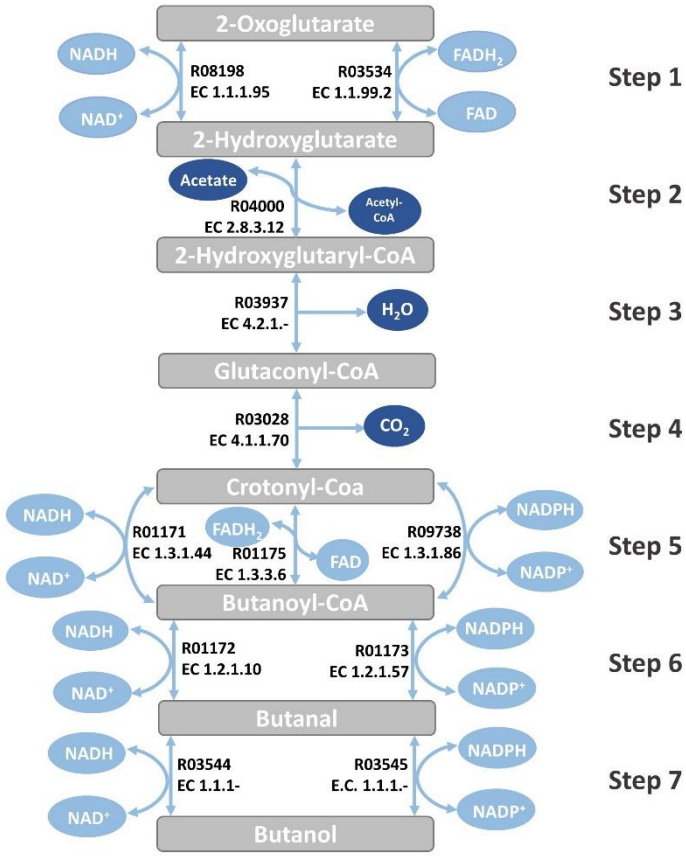


Figure 3.2: Set of the most promising pathways to express in E. coli*, able to convert 2-oxoglutarate into butanol through seven steps. This set possesses 24 possible pathways due to the possible combinations between the different reactions. The main metabolites involved are shown in grey square boxes; secondary metabolites are represented in dark blue circles, cofactors are represented in light blue circles; reactions are represented by bidirectional arrows with the respective KEGG ID and EC number.*

This set of solutions is constituted by 24 alternative pathways, because in four out of the seven steps there are alternative reactions using different cofactors (2x1x1x1x3x2x2=24). Since in the KEGG database all reactions are reversible by default, it was necessary to check if each reaction can occur in the desired direction towards butanol; or if it is more likely to happen in the opposite way. The change in Gibbs Energy (Δ_rG) is a practical indicator for reaction directionality: a reaction will be favorable in a certain direction if Δ_rG is negative, meaning Gibbs energy decreases during a reaction as ruled by the second law of thermodynamics (Flamholz et al., 2012).

Δ_rG^0 is the change in Gibbs free energy due to a chemical reaction in standard conditions and without accounting for pH, ionic strength or any other cellular factors. However, biochemical reactions

depend on the physiological conditions. $\Delta_r G'^{0}$ is the change in Gibbs free energy due to a reaction (at a particular pH and ionic strength), considering the reactants concentrations equal to 1 M. These concentrations are not realistic at a cell context, where metabolite concentrations typically range from 1 nM to 10 mM (Bar-Even et al., 2012), For this reason, we calculated $\Delta_r G'^{m}$, the change is Gibbs energy considering reactants concentrations of 1 mM.

The eQuilibrator web interface (http://equilibrator.weizmann.ac.il/) allows to easily estimate ($\Delta_r G'$) of a biochemical reaction, since it accommodates a comprehensive and accurate database of thermodynamic data (Flamholz et al., 2012). For all the reactions presented in Figure 3.2, $\Delta_r G'^{m}$ and K'_{eq} were calculated using eQuilibrator, these values were estimated using Component Contribution (Noor et al., 2013) considering a pH of 7 and an ionic strength of 0.1 M.

Table 3.3: Change in Gibbs free energy ($\Delta rG'^{m}$) under reactant concentrations of 1 mM and Equilibrium constant (K'_{eq}) values at a pH of 7 and an ionic strength of 0.1 M estimated with eQuilibrator (Flamholz et al., 2012) *for each of the reactions constituting the set of pathways and respective KEGG ID and EC number.*

Catalytic step	Reaction	Enzymatic Activity	$\Delta_r G'^{m}$ (*kJ.mol^{-1}*)	K'_{eq}
1	**2-Oxoglutarate** + NADH + H+<=> 2-Hydroxyglutarate + NAD^+	EC 1.1.1.95	-22.6±3.6	1.1×10^{-4}
	2-Oxoglutarate + $FADH_2$ <=> 2-Hydroxyglutarate + FAD	EC 1.1.99.2	-5.6±6.9	0.104
2	Acetyl-CoA + **2-Hydroxyglutarate** <=> Acetate + 2-Hydroxyglutaryl-CoA	EC 2.8.3.12	-8.6±15.3	32.1
3	**2-Hydroxyglutaryl-CoA** <=> Glutaconyl-CoA + H_2O	EC 4.2.1.167	0.7±2.9	0.751
4	**Glutaconyl-CoA** <=> Crotonyl-CoA + CO_2	EC 4.1.1.70	-35.8±17.4	1.9×10^{3}
5	**Crotonyl-CoA** + NADPH + H^+ <=> Butanoyl-CoA + $NADP^+$	EC 1.3.1.44	-57.0±16.0	9.6×10^{9}
	Crotonyl-CoA + NADH + H^+ <=> Butanoyl-CoA + NAD^+	EC 1.3.1.44	-40.0±17.0	1.0×10^{7}
	Crotonyl-CoA + $FADH_2$ <=> Butanoyl-CoA + FAD	EC 1.3.1.86	-57.9±16.0	1.40×10^{10}
6	**Butanoyl-CoA** + NADPH + H^+<=> Butanal + CoA + $NADP^+$	EC 1.2.1.10; EC 1.2.1.87	-9.4±16.1	0.0446
	Butanoyl-CoA + NADH + H^+<=> Butanal + CoA + NAD^+	EC 1.2.1.57	-10.4±16.1	0.0665
7	**Butanal** + NADH + H^+ <=> 1-Butanol + NAD^+	EC 1.1.1.-	-24.2±4.8	1.8×10^{4}
	Butanal + NADPH + H^+ <=> 1-Butanol + $NADP^+$	EC 1.1.1.-	-25.2±4.8	2.60×10^{4}

The $\Delta_r G'^{m}$ estimated values, as shown in Table 3.3, were negative for all reactions with exception of R03937. In this case, $\Delta_r G'^{m}$ value was close to 0 (0.7±2.9 KJ.mol^{-1}) and K'_{eq} value (0.751) close to

1, indicating that probably the reaction is reversible. The flux can still be forced through the forward direction by increasing the concentration of 2-hydroxyglutatyl-CoA sufficiently above the concentration of the products (Bar-Even et al., 2012). The subsequent step (R03028) is faster with a $\Delta_r G'^m$ of -35.8±17.4 KJ.mol^{-1} and K'_{eq} 1.9×10^{3}. So, we expect that this reaction ensures the concentration difference needed to drive the flux of reaction R03937 in the desired direction.

For step 1, the reaction R08198 (which uses NADH/NAD^+ as cofactors) has a $\Delta_r G'^m$ of -22.6±3.6 kJ.mol^{-1}, 4-fold lower than the alternative R03534, which uses $FADH_2$/FAD as cofactors. Also, the value of the Equilibrium constant is much higher (953-fold) for R08198 than R03534, indicating that, in the first case, almost all the substrate (2-oxoglutarate) should be converted into the product (2-hydroxyglutaryl-CoA). Therefore, R08198 appears to be thermodynamically more favorable than the alternative reaction.

Regarding the other steps (5, 6 and 7) with alternative reactions, no significant discrepancies were identified among them, considering $\Delta_r G'^m$ and K'_{eq} values.

3.3.3 *In silico* strain optimization and analysis

The simulated annealing and evolutionary algorithms supported the search for an optimum phenotype where the production of butanol was maximized, and byproduct formation minimized. The output could involve gene knockouts and cofactor swapping for the heterologous reactions.

For four of the seven steps that constitute the butanol production route, there were alternative reactions using different cofactors. Two different environmental conditions (aerobic and anaerobic) were tested using a glucose uptake rate equal to 10 mmol.(g_{DW}.h)$^{-1}$. For each condition, the algorithm selected the more advantageous reactions for butanol production.

The optimization was performed in the digital platform of *SilicoLife*. The different phenotypes of the optimization process were ranked by value of BCPY and number of knock-outs. The knock-out reactions included the inactive alternative reactions from the heterologous pathway.

For each environmental condition tested, the simplest (i.e. with the lowest number of inactivated reactions) optimization solution combined with the maximum BPCY was simulated was simulated

using *OptFlux* and subjected to further analysis to understand changes induced by the modifications on the flux distribution.

Table 3.4 shows the reference simulation (without any knock-outs) results under anaerobic conditions (simulation 1), the simulation of the top strain design obtained by the optimization process (simulation 2) and the simulation of two other mutants. The *in silico* growth rate is shown, as well as the main products excreted and the reactions used in each simulation for butanol synthesis. In the steps with alternative reactions, the cofactor of the active reaction is indicated.

Table 3.4: Simulation results using FBA for the different strains obtained using Optflux, *under anaerobic environmental conditions and glucose uptake of 10 mmol.(g_{DW}.h)$^{-1}$. For each simulation, the genetic manipulations, the active reactions from the added heterologous pathway, the biomass growth rate μ (h^{-1}) and the main products excreted (mmol.(g_{DW}.h)$^{-1}$) are shown.*

Simulation	Genetic manipulations	Active reactions for the production of butanol	μ (h^{-1})	Main products excreted (mmol.(g_{DW}.h)$^{-1}$)	
1	-	–	0.244	Succinate	0.08
				Acetate	8.26
				Formate	17.26
				Ethanol	8.32
				1-Butanol	-
2	R_ALCD2x=0 R_PTAr=0 R_ACALD=0 R_FUM=0 R_LDH_D=0	R08198 **(NADH)** R04000 R03937 R03028 R01171 **(NADH)** R01172 **(NADH)** R03544 **(NADH)**	0.102	Succinate	0.11
				Acetate	5.01
				Formate	10.27
				Valine	1.83
				CO_2	6.58
				1-Butanol	4.94
3	R_ALCD2x=0 R_PTAr=0 R_ACALD=0 R_LDH_D=0	R08198 **(NADH)** R04000 R03937 R03028 R01171 **(NADH)** R01172 **(NADH)** R03544 **(NADH)**	0.102	Succinate	0.18
				Acetate	4.98
				Formate	10.12
				Valine	1.84
				CO_2	6.56
				1-Butanol	4.90
4	R_ALCD2x=0 R_PTAr=0 R_ACALD=0 R_LDH_D=0 (["R_KARA1"],{"M_nadph_c" =>"M_nadh_c"})	-	0.103	Succinate	0.25
				Acetate	0.10
				Formate	0.45
				Valine	8.87
				CO_2	8.62
				1-Butanol	-

In simulation 1, no reactions were inactivated and the desired product - butanol - is not excreted. As expected for the growth of *E. coli* in anaerobic conditions the native mixed-acid fermentations products (acetate, ethanol, lactate, formate and succinate) are produced instead. It is worth noting that in simulation 1, the butanol production pathway is present in the model but when biomass is maximized the production is zero because producing butanol is less advantageous than the other fermentation products. Therefore, in order make *E. coli* prefer butanol as a fermentation product, some of the alternative NADH recycling pathways need to be inactivated. Even though the butanol production flux is zero in simulation 1, it is expected that even this strain should produce some butanol when the heterologous genes are present. Algorithms like the turnover dependent phenotype simulation (Pereira et al., 2015) can help quantify those basal production levels.

The best solution of the optimization process, simulation 2, consisted of five inactivated reactions (R_ALCD2x R_PTAr R_ACALD R_FUM R_LDH_D). Comparing with Simulation 1, the biomass growth decreased 2.4-fold from 0.244 to 0.102 h^{-1}.

The role of each knock-out was examined by sequential activation of each reaction (data not shown). Only the inactivation of R_FUM was not crucial for butanol production. For this reason, R_FUM was re-activated, creating the mutant design depicted in simulation 3. Comparing simulation 3 with the previous mutant, it is possible to observe slight changes in the fluxes of excreted products. Specifically, butanol production decreased from 4.94 to 4.90 mmol.$(g_{DW}.h)^{-1}$ and the value of biomass growth was maintained in 0.102 h^{-1}. Since the difference in butanol production between simulations 3 and 2 is negligible, the strain design of the mutant from simulation 3 has more potential to be applied *in vivo* because it requires one less knock-out.

Looking at the mechanism behind butanol production in the simulations 2 and 3, NADH appears to be used as driving force. This observation is supported by the deletion of the main reactions used in anaerobiosis for NADH recycling (namely R_ALCD2x and R_ACALD, used in ethanol production, and lactate synthesis through reaction R_LDH_D). Given the lack of the native NADH sinks, this co-factor must be recycled through the heterologous butanol pathway. This observation was further analyzed by adding a drain to butanal (the precursor of butanol) in both the mutants shown in simulations 2 and 3. The goal was to infer if, for the cell, it would be more advantageous to produce butanol (and recycle any extra NADH) or immediately excrete butanal. The butanol production results were identical to the ones obtained without this drain, showing that the cell uses butanol as a way of recycling NADH.

In order to investigate valine production and the role of cofactors in the butanol producing mutant from simulation 3, the cofactors in the (R)-2,3-dihydroxy-3-methylbutanoate oxidoreductase (R_KARA1), part of valine synthesis pathway, were shifted from NADPH/$NADP^+$ to NADH/NAD^+, originating the results shown for simulation 4. This simulation presented several differences when compared with simulation 3: first, the biomass growth increased 1.4-fold from 0.102 to 0.143 h^{-1}; then, butanol production became zero; valine production increased 4.82-fold from 1.84 to 8.87 mmol.$(g_{DW}.h)^{-1}$. Additionally, the reaction R_THD2pp, a $NAD(P)^+$ transhydrogenase, presented a flux of 10.45 mmol.$(g_{DW}.h)^{-1}$, while in the previous mutants it was inactive.

The significant changes observed show the importance of redox balance in metabolism, since changing only a cofactor in a single reaction completely altered the flux distribution. In simulation 3, a 19.64 mmol.$(g_{DW}.h)^{-1}$ total flux of NADH was recycled with an individual flux of 4.91 mmol.$(g_{DW}.h)^{-1}$ in the four NADH consuming reactions composing the butanol pathway (namely R08198, R01171, R01172 and R03544). In simulation 4, instead of the heterologous butanol pathway, most of the NADH was recycled through reaction R_KARA1 (8.87 mmol.$(g_{DW}.h)^{-1}$) and reaction R_THD2pp (10.45 mmol.$(g_{DW}.h)^{-1}$). Originally R_KARA1 (reaction involved in valine synthesis) uses NADPH/$NADP^+$ as cofactors and was active in simulation 3, where 1.84 mmol.$(g_{DW}.h)^{-1}$ of valine was excreted. For that reason, it seems that the redox balance established in simulation 3, required that NADPH was recycled through the valine pathway and NADH through butanol pathway. In simulation 4, the extra NADH that could be recycled in the valine pathway led to changes in redox balance that increased valine production and abolished butanol production.

To investigate if cell growth is dependent on the production of butanol or if butanol production is robust, we analyzed the flexibility of the fluxes. For that, using FVA, a graph was created correlating the maximum butanol production for different levels of biomass formation, as shown in Figure 3.3.

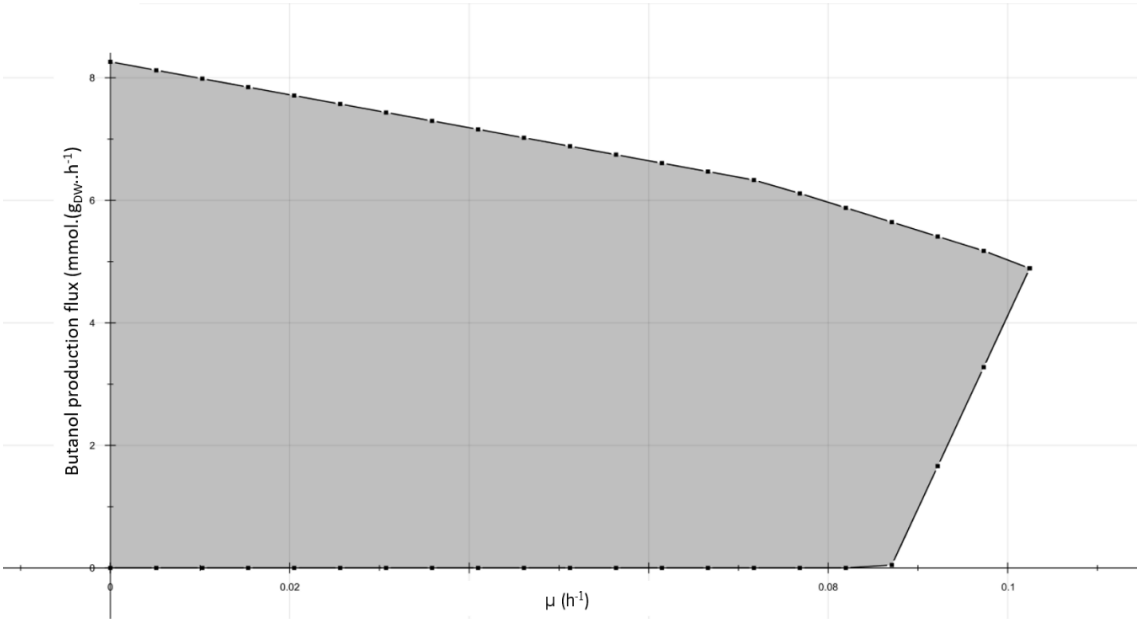


Figure 3.3. Graph created using Flux Variability analysis: butanol production flux mmol.(g_{DW}.h)$^{-1}$ for different levels (20 intervals) of growth rate (h^{-1}).

Analyzing Figure 3.3, it is possible to observe that, for the highest value of biomass formation (0.103 h^{-1}), there were no alternatives in the optimal solution space besides the butanol production. Therefore, for maximum growth, the cell is required to excrete butanol and this production is robust. However, for lower growth rates, the butanol production can drop and at growth rates below 0.087 h^{-1}, the production of butanol is not the only option. Ideally, butanol production should be robust for any growth rate, but as mentioned above, for valine production, there are other native *E. coli* pathways that can recycle NADH and act as alternatives to butanol production.

The optimization process using aerobic environmental conditions (also with glucose uptake equal to 10 mmol.(g_{DW}.h)$^{-1}$) was performed in the same fashion as under anaerobic conditions. The obtained solutions had a much larger number of knock-outs and the simplest solution found had 15 knock-outs. In Table 3.5, the growth rate and the main products excreted are presented for the reference strain (simulation 5) and for the best *in silico* strain design of the optimization process (simulation 6).

Table 3.5: Simulation results using FBA for the wild-type strain and the best mutant strain obtained with the optimization process in Optflux, with aerobic environmental conditions and glucose uptake of 10 mmol.(g_{DW}.h)$^{-1}$. For each simulation, the genetic manipulations are shown, the active reactions from the added heterologous pathway, the biomass growth μ (h^{-1}) and the main products excreted (mmol.(g_{DW}.h)$^{-1}$) are shown.

Simulation	Genetic manipulations		Active reactions for the production of butanol	μ (h^{-1})	Main products excreted (mmol/g_{DW}.h)	
5	-		.	1.07	CO_2	20.62
					H_2O	47.87
					1-Butanol	-
6	R_HACD2=0 R_CTECOAI7=0 R_CTECOAI6=0 R_ME2=0 R_RPI=0 R_SUCD=0 R_FESR=0	R_ALDD4=0 R_AACPS1=0 R_3OAR120=0 R_AACPS3=0 R_FTHFD=0 R_CPPPGO=0 R_FE3Ri=0	R03534 **($FADH_2$)** R04000 R03937 R03028 R_ACOAD1f **($FADH_2$)** R01172 **(NADH)** R03544 **(NADH)**	0.25	CO_2	22.82
					Kdo2-LipidA (KLA)	0.28
					H_2O	31.8
					1-Butanol	1.14

Comparing simulation 5, in Table 3.5, with simulation 1, in Table 3.4, (both without any inactive reaction, differing only in the environmental conditions), biomass growth had a 4-fold increment from 0.25 to 1.05 h^{-1}. Furthermore, since oxygen availability implies that NADH recycling is not necessary at the product level, there are no fermentation products excreted in simulation 5.

In simulation 6, the inactivation of 15 reactions led to a 4.16-fold reduction in biomass growth (from 1.05 to 0.25 h^{-1}) achieving the same value as the one obtained using anaerobic conditions without any knock-outs (simulation 1). This mutant excreted 1.14 mmol.(g_{DW}.h)$^{-1}$ of butanol using two reactions with $FADH_2$ as cofactor: one non-native reaction converting 2-oxoglutarate into 2-hydroxyglutarate (R03534) and one native reaction converting crotonyl-CoA into butanoyl-CoA (R_ACOAD1f).

From the six mutants simulated (Table 3.4 and Table 3.5), only simulations 2, 3 and 6 showed butanol production. Simulation 3 had the best performance, considering biomass growth, minimal number of inactivated reactions and butanol flux. In Figure 3.4, a schematic representation of the genetic alterations proposed for this strain design is shown.

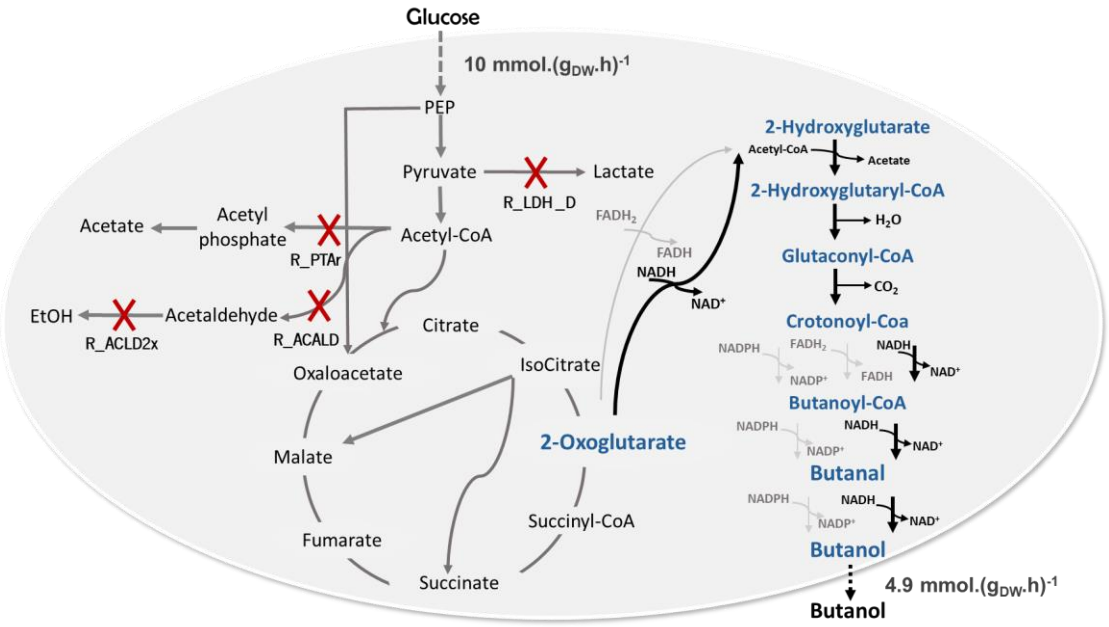


Figure 3.4: Schematic representation of the mutant from simulation 3, the best optimization result. Active heterologous reactions are in black arrows; inactive heterologous reactions are in light grey, red cross marks indicate knocked-out reactions. PEP: phosphoenolpyruvate;

3.3.4 Enzyme selection

After analyzing all the strain optimization solutions, the mutant detailed in simulation 3 had the best performance, providing a compromise between number of knock-outs, biomass growth and butanol production. Nevertheless, to apply this solution *in vivo*, it is essential to have curated gene sequences coding for the enzymes that need to be expressed in *E. coli*. If gene sequences are available, the literature must be reviewed to select the best option, preferably an enzyme with experimental validation in the selected host, *E. coli*.

The first four reactions of the proposed pathway in Figure 3.2, are part of the glutamate fermentation via 2-hydroxyglutarate pathway. This pathway, firstly described in 1974 (Buckel and Barker, 1974), is present in different bacteria, the most well-known is the strictly anaerobic gut bacterium *Acidaminococcus fermentans* (Buckel and Barker, 1974).

The first three steps corresponding to the conversion of 2-oxoglutarate into glutaconyl-CoA were already successfully expressed to produce glutaconate in *E. coli* (Djurdjevic et al., 2011). In the mentioned work, the first reaction is the reduction of 2-oxoglutarate to 2-hydroxyglutarate by 2-hydroxyglutarate dehydrogenase (*hgdH*) (Martins et al., 2005), followed by the CoA transfer from

acetyl-CoA to 2-hydroxyglutarate, forming 2-hydroxyglutaryl-CoA and releasing acetate, catalyzed by glutaconate-CoA transferase *gctAB* (Buckel et al., 1981). Lastly, 2-Hydroxyglutaryl-CoA dehydratase from *Clostridium symbiosum* with respective activator (*hgdC*) from *A. fermentans*, converts 2-hydroxyglutaryl-CoA in (E)-glutaconyl-CoA, by *syn*-elimination of water (Hans et al., 1999).

The decarboxylation of glutaconyl-CoA into crotonyl-CoA is intermediated by a transmembranal protein constituted by four subunits, hindering its expression in *E. coli*. This enzyme works as an ion pump, conserving free energy. Although the actual enzymatic activity is catalyzed solely by the α-subunit, its single expression greatly reduces its activity (Bendrat and Buckel, 1993). On the other side, glutaryl-CoA carboxylase also catalyzes the same reaction but using glutaryl-CoA as the initial substrate and glutaconyl-CoA as an intermediate. This reaction is present in microorganisms such as *Pseudomonas aeruginosa PAO1* (Indurthi et al., 2018).

The last steps of the pathway from crotonyl-CoA to butanol are common to clostridial pathway. For this reason, different enzymes catalyzing these reactions have been tested in *E. coli*. The reduction of crotonyl-CoA to butyryl-CoA is catalyzed by butyryl-CoA dehydrogenase (gene *bcd*) requiring an electron transfer flavoprotein (*etfAB*). This mechanism appears to be associated with native clostridial ferredoxins, hindering its efficient expression in *E. coli* (Dong et al., 2016). A crotonyl-CoA reductase (*ccr*) from *Streptomyces coelicolor*, has been tested but impaired butanol production (Atsumi et al., 2008a). *Trans*-enoyl-CoA reductase (*ter*) irreversibly catalyzes the reduction of crotony-CoA using directly NADH as a reducing equivalent instead of ferredoxins or flavoproteins. Several *ter* from various sources (*Treponema denticola*, *Treponema vincentii*, *Flavobacterium johnsoniae* and *Fibrobacter succinogenes*) were also tested to catalyze this step. *Ter* from *Treponema denticola* had the best increment in butanol titer (Shen et al., 2011).

In clostridia species, the last two steps are converted by a bifunctional enzyme, with both aldehyde and alcohol dehydrogenase activity encoded by *adhE*1 or *adhE*2. The last enzyme, active during alcohol production phase, had more activity when expressed in *E. coli* (Dong et al., 2016; Inui et al., 2008)

In Table 3.6, for each one of the steps forming the proposed novel pathway, the enzyme name, the selected encoding gene and NCBI accession numbers, as well the source microorganisms are shown.

Table 3.6: NCBI accession number and respective microorganism proposed for each reaction from the new butanol synthesis pathway.

Reaction	Enzyme	Gene	Reference	Microorganism
R08198	2-Hydroxyglutarate dehydrogenase	*hgdH*	EC 1.1.99.2 **NCBI GI:** >gb|CP001859.1|:1104274-1105269 **NCBI GeneID:** Acfer_0977	*Acidaminococcus fermentans* ATCC 25085
R04000	Glutaconate-CoA transferase	*gctAB*	gctA EC 2.8.3.12 **NCBI GI** >>gi|284047386:2019003-2019965 **NCBI** GeneID: Acfer_1820 gctB **NCBI GI:** >gi|284047386:2018200-2019000 **NCBI geneID:** Acfer_1819	*Acidaminococcus fermentans* ATCC 25085
R03937	(R)-2-hydroxyglutaryl-CoA dehydrogenase subunits A and B	*hgdAB*	EC 4.2.1.- **NCBI GI:** > AF123384.1:1241-2683 **NCBI geneID:** AF123384	*Clostridium symbiosum* ATCC 14940
	(R)-2-hydroxyglutaryl-CoA dehydrogenase Subunit C	*hgdC*	EC 4.2.1.- **NCBI GI:** >gi|284047386:2015608-2016390 **NCBI GeneID:** Acfer_0168	*Acidaminococcus fermentans* ATCC 25085
R03028	Glutaryl-CoA dehydrogenase	*gcdH*	EC 1.3.8.6 **NCBI GI:** >NP_249138.1 **NCBI GeneID:** PANN_05040	*Pseudomonas aeruginosa* PAO1
	Glutaconyl-CoA decarboxylase	*gcdA*	EC 4.1.1.70 **NCBI GI:** FJ390113.1: 740-2506 **NCBI GeneID:** ACJ24327.1	*Clostridium symbiosum* ATCC 25085
R01171	*trans*-enoyl-CoA reductase	*ter*	EC 1.3.1.44 **NCBI GI:** >AE017226: 636109-637302 **NCBI GeneID:** TDE_0597	*Treponema denticola.* ATCC 35405
R01172 & R03544	Bifunctional Aldehyde / Alcohol dehydrogenase	*adhE*	EC 1.1.1.11 / 1.2.1.3 **NCBI GI:** >CP002661.1:33722-36298 **NCBI GeneID:** adhe ATCC ID: DSM 1731	*Clostridium acetobutilicum* pSMBa - DSM 1731

3.4 Discussion

The main objective of this study is to rationally design a new microbial cell factory able to produce butanol with a novel pathway, generated by a hyper-graph algorithm. Furthermore, the exhaustive analysis carried using several *in silico* tools aims to save time in the lab by avoiding a trial and error experimental approach.

From the initial set of 105,954 different pathways, only 40,608 were stoichiometrically feasible. The successive filters applied (based on size, butanol flux and novelty) allowed the identification of a set

of 24 pathways. These pathways catalyze, with some alternative reactions, the conversion of 2-oxoglutarate into n-butanol.

The change in free Gibbs energy ($\Delta_r G'^m$) is a good indicator of reaction directionality and it should be negative for a biochemical reaction to occur in the set direction (Bar-Even et al., 2012). The first step of the clostridial butanol pathway catalyzed by a thiolase – the condensation of two molecules of acetyl-CoA into one of acetoacetyl-CoA – is thermodynamically unfavorable (Gheshlaghi et al., 2009) with a $\Delta_r G'^m$ equal to 26.1 ± 1.7 kJ.mol^{-1} and $K'_{eq} = 2.6\times10^{-5}$. Several strategies have been proposed to overcome this issue when expressing clostridial pathway in *E. coli*, such as increasing acetyl-CoA pool or testing the expression of alternative thiolases (Shen et al., 2011). In comparison, the estimation of $\Delta_r G'^m$ (Table 3.3) using eQuilibrator (Flamholz et al., 2012) suggests that the novel pathways described here can occur through the desired direction, forming butanol. The first reaction of the proposed pathway (R08198) had a $\Delta_r G'^m$ of -22.6±3.6 kJ.mol^{-1} and an Equilibrium constant (K'_{eq}) of 9.2×10^{3}. Therefore, no issues are expected regarding reaction reversibility for this step. 2-oxoglutarate is an intermediate in TCA cycle and acts as a carbon skeleton for nitrogen-assimilatory reactions, being the intersection between the two major nutrients needed by microorganisms: carbon and nitrogen (Commichau et al., 2006). Although the regulatory events associated with 2-oxoglutarate have not been totally disclosed, experimental data show its concentration varies according to carbon and nitrogen availability (Huergo and Dixon, 2015). There are examples of successfully heterologous pathways expressed in *E. coli* using 2-oxoglutarate as substrate, for the production of glutarate (Yu et al., 2017) and glutaconate (Djurdjevic et al., 2011). Considering the remaining reactions of the heterologous pathway, we predict no issues regarding thermodynamics. All the reactions constituting the pathway had a negative value of $\Delta_r G'^m$, with exception of R03937. This reaction is probably reversible and we expect that the subsequent reaction ensures its flux.

From all the simulations performed using FBA, the ones obtained using anaerobic conditions (Table 3.4) had a better performance than the ones obtained using aerobic conditions (Table 3.5). First, the simplest solution using aerobic conditions had a greater number of inactivated reactions than the design shown in simulation 3 for anaerobiosis (15 against 5). Also, the maximum production of butanol in aerobic conditions was 1.14 mmol.(g_{DW}.h)$^{-1}$, while under anaerobic conditions it was 4.94 mmol.(g_{DW}.h)$^{-1}$.

This is supported by the observation that anaerobic conditions favor alcohols production (Shen et al., 2011). During fermentation, the conversion of glucose to other products is accompanied by electrons being transferred to endogenous acceptors (such as NAD^+) forming the reducing equivalents (like NADH) (Trinh et al., 2011). The NADH molecules are then recycled through the synthesis of mixed-acid fermentation products such as ethanol, lactate and formate. While in aerobic conditions, directly regeneration of NADH by oxidation is metabolically more advantageous to the cell, since ATP is produced (Trinh et al., 2011). Overall, the best strain design found is shown in simulation 3 because it presents the best compromise between number of inactivated reactions, flux of butanol and biomass growth. In the proposed pathway in simulation 3, for one molecule of butanol produced, four molecules of NADH are recycled into NAD^+. So, it makes sense that butanol production is more favorable under anaerobic conditions (when NADH cannot be directly recycled using oxidative phosphorylation). Moreover, the knock-outs suggested by the optimization process, R_ALCD2x; R_ACALD; R_PTAr; and R_LDH_D, are the ones catalyzing the production of ethanol, acetate and lactate. These reactions are the ones natively used by *E. coli* to recycle NADH during fermentation. So, to force the flux towards the heterologous pathway, the native alternatives must be inactive. This strategy creates a source of reducing power that must be recycled through the heterologous pathway, i.e. NADH pool serves as a driving force to produce butanol. Also, in simulation 3 the possibility of excreting butanal instead of butanol was evaluated by adding its drain to the GSMM. Under those conditions, only butanol was secreted and the flux distribution was maintained unchanged. Indicating that butanol excretion is more advantageous to the microorganism in this phenotype. A similar strategy was already applied in the expression of *Clostridium acetobutylicum* pathway in *E. coli* (Shen et al., 2011).

In fact, the redox balance seems to be extremely important in the phenotype of the mutant of simulation 3. This was demonstrated by the totally different flux distribution in simulation 4, as the effect of changing a cofactor in a single reaction. When expressing clostridial butanol pathway in *E. coli*, besides knocking-out competing pathways, different approaches have been applied to increase the cofactors pool and, consequently, butanol production. For instance medium supplementation with formate and expression of *fdh1*, a formate dehydrogenase from *S. cerevisiae* that converts formate to CO_2 releasing one molecule of NADH (Nielsen et al., 2009).

Additionally, the first heterologous reaction from the butanol pathway used in simulation 3 (R08198) requires NADH as cofactor, thermodynamically more favorable than the alternative R03534 using

$FADH_2$ (Table 3.3). Since the change on the free Gibbs energy in reaction R08198 is 4-fold lower than the reaction R03534.

The robustness of the best strain design (see simulation 3) was also analyzed. Since FBA provides a unique "optimal" solution, alternative flux distributions may exist and remain unnoticed (Maia et al., 2016). FVA allows to check the solution space by creating a graph correlating the maximum/minimum butanol yield for different levels of biomass growth. Although the FVA graph revealed that butanol production was robust in simulation 3, it showed that for a growth rate below 0.087 h^{-1} butanol excretion is not essential for the cell. Further deletions in alternative NADH recycling reactions should improve the robustness of this strain design but would likely result in even lower growth rates and increase work for the *in vivo* implementation of this design.

One of the requirements to force butanol production through the heterologous pathway was the inactivation of ECOAH1. Although this inactivation was critical *in silico* to redirect the flux trough the desired pathway, it seems this would be unnecessary *in vivo*. After analyzing the literature, this reaction is only active under certain environmental conditions such as the presence in the medium of long-chain fatty acids (Campbell, Morgan-Kiss, & Cronan, 2003). Moreover, in other studies expressing the butanol pathway from *C. acetobutylicum* in *E. coli* (Garza et al., 2012; Inui et al., 2008; Nielsen et al., 2009), the enzymes responsible for the conversion of acetoacetyl-CoA into crotonyl-CoA are also expressed, indicating these reactions are inactive in *E. coli*.

After analyzing the literature, it was possible to conclude that all the catalytic steps constituting the proposed pathway have gene sequences available (Table 3.6) and experimental validation in *E. coli*, although in other contexts and for the production of different products.

The best solution of the optimization process (simulation 3) was supported by a profound *in silico* analysis alongside manual curation. The analysis of the simulation results; the availability of curated gene sequences encoding the respective enzymes and experimental validation in the selected host reinforce the potential of this novel pathway to produce butanol in *E. coli*.

The main goal here was to understand the predicted phenotype before experimental work. Consequently, with this analysis it is expected to develop *in vivo* a microbial cell factory able to produce butanol by 2-oxoglutarate in an unprecedent route.

CHAPTER 4 *In vivo* implementation of a novel pathway to produce n-butanol in *Escherichia coli*

The growing environmental awareness has motivated the search for more sustainable processes, in fact one of the European Union directives indicates that all the chemical processes should be shifted to biological processes and, particularly, 10 % of all fuels must be bio-synthesized by 2020. In this regard, butanol poses as one of the most promising alternative fuel, presenting more suitable properties for the existing oil-infrastructure than ethanol (leading biofuel). The search for novel pathways, alternatives to the clostridial pathway, to express in a more robust microorganism – such as *Escherichia coli* – is one of the metabolic engineering-driven strategies for efficient butanol production.

In this work, we developed novel strains of *E. coli* able to produce butanol through a novel pathway using 2-oxoglutarate as precursor. The implementation of this pathway, discovered using a hyper-graph algorithm (Chapter 3), involved seven catalytic steps and required the insertion of nine heterologous genes in *E. coli* distributed in three plasmids. The engineered strains were cultivated in HDM medium, the heterologous genes were induced with 0.5 mM of isopropyl 1-thio-β-D-galactopyranoside (IPTG) and after 12h transferred to anaerobic conditions. The maximum butanol titer obtained was 128.96±7.74 $mg.L^{-1}$ in strains expressing *gcdH* for the decarboxylation of glutaconyl-CoA.

4.1 Introduction

The limited amount of natural resources and the consequent endeavor for more sustainable processes remain as one of the major challenges nowadays, stimulating the interest towards Industrial Biotechnology. In this regard, butanol poses as a promising alternative to ethanol due to its superior properties as a fuel, such as higher energy content, less corrosiveness and higher blending capacity with gasoline (Koppolu and Vasigala, 2016).

Contrary to ethanol, which is naturally produced by diverse microorganisms, butanol is natively produced together with ethanol and acetone through Acetone-Butanol-Ethanol (ABE) fermentation by *Clostridium* species. The ABE fermentation consists in the conversion of carbohydrates into acetone-butanol-ethanol in a ratio of 3:6:1, comprising two stages. The first, named as acidogenic phase, is characterized by exponential bacterial growth along with the production of acetic and butyric acids. These acids are excreted, lowering the external pH. Next, growth rate decreases and bacteria start the solventogenic phase, partly re-assimilating the previously produced acids, which work as substrates for the final solvent production (Amador-Noguez et al., 2011; Gheshlaghi et al., 2009).

Although ABE fermentation has been studied and widely used for a long time, some challenges remain nowadays. These include slow growth rates, since clostridial strains are strictly anaerobic; low butanol yield due to its toxicity; spore-forming capacity with possible loss of butanol production ability; difficulties for continuous culture due to phenotypic instabilities; high costs of substrates and product recovery process (Dürre, 2008; Garza et al., 2012; Tashiro et al., 2013).

In this regard, the expression of *Clostridia* butanol pathway using the several genetic tools available in a more robust microbial host, such as *Escherichia coli* (Inui et al., 2008) or *Saccharomyces cerevisiae* (Steen et al., 2008), was successfully implemented. Nevertheless, some challenges have remained, namely the first reaction of the clostridia pathway – the condensation of two molecules of acetyl-CoA into one molecule of acetoacetyl-CoA – being thermodynamically unfavorable (Shen et al., 2011). Also, some enzymes of this pathway appear to be poorly expressed in *E. coli*, such as BCD, the enzyme that catalyzes the reduction of butanoyl-CoA into crotonoyl-CoA (Zheng et al., 2009). Finally, the high requirement of reducing power and the consequent competition with native fermentation pathways (Garza et al., 2012) have also to be taken into consideration when optimizing butanol production in heterologous hosts. Several rational strategies were performed in *E. coli* to

increase butanol productivity, especially metabolic engineering approaches such as manipulation of central carbon metabolism, elimination of competing pathways, gene overexpression, cofactor balancing, expression of novel enzymes and consumption of different substrates (H. Chan et al., 2013). An alternative strategy to tackle some of the issues pointed above is to identify non-natural butanol biosynthetic pathways more suitable for expression in standard microbial production hosts.

Some approaches to identify novel pathways have been reported in the literature, specifically for the production of butanol. Maranas group developed a study combining a hyper-graph algorithm with an optimization-based approach resulted in the identification of novel pathways to produce n-butanol in *E. coli*. Besides the discovery of pathways already explored for butanol production, one novel pathway through thiobutanoate intermediate was pinpointed (Ranganathan and Maranas, 2010). Another *in silico* approach combining a graph-method algorithm alongside with structural analysis based on docking, allowed to identify novel routes to produce butanol from pyruvate in *Clostridum* strains (Wu et al., 2011). Experimentally, Liao's group reported the exploration of the 2-keto-acid pathway to produce biofuels(Shen and Liao, 2008).

In Chapter 3, several heterologous pathways to produce butanol in *E. coli* generated in our group were subjected to an exhaustive *in silico* analysis, combining computational tools with manual curation. As a result, a novel biosynthetic pathway converting 2-oxoglutarate into butanol was considered the most promising route to apply *in vivo*. As illustrated in Figure 4.1, this route is composed by seven steps converting successively 2-oxoglutarate into n-butanol. Instead of the commonly used precursor acetyl-CoA, in this pathway butanol is synthesized from 2-oxoglutarate, the keto-acid formed by deamination of glutamate and an intermediate in TCA cycle. This first reaction is thermodynamically very favorable with a $\Delta_r G'^m$ of -22.6±3.6 KJ.mol^{-1}, hence, overcoming the thermodynamic issues identified by expressing the native clostridium butanol biosynthetic pathway. To the best of our knowledge, this 2-oxoglutarate pathway was not exploited before and or discovered in the study conducted by Maranas group (Ranganathan and Maranas, 2010).

For each of the steps composing the novel pathway, enzymes were selected considering curated gene sequences and the experimental data available in the literature, in order to express the most suitable enzyme for each catalytic step. The resulting set of genes catalyzing the successive reactions from 2-oxoglutarate to butanol originates from diverse microorganisms (namely *Acidaminococus fermentans*, *Clostridium symbiosum*, *Pseudomonas aeruginosa* PAO1, *Treponema denticola* and *Clostridium acetobutylicum*).

The goal of this study is to implement *in vivo* the selected butanol pathway from the *in silico* analysis in Chapter 3, constructing new strains of *E. coli* able to produce butanol through this novel pathway.

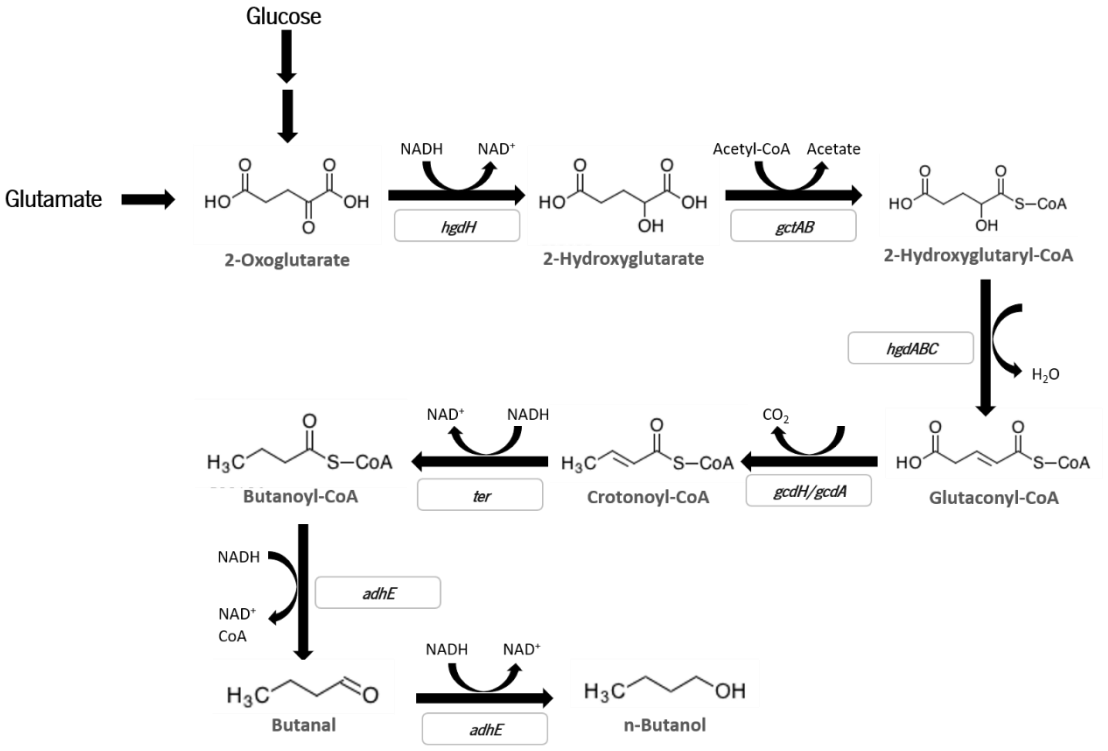


Figure 4.1: Proposed n-butanol biosynthetic pathway in E. coli *with indication of the genes encoding the enzymes catalyzing the different reactions. hgdH – 2-oxoglutarate reductase reductase; gctAB – glutaconate-CoA transferase; hgdABC - 2-hydroxyglutaryl-CoA dehydratase; gcdH – glutary-CoA dehydrogenase; gcdA – glutaconyl-CoA decarboxylase; ter* - trans-*enoyl-CoA reductase, adhE - aldehyde dehydrogenase/alcohol dehydrogenase*

4.2 Materials and methods

4.2.1 Cloning procedure

E. coli NEB 5-alpha cells were used for gene cloning and vector propagation. These strains were cultured in Luria-Bertani (LB) medium (10 $g.L^{-1}$ of peptone; 5 $g.L^{-1}$ yeast extract and 5 $g.L^{-1}$ of NaCl) with the appropriate antibiotics concentration. The solid version of this medium included 15 $g.L^{-1}$ agar. All cultivations were performed at 37 °C and, in the case of liquid cultures, under shaking conditions (200 rpm).

For long-term storage, glycerol was added to a final concentration of 30 % (v/v) to overnight cultures in selective media and kept in a -80 °C freezer.

The genes used in this study were amplified by polymerase chain reaction (PCR) using Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Waltham, USA) in a LifeECO Thermal Cycler (Bioer Technology, Zhejiang, China). All primers were purchased from Metabion (Munich, Germany). DNA fragments were purified using DNA Clean and Concentrator DNA Kit (Zymo Research, Irvine, USA).

Plasmids were extracted using Plasmid Miniprep kit (Zymo Research). All digestions were performed using the appropriate FastDigest® restriction endonucleases (Thermo Scientific). Ligations were performed with T4 DNA Ligase (Thermo Scientific) and transformed by heat-shock in chemically competent *E. coli* NEB 5-alpha cells (New England BioLabs, Massachusetts, USA). The success of ligation was checked through Colony PCR using DreamTaq (Thermo Scientific) and further confirmed by sequencing (StabVida, Lisbon, Portugal). Protocols were performed in accordance with manufacturer's instructions.

hgdH, *gcdH*, *hgdABC* and *gctAB* genes were codon-optimized through ATGenium for *E. coli*, synthesized and cloned in pHTP0 vector by NZYTech (Lisbon, Portugal). *gcdA-CS* was amplified from *Clostridium symbiosum* ATCC 14940 genomic DNA obtained from DSMZ (Braunschweig, Germany).

4.2.2 Plasmid construction

Compatible vectors pETDuet pCDFDuet, pCOLADuet and pRSFDuet (Novagen, Darmstadt, Germany) were used to provide individual expression of each protein under the control of the T7lac promoter and a ribosome binding site (RBS).

The plasmid pCDFDuet (Novagen), with spectinomycin resistance marker, was used to clone the codon-optimized genes encoding the first two reactions of the proposed pathway (*gctAB* and *hgdH*.). *hgdH* was amplified using the primers *hgdH_fw* and *hgdH_rev* (primers are shown in Table 4.1) with flanking restriction sites for *Kpn*I and *Xho*I and cloned into pCDFDuet. The PCR product for *gctAB*, amplified using primer *gctAB_fw* and *gctAB_rev*, was restricted and ligated into *BamH*I and *Hind*III restriction sites of the previous construction originating pCDFDuet_hgdH_gctAB. Colony PCR with appropriate primers was used to find successful clones and the final plasmid was sent for sequencing to confirm if the sequence was correct.

The plasmid pRSFDuet (Novagen), containing a kanamycin resistance gene, was used to clone the codon optimized genes *hgdABC* and *gcdH*, corresponding to the two intermediate steps of the proposed pathway. *gcdH* was amplified using primers *gcdH_fw* and *gcdH_rev* with restriction sites to *Nde*I and *Xho*I and cloned in pRSFDuet. Then, *hgdABC* was inserted in the previous construction yielding pRSFDuet_hgdABC_gcdH. This gene was amplified using primers *hgdABC_fw* and *hgdABC_rev* with restriction sites for *Sac*I and *Not*1, respectively. Colony PCR with appropriate primers was used to find successful clones and the final plasmid was sent for sequencing to confirm if the sequence was correct.

An alternative gene (*gcdA*) to convert glutaconyl-CoA into crotonyl-CoA was also cloned. This gene encodes the α-subunit of a glutaconyl-CoA carboxylase and was amplified from genomic DNA of *Clostridium symbiosum* ATCC 14940 using the primers *gcdA-CS_fw* and *gcdA-CS_rev*. The plasmid pCOLADuet (Novagen), expressing a kanamycin resistance marker, was used to clone *gcdA* by restricting and ligating between the sites *Nde*I and *Xho*I. This construction was then digested using *Sac*I and *Not*1, restriction sites, to clone the *hgdABC* operon amplified as detailed above, resulting in the plasmid pCOLADuet_hgdABC_gcdA. Colony PCR with appropriate primers was used to find successful clones and the final plasmid was sent for sequencing.

The plasmid pETDuet (Novagen), ampicillin-resistant, was used to clone the genes *adhE* and *ter*, corresponding to the last two genes of the proposed pathway. The *adhE* gene was amplified from template plasmid pmTA1 (Nielsen et al., 2009) using primers *adhE_fw* and *adhE_rev* with restriction sites for *Eco*RI and *Not*1, respectively. The synthetic gene *ter* (ATG:biosynthesis, Freiburg, Germany) was amplified using primers *ter_fw* and *ter_rev*, restricted and ligated into *Nde*I and *Xho*I restriction sites of the previous construction pETDuet_adhe, resulting in the plasmid pETDuet_adhE_ter. Colony PCR with appropriate primers was used to find successful clones and the final plasmid was sent for sequencing to confirm if the sequence was correct.

In Table 4.1 the primers used in this study for PCR amplification are shown.

*Table 4.1: Sequences of primers used in the cloning procedures of this study (*restriction sites are underlined). fw-forward; rev - reverse*

Primer	Sequence	Restriction Sites *
adhE_fw	CCGAATTCATGAAAGTCACAACAGTAAAGG	*EcoR*I
adhE_rev	CCGCGGCCGCTTAAGGTTGTTTTTTAAAACAATT	*Not*1

ter_fw	CCCATATGATTGTAAAACC	*Nde*I
ter_rev	CCCTCGAGTTAAATC	*Xho*I
hgdABC_fw	CCGAGCTCATGAGTATCTATACCCTGGGC	*Sac*I
hgdABC_rev	CCGCGGCCGCTTATTTTTGCATCTCCAAAAC	*Not*I
gcdH_fw	CCCATATGGCAACCAAAGCAAG	*Nde*I
gcdH_rev	CCCTCGAGTCAAAAGAACGCTTGAATACC	*Xho*I
hgdH_fw	CCGGTACCATGAAAGTGCTGTGCTACGG	*Kpn*I
hgdH_rev	CCCTCGAGTTATTTGATTTTGTTCGGGC	*Xho*I
gctAB_fw	CCGGATCCATGAGCAAAGTCATGACCC	*BamH*I
gctAB_rev	CCAAGCTTTTATTTGGCTTCAGTTGGAAC	*Hind*III
gcdA-CS_fw	CCCATATGAATATGTATTCAATGCCAGGAT	*Nde*I
gcdA-CS_rev	CCCTCGAGTTATTTACCAAATGTCTCGAATTCACG	*Xho*I

The success of the plasmid constructions was confirmed by sequencing the regions of interest with the appropriate primers. In Figure 4.2, the plasmids used or constructed in this study, as well as the respective major features are depicted.

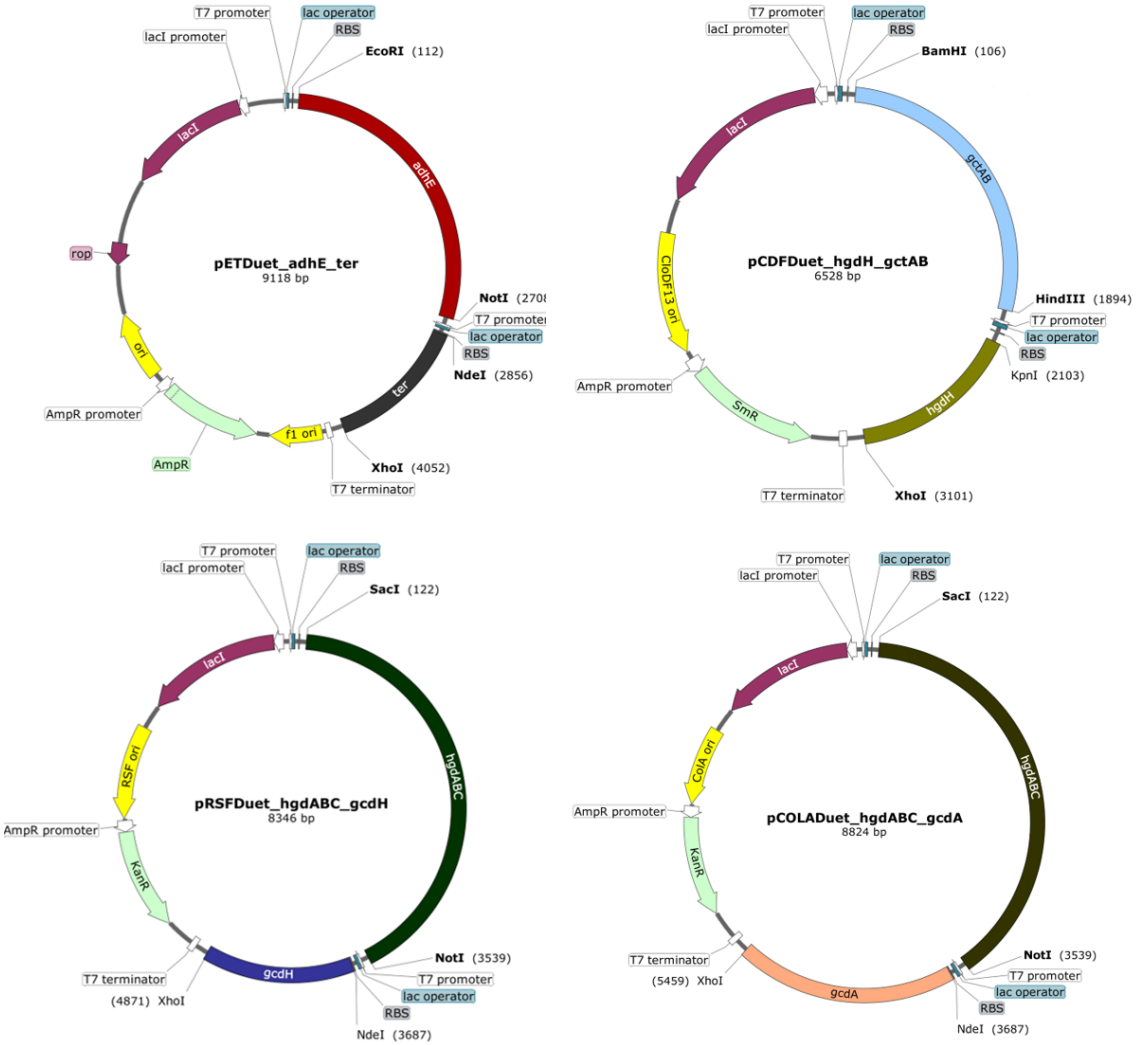


Figure 4.2: Schematic representation of the plasmid maps of all the cloning constructions developed, with indication of the main features. The restriction sites used to clone the genes are indicated.

4.2.3 Bacterial strains

E. coli K12 MG1655 (DE3) and *E. coli* BL21 (DE3) were used as hosts for gene expression under control of the T7 promotor.

BUT_OXG1 and BUT_OXG2 strains were obtained by transforming *E. coli* BL21 (DE3) and *E. coli* K12 MG1655 (DE3), respectively, with both pCDFDuet_gctAB_hgdH; pRSFDuet_hgdABC_gcdH and pETDuet_adhE_ter by electroporation using 0.1 cm-gap electroporation cuvettes at a voltage of 1.8 KV. BUT_OXG1_GCDA and BUT_OXG2_GCDA strains were obtained by transforming *E. coli* BL21 (DE3) and *E. coli* K12 MG1655 (DE3), respectively, with pCDFDuet_gctAB_hgdH; pCOLADuet_hgdABC_gcdA and pETDuet_adhE_ter by electroporation. Electrocompetent cells were

prepared using the protocol developed by (Dower et al., 1988). Positive transformants were isolated in LB (containing 10 $g.L^{-1}$ of peptone; 5 $g.L^{-1}$ yeast extract and 5 $g.L^{-1}$ of NaCl) agar (15 $g.L^{-1}$) plates, containing the appropriate antibiotic concentrations (50 $\mu g.mL^{-1}$ ampicillin, 50 $\mu g.mL^{-1}$ spectinomycin and 30 $\mu g.mL^{-1}$ kanamycin) and incubated at 37 °C, overnight. To confirm the success of the transformation, a few transformant colonies were cultivated in LB medium with antibiotics, overnight. After, plasmids were extracted and digested with appropriate restriction enzymes. The correct fragment lengths were confirmed by running the digestion in a 1 % (w/v) agarose gel.

Table 4.2 summarizes the strains of *E. coli* used or engineered for this study.

Table 4.2: List of strains and genomic DNA used or engineered for this study.

Strains	Relevant genotype	Source
***E. coli* K12 MG1655 (DE3)**	F - λ - ilvGrfb- 50 rph- 1 λ(DE3)	(Nielsen et al., 2009)
***E. coli* BL21 (DE3)**	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5	New England Labs
BUT_OXG1	*E. coli* BL21 DE3 pETDuet_adhE_ter; pCDFDuet_gctAB_hgdH; pRSFDuet_gcdH_hgdABC	This study
BUT_OXG2	*E. coli* K12 MG1655 DE3 pETDuet_adhE_ter; pCDFDuet_gctAB_hgdH; pRSFDuet_gcdH_hgdABC	This study
BUT_OXG1_gcdA	*E. coli* BL21 DE3 pETDuet_adhE_ter; pCDFDuet_gctAB_hgdH; pRSFDuet_gcdA_hgdABC	This study
BUT_OXG2_gcdA	*E. coli* K12 MG1655 DE3 pETDuet_adhE_ter; pCDFDuet_gctAB_hgdH; pCOLADuet_gcdA_hgdABC	This study
Control_OXG1	*E. coli* BL21 DE3 pETDuet_adhE_ter; ; pRSFDuet_gcdH_hgdABC	This study
Control_OXG2	*E. coli* K12 MG1655 DE3 pETDuet_adhE_ter; pRSFDuet_gcdH_hgdABC	This study

4.2.4 Butanol production experiments in complex medium

The strains BUT_OXG1 and BUT_OXG2 were cultivated in Terrific Broth (TB) medium supplemented with glucose, glutamate, riboflavin and iron (III) citrate according to composition shown in Table 4.3. The pH of this medium was 7.2 ± 0.2 at 25°C.

Table 4.3: Medium composition of Terrific Broth.

Component	Amount per Liter	Unit
Tryptone	12	g
Yeast extract	24	g
Glycerol	4	mL
Monobasic potassium phosphate	2.31	g
Dibasic potassium phosphate	12.54	g
Glutamate	0.468	g
Riboflavin	0.07529	g
Iron (III) citrate	0.525	g
Glucose	10	g

A single colony was picked from LB plates and inoculated in 10 mL of liquid LB medium. Cultivation was performed with the addition of suitable antibiotics according to the employed plasmids (50 $\mu g.mL^{-1}$ ampicillin, 50 $\mu g.mL^{-1}$ spectinomycin, and 30 $\mu g.mL^{-1}$ kanamycin). The pre-cultures were grown aerobically on a rotary shaker at 37 °C and 200 rpm, overnight.

500 mL shake flasks with 100 mL of TB medium, containing appropriate antibiotics, were inoculated with pre-cultures to obtain an initial optical density OD_{600} of 0.1. Cultivation was carried on a rotary shaker at 200 rpm at 37 °C. The butanol production genes were induced by the addition of 0.1, 0.5 or 1 mM IPTG to the culture medium when an optical density OD_{600} of 0.4-0.5 was reached.

To promote butanol production, after induction, the cells were switched to anaerobic conditions by transferring 60 mL of culture to 120 mL sealed serum flasks. The culture was supplemented with 600 µL of a 0.01 M stock solution of sodium bicarbonate to achieve a final concentration of 10 mM, since it reduces long lag phases in *E. coli* anaerobic growth (Hornsten, 1995)).

The cultures were incubated at 30 °C and 180 rpm, for 96 hours. Samples of broth were collected at time 0, induction time and 96 h. All the experiments were performed in triplicate and the samples were analyzed by High-Performance Liquid Chromatography (HPLC) and Gas Chromatography (GC).

4.2.5 Butanol production experiments in defined medium

The strains BUT_OXG1 and BUT_OXG2 were also cultivated in High Density Medium (HDM) adapted from (Sivashanmugam et al., 2009), supplemented with a solution of amino acids, extra glutamate, riboflavin and iron citrate (III), according to Table 4.4. The pH of the medium was adjusted to 7.1 using 2 M NaOH.

Table 4.4: Medium composition of HDM, adapted from (Sivashanmugam et al., 2009).

Component	Amount per Liter	Unit
Glucose	10	g
Dibasic sodium phosphate dihydrate	8.89	g
Monobasic potassium phosphate	6.8	g
Sodium chloride	0.58	g
Magnesium sulphate	1.35	g
Calcium chloride dihydrate	0.038	g
Ammonium chloride	1	g
Trace metals	250	µL
Vitamins BME100x	250	µL
Amino acid mix	2	g
Glutamate	0.468	g
Riboflavin	0.07529	g
Iron (III) citrate	0.525	g

The trace metals solution contained (per liter): $FeSO_4.7H_2O$ (30 mg); $ZnSO4.7H_2O$ (45 mg); $CaCl_2.2H_2O$ (45 mg); $MnCl_2.2H_2O$ (100 mg); $CoCl_2.6H_2O$ (30 mg); $CuSO_4.5H_2O$ (30 mg); $Na_2MoO_4.2H_2O$ (40 mg); H_3BO_3 (10 mg); KI (10 mg) and Na_2EDTA (1.5 g). The amino acid mix contained 1 g of adenine and 4 g of arginine, aspartate, glutamate, histidine, isoleucine, lysine, methionine, phenylalanine, serine, threonine, tryptophan, tyrosine and valine. The vitamin BME 100 x solution (Sigma Aldrich, St. Louis, MO, USA) contained (per liter): D-biotin (0.1 g); choline chloride (0.1 g); folic acid (0.1 g); myo-inositol (0.2 g); niacinamide (0.1 g); D-pantothenic acid.½Ca (0.1 g); riboflavin (0.01 g); thiamine.HCl (0.1 g) and NaCl (8.5 g).

A single colony was picked from LB plates and inoculated in 10 mL of LB medium. Cultivation was performed with the addition of suitable antibiotics according to the employed plasmids (50 $\mu g.mL^{-1}$ ampicillin, 50 $\mu g.mL^{-1}$ spectinomycin, and 30 $\mu g.mL^{-1}$ kanamycin). The pre-cultures were grown aerobically on a rotary shaker at 37 °C and 200 rpm, overnight. Cells were washed and harvested

by centrifugation (10 min at 3000 × *g*). Afterwards, an appropriate volume of pre-culture was transferred to 500 mL shake flasks with 100 mL of HDM medium, containing the appropriate antibiotics, yielding an initial OD_{600} of 0.1. This culture was cultivated on a rotary shaker at 200 rpm at 37 °C. The butanol production genes were induced with 0.1, 0.5 or 1 mM IPTG at an OD_{600} of 0.4-0.5 or, in some experiments, 0.6-0.7.

After induction, 60 mL of the culture were transferred to 120 mL sealed serum flasks to promote butanol production under anaerobic conditions. The culture was supplemented with 600 µL of a 0.01 M stock solution of sodium bicarbonate to achieve a final concentration of 10 mM. Selenium, nickel and molybdenum are part of the formate hydrogen lyase (FHL) complex, which is induced under anaerobic conditions (Soini et al., 2008). For this reason, 60 µL of a solution of extra trace metals ($NiCl_2$ (1.7 $mg.L^{-1}$); $(NH_4)_6Mo_7O_{24}$ (14.5 $mg.L^{-1}$); $4H_2O$ Na_2SeO_3 (2.4 $mg.L^{-1}$)) was supplied to the medium.

The cultures were incubated at 30 °C and 180 rpm, for 96 hours. Samples of supernatant were collected at time 0, induction time and 96 h. All the experiments were performed in triplicate and the samples were analyzed by HPLC and GC.

To infer the importance of anaerobic conditions for butanol production, in some of the experiments, using the engineered strains BUT_OXG1 and BUT_OXG2, after induction the transfer to serum bottles was delayed by 4h or 12h, as illustrated in Figure 4.3.

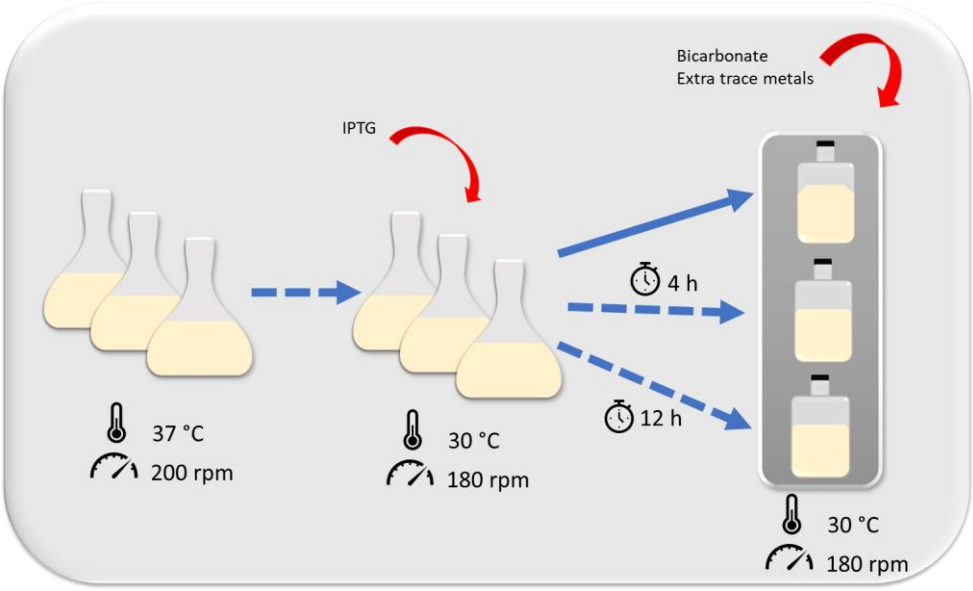


Figure 4.3: Schematic representation of the butanol-production experiments. Cells were cultured in shake-flasks until achieving mid-exponential phase, when IPTG was added to the medium and cells either immediately, after 4 or 12 h were switched to serum bottles.

4.2.6 Bioreactor cultivations

Bioreactor fermentations were performed in HDM medium. Cells were pre-grown overnight in 500 mL shake flasks containing 100 mL of the same medium, at 37 °C and 200 rpm. Each fermenter was inoculated at an initial OD_{600} of 0.15. The fermentations were performed in the Eppendorf DASGIP Parallel Bioreactor System (Switzerland) using 2-L culture vessels. The operating volume for the fermentations was 0.5 L, temperature was maintained at 37 °C, airflow at 1 VVM, pH was kept at 7.0 controlled by addition of 2 M NaOH, and dissolved oxygen was kept above 30 % of saturation by feedback control of the stirring speed from 200 rpm up to 400 rpm. Expression of the butanol genes was induced with 0.5 mM of IPTG when an OD_{600} of 0.4-0.5 was reached. After IPTG induction, temperature was decreased to 30 °C and stirring speed to 180 rpm. Anaerobic conditions were created by turning off the air flow and waiting for the leftover oxygen to be consumed. Samples were taken every two hours for the first 12 hours of the fermentation and, then, every 12 hours. At each time point the optical density was measured, and the supernatant was analyzed by HPLC and GC-FID.

4.2.7 Analytical methods

Butanol was quantified by GC and organic acids, ethanol and glucose by HPLC.

Samples were centrifuged at 6000 × g for 10 min to separate cells from the medium. Afterwards, the supernatant was filtered with a 0.22 µm pore filter membrane to glass vials and stored at -20 °C until analyzed.

Quantitative analysis of organic acids and glucose was performed using HPLC apparatus from Jasco (Japan) model LC-NetII/ADC equipped with UV-2075 Plus and RI-4030 Plus detectors, also from Jasco. The samples were analyzed using an Aminex HPX-87H column (300 mm x 7.7mm) from Bio-Rad, which was kept at 60 °C and 5 $mmol.L^{-1}$ H_2SO_4 was used as mobile phase with a flow rate of 0.5 $mL.min^{-1}$. Glucose and ethanol were detected with a refractive index (RI) detector (4030, Jasco)

and organic acids (succinate, lactate, formate and acetate) were detected at 210 nm using the UV detector. Calibration curves were obtained by injecting standards with known concentrations for each metabolite. Metabolite concentrations in samples were calculated by comparing the peak areas of the samples with the calibration curves.

Butanol concentration was quantified by a GP-9000 system (Chrompack) with a Meta-WAX capillary column (30 m X 0.25 mm X 0.25 µm) equipped with a flame ionization detector (FID), helium was used as carrier gas with a flow rate of 1 $mL.min^{-1}$. The filtered supernatant (900 µL) was mixed with 100 µL of a 5 $g.L^{-1}$ solution of isobutanol, the internal standard, yielding a final concentration of 0.5 $g.L^{-1}$, and 1 µL of this mixture was injected. The temperature of injector and detector was maintained at 250° C. The column temperature was initially at 50° C, heated to 177.5° C at a 5° $C.min^{-1}$ rate and then heated to 230° C at 10° $C.min^{-1}$, which was held for 15 minutes. A calibration curve was obtained by injecting standards with several concentrations of butanol and a fixed concentration of internal standard (0.5 $g.L^{-1}$ of isobutanol). Butanol concentration was calculated by comparing the ratio between its peak area and internal standard peak area with calibration curves.

All cell optical density measurements at 600 nm (OD_{600}) were performed using the spectrophotometer Ultrospec 10 from Biochrom (Cambridge, UK).

4.3 Results and discussion

4.3.1 Enzyme selection: decarboxylation of glutaconyl-CoA

The novel butanol production pathway depicted in Figure 4.1, includes the decarboxylation of glutaconyl-CoA into crotonyl-CoA. This step is of increased importance because it is at the intersection between the upstream reactions – already experimentally validated for the production of glutaconate in *E. coli* (Djurdjevic et al., 2011) – and the downstream reactions, validated for the production of butanol (Shen et al., 2011). Two genes were found to reportedly catalyze this enzymatic step, but neither gave us enough confidence to apply *in vivo* because since it was the only step without previously experimental validation. For this reason, two different constructions were designed to test two different enzymes for this particular step

There are two different enzymes that catalyze the decarboxylation of glutaconyl-CoA into crotonyl-CoA: one directly, glutaconyl-CoA decarboxylase, and the other, glutaryl-CoA dehydrogenase, as an intermediate step. In this study, we expressed a glutaryl-CoA dehydrogenase from *Pseudomonas aeruginosa* (Indurthi et al., 2018) and, as an alternative, the α-subunit of glutaconyl-CoA decarboxylase from *Clostridium symbiosum* (Kress et al., 2009).

Glutaryl-CoA dehydrogenase is a bifunctional enzyme that catalyzes the decarboxylation of glutaconyl-CoA into crotonyl-CoA, by using glutaryl-CoA as the initial substrate and glutaconyl-CoA as an intermediate (Blázquez et al., 2008). Enzymatic assays showed that this enzyme was able to catalyze directly the conversion of glutaconyl-CoA into crotonyl-CoA with an activity 1.2-1.6 times higher than for the first step (Ulrich et al., 1993). This enzyme was identified in cell-free extracts of *Pseudomonas sp* when grown anaerobically in aromatic compounds (Ulrich et al., 1993). A curated gene sequence encoding this enzyme was available for *Pseudomonas aeruginosa* PAO1, as well as reported expression in *E. coli* (Ulrich et al., 1993). In most microorganisms, the decarboxylation of glutaryl-CoA is catalyzed by the bifunctional glutaryl-CoA dehydrogenase, using glutaconyl-CoA as an reaction intermediate (Blázquez et al., 2008).

Glutaconyl-CoA decarboxylase is a four-subunit membrane-bound enzyme that catalyzes the conversion of glutaconyl-CoA into crotonyl-CoA, working as a biotin-dependent sodium pump (Chang et al., 2010). This enzyme is involved in the fermentation of glutamate by strictly anaerobic bacteria.

Considering that the butanol production pathway tested here requires a high number of heterologous genes, we decided to express only the α-subunit of the glutaconyl-CoA decarboxylase, which is able to catalyze the reaction as long as biotin is also supplied. Regarding the microorganism source, after reviewing the literature, we have selected *C. symbiosum* because the K_m for biotin was 14-times lower than the value obtained for *gcdA* from *A. fermentans* (Kress et al., 2009).

To test the decarboxylation of glutaconyl-CoA, the strains BUT_OXG1, BUT_OXG2, BUT_OXG1_GCDA and BUT_OXG2_GCDA were cultivated using TB and HDM medium, induced with 0.5 of IPTG and immediately switched to sealed serum bottles. The results are summarized in Table 4.5.

Table 4.5: Butanol titer (mg.L^{-1}) for strains BUT_OXG1, BUT_OXG2, BUT_OXG1_GCDA and BUT_OXG2_GCDA in TB medium and HDM medium with indication of what enzyme catalyzed the decarboxylation of glutaconyl-CoA into crotonoyl-CoA . In all experiments, strains were grown in shake-flasks until reaching 0.4-0.5 OD_{600}, 0.5 mM of IPTG was then added to the medium and cells transferred to sealed serum bottles. Data are shown as mean ± S.D from three independent experiments. n.d.: not detected

Enzyme		Strain	Butanol final titer (mg.L^{-1})	
Glutaryl-CoA dehydrogenase (*gcdH*) *Pseudomonas aerugionosa*	Glutaconyl-CoA decarboxylase (*gcdA*) *Clostridium symbiosum*		TB	HDM
+	-	BUT_OXG1	6.8±0.46	29.05±1.72
+	-	BUT_OXG2	24.5±4.6	75.32±4.21
-	+	BUT_OXG1_GCDA	n.d.	n.d.
-	+	BUT_OXG2_GCDA	n.d.	n.d.

We only detected butanol by cultivating the strains where glutaryl-CoA dehydrogenase (BUT_OXG1 and BUT_OXG2) was expressed. The method detection limit is 3 mg.L^{-1}.

Although the sole expression of the α-subunit had a V_{max} 1000 times lower than the full enzyme (Kress et al., 2009), it is not possible to affirm this enzyme is not a viable option for the proposed pathway. Several differences exist between the two constructions: *gcdH* was codon-optimized to be expressed in *E. coli*, while *gcdA* was directly cloned from genomic DNA. Moreover, the selected plasmids had different levels of copy-number: pCOLADuet possesses 20-40 copy number per cell, while pRSFDuet (the plasmid used to express *gcdH*) has a copy number higher than 100. As already mentioned, we did not consider the expression of the complete enzyme due to the high amount of heterologous DNA already expressed. In order to confirm that *gcdH* is more suitable for this pathway, both genes should be cloned using the same conditions. Nevertheless, since *gcdH* expression was enough to produce butanol, the remaining experiments were pursued only using the strains expressing this gene.

4.3.2 Butanol production in serum bottles – Effect of medium conditions

E. coli BL21 (DE3) and K12 (DE3) were genetically modified to express heterologous enzymes capable of catalyzing the successive reactions from 2-oxoglutarate to butanol originating the strains BUT_OXG1 and BUT_OXG2, respectively.

BUT_OXG1 and BUT_OXG2 were cultivated at 37 °C in a rotary shaker at 200 rpm in complex medium (TB) and in defined medium (HDM). The cells were grown in shake-flasks (under aerobic conditions) until the mid-exponential phase (0.4-0.5 OD_{600}) was reached. At this moment, IPTG was added to the medium and cells were shifted to sealed serum bottles and grown for 96 h at 30 °C and 180 rpm.

The medium composition is one of the major aspects considered during the optimization of a bioprocess. The goal of this experiment was to evaluate the influence of the medium and IPTG concentration on butanol titer. In Figure 4.4, the final butanol concentrations obtained are summarized for each strain, cultivated either in TB or HDM medium and induced with different IPTG concentrations.

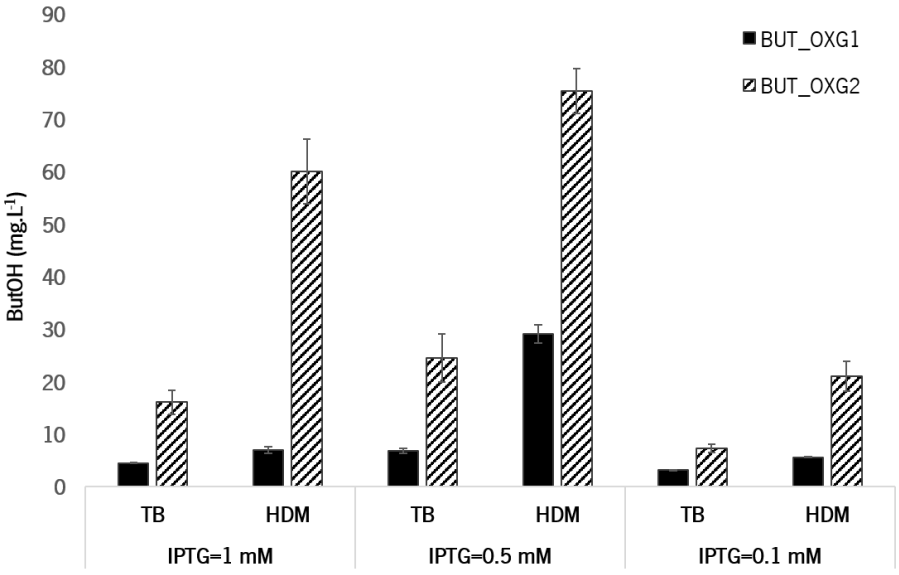


Figure 4.4: Butanol titer (mg.L^{-1}) for strains BUT_OXG1 and BUT_OXG2 in TB medium and HDM medium for different concentrations of IPTG (1, 0.5 and 0.1 mM). In all experiments, strains were grown in shake-flasks until 0.4-0.5 OD_{600} , IPTG was then added to the medium and cells transferred to sealed serum bottles. Titers are shown as mean ± S.D from three independent experiments.

In Figure 4.4, it is possible to observe that, in all experiments, butanol final titer was higher for BUT_OXG2 than for BUT_OXG1 strains. Also, for all strains and IPTG concentrations, butanol final concentration was higher in HDM than in TB medium. The maximum titer was 75.32±4.21 $mg.L^{-1}$, obtained by cultivating BUT_OXG2 in HDM medium and inducing with 0.5 mM of IPTG. Regarding the concentration of IPTG, both maximum concentrations for BUT_OXG1 and BUT_OXG2 strains were obtained inducing the cells with 0.5 mM of IPTG: 29.05±1.72 and 75.32±4.21 $mg.L^{-1}$, respectively.

The highest titer obtained for BUT_OXG2 (75.32±4.21 $mg.L^{-1}$) was 2.6-fold greater than the one obtained for BUT_OXG1 (29.05±1.72 $mg.L^{-1}$). Overall, HDM medium and IPTG concentration of 0.5 mM seem to favor butanol production.

It was shown before that the production of butanol through the clostridial pathway in *E. coli* benefits from the presence in the medium of complex components, such as tryptone and yeast extract (Wen and Shen, 2016). Considering the results obtained here, in the case of this novel heterologous pathway, the presence in the medium of complex components (TB medium) did not increase butanol production comparing with HDM medium. Actually, in all experiments, BUT_OXG1 and BUT_OXG2 accumulated more butanol in HDM than in TB medium. Hence, the production of butanol seems to be positively affected by HDM formulation. 2-Oxoglutarate, the initial substrate of the proposed pathway, is the intersection between the two major metabolic pathways of a microorganism: carbon and nitrogen. Its concentration depends on several factors and fluctuates according to nitrogen and carbon availability (Huergo and Dixon, 2015). We hypothesize that by supplying amino acids (common to both media), 2-oxoglutarate is not consumed during nitrogen assimilation. Nevertheless, in HDM medium the amino acids are supplied in the free form, while in TB medium tryptone could contain small peptides. Depending on the type of amino acids supplied, we do not exclude a possible influence on 2-oxoglutarate accumulation. Also, its concentration varies depending on carbon availability, i.e., the supply of carbon sources such as glucose and glycerol increases the intracellular concentrations of 2-oxoglutarate. Since TB composition includes complex components and reportedly favors heterologous protein expression, we hypothesize that some of the additional compounds present in TB medium might be repressing the production of 2-oxoglutarate and lowering butanol titers. Nevertheless the number of known pathways metabolically regulated by 2-oxoglutarate concentration has been increasing (Huergo and Dixon, 2015). For this reason, without additional

data, it is difficult to pinpoint exactly the influence of the media composition on its concentration and consequent regulatory effects.

The intermediate concentration of IPTG tested, 0.5 mM, provided the best compromise between high level of heterologous enzymes expression, the metabolic burden imposed affecting biomass growth and potential IPTG toxicity effects. The experiments with the lowest concentration (0.1 mM) of IPTG tested resulted in the lowest accumulated butanol concentrations, for both strains and media. So, this concentration appears to be insufficient to induce an elevated level of enzyme expression. On the other hand, although the experiments where cells were induced with 1 mM IPTG achieved higher concentrations of butanol when compared with 0.1 mM, they were still lower than the ones induced with 0.5 mM. One possible explanation is the metabolic burden imposed by high concentrations of plasmid replication, affecting biomass growth and plasmid stability (Rosano and Ceccarelli, 2014). Moreover, IPTG could be toxic to the cells causing cell death (Baneyx, 1999).

E. coli K12 MG1655 was used as host for expressing the studied pathway, considering that all the *in silico* simulations were based on *i*JO1366 (Orth et al., 2011), a genome scale metabolic model developed for this strain (Chapter 3). We also tested *E. coli BL21* (DE3) to produce butanol for being considered more suitable for heterologous gene expression. In fact, *E. coli* BL21 genome was modified to improve protein expression, lacking genes encoding proteases (such as *Lon* and *OmpT*) and preventing plasmid loss by the mutation in gene *hsdB* (Rosano and Ceccarelli, 2014). Nevertheless, the results (Figure 4.4) showed that *E. coli* K12 produced higher titers than BL21 cells. Since recombinant protein production is expected to be favored in *E. coli* BL21, a possible explanation for the contradictory results may be the metabolism of the two different strains. Reportedly, in BL21 cells the enzymes responsible for the glyoxylate bypass are constitutively expressed, while in K12 cells, their expression is extremely regulated (Phue and Shiloach, 2004). This pathway converts directly oxaloacetate and acetyl-CoA into malate and succinate, avoiding the two successive decarboxylations of isocitrate into 2-oxoglutarate and succinyl-CoA. (Phue and Shiloach, 2004). Assuming that the glyoxylate bypass is only active in BUT_OXG1 strains, the respective flux through the oxidative branch of the TCA cycle will be lower than in BUT_OXG2. Consequently, the accumulation of 2-oxoglutarate will be reduced (Li et al., 2013). We hypothesize that, in BUT_OXG2 strains, the 2-oxoglutarate pool is bigger, increasing the flux towards the heterologous production of butanol.

Since butanol production experiments had a better performance using HDM medium (Figure 4.4), this medium was used to evaluate the effect of the cell density (OD_{600}) at induction time on butanol production. To do so, BUT_OXG1 and BUT_OXG2 were induced at values of OD_{600} slightly higher (0.6-0.7) than in the previous experiments (0.4-0.5) and the obtained results are summarized in Figure 4.5.

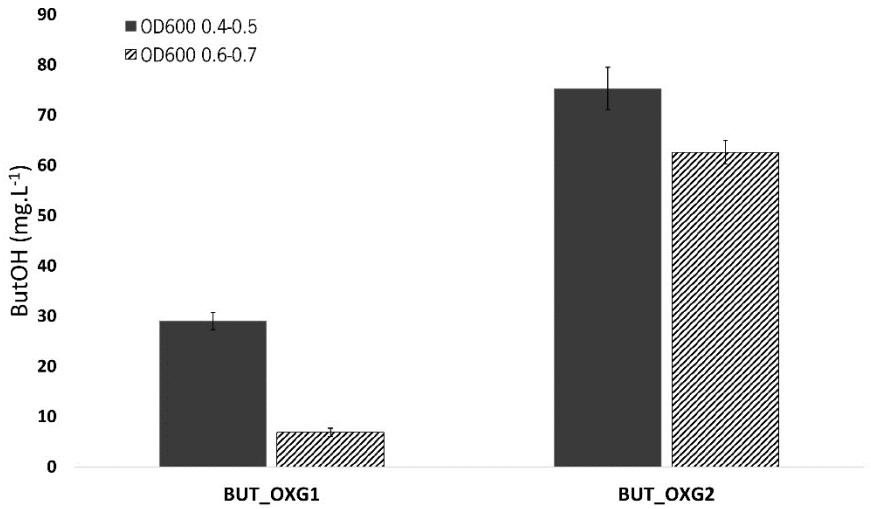


Figure 4.5: Final extracellular butanol titer (mg.L^{-1}) for strains BUT_OXG1 and BUT_OXG2 at different induction times. Cells were grown in shake-flasks at 37 °C and at the OD_{600} shown 0.5 mM of IPTG *was added, and cells were transferred to sealed serum bottles. Data are shown as mean ± S.D from three independent experiments.*

It is possible to observe that, for both strains, a higher number of cells at the induction time decreased the final butanol titer. For BUT_OXG1, butanol titer was 9.05-fold greater when inducing cells with an OD_{600} of 0.4-0.5 when compared with 0.6-0.7, while for BUT_OXG2 the difference was 2.59-fold higher with a lower number of cells at induction time. The drop on butanol titer by inducing the strains at a higher cell density was more significant in *E. coli* BL21 DE3. Also, considering the previously obtained results (Figure 4.4), in which K12 outperformed BL21 strains in terms of butanol titer. So, further optimization still needs to be performed with BUT_OXG1 strains.

To understand how the supplementation of glutamate, iron (III) citrate and amino acids influenced butanol production, different HDM media formulations, in which these supplements were omitted were used to grow BUT_OXG1 and BUT_OXG2. In Figure 4.6, the final butanol titer and OD_{600} achieved for different media formulations are represented.

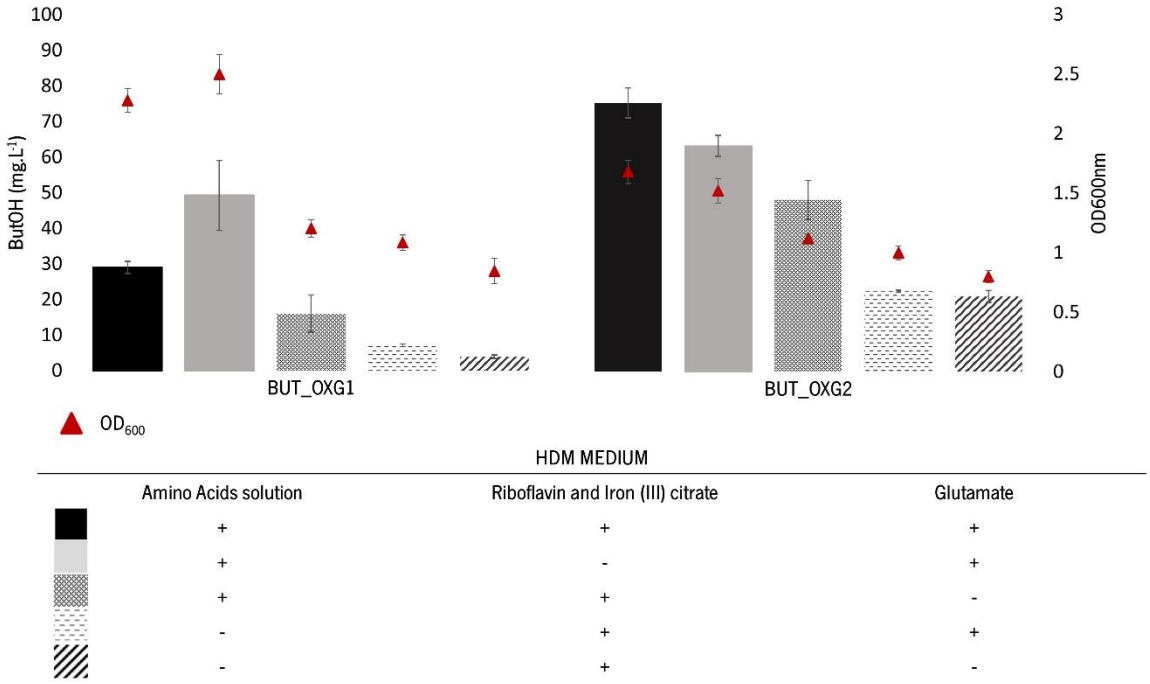


Figure 4.6: Final extracellular butanol titer (mg.L^{-1}) and OD_{600} value for strains BUT_OXG1 and BUT_OXG2 for different formulations of HDM medium. Cells were grown in shake-flasks at 37 °C until 0.4-0.5 OD_{600}, when 0.5 mM of IPTG was added and cells were transferred to sealed serum bottles. Data are shown as mean ± S.D from three independent experiments. ButOH - Butanol

For each medium formulation, final butanol titer was superior for BUT_OXG2 than for BUT_OXG1 strains, similarly to the results previously obtained (Figure 4.4). The formulation that provided the higher butanol titer (75.32±4.21 mg.L^{-1}), and cell density (OD_{600} = 1.68±0.10) for BUT_OXG2 was the medium formulation including all supplements. Surprisingly, for BUT_OXG1, the omission of iron (III) citrate and riboflavin had a positive impact in butanol titer (44.26±2.40 mg.L^{-1}) and number of cells (OD_{600} =2.4) when compared with the complete formulation of HDM medium (final butanol titer of 29.05±2.28 mg.L^{-1} and OD_{600} of 2.28). These compounds were added to the medium because of the reported iron and riboflavin requirements of 2-hydroxyglutarate dehydratase (*hgdABC*) . This enzyme possesses two [4Fe-4S] clusters and uses riboflavin and riboflavin-5'-phosphate (FMN) as prosthetics groups (Hans et al., 1999). One possible explanation for these results is that the cells used citrate as carbon source – as reported for *E. coli* under anaerobic conditions in the presence of other available substrates such as glucose (Lutgens and Gottschalk, 1980) – decreasing butanol titer. Experiments where only iron (III) citrate is omitted from the medium should be performed to

confirm this hypothesis. Nevertheless, the fact that this behavior is only observed in *E. coli* BL21 DE3 cells remains inexplicable.

The omission of the amino acids solution caused the highest reduction on both butanol titer and final OD_{600}. For BUT_OXG2, both formulations where amino acids were omitted from the medium led to final butanol titers of 22.52±0.31 $mg.L^{-1}$ and 21.09±1-67 $mg.L^{-1}$. Therefore, at least for the HDM medium, the presence of amino acids favors butanol production. The lower butanol titers in TB medium (Figure 4.4) should be a consequence from the presence of complex nutrients in the medium as mentioned before.

Given the direct connection between butanol production in the strains tested and 2-oxoglutarate, we expected that the omission of glutamate would cause a greater impact on butanol titer. For BUT_OXG1 and BUT_OXG2 strains, the omission of glutamate led, respectively, to a 1.71-fold and a 1.56-fold decrease in butanol accumulation comparing with the complete HDM formulation. However, when the omission of glutamate was coupled to the omission of the amino acids solution, the results showed similar values to the ones obtained with the single omission of amino acids. This suggests that extra glutamate in the medium only has a considerable impact on butanol production when extra amino-acids are supplemented.

Glutamate is intrinsically correlated with 2-oxoglutarate, being produced by action of glutamate dehydrogenase through assimilation of ammonium into 2-oxoglutarate (Huergo and Dixon, 2015). So, by supplying this amino acid we would expect to increase the availability of 2-oxoglutarate. Inside the cell, glutamate acts as an intracellular nitrogen donor for other nitrogen-containing compounds such as other amino acids (Huergo and Dixon, 2015). In fact, glutamate is the donor of at least one molecule of nitrogen for eleven amino acids (Reitzer, 2003). Therefore, by supplementing amino acids to the medium, glutamate dehydrogenase is inactive leading to the accumulation of 2-oxoglutarate. The extra supply of glutamate probably increases even more the 2-oxoglutarate pool, rewiring the flux towards production of butanol.

It is also possible to observe that cell density is directly correlated with the titer of butanol for each strain. However, even though *E. coli* BL21 strains achieved higher values of OD_{600} compared with their K12 equivalent, the specific biomass yield on butanol is lower for these strains, similarly to what was observed in previous experiments (Figure 4.4 and Figure 4.5).

It has been described that IPTG induction causes a metabolic burden imparing biomass growth (Mairhofer et al., 2013). In the butanol prodution experiments, an additional stress is caused to the cells by transferring them, at induction time, to anaerobic conditions. To infer the influence of aerobic growth after IPTG induction, we tested different intervals of time between IPTG induction and switch to anaerobic conditions as depicted in Figure 4.7.

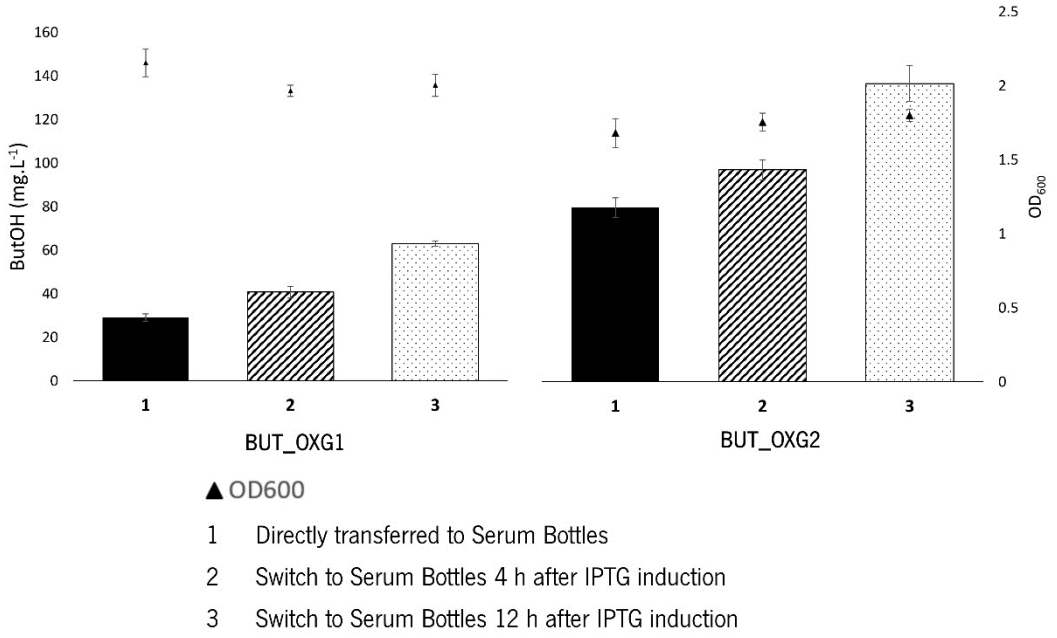


Figure 4.7: Final extracellular butanol titer (mg.L^{-1}) and OD_{600} value for strains BUT_OXG1 and BUT_OXG2 by culturing in HDM medium. Cells were grown in shake-flasks at 37 °C until 0.4-0.5 OD_{600}, when 0.5 mM IPTG was added and cells were transferred immediately (1); after 4 h (2); or after 12h (3) to sealed serum bottles. Data are shown as mean ± S.D from three independent experiments. ButOH - Butanol

Most of the enzymes constituting the proposed pathway ae expressed under anaerobic conditions. In particular, the activator of 2-hydroxyglutaryl-CoA dehydratase (*hgdC*) is oxygen-sensitive (Djurdjevic et al., 2011). In this regard, we selected 4 h and 12 h intervals between induction and anaerobic switch to test if the pathway would still functional and the respective effect on butanol production.

From Figure 4.7, it is possible to see that the delay of cell transfer to serum bottles had a positive impact on final butanol titer for both strains. Similarly to previous experiments (Figure 4.4, Figure 4.5 and Figure 4.6), final butanol concentration was higher for BUT_OXG2 strains than for BUT_OXG1. In fact, the minimal butanol accumulation achieved in BUT_OXG2 experiments (75.32±4.21 mg.L^{-1}) is still higher than the maximal value achieved for BUT_OXG1 (62.94±1.22 mg.L^{-1}).

The greater butanol titer (128.95±7.73 mg.L^{-1}) was achieved by culturing BUT_OXG2 and delaying, after IPTG induction, for 12 h the transfer to serum bottles. In this experiment, samples were taken at the moment of the anaerobic switch and analyzed for OD_{600} and production of butanol. In all experiments, butanol was not detected at the induction time.

Although, at the moment of the anaerobic switching, BUT_OXG1 and BUT_OXG2 strains had a higher value of OD_{600} (data not shown), in the end of the experiments (96 h) this value, when compared with the value obtained for immediately transference to serum bottles, was similar for BUT_OXG2 (with overlapping confidence intervals (1.68±0.10; 1.75±0.06 and 1.8±0.04)) or, in the case of BUT_OXG1, even lower.

The final values of OD_{600} were similar among the three different tested conditions, but gradually higher at the beginning of anaerobic growth with the delay of 4 h and 12 h between induction and transfer to serum bottles. So, we may conclude that delaying the anaerobic switch led to a lower growth rate under anaerobic conditions. Nevertheless, the final concentration of butanol increased for both strains. We hypothesize that, during aerobic growth after IPTG induction (4 h or 12 h), the cells were able to better express the heterologous proteins without affecting their stability. So, when the strains actually started to grow anaerobically, all the heterologous enzymes were available, increasing the activity of heterologous pathway and, consequently, butanol titers.

From previous experiments a correlation between final value of OD_{600} and butanol production was evident but in these experiments the increment on butanol production was not matched with an increase in the final values of cell density.

From the results obtained by cultivating the cells in serum bottles (Figure 4.4, Figure 4.5, Figure 4.6 and Figure 4.7), we may conclude that HDM formulation favors butanol production in detriment of TB composition. In this medium, the supplementation of amino acids provoked the greater increment on butanol accumulation. The cell density at the moment of IPTG induction influences butanol final titer, 0.4-0.5 OD_{600} led to higher butanol final concentrations than 0.6-0.7 OD_{600} and the optimal concentration of IPTG was 0.5 mM. Finally, *E. coli* K12 was able to produce higher titers of butanol in all the conditions tested than BL21 cells.

4.3.3 Butanol production in Bioreactors

Under anaerobic conditions, *E. coli* produces mixed-acid fermentation products decreasing the pH of the medium, which can slow, or even stop growth. For this reason, BUT_OXG2 (the strain that provided the greater titers in the serum bottle experiments) was cultivated under controlled conditions (including pH) in a bioreactor.

BUT_OXG2 strains were cultivated with 0.5 L-working volume in a 2 L-bioreactor using HDM medium. Samples were taken to measure cell density, concentration of excreted metabolites with HPLC and GC-FID analysis. Figure 4.8 summarizes the obtained results.

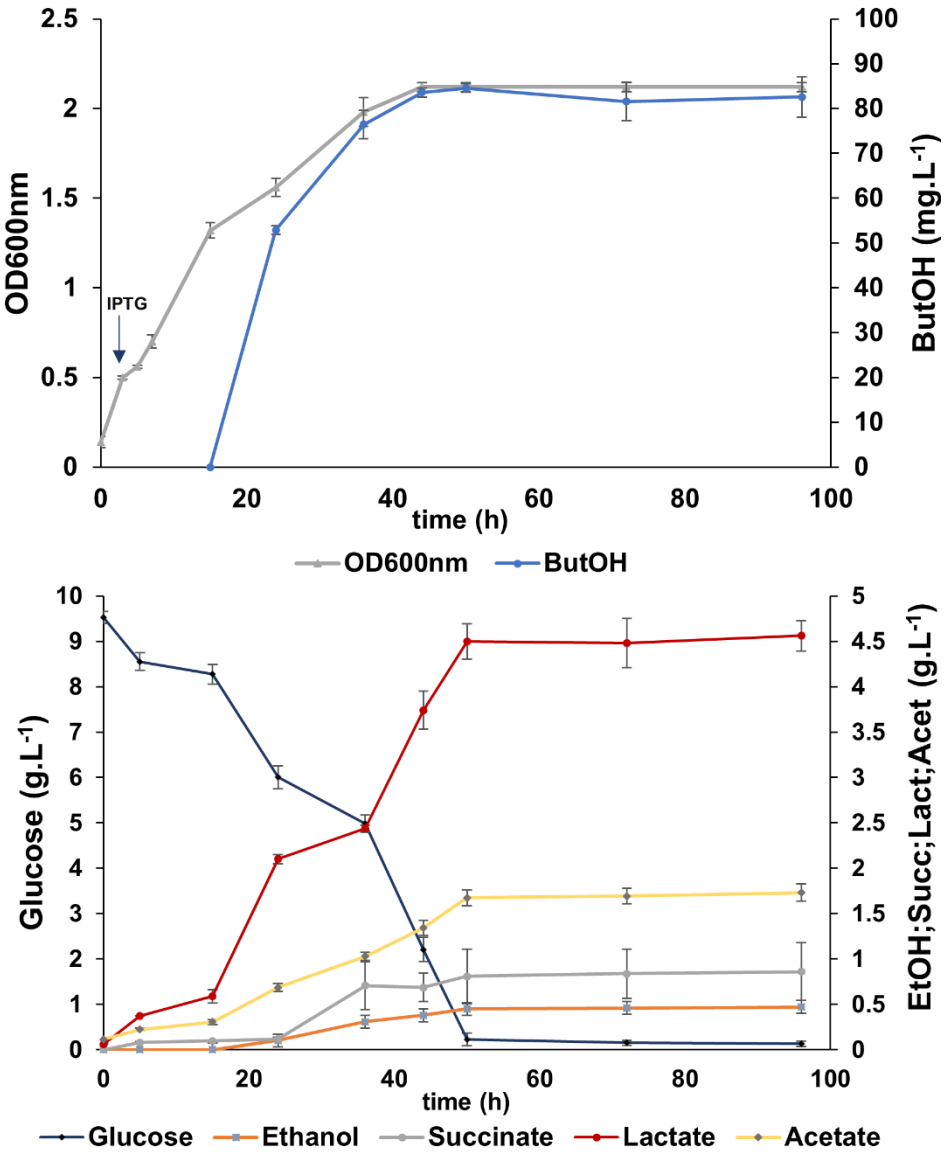


Figure 4.8: Growth-curve profile and butanol production (a), glucose consumption profile and end-products formation (b) of strain BUT_OXG2 in controlled bioreactors. Cells were cultivated in HDM medium and induced with 0.5 mM of IPTG at 0.4-0.5 OD_{600}. Data

are shown as mean ± SD of three independent experiments. ButOH: Butanol; EtOH: Ethanol; Succ: Succinate; Acet: Acetate; Lact: Lactate;.

It is possible to observe that the maximum butanol titer (84.57±0.99 $mg.L^{-1}$), obtained after 50 h, represented a 1.12-fold improvement, comparing with the results obtained in serum bottles for BUT_OXG2 strains (Figure 4.4). We can see a biomass-product coupled behavior by analyzing the growth-curve and the butanol production profile. These strains presented several stages of growth: first, during the first 3 h, under aerobic conditions. At this moment, cells achieved mid-exponential phase, and 0.5 mM of IPTG was added to the medium. Between 5 h and 12 h the growth rate and glucose consumption decreased. The switch to anaerobic conditions and the metabolic burden imposed by IPTG induction can explain this stagnation. Next, between 12 h and 50 h cells continuously grew, consuming glucose, producing butanol and mixed-acid fermentation products until reaching a stationary phase.

Butanol was first detected at 24 h of fermentation and its accumulation increased until 50 h. For this moment on, butanol concentration remained constant, coinciding with the glucose exhaustion and biomass growth plateau. In the experiments using serum bottles, the greatest titer was only achieved after 96 h of cell growth (data not shown). So, we can conclude that cultivating the cells under controlled conditions has accelerated the growth rate and butanol production, although the increase in titers has been modest. Nevertheless, even in bioreactors the growth rate (μ) was quite low (0.035±0.002 h^{-1}), representing a duplication time of 19.9±1.3 h. Lactate and acetate were the metabolites accumulated by cells at higher concentrations. Succinate and ethanol were also detected in the fermentation broth. The growth rate (μ), the duplication time (td) and the yields on substrate for the different products (butanol, ethanol, succinate, acetate and lactate) detected in the broth are summarized in Table 4.6

Table 4.6: Specific growth rate (μ), duplication time (td) and butanol. ethanol, succinate, acetate and lactate yields (Y) on glucose. BUT_OXG2 cells were cultivated under controlled conditions in 2 L bioreactors in HDM medium for 96 h. Gluc: Glucose; ButOH: Butanol; EtOH: Ethanol; Succ: Succinate; Acet: Acetate; Lact: Lactate

Parameter	Value
μ (h^{-1})	0.035±0.002
t_d (h)	19.9±1.3
$Y_{ButOH/S}$ ($mg_{ButOH}.g_{Gluc}^{-1}$)	9.51±0.11
$Y_{EtOH/S}$ ($g_{EtOH}.g_{Gluc}^{-1}$)	0.05±0.01
$Y_{Succ/S}$ ($g_{Succ}.g_{Gluc}^{-1}$)	0.09±0.03
$Y_{Acet/S}$ ($g_{Acet}.g_{Gluc}^{-1}$)	0.18±0.01
$Y_{Lact/S}$ ($g_{Lact}.g_{Gluc}^{-1}$)	0.48±0.02

Analyzing Table 4.6 and Figure 4.8, it is clear that lactate was the major fermentation product released by the cells. In fact, the yield of lactate on glucose was 0.48±0.02 $g_{Lat}.g_{gluc}^{-1}$, 52.5-fold greater that the yield of butanol on glucose (9.51±0.11 $mg_{ButOH}.g_{gluc}^{-1}$). Actually, butanol yield on substrate was the lowest one, within the detected products, namely succinate, lactate and acetate. So, during the anaerobic growth, *E. coli* preferably uses the native pathways to recycle the excess of NADH instead of the butanol pathway, even with the high amount of heterologous proteins being expressed. The results obtained in Chapter 3 also indicated that the mixed-acid fermentation pathways should be knocked-out to redirect the metabolism towards the heterologous pathway.

4.3.4 Pathway validation

Since the *in silico* driven approach from Chapter 3 showed that butanol was only produced under anaerobic conditions, as a control experiment, BUT_OXG1 and BUT_OXG2 were cultivated under aerobic conditions in shake-flasks, in the same fashion as described before, but skipping the anaerobic switch step after IPTG induction. As expected, butanol was not detected in the aerobic experiments (data not shown). The detection limit of the analytical method is 3 $mg.L^{-1}$.

This can be explained by two distinct reasons. First, the exigent demand of this heterologous pathway for NADH molecules. For each molecule of butanol synthesized, four molecules of NAD(P)H are

required. Although most of the enzymes have more affinity for NADH, some of them can also recycle NADPH, as already studied for alcohol dehydrogenase from *C. acetobutylicum* (Welch et al., 1989). Since under aerobic conditions it is metabolically more advantageous for the cells to directly oxidize those molecules producing ATP (Trinh et al., 2011), no butanol production was expected under these conditions, as in fact happens for most fermentation products. This was predicted by the *in silico* analysis (Chapter 3), as well. Also, the enzymes that constitute the proposed pathway are natively expressed in microorganisms either strictly anaerobic (*A. fermentans*, *C. symbiosum and C. acetobutylicum*) or facultative anaerobes (namely, *Treponema denticola)*, and some of them are sensitive to oxygen. Nevertheless, *E. coli* strains expressing the clostridial pathway have accumulated butanol in experiments under aerobic conditions although at lower concentrations than under (semi-)anaerobic conditions (Atsumi et al., 2008a; Shen et al., 2011). It was also observed that both strains achieved a final OD_{600} much higher than the values achieved for anaerobic conditions (data not shown), but butanol was not detected. This confirms that butanol production benefits from anaerobic conditions and that this pathway can be used as alternative to the endogenous NADH recycling mechanisms.

To validate that the pathway introduced into *E. coli* was using 2-oxoglutarate as starting precursor, strains expressing a part of the pathway, Control_OXG1 and Control_OXG2, were cultured in the same fashion as described before for experiments using TB and HDM media. These strains only expressed the final five catalytic steps of the pathway. The main goal was to infer if, somehow, some substrates were being used directly from the media. Since butanol was not detected (the method detection limit is 3 $mg.L^{-1}$) in all the performed experiments, we can conclude that all the three plasmids are needed to produce a detectable amount of butanol (data not shown).

4.4 Conclusions

In this chapter, we were able to design new strains of *E. coli* capable of producing butanol through a novel pathway, by inserting the pathway selected from the *in silico* analysis carried in Chapter 3.

Two different genetic constructions were designed: one catalyzing the decarboxylation of glutaconyl-CoA into crotonyl-CoA by using the α-subunit of a glutaconyl-decarboxylase and another expressing

a glutaryl-CoA dehydrogenase, using glutaconyl-CoA as an intermediate. Only strains expressing *gcdH* (glutaryl-CoA dehydrogenase) were able to accumulate butanol.

E. coli BUT_OXG2, using *E. coli* K12 MG1655 DE3 as host, provided greater final titers of butanol than using BL21. The greatest titer (128.95±7.73 $mg.L^{-1}$) was obtained by cultivating BUT_OXG2 strains aerobically in HDM medium, inducing with 0.5 mM of IPTG and, after 12 h, transferring the cells to serum bottles.

The growth delay between IPTG induction and transfer to serum bottles increased butanol production (Figure 4.7). Although 12 h was the largest tested interval, switching to anaerobic conditions is mandatory to produce butanol, since no butanol was detected in the fully aerobic experiments.

In the control experiments, with strains BUT_OXG1 and BUT_OXG2 cultured under aerobic conditions and strains expressing only the last steps of the proposed pathway (Control_OXG1 and Control_OXG2), no butanol was detected.

The bioreactor experiments did not increase the butanol final titer by much (Figure 4.8), representing a 1.12-fold increment on butanol titer comparing with serum bottle experiments. The maximum concentration was achieved after 50 h of fermentation, while in serum bottle experiments the maximum value was achieved after 96h.

As observed in Figure 4.5, just by slightly increasing the number of cells at induction time, the final butanol concentration decreased. These results show that the optimization of a bioprocess involves various stages and it is an exhaustive task. In this regard, for BUT_OXG2 strain, from an initial butanol titer of 7.25±0.8 $mg.L^{-1}$ (obtained by culturing these strains in TB medium and induce with 0.1 mM of IPTG), we were able to improve butanol production by 17.8-fold, achieving the maximum value of 128.95±7.73 $mg.L^{-1}$. Nevertheless, there is still a lot of room for improvement, turning this bioprocess more efficient.

Actually, the maximum titers obtained are still far below from those required for industrial purposes. The *in silico* simulations of Chapter 3 have indicated that, under steady-state conditions, from a 10 $mmol.(g_{DW}.h)^{-1}$ glucose uptake, 4.94 $mmol.(g_{DW}.h)^{-1}$ of butanol could be excreted. However, to achieve the optimum genotype indicated, the knock-out of mixed-acid fermentation pathways is necessary to force NADH recycling using the heterologous pathways.

CHAPTER 5 Engineering the mixed-acid fermentation pathways in *Escherichia coli* to improve anaerobic butanol production

The genome manipulation of a microorganism is one of the most common strategies to redirect the cellular metabolism to the production of a target compound. In this regard, the advent of better genome-editing techniques allows to rapidly implement engineering strategies in the selected host.

In this chapter, we have implemented an *in silico* metabolic engineering strategy (Chapter 3) in *Escherichia coli* to improve butanol production, which consisted in deleting the genes encoding the major mixed-acid fermentation enzymes: *adhE*, *adhP* and *mhpF* (involved in ethanol production); *ldhA* (lactate production), *pta* and *ackA* (acetate production).

The *adhP*, *ldhA*, *adhE*, *pta* quadruple knock-out strain achieved the highest butanol titer when cultivated in HDM medium, using 0.5 mM of isopropyl 1-thio-β-D-galactopyranoside (IPTG) to induce the heterologous genes and by delaying the post-induction transfer to anaerobiosis for 12h. The maximum butanol titer obtained in these conditions was 49.52±0.23 $mg.L^{-1}$, which was still lower than the titer achieved with the reference strain without any knock-outs.

Mixed-acid fermentation products compete with butanol for reducing power in anaerobic conditions. Although butanol production could not be improved by removing the competing pathways, we were able to cease ethanol production and reduce lactate and acetate accumulation. Even though the experimental results did not fully match the computational predictions (Chapter 3), adaptive laboratory evolution might be the answer to recover some of the growth deficit shown by the deletion mutants and promote the convergence of the simulated phenotype with the results showed here.

5.1 Introduction

The creation of an efficient microbial cell-factory often requires genome modifications in order to redirect the metabolism to product synthesis and minimization of by-products formation. Before the availability of whole genome sequences, this process relied mostly on isolation from the environment of microorganism with the desired phenotype or on random mutation protocols followed by selection of strains with improved characteristics.

In recent years, the availability of whole genomes has allowed to invert this process, allowing rational strategies to first identify target genes whose modification can increment the production of a certain compound or lead to another relevant phenotype (Nakashima and Miyazaki, 2014). In this regard, specific genes can be modified using different types of mutations including allelic exchange, knock-ins (insertion) and knock-outs (disruption) (Nakashima and Miyazaki, 2014). The advent of techniques to modify the genomic DNA have accelerated the application of these engineering strategies into *E. coli*. The easiness of modifying the genome of *E. coli* makes this organism a good alternative to produce biobutanol when compared with the native producer species (Nielsen et al., 2009).

The efficient development of a biofuel microbial cell factory can include: the selection of the most suitable production genes; the identification of the rate-limiting steps and fine-tuning the biosynthetic pathway expression; the disruption or downregulation of competing pathways to increase the availability of precursors and cofactors (Zhao et al., 2017). Specifically, the production of highly reduced compounds such as alcohols in *E. coli* is challenging due to the great demand of reducing power (Saini et al., 2016).

In order to increase NADH availability and, consequently, butanol titers, several approaches have been applied mostly by rewiring the central carbon metabolism. In *E. coli*, pyruvate is oxidized into acetyl-CoA, under anaerobic conditions, by the pyruvate formate-lyase (PFL), releasing formate as a byproduct. In order to increase NADH availability in *E. coli*, the formate produced by PFL can be oxidized to CO_2 by a formate dehydrogenase (FDH) found in other microorganisms, which releases NADH. By expressing in *E. coli* the FDH from *Saccharomyces cerevisiae* (Nielsen et al., 2009) and *Candida boidinii* (Ohtake et al., 2017) the respective butanol titers were improved. On the other

hand, under aerobic conditions, pyruvate is oxidized into acetyl-CoA by pyruvate dehydrogenase (PDH) generating NADH. For this reason, the activation of the PDH complex under anaerobic conditions, by promoter replacement, implies the generation of two more molecules of NADH per glucose consumed, which can improve butanol titers (Bond-watts et al., 2011; Garza et al., 2012; Lim et al., 2013).

The inactivation of the mixed-acid fermentation pathways can also increase the NADH pool by deleting the native NADH recycling pathways of *E. coli*. The Liao group reported in 2008 the first successful implementation of the butanol biosynthetic pathway in *E. coli*. By deleting the genes responsible for ethanol, lactate and succinate production (*adhE*, *ldhA* and *frd,* respectively), the butanol titer almost doubled. The additional deletion of *pta* or *fnr* decreased butanol production, but butanol titer improved when the two genes were deleted simultaneously together in the triple knock-out strain. On the other hand, by disrupting *plfB*, butanol production was almost abolished, indicating that PFL was, in the tested conditions, the main enzyme responsible for the acetyl-CoA accumulation (Atsumi et al., 2008a). Several other knock-out combinations have been reported in the literature, such as the *adhE*, *ackA*, *ldhA*, *pfl*, *frd* knock-out strains (Shen et al., 2011). Generally, the reported genetic interventions target the genes responsible for the production of ethanol, acetate, lactate, and succinate.

In Chapter 4, we had successfully developed strains of *E. coli* able to produce butanol through a novel pathway. However, the maximum butanol titer obtained, 128.95±7.73 $mg.L^{-1}$, was still modest and too low for a strain suitable for industrial purposes.

Consequently, in this chapter, we aim to apply the optimum genotype provided by the *in silico* optimizations performed in Chapter 3, in order to improve the butanol titers, abolish ethanol and lactate production and reduce acetate accumulation. This genotype included the knock-out of the main mixed-acid fermentation pathways, active during anaerobicity, obligating *E. coli* to recycle the excess of NADH through the heterologous butanol production pathway. A schematic representation of the strain design is depicted in Figure 5.1.

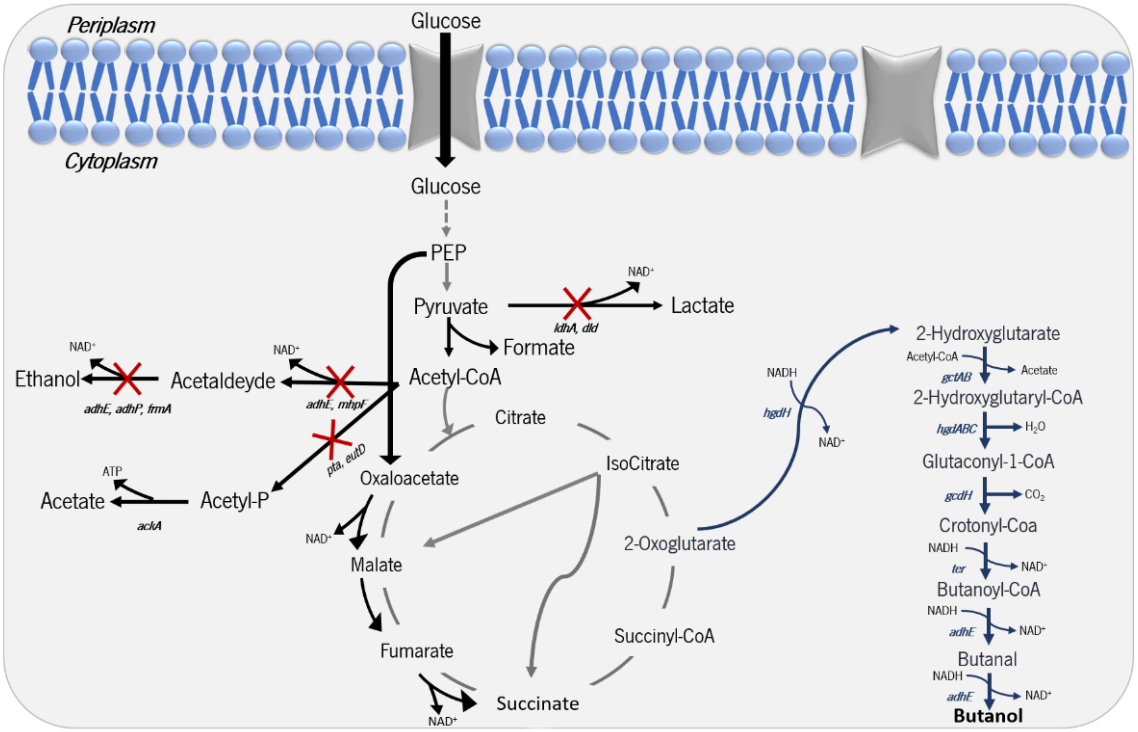


Figure 5.1: Schematic representation of the engineering strategy to improve butanol production. Dashed lines represent multiple reactions; Heterologous reactions are represented by dark blue arrows; red cross marks indicate knocked-out reactions; black arrows indicate fermentative reactions. PEP: phosphoenolpyruvate; Acetyl-P: Acetyl-phosphate

5.2 Materials and methods

5.2.1 Bacterial strains and plasmids

E. coli K12 MG1655 (DE3) was used as host in all experiments. The plasmids used in this chapter and their source are summarized in Table 5.1. pTKRED (Addgene Plasmid #41062) and pTKS/CS (Addgene Plasmid #41063) were a gift from Edward Cox and Thomas Kuhlman (Kuhlman and Cox, 2010).

Table 5.1: Plasmids used or engineered for this study.

Plasmid	Features	Source
pETDuet_adhE_ter	pETDuet[-1] carrying *adhE* from *C. acetobutylicum* and *ter* from *T. denticola*	Chapter 4
pCDFDuet_gctAB_hgdH	pCDFDuet[-1] carrying codon-optimized *gctAB* and *hgdH* from *A. fermentans*	Chapter 4
pRSFDuet_gcdH_hgdABC	pRSFDuet[-1] carrying codon-optimized *hgdC* from *A. fermentans*; *hgdAB* from *Clostridium symbiosum* and *gcdH* from *Pseudomonas aeruginosa PAO1*	Chapter 4
pTKRED	pSC101 ori, AraC, araBAD promotor, lac UV5 promotor, Rep101, lacI ,*recA*, *gam*, *beta, exo*, Sm^R	Addgene
pTKS/CS	p14A ori, cat Promotor, Tc^R, Cm^R,	Addgene

5.2.2 DNA manipulation

Polymerase chain reaction (PCR) was performed using Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Waltham, USA) in a LifeECO Thermal Cycler (Bioer Technology, Zhejiang, China). All primers were purchased from Metabion (Munich, Germany). DNA fragments were purified using DNA Clean and Concentrator DNA Kit (Zymo Research, Irvine, USA).

Plasmids were extracted using Plasmid Miniprep kit (Zymo Research). Colony PCR was performed using DreamTaq (Thermo Scientific) and sequences were further confirmed by sequencing (StabVida, Lisbon, Portugal). Protocols were performed in accordance with manufacturer's instructions.

5.2.3 Gene knock-outs

Gene knock-outs were introduced in *E. coli* MG1655 K12 (DE3) and its derivatives by first transforming the target strain with the plasmid pTKRED and following the method developed by (Kuhlman and Cox, 2010). The plasmid pTKS/CS (Tas et al., 2015) was used as PCR template for the amplification of the 1.3 Kb deletion cassette, which includes a gene conferring resistance to tetracycline (*tetA*) that is flanked by I-*Sce*I restriction sites. The plasmid pTKRED, with a spectinomycin resistance marker, included a constitutively expressed *recA* gene and the threeλ-Red genes (*gam*, *bet* and *exo*), inducible by IPTG. These genes were necessary for the integration of the deletion cassette into the desired integration site and for the enhancement of fragment insertion via recombination. This plasmid also included a P_{araBAD}-driven I-*Sce*I gene, inducible with L-arabinose, which allowed to excise the deletion cassette with no scars from the genome.

To construct the deletion cassette and check the respective integration into the genome, for each target gene, five primers were designed: Primer A, B, C, D and E. Primer A, included in its 5' end around 50 bp homologous to the region upstream of the target gene (including the start codon) and 14 bps on the 3' end that are reverse complementary to the 3' end of primer B. The 5' end of primer B comprised 25 bps reverse complementary to the beginning of the cassette to be amplified from the plasmid pTKS/CS. Primer B also included 50 bps reverse complementary to the sequence downstream the target gene (including the stop codon). Primer C contained in the 5' end, 50 bps reverse complementary to the last 50 bps of the coding sequence of the targeted gene. Moreover, 25 bps, on the 3' end, were reverse complementary to the end of the cassette region on plasmid pTKS/CS. Primers D and E included 20-25 bps flanking the gene region and were obtained using the Primer Designing Tool from NCBI, in order to design specific primers (Ye et al., 2012). These primers were used amplify the cassette insertion region to check different sizes of the wild-type (gene), insertion of the deletion cassette and, finally, cassette excision. The list of primers designed for this study are summarized in Table 5.2.

Table 5.2: List of primers used in the knock-out protocol of each target gene.

Primer	Sequence
*ackA_*	TGGCTCCCTGACGTTTTTTTAGCCACGTATCAATTATAGGTACTTCCATGTGATTTCACACCG
*ackA_*	TCACCGTTTGGACCTTGGGGCCGTATTGGCGGGTTACAAAACAGCACCGCCAGCTGAGCTG
*ackA_*	AGTCAGGCGGCTCGCGTCTTGCGCGATAACCAGTTCTTCGTTGGTTGGGATTGGCTTCAGG
*ackA_*	GCGGGACAACGTTCAATAAACA
*ackA_*	GATGCAGCGCAGTTAATAGCAA
*adhE_*	CGAGCAGATGATTTACTAAAAAAGTTTAACATTATCAGGAGAGCATTATGTAATCAGTAGCGC
*adhE_*	TCACCGTTTGGACCTTGGGGCCGTAGGCATTGCCCAGAAGGGGCCGTTTATGTTGCCAGAC

*adhE_*	AGCGGATTTTTTCGCTTTTTTCTCAGCTTTAGCCGGAGCAGCTTCTTTCTTTGGCTTCAGGGA
*adhE_*	TGAGTGTGAGCGCGAGTAAG
*adhE_*	AACCATGTCCCCCGGTAAAC
*adhP_*	CATTTGCCTCACCTGCTATGCAGAACATCATCCGAAAAGGAGGAACTATGTAAGAGGCCTTT
*adhP_*	TCACCGTTTGGACCTTGGGGCCGTAAGTGTGCGATGCTGCGACCCGAACATGGCAGTCGCA
*adhP_*	GTGACGGAAATCAATCACCATGCGGCCACGGATTTTGCCTTCTTCCATCTTTGGCTTCAGGG
*adhP_*	ATGCTGATGTCGGCAAGTGA
*adhP_*	TTCCTGTCAGCAATCGGCTT
*ldhA_*	TATTTTTAGTAGCTTAAATGTGATTCAACATCACTGGAGAAAGTCTTATGTAATCTTGCCGCTC
*ldhA_*	TCACCGTTTGGACCTTGGGGCCGTAGGGGATTATCTGAATCAGCTCCCCTGGAATGCAGGG
ldhA_c	TCGTTCGGGCAGGTTTCGCCTTTTTCCAGATTGCTTAAGTTTTGCAGCGTTTGGCTTCAGGG
*ldhA_*	CCGTTCAGTTGAAGGTTGCG
*ldhA_*	TAGCGCACATCATACGGGTC
mhpF	TCAGTTGCTGCGACATTTTCAAGCGCAGCCCCAAAAGGAAGTCTGTCATGTGAACGGTAAAA
mhpF	TCACCGTTTGGACCTTGGGGCCGTACATACCGTCACGCAATGTGACGTCCGAGATATAAAGT
mhpF	GCTTCTCCTGCCTTGCGCGCCAGTGACTGGGCCATTTTTTCCGCTGTCGCTTGGCTTCAGG
mhpF	GGCGGGGTTAGACCTGAAAA
mhpF	GATTCCTGGCAGCAACAACG
pta_a	GCTGTTTTGTAACCCGCCAAATCGGCGGTAACGAAAGAGGATAAACCGTGTAATCTCGTCAT
pta_b	TCACCGTTTGGACCTTGGGGCCGTATTTCCGGTTCAGATATCCGCAGCGCAAAGCTGCGGA
pta_c	CTGCTGCTGTGCAGACTGAATCGCAGTCAGCGCGATGGTGTAGACGATATTTGGCTTCAGG
pta_d	GGTATCGGTGAAAATGCCGC
pta_e	GATGCAGCGCAGTTAAGCAA

The gene deletion cassette was constructed by first fusing 20 pmol of primer A and B with 15 cycles of PCR. 1 µL of the primer AB fusion, together with 20 pmol of primer A and primer C, were used to amplify the deletion cassette from plasmid pTKS/CS (100 ng).

For each target gene, 100 ng of the purified deletion cassette were transformed by electroporation in the target host using 0.1 cm-gap electroporation cuvettes at a voltage of 1.8 KV. Electrocompetent cells were prepared using the protocol developed by (Dower et al., 1988). Transformants were plated on solid medium with 10 $mg.L^{-1}$ of tetracycline and 100 $mg.L^{-1}$ spectinomycin. In order to check the success of the deletion protocol, colony PCR was performed using the respective primers D and E to check the insertion of the deletion cassette. Afterwards, the positive transformants were cultivated overnight in liquid Luria-Bertani (LB) supplemented with 0.2 % (w/v) of arabinose to induce the expression of I-*Sce*I. The cassette excision was checked with the same primer pair. The final sequence was confirmed through sequencing (StabVida, Portugal). The plasmid pTKRED was curated

by cultivating the cells overnight at 42 °C in an orbital shaker and, afterwards, colonies were picked onto LB agar (15 $g.L^{-1}$) plates with 100 $\mu g.mL^{-1}$ spectinomycin to verify the loss of pTKRED

.

5.2.4 Butanol production strains

E. coli K12 MG1655 (DE3) was used as host for gene deletion using the protocol described above. The first deleted gene was *adhP* originating *E. coli* 1KO, followed by the knock-out of *ldhA* yielding *E. coli* 2KO. The strain *E. coli* 3KO was obtained by deleting *adhE* from the previously developed strain. The successive knock-out of *pta* gene originated *E. coli* 4KO. This strain was the host for the *mhpF* deletion, forming *E. coli* 5KO. Finally, *ackA* was deleted from the last strain, originating *E. coli* 6KO.

E. coli K12 MG1655 (DE3) and its derivatives were used as hosts for gene expression under control of T7 promotor. BUT_OXG2, BUT_OXG3, BUT_OXG4, BUT_OXG5 and BUT_OXG6 strains were obtained by transforming *E. coli* K12 MG1655 (DE3), *E. coli* 3KO, *E coli* 4KO, *E. coli* 5KO and *E. coli* 6KO respectively, with pCDFDuet_gctAB_hgdH; pRSFDuet_hgdABC_gcdH and pETDuet_adhE_ter (see Chapter 4 for more information) by electroporation using 0.1 cm-gap electroporation cuvettes at a voltage of 1.8 KV. Electrocompetent cells were prepared using the protocol developed by (Dower et al., 1988). Positive transformants were isolated in LB agar plates, containing the appropriate antibiotics concentrations (50 $\mu g.mL^{-1}$ ampicillin, 50 $\mu g.mL^{-1}$ spectinomycin and 30 $\mu g.mL^{-1}$ kanamycin) and incubated at 37 °C, overnight. To confirm the success of the transformation, in the next day some colonies were inoculated in LB medium with antibiotics, overnight. Afterwards, plasmids were extracted and digested. The resulting fragments lengths were confirmed by running a 1 % (w/v) agarose gel.

For long-term storage, glycerol was added to a final concentration of 30 % (v/v) to overnight cultures in selective media and kept in -80 °C freezer.

Table 5.3 summarizes the strains of *E. coli* used or engineered for this study.

Table 5.3: List of strains and genomic DNA used or engineered for this study.

Strains	Relevant genotype	Source
***E. coli* 1KO**	F - λ - ilvGrfb- 50 rph- 1 λ(DE3) Δ*adhP*	This study

E. coli 2KO	F - λ - ilvGrfb- 50 rph- 1 λ(DE3) Δ*adhP* Δ*ldhA*	This study
E. coli 3KO	F - λ - ilvGrfb- 50 rph- 1 λ(DE3) Δ*adhP* Δ*ldhA* Δ*adhE*	This study
E. coli 4KO	F - λ - ilvGrfb- 50 rph- 1 λ(DE3) Δ*adhP* Δ*ldhA* Δ*adhE* Δ*pta*	This study
E. coli 5KO	F - λ - ilvGrfb- 50 rph- 1 λ(DE3) Δ*adhP* Δ*ldhA* Δ*adhE* Δ*pta* Δ*mhpF*	This study
E. coli 6 KO	F - λ - ilvGrfb- 50 rph- 1 λ(DE3) Δ*adhP* Δ*ldhA* Δ*adhE* Δ*pta* Δ*mhpF* Δ*ackA*	This study
BUT_OXG2	*E. coli* K12 MG1655 pETDuet_adhE_ter; pCDFDuet_gctAB_hgdH; pRSFDuet_gcdH_hgdABC	This study
BUT_OXG3	*E. coli* 3KO pETDuet_adhE_ter; pCDFDuet_gctAB_hgdH; pRSFDuet_gcdH_hgdABC	This study
BUT_OXG4	*E. coli* 4KO pETDuet_adhE_ter; pCDFDuet_gctAB_hgdH; pRSFDuet_gcdH_hgdABC	This study
BUT_OXG5	*E. coli* 5KO pETDuet_adhE_ter; pCDFDuet_gctAB_hgdH; pRSFDuet_gcdH_hgdABC	This study
BUT_OXG6	*E. coli* 6KO pETDuet_adhE_ter; pCDFDuet_gctAB_hgdH; pRSFDuet_gcdH_hgdABC	This study

5.2.5 Butanol production experiments in shake flasks

The resulting genetically engineered strains BUT_OXG2, BUT_OXG3, BUT_OXG4, BUT_OXG5 and BUT_OXG6 were tested for butanol production by cultivation in High Density Medium (HDM) adapted from (Sivashanmugam et al., 2009) reported to provide high level of heterologous protein expression, supplemented with a solution of amino acids, glutamate, riboflavin and iron citrate (III), according to the optimization results from Chapter 4. Medium formulation is presented in table 5.4.

.

Table 5.4: Medium composition of HDM, adapted from (Sivashanmugam et al., 2009).

Component	Amount per Liter	Unit
Glucose	10	g
Dibasic sodium phosphate dihydrate	8.89	g
Monobasic potassium phosphate	6.8	g
Sodium chloride	0.58	g
Magnesium sulphate	1.35	g
Calcium chloride dihydrate	0.038	g
Ammonium chloride	1	g
Trace metals	250	µL
Vitamins BME100x	250	µL
Amino acids mix	2	g
Glutamate	0.468	g
Riboflavin	0.07529	g
Iron (III) citrate	0.525	g

The trace metals solution contained (per liter): $FeSO_4.7H_2O$ (30 mg); $ZnSO_4.7H_2O$ (45 mg); $CaCl_2.2H_2O$ (45 mg); $MnCl_2.2H_2O$ (100 mg); $CoC_{l2}.6H_2O$ (30 mg); $CuSO_4.5H_2O$ (30 mg); $Na_2MoO_4.2H_2O$ (40 mg); H_3BO_3 (10 mg); KI (10 mg) and Na_2EDTA (1.5 g). The amino acid mix contained 1 g of adenine and 4 g of arginine, aspartate, glutamate, histidine, isoleucine, lysine, methionine, phenylalanine, serine, threonine, tryptophan, tyrosine and valine. The vitamin BME 100 x solution (Sigma Aldrich, St. Louis, MO, USA) contained (per liter): D-biotin (0.1 g); choline chloride (0.1 g); folic acid (0.1 g); myo-inositol (0.2 g); niacinamide (0.1 g); D-pantothenic acid.½Ca (0.1 g); riboflavin (0.01 g); thiamine.HCl (0.1 g) and NaCl (8.5 g). Cultivation was performed with the addition of suitable antibiotics according to the employed plasmids (50 µg.mL^{-1} ampicillin, 50 µg.mL^{-1} spectinomycin, and 30 µg.mL^{-1} kanamycin).

A single colony was picked from LB plates and inoculated in 10 mL of LB medium. The pre-cultures were grown aerobically on a rotary shaker at 37 °C and 200 rpm, overnight. Cells were washed and harvested by centrifugation (10 min at 3000 × g). Afterwards, an appropriate volume of pre-culture was transferred to 500 mL shake flasks with 100 mL of HDM medium, containing the appropriate antibiotics, yielding an initial OD_{600} of 0.1. This culture was cultivated on a rotary shaker at 200 rpm at 37 °C. The butanol production genes were induced with 0.5 mM IPTG at an OD_{600} of 0.4-0.5.

After induction, 60 mL of the culture were transferred to 120 mL sealed serum flasks to promote butanol production under anaerobic conditions. In some experiments the transfer to anaerobic conditions was delayed for 4 h or 12 h, and in that period the cells grew under aerobic conditions.

The culture was supplemented with 600 µL of a 0.01 M stock solution of sodium bicarbonate to achieve a final concentration of 10 mM (to reduce lag phases in *E. coli* anaerobic growth (Hornsten, 1995). Selenium, nickel and molybdenum are part of the formate hydrogen lyase (FHL) complex, which is induced under anaerobic conditions (Soini et al., 2008). For this reason, 60 µL of a solution of extra trace metals ($NiCl_2$ (1.7 $mg.L^{-1}$); $(NH_4)_6Mo_7O_{24}$ (14.5 $mg.L^{-1}$); $4H_2O$ Na_2SeO_3 (2.4 $mg.L^{-1}$)) was supplied to the medium.

The cultures were incubated at 30 °C and 180 rpm, for 96 hours. Samples of broth were collected at time 0, induction time and 96 h. All experiments were performed in triplicate and the samples were analyzed by High-Performance Liquid Chromatography (HPLC) and Gas Chromatography (GC).

5.2.6 Butanol production experiments in bioreactors

Bioreactor fermentations were performed in HDM medium. Cells were pre-grown overnight in 500 mL shake flasks containing 100 mL of the same medium, at 37 °C and 200 rpm. Each fermenter was inoculated at an initial OD_{600} of 0.15. The fermentations were performed in the Eppendorf DASGIP Parallel Bioreactor System (Switzerland) using 2-L culture vessels. The operating volume for the fermentations was 0.5 L, temperature was maintained at 37 °C, airflow at 1 VVM, pH was kept at 7.0 by addition of 2 M NaOH, and dissolved oxygen was kept above 30 % of saturation by feedback control of the stirring speed from 200 rpm up to 400 rpm. Expression of the butanol genes was induced with 0.5 mM of IPTG when an OD_{600} of 0.4-0.5 was reached. After IPTG induction, temperature was decreased to 30 °C and stirring speed to 180 rpm. Anaerobic conditions were created by turning off the air flow after 4 h or after 12 h and waiting for the leftover oxygen to be consumed. Samples were taken every two hours for the first 12 hours of the fermentation and, then, every 12 hours. At each time point the optical density was measured, and the supernatant was analyzed by HPLC and GC.

5.2.7 Analytical methods

Butanol was quantified by GC and organic acids, ethanol and glucose by HPLC. Samples were centrifuged at 6000 × g for 10 min to separate cells from the medium. Afterwards, the supernatant was filtered with a 0.22 µm pore filter membrane to glass vials and stored at -20 °C until analyzed.

Quantitative analysis of organic acids and glucose was performed using HPLC apparatus from Jasco (Japan) model LC-NetII/ADC equipped with UV-2075 Plus and RI-4030 Plus detectors, also from Jasco. The samples were analyzed using an Aminex HPX-87H column (300 mm x 7.7mm) from Bio-Rad, which was kept at 60 °C and 5 mmol.L^{-1} H_2SO_4 was used as mobile phase with a flow rate of 0.5 mL.min^{-1}. Glucose and ethanol were detected with a refractive index (RI) detector (4030, Jasco) and organic acids (succinate, lactate, formate and acetate) were detected at 210 nm using a UV detector (Jasco). Calibration curves were obtained by injecting standards with known concentrations for each metabolite. Metabolite concentrations in samples were calculated by comparing the peak areas of the samples with the calibration curves.

Butanol concentration was quantified by GP-9000 system (Chrompack) with a Meta-WAX capillary column (30 m X 0.25 mm X 0.25 µm) equipped with a flame ionization detector (FID); helium was used as carrier gas with a flow rate of 1 mL.min^{-1}. The filtered supernatant (900 µL) was mixed with 100 µL of a 5 g.L^{-1} solution of isobutanol, the internal standard, yielding a final concentration of 0.5 g.L^{-1}, and 1 µL of this mixture was injected. The temperature of injector and detector was maintained at 250° C. The column temperature was initially at 50° C, heated to 177.5° C at a 5° C.min^{-1} rate and then heated to 230° C at 10° C.min^{-1}, which was held for 15 minutes. A calibration curve was obtained by injecting standards with several concentrations of butanol and a fixed concentration of internal standard (0.5 g.L^{-1} of isobutanol). Butanol concentration was calculated by comparing the ratio between its peak area and internal standard peak area with calibration curves.

All cell optical density measurements at 600 nm (OD_{600}) were performed using the spectrophotometer Ultrospec 10 from Biochrom (Cambridge, UK).

5.3 Results and discussion

5.3.1 Selection of target genes to knock-out

In Chapter 3, *in silico* strain optimization algorithms suggested a series of reaction knock-outs to improve the butanol accumulation in *E. coli* expressing a novel butanol biosynthetic pathway. One of the mutants obtained provided the greatest compromise between number of knock-outs and biomass-product coupled yield. Table 5.5 shows for each inactivated reaction from the optimum simulation, the respective enzyme and encoding genes.

Table 5.5: Inactivated reactions in the optimization process to improve butanol production and minimize by-products formation (Chapter 3). The reaction ID corresponds to the i*JO1366 GSMM of* E. coli. *For each reaction the enzyme name and the genes encoding the respective enzyme are shown.*

Reaction ID	Enzyme	Reaction	Gene
R_ALCD2x	Alcohol dehydrogenase	Ethanol + NAD^+ <=> Acetaldehyde + NADH + H^+	*adhP* and *adhE* and *frmA*
R_ACALD	Acetaldehyde dehydrogenase	Acetaldehyde + CoA + NAD^+ <=> Acetyl-CoA + NADH + H^+	*mhpF* and *adhE*
R_PTAr	Phosphotransacetylase	Acetyl-CoA + Orthophosphate <=> CoA + Acetyl phosphate	*pta* and *eutD*
R_LDH_D	D-lactate dehydrogenase	(R)-Lactate + NAD^+ <=> Pyruvate + NADH + H^+	*dld* and *ldhA*

We can see that, according to the GSMM, for each reaction to be inactivated, there are alternative genes encoding enzymes capable of catalyzing the respective reaction. Considering the large amount of genomic interventions necessary to fully inactivate the target reactions, a review of databases (Metacyc, BRENDA) and relevant literature was performed to select the most relevant genes to achieve the desired goal.

AdhE is the main active alcohol dehydrogenase in *E. coli*, but *adhP* also encodes an alcohol dehydrogenase and *mhpF* an acetaldehyde dehydrogenase. *adhP* and *mhpF* were suggested as knock-out targets by other authors who deleted *adhE* and still observed significant ethanol production (Garza et al., 2012). Regarding lactate formation, *ldhA* is the only fermentative lactate dehydrogenase present in *E. coli*. Additionality *dld* encodes for a membrane-bound D-lactate dehydrogenase, but it preferably oxidizes lactate into pyruvate (Bunch et al., 1997). Finally, previous works have demonstrated that deleting *pta* prevented carbon leakage from acetyl-CoA to acetate (Ohtake et al.,

2017). EutD is involved in ethanolamine utilization and is also an isozyme of *pta*. Nevertheless, *eutD* was never considered as a target gene even in works whose main goal was to significantly decrease acetate accumulation in *E. coli* (De Mey et al., 2007; Kim et al., 2015). Acetate kinase (*ackA*) and phosphotransacetylase (*pta*) are part of the same operon and constitute the *ackA-pta* pathway with acetyl-phosphate as the intermediate. Since this metabolite can be synthesized by other enzymes (such as EutD) and consequently be used as substrate by acetate kinase, we decided to delete *ackA* as well (Klein et al., 2007)

In conclusion, it was decided that it was unnecessary to disrupt the following genes: *frmA*; *eutD* and *dld*. The remaining genes were deleted in the following order: *adhP*, *ldhA*, *adhE*, *pta*, *mhpF*, *ackA*, originating *E. coli* 1KO, *E. coli* 2KO, *E. coli* 3KO, *E. coli* 4KO, *E. coli* 5KO and *E. coli* 6KO strains.

E. coli K12 MG1655 (DE3), *E. coli* 3KO, *E. coli* 4KO, *E. coli* 5KO and *E. coli* 6KO strains were genetically modified to express heterologous enzymes capable of catalyzing the successive reactions from 2-oxoglutarate to butanol (Chapter 4) originating BUT_OXG2, BUT_OXG3, BUT_OXG4, BUT_OXG5 and BUT_OXG6, respectively.

5.3.2 Physiological characterization of the knock-out strains in serum bottles

BUT_OXG2, BUT_OXG3, BUT_OXG4, BUT_OXG5 and BUT_OXG6 were cultivated aerobically at 37 °C in a rotary shaker at 200 rpm in defined medium (HDM). The cells were grown in shake-flasks until reaching the mid-exponential phase (0.4-0.5 OD_{600}). At that moment, IPTG was added to the medium and cells were either immediately, after 4 h or after 12 h, shifted to sealed serum bottles and grown for 96 h at 30 °C and 180 rpm. The growth-profile graphs for all strains in the different conditions tested are depicted in Figure 5.2.

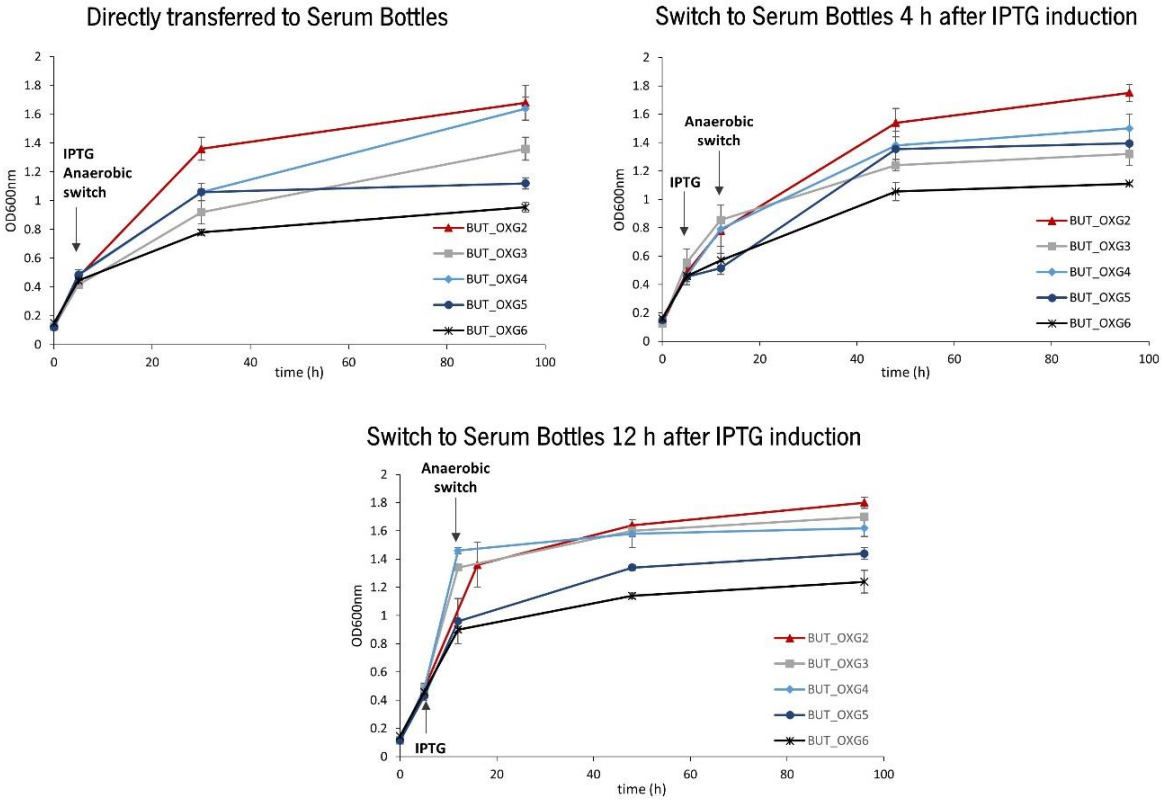


Figure 5.2: Growth-curve profile of strains BUT_OXG2, BUT_OXG3, BUT_OXG4, BUT_OXG5 and BUT_OXG6 by immediately, after 4h or after 12 h transferring to serum bottles after IPTG induction. Cells were cultivated in HDM medium and induced with 0.5 mM IPTG. Data are shown as mean ± SD of three independent experiments.

Analyzing the growth-profiles of the different strains, it is possible to see that the final cell density increased for all the knock-out mutant strains when the switch to anaerobic conditions was delayed. However, this behavior was not observed in the wild-type strain, BUT_OXG2, in which, for the three conditions tested, the final values of cell density overlap within the confidence intervals (1.68±0.10; 1.75±0.06 and 1.8±0.04). However, regarding the knock-out strains, it seems that the biomass growth was not supported under anaerobic conditions by the heterologous pathway. Hence, by increasing the biomass at the moment of the anaerobic switch, the cells were able to achieve higher final values of cell density.

Although under aerobic conditions and before IPTG induction, the different strains achieved mid-exponential phase at the same time (around 4 h), the growth-profile, even under aerobic conditions, is distinct after this moment. So, we may conclude that the metabolic burden imposed by IPTG induction had more effect on the strains with higher the number of disrupted genes, affecting the production of biomass.

The main goal of developing the knock-out mutant strains was to maximize the production of butanol and to minimize the by-products formation. For this reason, metabolites accumulation during anaerobic phase are summarized in Table 5.6.

Table 5.6: Physiological properties during anaerobic growth on the fermentation products of the strains engineered for butanol production after 96 h.

Strain	Anaerobic switch	Ethanol ($g.L^{-1}$)	Succinate ($g.L^{-1}$)	Lactate ($g.L^{-1}$)	Formate ($g.L^{-1}$)	Acetate ($g.L^{-1}$)
BUT_OXG2	0 h	0.070±0.002	0.040±0.009	3.670±0.102	-	0.550±0.005
	4 h	0.080±0.002	0.611±0.055	3.210±0.052	-	0.618±0.013
	12 h	0.200±0.001	0.965±0.021	3.110±0.064	-	1.031±0.052
BUT_OXG3	0 h	-	0.727±0.033	-	0.450±0.006	0.457±0.016
	4 h	-	0.698±0.024	-	0.433±0.003	0.460±0.080
	12 h	-	0.622±0.005	-	0.562±0.003	0.659±0.023
BUT_OXG4	0 h	1.240±0.003	0.645±0.039	0.132±0.001	0.151±0.003	0.901±0.064
	4 h	0.950±0.004	0.590±0.023	0.171±0.005	0.234±0.001	0.986±0.023
	12 h	0.610±0.004	0.574±0.026	0.174±0.003	1.125±0.005	1.131±0.090
BUT_OXG5	0 h	-	0.757±0.053	0.426±0.002	3.569±0.003	0.734±0.021
	4 h	-	0.724±0.045	0.414±0.001	3.593±0.006	0.742±0.031
	12 h	-	0.638±0.010	0.335±0.003	3.675±0.004	0.767±0.027
BUT_OXG6	0 h	-	0.815±0.023	0.547±0.005	0.123±0.003	0.225±0.023
	4 h	-	0.847±0.032	0.553±0.005	0.137±0.003	0.242±0.080
	12 h	-	0.805±0.034	0.554±0.006	0.113±0.002	0.271±0.016

From Chapter 3, the optimum genotype of the *in silico* simulation required the inactivation of the reactions catalyzing the production of ethanol, acetate and lactate (Table 5.5). In this mutant, from a glucose uptake of 10 $mmol.(g_{DW}.h)^{-1}$, the main excreted products were succinate (0.11 $(mmol.(g_{DW}.h)^{-1})$, acetate (5.01 $(mmol.(g_{DW}.h)^{-1})$, formate (10.27 $(mmol.(g_{DW}.h)^{-1})$, valine (1.83 $(mmol.(g_{DW}.h)^{-1})$, CO_2 (6.58 $(mmol.(g_{DW}.h)^{-1})$ and butanol (4.94 $mmol.(g_{DW}.h)^{-1}$). We can observe that, contrarily to what was predicted by the *in silico* simulations, the disruption of the genes did not totally abolish the production of ethanol and lactate.

Analyzing the phenotypic results from Table 5.6, we can see that in the *adhP*, *ldhA* and *adhE* triple knock-out mutant, BUT_OXG3, ethanol and lactate were not detected. In these mutant strains, two

genes encoding enzymes responsible for ethanol accumulation (*adhP* and *adhE*) and one for lactate production (*ldhA*) were disrupted. Within the detected products, succinate is the only one that allows NADH recycling. So, it appears that the excess of NADH was recycled by redirecting the carbon flux to the production this compound. On the other side, the cells accumulated formate, contrarily to what was observed in the reference strains BUT_OXG2. By impairing the production of lactate from pyruvate, a higher flux of this metabolite can be converted by PFL into formate and acetyl-CoA, whose conversion into acetate through *ackA-pta* pathway yields one molecule of ATP.

BUT_OXG4 was developed by deleting the phosphotransacetylase (*pta*) from the previous triple knock-out mutant, a gene encoding the conversion of acetyl-CoA into acetyl-phosphate, which is then converted to acetate by acetate kinase (*ackA*). Contrarily to what we expected, acetate was still produced in BUT_OXG4, even in higher concentrations than BUT_OXG3, as well as lactate and ethanol in all the conditions tested. Acetate can either be produced by pyruvate oxidase (*poxB*) or from acetyl-CoA via acetyl phosphate by the *pta-ackA* pathway, which is reversible and constitutively expressed (De Mey et al., 2007; Phue et al., 2010). So, acetate accumulation in Δ*pta* mutants, can be explained by the activation of other genes such as *poxB*, which is usually more active under aerobic conditions, and *ackA* (Chang et al., 1999). Nevertheless, even in *ackA*, *pta*, *poxB* triple knock-out mutants, acetate (in lower amounts) was reportedly still produced (Phue et al., 2010). In the referred study, the authors have concluded that acetate accumulation, besides as a by-product of the central carbon metabolism, can be due to its production via anabolic pathways, by the action of enzymes such as acetylornithine deacetylase (*argE*), cysteine synthases (*cysM* and *cysK*) and acetoacetyl-CoA transferases (*atoA* and a *toD*) (Phue et al., 2010). Moreover, in the reaction catalyzed by the glutaconate-CoA transferase, Coenzyme A is transferred from acetyl-CoA to 2-hydroxyglutarate yielding 2-hydroxyglutaryl-CoA and releasing acetate. In fact, the FBA simulation of the optimum genotype indicated the production of acetate only through this reaction. Since in BUT_OXG4 genes encoding two alcohol dehydrogenases (*adhP* and *adhE*) were disrupted, the production of ethanol could be explained either by the promiscuous activity of the heterologous gene *adhE* or by the native gene *mhpF*. By knocking-out this last gene, we have developed BUT_OXG5 and, as result, in this strain ethanol was not accumulated. So, it appears that in BUT_OXG4, this was the enzyme responsible for ethanol accumulation. The accumulation of lactate in BUT_OXG4, although much lower than the values obtained in the wild-type, increased after *pta* deletion. So, we may assume that the *pta* disruption led somehow to a rewiring of the flux and possible activation of latent pathways. *E. coli* expresses three lactate dehydrogenases, one fermentative (encoded by the gene *ldhA*,) and

two respiratory (encoded by *lldD* and *dld*) (Toyoda et al., 2009). Although, reportedly, *ldhA* is the only fermentative D-lactate dehydrogenase, its inactivation is often insufficient to totally abolish lactate production under anaerobic conditions (Atsumi et al., 2008a; Causey et al., 2003; Zhou et al., 2014). The other two lactate dehydrogenases are membrane-bound. The one encoded by *dld* gene is D-(+)-chiral-specific, while *lldD* is L-(+)- chiral-specific and catalyzes the conversion of lactate into pyruvate. Considering that the lactate detected was chiral-specific D-(+)-, we assume the enzyme Dld catalyzed this reaction. Since the deletion of acetate pathways causes pyruvate accumulation (Dittrich et al., 2005), it is possible that after deleting *pta*, an accumulation of pyruvate has driven the reaction in the opposite direction and resulted in the minor accumulation of lactate observed.

Regarding the quintuple gene knock-out mutant, BUT_0XG5, formate was the major product accumulated (achieving titers higher than 3 $g.L^{-1}$ in the tree conditions tested) and ethanol was not detected. Since after disrupting *ackA*, formate levels drastically decrease, we may assume that the formate accumulation was due to the PFL activity. Apparently, acetyl-phosphate was being formed by an alternative enzyme to Pta and, subsequently, converted into acetate by activity of acetate kinase. Disrupting *ackA*, implies that the catalytic step that releases ATP is impaired, turning the conversion of pyruvate into formate and acetyl-CoA no more metabolically advantageous.

Nevertheless, the disruption of *ackA* in BUT_0XG6 was insufficient to cease acetate production, so we may conclude, as discussed before, that *poxB* or anabolic pathways were partially responsible for acetate production. We also observed that, in these strains, succinate was the major accumulated product, contrarily to what was computationally simulated, and lactate accumulation increased. Under anaerobic conditions, cells do not induce the full TCA cycle. In this branched version, succinate is produced via a reductive route. In some situations, a glyoxylate bypass is activated and succinate can be synthesized through this pathway (Skorokhodova et al., 2015).

Regarding lactate accumulation in BUT_0XG6, considering, as discussed above, that lactate production was due to the activity of Dld, the higher availability of pyruvate (as a consequence of the lower formate accumulation) can explain this increment.

The different combinations of gene knock-outs resulted in distinct phenotypes, but for the strain with all the genetic interventions (BUT_0XG6) the phenotype was quite different from the *in silico* predictions. Even though the flux unit values obtained computationally are distinct from the experimental ones, a qualitative comparative analysis is still possible. In this regard, only ethanol

production was completely abolished by the knock-outs performed, while lactate was still produced. Formate and succinate were accumulated, as predicted. Nevertheless, in the sextuple knock-out mutant (BUT_OXG6), succinate was the major accumulated product, in the model predictions, succinate was the product accumulated in minor concentrations. We would expect that the deletion of the target genes for the production of ethanol (*adhE*, *adhP*, *mhpF*), lactate (*ldhA*) and acetate (*pta*, *ackA*) would impair or, at least, greatly reduce the respective product formation. However, it seems the higher the number of disrupted genes, the more unpredictable the phenotype is. Usually, the disruption of one or two of the genes active under anaerobic conditions, implies an increment of the metabolic flux through the alternative native pathways. An approach applied in the individual production of succinate, ethanol and lactate (Long et al., 2018). All the genetic perturbations imposed in these strains possibly forced a rewire of the flux in the cell resulting in a phenotype with compromised growth under anaerobic conditions.

In this regard, after an external intervention such as gene deletion, microorganisms can adjust their functionality using diverse strategies, such as: activation of silent genes or down-regulated pathways; redistribution of metabolic fluxes; local bypass of the deleted reaction; or conversion of the deleted reaction by other enzymes (Long et al., 2018). So, without additional data, it is hard to disclose which were the metabolic alterations provoked by disrupting genes. Techniques such as Metabolic Flux Analysis using labelled carbon (^{13}C-MFA) could give insight on the central carbon metabolism rewiring of the gene-deleted mutant strains (Long and Antoniewicz, 2014). Transcriptomics can also help to identify the genes overexpressed in the different knock-out mutants by comparing with the reference strain (Park et al., 2007).

5.3.3 Butanol production experiments in serum bottles

The *in silico* analysis (Chapter 3) indicated that the knock-out of the NADH-competing pathways (namely ethanol, acetate and lactate) would lead to an excretion of 4.94 mmol.(g_{DW}.h)$^{-1}$ of butanol from an initial uptake flux of 10 mmol.(g_{DW}.h)$^{-1}$ glucose. In fact, this is the application of the best result from the optimization process, in which the synthesis of the target product was maximized, and by-products formation minimized. For the same experiments described before, butanol final concentrations were also quantified. The results obtained for the different strains are depicted in Figure 5.3.

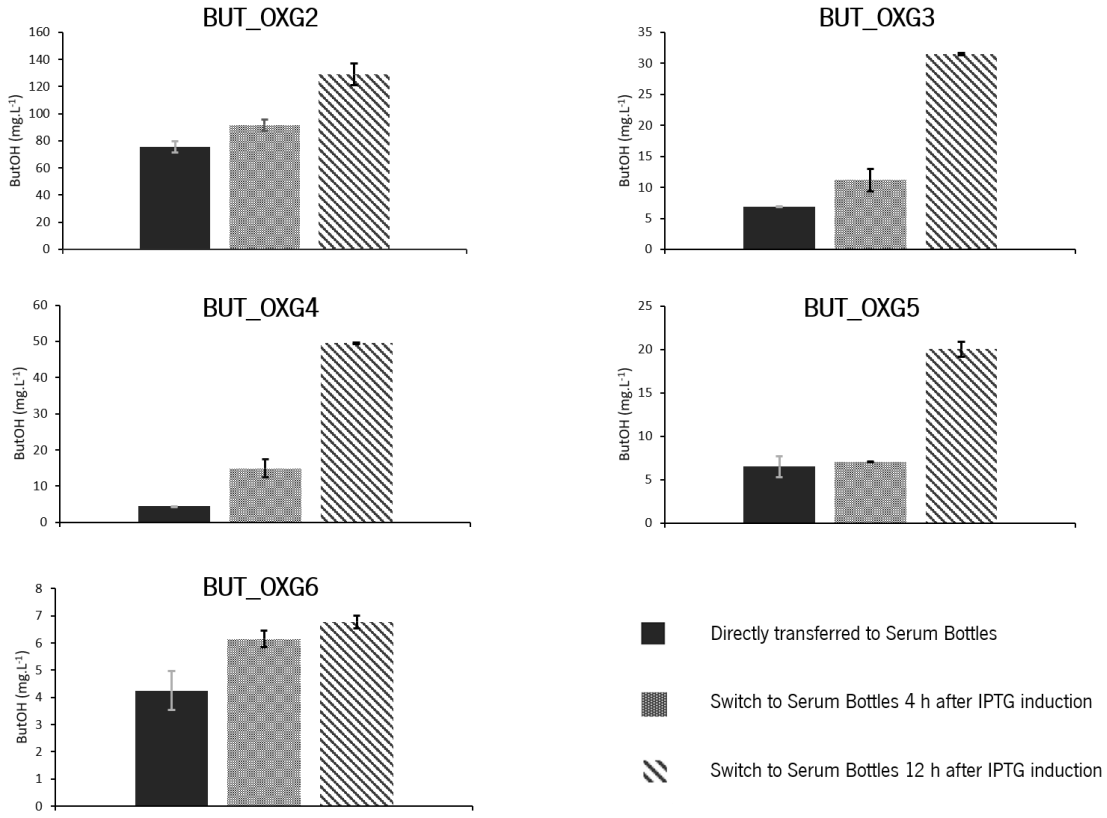


Figure 5.3: Final extracellular butanol titer (mg.L^{-1}) for strains BUT_OXG2, BUT_OXG3, BUT_OXG4, BUT_OXG5 and BUT_OXG6 by culturing in HDM medium. Cells were grown in shake-flasks at 37 °C until achieving 0.4-0.5 OD_{600}, when 0.5 mM IPTG was added, and cells were transferred immediately, after 4 h or after 12h to sealed serum bottles. Data are shown as mean ± S.D from three independent experiments. ButOH - Butanol

It is possible to observe that, for the knock-out mutants, very low butanol titers were obtained by immediately switching the cells to anaerobicity, namely 6.91±0.01 mg.L^{-1} for BUT_OXG3; 4.42±0.01 mg.L^{-1} for BUT_OXG4; 6.49±1.23 mg.L^{-1} for BUT_OXG5 and 4.25±0.72 mg.L^{-1} for BUT_OXG6, comparing with the wild-type strain (75.32±4.21 mg.L^{-1}). By deleting the competing pathways, the major native alternatives to recycle NADH were blocked, so we expected an increase on butanol accumulation. However, the gene knock-outs had the opposite effect, leading to lower final butanol titers when compared with BUT_OXG2 strains. The low butanol titers obtained by immediately switching cells to the serum bottles indicate that the oxygen left in the medium before the switch was insufficient to allow enough heterologous protein expression. In accordance to what was observed in Figure 5.2, the low butanol yields support the observation that the butanol production pathway could not compensate the role of the deleted fermentative pathways in sustaining anaerobic growth, possibly because there was insufficient presence of the heterologous enzymes.

Increasing the delay of the anaerobic switch after IPTG induction (similarly to what was previously observed for BUT_OXG1 and BUT_OXG2 in Chapter 4) had a positive impact on butanol final titer for all strains. Nevertheless, the correlation between increment on butanol titer and increasing delay of anaerobic switch was not linear for all the strains tested. For instance, in the case of *adhP*, *ldhA*, *adhE* triple mutant, BUT_OXG3, the 4 h and the 12 h gap led to, respectively, a 1.62-fold and a 4.55-fold increment on butanol production comparing with the immediate transfer to serum bottles. Regarding BUT_OXG4 (in which *pta* was also deleted), the maximum butanol titer was achieved by delaying the anaerobic switch for 12 h (49.52±0.23 $mg.L^{-1}$), representing a 11.2-fold improvement relative to the direct transfer to serum bottles after IPTG induction. Delaying for 4 h also improved the butanol titer, but in a smaller rate (3.4-fold).

For BUT_OXG5, the increment on butanol titer was 1.1-fold and 3.1-fold for an interval of 4 h and 12 h, respectively, between IPTG induction and switch to anaerobic bottles. Finally, for the strain with more knocked-out genes, the effect of delaying the anaerobic switch for 4 h and 12 h had the less impact on butanol titer with an increment of 1.44-fold and 1.59-fold, respectively.

By delaying the interval between IPTG induction and anaerobic switch, it was expected that the gene-deleted mutant strains would be able to express the enzymes catalyzing the butanol heterologous pathway. Since in these strains the main mixed acid fermentation pathways are impaired, the anaerobic growth would be supported by the butanol pathway. It seems that for the quintuple and sextuple knock-out strains, the longest interval tested, either was insufficient or the gene knock-outs led to severe growth impairment under anaerobic conditions.

Besides the metabolic effect of disrupting native genes, the overexpression of non-native pathways can also disturb the host's metabolism, by competing for precursors necessary for growth or maintenance. In addition, pathway engineering often leads to imbalanced gene expression, creating a bottleneck in the biosynthetic pathway that reduces production of the target product (Shen and Liao, 2008).

Nevertheless, the strategy here applied has been previously implemented in strains of *E. coli* expressing the butanol production pathway from clostridia, with reported increments on butanol titer by deleting mixed-acid fermentation pathways. For instance, in the work developed by Atsumi et al., the disruption of several genes (*adhE*, *ldhA*, *frdBC*, *fnr* and *pta*) increased 2.6-fold butanol titer and reduced the accumulation of acetate, ethanol and lactate (Atsumi et al., 2008a). Liao group

developed the *E. coli* strain able to achieve the maximum butanol titer so far (30 $g.L^{-1}$) and also observed a 4-fold improvement on butanol accumulation after knocking-out the genes *adhE*, *ldhA*, *frdBC* an *pta*. In this study, an enzyme converting formate into CO_2 and NADH was also expressed to provide extra NADH (Shen et al., 2011). It is important to point out that the clostridial pathway is quite different from the novel butanol production pathway used here. One of the main differences between the two butanol production pathways is the initial precursor used: acetyl-CoA for the clostridial pathway and 2-oxoglutarate for the pathway implemented here. By analyzing the final products accumulation (Table 5.6) and comparing with the butanol titers (Figure 5.3), it appears that the knock-out mutant strains with the greatest butanol titer (BUT_OXG4) had also produced other mixed-acid fermentation products, such as ethanol, lactate, formate and acetate. This strongly indicates that the heterologous pathway was not sufficient to support growth under anaerobic conditions and the accumulation of native fermentation products led to higher cell densities and improved butanol final titers.

In fact, NADH availability in anaerobiosis should not be a limiting factor for butanol production after deleting the NADH-competing pathways. Therefore, since the precursor of this novel pathway is 2-oxoglutarate, an intermediate in the TCA cycle, it seems that the bottleneck of butanol production relies in the 2-oxoglutarate concentration. As already shown before (Chapter 4), 2-oxoglutarate concentration fluctuates according to several factors and, by supplying amino acids and extra glutamate, its concentration appears to be positively affected. Nevertheless, if the metabolic flux through the oxidative pathway of the TCA cycle is reduced, the concentration of 2-oxoglutarate will concomitantly decrease.

Although the glyoxylate bypass is usually only active in *E. coli* under aerobic conditions and in the presence in the medium of carbon sources like acetate (De Mey et al., 2007), there are reported evidences of its activation under anaerobic conditions after genetic manipulations as gene disruptions (Skorokhodova et al., 2015; Zhu et al., 2013). In fact, the activation of the glyoxylate pathway is a common strategy to improve succinate production in *E. coli* during fermentation (Li et al., 2013). The activation of this pathway in non-adaptive evolved strains with an inefficient flux distribution was also found. This pathway was activated and, after adaptive evolution, its flux decrease and the growth rate was restored (Long et al., 2018). The glyoxylate bypass is characterized to skip the decarboxylation steps, including the formation of 2-oxoglutarate. After the genetic interventions implemented, the cells had to adapt their physiology to the loss of important anaerobic genes, which

can lead in some cases to lower efficiency metabolic adjustments and inefficient growth (Wolfe, 2005). Considering that succinate accumulation observed in Table 5.6 resulted from glyoxylate bypass, this could be an explanation for the slow growth rates observed in Figure 5.2.

The synthesis of 2-oxoglutarate via the oxidative branch of the TCA cycle shares the same initial precursors (acetyl-CoA and oxaloacetate) of the glyoxylate bypass. Assuming that the activation of glyoxylate shunt contributed to the accumulation of succinate in the mutant strains (Table 5.6), a lower metabolic flux will be redirected to the oxidative branch of the TCA cycle. Since the oxidative branch of the TCA cycle is only active in anaerobic conditions to supply biomass precursors, it is not surprising that the reduced flux to 2-oxoglutarate and consequently lower values of butanol production (Figure 5.3) are associated with the severe growth phenotypes observed in the multiple knock-out strains (Figure 5.2). Probably, in these strains, even after disrupting NADH-competing pathways, the butanol pathway was not be able to support cellular growth under anaerobic conditions (by recycling the excess of NADH) due to the reduced pool of the respective precursor: 2-oxoglutarate.

An ALE experiment could contribute to a more efficient flux distribution in the TCA cycle, recovering the growth deficit and, simultaneously, to increase the 2-oxoglutarate pool, fine-tuning the heterologous pathway to support growth under anaerobic conditions. In fact, considering the results summarized in Table 5.6, we may conclude that the metabolic model failed to accurately predict the phenotype of the knock-out mutant strains. Despite the reported limitations of GSMMs, (for instance, these models do not consider kinetic and regulatory effects (Rocha et al., 2008)), the similarity between *in silico* predictions and experimental results could be improved by subjecting the developed strains to adaptive laboratory evolution (ALE) experiments. ALE is a powerful technique in which a microorganism is cultivated continuously for several generations, improving its fitness in a response to a certain condition by natural selection (Fong et al., 2005). In this regard, often *in silico* designs are coupled with ALE experiments, since this approach approximates the behavior of the microorganism, to an optimum growth phenotype, resulting in a better match between experimental growth phenotypes and model predictions (Fong et al., 2006).

In fact, the predicted *in silico* growth phenotype of gene-knock-out strains reportedly coincide with the end-point of the evolution process (Fong et al., 2006). In the constraint-based computational models, simulations are often performed by assuming optimal growth rates for the defined environment and genetic modifications, directly accounting for the fundamental nature of adaptive evolution. (Ibarra et al., 2002)

ALE can increase the growth rate by allowing that each strain fine-tunes its metabolic functionality in response to the disrupted genes (Fong et al., 2005). Since the *in silico* optimization process was performed using as an objective function the biomass-product coupled yield, by improving growth rate trough ALE experiments, the butanol titer (coupled to growth rate) would also be improved.

5.3.4 Bioreactor cultivations

Since, among the gene-deleted mutant strains, the quadruple mutant, BUT_OXG4 achieved the maximum production of butanol in the serum-bottle experiments, we have selected this strain to be cultivated under controlled conditions. We have tested a delay of 4 h and 12 h between IPTG induction and anaerobic switch.

BUT_OXG4 strains were cultivated with 0.5 L-working volume in a 2 L-bioreactor in HDM medium. Samples were taken to measure cell density and for HPLC and GC-FID analysis. Figure 5.4 summarizes the obtained results.

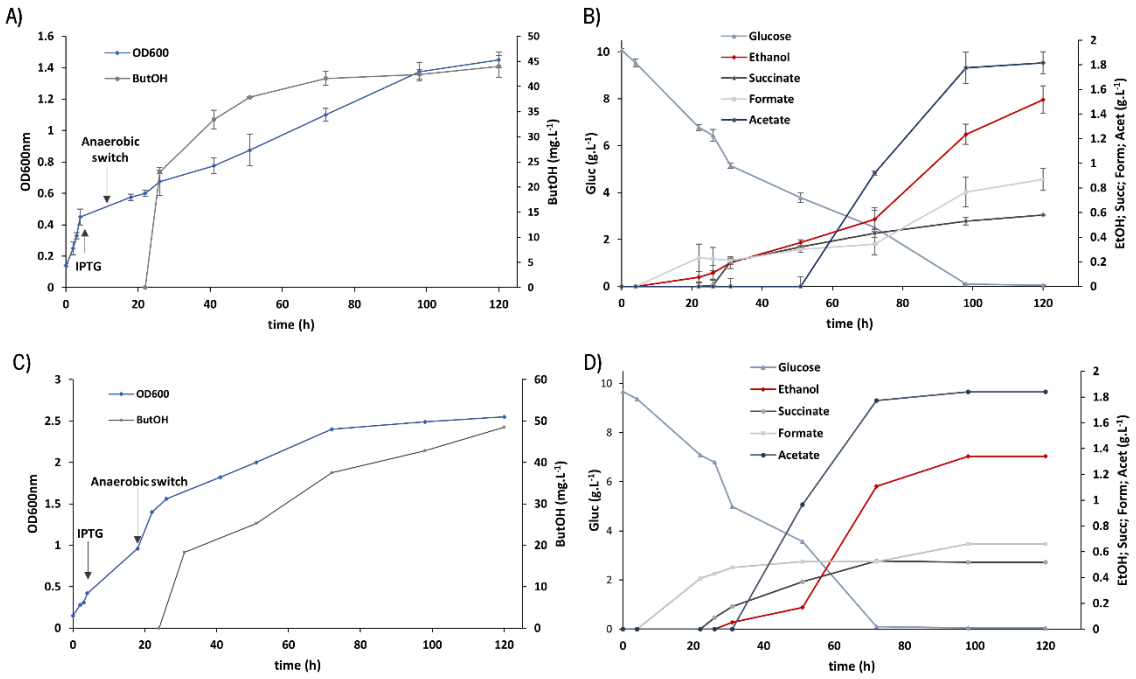


Figure 5.4: Results of fermentation in bioreactors for strain But_OXG4. The top results correspond to a delay of 4 h between IPTG induction and anaerobic switch, while the bottom results were obtained by postponing for 12 h. Growth-curve profile and butanol production (A) and (C); and glucose consumption profile and end-products formation (B) and (D). Cells were cultivated in HDM medium and induced with 0.5 mM of IPTG. Data for A) and B) are shown as mean ± SD of two independent experiments. ButOH: Butanol; EtOH: Ethanol; Succ: Succinate; Acet: Acetate; Form: Formate.

Observing the butanol production profile obtained by delaying the anaerobic switch for 12 h (Figure 5.4 C)), it is possible to observe that cultivating the cells under controlled conditions achieved around the same maximum butanol accumulation (48.52 $mg.L^{-1}$), than in serum bottles (49.52±0.23 $mg.L^{-1}$).

Also, under controlled conditions, the impact on butanol accumulation, of increasing the switch to anaerobic conditions for 4 h or 12 h, was much less noticeable than in the serum bottle experiments, where the increment was 3.31-fold higher (Figure 5.3). In bioreactors, the gap of 12 h led to a butanol titer 1.1-fold higher (44.01 $mg.L^{-1}$), than of 4 h (48.52±0.99 $mg.L^{-1}$).

Similarly to what was observed for BUT_OXG2 strains (Chapter 4), a growth-coupled butanol synthesis behavior was evident for both conditions. The cells that were able to grow for a longer period under aerobic conditions achieved a final higher cell density, OD_{600}=2.55, over OD_{600}=1.45±0.05. Even though the cells were able to achieve higher densities by increasing the interval between induction and anaerobic switch, this was not reflected on final butanol titer.

In serum bottles, BUT_OXG4 strains able to grow under aerobic conditions after IPTG induction for 4 h and 12 h, achieved a final cell density of 1.5±0.1 and 1.62±0.06, respectively (Figure 5.2). Regarding glucose consumption, its exhaustion was faster by delaying the anaerobic switch for 12 h (at 72 h) than for 4 h (at 100 h). However, in serum bottle experiments, these strains were not able to extinguish glucose (data not shown). Hence, by cultivating these strains in bioreactors, the increasing growth under aerobic conditions seems only to impact biomass formation, and this increment had no effect on butanol final titer.

Since the major product accumulated by BUT_OXG4 strains was acetate, which lowers the pH of the medium, we may assume that in serum bottle experiments the strains tested could not achieve higher biomass values due to growth impairment. Also, in the bioreactor experiments, lactate was not detected, contrarily to what was observed in serum bottles (Table 5.6). Usually, lactate production is induced at low pH (Higgins and Johnson, 1970); this could explain the fact that lactate was only accumulated in serum bottles, and not under pH controlled conditions. However, in this mutant the fermentative lactate dehydrogenase (*ldhA*) was disrupted and, as discussed before, we assume the protein product of *dld* gene was responsible for the lactate accumulation. Hence, the regulatory mechanisms that led to the lactate excretion need to be further investigated.

The growth rate (μ), the duplication time (t_d) and the yields on substrate for the different products (butanol, ethanol, succinate, acetate and formate) detected in the broth are summarized in Table 5.7.

Table 5.7: Specific growth rate (μ), duplication time (td) and butanol. ethanol, succinate, acetate and lactate yield (Y) on glucose. BUT_OX4 cells were cultivated under controlled conditions in 2 L bioreactors in HDM medium for 120 h. Gluc: Glucose; ButOH: Butanol; EtOH: Ethanol; Succ: Succinate; Acet: Acetate; Form: Formate

Parameter	Delaying of anaerobic switch after IPTG induction	
	4 h	12 h
μ (h^{-1})	0.009±0.001	0.013
t_d (h)	79.9±2.8	46.2
$Y_{ButOH/S}$ ($mg_{ButOH} \cdot g_{Gluc}^{-1}$)	4.38±0.21	5.04
$Y_{EtOH/S}$ ($g_{EtOH} \cdot g_{Gluc}^{-1}$)	0.151±0.011	0.139
$Y_{Succ/S}$ ($g_{Succ} \cdot g_{Gluc}^{-1}$)	0.057±0.001	0.054
$Y_{Acet/S}$ ($g_{Acet} \cdot g_{Gluc}^{-1}$)	0.180±0.002	0.191
$Y_{Form/S}$ ($g_{Form} \cdot g_{Gluc}^{-1}$)	0.086±0.009	0.068

As expected, in the experiments where the switch to anaerobic conditions was delayed by 12h, the growth rate was higher (0.0127 h^{-1}) than by 4 h (0.009±0.0001 h^{-1}), since cells were able to grow aerobically for a longer period of time. Nevertheless, for both conditions tested, the growth rates were still low and, consequently, the respective duplication time was quite long (46.2 h and 79.9±2.8 h). The yields on butanol were the lowest values within the other mixed-acid fermentation products detected (namely ethanol, succinate, acetate and formate). This reinforces the conclusion that the butanol heterologous pathway is insufficient to support the anaerobic growth of the mutant stains and, probably, 2-oxoglutarate pool is insufficient to work as a driving force to push the carbon flux through the heterologous pathway.

Comparing both conditions, the native fermentation products yield on substrate were lower for all the analyzed compounds (with exception of acetate), when the cells were able to grow 12 h anaerobically after IPTG induction. However, the butanol yield on glucose was 1.15-fold higher, reinforcing the assumption that the delay of anaerobic switch allowed a better heterologous protein

expression. As discussed before, ALE experiments could improve growth rate and butanol production by improving the growth rate of the multiple knock-out strains.

5.4 Conclusions

After analyzing the obtained results, we may conclude that the gene knock-outs implemented in *E. coli* K12 were not sufficient to abolish and decrease the production of lactate and acetate, respectively, as predicted by the *in silico* FBA simulations, using the GSMM *i*JO1366 (Table 5.6). Consequently, the carbon flux was not redirected through butanol heterologous pathway. The maximum titers obtained for the wild-type were greater than the ones obtained with gene knock-out mutant strains (Figure 5.3).

As already observed by cultivating BUT_OXG2 strains, the wild-type strain, butanol accumulation was incremented by delaying the anaerobic switch. This was observed for all the strains, nonetheless the increment level on butanol accumulation was not similar to all the strains (Figure 5.3). By delaying the switch to serum bottles by 12 h, BUT_OXG4 strains were able to increase the respective butanol titer in 11.2-fold, achieving the maximum value within gene-deletion mutant strains (49.52±0.23 $mg.L^{-1}$). Since in this strain other mixed-acid fermentation products were also accumulated, we may conclude that, more than providing a NADH sink, we need to increase the flux through the oxidative branch of the TCA cycle and, consequently, increment the 2-oxoglutarate pool to increase butanol titers.

Regarding the bioreactor experiments by cultivating BUT_OXG4 strains (Figure 5.4 and Table 5.7), no increment on butanol titer was observed by delaying the anaerobic switch for 12 h, comparing with serum bottles. On the other side, in the experiments where the anaerobic switch was delayed only for 4 h, a final butanol titer of 44.01±2.21 $mg.L^{-1}$ was achieved, representing a 2.95-fold increment.

Further optimization to achieve industrial requirements is still needed. Metabolomics analysis could provide a better picture on the metabolism of the gene-deleted mutant strains and help determining the rate-limiting steps. Also, with the increasing number of whole genome sequences, often novel enzymes sequences become available and testing alternative genes could also be a hypothesis.

The *in silico* predictions for the multiple knock-out mutants proved to be quite different from the results obtained experimentally. One of the reasons for the big disparity can be the results of the optimum growth assumed in the simulations using CBM. When simulations are performed for the multiple knock-out strains, the algorithm assumed that the butanol production pathway could take over the NADH recycling role in anaerobic conditions. Although stoichiometrically this is true, in reality there are other factors that were not taken into account, such as enzyme kinetics and substrate availability. Often the metabolic predictions obtained with CBM are only matched experimentally after evolving the mutant strains for long periods to recover the growth deficit. Therefore, the strains build here might require ALE to achieve their full butanol production potential.

In parallel, another good option to improve the butanol production of these strains, would be to integrate the butanol production genes into the genome under the control of native fermentation regulatory elements (FRE) of the major fermentative genes, in the same fashion as previously reported for clostridial butanol pathway (Wen and Shen, 2016). This would allow to construct self-regulated butanol production system in *E. coli*, a strain able to auto-induce butanol production under anaerobic conditions and, simultaneously, with the native fermentation alternatives impaired. In fact, from an industrial-perspective, this would be much more economical, avoiding the use of IPTG.

CHAPTER 6 Development of *Escherichia coli* butanol producing strains by engineering the clostridial native pathway

The potential of engineering new butanol producing microorganism has been evaluated by transferring the clostridial pathway to different hosts. *Escherichia coli* presented one of the most promising results within the recombinant producers; nevertheless, several efforts have been made to improve butanol titers. Feasible strategies include testing alternative enzymes (more suitable to the host) and/or engineering the expression system.

In this work, we have designed strains of *E. coli* able to produce butanol using engineered clostridial pathways by addressing two identified rate-limiting steps. For the first step of this pathway, we have replaced the native thiolase from *Clostridium* (*thl*) for *phbA*, from *Cupriavidus necator*. We also tested two enzymes for the reduction of crotonyl-CoA into butanoyl-CoA: the clostridial complex *bcd-etfAB* and *ter* from *Treponema denticola*. Finally, we evaluated the effect of using a single or an individual promoter to express the genes constituting the BCS (butyryl-CoA synthesis) operon.

We were able to successfully implement the butanol pathway variations in *E. coli*, but the titers were still low. Contrarily to the results obtained in chapter 4, BL21 hosts outperformed the K12 equivalents in terms of butanol accumulation in all the tested designs. Moreover, the use of Terrific Broth (TB) favored butanol production comparing with the defined media HDM.

The maximum butanol titer was 58.41±2.88 $mg.L^{-1}$ by cultivating *E. coli* BL21 strains expressing *phbA*, *hbd*, *crt*, *ter* and *adhE* in TB medium. Further optimization is still required to achieve significant butanol accumulation.

6.1 Introduction

Butanol, together with acetone and ethanol, is naturally produced by some clostridia species through a process named ABE fermentation. Although clostridia-based bioprocesses have been widely used for butanol production, some challenges remain nowadays. The increasing butanol global market and environmental awareness have stimulated the interest into engineering microbial cell factories for butanol production. The lack of efficient genetic tools for engineering clostridia impairs the development of more efficient processes based on this microorganism (Zheng et al., 2009). An alternative strategy to using natural butanol producers relies on the expression of the butanol biosynthetic pathway in a biological work-horse such as *Escherichia coli*. For the first time in 2008, the Liao group successfully implemented the clostridial butanol production pathway in *E. coli*, achieving a maximum titer below 1 $g.L^{-1}$, even after disrupting competing pathways (Atsumi et al., 2008a). Since then, several efforts were reported in order to increase butanol titers in *E. coli*. Factors such as the redox balance, identification of rate-limiting steps, the limited activity of some enzymes and the gene expression system must be taken into consideration when developing recombinant butanol-producing strains. Therefore, efficient butanol production in *E. coli* requires more than simply expressing several clostridial genes (Dong et al., 2016).

The butanol clostridial pathway is constituted by six steps catalyzed by five different enzymes. It starts with the condensation of two molecules of acetyl-CoA into one of acetoacetyl-CoA, catalyzed by a thiolase (*thl*). Followed by its dehydration into 3-hydroxybutyryl-CoA – catalyzed by 3-hydroxybutyryl-CoA dehydrogenase (*hbd*). Then, this compound is reduced into crotonyl-CoA by the action of crotonoyl-CoA reductase (*crt*). The butyryl-CoA dehydrogenase/electron-transferring flavoprotein complex (*bcd-etfAB*) reduces crotonyl-CoA into butyryl-CoA. Lastly, butyryl-CoA is reduced to butyraldehyde and then to butanol, by the bifunctional aldehyde/alcohol dehydrogenase(*adhE*) (Chen and Liao, 2016).

The first and the fourth reactions are reported to be rate-limiting steps when expressing this pathway in *E. coli*. Since the first reaction is thermodynamically unfavorable (Ohtake et al., 2017) and the activity of the *bcd-etfAB* complex requires the native ferredoxin system from *Clostridium*, complicating its expression in *E. coli* (Shen et al., 2011).

The thermodynamics of a reaction is not dependent on the enzyme catalyzing it. Instead, the concentration of the reactants and products has the biggest impact on the directionality and rate of the reaction. Regarding the first step of the butanol production pathway in clostridia (thiolase), the increase in acetyl-CoA availability can drive the reaction towards butanol production. Nevertheless, the activity of different thiolases has been tested in several hosts and expression systems. *E. coli* expresses two 3-ketoacyl-CoA thiolase proteins: *atoB* and *fadA.* Only the first has been tested for butanol synthesis, resulting in higher butanol titers when replacing the clostridial thiolase (Atsumi et al., 2008a). In two other publications, the expression of the *phaA* gene from *Cupriavidus necator* was tested and proved to be a functional alternative (Bond-watts et al., 2011).

Regarding the fourth step, by replacing *bcd-etfAB* with the *ccr* gene from *Streptomyces coelicolor* a negative impact on butanol titer was observed by Atsumi and coworkers (Atsumi et al., 2008a). Another alternative to *bcd-etfAB* consists in using *trans*-enoyl-CoA reductase (Ter) genes, whose protein product can also catalyze the reduction of crotonyl-CoA. The authors in (Shen et al., 2011) tested the impact of expressing *ter* genes from several microorganisms (*Treponema denticola*, *T. vincentii*, *Flavobacterium johnsoniae*, and *Fibrobacter succinogenes*) on butanol accumulation in an *E. coli* strain expressing the butanol production pathway from clostridia. Among the enzymes tested, the *ter* gene from *T. denticola* provided the greatest increment on butanol production.

In this study, we aimed to test different enzymes to improve the flux in the reported bottlenecks for butanol production in *E. coli*. We replaced the clostridial *thl* for a gene encoding a thiolase never tested before, *phbA* (*Cupriavidus necator*) and designed strains expressing *bcd-etfAB* and *ter.* The expression of the BCS operon with a single promoter and with individual promoters for each enzyme was also evaluated. Moreover, we aimed to compare the results with the ones obtained for strains expressing 2-oxoglutarate pathway to produce butanol (Chapters 4 and 5). In Figure 6.1 the clostridial butanol pathway is depicted and the alternative enzymes tested in this study are underlined.

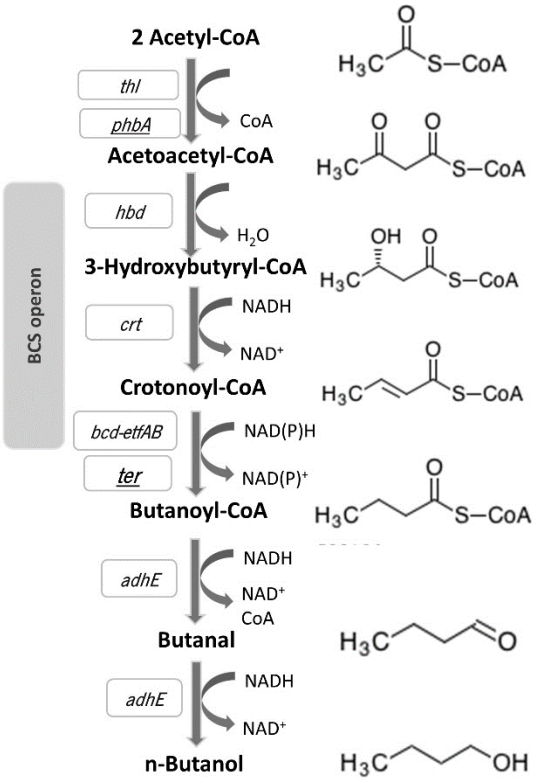


Figure 6.1: Clostridial n-butanol biosynthetic pathway with indication of the genes encoding the enzymes catalyzing the different reactions and with indication of the BCS (butyryl-CoA synthesis) operon. The alternative genes tested in this work are underlined. thl: thiolase; phbA: β-ketothiolase; hbd; 3-hydroxybutyryl-CoA dehydrogenase; crt: 3-hydroxybutyryl-CoA dehydratase; bcd: butyryl-CoA dehydrogenase; etfAB: electron transferring flavoprotein ter - trans-*enoyl-CoA reductase, adhE - aldehyde dehydrogenase/alcohol dehydrogenase*

6.2 Materials and methods

6.2.1 Cloning procedure

E. coli NEB 5-alpha cells (New England BioLabs, Massachusetts, USA) were used for gene cloning and vector propagation. These strains were cultured in Luria-Bertani (LB) medium (10 $g.L^{-1}$ of peptone; 5 $g.L^{-1}$ yeast extract and 5 $g.L^{-1}$ of NaCl) with the appropriate antibiotic concentration. The solid version of this medium included 15 $g.L^{-1}$ agar. All cultivations were performed at 37 °C and, in the case of liquid cultures, under shaking conditions (200 rpm).

For long-term storage, glycerol was added to a final concentration of 30 % (v/v) to overnight cultures in selective media and kept in a -80 °C freezer.

The genes used in this study were amplified by polymerase chain reaction (PCR) using Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Waltham, USA) in a LifeECO Thermal Cycler (Bioer Technology, Zhejiang, China). All primers were purchased from Metabion (Munich, Germany). DNA fragments were purified using DNA Clean and Concentrator DNA Kit (Zymo Research, Irvine, USA).

Plasmids were extracted using Plasmid Miniprep kit (Zymo Research). All digestions were performed using the appropriate FastDigest® restriction endonucleases (Thermo Scientific). Ligations were performed with T4 DNA Ligase (Thermo Scientific) and transformed by heat-shock in chemically competent cells *E. coli* NEB 5-alpha. The success of ligation was checked through Colony PCR using DreamTaq (Thermo Scientific) and further confirmed by sequencing (StabVida, Lisbon, Portugal). Protocols were performed in accordance with manufacturer's instructions.

6.2.2 Plasmid constructions

Compatible plasmids pACYDDuet-1, pCOLADuet-1, pETDuet-1 and pCDFDuet-1 (Novagen, Germany) were used to express the genes constituting the different biosynthetic butanol pathways here proposed.

The plasmid pETDuet (Novagen), ampicillin-resistant, was used to clone the genes *adhE* and *phbA*, originating peAP. The *adhE* gene was amplified from template plasmid pmTA1 (Nielsen et al., 2009) using primers *adhE_fw* and *adhE_rev* flanked with restriction sites for *Eco*RI and *Sal*I, respectively. *phbA* was amplified from *Cupriavidus necator* genomic DNA using *phbA_fw* and *phbA_rev* and inserted between *Nde*I/*Xho*I recognition sites of the previous construction.

The BCS operon (*crt-bcd-etfB-etfA-hbd*) was cloned in pACYCDuet. BCS operon was amplified with the primer pair *BCS_fw*/*BCS_rev* using as template the plasmid prBCS (Nielsen et al., 2009). The BCS operon was cloned between the recognition sites *Sac*I and *Sal*I, yielding paBCS.

Each of the genes constituting the BCS operon were also cloned separately in two different plasmids: pCOLADuet and pACYCDuet. The PCR amplification of each gene was performed using as template the plasmid prBCS (Nielsen et al., 2009). pCOLADuet, containing the kanamycin resistance gene, was used to clone the gene *hbd* (amplified using the primers *hbd_fw* and *hbd_rev*) between the *BamH*I/*Sac*I recognition sites. The PCR product for *crt*, amplified using primer *crt_fw* and *crt_rev*,

was restricted and ligated into *Kpn*I and *Hind*III restriction sites of the previous construction, yielding pcHC.

The plasmid pACYCDuet (Novagen), expressing the chloramphenicol-resistance marker, was used to clone the genes *bcd* and *etfAB*. *Bcd* was amplified using the primers *bcd_fw* and *bcd_rev* with restriction sites to *BamH*I and *Sac*I and cloned in pACYCDuet. Then, *etfAB* was inserted in the previous construction. These two genes were simultaneously amplified using primers *etfAB_fw* and *etfAB_rev* with restriction sites for *Kpn*I and *Xho*I, originating paBE.

The synthetic gene *ter* (ATG:biosynthesis, Freiburg, Germany) from *Treponoema denticola* (NCBI Sequence ID: AE017226.1) was amplified using primers *ter_fw* and *ter_rev*, restricted and ligated into *Nde*I and *Xho*I restriction sites of the plasmid pCDFDuet, with spectinomycin resistance marker, originating pcTER.

For each constructed plasmid, colony PCR with appropriate primers was used to find successful clones and the final plasmid was sent for sequencing to confirm the sequence was correct. All the primers used in this study are shown in Table 6.1.

*Table 6.1: Sequences of primers used in the cloning procedures of this study (*restriction sites are underlined). fw-forward; rev – reverse.*

Primer	Sequence	Restriction Sites *
adhE_fw	CCGAATTCATGAAAGTCACAACAGTAAAG	*EcoR*I
adhE_rev	CCGTCGACTTAAGGTTGTTTTTTAAAACAA	*Sal*I
ter_fw	CCCATATGATTGTAAAACC	*Nde*I
ter_rev	CCGCTCGAGTTAAATC	*Xho*I
BCS_fw	CCGAGCTCATGGAACTAAACAATGTC	*Sac*I
BCS_rev	CCGTCGACTTATTTTGAATAATCGTAGAA	*Sal*I
phbA_fw	CCCATATGACTGACGTTGTCATCG	*Nde*I
phbA_rev	CCCTCGAGTTATTTGCGCTCG	*Xho*I
hbd_fw	CCGGATCCATGAAAAAGGTATGTGTTATAG	*BamH*I
hbd_rev	CCGAGCTCTTATTTTGAATAATCGTAGAAAC	*Sac*I
crt_fw	CCGGTACCATGGAACTAAACAATGTC	*Kpn*I
crt_rev	CCCTCGAGCTATCTATTTTTGAAGCCTT	*Xho*I
bcd_fw	CCGGATCCATGGATTTTAATTTAACAAGAG	*BamH*I
bcd_rev	CCGAGCTCTTATCTAAAAATTTTTCCTGAA	*Sac*I
etfAB_fw	CCGGTACCATGAATATAGTTGTTTGTTTAA	*Kpn*I
etfAB_rev	CCCTCGAGTTAATTATTAGCAGCTTTAACT	*Xho*I

The plasmids constructed in this study and the respective major features are depicted in Figure 6.2.

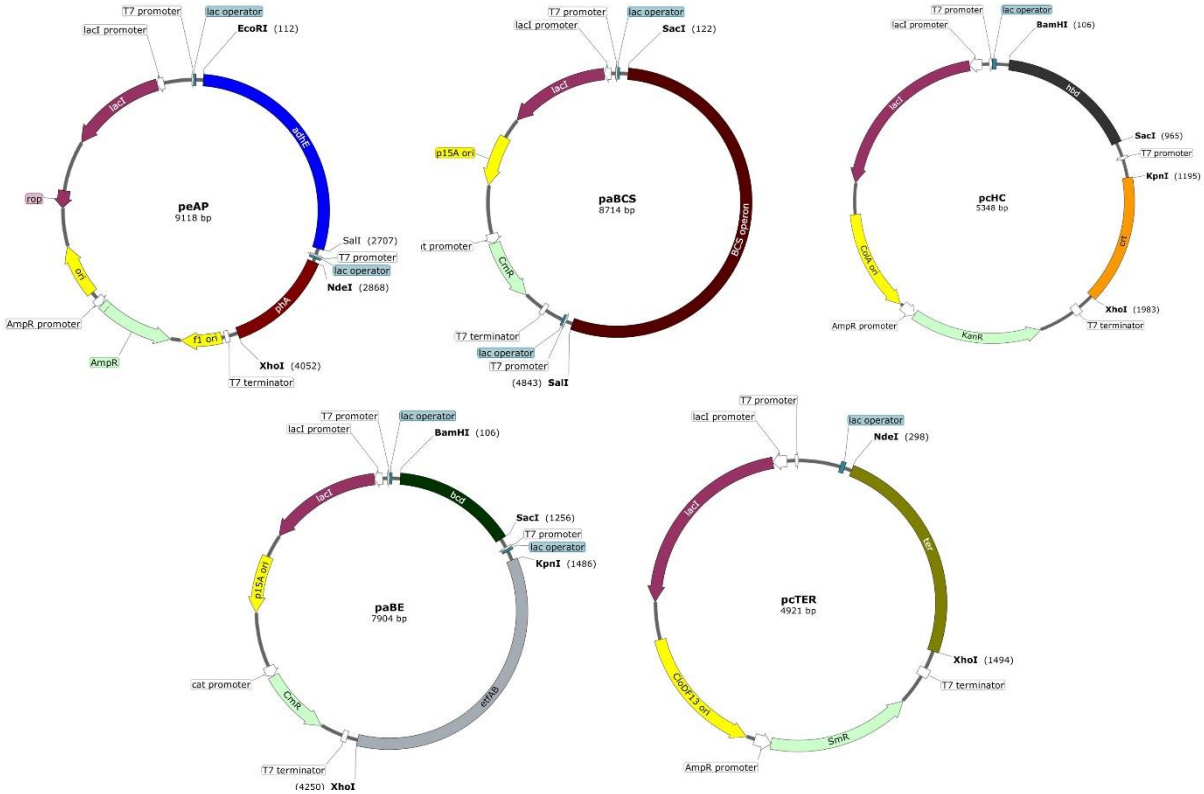


Figure 6.2: Schematic representation of the plasmid maps of all the cloning constructions developed, with indication of the main features. The restriction sites used to clone the genes are indicated

6.2.3 Bacterial strains

Besides *E. coli* BL21 (DE3) and *E. coli* K12 MG1655 (DE3), the multiple knock-out strains constructed in chapter 5 were also used as hosts for gene expression under control of T7 promotor. Three groups of strains were created by transforming the host strains with different combinations of plasmids: PB strains were generated by transforming *E. coli* BL21 (DE3), *E. coli* K12 MG1655 (DE3) and the knock-out mutants with peAP and paBCS by electroporation; PHCB strains were developed by transforming *E. coli* strains with peAP, pcHC, paBE; and PHCT strains were transformed with the plasmids peAP, pcHC and pcTER.

All host strains were made electrocompetent using the protocol developed by (Dower et al., 1988) and transformed using 0.1 cm-gap electroporation cuvettes at a voltage of 1.8 KV. Positive transformants were isolated in LB agar plates, containing the appropriate antibiotics concentrations (50 $\mu g.mL^{-1}$ ampicillin, 50 $\mu g.mL^{-1}$ spectinomycin, 34 $\mu g.mL^{-1}$ chloramphenicol, 30 $\mu g.mL^{-1}$ kanamycin) and incubated at 37 °C, overnight. To confirm the success of the transformation, in the

next day some colonies were inoculated in LB medium with antibiotics, overnight. Afterwards, plasmids were extracted and digested. The resulting fragments lengths were confirmed by running a 1 % (w/v) agarose gel.

For long-term storage, glycerol was added to a final concentration of 30 % (v/v) to overnight cultures in selective media and kept in -80 °C freezer.

Table 6.2 summarizes the strains of *E. coli* used or engineered for this study.

Table 6.2: List of strains and genomic DNA used or engineered for this study.

Strains	Relevant genotype	Source
***E. coli* K12 MG1655 (DE3)**	F - λ - ilvGrfb- 50 rph- 1 λ(DE3)	(Nielsen et al., 2009)
***E. coli* BL21 (DE3)**	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5	New England Labs
***E. coli* 3KO**	F - λ - ilvGrfb- 50 rph- 1 λ(DE3) Δ*adhP* Δ*ldhA* Δ*adhE*	Chapter 5
***E. coli* 4KO**	F - λ - ilvGrfb- 50 rph- 1 λ(DE3) Δ*adhP* Δ*ldhA* Δ*adhE* Δ*pta*	Chapter 5
***E. coli* 5KO**	F - λ - ilvGrfb- 50 rph- 1 λ(DE3) Δ*adhP* Δ*ldhA* Δ*adhE* Δ *pta* Δ*mhpF*	Chapter 5
***E. coli* 6KO**	F - λ - ilvGrfb- 50 rph- 1 λ(DE3) Δ*adhP* Δ*ldhA* Δ*adhE* Δ*pta* Δ*mhpF* Δ*ackA*	Chapter 5
PB1	*E. coli* BL21 (DE3) peAP paBCS	This study
PB2	*E. coli* K12 MG1655 (DE3) peAP paBCS	This study
PB3	*E. coli* 3KO peAP paBCS	This study
PB4	*E. coli* 4KO peAP paBCS	This study
PB5	*E. coli* 5KO peAP paBCS	This study
PB6	*E. coli* 6KO peAP paBCS	This study
PHCB1	*E. coli* BL21 (DE3) peAP pcHC paBE	This study
PHCB2	*E. coli* K12 MG1655 (DE3) peAP pcHC paBE	This study
PHCB3	*E. coli* 3KO peAP pcHC paBE	This study
PHCB4	*E. coli* 4KO peAP pcHC paBE	This study
PHCB5	*E. coli* 5KO peAP pcHC paBE	This study
PHCB6	*E. coli* 6KO peAP pcHC paBE	This study
PHCT1	*E. coli* BL21 (DE3) peAP pcHC pcTER	This study
PHCT2	*E. coli* K12 MG1655 (DE3) peAP pcHC pcTER	This study
PHCT3	*E. coli* 3KO peAP pcHC pcTER	This study
PHCT4	*E. coli* 4KO peAP pcHC pcTER	This study

PHCT5	*E. coli* 5KO peAP pcHC pcTER	This study
PHCT6	*E. coli* 6KO peAP pcHC pcTER	This study

6.2.1 Butanol production experiments in complex medium

The PB, PHCB and PHCT strains were cultivated in Terrific Broth (TB) medium supplemented with glucose according to the composition shown in Table 6.3. The pH of this medium was 7.2 ± 0.2 at 25°C.

Table 6.3: Medium composition of Terrific Broth.

Component	Amount per Liter	Unit
Tryptone	12	g
Yeast extract	24	g
Glycerol	4	mL
Monobasic potassium phosphate	2.31	g
Dibasic potassium phosphate	12.54	g
Glucose	10	g

For the pre-cultures, a single colony was picked from LB plates and inoculated in 10 mL of LB medium. Cultivation was performed with the addition of suitable antibiotics according to the employed plasmids (50 μg.mL^{-1} ampicillin, 50 μg.mL^{-1} spectinomycin, 34 μg.mL^{-1} chloramphenicol and 30 μg.mL^{-1} kanamycin). The pre-cultures were grown aerobically on a rotary shaker at 37 °C and 200 rpm, overnight.

500 mL shake flasks with 100 mL of TB medium, containing appropriate antibiotics, were inoculated with pre-cultures to obtain an initial optical density OD_{600} of 0.1. Cultivation was carried on a rotary shaker at 200 rpm at 37 °C. The butanol production genes were induced by the addition of 0.5 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG) to the culture medium when an optical density OD_{600} of 0.4-0.5 was reached.

To promote butanol production, after induction, the cells were switched to anaerobic conditions by transferring 60 mL of culture to 120 mL sealed serum flasks. The culture was supplemented with 600 μL of a 0.01 M stock solution of sodium bicarbonate to achieve a final concentration of 10 mM,

since it reduces long lag phases in *E. coli* anaerobic growth (Hornsten, 1995). The cultures were incubated at 30 °C and 180 rpm, for 96 hours.

Samples of the broth were collected at time 0, induction time and 96 h. All experiments were performed in triplicate and the samples were analyzed by Gas Chromatography (GC).

6.2.2 Butanol production experiments in defined medium

The strains PB, PHCB and PHCT were also tested for butanol production by cultivating them in High Density Medium (HDM) adapted from (Sivashanmugam et al., 2009) reported to provide high level of heterologous protein expression. HDM formulation is shown in Table 6.4.

Table 6.4: Medium composition of HDM, adapted from (Sivashanmugam et al., 2009).

Component	Amount per Liter	Unit
Glucose	10	g
Dibasic sodium phosphate dihydrate	8.89	g
Monobasic potassium phosphate	6.8	g
Sodium chloride	0.58	g
Magnesium sulphate	1.35	g
Calcium chloride dihydrate	0.038	g
Ammonium chloride	1	g
Trace metals	250	µL
Vitamins BME100x	250	µL
Amino acids mix	2	g

The trace metals solution contained (per liter): $FeSO_4.7H_2O$ (30 mg); $ZnSO_4.7H_2O$ (45 mg); $CaCl_2.2H_2O$ (45 mg); $MnCl_2.2H_2O$ (100 mg); $CoCl_2.6H_2O$ (30 mg); $CuSO_4.5H_2O$ (30 mg); $Na_2MoO_4.2H_2O$ (40 mg); H_3BO_3 (10 mg); KI (10 mg) and Na_2EDTA (1.5 g). The amino acid mix contained 1 g of adenine and 4 g of arginine, aspartate, glutamate, histidine, isoleucine, lysine, methionine, phenylalanine, serine, threonine, tryptophan, tyrosine and valine. The vitamin BME 100 x solution (Sigma Aldrich, St. Louis, MO, USA) contained (per liter): D-biotin (0.1 g); choline chloride (0.1 g); folic acid (0.1 g); myo-inositol (0.2 g); niacinamide (0.1 g); D-pantothenic acid.½Ca (0.1 g); riboflavin (0.01 g); thiamine.HCl (0.1 g) and NaCl (8.5 g).

A single colony was picked from LB plates and inoculated in 10 mL of LB medium. Cultivation was performed with the addition of suitable antibiotics according to the employed plasmids (50 $\mu g.mL^{-1}$

ampicillin, 50 μg.mL^{-1} spectinomycin, and 30 μg.mL^{-1} kanamycin). The pre-cultures were grown aerobically on a rotary shaker at 37 °C and 200 rpm, overnight. Cells were washed and harvested by centrifugation (10 min at 3000 × g). Afterwards, an appropriate volume of pre-culture was transferred to 500 mL shake flasks with 100 mL of HDM medium, containing the appropriate antibiotics, yielding an initial OD_{600} of 0.1. This culture was cultivated on a rotary shaker at 200 rpm at 37 °C. The butanol production genes were induced with 0.5 mM IPTG at an OD_{600} of 0.4-0.5.

After induction, 60 mL of the culture were transferred to 120 mL sealed serum flasks to promote butanol production under anaerobic conditions. The culture was supplemented with 600 μL of a 0.01 M stock solution of sodium bicarbonate to achieve a final concentration of 10 mM. Selenium, nickel and molybdenum are part of the formate hydrogen lyase (FHL) complex, which is induced under anaerobic conditions (Soini, Ukkonen and Neubauer, 2008). For this reason, 60 μL of a solution of extra trace metals ($NiCl_2$ (1.7 mg.L^{-1}); $(NH_4)_6Mo_7O_{24}$ (14.5 mg.L^{-1}); $4H_2O$ Na_2SeO_3 (2.4 mg.L^{-1})) was supplied to the medium. The cultures were incubated at 30 °C and 180 rpm, for 96 hours.

Samples of broth were collected at time 0, induction time and 96 h. All the experiments were performed in triplicate and the samples were analyzed by GC.

6.2.3 Analytical methods

For chromatographic analysis, samples were centrifuged at 6000 × g for 10 min to separate cells from the medium. Afterwards, the supernatant was filtered with a 0.22 μm pore filter membrane to glass vials and stored at -20 °C until analyzed.

Butanol concentration was quantified by a GP-9000 system (Chrompack) with a Meta-WAX capillary column (30 m X 0.25 mm X 0.25 μm) equipped with a flame ionization detector (FID); helium was used as carrier gas with a flow rate of 1 mL.min^{-1}. The filtered supernatant (900 μL) was mixed with 100 μL of a 5 g.L^{-1} solution of isobutanol, the internal standard, yielding a final concentration of 0.5 g.L^{-1}, and 1 μL of this mixture was injected. The temperature of injector and detector was maintained at 250° C. The column temperature was initially at 50° C, heated to 177.5° C at a 5° C.min^{-1} rate and then heated to 230° C at 10° C.min^{-1}, which was held for 15 minutes. A calibration curve was obtained by injecting standards with several concentrations of butanol and a fixed concentration of

internal standard (0.5 $g.L^{-1}$ of isobutanol). Butanol concentration was calculated by comparing the ratio between its peak area and internal standard peak area with calibration curves.

All cell optical density measurements at 600 nm (OD_{600}) were performed using the spectrophotometer Ultrospec 10 from Biochrom (Cambridge, UK).

6.3 Results and discussion

6.3.1 Butanol production experiments

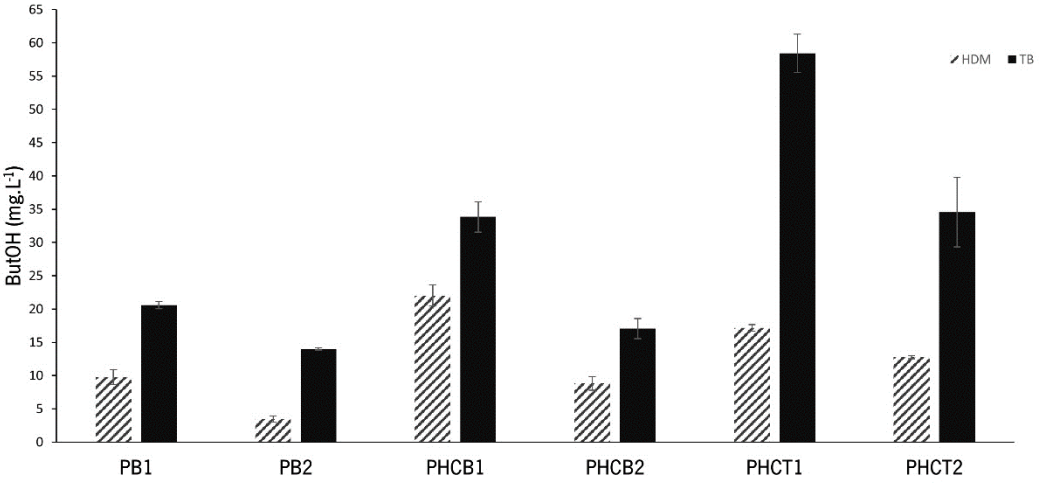
The first step of the clostridial butanol biosynthetic pathway is the condensation of two acetyl-CoA into one molecule of acetoacetyl-CoA by a thiolase. Acetyl-CoA has a central role in metabolism and several pathways in *E. coli* compete for its consumption. In order to optimize the flux of acetyl-CoA to butanol in *E. coli* expressing the clostridial pathway, several thiolases have been tested (Bond-watts et al., 2011; Dellomonaco et al., 2011; Ohtake et al., 2017). Given the high level of polyhydroxybutyrate accumulated in *Cupriavidus necator* and the successful implementation of the gene *phaA* to produce butanol, we hypothesized that the homologue enzyme beta-ketothiolase (*phbA*) from this organism could be an interesting alternative to optimize the flux to butanol production. The strains containing the prefix PB express *phbA* from *C. necator* together with the BCS operon from clostridia in the operon format (expressed using a single promoter).

Given the higher butanol production observed by other authors in strains that have the individual expression of the genes constituting the BCS operon (Nielsen et al., 2009), we have developed strains (PHCB prefix) that, besides the *phbA* from *C. necator*, also express the genes constituting the BCS operon with individual promoters-RBSs.

Butyryl-CoA dehydrogenase electron-transferring flavoprotein (Bcd/Etf) complex is sensitive to oxygen and requires ferredoxin as an additional redox partner, explaining its poor activity in *E. coli* (Lan and Liao, 2013). For this reason, we replaced this gene by the *ter* gene from *Treponema denticola* in the strains with the PHCT prefix. The enzyme, encoded by the gene *ter* is NADH-dependent and in *E. coli* presented the highest activity within the reported alternatives tested by (Shen et al., 2011).

E. coli BL21 (DE3), *E. coli* K12 (DE3), *E. coli* 3 KO, *E. coli* 4 KO, *E. coli* 5 KO and *E. coli* 6KO were genetically modified to express heterologous enzymes capable of catalyzing the successive reactions from acetyl-CoA to butanol originating the strains PB, PHCB and PHCT, respectively.

Since medium composition and type of *E. coli* strain showed to be a significant factor on the butanol titers obtained in Chapter 4, we tested all strain designs in two different media (TB and HDM) and for two different strain backgrounds (BL21 and K12). Figure 6.3 shows the final butanol concentrations obtained for BL21 based strains (suffix 1) and K12 based strains (suffix 2), cultivated either in TB or HDM.



	PB1	PB2	PHCB1	PHCB2	PHCT1	PHCT2
peAP	+	+	+	+	+	+
paBCS	+	+	-	-	-	-
pcHC	-	-	+	+	+	+
pABE	-	-	+	+	-	-
pcTER	-	-	-	-	+	+

Figure 6.3: Butanol titer (mg.L^{-1}) for PB, PHCB and PHCT strains in TB medium and HDM medium. BL21 based strains end with the suffix 1 and K12 based end with suffix 2. In all experiments, strains were grown in shake-flasks until 0.4-0.5 OD_{600} , 0.5 mM of IPTG was added to the medium and cells were transferred to sealed serum bottles. Titers are shown as mean ± S.D from three independent experiments.

The first observation from the engineered strains is that there is accumulation of butanol using the *phbA* gene, showing that the *C. necator* enzyme is capable of substituting the clostridial enzyme in the butanol production pathway (Figure 6.3). Furthermore, we observed that final butanol titers were improved for BL21 and K12 based strains in both media by individually expressing the genes from BCS operon (PHBC versus PB strains). Regarding PHCB1 cells, butanol increased 1.64-fold for TB medium and 2.25-fold for HDM medium, while for PHCB2 strains, the increment was 1.21-fold by

cultivating the cells in TB medium and 2.53-fold in HDM medium. PHCB1 grown in TB medium excreted the greatest amount of butanol (33.85±2.27 $mg.L^{-1}$).

By promoting individual gene expression, it was expected that the amount of heterologous protein would increase. The increment observed on the butanol final titers support this assumption. However, the level of increment on butanol titer was quite modest in comparison to another study that used the same strategy and obtained a 6-fold improvement on butanol production (Nielsen et al., 2009). Possibly, fine-tuning the expression of each gene individually by testing different promoter strengths could help optimizing further butanol production (Bond-watts et al., 2011).

By replacing *bcd-etfAb* for *ter*, we were able to increase butanol accumulation in both *E. coli* BL21 and K12 based strains (PHCT strains in Figure 6.3). The highest amount of butanol was achieved by cultivating PHCT1 in TB medium (58.41±2.88 $mg.L^{-1}$).

Contrarily to what was observed for BUT_OXG strains (Chapter 4), TB formulation seems to favor butanol production in all strains in Figure 6.3 in detriment of HDM composition. Other publications have demonstrated that the presence in the medium of complex nutrients is beneficial for butanol production through clostridial pathway and its derivatives. By removing complex nutrients such as tryptone and yeast extract from the medium, the authors observed a reduction in butanol production (Shen et al., 2011; Wen and Shen, 2016). The contrasting behavior between the two butanol production pathways is likely to be a consequence of the different precursors used by each pathway and the influence of the complex nutrients on their availability.

Regarding the host strain used for butanol production, we observed that in all cases BL21 strains provided the highest butanol titers. Again, this contrasts with the results from Chapter 4 where BUT_OXG strains (K12 based) showed increased butanol production in comparison to the same strain designs in BL21 hosts.

In Chapter 4, we hypothesized that K12 cells were better butanol producers than BL21 due to the constitutively expression of the glyoxylate bypass on the later and consequent lower flux towards 2-oxoglutarate. Since acetyl-CoA is the main precursor of the pathway tested here and BL21 cells are able to better express heterologous enzymes (Rosano and Ceccarelli, 2014), the superior butanol production results for BL21 based strains support this hypothesis.

Considering the improved butanol titers obtained with BUT_OXG strains by delaying the anaerobic switch for 12 h (Chapter 4 and Chapter 5), it is very likely that similar improvements would be observed for strains PB, PHCB and PHCT. Although we did not test this culture method here, we expect an increment on butanol titer at least for the PCHT strains (expressing Ter). Since PB and PHCB strains express the oxygen sensitive Bcd-EtfAB, expressing these proteins for too long under aerobic condition might not result in increased butanol titers. These experiments should be performed with TB medium since it provided the highest butanol titers. Considering that the growth rate is higher in complex media (data not shown), the intervals between induction and anaerobic switch may need to be shortened to avoid nutrient exhaustion before the switch to anaerobic conditions.

When we compare the highest butanol production obtained here for the PHCT1 strain (58.41±2.88 $mg.L^{-1}$), this value is still below the best butanol production results obtained for BUT_OXG2 when the cells were immediately switched to anaerobic conditions after induction (75.32±4.21 $mg.L^{-1}$). Although some of the enzymes constituting the 2-oxoglutarate pathway were codon-optimized, this is a still a good indicator of the potential of this novel pathway.

It is important to point out that the major difference between the two pathways is the precursor. The immediate advantage of 2-oxoglutarate pathway is the thermodynamically favorable first step to avoid avoiding the unfavorable condensation of two molecules of acetyl-CoA. Nevertheless, both metabolites are part of the central carbon metabolism and intermediates in fundamental biochemical pathways. Although we did not perform fluxomics, we believe that, depending on the pathway expressed, different flux distributions are obtained. This could explain the differences observed regarding the medium and strain that provided the greater butanol titers.

Both 2-oxoglutarate pathway implemented in Chapter 4 and the clostridial pathway (and variations characterized here) require, per molecule of butanol synthesized, four molecules of NADH. As discussed with the results obtained for the 2-oxoglutarate pathway (Chapter 5), it seems that the heterologous pathways for butanol production tested here could not fulfil the NADH recycling role under anaerobic conditions. Besides elimination of competing pathways, some of the approaches used to increase NADH pool in strains expressing clostridial pathway (gathered in Chapter 2) could also be applied to improve 2-oxoglutarate pathway, such as the expression of *fdh* (which converts formate into CO_2 and NADH) (Zhou et al., 2014) or activation of PDH complex under anaerobic conditions (Garza et al., 2012).

The same knock-out mutants tested in Chapter 5 were also tested in combination with clostridial pathways evaluated here. However, the results showed that in most cases butanol production was absent or considerably lower than for the reference strains (data not shown). One possible solution to improve butanol titers on the deletion mutants would be to perform Adaptive Laboratory Evolution to recover from the growth deficit and improve butanol accumulation (Fong et al., 2005). Although it is not clear if the knock-out strains used to produce butanol via clostridial pathway, reported in the literature, were subjected to adaptive evolution, we believe that strains expressing both butanol biosynthetic pathways would benefit from the implementation of this protocol.

6.4 Conclusions

In this chapter, we were able to design new strains of *E. coli* able to produce butanol though a pathway based on the clostridial route but expressing the enzyme β-ketothiolase. Within the developed strains, we also intended to analyze the effect on butanol production of expressing the enzymes constituting the BCS operon with individual promoters, instead of a single one, and of replacing the clostridial *bcd-etfAB* complex by *ter* from *Treponema denticola*.

In PHCB strains, the individual promoter-RBS expression of the enzymes constituting the BCS operon resulted, as expected, in higher butanol titers (Figure 6.3). Nevertheless, different promoter strengths can be applied to fine-tune the expression of heterologous enzymes, finding an optimal balance.

The replacement of *bcd-etfAB* by *ter* had a positive impact on butanol titer. From the reported works in the literature, *ter* from *Treponema denticola* provided the highest increment on butanol titer. Still, with the increasing availability of new enzymes sequence, other options can also be tested.

The highest butanol titer, 58.41±2.88 $mg.L^{-1}$, was achieved by cultivating PHCT1 strain in TB medium. In general, TB medium formulation and using *E. coli* BL21 as host seem to favor butanol production in all the designs (Figure 6.3). Still, the strains expressing 2-oxoglutarate pathway (Chapter 4) were able to produce greater concentrations of butanol, when compared with the titers here obtained.

CHAPTER 7 General conclusions and future perspectives

In this chapter we summarize the general conclusions of this thesis. Additionally, perspectives and recommendations for future work are presented here as well.

7.1 General conclusions

The main goal of this thesis was to design novel strains of *Escherichia coli* able to produce butanol. Particularly, we aimed at implementing *in vivo* a novel butanol pathway generated by a hyper-graph algorithm and further supported by a thorough *in silico* analysis. Moreover, the clostridial butanol biosynthetic pathway was also engineered and implemented in *E. coli*. The general conclusions of this work are as follows:

- Reviewing the literature allowed to gather different rational strategies to improve butanol production in *E. coli*. Several strategies haven been applied before to tackle the issues previously identified when expressing the clostridial butanol pathway in *E. coli*, such as cofactor balancing and elimination of competing pathways. Among the promising strategies identified in the literature, the implementation of novel routes more suitable to the host may be of particular interest. In this regard, pathway discovery tools can generate pathways with improved performance that solve some of the problems associated with the clostridial butanol biosynthesis pathway (e.g., poor thermodynamics for some enzymatic steps)

- Pathway enumeration algorithms are a powerful tool to generate novel routes. Nevertheless, these novel pathways must be subjected to an exhaustive *in silico* analysis alongside with manual curation. By successively ranking the butanol production pathways generated in a previous study (Liu *et al.*, 2015), we were able to filter 24 promising pathways from an initial set of 105,954 routes. The integration of these routes in a genome-scale model of *E. coli* allowed a detailed analysis of each candidate pathway. The use of strain optimization algorithms led to the computation of optimum genotypes that maximized butanol production *in silico* in *E. coli,* which promised to avoid long trial and error strain construction cycles.

- For the first time, butanol was produced through a pathway whose precursor is 2-oxoglutarate. Among the different strains tested, the maximum butanol titer obtained was 128.95±7.73 $mg.L^{-1}$. Expressing butanol genes in *E. coli* K12, cultivation in High Density Medium (HDM) formulation and delaying anaerobic switch after IPTG induction by 12 h seem to favor butanol accumulation via the 2-oxoglutarate pathway.

- As predicted by the *in silico* analysis, butanol production is favored under anaerobic conditions. This was validated by *in vivo* experiments, which did not show any butanol accumulation by cultivating the strains expressing butanol genes under aerobic conditions.

- The optimum phenotype predicted by the *in silico* analysis included the knock-out of ethanol, acetate and lactate production pathways. Nevertheless, butanol accumulation decreased by disrupting the respective genes. Although the phenotype of the knock-out strains did not match fully the *in silico* predictions, we were able to cease ethanol production and reduce lactate and acetate accumulation. The results suggest that the engineered strains, as a response to genetic interventions, are probably in a non-optimal state showing a growth-deficit.

- We were able to validate the *in vivo* potential of the solutions generated by the hyper-graph algorithm previously implemented by Liu *et al.*, 2015. This validation encourages further research into the results generated in the same work for the non-native production of vanillin in *Saccharomyces cerevisiae* and of curcumin in *E. coli*.

- We developed strains of *E. coli* able to produce butanol though a pathway based on the clostridial route modified to use the enzyme β-ketothiolase in the first step. Within the developed strains, we intended to analyze the effect on butanol production of expressing the enzymes constituting the BCS (butyryl-CoA synthesis) operon with individual promoters, instead of a single one and replacing the clostridial *bcd-etfAB* complex by *ter* from *Treponema denticola*. Both modifications led to greater butanol accumulation. The maximum butanol titer obtained was 58.41±2.88 $mg.L^{-1}$. Contrarily to what was observed by cultivating cells expressing 2-oxoglutarate pathway, Terrific Broth medium formulation and using *E. coli* BL21 as host favored butanol accumulation. Overall, strains expressing 2-oxoglutarate pathway produced greater concentrations of butanol, which is a promising indicator of this new route's potential.

7.2 Recommendations for future work

The replacement of chemical-based processes by biological-based ones plays a key role in the quest for sustainability. Although in this work we were able to successfully implement new pathways in *E. coli* to produce butanol, the obtained titers still far below the industrial requirements. The optimization of a bioprocess requires several rounds of optimization. So, the results here obtained are the starting point for the development of an efficient and sustainable process. To do so, several strategies could be implemented:

- The use of RBS strength prediction algorithms to fine-tune gene-expression. By doing so, an optimal balance of gene expression with increased butanol production could be found. Codon-optimization of the genes *ter* and *adhE* converting the three last steps of 2-oxoglutarate pathway and of all the genes constituting clostridial derivative pathways could enhance gene expression in *E. coli* and, consequently, butanol production.

- Application of techniques such as transcriptomics, fluxomics and metabolomics could give insight on the metabolism of the different strains, allowing to identify possible metabolic engineering strategies to increase butanol accumulation. Further exploration of strategies to improve the accumulation of the precursors 2-oxoglutarate and acetyl-CoA, could help to improve butanol titers obtained through the respective pathways.

- Detection of pathway intermediates could allow to identify the bottleneck(s), providing insight for the next steps for strain engineering. Measuring the activity of the enzymes could also help to identify the rate-limiting step(s) and act in accordance by, for example, testing enzyme homologs. This could be helped by the ever increasing number of whole genome sequences available, which can include novel enzymes to catalyze the enzymatic reactions explored in this thesis.

- ALE experiments could help the knock-out strains to recover some of the growth deficit shown, which could make the experimental phenotype converge to the simulated data and result in improved butanol production.

- The research on the conversion of cheap and widely available substrates in an important task in the endeavor of developing a sustainable process to produce biobutanol able to compete with the petrochemical industry.

- The integration of butanol production genes into the genome under the control of native fermentation regulatory elements from the major fermentative genes could also improve butanol production by creating a self-regulated butanol production system in *E. coli*. This strain would be able to auto-induce butanol production under anaerobic conditions and, simultaneously, inactivate alternative fermentation pathways. From an industrial-perspective, this would be much more economical because it would avoid the use of antibiotics and IPTG.

Agren, R., Liu, L., Shoaie, S., Vongsangnak, W., Nookaew, I., Nielsen, J., 2013. The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for *Penicillium chrysogenum* . PLoS Comput. Biol. 9.

Alper, H., Miyaoku, K., Stephanopoulos, G., 2005. Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. Nat. Biotechnol. 23, 612–616.

Amador-Noguez, D., Brasg, I. a, Feng, X.-J., Roquet, N., Rabinowitz, J.D., 2011. Metabolome remodeling during the acidogenic-solventogenic transition in *Clostridium acetobutylicum* . Appl. Environ. Microbiol. 77, 7984–97.

Archer, C.T., Kim, J.F., Jeong, H., Park, J.H., Vickers, C.E., Lee, S.Y., Nielsen, L.K., 2011. The genome sequence of *E. coli* W (ATCC 9637): Comparative genome analysis and an improved genome-scale reconstruction of *E. coli*. BMC Genomics 12, 9.

Arita, M., 2003. *In Silico* Atomic Tracing by Substrate-Product Relationships in *Escherichia coli* Intermediary Metabolism. Genome Res. 13, 2455–2466.

Atsumi, S., Cann, A.F., Connor, M.R., Shen, C.R., Smith, K.M., Brynildsen, M.P., Chou, K.J.Y., Hanai, T., Liao, J.C., 2008a. Metabolic engineering of *Escherichia coli* for 1-butanol production. Metab. Eng. 10, 305–11.

Atsumi, S., Hanai, T., Liao, J.C., 2008b. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature 451, 86–9.

Baneyx, F., 1999. Recombinant protein expression in *Escherichia coli*. Curr. Opin. Biotechnol.

Bar-Even, A., Flamholz, A., Noor, E., Milo, R., 2012. Thermodynamic constraints shape the structure of carbon fixation pathways. Biochim. Biophys. Acta 1817, 1646–59.

Baumler, D.J., Peplinski, R.G., Reed, J.L., Glasner, J.D., Perna, N.T., 2011. The evolution of metabolic networks of *E. coli* . BMC Syst. Biol. 5, 182.

Bendrat, K., Buckel, W., 1993. Cloning, sequencing and expression of the gene encoding the carboxytransferase subunit of the biotin-dependent Na^+ pump glutaconyl-CoA decarboxylase from *Acidaminococcus fermentans* in *Escherichia coli*. Eur. J. Biochem. 211, 697–702.

Blázquez, B., Carmona, M., García, J.L., Díaz, E., 2008. Identification and analysis of a glutaryl-CoA

dehydrogenase-encoding gene and its cognate transcriptional regulator from *Azoarcus* sp. CIB. Environ. Microbiol. 10, 474–482.

Bokinsky, G., Peralta-yahya, P.P., George, A., Holmes, B.M., Steen, E.J., Dietrich, J., 2011. Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered Escherichia coli. PNAS 108, 19949–19954.

Bond-watts, B.B., Bellerose, R.J., Chang, M.C.Y.Y., 2011. Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. Nat. Chem. Biol. 7, 222–227.

Bordbar, A., Monk, J.M., King, Z.A., Palsson, B.O., 2014. Constraint-based models predict metabolic and associated cellular functions. Nat. Rev. Genet. 15, 107–120.

Brochado, A.R., Andrejev, S., Maranas, C.D., Patil, K.R., 2012. Impact of Stoichiometry Representation on Simulation of Genotype-Phenotype Relationships in Metabolic Networks. PLOS Comput. Biol. 8, e1002758.

Buckel, W., Barker, H.A., 1974. Two pathways of glutamate fermentation by anaerobic bacteria. J. Bacteriol. 117, 1248–1260.

Buckel, W., Dorn, U., Semmler, R., 1981. Glutaconate CoA-transferase from *Acidaminococcus fermentans*. Eur. J. Biochem. 118, 315–321.

Buehler, E.A., Mesbah, A., 2016. Kinetic study of acetone-butanol-ethanol fermentation in continuous culture. PLoS One 11, 1–21.

Bunch, P.K., Mat-Jan, F., Lee, N., Clark, D.P., 1997. The *ldhA* gene encoding the fermentative lactate dehydrogenase of <i>Escherichia coli<. Microbiology 143 Pt 1, 187–195.

Burgard, A.P., Pharkya, P., Maranas, C.D., 2003. OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization. Biotechnol. Bioeng. 84, 647–657.

Carbonell, P., Fichera, D., Pandit, S.B., Faulon, J.-L., 2012. Enumerating metabolic pathways for the production of heterologous target chemicals in chassis organisms. BMC Syst. Biol. 6, 10.

Caspi, R., Altman, T., Dreher, K., Fulcher, C.A., Subhraveti, P., Keseler, I.M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A., Ong, Q., Paley, S., Pujar, A., Shearer, A.G., Travers, M., Weerasinghe, D., Zhang, P., Karp, P.D., 2012. The MetaCyc database of metabolic

pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 40, D742–D753.

Causey, T.B., Zhou, S., Shanmugam, K.T., Ingram, L.O., 2003. Engineering the metabolism of *Escherichia coli* W3110 for the conversion of sugar to redox-neutral and oxidized products: Homoacetate production. Proc. Natl. Acad. Sci. 100, 825–832.

Chang, D., Shin, S., Rhee, J., 1999. Acetate Metabolism in a pta Mutant of *Escherichia coli* W3110 : Importance of Maintaining Acetyl Coenzyme A Flux for Growth and Survival 181, 6656–6663.

Chang, Y.-J., Pukall, R., Saunders, E., Lapidus, A., Copeland, A., Nolan, M., Glavina Del Rio, T., Lucas, S., Chen, F., Tice, H., Cheng, J.-F., Han, C., Detter, J.C., Bruce, D., Goodwin, L., Pitluck, S., Mikhailova, N., Liolios, K., Pati, A., Ivanova, N., Mavromatis, K., Chen, A., Palaniappan, K., Land, M., Hauser, L., Jeffries, C.D., Brettin, T., Rohde, M., Göker, M., Bristow, J., Eisen, J. a, Markowitz, V., Hugenholtz, P., Kyrpides, N.C., Klenk, H.-P., 2010. Complete genome sequence of *Acidaminococcus fermentans* type strain (VR4). Stand. Genomic Sci. 3, 1–14.

Chatsurachai, S., Furusawa, C., Shimizu, H., 2012. An *in silico* platform for the design of heterologous pathways in nonnative metabolite production. BMC Bioinformatics 13, 93.

Chen, C.T., Liao, J.C., 2016. Frontiers in microbial 1-butanol and isobutanol production. FEMS Microbiol. Lett. 363, 1–13.

Clomburg, J.M., Gonzalez, R., 2010. Biofuel production in *Escherichia coli* : The role of metabolic engineering and synthetic biology. Appl. Microbiol. Biotechnol. 86, 419–434.

Commichau, F.M., Forchhammer, K., Stülke, J., 2006. Regulatory links between carbon and nitrogen metabolism. Curr. Opin. Microbiol. 9, 167–172.

Conrad, T.M., Lewis, N.E., Palsson, B.Ø., 2011. Microbial laboratory evolution in the era of genome-scale science. Mol. Syst. Biol. 7, 509.

Copeland, W.B., Bartley, B.A., Chandran, D., Galdzicki, M., Kyung, H., Sleight, S.C., Maranas, C.D., Sauro, H.M., 2013. Computational Tools for Metabolic Engineering 14, 270–280.

Cotten, C., Reed, J.L., 2013. Constraint-based strain design using continuous modifications (CosMos) of flux bounds finds new strategies for metabolic engineering. Biotechnol. J. 8, 595–604.

Croes, D., Couche, F., Wodak, S.J., van Helden, J., 2005. Metabolic PathFinding: inferring relevant pathways in biochemical networks. Nucleic Acids Res. 33, W326–W330.

de Graef, M.R., Alexeeva, S., Snoep, J.L., Teixeira de Mattos, M.J., 1999. The Steady-State Internal Redox State (NADH/NAD) Reflects the External Redox State and Is Correlated with Catabolic Adaptation in *Escherichia coli*. J. Bacteriol. 181, 2351–2357.

De Mey, M., De Maeseneire, S., Soetaert, W., Vandamme, E., 2007. Minimizing acetate formation in *E. coli* fermentations, Journal of Industrial Microbiology and Biotechnology.

Dellomonaco, C., Clomburg, J.M., Miller, E.N., Gonzalez, R., 2011. Engineered reversal of the β-oxidation cycle for the synthesis of fuels and chemicals. Nature 476, 355–359.

Dellomonaco, C., Rivera, C., Campbell, P., Gonzalez, R., 2010. Engineered Respiro-Fermentative Metabolism for the Production of Biofuels and Biochemicals from Fatty Acid-Rich Feedstocks. Appl. Environ. Microbiol. 76, 5067–5078.

Devoid, S., Overbeek, R., DeJongh, M., Vonstein, V., Best, A.A., Henry, C., 2013. Automated Genome Annotation and Metabolic Model Reconstruction in the SEED and Model SEED, in: Alper, H.S. (Ed.), Systems Metabolic Engineering: Methods and Protocols. Humana Press, Totowa, NJ, pp. 17–45.

Dias, O., Rocha, M., Ferreira, E.C., Rocha, I., 2015. Reconstructing genome-scale metabolic models with merlin. Nucleic Acids Res. 43, 3899–3910.

Dittrich, C.R., Vadali, R. V., Bennett, G.N., San, K.Y., 2005. Redistribution of metabolic fluxes in the central aerobic metabolic pathway of *E. coli* mutant strains with deletion of the *ackA-pta* and *poxB* pathways for the synthesis of isoamyl acetate. Biotechnol. Prog. 21, 627–631.

Djurdjevic, I., Zelder, O., Buckel, W., 2011. Production of Glutaconic Acid in a Recombinant *Escherichia coli* Strain 77, 320–322.

Dong, H., Zhao, C., Zhang, T., Lin, Z., Li, Y., Zhang, Y., 2016. Engineering *Escherichia coli* Cell Factories for n-Butanol Production, in: Ye, Q., Bao, J., Zhong, J.-J. (Eds.), Bioreactor Engineering Research and Industrial Applications I: Cell Factories. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 141–163.

Dower, W.J., Miller, J.F., Ragsdale, C.W., 1988. High efficiency transformation of *E. coli* by high

voltage electroporation. Nucleic Acids Res. 16, 6127–6145.

Dürre, P., 2008. Fermentative butanol production: bulk chemical and biofuel. Ann. N. Y. Acad. Sci. 1125, 353–62.

Dürre, P., 2007. Biobutanol: An attractive biofuel. Biotechnol. J. 2, 1525–1534.

Edwards, J.S., Palsson, B.O., 2000. The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. Proc. Natl. Acad. Sci. 97, 5528–5533.

Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.Ø., 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol. 3, 121.

Feist, A.M., Herrgård, M.J., Thiele, I., Reed, J.L., Palsson, B., 2009. Reconstruction of biochemical networks in microorganisms. Nat. Rev. Microbiol. 7, 129–143.

Flamholz, A., Noor, E., Bar-Even, A., Milo, R., 2012. EQuilibrator - The biochemical thermodynamics calculator. Nucleic Acids Res. 40, 770–775.

Fong, S.S., Burgard, A.P., Herring, C.D., Knight, E.M., Blattner, F.R., Maranas, C.D., Palsson, B.O., 2005. *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. Biotechnol. Bioeng. 91, 643–648.

Fong, S.S., Nanchen, A., Palsson, B.O., Sauer, U., 2006. Latent pathway activation and increased pathway capacity enable *Escherichia coli* adaptation to loss of key metabolic enzymes. J. Biol. Chem. 281, 8024–8033.

Friedler, F., Varga, J.B., Fan, L.T., 1995. Decision-mapping: A tool for consistent and complete decisions in process synthesis. Chem. Eng. Sci. 50, 1755–1768.

García, V., Päkkilä, J., Ojamo, H., Muurinen, E., Keiski, R.L., 2011. Challenges in biobutanol production: How to improve the efficiency? Renew. Sustain. Energy Rev. 15, 964–980.

Garza, E., Zhao, J., Wang, Y., Wang, J., Iverson, A., Manow, R., Finan, C., Zhou, S., 2012. Engineering a homobutanol fermentation pathway in *Escherichia coli* EG03. J. Ind. Microbiol. Biotechnol. 39, 1101–7.

Gheshlaghi, R., Scharer, J.M., Moo-Young, M., Chou, C.P., 2009. Metabolic pathways of clostridia

for producing butanol. Biotechnol. Adv. 27, 764–81.

Gonçalves, E., Pereira, R., Rocha, I., Rocha, M., 2012. Optimization Approaches for the *In Silico* Discovery of Optimal Targets for Gene Over/Underexpression. J. Comput. Biol. 19, 102–114.

Goto, S., Bono, H., Ogata, H., Fujibuchi, W., Nishioka, T., Sato, K., Kanehisa, M., 1997. Organizing and computing metabolic pathway data in terms of binary relations. Pac. Symp. Biocomput. 175–186.

Green, E.M., 2011. Fermentative production of butanol–the industrial perspective. Curr. Opin. Biotechnol. 22, 337–43.

H. Chan, W., S. Mohamad, M., Deris, S., M. Illias, R., 2013. A Review of Computational Approaches for *In Silico* Metabolic Engineering for Microbial Fuel Production. Curr. Bioinform. 8, 253–258.

Hans, M., Sievers, È., Mu, U., Bill, E., Vorholt, J.A., Linder, D., Buckel, W., 1999. 2-Hydroxyglutaryl-CoA dehydratase from *Clostridium symbiosum* 414, 404–414.

Harris, L.M., Blank, L., Desai, R.P., Welker, N.E., Papoutsakis, E.T., 2001. Fermentation characterization and flux analysis of recombinant strains of *Clostridium acetobutylicum* with an inactivated solR gene. J Ind Microbiol Biotechnol 27, 322–328.

He, B.Q., Liu, M. Bin, Zhao, H., 2015. Comparison of combustion characteristics of n-butanol/ethanol-gasoline blends in a HCCI engine. Energy Convers. Manag. 95, 101–109.

Heath, A.P., Bennett, G.N., Kavraki, L.E., 2010. Finding metabolic pathways using atom tracking. Bioinformatics 26, 1548–1555.

Higgins, T.E., Johnson, M.J., 1970. Pathways of anaerobic acetate utilization in *Escherichia coli* and *Aerobacter cloacae*. J. Bacteriol. 101, 885–891.

Hornsten, E.G., 1995. On culturing *Escherichia coli* on a mineral salts medium during anaerobic conditions. Bioprocess Eng. 12, 157–162.

Huergo, L.F., Dixon, R., 2015. The Emergence of 2-Oxoglutarate as a Master Regulator Metabolite. Microbiol. Mol. Biol. Rev. 79, 419–35.

Ibarra, R.U., Edwards, J.S., Palsson, B.O., 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. Nature 420, 20–23.

Indurthi, S.M., Chou, H., Lu, C., Lu, C., 2018. Molecular characterization of *lysR-lysXE*, *gcdR-gcdHG* and *amaR-amaAB* operons for lysine export and catabolism : a comprehensive lysine catabolic network in *Pseudomonas aeruginosa* PAO1 876–888.

Inui, M., Suda, M., Kimura, S., Yasuda, K., Suzuki, H., Toda, H., Yamamoto, S., Okino, S., Suzuki, N., Yukawa, H., 2008. Expression of *Clostridium acetobutylicum* butanol synthetic genes in *Escherichia coli*. Appl. Microbiol. Biotechnol. 77, 1305–16.

Jang, Y.S., Malaviya, A., Cho, C., Lee, J., Lee, S.Y., 2012. Butanol production from renewable biomass by clostridia. Bioresour. Technol. 123, 653–663.

Jantama, K., Zhang, X., Moore, J.C., Shanmugam, K.T., Svoronos, S.A., Ingram, L.O., 2008. Eliminating side products and increasing succinate yields in engineered strains of *Escherichia coli* C. Biotechnol. Bioeng. 101, 881–893.

Kim, S.M., Peña, M.I., Moll, M., Bennett, G.N., Kavraki, L.E., 2017. A review of parameters and heuristics for guiding metabolic pathfinding. J. Cheminform. 9, 1–13.

Kim, T.S., Jung, H.M., Kim, S.Y., Zhang, L., Li, J., Sigdel, S., Park, J.H., Haw, J.R., Lee, J.K., 2015. Reduction of acetate and lactate contributed to enhancement of a recombinant protein production in *E. coli* BL21. J. Microbiol. Biotechnol. 25, 1093–1100.

Kim, Y., Ingram, L.O., Shanmugam, K.T., 2008. Dihydrolipoamide Dehydrogenase Mutation Alters the NADH Sensitivity of Pyruvate Dehydrogenase Complex of *Escherichia coli* K-12. J. Bacteriol. 190, 3851–3858.

Kim, Y., Ingram, L.O., Shanmugam, K.T., 2007. Construction of an *Escherichia coli* K-12 mutant for homoethanologenic fermentation of glucose or xylose without foreign genes. Appl. Environ. Microbiol. 73, 1766–1771.

King, Z.A., Feist, A.M., 2013. Optimizing Cofactor Specificity of Oxidoreductase Enzymes for the Generation of Microbial Production Strains—OptSwap. Ind. Biotechnol. 9, 236–246.

Klein, A.H., Shulla, A., Reimann, S.A., Keating, D.H., Wolfe, A.J., 2007. The intracellular concentration of acetyl phosphate in *Escherichia coli* is sufficient for direct phosphorylation of two-component response regulators. J. Bacteriol. 189, 5574–5581.

Koppolu, V., Vasigala, V.K., 2016. Role of *Escherichia coli* in Biofuel Production. Microbiol. Insights

9, MBI.S10878.

Kress, D., Bru, D., Schall, I., Linder, D., Buckel, W., Essen, L., 2009. An Asymmetric Model for Na^+-translocating Glutaconyl-CoA Decarboxylases 284, 28401–28409.

Kuhlman, T.E., Cox, E.C., 2010. Site-specific chromosomal integration of large synthetic constructs. Nucleic Acids Res. 38.

Lan, E.I., Liao, J.C., 2012. ATP drives direct photosynthetic production of 1-butanol in cyanobacteria. Proc. Natl. Acad. Sci. 109, 6018–6023.

Lee, D.-Y., Fan, L.T., Park, S., Lee, S.Y., Shafie, S., Bertók, B., Friedler, F., 2005. Complementary identification of multiple flux distributions and multiple metabolic pathways. Metab. Eng. 7, 182–200.

Lee, K.H., Park, J.H., Kim, T.Y., Kim, H.U., Lee, S.Y., 2007. Systems metabolic engineering of *Escherichia coli* for L-threonine production. Mol. Syst. Biol. 3, 149.

Li, N., Zhang, B., Chen, T., Wang, Z., Tang, Y.J., Zhao, X., 2013. Directed pathway evolution of the glyoxylate shunt in *Escherichia coli* for improved aerobic succinate production from glycerol. J. Ind. Microbiol. Biotechnol. 40, 1461–1475.

Lim, J.H., Seo, S.W., Kim, S.Y., Jung, G.Y., 2013. Model-driven rebalancing of the intracellular redox state for optimization of a heterologous n-butanol pathway in *Escherichia coli* . Metab. Eng. 20, 56–62.

Liu, F., Vilaça, P., Rocha, I., Rocha, M., 2015. Development and application of efficient pathway enumeration algorithms for metabolic engineering applications. Comput. Methods Programs Biomed. 118, 134–146.

Long, C.P., Antoniewicz, M.R., 2014. Metabolic flux analysis of *Escherichia coli* knockouts: Lessons from the Keio collection and future outlook. Curr. Opin. Biotechnol. 28C, 127–133.

Long, C.P., Gonzalez, J.E., Feist, A.M., Palsson, B.O., Antoniewicz, M.R., 2018. Dissecting the genetic and metabolic mechanisms of adaptation to the knockout of a major metabolic enzyme in *Escherichia coli*. Proc. Natl. Acad. Sci. 115, 222–227.

Luque, R., Herrero-Davila, L., Campelo, J.M., Clark, J.H., Hidalgo, J.M., Luna, D., Marinas, J.M., Romero, A.A., 2008. Biofuels: a technological perspective. Energy Environ. Sci. 1, 542.

Lutgens, M., Gottschalk, G., 1980. Why a co-substrate is required for anaerobic growth of *Escherichia coli* on citrate. J Gen Microbiol 119, 63–70.

Maia, P., Rocha, M., Rocha, I., 2016. In Silico Constraint-Based Strain Optimization Methods: the Quest for Optimal Cell Factories. Microbiol. Mol. Biol. Rev. 80, 45–67.

Mairhofer, J., Scharl, T., Marisch, K., Cserjan-Puschmann, M., Striedner, G., 2013. Comparative transcription profiling and in-depth characterization of plasmid-based and plasmid-free *Escherichia coli* expression systems under production conditions. Appl. Environ. Microbiol. 79, 3802–3812.

Maranas, C.D., Zomorrodi, A.R., 2016. Computational Strain Design, in: Optimization Methods in Metabolic Networks. John Wiley & Sons, Inc, pp. 155–172.

Martins, B.M., Macedo-Ribeiro, S., Bresser, J., Buckel, W., Messerschmidt, A., 2005. Structural basis for stereo-specific catalysis in NAD^+-dependent (R)-2-hydroxyglutarate dehydrogenase from *Acidaminococcus fermentans* 272, 269–281.

McCloskey, D., Palsson, B., Feist, A.M., 2013. Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli* . Mol. Syst. Biol. 9, 1–15.

Medema, M.H., Raaphorst, R. Van, Takano, E., Breitling, R., 2012. Computational tools for the synthetic design of biochemical pathways. Nat. Publ. Gr. 10, 191–202.

N-butanol market: Global trends and forecasts to 2018 by applications (butyl acrylate, butyl acetate, glycol ethers, and others) and geography, 2015.

Nakashima, N., Miyazaki, K., 2014. Bacterial cellular engineering by genome editing and gene silencing. Int. J. Mol. Sci. 15, 2773–2793.

Ndaba, B., Chiyanzu, I., Marx, S., 2015. N-Butanol derived from biochemical and chemical routes: A review. Biotechnol. Reports 8, 1–9.

Nielsen, D.R., Leonard, E., Yoon, S.-H., Tseng, H.-C., Yuan, C., Prather, K.L.J., 2009. Engineering alternative butanol production platforms in heterologous bacteria. Metab. Eng. 11, 262–73.

Noor, E., Haraldsdóttir, H.S., Milo, R., Fleming, R.M.T., 2013. Consistent Estimation of Gibbs Energy Using Component Contributions. PLOS Comput. Biol. 9, 1–11.

Ohta, K., Beall, D.S., Mejia, J.P., Shanmugam, K.T., Ingram, L., 1991. Genetic Improvement of

Escherichia coli for Ethanol Production : Chromosomal Integration of Zymomonas mobilis Genes Encoding Pyruvate Decarboxylase and Alcohol Dehydrogenase I. Appl. Environ. Microbiol 57, 893–900.

Ohtake, T., Pontrelli, S., Laviña, W.A., Liao, J.C., Putri, S.P., 2017. Metabolomics-driven approach to solving a CoA imbalance for improved 1- butanol production in *Escherichia coli* . Metab. Eng. 41, 135–143.

Orth, J.D., Conrad, T.M., Na, J., Lerman, J. a, Nam, H., Feist, A.M., Palsson, B.Ø., 2011. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism-2011. Mol. Syst. Biol. 7, 1–9.

Orth, J.D., Palsson, B.Ø., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? Nat. Biotechnol. 28, 245–248.

Park, J.H., Lee, K.H., Kim, T.Y., Lee, S.Y., 2007. Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and *in silico* gene knockout simulation. Proc. Natl. Acad. Sci. U. S. A. 104, 7797–7802.

Pasteur, L., 1862. Quelques résultats nouveaux relatifs aux fermentations acétique et butyrique. Bull. la Société Chim. Paris 52–53.

Patakova, P., Linhova, M., Rychtera, M., Paulova, L., Melzoch, K., 2013. Novel and neglected issues of acetone-butanol-ethanol (ABE) fermentation by clostridia: *Clostridium* metabolic diversity, tools for process mapping and continuous fermentation systems. Biotechnol. Adv. 31, 58–67.

Patil, K.R., Rocha, I., Förster, J., Nielsen, J., 2005. Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinformatics 6, 308.

Pereira, R., Nielsen, J., Rocha, I., 2015. Improvement of *in silico* strain engineering methods in *Saccharomyces cerevisiae* (Doctoral thesis). Universidade do Minho.

Pfromm, P.H., Amanor-Boadu, V., Nelson, R., Vadlani, P., Madl, R., 2010. Bio-butanol vs. bio-ethanol: A technical and economic assessment for corn and switchgrass fermented by yeast or *Clostridium acetobutylicum* . Biomass and Bioenergy 34, 515–524.

Pharkya, P., Burgard, A.P., Maranas, C.D., 2004. OptStrain: A computational framework for redesign of microbial production systems. Genome Res. 14, 2367–2376.

Phue, J.N., Lee, S.J., Kaufman, J.B., Negrete, A., Shiloach, J., 2010. Acetate accumulation through alternative metabolic pathways in *ackApta poxB* triple mutant in *E. coli* B (BL21). Biotechnol. Lett. 32, 1897–1903.

Phue, J.N., Shiloach, J., 2004. Transcription levels of key metabolic genes are the cause for different glucose utilization pathways in *E. coli* B (BL21) and *E. coli* K (JM109). J. Biotechnol. 109, 21–30.

Ponce De León, M., Cancela, H., Acerenza, L., 2008. A strategy to calculate the patterns of nutrient consumption by microorganisms applying a two-level optimisation principle to reconstructed metabolic networks. J. Biol. Phys. 34, 73–90.

Rabinovitch-Deere, C. a, Oliver, J.W.K., Rodriguez, G.M., Atsumi, S., 2013. Synthetic biology and metabolic engineering approaches to produce biofuels. Chem. Rev. 113, 4611–32.

Ranganathan, S., Maranas, C.D., 2010. Microbial 1-butanol production: Identification of non-native production routes and in silico engineering interventions. Biotechnol. J. 5, 716–25.

Ranganathan, S., Suthers, P.F., Maranas, C.D., 2010. OptForce: An Optimization Procedure for Identifying All Genetic Manipulations Leading to Targeted Overproductions. PLoS Comput. Biol. 6, e1000744.

Reed, J.L., Vo, T.D., Schilling, C.H., Palsson, B.O., 2003. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). Genome Biol. 4, R54.

Reitzer, L., 2003. Nitrogen Assimilation and Global Regulation in *Escherichia coli*. Annu. Rev. Microbiol. 57, 155–176.

Reyes, L.H., Cardona, C., Pimentel, L., Rodríguez-López, A., Alméciga-Díaz, C.J., 2017. Improvement in the production of the human recombinant enzyme N-acetylgalactosamine-6-sulfatase (rhGALNS) in *Escherichia coli* using synthetic biology approaches. Sci. Rep. 7, 1–14.

Rezaei, M., Zarkesh-Esfahani, S.H., 2012. Optimization of production of recombinant human growth hormone in *Escherichia coli* . J. Res. Med. Sci. 17, 681–685.

Rocha, I., Maia, P., Evangelista, P., Vilaça, P., Soares, S., Pinto, J.P., Nielsen, J., Patil, K.R., Ferreira, E.C., Rocha, M., 2010. OptFlux: an open-source software platform for *in silico* metabolic engineering. BMC Syst. Biol. 4, 45.

Rocha, M., Maia, P., Mendes, R., Pinto, J.P., Ferreira, E.C., Nielsen, J., Patil, K.R., Rocha, I., 2008. Natural computation meta-heuristics for the in silico optimization of microbial strains. BMC Bioinformatics 9, 1–16.

Rosano, G.L., Ceccarelli, E.A., 2014. Recombinant protein expression in *Escherichia coli*: Advances and challenges. Front. Microbiol. 5, 1–17.

Saha, R., Chowdhury, A., Maranas, C.D., 2014. Recent advances in the reconstruction of metabolic models and integration of omics data. Curr. Opin. Biotechnol. 29, 39–45.

Saini, M., Li, S.Y., Wang, Z.W., Chiang, C.J., Chao, Y.P., 2016. Systematic engineering of the central metabolism in *Escherichia coli* for effective production of n-butanol. Biotechnol. Biofuels 9, 1–10.

Sajo Mienda, B., Shahir Shamsir, M., 2015. An overview of pathway prediction tools for synthetic design of microbial chemical factories. AIMS Bioeng. 2, 1–14.

Salis, H.M., 2011. Chapter two - The Ribosome Binding Site Calculator, in: Voigt, C.B.T.-M. in E. (Ed.), Synthetic Biology, Part B. Academic Press, pp. 19–42.

Sauer, M., 2016. Industrial production of acetone and butanol by fermentation-100 years later. FEMS Microbiol. Lett. 363, 1–4.

Schadeweg, V., Boles, E., 2016. Increasing n-butanol production with *Saccharomyces cerevisiae* by optimizing acetyl-CoA synthesis, NADH levels and *trans*-2-enoyl-CoA reductase expression. Biotechnol. Biofuels 9, 1–11.

Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., Söhngen, C., Stelzer, M., Thiele, J., Schomburg, D., 2011. BRENDA, the enzyme information system in 2011. Nucleic Acids Res. 39, D670–D676.

Schellenberger, J., Que, R., Fleming, R.M.T., Thiele, I., Orth, J.D., Feist, A.M., Zielinski, D.C., Bordbar, A., Lewis, N.E., Rahmanian, S., Kang, J., Hyduke, D.R., Palsson, B., 2011. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. Nat. Protoc. 6, 1290–1307.

Schuetz, R., Kuepfer, L., Sauer, U., 2007. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. Mol. Syst. Biol. 3, 119.

Segrè, D., Vitkup, D., Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. 99, 15112 LP-15117.

Shen, C.R., Lan, E.I., Dekishima, Y., Baez, A., Cho, K.M., Liao, J.C., 2011. Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli* . Appl. Environ. Microbiol. 77, 2905–2915.

Shen, C.R., Liao, J.C., 2008. Metabolic engineering of *Escherichia coli* for 1-butanol and 1-propanol production via the keto-acid pathways. Metab. Eng. 10, 312–320.

Shlomi, T., Berkman, O., Ruppin, E., 2005. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc. Natl. Acad. Sci. U. S. A. 102, 7695 LP-7700.

Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J., Li, Q., 2009. Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli* 18, 936–948.

Skorokhodova, A.Y., Morzhakova, A.A., Gulevich, A.Y., Debabov, V.G., 2015. Manipulating pyruvate to acetyl-CoA conversion in *Escherichia coli* for anaerobic succinate biosynthesis from glucose with the yield close to the stoichiometric maximum. J. Biotechnol. 214, 33–42.

Soini, J., Ukkonen, K., Neubauer, P., 2008. 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. 7, 26.

Steen, E.J., Chan, R., Prasad, N., Myers, S., Petzold, C.J., Redding, A., Ouellet, M., Keasling, J.D., 2008. Metabolic engineering of *Saccharomyces cerevisiae* for the production of n-butanol. Microb. Cell Fact. 7, 36.

Tas, H., Nguyen, C.T., Patel, R., Kim, N.H., Kuhlman, T.E., 2015. An integrated system for precise genome modification in *Escherichia coli*. PLoS One 10, 1–19.

Tashiro, Y., Yoshida, T., Noguchi, T., Sonomoto, K., 2013. Recent advances and future prospects for increased butanol production by acetone-butanol-ethanol fermentation. Eng. Life Sci. 13, 432–445.

Tepper, N., Shlomi, T., 2010. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics 26, 536–543.

Tian, P., Wang, J., Shen, X., Rey, J.F., Yuan, Q., Yan, Y., 2017. Fundamental CRISPR-Cas9 tools and

current applications in microbial systems. Synth. Syst. Biotechnol. 2, 219–225.

Tice, R.R., Austin, C.P., Kavlock, R.J., Bucher, J.R., 2013. Improving the Human Hazard Characterization of Chemicals: A Tox21 Update. Environ. Health Perspect. 121, 756–765.

Toyoda, K., Teramoto, H., Inui, M., Yukawa, H., 2009. The *ldhA* gene, encoding fermentative L-lactate dehydrogenase of *Corynebacterium glutamicum*, is under the control of positive feedback regulation mediated by LldR. J. Bacteriol. 191, 4251–4258.

Trinh, C.T., Li, J., Blanch, H.W., Clark, D.S., 2011. Redesigning *Escherichia coli* metabolism for anaerobic production of isobutanol. Appl. Environ. Microbiol. 77, 4894–4904.

Tudor, R., Ashley, M., 2007. Enhancement of industrial hydroformylation processes by the adoption of rhodium-based catalyst: Part II. Platin. Met. Rev. 51, 164–171.

Ulrich, H., Eekel, E., Koch, J., Fuchs, G., Linder, D., Buckel, W., 1993. Purification of glutaryl-CoA dehydrogenase from *Pseudomonas sp.* , an enzyme involved in the anaerobic degradation of benzoate 174–181.

Visioli, L.J., Enzweiler, H., Kuhn, R.C., Schwaab, M., Mazutti, M. a, 2014. Recent advances on biobutanol production. Sustain. Chem. Process. 2, 15.

Wang, L., Dash, S., Ng, C.Y., Maranas, C.D., 2017. A review of computational tools for design and reconstruction of metabolic pathways. Synth. Syst. Biotechnol. 2, 243–252.

Wang, Q., Ding, Y., Liu, L., Shi, J., Sun, J., Xue, Y., Ding, Y., Liu, L., Shi, J., Sun, J., Xue, Y., 2015. Engineering *Escherichia coli* for autoinducible production of n-butanol. Electron. J. Biotechnol. 18, 138–142.

Wasels, F., Jean-Marie, J., Collas, F., López-Contreras, A.M., Lopes Ferreira, N., 2017. A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. J. Microbiol. Methods 140, 5–11.

Weizmann, C., 1919. Improvements in the Bacterial Fermentation of Carbohydrates and in Bacterial Cultures for the same. 4845.

Welch, R.W., Rudolph, F.B., Papoutsakis, E.T., 1989. Purification and characterization of the NADH-dependent butanol dehydrogenase from *Clostridium acetobutylicum* (ATCC 824). Arch. Biochem. Biophys. 273, 309–318.

Wen, R.C., Shen, C.R., 2016. Self-regulated 1-butanol production in *Escherichia coli* based on the endogenous fermentative control. Biotechnol. Biofuels 1–15.

Wintermute, E.H., Lieberman, T.D., Silver, P.A., 2013. An objective function exploiting suboptimal solutions in metabolic networks. BMC Syst. Biol. 7, 98.

Wolfe, A.J., 2005. The Acetate Switch. Microbiol. Mol. Biol. Rev. 69, 12–50.

Wu, D., Wang, Q., Assary, R.S., Broadbelt, L.J., Krilov, G., 2011. A computational approach to design and evaluate enzymatic reaction pathways: Application to 1-butanol production from pyruvate. J. Chem. Inf. Model. 51, 1634–1647.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T.L., 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13, 134.

Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J.D., Osterhout, R.E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H.B., Andrae, S., Yang, T.H., Lee, S.Y., Burk, M.J., Van Dien, S., 2011. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. Nat. Chem. Biol. 7, 445–452.

Yu, A., Alexandra, G., Skorokhodova, Y., Debabov, V.G., 2012. Metabolic engineering of *Escherichia coli* for 1-butanol biosynthesis through the inverted aerobic fatty acid b -oxidation pathway 463–469.

Yu, J.-L., Xia, X.-X., Zhong, J.-J., Qian, Z.-G., 2017. A novel synthetic pathway for glutarate production in recombinant *Escherichia coli* . Process Biochem. 59, 167–171.

Zhang, H., Cheng, Q.X., Liu, A.M., Zhao, G.P., Wang, J., 2017. A novel and efficient method for bacteria genome editing employing both CRISPR/Cas9 and an antibiotic resistance cassette. Front. Microbiol. 8, 1–11.

Zhang, X., Jantama, K., Moore, J.C., Shanmugam, K.T., Ingram, L.O., 2007. Production of L-alanine by metabolically engineered *Escherichia coli* . Appl. Microbiol. Biotechnol. 77, 355–366.

Zhao, C., Zhao, Q., Li, Y., Zhang, Y., 2017. Engineering redox homeostasis to develop efficient alcohol-producing microbial cell factories. Microb. Cell Fact. 16.

Zheng, Y.N., Li, L.Z., Xian, M., Ma, Y.J., Yang, J.M., Xu, X., He, D.Z., 2009. Problems with the

microbial production of butanol. J. Ind. Microbiol. Biotechnol. 36, 1127–1138.

Zhou, P., Zhang, Y., Wang, P., Xie, J., Ye, Q., 2014. Butanol production from glycerol by recombinant *Escherichia coli* . Ann. Microbiol. 64, 219–227.

Zhu, L.W., Li, X.H., Zhang, L., Li, H.M., Liu, J.H., Yuan, Z.P., Chen, T., Tang, Y.J., 2013. Activation of glyoxylate pathway without the activation of its related gene in succinate-producing engineered *Escherichia coli*. Metab. Eng. 20, 9–19.

Appendix

Appendix A: List of reactions suggested as knock-outs in the optimization process to improve butanol production from Chapter 3

Table A 1: List of reactions removed in the optimization process from Chapter 3.

ID	Reaction
R_ALCD2x	ethanol + NAD^+ <=> acetaldehyde + NADH + H^+
R_PTAr	phosphate + acetyl-CoA <=> acetyl-phosphate + coenzyme A
R_ACALD	NAD^+ + acetaldehyde + coenzyme A <=> NADH + acetyl-CoA + H^+
R_FUM	Fumarate + H_2O<=> (L)-malate
R_LDH_D	NAD^+ + (R)-lactate <=> NADH + pyruvate + H^+
R_KARA1	$NADP^+$ + (R)-2,3-Dihydroxy-3-methylbutanoate <=> H^+ + NADPH + (S)-2-Acetolactate
R_ACOAD1f	Butanoyl-CoA + FAD <=> Crotonyl-CoA + $FADH_2$ + H+
R_HACD2	(S)-3-hydroxyacyl-CoA + NAD^+ = 3-oxoacyl-CoA + NADH + H+
R_CTECOAI7	3-*cis*-Dodecenoyl-CoA <=> 2-trans-Dodecenoyl-CoA
R_CTECOAI6	3-*cis*-Dodecenoyl-CoA <=> 2-trans-Dodecenoyl-CoA
R_ME2	(S)-Malate + $NADP^+$ <=> Pyruvate + CO_2 + NADPH + H+
R_RPI	D-Ribose 5-phosphate <=> D-Ribulose 5-phosphate
R_SUCD	Quinone + Succinate <=> Hydroquinone + Fumarate
R_FESR	[3Fe-4s]+ Fe_{2+} <=> [4Fe-4s]
R_ALDD4	Butanal + H_2O+ NAD+ <=> Butyrate+ 2H + NADH
R_AACPS1	(3R)-3-Hydroxyacyl-[acyl-carrier protein] + $NADP^+$ <=> 3-Oxoacyl-[acyl-carrier protein] + NADPH + H^+
R_3OAR120	(3R)-3-Hydroxyacyl-[acyl-carrier protein] + $NADP^+$ <=> 3-Oxoacyl-[acyl-carrier protein] + NADPH + H^+
R_AACPS3	ATP + Long-chain fatty acid + Acyl-carrier protein <=> AMP + Diphosphate + Long-chain acyl-[acyl-carrier protein]
R_FTHFD	10-Formyltetrahydrofolate + H2O <=> Formate + Tetrahydrofolate
R_CPPPGO	Coproporphyrinogen III + Oxygen <=> Protoporphyrinogen IX + 2 CO2 + 2 H2O
R_FE3Ri	$FADH_2$ + $2Fe^{3+}$ <=> FAD + 2 Fe^{2+} + $2H^+$

Appendix B: DNA Sequences of the Codon-Optimized Genes

Table B 1: DNA Sequences of the codon-optimized genes.

Gene	Sequence
hgdH	ATGAAAGTGCTGTGCTACGGCGTTCGCGATGTGGAACTTCCGATCTTCGAGGCATGTAACAAAGAATTTGGTTATGATATTAAAT GCGTCCCGGATTATCTGAATACCAAAGAAACGGCAGAAATGGCCGCCGGCTTTGACGCAGTGATTCTCCGTGGCAACTGCTTCG CAAACAAACAGAACCTGGATATCTACAAAAAACTGGGCGTCAAATATATTCTGACGCGTACCGCAGGCACGGACCACATCGACA AAGAATACGCGAAAGAACTTGGCTTTCCGATGGCCTTCGTGCCACGCTACAGTCCGAACGCCATCGCCGAGCTGGCCGTTACC CAGGCGATGATGCTGCTCCGCCATACCGCGTATACGACCTCCCGCACCGCAAAAAAGAACTTCAAGGTGGATGCGTTTATGTTC AGTAAAGAAGTGCGCAATTGCACCGTCGGTGTTGTTGGCTTAGGCCGTATTGGCCGCGTCGCGGCGCAGATCTTTCATGGTATG GGTGCCACTGTGATCGGCGAAGACGTTTTCGAAATCAAAGGCATTGAGGATTACTGCACTCAAGTGAGCTTGGACGAGGTTCTG GAAAAGAGTGATATTATCACCATCCACGCGCCATATATTAAAGAAAACGGTGCAGTTGTCACCCGCGATTTTCTGAAAAAGATGA AAGACGGTGCGATTTTGGTGAACTGCGCACGCGGTCAGCTTGTTGACACTGAGGCAGTTATTGAAGCGGTGGAAAGCGGTAAAC TTGGCGGCTACGGTTGTGATGTGTTGGATGGTGAGGCGAGCGTCTTTGGTAAGGATCTCGAAGGCCAAAAACTGGAAAACCCGT TATTCGAGAAACTTGTCGATCTTTATCCGCGTGTGCTGATTACGCCGCATCTCGGCTCCTATACGGACGAGGCCGTGAAAAATAT GGTGGAAGTGAGCTACCAAAATCTCAAAGATCTGGCCGAAACTGGCGATTGCCCGAACAAAATCAAATAA
gctA	ATGAGCAAAGTCATGACCCTGAAAGACGCAATCGCAAAATACGTCCACAGCGGCGACCACATCGCCCTCGGCGGCTTCACCAC CGATCGCAAACCATACGCCGCGGTGTTCGAAATCCTGCGCCAAGGCATCACCGATCTGACTGGTTTGGGCGGCGCCGCCGGTG GTGATTGGGACATGCTGATTGGTAACGGTCGCGTGAAGGCGTATATCAACTGCTACACCGCGAACAGTGGCGTTACCAACGTGA GTCGCCGTTTTCGCAAATGGTTCGAAGCGGGTAAACTGACTATGGAAGACTATTCGCAAGATGTGATCTACATGATGTGGCATGC GGCAGCGTTGGGTCTTCCATTTTTACCGGTTACGTTGATGCAGGGCAGCGGCCTGACTGATGAATGGGGCATTAGCAAGGAGGT TCGCAAAACTCTTGATAAGGTTCCGGATGACAAGTTCAAATACATTGATAATCCGTTCAAACCGGGTGAAAAAGTTGTGGCGGTT CCGGTTCCGCAGGTGGATGTCGCCATTATTCACGCACAGCAGGCATCGCCGGATGGTACCGTCCGTATCTGGGGTGGTAAATTT CAGGACGTTGACATTGCGGAAGCGGCCAAATATACGATTGTGACGTGCGAGGAGATTATCTCCGATGAGGAGATCCGTCGTGAT CCGACCAAAAACGATATTCCGGGCATGTGCGTGGATGCGGTTGTGCTTGCCCCATATGGTGCGCATCCGTCCCAGTGCTATGG CCTGTATGATTACGATAACCCGTTTCTGAAAGTGTATGATAAAGTGTCGAAAACCCAGGAAGACTTTGATGCCTTTTGTAAGGAAT GGGTCTTCGACCTGAAAGATCACGATGAATATCTTAATAAACTTGGCGCAACCCGTCTGATTAATTTAAAAGTGGTGCCGGGCCT GGGCTATCACATTGATATGACGAAAGAAGATAAATAA
gctB	ATGGCCGACTACACCAACTACACCAACAAAGAAATGCAGGCCGTGACCATCGCCAAACAGATCAAAAACGGCCAGGTGGTGACT GTTGGCACCGGCCTGCCGCTGATCGGTGCGAGCGTCGCAAAGCGCGTTTATGCGCCGGATTGCCACATTATCGTCGAAAGTGG CCTGATGGATTGCAGTCCGGTGGAAGTTCCGCGCTCGGTGGGTGATTTGCGCTTCATGGCACACTGCGGCTGTATTTGGCCAAA TGTCCGTTTCGTGGGCTTTGAGATCAACGAATATCTGCATAAAGCCAATCGCCTTATTGCGTTTATTGGTGGTGCCCAGATTGAC CCGTACGGCAACGTCAACAGCACGTCCATCGGCGATTATCATCACCCAAAAACCCGTTTCACCGGCAGCGGTGGCGCAAATGG TATCGCCACTTACAGTAATACGATCATTATGATGCAACATGAAAAACGCCGCTTTATGAACAAAATCGACTATGTGACCTCGCCG GGCTGGATTGATGGCCCGGGTGGCCGTGAACGTCTCGGTCTGCCGGGTGATGTGGGTCCACAGTTGGTTGTGACCGATAAGGG TATCCTGAAATTCGACGAAAAAACCAAACGCATGTATCTGGCAGCCTATTACCCGACGTCCAGCCCAGAGGACGTGCTGGAAAA CACCGGCTTTGACCTGGATGTGAGCAAAGCAGTGGAACTGGAAGCGCCGGACCCAGCAGTGATCAAACTGATTCGCGAAGAAA TCGATCCGGGCCAAGCATTTATTCAGGTTCCAACTGAAGCCAAATAA

hgdA	ATGGCAAAACAAGTCAGCCCAGGCGTGCTGGCCCTGCGCAAAGTTGTGGACGATGTGCACAAGGAAGCGCGTGAAGCCAAAGC CCGTGGCGAGTTGGTGGGTTGGAGTTCGTCCAAATTCCCGTGCGAACTTGCGGCAGCCTTTGATCTGAACGTCATGTACCCGGA AAATCAGGCGGCAGGTATCGCAGCAAATCGTTATGGCGAAATGATGTGCCAGGCAGCAGAAGATCTTGGTTATGATAATGATATT TGTGGCTATGCACGCATTAGTCTCGCATACGCCGCCGGCGTGCGTGTTTCCCGCAAATATGACGCAGAAACCGGTGAGTACATT ATCGATCCGGCAACGGGCAAACCGCTGAAAGATGCGGAAGGCAACGTCGTCATTGACGAAGCAACTGGCAAGCCGAAAAAAGA CCCAAAAACGCAGACCCCGTACTTAGTCCTGGACAATCTGCTGGAAATCGAAGCACTGCCGGACGGCCCAGAAAAAGAACGCC GTCTGGAAGCAATCAGCCCAATTCGCCAGATGCGCATCCCACAGCCAGACTTTGTCCTCTGTTGCAACAACATCTGTAATTGCA TGACGAAATGGTATGAAAACATCGCACGCATGTGCAACGTCCCACTCATCATGATCGACATCCCATACAACAACACCGTCGAAG TCCATGACGATAACGTGAAATACGTGCGCGCGCAATTTGACAAAGCAATCAAACAGCTGGAAGAACTGACCGGCAAAAAATTCG ACGAAAAAAAATTTGAAAAAGCCTGTAGTAACGCGAACCGCACTGCGCAGGCCTGGCTCAAAGTCTGCGACTACCTCCAATACA AACCAGCACCGTACTCGGGCTTCGATCTTTTCAACCATATGGCCGACGTCGTGACCGCGCGCGCGCGCGTCGAAGCGGCAGAA GCGTTTGAACTGCTGGCCGACGACCTCGAGGAAACCGTCAAAAAGGGCGAAACCACCACCCCGTTCCCAGAAAAATACCGCGT CATGTTCGAAGGCATTCCATGCTGGCCGAAACTGCCAAACTTATTCAAACCACTCAAAGAACATGGCGTGAACGTTACCGCAGT CGTGTACGCCCCAGCGTTCGGCTTCGTTTATAACAATATCGACGAAATGGCACGCGCCTACTACAAAGCCCCAAACAGCGTGTG CATCGAACAAGGCGTGGATTGGCGCGAGGGTATCTGCCGCGACAACAAAGTCGATGGCGTGCTGGTTCACTACAACCGTAGCT GCAAACCGTGGAGCGGCTATATGGCAGAGATGCAACGCCGCTTCACCGAGGATCTTGGCGTTCCATGCGCCGGCTTCGATGGT GATCAAGCCGACCCGCGCAACTTTAATGCAGCCCAATACGAAACCCGCGTTCAAGGCCTTGTGGAGGCAATGGAAGCCAACAA GCAAGCGAAGGAAGCAAAGTAA
hgdB	ATGAGCATCAACGCACTGTTGGACGAATTTAAAGTGAAAGCGGCCACGCCGAAACAACAACTTGCGGAGTATAAAGCGCAGGGT AAAAAGGTTATCGGCGTCCTGCCATACTACGCGCCAGAAGAATTGGTGTATGCGGCCGGCATGGTGCCGATGGGCATCTGGGG CTCCAATAATAAAACCATCTCCCGTGCGAAAGAGTACTGCGCGACCTTCTATTGCACGATCGCCCAATTGGCACTGGAGATGCT GCTGGATGGCACGATGGACCAGCTGGACGGCATTATCACCCCGACGATTTGCGACACCTTGCGCCCAATGAGCCAGAACTTTC GCGTGGCGATGGGCGATAAGATGGCGGTCATTTTCTTGGCGCAACCGCAGAATCGTTTCGAAGACTTCGGCCTCCAATTTAGCG TTGATCAGTATACCAACGTCAAGAAAGAACTGGAAAAAGTGGCCGGTAAAGAAATTACCAACGAGGCGATTCAAGATGCCATTAA AGTTTACAACAAGAGTCGCGCCGCACGTCGTAAGTTCGTTGAGTTGGCGTCCGCCCATTGCGATGTGATTACCCCGACCAAACG CAGCGCGGTGTTGAAATCCTTTTTTTTCATGGAAAAGCCGGAATACATTGAGAAGCTGGAGGAACTCAACGCAGAACTGGAAAAA TTACCAGTGTGTGATTGGCAGGGCACCAAAGTGGTCACCAGCGGCATCATTTGCGATAACCCAAAACTGCTTGAGATTTTTGAAG AGAACAACATTGCAATTGCGGCGGATGATGTGGGCCACGAGAGTCGCAGCTTCCGTGTGGACGCACCAGAGGATGAGGCCGAT GCCCTTATGGCGTTGGCCAAGCAATTCGCAAACATGGATTATGATGTTTTGCTGTATGATCCGAAAAGCACCGAAAATCGTCGCG GCGAGTTCATCGCCAACATGGTGAAAGAGAGCGGCGCCCAGGGTCTTGTGTTGTTTATGCAGCAGTTCTGCGATCCAGAAGAAA TGGAGTACCCGTATCTGAAAAAGGCCCTGAACAACGCAGGCATTCCGCACATTAAGTTGGGCATTGATCAGCAGATGCGTGATT TTGGTCAGGCGAGTACTGCAATCCAGGCGTTTGCCGACGTTTTGGAGATGCAAAAATAA
hgdC	ATGAGTATCTATACCCTGGGCATCGATGTGGGCAGCACCGCCAGTAAATGCATCATTCTGAAAGACGGCAAAGAAATCGTGGCC AAAAGTCTGGTGGCCGTGGGCACCGGCACCAGTGGCCCGGCCCGCAGTATTAGCGAAGTGCTGGAAAACGCCCACATGAAAAA AGAAGATATGGCCTTCACCCTGGCAACGGGCTACGGTCGCAACAGTCTGGAAGGCATCGCCGATAAACAGATGAGCGAACTGT CCTGCCACGCAATGGGTGCGTCCTTCATTTGGCCGAACGTGCATACCGTCATCGATATTGGTGGCCAAGATGTGAAAGTCATTC ATGTTGAAAACGGTACGATGACCAACTTTCAGATGAACGACAAATGCGCGGCAGGCACTGGCCGCTTTCTGGATGTTATGGCGA ACATCCTGGAAGTCAAGGTTAGCGACTTAGCAGAACTTGGCGCAAAATCCACGAAACGCGTCGCAATTAGCAGCACCTGCACTG TCTTCGCAGAGAGCGAGGTCATCAGCCAGCTCAGCAAGGGCACCGATAAAATCGATATCATCGCAGGCATCCACCGCAGTGTG GCCAGCCGCGTCATCGGCTTAGCAAATCGCGTGGGCATCGTCAAAGACGTTGTCATGACCGGTGGCGTTGCGCAGAACTATGG TGTTCGTGGTGCACTGGAAGAGGGCCTCGGCGTCGAAATTAAAACCTCCCCGCTCGCGCAGTATAATGGCGCGCTGGGTGCAG CCCTGTACGCATATAAAAAAGCGGCAAAATAA

gcdH	ATGGCAACCAAAGCAAGCTTCAACTGGGAAGACCCACTGCTCTTGGACCAACAACTGACCGAAGAGGAACGCATGGTCCGCGA TAGCGCGCAGCAATTCGCCCAGGACAAACTGGCCCCACGTGTCCTTGAAGCGTTTCGCCACGAACAGACTGATCCGAAGATCT TCCGCGAAATGGGCGAAACCGGCCTGCTGGGTGCCACCATCCCGGAAGCCTATGGCGGCTCGGGTCTGAACTATGTTTGCTAC GGTTTGATTGCACGTGAGGTGGAACGTGTGGATAGCGGCTATCGTAGTATGATGTCCGTGCAGTCCAGTCTGGTGATGGTGCCG ATCCACGAATTTGGCAATGAGGCGACGCGCCAGAAATACCTGCCGAAATTAGCCAGCGGTGAGTATATCGGCTGCTTTGGCTTA ACGGAACCGAACCATGGTTCGGATCCGGGTAGCATGGTGACGCGTGCCAAAAAAGTTGATGGCGGTTACCGCCTGAGTGGCTC GAAAATGTGGATTACCAACTCCCCAATTGCGGACGTGTTTGTGGTTTGGGCGAAAGATGACGAGGGCCAAATTCGCGGCTTTGT TTTGGAAAAGGGTTGGGAAGGTCTGAGCGCCCCGGCGATCCACGGCAAAGTGGGCCTCCGTGCCTCGATTACTGGTGAAATCG TTATGGATAATGTGTTCGTCCCAGAGGAGAATGCATTTCCGGAAGTCCGTGGTCTGCGCGGTCCGTTCACCTGTTTAAATAGTGC ACGTTACGGTATTAGCTGGGGCGCGTTAGGCGCGGCAGAATTTTGCTGGCATACCGCCCGCCAGTACGTTCTCGATCGCCAAC AGTTTGGCCGCCCGCTGGCGGCGAACCAGTTAATTCAGAAAAAGTTAGCGGATATGCAAACTGAGATTACGTTAGCCCTGCAGG GCTGCCTGCGTCTGGGTCGTATGAAGGACGAGGGTACCGCAGCAGTTGAGATCACTAGCATTATGAAACGCAATTCGTGTGGCA AGGCACTGGATATCGCACGCCTGGCGCGTGATATGCTTGGTGGTAACGGCATCTCGGACGAATTTGGTATCGCGCGTCATTTAG TCAACCTGGAAGTGGTTAATACCTATGAAGGTACCCATGACGTTCATGCACTGATTTTAGGTCGTGCCCAGACCGGTATTCAAGC GTTCTTTTGA