Understanding and Engineering Cofactor Metabolism

A Quest for Improved Biochemical Production

A THESIS PRESENTED
BY
MISS LAURA DE ARROYO GARCIA
TO
THE DEPARTMENT OF LIFE SCIENCES

FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

IMPERIAL COLLEGE LONDON
LONDON, UK

SEPTEMBER 2020

Abstract

In the field of metabolic engineering, where cells are treated as "factories" that synthesise industrial compounds, it is essential for cell metabolism to accommodate the energy and redox cofactor demands of synthetic pathways. A balanced supply and consumption of ATP and NAD(P)H directly influences biotechnological performance. This study develops computational and experimental frameworks to explore how ATP and NAD(P)H limit the yield of synthetic pathways during bioproduction.

Constraint-based modelling was used to develop a novel computational protocol, CBA (Cofactor Balance Assessment), which tracks how ATP and NAD(P)H contribute to cell target production, as opposed to cell maintenance, biomass and waste release, in the presence of a synthetic pathway. Using butanol pathways (a non-native product in *E.coli*) with varying cofactor demands, CBA discerned the network-wide effects of cofactor variations on butanol yield. Results indicate that yields could be boosted by up to 13% if the introduced pathway is balanced both in terms of energy and redox. CBA simplified cofactor balance assessments and provided insights into how to improve the efficiency of recombinant strains.

Physiological and metabolic responses to cofactor perturbations were also experimentally assessed. The predominant phenotypes of strains harbouring the ATP synthase and PCK knockouts included high glycolytic flux, lower biomass and ATP. These strains were used to improve ethanol production, resulting in yields 10% and 29% higher than the WT overproducing ethanol and reaching over 70% of the theoretical maximum. The low but positive ATP yields boosted ethanol production and minimised unrestricted growth.

This research posits that early-stage *in silico* cofactor usage profiling serves as an instrument to select better performing pathways. Significant yield improvements can be achieved experimentally with a small number of cofactor-driven modifications that reduce the waste of cofactors, illustrating the potential of these strains as platforms to improve bioproduction of cofactor-neutral or cofactor-surplus synthetic pathways.

Acknowledgements

Many remarkable individuals have contributed to the completion of this thesis.

First and foremost, I would like to thank my sponsors, the UK's Biotechnology and Biological Sciences Research Council (BBSRC), who fully funded this work. This study could not have come to fruition without their financial support. Imperial College London also deserves a very special mention for providing such a powerful research ecosystem, and facilitating an endless number of enriching academic opportunities that complemented my studies and fueled my curiosity. Having completed my Undergraduate, Masters and Doctoral studies at the South Kensington campus, Imperial College will always have a very special place in my heart.

I would like to express my most sincere gratitude to my supervisor, Dr Patrik Jones, for believing in me to carry out this research, encouraging me to be independent, and helping me shape and develop my research questions. But most importantly, I thank you for your support with my career decisions and for never losing faith in my work.

I would like to thank John Pinney and the Bioinformatics and Theoretical Systems Biology group at Imperial College London for co-supervising this work, Dr Suhail Islam for always making himself available to resolve any computing-related issues, and Dr Virginia Fairclough, whose guidance, support and friendship I will cherish forever. I also thank the members of my progress review panel, Dr John Heap and Prof. Guy-Bart Stan, for their ongoing feedback throughout my PhD studies. Many more research groups contributed to this research, from providing resources, to sharing knowledge and expertise. For convenience, all technical contributions are acknowledged in detail in the relevant sections of this thesis.

I am very grateful to Allen & Overy LLP for hosting my 3-month PhD internship, in particular the International IP Litigation team for uncovering a career path that I love. Special mentions to Shohta, Daniel, Neville, Amanda and Oscar for making me feel part of the team. And Patricia, I could simply not imagine this experience without you.

Huge thanks to every single one of my research colleagues in the Microbial and Metabolic Engineering Group for the friendly atmosphere, invaluable discussions, and for committing your time, effort and patience to improving my research. In no particular order: Dr Giorgio Perin, Marine Valton (then a PhD student), Dr Daniel Zabala, Dr Ian Yunus, Jessica Rollit (then a PhD student), Pachara Sattayawat (then a PhD student), Jonathan Muller, Arianna Palma, Mathieu Bousquet (then a PhD student), Dr Emanga Alobwede and Dean Keck. You will be greatly missed. Especial thanks to Dr John Rowland, who moved on just as this thesis began to come to life; Dani, for all the coffee chats; and Marine, I am certain this is only the beginning of a beautiful and long-lasting friendship. Thank you also to my friends in the PhD cohort, SAF was a better place because of all of you.

My warmest and deepest appreciation goes to my close and loving friends, family and family friends, my parents and brother for unconditionally believing in me. Especially, to my housemate Cristina, for joining along for the ride and for keeping me in check every time life got difficult; and Lara, you quite literally carried me through the final few months of writing when I had no strength to carry myself. I thank you both endlessly.

Borja, Papi, Mami, gracias por ser el mejor equipo, por enseñarme que la actitud es lo más importante y que no hay piedra que se interponga en el camino que no pueda levantar. Gracias por vuestro apoyo incondicional.

Truly, this thesis would not exist if it wasn't for all of you.

Para Borja y mis padres.

Se cierra una puerta. Se abre un ventanal.

"Your soul needs time for solitude and self-reflection in order to love, lead, heal and create"

Linda Joy

Declaration: I, Laura de Arroyo Garcia, hereby confirm that this dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. I further state that no substantial part of my dissertation has been previously submitted, or is being concurrently submitted, for a degree or diploma or other qualification at Imperial College London or any other University or similar institution except as specified in the text. I also confirm that this dissertation does not exceed the word limit set by the Degree Committee.

Publications

Material included in this thesis has been published and presented at conferences and other events.

Journal papers

1. **de Arroyo Garcia, L.** and Jones, P. R. (2020) *In silico* co-factor balance estimation using constraint-based modelling informs metabolic engineering in *Escherichia coli*. PLoS Comput Biol 16(8): e1008125. https://doi.org/10.1371/journal.pcbi.100812

Conferences

1. **de Arroyo Garcia, L.**, Pinney, J. W. and Jones. P. R. A systems approach to evaluating the balance of synthetic metabolic pathways. Abstract and Poster presenter at the *Data-driven Biotechnology*, Hillerod, Denmark, May, 2017.

Other presentations and meetings

1. *In silico* evaluation of cofactor balancing in synthetic metabolic pathways. Poster presenter at the *Imperial College Natural Sciences Research Day*, London, United Kingdom, September, 2018. L. de Arroyo Garcia, J. W. Pinney and P. R. Jones.

2. Systems-based evaluation of the balance of synthetic metabolic pathways. Poster presenter at the *BBSRC visit at Imperial College London*, London, United Kingdom, December, 2018. L. de Arroyo Garcia, J. W. Pinney and P. R. Jones.

3. The role of cofactor (im)balance in engineered metabolism. Presented my work to a panel of judges and the Life Sciences department at Imperial College London, London, United Kingdom, April, 2019. L. de Arroyo Garcia, J. W. Pinney and P. R. Jones.

Contents

List of Figures

List of Tables

Glossary

List of Abbreviations

ACP	Acyl Carrier Protein	GS	Gas Spectrometry
ADP	Adenosine Diphosphate	GSM	Genome-scale Model
AMP	Adenosine Monophosphate	HPLC	High-performance Liquid Chromatography
ATP	Adenosine Triphosphate	ID	Identifier
ATPase	ATP synthase	IPCC	Intergovernmental Panel on Climate Change
BBSRC	Biotechnology and Biological Sciences Research Council	KO	Knockout
BiGG	biochemically, genomically and genetically	LB	Luria Broth
CAR	Carboxylic Acid Reductase	MFA	Metabolic Flux Analysis
CCM	Central Carbon Metabolism	MS	Mass Spectrometry
CFS	Co-factor Flux Score	NAD	Nicotinamide Adenine Dinucleotide
COBRA	Constraint-based Reconstruction and Analysis	NADP	Nicotinamide Adenine Dinucleotide Phosphate
COBRApy	Constraint-based Reconstruction and Analysis for Python	OECD	Organization for Economic Cooperation and Development
COP	Conference of the Parties	PCR	Polymerase Chain Reaction
DMF	Dimethylformamide	PCK	phosphoenolpyruvate decarboxylase
DNA	Deoxyribonucleic Acid	PDC	pyruvate decarboxylase
E.coli	*Escherichia coli*	PDH	Pyruvate Dehydrogenase
EOD	Earth's Overshoot Day	pFBA	parsimonious Flux Balance Analysis
ETC	Electron Transport Chain	PPP	Pentose Phosphate Pathway
FAS	Fatty Acid Synthesis	r.p.m	revolutions per minute
FBA	Flux Balance Analysis	RPC	Relative Potential Calculation
FVA	Flux Variability Analysis	SDGs	Sustainable Development Goals
GENRE	Genome-scale Network Reconstruction	SBML	Systems Biology Markup language
GFP	Green Fluorescent Protein	TCA	Tricarboxylic Acid
		WT	Wild Type

"At the heart of the challenge is the need to decouple economic growth from environmental degradation [...] Grand challenges need not be insurmountable obstacles leading to economic despair, but rather the chance to rebuild industry and society in a sustainable manner" - (OECD, 2018)

1 Motivation

1.1 Setting the scene:

1.1.1 Global challenges faced by society's overall sustainability

Environmental, political and socio-economic factors, as well as humanity's unprecedented demand on global resources, are transforming our manufacturing needs. Our planet's population growth has sky-rocketed to an annual rate of 83 million people per year (United Nations, 2017). People are living longer, expecting greater mobility, and quicker access to improved products and services (HM Government UK, 2018). As technological developments flourish, so increases our global carbon footprint.

There is growing awareness of the negative environmental effects of fossil fuels, as well as their finite nature and fast diminishment (OECD, 2018). Still, oil, coal and natural gas continue to account for around 80% of the total energy supplied globally (BP, 2019). However, due to the fast accumulation of global atmospheric greenhouse gas emissions, the increase in air pollution, urban development and the industrialisation of developing countries, as well as the energy demand from an ever wider range of industries (IPCC, 2018), we cannot rely on fossil fuels to meet our demands any longer (HM Government UK, 2018). It is our duty to find sustainable alternatives, protect the environment and tackle the effects of climate change.

Growing calls for more sustainable industrial processes have fed the quest for sustainable energy

sources. After making it into the United Nations agenda as part of the global sustainable development goals (SDGs), first proposed in 2015 (United Nations, 2015), and becoming a crucial theme in the 2015 Paris Climate Agreement (UNFCC, 2015), this search is now a major global milestone. However, this issue moves beyond politics. 2019 marked the year with the highest number of wildfires in the Amazon (Escobar, 2019), the highest temperatures on record (NOAA, 2019, WMO, 2020), and the earliest Overshoot Day [1] thus far (WWF, 2019). As extreme weather gauged mediatic protagonism, exemplified by the catastrophic floods in East Africa in early 2019 through to the Bahamas devastation following Hurricane Dorian only a few months later, it is not surprising that the generations of today and tomorrow are demanding urgent climate action. The global climate urgency is faced up against conflicting national priorities and political agendas, making it hard to align environmental policies at a global scale. While some world leaders have attempted to enforce regulations and policies to maintain low emissions through the Conference of Parties (COP), countries like the United States (USA) have ceased participation in the Paris Agreement, and cut oil prices and funding for clean energy projects. Moreover, oil prices have been subject to great price fluctuations in recent years, driven by volatile market dynamics and the tense geopolitical landscape of supplier countries. Further uncertainty as to what the petrochemical industry might look like in the years to come has been triggered by the widespread industry and travel restrictions arising from the covid-19 crisis, which caused the oil prices to plunge to record negative levels in 2020 (IEA, 2020)

Therefore, if we are to sustain the global population whilst remaining below the 2°C threshold set out by the Paris Agreement, and ease any political tensions while achieving energy independence, it is in everyone's interest that the next generation of energy sources is pursued, having minimal carbon emissions (or ideally are carbon negative), improved chemical and physical properties (Nielsen and Keasling, 2016), and entail less competition for natural resources (Kitney et al., 2012, Zhou et al., 2008). This search will define our generation, and have an unprecedented impact on our industries and how we shape our future development. Investing in the bioeconomy is one such strategy to realise sustainable industries and societies.

1.1.2 The emergence of the bioeconomy

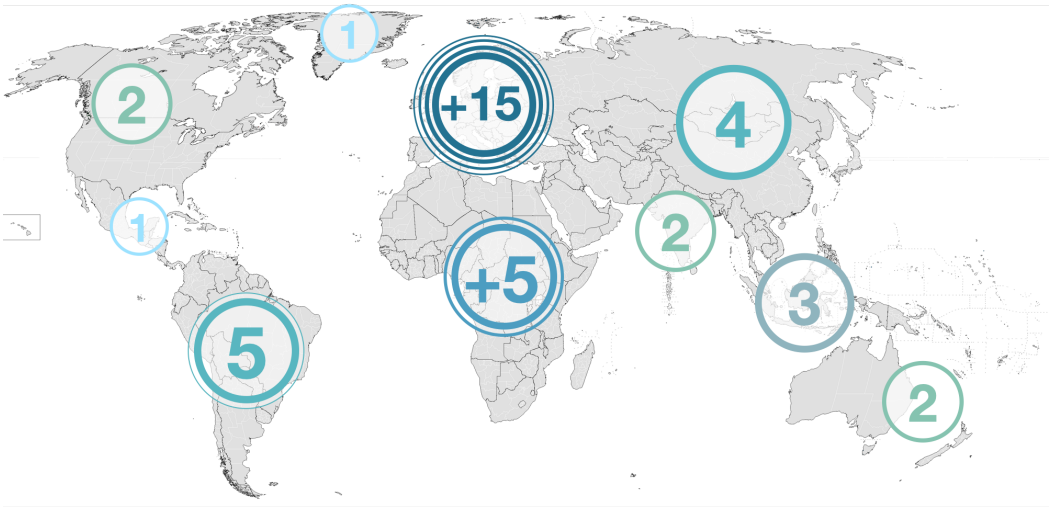
The bioeconomy is the set of economic activities that harnesses bio-based processes based on biotechnology, instead of on their fossil-fuel-derived counterparts. In other words, it is the econ-

[1] Earth Overshoot Day (EOD), is commonly defined as the calculated date on which society's resource consumption exceeds Earth's capacity to regenerate those resources that year.

omy's smaller pocket based on biology and the biosciences. Since 2009, the bioeconomy has been at the forefront of politics (OECD, 2018). In 2014, 13% of the world's trade involved bio-based products, worth an average of $2trillion (El-Chichakli, 2016). More specifically at home, the UK bioeconomy contributed over 5.2 million jobs in the country and was worth £220bn gross value added that same year (HM Government UK, 2018). By now, over 50 nations globally have proposed to boost their bioeconomy (Figure 1.1) (OECD, 2018), with over 35% of all products expected to be partly or fully manufactured using industrial biotechnology by 2030 (Maia et al., 2016). The bioeconomy has also grown from being biologics-centric to becoming a platform technology spreading across a vast number of sectors such as agri-food, chemicals, materials, health and biologics, energy and fuel, and the environment (Carlson, 2016). In the UK, the sector is growing at a rate of 8-10% per year and is predicted to continue beyond 2025 (BBSRC, 2019).
A world-class, robust and diversified bioeconomy would remove reliance on finite resources (i.e. fossil fuels) and disrupt how we address societal challenges. These developments are expected to deliver healthier foods, more productive agricultural systems, better medicines, sustainable fuels, and novel materials whilst contributing to a cleaner environment. The potential benefits are so great that the UK government has identified it as one of the four "Grand Challenges" within its own Industrial Strategy, and is expecting to double its size within the next 10-15 years (HM Government UK, 2018).

1.1.3 The potential of microbial production with metabolic engineering

The shift from chemical to biotechnological synthesis of desired chemicals and fuels (hereinafter called "biosynthetic production", or "bioproduction") represents one of the cornerstones of sustainable production processes. Biosynthetic production may proceed via isolated enzymatic reactions (i.e. *"in vitro"*) or via whole-cell systems (*"in vivo"*, also called microbial production) (Dai et al., 2015, Stephanopoulos, 2012). The use of biological organisms for chemical production, although it has existed for thousands of years, has been increasingly exploited as an alternative to chemical conversions (Nielsen and Keasling, 2016). Microbes are inherently advantageous factories because of their self-regeneration capabilities and extended life cycles (Siedler et al., 2011). They have naturally evolved over millions of years to carry out essential biochemical conversions to guarantee their own survival, without the need for enzyme purification and processing. Their metabolic diversity also enables them to thrive on a large repertoire of substrates. Numerous microorganisms can produce a range of commercially relevant molecules, or their precursors, by harnessing their existing catalytic machinery and metabolism (Clomburg et al., 2017), albeit at low production



The Americas	Europe			Africa	Asia	Oceania
• Canada	• Portugal	• Denmark		• Senegal	• India	• Australia
• United States	• Spain	• Norway		• Mali	• Sri Lanka	• New Zealand
• Mexico	• France	• Finland		• Nigeria	• China	
• Colombia	• Italy	• Sweden		• Uganda	• Korea	
• Greenland	• United Kingdom	• Luxembourg		• Kenya	• Russia	
• Brazil	• Belgium	• Estonia		• Tanzania	• Thailand	
• Paraguay	• Ireland	• Lithuania		• Mozambique	• Malaysia	
• Argentina	• Iceland	• Germany		• South Africa	• Indonesia	
• Uruguay	• Austria	• Turkey		• Namibia	• Japan	

Figure 1.1: Regions implementing bioeconomy policies span 5 continents. This includes countries with dedicated bioeconomy strategies and bioeconomy-related strategies. It also accounts for countries with strategies that are currently under development. (Adapted from OECD (2018))

rates (Liao et al., 2016). The idea of using genetic modifications to improve production has been widely employed as an exciting avenue for higher yield accumulation since the 1990s, when the field of metabolic engineering was first coined (Bailey, 1991).

Metabolic engineering, also recently termed synthetic metabolism (Erb et al., 2017), involves rerouting a microbe's metabolism to drive the carbon flux towards particular pathways for optimal production of selected chemicals (Nielsen and Keasling, 2016). Metabolic engineering strategies may involve fine-tuning existing microbial pathways to optimise native production, or introducing new pathways and enzymes into an organism of interest. Even though native producers may exist, heterologous production has unarguably taken much of the recent focus, because selected organisms may be easier to grow from a practical and economic perspective, and vast amounts of

literature may facilitate their use, application, engineering and characterisation.

1.1.4 Biosynthetic production in *Escherichia coli*

Metabolic engineering has popularised the use of model systems as platform strains, such as *Saccharomyces cerevisiae*, *Escherichia coli* or *Bacillus subtilis*. The gram negative bacterium *Escherichia coli* (*E.coli*) is one of the most important organisms used in the industry, and was selected as the organism of choice for this PhD study. *E.coli* has been extensively studied, both at the fundamental and applied levels and, as a result, a myriad of molecular cloning and genome editing techniques are now available. Furthermore, *E.coli* has excellent properties, including rapid proliferation and growth rates, high cell density, and low production costs (Pontrelli et al., 2018), as well as the fact that it is easy to handle and culture and is robust at industrial scale (Burdette et al., 2018). These characteristics make *E. coli* one of the bacteria of choice for for producing a wide range of molecules, from native products, such as amino acids or organic acids (Figure 1.2), to heterologous compounds, including biofuels, both at the proof-of-concept and commercial-scale (Chen et al., 2013). The physiology, biochemistry, genomics and regulation of this microorganism, as well as its suitability as laboratory workhorse and industrial production platform have been reviewed in great detail by Chen et al. (2013) and Pontrelli et al. (2018), and thus fall outside of the scope of this introduction.

1.2 The problem:

1.2.1 Economic and technical bottlenecks in metabolic engineering

Microbial production is selective, and proceeds at low temperature and pressure, presumably making any bio-based approach technologically simpler and more attractive in economic terms (Clomburg et al., 2017). The financial commitment, capital investment and risks associated with bio-processing are theoretically lower than large-scale, capital-intensive, infrastructure-heavy chemical projects (Carlson, 2016, OECD, 2018). It is believed, that with the right incentives, enough funding opportunities and international cooperation, biotechnological production will lead to faster innovation, a more circular economy, and prompt response to market drives and policy changes (Clomburg et al., 2017). However, it is fair to say that developing new commercial strains (i.e. strains that meet the economic and scale requirements for industrial production) is still difficult due to the time and financial investment still required to develop a viable working system, requiring an average project duration of 6–8 years and over $50 million investment (Nielsen

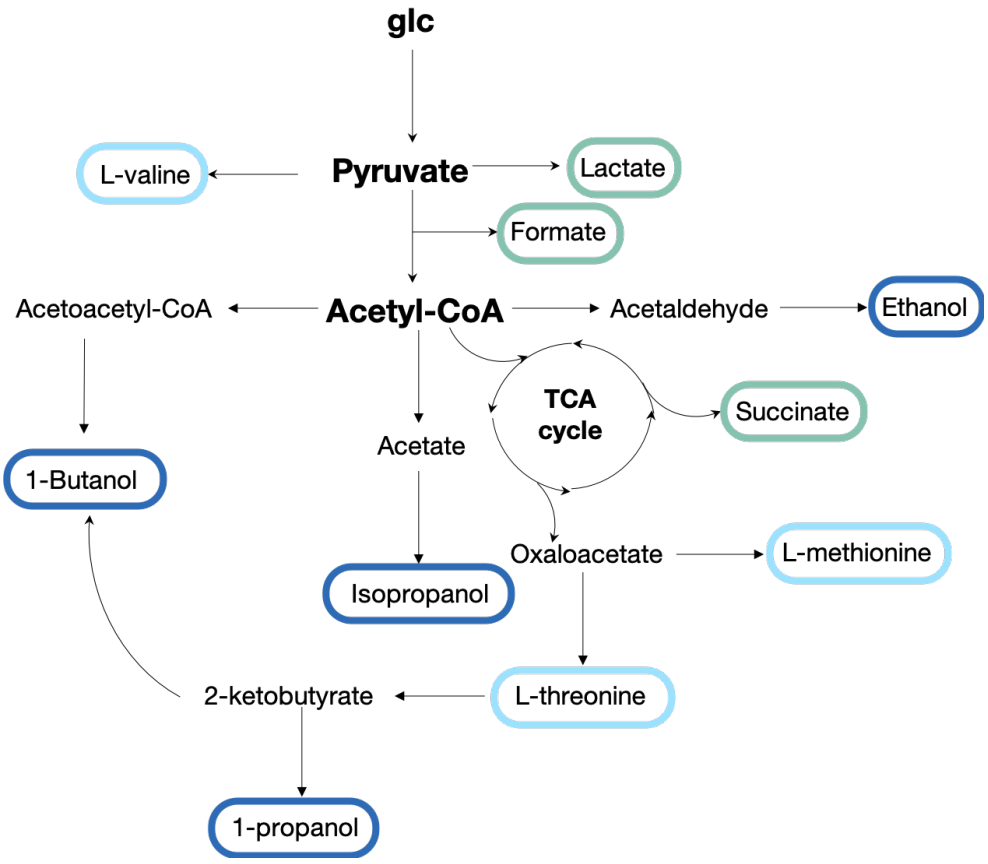


Figure 1.2: Schematic diagram of synthetic pathways for the production of a wide variety of chemicals in *Escherichia coli* (adapted from Chen et al. (2013)). *E.coli* strains have been engineered to produce a wide variety of chemicals in the past two decades (indicated by black arrows). In this thesis, I will focus primarily on butanol and ethanol as case studies for further improvements in biochemical production. TCA, tricarboxylic acid; CoA, coenzyme A.

and Keasling, 2016).

Limitations span several fold. A major challenge in commercialising new biotechnological processes is economic sustainability. For the industry to capitalise on this opportunity, not only should cell factories be attractive alternatives from the energetic and carbon efficiency perspective (Zomorrodi et al., 2012), but such a shift can only be successful if the designed processes are as financially viable as their petrochemical counterparts. It is crucial that biological conversion rates are higher than, or at least match, chemical conversion rates. To materialise this, microbial production shall reach near optimal yields (gram of product per gram of substrate) and productiv-

ities (production rate per volume) (Hädicke and Klamt, 2015). More efficient bio-based processes would lower production costs and make it more cost-competitive (OECD, 2018). The lower the market value of the competing, fossil-fuel derived equivalents, the greater this challenge becomes. To exacerbate this issue, given the recent price volatility of fossil fuels, additional pressures on the biotech industry to operate at near optimal values have been imposed. This, certainly, raises challenges in selecting the best approach to achieve the highest productivity levels.

Technical difficulties include scale-up, which is often hampered by lower microbial performance at larger scale, unlike the reliable economies of scale of the petrochemical industry. Identification of the best choice of target chemical, and also what microorganism to use for production is also not trivial. *E.coli* may also not always be a suitable host. This platform strain does not perish in extremes of pH or temperature, and it is also incapable of producing glycosylated products and/or complex proteins (Pontrelli et al., 2018). It is for these reasons that researchers are hunting for alternative microbes that can withstand more extreme manufacturing conditions (Eisenstein, 2016).

Arguably the most important issue is how internal cellular complexity also plays a key role. In reality, some microbial systems are more efficient than others, even when they have been designed to produce the same target chemical (Pasztor et al., 2015). This study focuses on the ability of the cells' native metabolism to accommodate the demands of synthetic pathways, as these pathways will alter the homeostasis of cellular energy and electron metabolism.

1.3 Defining a solution framework:

1.3.1 Metabolic cofactors: suitable targets for metabolic engineering

For any complex biological system, it is difficult to precisely pin-point the most important factors, and parameters thereof, influencing optimal catalytic performance. Traditionally, the focus has been on enhancing carbon flow towards the product by, for instance, targeting rate-limiting enzymes through either genetic modifications, enzyme engineering or synthetic biology. However, carbon metabolism alone cannot achieve efficient accumulation of target products. One reason for this is that the efficient production of a target chemical requires redox and energy as well as carbon (De Kok et al., 2012, Wang et al., 2017, **?**). Cofactors such as NAD^+ (nicotinamide adenine dinucleotide), $NADP^+$ (nicotinamide adenine dinucleotide phosphate) and ATP (adenosine triphosphate) provide the energy and electrons needed for carbon metabolism to proceed effectively.

ATP is the energy currency of the cell. Although it is not always classified as a cofactor (Fischer et al., 2010), ATP is required for DNA replication, biosynthesis, protein assembly, biochemical transport, energy supply and cell maintenance (Hara and Kondo, 2015). This makes ATP an indispensable building block in cellular metabolism. On the other hand, the fundamental role of the nicotinamide cofactors NAD^+ and $NADP^+$ is the smooth transfer of electrons between chemical reactions in the cell (Fischer et al., 2010). Conjointly referred to as NAD(P)(H) in this study, the NADH/NAD^+ and NADPH/$NADP^+$ cofactor pairs adjust the intracellular redox state, control energy metabolism, and regulate carbon flux and the cell life cycle more generally (Chen et al., 2014).

In any biological system, carbon assimilation inevitably results in the phosphorylation of ADP (adenosine diphosphate) and AMP (adenosine monophosphate) to form ATP, and the reduction of NAD^+ and $NADP^+$. These cofactors are subsequently hydrolysed and oxidised, respectively, when carbon source(s) is converted into biomass and by-products. Cofactor recycling is essential to allow central carbon metabolism to continue, i.e. to enable homeostasis (Varma et al., 1993). It follows that a balanced supply and consumption of both ATP and NAD(P)H, here termed cofactor balance, will have great influence in biotechnological performance. It is for these reasons that the ATP and NAD(P)(H) pools have become attractive targets for engineering, as I discuss in detail in Chapter 2. However, due to their functional relevance, both ATP and NAD(P)H are highly connected molecules, and consequently any small change in cofactor availability can result in substantial downstream ramifications in other areas of metabolism (Holm et al., 2010). Identifying the optimal cofactor profile for the production of a particular target can thus be very exhaustive and time-consuming work experimentally. So, how can we overcome this challenge while minimising experimental testing?

1.3.2 Modelling cofactor systems aids biosynthetic production

Fortunately, the physiology, biochemistry, genetics and metabolism of *E.coli* have been extensively studied. The advent of high-throughput technologies and their exponential decrease in price have facilitated the integration of biological information to capture the inner works of this organism (Bordbar et al., 2014). Genome-scale network reconstructions (or GENREs) are structured knowledge bases that contain the majority of all identified metabolic reactions in an organism of interest, including substrate and product information, stoichiometric coefficients, reaction reversibility and intracellular location, associated genes and gene products (Feist et al., 2008). To our benefit, these databases also provide reaction-specific cofactor information.

Building and curating a GENRE requires collating species-specific information from genome annotations (genomics), high-throughput experiments, literature searches and publicly available databases (O 'brien et al., 2015). GENREs can now be generated in an automated manner (Feist et al., 2008), although they require extensive validation against experimental data to improve their quality and accuracy. Once a reconstruction has been curated to a satisfactory level, it can be converted into a mathematical format (known as Genome-scale Model, or GSM), which can be queried by a myriad of modelling tools (Section 2.3.1) to predict potential phenotypes, analyse network properties and evaluate the effect of perturbations such as gene knockouts or the implementation of biosynthetic pathways (King et al., 2015). Most curated reconstructions are now available in standard formats such as SBML (systems biology markup language), which can be imported into most software applications. To our advantage, the *E.coli* metabolic network has reached the genome-scale, and the available genome-scale reconstructions can be used to computationally assess phenotypic properties.

The growing completeness of computational models has improved our understanding of the holistic approach to cell dynamics. Together with the availability of experimental studies we have been able to better depict the flow, roles and relevance of cofactors in cell physiology and allowed these insights to inform metabolic engineering. But our understanding is still incomplete. Using both computational and experimental frameworks, *E.coli* as a model system, and the attractive fuel molecules butanol and ethanol as proof-of-concept targets, the work underpinning this study focuses specifically on how network-wide cofactor usage profiles across biosynthetic pathway variants impacts bioproduction, and uses cofactor manipulations as a tool to improve bioproduction.

1.4 Research summary

As outlined in the Research Plan (Figure 1.3), this PhD study can be broadly grouped into three components, which directly align with the three core objectives of this project:

- to evaluate existing frameworks for the assessment of cofactor balancing, identify limitations and areas for improvement;
- to develop a novel, more versatile cofactor balancing algorithm that can help users better select optimal pathway variants; and
- to evaluate metabolic and physiological responses to selected cofactor perturbations, and harness these modifications to improve bioproduction.

Table 1.1: Outline of thesis structure

Chapter 2	**Background: cofactor manipulations for advancing bioproduction** *Refining the initial research questions, Q1: what cofactor manipulations and modelling approaches have been proposed for the assessment of cofactor balancing of metabolic engineering designs?*
Chapter 3	**Materials and Methods**
Chapter 4	**Energy and redox impact on butanol production** *Q2: how does varying cofactor demand affect the performance of butanol production pathways in E.coli?*
Chapter 5	**Development of a Cofactor Balance Assessment (CBA) protocol** *Q3: Is it possible to propose a new pathway selection method to track cofactor usage and classify synthetic pathway variants?* *Q4: How can the underdetermined nature of cofactor-related reactions be minimized?*
Chapter 6	**Energy and redox perturbations for enhanced bioproduction** *Q5: How robust is central carbon metabolism in E.coli to changes in ATP and NAD(P)H availability?* *Q6: Can bioethanol production be improved by reducing the ATP pool and targeting futile cycles?*
Chapter 7	**Conclusions and Future Work** Overall conclusions from Chapters 4, 5 and 6, including opportunities for future research

The thesis structure outline is summarised in Table 1.1. I begin in Chapter 2 with a detailed review of the existing cofactor manipulation strategies used in *Escherichia coli* that target the ATP and NAD(P)H pools specifically. Further, I present the state-of-the-art computer-aided and analytical tools that have facilitated the study of cofactor manipulation and cofactor balancing. Following a comprehensive outline of the materials and methods used in this study in Chapter 3, I set off my modelling analysis in Chapter 4 by exploring how varying cofactor demands limit butanol and butanol precursor production in *E.coli*. Using 8 different pathway variations with different demands for ATP and NAD(P)H, the relationship between cofactor usage and final yield is evaluated using two existing methods: First, I apply the analytical method developed by Dugar and Stephanopoulos (2011), which is pathway-specific and adjusts the maximal theoretical pathway yield based on ATP and NAD(P)H imbalances. Second, I use Flux Balance Analysis

(FBA) to estimate butanol production in *E.coli* at the network scale, and evaluate how changes in ATP and NAD(P)H demands across pathways routes trigger changes both in the internal fluxes and target production in this organism. The maximum theoretical yields obtained with these two methods are compared.

Chapter 5 presents the first attempt at developing a python-based Cofactor Balance Assessment (CBA) protocol that tracks cofactor production across central metabolism and estimates cofactor contributions to biomass and waste release, target production and metabolic maintenance. CBA can be used to evaluate how variations in ATP and redox demands contribute to yield efficiency when a synthetic pathway is introduced into a host. Using the butanol and butanol precursor pathways from Chapter 4, the CBA protocol helped explain why some pathways were predicted to achieve higher yields than others from a cofactor usage perspective. The study highlighted the underdetermined nature of some of the modelling systems (meaning there are several possible flux patterns that reach the same theoretical maximum), demonstrated by considerable dissipation of any excess ATP and NAD(P)H in high-flux futile cycles. In this chapter, a manual curation method to minimise this *in silico* flexibility was developed, as well as the use of experimentally-derived flux ranges as flux constraints. Adjusted yields from these two curation methods were compared to the estimates calculated in Chapter 4.

Chapter 6 then turns to assess cofactor futile cycling experimentally, as well as the connectivity between the ATP and NADP)H pools. First, I evaluated the physiological and metabolic responses of *E.coli* to cofactor perturbations induced by knocking out the ATP synthase and phosphoenolpyruvate kinase (PCK) reactions, which are directly associated with ATP production (and NAD^+ recycling) and ATP burning through futile cycling, respectively. Redirected internal fluxes led to industrially-attractive features such as increased glucose consumption and glycolysis flux, at the expense of lower ATP production and biomass formation. These phenotypes were then tested as potential bioproduction platforms for the accumulation of ATP-neutral targets, using ethanol as proof-of-concept. Finally, Chapter 7 concludes this thesis with a discussion of the impact of these results, contributions to the field, a critical analysis of this study's limitations and areas for improvement, as well as recommendations for further studies.

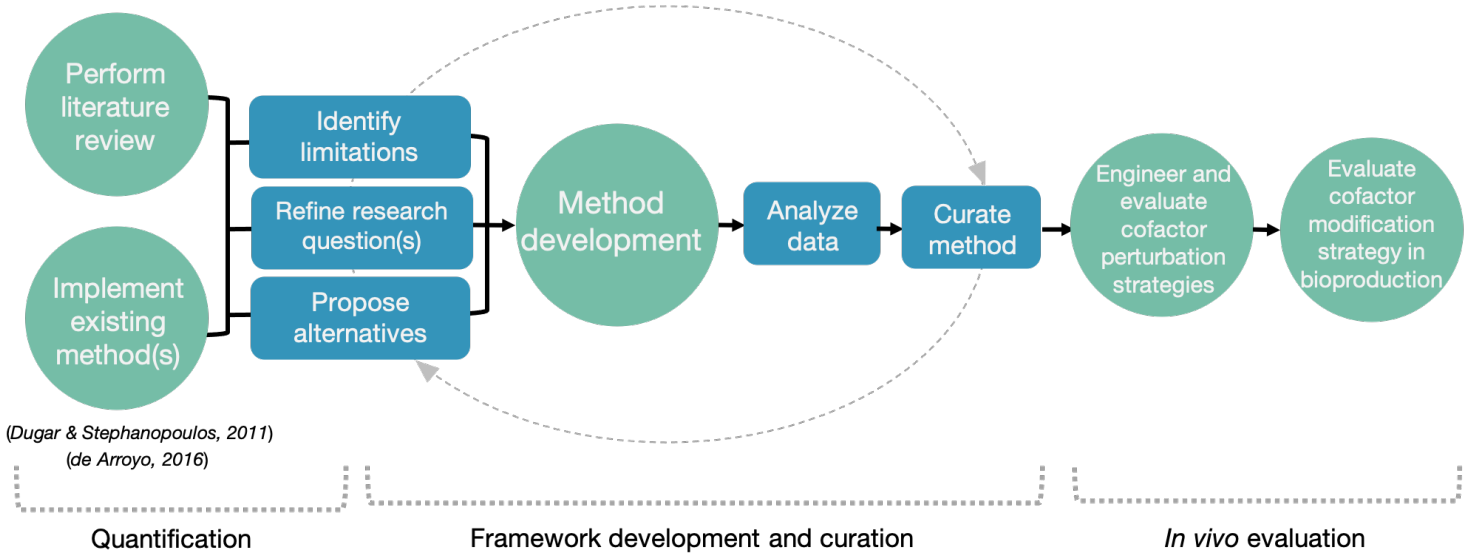


Figure 1.3: Summary of Research Plan.

"If I have seen further it is by standing on the shoulders of giants" - Isaac Newton, 1675.

2

Background: Cofactor Manipulation for Advancing Bioproduction

2.1 The relevance of cofactors in cellular homeostasis

It is widely understood that enzyme function often relies on the enzyme's association with one or more cofactors. Cofactors are catalytically-relevant organic molecules or ions that associate with enzymes to enable, regulate and/or fine-tune their activity (Broderick, 2001). With over half of all known proteins known to associate with at least one cofactor (Akhtar et al., 2014), it is undeniable that cofactors play a crucial role in catalysis (Fischer et al., 2010). As a result, manipulation of cofactor pools has long been identified as a valuable metabolic engineering strategy for enhanced bioproduction (Hädicke and Klamt, 2015).

Three of the most important, most connected cofactors are ATP, NADH and NADPH. ATP is the cell's energy-storage workhorse. Its dephosphorylation converts its high-energy form to low-energy molecules ADP or AMP, with chemical energy being fast released in the process (Equations 2.1 and 2.2). ATP is one of the most widely connected cofactors in any metabolic network (Vemuri and Aristidou, 2005), with more than 600 ATP-related reactions listed by KEGG since 2015 (Hara

and Kondo, 2015).

$$Pi + ADP \iff ATP \tag{2.1}$$

$$ATP + H_2O \rightarrow AMP + PPi \tag{2.2}$$

NAD^+ and $NADP^+$, which differ only on the presence of a phosphate group, are reversible electron carriers. The reduced forms, NADH and NADPH, respectively, carry two additional electrons. Collectively referred to as NAD(P)H in this study, the NADH/ NAD^+ and NADPH/$NADP^+$ cofactor pairs are involved in 740 and 887 biochemical reactions and interact with 433 and 462 enzymes, respectively (Chen et al., 2014). ATP and NAD(P)H play a crucial role in central metabolism and thus are popular targets for engineering. To achieve the desired metabolic changes, however, it is important to understand the core metabolic processes that are specifically controlled by these cofactors.

Substrate take-up by cellular metabolism results in the oxidation of sugars to generate metabolic building blocks, as well as the phosphorylation of ADP and AMP to form ATP and the reduction of electron carriers NAD^+ and $NADP^+$ to produce NADH and NADPH (Saini et al., 2016). For the sake of simplicity, I will hereinafter refer to these metabolic events as "synthesis" and "consumption" of ATP (also called "energy") and NAD(P)H (also called "redox"). ATP is primarily produced through glycolysis (via reactions PFK and PGK, Figure 2.1) and by oxidative phosphorylation (ATPS4r, Figure 2.1) powered by the electron transport chain, or ETC (reaction NADH11 in Figure 2.1), which is in turn fueled by NADH. During aerobic respiration of redox cofactors, free energy is conserved as a proton motive force across the cell's inner cytoplasmic membrane, used to power ATP production via the proton-translocating ATPase complex (hereinafter referred to as ATPase) (De Kok et al., 2012).

NADH is produced during catabolism, linking substrate utilization to biosynthesis and product formation. Alternatively, NADPH primarily drives anabolic reactions. To fulfill their distinct roles, these two redox carriers are generally not in thermodynamic equilibrium (Holm et al., 2010). NADPH can be produced by the isocitrate dehydrogenase reaction (ICD) in the Tricarboxylic Acid Cycle (TCA), but is primarily produced in the oxidative branch of the Pentose Phosphate Pathway (PPP) (Fuhrer and Sauer, 2009) by glucose-6-phosphate dehydrogenase (G6PDH2r in Figure 2.1, encoded by gene *zwf*) and 6-phosphogluconate dehydrogenase (reaction GND, encoded by *gnd*) to produce ribulose 5-phosphate (Siedler et al., 2011). Many organisms also have a transhydrogenase systems that can interconvert NADH and NADPH. *E.coli*, for example, heavily relies on its two

transhydrogenases (Fuhrer and Sauer, 2009). The soluble transhydrogenase encoded by the gene *sthA* catalyzes the reaction:

$$NAD^+ + NADPH \rightarrow NADP^+ + NADH \tag{2.3}$$

which oxidizes NADPH to produce extra NADH (labelled NADTRHD in Figure 2.1). The membrane-bound transhydrogenase encoded by *pntAB* couples reduction of $NADP^+$ with inward proton translocation, catalyzing the reaction:

$$NADP^+ + NADH + H^+_{out} \iff NADPH + NAD^+ + H^+_{in} \tag{2.4}$$

pntAB therefore contributes to an increased NADPH pool at the expense of NADH (reaction THD2 in Figure 2.1). Cofactor recycling is essential to enable homeostasis (Varma et al., 1993). NAD^+ is regenerated from NADH when intermediate metabolites in central catabolism are subsequently reduced. This process in *E.coli* usually leads to the production of various organic acids (such as lactate, succinate, acetate) and neutral compounds (like ethanol) (Ingram et al., 1987). By-product release enables *E.coli* to maintain the NAD^+/NADH and $NADP^+$/NADPH redox balance to ensure the continued operation of cellular metabolism.

Metabolic systems resulting in cofactor balance and robustness to environmental changes has evolved to facilitate survival of the species, rather than to act as a host for biocatalysis serving human objectives (Figure 2.2.A). Hence, it is not surprising that an organism engineered to include a synthetic metabolic pathway does not have optimal cofactor balance. Such an imbalance in the production and consumption of redox and/or energy by the engineered target pathway will result in the dissipation of cofactors by native metabolic processes such as cell maintenance and waste release, or promotion of growth over bioproduction (Figure 2.2.B). An increase in the waste product profile is therefore an indicator of imbalance in the metabolic network, compromising the overall efficiency of the biocatalytic conversion of carbon towards the target (Dugar and Stephanopoulos, 2011). In fact, even small changes in cofactor pools can have wide effects on metabolic networks and bio-production (Holm et al., 2010), and the inability of engineered systems to reach homeostasis can lead to partial or even full disruption of the cell's physiological state (Dugar and Stephanopoulos, 2011, Holm et al., 2010, Varma et al., 1993). In order to maximize our ability to engineer optimal bio-catalysts, it is therefore essential to carefully consider cofactor balancing and identify the constraints that pathway-specific cofactor imbalances may impose on the wider metabolic network.

Several articles have discussed this topic. For example, De Kok et al. (2012) suggested a positive

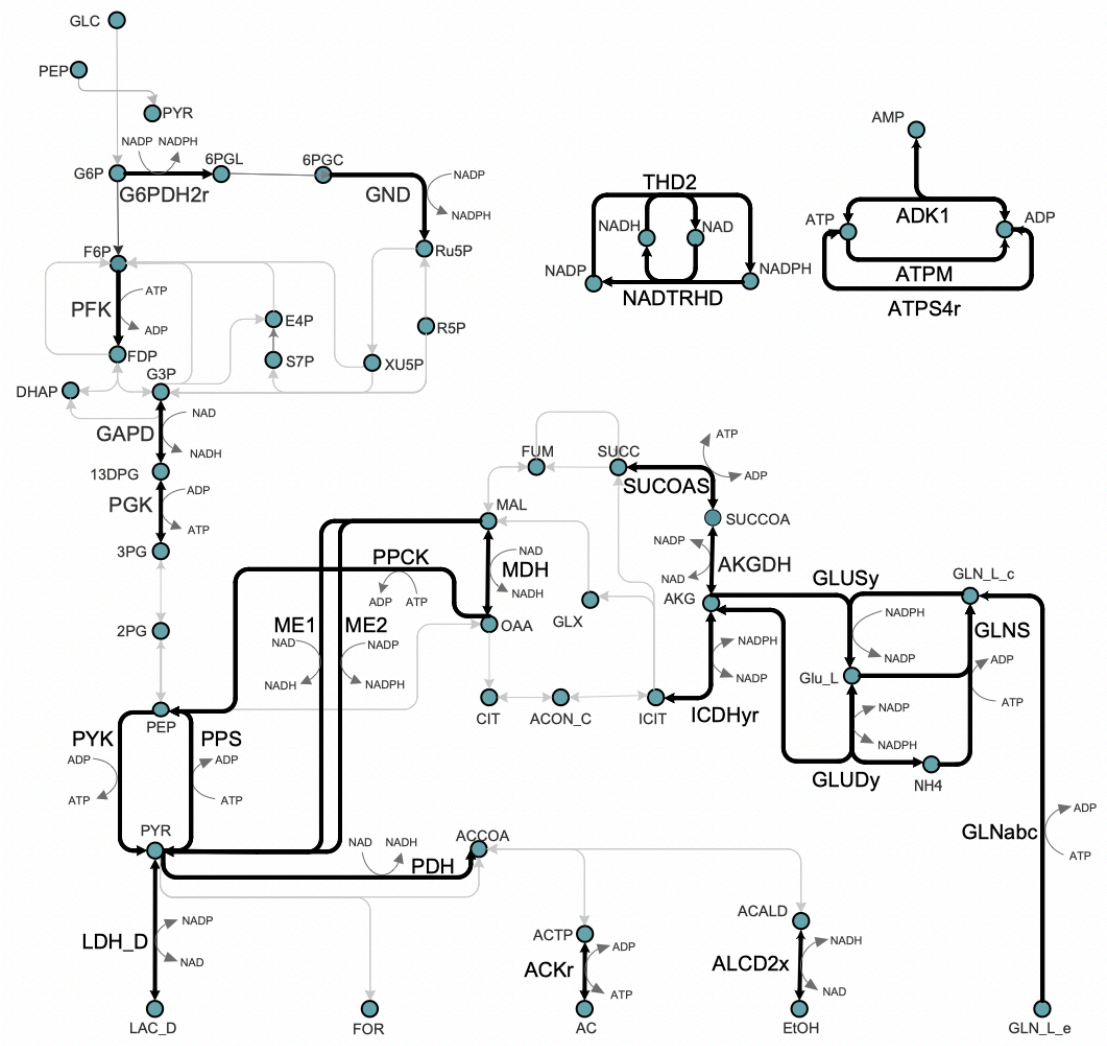


Figure 2.1: Central Carbon Metabolism. Key reactions that either consume or produce ATP or NAD(P)H are involved in relevant areas of central metabolism such as glycolysis, pentose phosphate pathway (PPP), tricarboxylic acid (TCA) cycle, oxidative phosphorylation, cofactor maintenance and waste production. These reactions have been highlighted in the diagram with thick black edges, and labelled according to standard GENRE labelling. Metabolites are shown as blue circles. No other cofactors are shown.

yield theory whereby optimal biosynthetic pathway flux is more likely when there is small, positive ATP excess remaining after the target pathway has utilised what it needs, enabling some but limited biomass production to keep the culture alive. To support this argument, they compared the low ethanol yield from a high-biomass producing *Saccharomyces cerevisiae* strain, against

the higher ethanol yield in a low-biomass producing *Zymomonas mobilis* strain. They argued that the cause of the difference was due to excess ATP production by the S.cerevisiae strain and that ATP surplus caused too much cell growth, burdening metabolism and decreasing product formation. Notably in this case, the authors referred to 'pathway' as the entire metabolic network of the cell (De Kok et al., 2012). Albeit with different terminology, Dugar and Stephanopoulos (2011) reached a similar conclusion based on a proposed theoretical framework (outlined in Section 2.3.3), concluding that the most effective equilibrium between substrate and product optimization was found in fully balanced (net zero) or ATP-requiring (negative ATP yield) pathways. In this article, 'pathway' referred to the leading route towards target production from a central carbon metabolite, not the entire metabolic network of the cell (Dugar and Stephanopoulos, 2011)

Both De Kok et al. (2012) and (Dugar and Stephanopoulos, 2011) illustrate the negative relationship between target product yield and bacterial growth, i.e. the more ATP is used for biomass production the less will be available for product synthesis (Hädicke and Klamt, 2015). Similarly, because of redox neutrality, pathways producing excess NAD(P)H are inefficient from a yield perspective as the surplus will need to rebalanced through either the activation of futile pathways or side-product(s) formation, which leads to diversion of additional substrate-generated energy away from the product (Dugar and Stephanopoulos, 2011). This basic principle is fundamental to any cofactor engineering project, because the choice of approach will depend on whether the target biosynthetic pathway is cofactor-neutral, cofactor-demanding or cofactor-generating. Targeting the ATP pool as an engineering strategy involves either increasing or decreasing the net amount of ATP that is available for biomass synthesis (Hara and Kondo, 2015). If the biosynthetic pathway is ATP-neutral or ATP-consuming, ATP conservation or ATP-generating strategies supplying the additional ATP required by the pathway shall be explored. Alternatively, if the pathway is ATP-producing, it would logically follow that ATP-wasting approaches would minimise the activation of futile cycles or generation of excess biomass. Similarly, if the biosynthetic pathway requires the supply of additional NAD(P)H (this is commonly the case for reduced products), the cell's network will need to generate sufficient NAD(P)H to supply the pathway, after it has utilised the redox and energy it needs to guarantee its own survival (Figure 2.2.C). Alternatively, if the pathway is redox generating, an redox-consuming sink that consumes surplus energy would be an effective strategy.

Redox and energy manipulation strategies have been extensively reviewed previously (Chen et al., 2014, De Kok et al., 2012, Hädicke and Klamt, 2015, Wang et al., 2013, Zhao and Van Der Donk, 2003). The aim of this background chapter is therefore not to provide an in-depth analysis of the strategies available. Instead, the objective is primarily to introduce the reader to the options

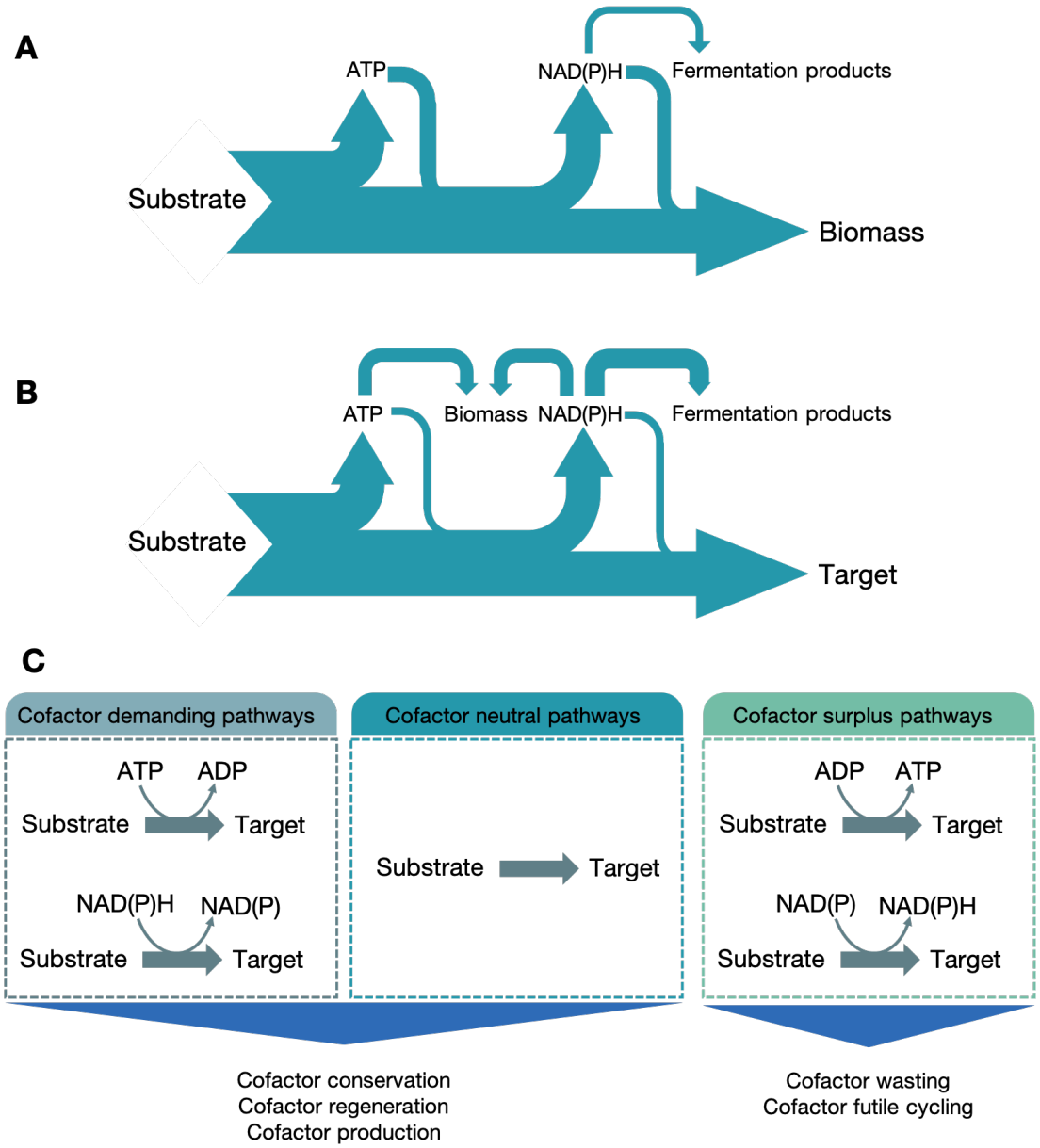


Figure 2.2: Understanding cofactor metabolism, adapted from Dugar and Stephanopoulos (2011). (A) Theoretical energy and redox flow in a non-engineered, biomass producing system. Substrate oxidation leads to the production of surplus ATP and redox, which is consumed by the system to guarantee cellular maintenance and biomass formation. (B) Theoretical energy and redox flow in an engineered, target-producing system. In this case, some ATP and redox released from substrate oxidation is used for target production, after the system has used what it needs for cellular maintenance and biomass formation. Imbalances are dealt with by the activation of futile cycles and release of side products. (C) Cofactor-neutral or cofactor-demanding pathways can use up any surplus cofactor generated from substrate oxidation, and thus are generally assumed to lead lower side product release and higher yields. Cofactor conservation or cofactor regeneration strategies are best when dealing with this sort of pathways. Pathways producing excess ATP or NAD(P)H are termed 'surplus' pathways and are deemed inefficient from a yield perspective because of the need to balance the excess cofactor, often met via futile pathways, side-product formation or surplus biomass formation.

available and challenges of modifying ATP and NAD(P)H metabolism within the context of *Escherichia coli*. This is followed by an overview of the existing computational tools built to this date to facilitate the study of cofactor metabolism and implementation of cofactor modification strategies for enhanced bioproduction. I will close this chapter by presenting the gaps and limitations identified during the review of the literature, which inspired the key areas of focus for this PhD study.

2.2 Experimental manipulation of cofactor pools

Cofactor engineering strategies can generally be subdivided into three key areas of focus: (i) genetic modifications for increased/decreased availability of desired cofactors (either by regulating endogenous cofactor systems or introducing heterologous ones), (ii) changes in enzyme-cofactor preferences (by engineering enzyme specificity), and (iii) introduction of novel, synthetic cofactor systems (Wang et al., 2017, 2013). This PhD study concentrates on the first of these areas, and therefore (ii) and (iii) are outside the scope of this review.

2.2.1 Manipulation of ATP pools

ATP supply is one of the most critical factors for bioproduction (Hara and Kondo, 2015). The negative relationship between growth and product yield outlined in Section 2.1 makes the ATP pool a highly relevant target for the design of metabolic engineering strategies for increased bioproduction (Hädicke and Klamt, 2015). Targeting the ATP pool involves either increasing or decreasing the net amount of ATP that is available for biomass synthesis (Hara and Kondo, 2015). As described previously, the choice of approach will depend on whether the target biosynthetic pathway is ATP-neutral, ATP-demanding or ATP-generating, as this will inform whether ATP conservation, ATP wasting or ATP futile cycling strategies apply.

Hädicke and Klamt (2015) suggest that for product synthesis pathways that are ATP-neutral or ATP-demanding, ATP conserving manipulations are particularly appropriate. ATP conservation guarantees that other ATP-forming reactions in the network meet the demands for sufficient growth and maintenance alongside product formation. Popular ATP conservation strategies include modifications of substrate uptake and product export (De Kok et al., 2012), manipulating exchanges between anaplerotic reactions around phosphoenolpyruvate (PEP), pyruvate, malate, or oxaloacetate (OAA) pools (Kim et al., 2012), and modulating the intracellular ATP/ADP ratio (Toya et al., 2012)

The carboxylation of phosphoenolpyruvate (PEP) to oxaloacetate (OAA) by phosphoenolpyruvate carboxylase (PPC) is one such anaplerotic reaction (Equation 2.5). Under glycolytic conditions in *E.coli*, PPC is the primary pathway for the production of OAA from PEP. This reaction is deemed irreversible because of the energy loss suffered from the release of inorganic phosphate (Zhang et al., 2009). Under gluconeogenic conditions, the reverse reaction is catalysed by the ATP-consuming phosphoenolpyruvate kinase enzyme (PCK) instead (Equation 2.6) (Chao et al., 1993). PCK is normally repressed by high glucose concentration, and is reported to theoretically function only under gluconeogenesis conditions in *E.coli*. However, succinate-producing bacteria present an analogous ATP-producing PEP carboxykinase that functions as the primary PEP carboxylating enzyme, with the consequent conservation of energy as ATP (Zhang et al., 2009).

$$Glycolysis \rightarrow\rightarrow\rightarrow PEP + CO_2 + H_2O \rightarrow OAA + Pi + H^+ \tag{2.5}$$

$$OAA + ATP \iff PEP + ADP + CO_2 \rightarrow\rightarrow\rightarrow gluconeogenesis \tag{2.6}$$

ATP-conserving PCK has been used as a target for higher intracellular energy content, because it can produce ATP instead of a phosphate group, unlike PPC. Its industry relevance was tested by Singh et al. (2011), who showed that heterologous overexpression of the ATP-generating PCK from *Actinobacillus succinogenes* in *E.coli* led to enhanced succinate production and higher cell growth in the presence of sufficient CO_2. Kwon et al. (2014) showed that PCK overexpression led to twice the intracellular ATP concentration. This was followed by Kim et al. (2012)'s demonstration of the biotechnological benefit of this approach by showing that a strain overexpressing PCK produced 85% more GFP (Green Fluorescent Protein) and 37% less biomass than the control strain. Albeit with no introduction of heterologous genes or elimination of competing pathways, Zhang et al. (2009) achieved a 5-fold increase in succinate yield in minimal media by increasing the expression of PCK and truncating the PEP-dependent phosphotransferase system. The PEP-dependent phosphotransferase system is the primary mechanism for glucose uptake in *E. coli*, and represses glucose uptake when there is high intracellular PEP. Zhang et al. (2009) reported that these changes, together, increased the intracellular PEP as well as net ATP production through PCK, thereby increasing PEP carboxylation and succinate production.

ATP conservation may also be achieved by indirect modulation of the ATP/ADP ratio through the truncation of global regulatory factors, such as the regulatory factor for anoxic response control (ArcA), which is known to be involved in NADH and ATP-dependent reactions and thus can change the cofactor turnover. To illustrate this, Toya et al. (2012) reported a 4.4-fold increase

in TCA cycle activity as well as a 2-fold increase in the ATP/ADP ratio after deleting ArcA in *E.coli*.

Alternatively, pathways that are ATP-generating can benefit from interventions that trigger a decrease in ATP availability (Hädicke and Klamt, 2015). For example, expressing the soluble F(1) part of the membrane-bound F(1)F(0) H^+-ATPase (hereinafter known as ATPase, or ATPs), which constantly consumes ATP without pumping protons, has been shown to increase substrate uptake, glycolytic flux, and acetate overflow, while decreasing biomass production (Holm et al., 2010, Koebmann et al., 2002). Similarly, mutant *E. coli* strains lacking respiratory chain enzymes exhibited up to 30% lower growth rates, accelerated generation of glycolytic ATP and enhanced production of pyruvate and acetate (Kihira et al., 2012).

Another well-known strategy for lowering the net ATP is triggering elevated ATP wasting (i.e. burning surplus ATP in the system) through ATP futile cycling. ATP burning through futile cycles enhances ATP turnover and increases biosynthetic production (Erdrich et al., 2014). As early as 1992, Patnaik et al. (1992) discovered a futile cycle in *E.coli* between PEP and pyruvate (Equations 2.7 and 2.8) which, if overexpressed, led to increased oxygen and glucose consumption, as well as higher excretion of pyruvate and acetate in minimal media. Similar findings were reported by Chao and Liao (1994*b*) two years later. Shortly after, Chao and colleagues reported another cyclic futile pathway between PEP and OAA by overexpressing PCK and PPC (Equations 2.5 and 2.6 above) in *E.coli*. Similar observations regarding increased oxygen and glucose uptake, reduced growth yields and pyruvate and acetate excretion were reported.

$$Pyruvate + ATP + H_2O \rightarrow PEP + AMP + 2H^+ \tag{2.7}$$

$$PEP + ADP + H^+ \rightarrow Pyruvate + ATP \tag{2.8}$$

Over two decades later, Hädicke et al. (2015) implemented an IPTG-inducible system encoding PEP synthase (PPS) and pyruvate kinase (PYK), giving rise to an ATP consuming cycle that exhibited 25% higher specific lactate productivity, 8% higher yield, and 14% higher substrate uptake rate.

Additional strategies include the addition of energy generating substrates such as ATP and citric acid into the media, and controlling the environmental pH (lower pH generates a proton-motive force between the inner and outer surfaces of prokaryotes and drives ATP synthase activity) (Hara and Kondo, 2015), but these fall outside the scope of this review.

2.2.2 Manipulation of NAD(P)H pools

Interventions targeting the NAD(P)H pool are mainly characterised by either one of the following strategies. First, knocking out competitive pathways involving reduced cofactor consumption (Wang et al., 2017) so that product excretion is the only possibility to regenerate the oxidized redox cofactors NAD(P) (Hädicke and Klamt, 2015). This approach is especially appropriate for relatively reduced products like organic acids, alcohols or terpenoids. Successful examples included already mentioned studies such as Singh et al. (2011), where they knocked out *ldhA* (lactate release) and *pflB* (formate release) to increase succinate production, or Zhou et al. (2008), whereby ethanol production was enhanced after knocking out all other NADH-dependent organic acid fermentation pathways. Similarly, a modified clostridial 1-butanol pathway was introduced into *E. coli*, where competing pathways had been removed to accumulate sufficient NADH and acetyl-CoA to help direct the flux towards the target. *E. coli*'s native AtoB reaction was used for the formation of acetoacetyl-CoA, which is one of the initial precursors required for butanol formation.

Alternatively, there are various other means to increase the net supply of reducing power (Hädicke and Klamt, 2015). This can be achieved either by overexpressing existing cofactor regeneration routes, or by introducing heterologous cofactor regeneration systems. Often, NADPH will be the reduced cofactor needed for catalysis (Celton et al., 2012), but it is well-known that the NADPH turnover in *E.coli* is lower than NADH (Fuhrer and Sauer, 2009). Great efforts have therefore been directed towards modulating the NADH/NADPH ratio and increasing the availability of the latter. Sources of NADPH include the PPP, ICD and NAD(P)-specific malic enzyme, and the following studies report that it is possible to increase NADPH availability either by increasing carbon flux into PPP, or decreasing the carbon flux through glycolysis. For example, direct overexpression of enzymes involved in PPP, such as glucose-6-phosphate dehydrogenase (*zwf*, or G6PDH2r in Figure 2.1) and/or gluconate-P dehydrogenase (*gnd*, or GND in Figure 2.1) have been reported to increase the carbon flux through PPP, and thus augment the overall NADP(H) availability (Lakshmanan et al., 2015). Another option to increase carbon flux through the PPP is the the deletion of phosphoglucose isomerase (*pgi*), which theoretically leads to the full disruption of glucose catabolism via the glycolytic pathway and results in 100% glucose being metabolized via the PPP, yielding 2mol NADPH per mol glucose.

An alternative approach to increase the overall NAD(P)H turnover involves the direct phosphorylation of NADH to form NADPH by overexpressing NADH kinase. Lee et al. (2013) expressed the NADH kinase from *S.cerevisiae* and enhanced the availability of NADPH by 51% and 96%

compared to the wild type strain (Lee et al., 2013). Another commonly-used strategy is to simply increase the activity of transhydrogenases that can convert excess NADH into NADPH (Wang et al., 2013). As described in detail in Section 2.1, two transhydrogenase systems exist in *E. coli*: (i) the membrane-associated PntAB and the soluble the Sth systems (Fuhrer and Sauer, 2009). Wang et al. (2013) is just one example of how this strategy can be combined alongside other host strain modifications to improve NADPH-dependent pathway performance where NADPH availability is measured in the presence of a NADPH utilizing reaction or 'sink'. Increased NADPH availability has also been used as a strategy to increase 1,3-propanediol production in Siedler et al. (2011). Moreover, there are times when it is useful to remove any excess NADH in its reduced form, i.e. to reduce the pool of NADH in the cell. NADH oxidase removes excess NADH and helps regenerate the oxidized product (Holm et al., 2010)

Although it will not be the focus of this study, modifying cofactor preferences has also been an important consideration in the field, and is worth mentioning briefly. Protein engineering of enzyme-cofactor specificity prompting the conversion of enzyme specificity into the desired characteristics is a promising technique, and the growing number of protein sequences and structures enable the precise identification of binding regions as potential targets for engineering. By creating an enzyme that reacts with $NADP^+$ instead of NAD^+, or vice versa, Wang et al. (2013) showed that replacement of the *E. coli* native NAD^+ dependent GAPDH *gapA* by $NADP^+$ dependent GAPDH *gapB* from *B. subtilis* increased product formation and the strain showed enhanced NADPH-dependent biosynthesis when NAD kinase was co-expressed (Martínez et al., 2008). Formation of artificially created active complexes are also commonly used strategies. Other alternatives include modification of ferrodoxin genes, cytochrome P450 reactions, or by altering the environmental conditions of the cultures (Chen et al., 2014, Wang et al., 2013), which have been successful strategies but also fall outside the scope of this review.

2.2.3 Limitations of cofactor manipulation strategies

A review of the experimental literature gave rise to the following observations. Firstly, it was noted that most studies focused on either energy or redox optimisation. Aside from Holm et al. (2010), no or very little research has so far focused on the shared regulatory roles of ATP and NAD(P)H. Secondly, the literature demonstrates that most (if not all) studies have taken a late-stage and somewhat narrow approach to cofactor manipulation, i.e. increasing flux through already selected pathways by improving cofactor availability or cofactor recycling in those systems for which maximum yield is the ultimate objective. Thirdly, target genes and reactions for

deletion/manipulation have often been selected based upon manual inspection of the pathway itself (Zhou et al., 2008). This simple, manual approach to cofactor manipulation is not always successful or straightforward. Achieving cofactor balance in more complex systems may not be a trivial task, especially because direct knockouts, overexpression or replacement of target reactions may lead to new metabolic burdens needing further manipulation (Liu et al., 2015). This is exacerbated by the fact that, in reality, some bio-catalysts are more efficient than others, even when they have been designed to produce the same target chemical (Dugar and Stephanopoulos, 2011, Pasztor et al., 2015). Given that pathway variants that produce the same target will not be energetically equivalent from a cofactor usage perspective, it becomes crucial to accurately predict the designs likely to be superior and identify the most optimal strategies for production at earlier stages of the design process. This would speed up pathway and host selection, minimise experimental testing and optimise the use of available resources.

2.3 Cofactor balancing analytical approaches

Computational studies have attempted to estimate cofactor balancing. Most of the work to date has been performed using genome-scale models (or GSMs, briefly introduced in Section 1.3.2). Using appropriate modelling tools, GSMs enable the systematic assessment of different types of growth media, pathway routes, end products, biological and environmental conditions, or even the use of different strains and host organisms. This type of approach enables the community not only to select between various pathways according to their potential, but also evaluate how these pathways integrate with selected hosts, before moving on to the experimental stages of bioproduction. GSMs have the advantage that they contain vast reaction information, providing the opportunity to track reactions that consume and produce energy and reducing equivalents.

2.3.1 Computer-aided tools for metabolic engineering

GSMs can be queried using a number of techniques for network analysis. Elementary mode analysis is one such approach, which identifies network structures by decomposing a highly interconnected, complex metabolic network into its discrete, organized components (Trinh et al., 2009). Information about the elementary pathways that make up the larger network can be used to characterize network properties, such as cellular phenotypes, network regulation and robustness, facilitating an understanding of cell physiology and implementation of metabolic engineering strategies. Some computational tools that perform elementary mode analysis are MetaTool and FluxAnalyzer, reviewed by Fernández-Castané et al. (2014).

Alternatively, Constraint-based Reconstruction and Analysis methods (COBRA) (Schellenberger, J. et al., 2011) is another widely-used approach that estimates the flux distributions metabolic networks. COBRA takes a stoichiometric model, such as a GSM, and puts it into mathematical form by generating a stoichiometric matrix (S), with rows indicating the available metabolites, columns indicating the reactions, and their associated stoichiometries acting as limits (or constraints) in the potential flux patterns that the network can take. Because most biological networks are underdetermined, i.e. the number of reactions greatly exceeds the number of metabolites (Gianchandani et al., 2010), there are insufficient linear-independent balances to solve the intracellular fluxes (Bonarius et al., 1997), so further constraints must be included to solve the linear problem. These are often either taken from the literature or experimentally-derived parameters, such as substrate uptake rates or secretion rates. Additional constraints may also be imposed by adding "upper" and "lower" bounds to any reaction, which represent the maximum and minimum allowable fluxes for such reaction. These constraints further shrink the possible range of phenotypes (Price et al., 2004), also referred to as solution space. These limits also ensure that the total amount of any compound being produced is equal to the total amount being consumed. Finally, a list of all flux values for each reaction in the network can be calculated. The so called "flux vector" (also known as flux distribution) represents a unique ''network state'' related to the physiological state of the metabolic network under the selected conditions. The computed candidate representation can then be directly compared to the experimental version of the target organism under the conditions considered (Price et al., 2004).

COBRA is widely used to predict phenotypic behaviours and study model properties using user-selected parameters. One of the most commonly-used applications of COBRA is Flux Balance Analysis (FBA) (Figure 2.3), which will receive particular attention and be regularly implemented throughout this PhD study. FBA is a mathematical framework that predicts optimal flux distributions in an organism of choice (Kauffman et al., 2003, Orth et al., 2010). FBA overcomes the issue of underdetermination by optimising for a particular "objective function" while ensuring mass balance. This objective function is generally assumed to be biomass production, given that cells thrive to optimise their own growth. An FBA solution will thus consist of both the maximal theoretical value achievable by the objective (e.g. maximal growth rate if biomass is set as the optimisation criterium), as well as the flux distribution of the resulting metabolic network solution.

One of the most widely used COBRA approaches is FBA. FBA is a simple yet powerful tool, which overcomes the underdetermination problem by specifying an objective function as an additional constraint. Examples of objective functions include target production maximization, optimization

of biomass formation, ATP production or minimization of O_2 consumption (Fernández-Castané et al., 2014). By harnessing only the stoichiometric information, FBA is able to provide a fairly accurate measure of optimal theoretical yields, insights into whether a specific modification to a genome-scale model is theoretically feasible or not, and a quick test of the consequence of a potential engineering strategy. As already discussed, biosynthesis of a product requires redox and energy as well as carbon. Redox and energy are generated during carbon conversions between substrate and products. Therefore, under the FBA framework, the maximum theoretical yield will be limited not only by the stoichiometry of the reactions in the metabolic network, but also by the consumption and regeneration of metabolic cofactors, because an inadequate production of redox and energy would in turn require the oxidation of additional substrate in order to provide the required redox and energy (Varma et al., 1993), and this would be reflected in the final solution. FBA solutions thus capture how biosynthetic pathways integrate with selected hosts, and how their efficiency is limited by cofactor dynamics in the wider network.

A caveat of FBA, however, is that it can still often lead to many, non-unique, optimal solutions that satisfy the imposed constraints and solve the mathematical objective. This is particularly relevant when the selected genome-scale model contains thousands of reactions, because optimisation will often lead to more than one flux distribution that satisfies the objective. Flux Variability Analysis (FVA) (Mahadevan and Schilling, 2003), also regularly used in this PhD study, was developed to address this issue. Instead of unique flux values, FVA calculates the minimum and maximum allowable flux ranges for each reaction in a model, whilst satisfying mass balance conditions and all other constraints imposed on the system. In other words, FVA discerns the complete flux range, capturing all possible phenotypes the network can generate. Several other extensions to FBA have been developed over the years that have allowed COBRA to be more broadly applicable to specific problems in metabolic engineering and systems biology. One of these extensions is MOMA (Minimisation of Metabolic Adjustment) (Segre et al., 2002), which was also of interest in this study. MOMA is particularly useful when modelling reaction knockouts, because it assumes that there is a transition period during which the cell is not necessarily optimising for the predetermined objective function, but it is instead undergoing a period of "cellular adjustment" in which it is trying to restore its original state. This is achieved by calculating the minimal distance between the network of the modified strain and the reference strain. Extensive reviews of the extensions to FBA and their application in metabolic engineering have now been published (Copeland et al., 2012, Gianchandani et al., 2010, Simeonidis and Price, 2015).

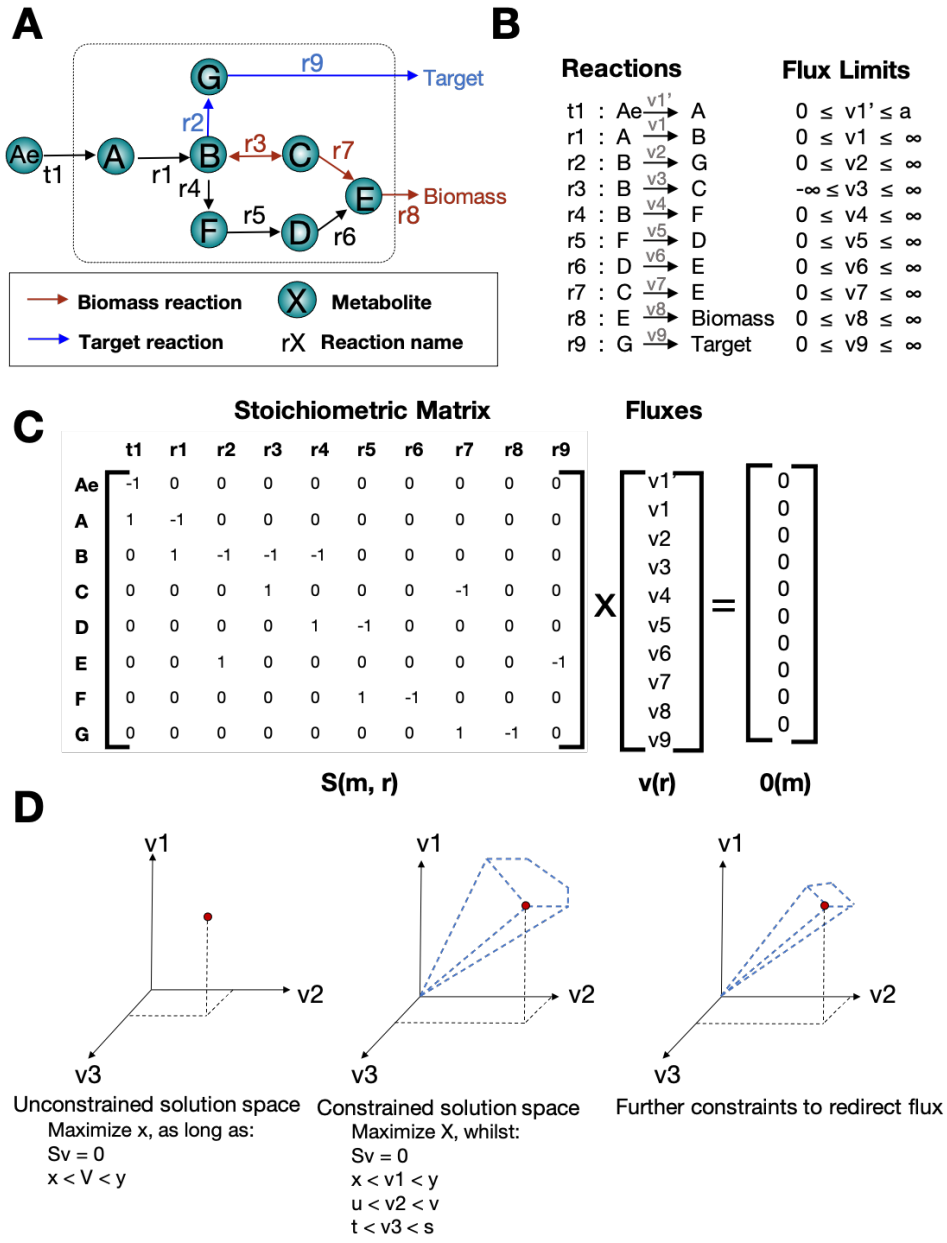


Figure 2.3: Introducing Flux Balance Analysis (FBA) (adapted from (Orth et al., 2010) and (Maia et al., 2016)). FBA is a mathematical framework that simulates fluxes through metabolic reactions in a biological system of choice and predicts the optimal growth rate (or yield). (A) FBA requires a metabolic network reconstruction, often a genome-scale model, containing stoichiometrically balanced metabolic reactions. (B) A Stoichiometric Matrix (S(m, r)) is derived, where rows represent metabolites and columns represent reactions. S is sparse, filled with positive or negative coefficients depending on the stoichiometry of each metabolite in each reaction. Assuming equilibrium, the flux distribution is $Sv = 0$, where v is the flux value through each reaction. (C) An objective function $Z = Cv$ must then be specified, where the contribution of each reaction to the total flux is accounted for via C, a vector of weights. Solutions can be constrained by manipulating flux values and flux ranges (lower and upper bounds). Constraints may be verified rates, intracellular fluxes or cellular growth rates, inferred from interpretation of gene expression data. A linear programming solver then computes the optimal flux distribution out of the available solution space defined by the constraint space, ensuring that steady state is satisfied.

2.3.2 COBRA methods for the study of cofactors

Given the extensive availability of cofactor-related information in stoichiometric models, a growing number of COBRA-based modelling approaches have been increasingly employed to guide metabolic engineering strategies to improve cofactor availability. Chin et al. (2009) represents one of the first early attempts at using constraint-based modelling to investigate how xylitol theoretical yields were affected by central metabolism knockouts directly affecting NAD(P)H supply in *E.coli*. While deletion of the membrane-bound transhydrogenase *pntAB*, TCA activity or glucose-6-phosphate dehydrogenase activity reduced xylitol yield by up to 60%, deleting phosphoglucose isomerase (PGI) increased xylitol yield around 59%. Although they identified possible metabolic sources of imbalance and their contributions towards overall NADP(H) supply, they did not report any specific strategies to improve cofactor balance in these specific systems experimentally. Later, Ahn et al. (2011) used a genome-scale model of *E.coli* to understand how a mutant strain lacking PGI activity grew better with fructose as a carbon source during enhanced shikimate production, and found that it was due to the tight coordination between PPP, TCA and transhydrogenase activity. Furthermore, Ghosh et al. (2011) used constraint-based modelling to study, for the first time, the effects of cofactor balancing for the specific case of yeast producing ethanol from L-arabinose and D-xylose. Further, Ghosh et al. (2011) explored the effect of altering the cofactor specificity of various enzymes among the non-native xylose catabolic pathways of *Saccharomyces cerevisiae* upon its cellular growth and ethanol production.

These initial applications of COBRA to analyse redox balancing, although successful and informative, were, again, particularly targeted. More comprehensive, widely-applicable *in silico* methods began to be proposed to improve the regeneration of desired cofactors at the system-wide scale, with a particular focus on cofactor swapping, reaction knockouts and overexpression strategies. Flux Sum Analysis (Kai et al., 2009) became a widely-used tool to calculate total cofactor turnovers. Cipher for Evolutionary Design was the first algorithm to propose single, double and triple gene deletion targets which can improve intracellular NADP(H) availability (Chemler et al., 2010). Two additional *in silico* methods, cofactor modification analysis (CMA) (Kai-Sheng Chung et al., 2013) and OptSwap (King and Feist, 2013), were also proposed to identify relevant cofactor specificity engineering targets which can augment the yield of several native and non-native products in *E. coli* and *S. cerevisiae* by improving the overall redox balance. Kai-Sheng Chung et al. (2013)'s CMA optimises for modifications of oxidoreductase specificity to improve the yield of terpenoids in yeast. On the other hand, King and Feist (2013)'s OptSwap is a bilevel optimisation method to identify modifications of oxidoreductase specificity using knockout strategies.

OptSwap's developers followed up with a more comprehensive, system-wide global analysis of cofactor swapping. They presented a COBRA approach that identifies optimal cofactor-specificity swaps that can increase the theoretical yields for a large number of molecules. They concentrated on two of the most important production hosts, *E. coli* and *S. cerevisiae*, and were able to identify the minimal number of cofactor swaps needed to maximise theoretical yields (King and Feist, 2014).

At this point in time, cofactor engineering targets were usually predicted by either simulating the redox cofactor swaps (Kai-Sheng Chung et al., 2013, King and Feist, 2013) or specific gene knockouts (Chemler et al., 2010). In Lakshmanan et al. (2015), several techniques were for the first time integrated, whereby flux-sum analysis was first used to quantify the NADPH turnover rates of four commonly used industrial workhorses , *E. coli*, *S. cerevisiae*, *B. subtilis*, and *Pichia pastoris*. They found that, unlike all other organisms, *E.coli* showed remarkably high ability to produce NADPH, thanks to its strong transhydrogenase systems. Using CMA (Kai-Sheng Chung et al., 2013), Lakshmanan et al. (2015) also identified the optimal enzyme targets that could improve the overall NADP(H) turnover, which increased by 20-fold with cofactor engineering. They were able to show that cofactor engineering of GAPD almost reverses the NAD(H)/NADP(H) ratios. GAPD is a central enzyme in glycolysis, so most of the carbon needs to be processed by this reaction in order to reach any downstream nodes relevant for bioproduction. Lakshmanan et al. (2015) suggest that generating NADPH-specific GAPD-modified strains can be a suitable strategy for NADPH-dependent bioproduction.This approach was followed by structural analysis to identify the cofactor binding sites in these target enzymes and propose mutation strategies to switch their cofactor specificity from NADH to NADPH. In fact, as was also briefly mentioned in Section 2.2.2, engineered strains bearing the NADPH-specific GAPD showed a substantial improvement in NADPH-dependent product yields (Martínez et al., 2008).

Despite these success stories, state-of-the-art techniques focused on optimisation, and only King and Feist (2014) seemed to have provided a comprehensive tool for the evaluation of the impact of changing cofactor specificity on a system-wide scale. No methods for the *a priori* assessment of system-wide cofactor balance exist. Flux sum analysis provides insights into the optimal cofactor turnovers, but does not inform the user of what these cofactors are doing, where they originate, or where they are going. The sector still lacked an easy-to-implement approach that could capture a system's cofactor usage profile at the network scale, in order to gain insights into which type of follow-up modifications may be more suitable.

Some attempts at capturing cofactor usage profiles discerning energy balancing have been attempted, particularly in the fluxomics sector. Fluxomics determines the intracellular metabolic

fluxes that most reasonably reproduce phenotypic behaviour. Flux estimates are often determined using ^{13}C Metabolic Flux Analysis (^{13}C-MFA, or simply MFA), which combines a feeding experiment using isotopically-labelled carbon source(s), measurements of physiological parameters gathered from such feeding experiment (growth rate, sugar uptake and product excretion rates) and fitting techniques to estimate the internal metabolic flux distribution that best describe the observed phenotype(s) (Crown and Antoniewicz, 2013, Gianchandani et al., 2010). These flux estimates and can be directly compared to predicted phenotypes gathered from COBRA models (Chen et al., 2011). Chen et al. (2011) shows some early attempts to verify energy balancing in *E.coli* using synergies between FBA and ^{13}C-MFA (Chen et al., 2011). Similarly, Garcia Martin et al. ((2015) also provided cofactor balance information by constraining genome-scale models using ^{13}C labelled data (Garcia Martin et al., (2015). He et al. (2016) found all ATP-generating and ATP-consuming reactions and their fluxes from MFA data and determined how much energy was consumed across biomass synthesis/amino acid production as opposed to energy used for cell maintenance. These model-driven studies provided valuable insights regarding intracellular cofactor usage. However, MFA is a fitting technique used to reproduce experimental observations. Still, a predictive *in silico* analysis of how metabolic pathway implementation can influence a system's redox usage was not available. Accessibility to such information would help experimentalists to rationally select the preferred microbial host, design the best-performing pathway, and optimise the culture media appropriate for the product of interest (Lakshmanan et al., 2015)

2.3.3 Analytical estimation of pathway potential

As a means to move away from targeted assessments and towards the ability to predict optimal catalysts, Dugar and Stephanopoulos (2011) published a theoretical framework that assessed the imbalance of metabolic pathways and the effect this has on theoretically optimal product yield (Dugar and Stephanopoulos, 2011). Using stoichiometric and energetic calculations, they quantified the relative potential of synthetic pathways, concluding that the most effective equilibrium between substrate and product optimisation was found in fully balanced (net zero) or ATP-requiring (negative ATP yield) pathways. Assuming the cell's goal of redox neutrality, driven by the laws of thermodynamics which govern chemical equilibria and determine the energetically favourable directionality of chemical reactions (Noor et al., 2014),Dugar and Stephanopoulos (2011) suggest that pathways generating excess ATP, NAD(P)H or CO_2 along with the product are inefficient from a carbon yield perspective because, on a per carbon-mol basis, excess carbon is being diverted towards biomass production, maintenance of redundant pathways and waste re-

lease, instead of contributing towards carbon accumulation in the form of target production. On the other hand, balanced or energy-requiring pathways are better at transducing substrate energy into product, resulting in better yields. Dugar and Stephanopoulos (2011) use both biomass and glycerol sinks to channel excess energy and electrons. In fact, limiting byproduct formation such as glycerol production during yeast fermentation is one such strategy shown to have improved bioproduction (Wang et al., 2013).

In Dugar and Stephanopoulos (2011), "pathway" referred to the leading route towards target production from a central carbon metabolite, not the entire metabolic network of the cell. Their calculations facilitated a comparison between different pathway yields after adjusting for any imbalances, providing insights into where the imbalance in question may be occurring and an adjusted theoretical yield estimate. This information can then be used to select better performing pathways and guide engineering strategies to render the pathway more balanced and thus more yield-efficient. However, their approach is built on a set of case-specific and not easily generalisable assumptions, does not consider various experimental conditions or biological settings, nor does the method scale up to larger metabolic networks or address the implications of pathway imbalance at the genome scale. Given that an understanding of cofactor metabolism is very useful and informative to predict the superiority of biosynthetic pathways, but the method published by Dugar and Stephanopoulos (2011) suffers from a lack of flexibility, I asked whether it would be possible to integrate both pathway-specific and network-specific balance assessments and carry out a similar analysis but using alternative computational frameworks.

2.3.4 Knowledge gap: early-stage cofactor profiling at the network scale

A review of the literature revealed that most state-of-the-art COBRA modelling approaches were either too narrow, or focused on the later stages of optimisation, i.e. biosynthetic pathways had already been selected and the focus had shifted to enhancing the carbon flux towards the target. Furthermore, those systematic, more comprehensive tools applicable to a wider number of scenarios, enabling the potential comparison between pathway candidates, focused on cofactor swaps more specifically, or provided insights into cellular turnover rates. Interestingly, most approaches also focused on redox availability or regeneration strategies, but there were no alternatives available for ATP.

Dugar and Stephanopoulos (2011) recognised that understanding the relevance of both ATP and NAD(P)H, as well as identifying the source of imbalance at an early stage would enable the metabolic engineering community to more easily select between various pathway variants

according to their maximal potential and their level of imbalance. In their study, Dugar and Stephanopoulos (2011) acknowledge that, beyond their proposed analytical method, genome-scale models (GSMs) may also be used to perform analogous assessments. In fact, GSMs would facilitate an evaluation of how biosynthetic pathways integrate with selected hosts and how they are limited by cofactor usage.

Having identified these gaps, this PhD study proposed a more pragmatic approach to cofactor engineering. Instead of using cofactor manipulations as an optimisation strategy for the later stages of a metabolic engineering project, a computational framework for the assessment of cofactor usage profiles, cofactor balancing techniques and potential cofactor-driven manipulations should be proposed from the early stages of pathway and host selection.

3

Materials & Methods

3.1 Computational methods

3.1.1 Programming packages

All work described in this study was done using the Constraint-Based Reconstruction and Analysis toolbox for Python (COBRApy version 0.13.3) (Ebrahim et al., 2013) and Gurobi solver (version 5.5.0) (Gurobi Optimization, Inc., 2016). All scripts and functions extensively used in this study were run using Python 3 (version 3.7.4) and the Anaconda environment (version 4.7.12) (Anaconda Software Distribution, 2020).
Packages MatPlotLib and Mplot3D (version 3.2.1) (Hunter, 2007), and Seaborn (0.9.0) (Waskom et al., 2020) and were used for result visualization.

3.1.2 Stoichiometric models of *Escherichia coli*

All simulations in this PhD study employed the original *Escherichia coli* Core Model created by Orth et al. (2010b), which contains 77 reactions and 63 metabolites from Central Carbon Metabolism (CCM). This stoichiometric model excluded all reactions for butanol production and fatty acid biosynthesis. To enable butanol and butanol precursor production *in silico*, reactions

corresponding to the biosynthetic pathways were added as per COBRApy standards (Table 3.1) into separate copies of the *E.coli* core model, to yield stoichiometric models iDAG85, iDAG87, iDAG86, iDAG88, iDAG91, iDAG83, iDAG84_butyr and iDAG84_butal, referred to as BuOH-0, BuOH-1, tpcBuOH, BuOH-2, fasBuOH, CROT, BUTYR, BUTAL for simplicity.
All engineered models included target-specific production, transport and sink reactions, i.e. reactions that drain the final product out of the metabolic network.

3.1.3 Parsimonious FBA

All models were simulated using parsimonious FBA (pFBA) for computing optimal phenotypes (Lewis et al., 2010). pFBA is a bi-level optimisation method that minimises the total sum of flux whilst optimizing for the selected objective using FBA. Net flux is minimised subject to optimal biomass as follows:

$$\begin{aligned} \min \sum_{j=1}^{m} v_{irrev,j} \\ \text{subject to } \max v_{objective} = v_{objective,lb} \\ \text{subject to } S_{irrev} \times v_{irrev} = 0 \\ 0 \leq v_{irrev,j} \leq v_{max} \end{aligned} \tag{3.1}$$

Where m = number of irreversible reactions in the network, S_{irrev} = stoichiometric matrix; v_{irrev} = non-negative, steady-state flux; $v_{objective}$ = approximates the theoretical objective; and $v_{objective,lb}$ = lower bound for the objective rate.
This is followed by the maximization of target per unit flux, which optimizes the ratio of the objective to the square of the total network flux:

$$\begin{aligned} max \frac{v_{objective}}{\sum_{i=1}^{n} v_i^2} \\ \text{subject to } Sxv = 0 \\ v_{min} < v_i < v_{max} \end{aligned} \tag{3.2}$$

Modified models were optimized for the drain of butanol or butanol precursor, whilst the wild type had maximal growth rate selected as the optimization principle. All models were initially

Table 3.1: Reactions added to the *E.coli* Core Model to implement the synthetic pathways for butanol and butanol production.

Reaction ID	Reaction Name	Gene(s)	EC	Reaction	Model(s)
ACCOAC	Acetoacetyl-CoA carboxylase	*acc*	6.4.1.2	accoa_c + hco3_c + atp_c => adp_c + pi_c + malcoa_c	BuOH-1, BuOH-2, fasBuOH
ACPS1	Acyl-ACP synthetase	*aas*	6.2.1.20	apoACP_c + coa_c => ACP_c + h_c + pap_c	fasBuOH
BPNT	3'(2'),5-biphosphate nuclotidase	*cysq*	3.1.3.7	pap_c + h2o_c => amp_c + pi_c	fasBuOH
BTBTAC	Acyl-CoA thioester hydrolase	*ycia*	3.1.2.20	butcoa_c + h2o_c => btac_c + coa_c	tpcBuOH, BuOH-2, BUTYR
BUT1	Acetyl-CoA acetyltransferase	*atob*	2.3.1.9	2 accoa_c => acoa_c + coa_c	BuOH-0, tpcBuOH, fasBuOH, CROT, BUTYR, BUTAL
BUT2	3-hydroxybutyryl-CoA dehydrogenase	*hbd*	1.1.1.157	acoa_c + nadh_c + h_c => 3hbcoa_c + nad_c	BuOH-0, BuOH-1, BuOH-2, tpcBuOH, CROT, BUTYR, BUTAL
BUT3	3-hydroxybutyryl-CoA dehydratase	*crt*	4.2.1.55	3hbcoa_c => b2coa_c + h2o_c	BuOH-0, BuOH-1, tpcBuOH, BuOH-2, CROT, BUTYR, BUTAL
BUT4	Trans-2-enoyl-CoA reductase	*ter*	1.3.1.44	b2coa_c + h_c + nadh_c => butcoa_c + nad_c	BuOH-0, BuOH-1, tpcBuOH, BuOH-2, BUTYR, BUTAL
BUT5	Aldehyde-alcohol dehydrogenase	*adhe2*	1.1.1.1	butcoa_c + h_c + nadh_c => btal_c + coa_c + nad_c	BuOH-0, BuOH-1, BUTAL
BUT6	Aldehyde reductase	*ahr*	1.1.1.2	btal_c + h_c + nadh_c => nad_c + nbutanol_c	BuOH-0, BuOH-1, tpcBuOH, BuOH-2, fasBuOH
CAR	Carboxylic acid reductase	*car*	1.2.99.6	btac_c + nadph_c + atp_c + h_c => btal_c + adp_c + nadp_c + amp_c + 2.0 pi_c + h2o_c	tpcBuOH, BuOH-2, fasBuOH
B2CTCRO	Acyl-CoA thioester hydrolase	N.A.	3.1.2.20	b2coa_c + h2o_c => coa_c + croac_c	CROT

Reaction ID	Reaction Name	Gene(s)	EC	Reaction	Model(s)
EAR40x	Enoyl-ACP reductase	N.A.	1.3.1.9	but2eACP_c + h_c + nadh_c => butACP_c + nad_c	fasBuOH
HCOE3	Carbonic anhydrase	*cynt.*	4.2.1.1	co2_c + h2o_c => hco3_c + h_c	BuOH-1, BuOH-2, fasBuOH
KAS15	Beta-ketoacyl-ACP synthase II	*fabf*	2.3.1.41	accoa_c + h_c + malACP_c => actACP_c + co2_c + coa_c	fasBuOH
MCOATA	Malonyl-CoA-ACP transacylase	*fabD*	2.3.1.39	ACP_c + malcoa_c <=> coa_c + malACP_c	fasBuOH
NPHT7	Acetoacetyl-CoA synthase	*npht7*	2.3.1.194	accoa_c + malcoa_c => aacoa_c + co2_c + coa_c	BuOH-1, BuOH-2
3HAD40	3-hydroxyacyl-ACP dehydratase	*fabF*	4.2.1.59	3haACP_c => but2eACP + h2o_c	fasBuOH
3OAR40	3-oxoacyl-ACP reductase	*fabG*	1.1.1.100	actACP_c + h_c + nadph_c <=> 3haACP_c + nadp_c	fasBuOH
5_BUT1	Acyl-ACP thioesterase	*aat*	N.A.	butACP_c + h2o_c => ACP_c + btac_c	fasBuOH
BTAC_sink	butyric acid drain from model	N.A.	N.A.	btac_e =>	BUTYR
BTAC_tr	butyric acid transport to extracellular space	N.A.	N.A.	btac_c => btac_e	BUTYR
BTAL_sink	butyraldehyde drain from model	N.A.	N.A.	btal_e =>	BUTAL
BTAL_tr	butyraldehyde transport to extracellular space	N.A.	N.A.	btal_c => btal_e	BUTAL
BTOH_sink	Butanol drain from model	N.A.	N.A.	nbutanol_e =>	BuOH-0, BUOH-1, tpcBuOH, BuOH-2, fasBuOH
BTOH_tr	Butanol transport to extracellular space	N.A.	N.A.	nbutanol_c => nbutanol_e	BuOH-0, BuOH-1, tpcBuOH, BuOH-2, fasBuOH
CROAC_sink	crotonic acid drain from model	N.A.	N.A.	croac_e =>	CROT
CROAC_tr	crotonic acid transport to extracellular space	N.A.	N.A.	croac_c => croac_e	CROT

unconstrained. Prior to optimization, the model was pre-processed to set the primary carbon source, glucose, to a maximum intake rate of -10 mmol gDW^{-1} hr^{-1}. Constraints necessary for computational minimal media conditions were set as default (Feist et al., 2007). The presence or absence of oxygen was also modulated to run simulations under both aerobic and anaerobic conditions. In aerobic simulations, the oxygen uptake rate was set to a maximum of -10 mmol gDW^{-1} hr^{-1}. Alternatively, oxygen uptake was constrained to zero under anaerobic conditions.

3.1.4 Flux Variability Analysis

Flux variability analysis (FVA) was used to calculate the minimum and maximum allowable flux values within the *E.coli* Core model (Orth et al., 2010b) whilst satisfying FBA mass balance conditions and all other constraints in the system (Mahadevan and Schilling, 2003). The feasible range of reaction fluxes by maximization and minimization was calculated at 100% of the maximal value of the objective function, unless explicitly stated otherwise. Similarly, FVA was run on all reactions in the model, unless stated otherwise. The mathematical formulations for maximization and minimization are shown below:

Maximization

$$\begin{aligned} &\max v_j \\ \text{subject to } &S_{ij} \times v_j = 0 \\ &C^T v = Z_{objective} \\ &v_j^{min} \leq v_j \leq v_j^{max} \end{aligned} \tag{3.3}$$

Minimization

$$\begin{aligned} &\min v_j \\ \text{subject to } &S_{ij} \times v_j = 0 \\ &C^T v = Z_{objective} \\ &v_j^{min} \leq v_j \leq v_j^{max} \end{aligned} \tag{3.4}$$

Where $Z_{objective}$ = value of the objective function. If n is the number of fluxes, then 2n LP problems are solved under FVA.

3.1.5 Minimization of Metabolic Adjustment

Minimization of Metabolic Adjustment (MOMA) (Segre et al., 2002) searches for a vector x that presents minimal distance from a given flux vector L so that the euclidean distance is minimized as:

$$minf(x) = L(x) + \frac{1}{2}x^T Qx \tag{3.5}$$

Where L = vector of length N; Q = NxN matrix defining the linear and quadratic part of the objective function; x^T = transpose of x; x = vector with minimal euclidean distance from L

3.1.6 Cofactor Balance Assessment

Cofactor Balance Assessment (CBA) is an algorithm built to track intracellular production of ATP, NADH and NADPH across a selected metabolic network and determine the total cofactor flux contributing to biomass, waste, cell maintenance and target chemical production (if a synthetic pathway is introduced and used as objective function). The algorithm was written as a Python script, and built using the COBRApy package. CBA requires two main inputs. First, a stoichiometric model, often a GSM, is the source of co-factor related reaction information. Second, FBA estimates or equivalent (pFBA or MOMA estimates, for example) provide flux information for all reactions contained in the selected model.

The CBA algorithm evaluates all reactions involving ATP or NAD(P), and determines whether such cofactors are either consumed or produced by multiplying the reaction flux (from FBA or equivalent) and cofactor stoichiometry (from the stoichiometric model). This value is recorded as a reaction-specific cofactor flux score (CFS), symbolising the level at which said cofactor is being consumed or produced by the reaction under evaluation. For total ATP and NAD(P)H production, all positive CFSs in the network are summed up. Then, cofactors fluxes are grouped into categories: (1) ATP produced (Net ATP produced within CCM); (2) ATP target (total ATP consumed for target optimization); (3) ATP waste (total ATP burned to produce ADP and/or AMP, or ATP consumed or produced during the release of CO2 or fermentation by-products); (4) ATP biomass (ATP consumed during biomass formation); (5) NAD(P)H produced (Net NAD(P)H produced within CCM); (6) NAD(P)H target (NAD(P)H released during target production); (7) NAD(P)H biomass (NAD(P)H released during biomass production); and (8) NAD(P)H waste (NAD(P)H generated or consumed during CO2 or fermentation by-product release). Finally, the scores are added together into summed flux values that describe the overall "weight" of each

category. Only ATP and NAD(P)H related reactions were considered within the scope of this study. Further information about CBA's foundations, function development and assumptions can be found in Chapter 5, Section 5.2.

3.2 Experimental methods

3.2.1 Strains, plasmids and primers

The lambda red recombinase-integrated *Escherichia coli* strain SIJ488 (hereinafter SIJ488)was used as the parental strain throughout this study. This *E.coli* strain, developed by Jensen et al. (2015) was kindly provided by the Bacterial Cell Factory Optimisation Laboratory, at the Technical University of Denmark (Kemitorvet, Denmark).
All engineered strains and plasmids developed and used in this study were included in Tables 3.2 and 3.3. Strains were maintained as glycerol-based cryo stocks (see below) and kept at $-80°C$. All primers were synthesized and ordered from Integrated DNA Technologies (IDT, Belgium), and are listed in Table 3.4. Primer stocks were maintained at $-20°C$.

Table 3.2: List of strains used in this study.

Strain Name	Description	Source
SIJ488	*E.coli* K-12 MG1655Tn7::para-exo-beta-gam;prha-FLP;xy1Spm-IsceI containing the genome-integrated lambda red recombinase system	(Jensen et al., 2015)
ATPs	As per SIJ488, inc. ΔatpIBEFHAGDC	This study
PCK	As per SIJ488, inc. Δpck	This study
ATPs-PCK	As per SIJ488, inc. ΔatpIBEFHAGDC Δpck	This study
PDC	DH5*α* carrying pyruvate decarboxylase (PDC) gene under the $P_{AlacO-1}$ promoter in the pCDF plasmid backbone	This study
pCDF	DH5*α* carrying pCDF plasmid	This study

Table 3.3: List of plasmids used in this study.

Plasmid name	Description	Source
pKD13	Ap^R, FRT-Km^R-FRT, *ori*R6K, includes a kanamycin resistance cassette	(Datsenko and Wanner, 2000)
pCDF-empty	pCDF plasmid backbone	This Study
pCDF-PDC	pCDF plasmid backbone carrying the pyruvate decarboxylase (PDC) gene under $P_{AlacO-1}$ promoter	This Study

Table 3.4: List of primers used in this study. All primers were designed during this study unless otherwise stated.

Primer Name	Primer Sequence	References
F_kt_val	CGG CCA CAG TCG ATG AAT CC	(Datsenko and Wanner, 2000)
R_kt_val	CGG TGC CCT GAA TGA ACT GC	
F1_ATPS4r	AAA AAA GCC AGC CTG TTT CCA GAC TGG CTT TTG TGC TTT TCA AGC CGG TGG TGT AGG CTG GAG CTG CTTC	
R2_ATPS4r	TTT TAT ACG ACA CGC GGC ATA CCT CGA AGG GAG CAG GAG TGA AAA ACG TGA TTC CGG GGA TCC GTC GACC	
F_PPCK	CAA AAA GAC TTT ACT ATT CAG GCA ATA CAT ATT GGC TAA GGA GCA GTG AAG TGT AGG CTG GAG CTG CTTC	
R_PPCK	CGT TTT GCT TTC TAT AAG ATA CTG GAT AGA TAT TCT CCA GCT TCA AAT CAA TTC CGG GGA TCC GTC GACC	
F_seq_ATPS4r	AGC GTC AGG TGG ATG TTT TTG	
R_ATPS4r	CGG GGG CGC ACC GTA TAA TT	
seq_F_PPCK	TCG TGA CAG GAA TCA CGG AG	
seq_R_PPCK	GCA GGG CAC GAC AAA AGA AG	

3.2.2 Media and reagents

3.2.2.1 LB, SOC and M9 media

Lysogeny Broth (LB) liquid medium was prepared by dissolving 6 g of LB Broth powder (g/L: 10 tryptone, 5 yeast extract, and 5 NaCl) (Sigma Aldrich, Dorset, UK) in 300mL deionized water in a 500mL duran bottle (Sigma Aldrich). The LB medium was sterilized by autoclaving for 15min at 15psi (121°C) and cooled down to room temperature prior to use.

SOC medium (0.5% yeast extract, 2% tryptone, 10mM NaCl, 2.5mM KCl, 10mM $MgSO_4$ and 20mM glucose), prepared and kindly provided by my colleague Jessica Rollit (then a PhD student), was utilized to recover mutants after electroporation. M9 minimal medium, prepared as per standard procedure using 5X M9 Salts ($MgSO_4$, $CaCl_2$) (Sigma Aldrich) was used for growth and labelling experiments.

Table 3.5: Antibiotic concentrations for antibiotic stock preparation

Antibiotic	Stock conc. (mg/mL)	Working conc. (mg/L)	Solvent
Kanamycin	50	50	deionized water
Spectinomycin	50	50	deionized water

3.2.2.2 LB Agar medium

LB agar medium (1.5% w/v) was prepared by adding 4.5g agar powder (VWR International, UK) to 300mL LB liquid medium in a 500mL duran bottle (Sigma Aldrich). The LB agar medium was sterilized by autoclaving for 20 min at 15psi (121°C). The hot mix solution was cooled down to 50-60°C before antibiotic(s) was added appropriately. After cooling, 10-20mL of LB agar were aseptically poured into 10cm petri dishes (VWR) to make "agar plates". The agar plates were then used when the medium had hardened completely after cool-down.

3.2.2.3 Antibiotic preparation

To create a stock solution of antibiotic (Table 3.5), antibiotic powder was dissolved in deionized water and filter-sterilized with sterile 0.2 m syringe filters. Stock solutions were stored at -20°C and thawed at room temperature prior to use. Antibiotics were added at 30 μg mL^{-1} or 50μg mL^{-1}, where appropriate.

3.2.2.4 Other chemicals

Other chemicals and reagents, unless otherwise stated, were ordered from Sigma-Aldrich, VWR or Thermo Fischer Scientific (Fisher Scientific Ltd. UK).

3.2.3 Molecular biology techniques

3.2.3.1 Plasmid DNA extraction

QIAprep Spin Miniprep Kit (QIAGEN) was used for routine plasmid DNA extraction from *E. coli* liquid cultures according to the manufacturer's instructions. In brief, 5mL of overnight culture was centrifuged in a Heraeus Megafuge 16R benchtop centrifuge (Thermo Fisher Scientific) at 4,000 x g for 5min. The supernatant was discarded and the cell pellet resuspended in 250μL Buffer P1 containing RNase A and transferred to a 1.5mL microcentrifuge tube (VWR). 250μL

Buffer P2 was added to lyse the cells and the microcentrifuge tube was inverted 4-6 times until the solution became clear. 350μL neutralization Buffer N3 was added and the sample was mixed immediately by inverting the tube 4-6 times. Samples were then centrifuged for 10min at 17,900 x g in an an Eppendorf 5424R microcentrifuge (Eppendorf, UK) to separate the cell debris and proteins from the supernatant. The supernatant was transferred to the provided QIAprep 2.0 spin column and centrifuged for 1min at 17,900 x g. The flow-through in the collection tube was discarded and the QIAprep 2.0 spin column was washed twice with 750μL and 500μL Buffer PE (containing ethanol) by centrifugation for 1min at 17,900 x g. The flow-through was discarded and the empty QIAprep 2.0 spin column was centrifuged one final time to remove any residual PE Buffer. The QIAprep 2.0 spin column was removed from the collection tube and placed onto a sterile 1.5mL microcentrifuge tube (VWR). To elute the plasmid DNA, 30μL of filter-sterilized deionized water was added to the QIAprep 2.0 spin column and incubated at room temperature for 1min before it was centrifuged for 1min at 17,900 x g. The concentration of the eluted plasmid DNA was then quantified using a NanoQuant plate (see Section 3.2.3.6).

3.2.3.2 Q5®High-Fidelity DNA Polymerase Chain Reaction

To amplify DNA fragments, Q5®High-fidelity DNA polymerase (New England BioLabs, Hitchin, UK) was used. Stock solution of 10mM dNTPs was also purchased from New England Bio-Labs. All solutions provided with the PCR kit were thawed on ice and mixed according to the manufacturer's protocol (Table 3.6) in a PCR tube (VWR).

Table 3.6: Q5®High-Fidelity DNA Polymerase PCR reaction components.

Component	Volume (μL)
5X Q5 Reaction Buffer	4
Forward primer (10 M)	1
Reverse primer (10 M)	1
10 mM dNTPs	0.4
Template DNA	1
Q5 High-fidelity DNA Polymerase	0.2
Sterile deionized water	up to 20μL

The PCR was then performed in a thermocycler (Biometra Professional Trio, Biometra) following

the thermal cycling conditions outlined in Table 3.7. The PCR products were then analysed with agarose gel electrophoresis (see Section 3.2.3.4).

Table 3.7: Q5®High-Fidelity DNA Polymerase thermal cycling conditions.

Step	Temperature (°C)	Time	Number of cycles
Initial denaturation	98	30s	1
Denaturation	98	10s	
Annealing	55-57	30s	
Extension	72	45 (30s/kb)	35
Final extension	72	2min	1
Hold	4	∞	

3.2.3.3 Colony PCR

To validate electroporation and knockout validation, DreamTaq Green PCR Master Mix (2X) (Thermo Fisher Scientific) was used. The DreamTaq Green PCR Master Mix (2X) solution was gently vortexed and briefly centrifuged after it was thawed on ice. The composition of PCR reaction solution is listed in Table 3.8.

Table 3.8: DreamTaq Green PCR Master Mix PCR reaction components.

Component	Volume (μL)
DreamTaq Green PCR Master Mix (2X)	7.5
Forward primer (10 M)	0.8
Reverse primer (10 M) Template DNA	0.8
Sterile deionized water	up to 20μL

To provide the template DNA, a single *E.coli* colony was picked from an LB agar plate using a a sterile disposable inoculating loop (VWR) and resuspended in the PCR reaction solution by scrubbing the tip on the wall of the PCR tube. The PCR was then carried out in a thermocycler (Biometra Professional Trio, Biometra) by following the thermal cycling conditions as outlined in

Table 3.9. The PCR products were then analysed with agarose gel electrophoresis (see Section 3.2.3.4).

Table 3.9: DreamTaq Green PCR Master Mix (2X) thermal cycling conditions.

Step	Temperature (°C)	Time	Number of cycles
Initial denaturation	95	3min	1
Denaturation	95	30s	
Annealing	55-57	30s	
Extension	72	1.5min (1min/kb)	35
Final extension	72	15min	1
Hold	4	∞	

3.2.3.4 Agarose gel electrophoresis

To prepare an agarose gel, 1g agarose (Sigma Aldrich) was mixed with 100mL 1X TAE buffer (40mM Tris, 20mM acetic acid, 1mM EDTA pH 8.3) in a 500mL duran erlenmeyer flask and dissolved by heating the mix until all agarose was completely dissolved. After cool-down, 7μL SYBR Safe DNA gel stain (Thermo Fisher Scientific) was added to the solution and mixed well. Next, the mix was poured into a gel casting tray, and an appropriate comb (Bio-Rad Laboratories, Watford, UK) was inserted to create the required number of wells for the loading of PCR products. The agarose solution was allowed to solidify for 30mins at room temperature, which was followed by the removal of the comb from the gel. Next, the gel (in its casting tray) was transferred to an electrophoresis tank (Bio-Rad Laboratories), and filled with 1X TAE buffer solution. Prior to gel loading, Q5®PCR products (20μ/L total volume) were mixed with 4μL of gel loading dye (6X) (New England BioLabs) and loaded into the gel wells. DreamTaq Green PCR products required no gel loading dye, so the PCR products were loaded onto the gel directly. To determine the size of the PCR products, 2μL of DNA ladder (New England BioLabs) were loaded onto the gel for reference. Finally, the gel tank was covered, and the gel was run at 100V for 30min. To visualize the DNA samples, the gel was placed on a blue-light transilluminator (Clare Chemical Research) or in a UV gel imager (GelDoc-It, Ultra-Violet Products, Cambridge, UK).

3.2.3.5 DNA extraction from agarose gel

To extract the PCR products from the agarose gel, the gel was placed on a blue-light transilluminator and the DNA bands of interest were excised using a washable metal scalpel. The excised gel was transferred to a 1.5mL microfuge tube (VWR) and the weight of the gel was determined using a weighing scale. The DNA was purified using QIAquick Gel Extraction Kit (QIAGEN), following the manufacturer's instructions. In brief, three volumes of QG Buffer were added to 1 volume of gel (e.g. 300μL of Buffer QG was mixed with 100mg gel). The gel slice was dissolved in the buffer by incubating the mixture at 50°C and vortexing occasionally until the solution became homogenous. One volume of isopropanol (e.g.100μL for 100mg gel) was added to the solution and mixed accordingly by pipetting. The solution was then transferred into the provided spin column placed in the provided collection tube. The spin column together with the collection tube were then centrifuged at 17,000 x g for 1min (Eppendorf). The flow-through solution in the collection tube was discarded and 700μL of the provided PE Buffer was added into the spin column followed by centrifugation at 17,000 x g for 1min. The flow-through solution was discarded and the spin column was centrifuged one more time to remove any left over PE Buffer. Finally, the spin column was placed into a fresh 1.5mL microcentrifuge tube and 30μL nuclease-free water (New England BioLabs) was added directly into the membrane matrix in the spin column and incubated at room temperature for 1min. To elute the DNA, the column was centrifuged for 1 min at 17,900 x g. The concentration of the purified DNA was determined using the NanoQuant plate (see Section 3.2.3.6).

3.2.3.6 Quantification of plasmid DNA concentration

The concentration of plasmid DNA was quantified in a NanoQuant plate using a Tecan Infinite M200 Pro plate reader (Tecan AG, Reading, UK). Briefly, 2μL of filter-sterilized deionized water were loaded as blank to the sample well in a NanoQuant plate to calibrate the machine. After wiping the water droplet with lint-free Kimtech Science Kimwipes (Kimberly-Clark) 2μL of sample was loaded onto the sample well. Absorbance of nucleic acid sample was measured at 260nm, as well as an additional measurement at 280nm to assess the purity of the nucleic acid and to indicate the presence of proteins. Samples with 260/280 ratio between 1.8-2.0 were used for further experiments.

3.2.4 Culturing conditions

3.2.4.1 Cryopreservation of *E.coli* liquid cultures

70% glycerol was used to cryopreserve all *E.coli* strains used in this study. A 70% (v/v) glycerol stock solution was prepared by mixing 30mL of glycerol (Sigma Aldrich) with 70mL deionized water. The glycerol solution was then sterilized by autoclaving for 20min at 15 psi (121°C) and cooled down to room temperature prior to use. To prepare the *E.coli* glycerol stocks, 750μL of 70% (v/v) glycerol were mixed aseptically with 250μL of *E.coli* overnight culture in a 2mL cryogenic vial. Glycerol stocks were stored in -80°C freezer.

3.2.4.2 Bacterial cultivations

Under aseptic conditions, performed in class I or class II biosafety cabinets under continuous laminar flow, a sterile disposable inoculating loop (VWR) was used to restreak the *E. coli* glycerol stock on an LB agar plate containing appropriate antibiotic(s). LB agar plates were incubated in a Labnet Mini Incubator (Appleton Woods Ltd., UK) overnight at 37°C. The next day, liquid cultures were inoculated by picking individual colonies using a 1mm plastic inoculation needle and mixing it with 5mL liquid LB medium supplemented with the appropriate antibiotics in 50mL centrifuge tubes. Agar plates were sealed with plastic paraffin film (VWR), stored at 4°C fridge and discarded when no longer needed. Cultures were grown overnight at 37°C and 200 r.p.m. orbital shaking in a shaker incubator (New Brunswick 44R, Eppendorf, UK). Tube caps were loosened to allow proper aeration of the cultures. Liquid cultures were then processed for further experimental use.

To cultivate *E. coli* for the production of ethanol, the overnight pre-cultures were washed twice by centrifugation at 4,000 x g for 5min and resuspended in fresh M9 minimal medium. The OD_{600} of resuspended cultures in M9 minimal medium was measured and adjusted to a starting OD of 0.1 in 25mL M9 minimal medium supplemented with 2% (w/v) glucose in 100 mL erlenmeyer flasks. All strains were incubated at 37°C and 180 r.p.m. orbital shaking for 4h before inducing with 0.25mM, 0.5mM or 0.75mM IPTG. After induction, culturing proceeded at 30°C and 150 r.p.m. orbital shaking for 48h. Samples were taken every 24h.

3.2.5 Generation of mutant strains and bioproduction strains

Knockout strains were constructed using the method specified by Jensen et al. (2015), and validated as suggested by Datsenko and Wanner (2000). Briefly, recombination primers including

a 50bp homology arms complementary to the upstream and downstream regions flanking the target gene(s) were designed (Table 3.4), and used to amplify by Q5®PCR (Section 3.2.3.2) the FRT-flanked kanamycin resistance (kan^R) gene of plasmid pKD13 (purified from DH5α *E.coli* following the plasmid extraction instructions outlined in Section 3.2.3.1). After validation using gel electrophoresis and PCR product gel extraction (Sections 3.2.3.4 and 3.2.3.5), gene deletions were performed by mixing 50μL aliquots of electrocompetent SIJ488 with up to 500ng of purified PCR product (containing the 50-bp arms homologous to the 5' and 3'-ends of the target gene(s)) on ice, electroporated at 2.5kV, 25F, 100Ω and recovered in 1mL of SOC or LB medium at 30°C for 4h prior to positive selection.

Positive selection of successful mutant strains on LB Agar + kan plates (prepared at the appropriate antibiotic concentration as per Table 3.5) provided an initial preliminary confirmation that kan^R PCR products (flanked by the target gene's homology regions) had been successfully integrated into the *E.coli* genome. This was followed by further validation using DreamTaq Green Mix (2X) colony PCR (Section 3.2.3.3) using validation sequencing primers (Table 3.4). Validated colonies were then subjected to removal of their antibiotic resistance cassette, by inducing the expression of the genome-integrated flipase gene contained in the SIJ488 strain, which is controlled by a rhamnose-inducible promoter. Briefly, colonies were picked and grown in 5mL liquid cultures overnight in LB medium as per Section 3.2.4.2 and then diluted to a starting OD of 0.05. Cultures were incubated in the same conditions until they reached an OD 0.1-0.4. Then, L-rhamnose was added to a final concentration of 50 mM L-rhamnose (Sigma Aldrich) and induction was performed at 30°C, in LB medium for 4-6h. Finally, cassette removal was verified by negative selection on LB Agar + kan plates, GreenTaq Colony PCR and finally DNA sequencing (Appendix D Figures 1-9).

Ethanol bioproduction strains harbouring the pCDF-empty and pCDF-PDC plasmid systems were implemented by electroporation (2.5kV, 25F, 100Ω). Electroporated *E.coli* were recovered in 1mL of LB at 30°C for 4h prior to positive selection in LB-Agar plates containing spectinomycin at the required concentration (Table 3.5).

3.2.6 Analytical methods

3.2.6.1 Amino acid labelling and GC-MS detection of ^{13}C-labelled amino acids

For estimation of intracellular fluxes, a tracer experiment using [1–^{13}C] D-glucose and [U–^{13}C] D-glucose and ^{13}C Metabolic Flux Analysis (MFA) was performed following the widely recognized protocol by Zamboni et al. (2009). Briefly, colonies were first grown on LB plates from cryo

stocks, and biological replicate colonies (n=4) were picked and cultivated in 5mL LB medium at 37°C and 180 r.p.m. orbital shaking in Innova-44 incubators overnight. 50μL of pre-culture were transferred into 5mL of M9 medium supplemented with 2% (w/v) unlabelled glucose and grown at 37°C overnight before growth analysis and ^{13}C labelling experiments.

The labelling experiment was performed out in 300mL shake flasks with a working volume of 30mL at 37°C, using a mixture of 20% (wt/wt) 99% pure [U–^{13}C] D-glucose (Cambridge Isotopes, Cambridge, UK) and 80% [1–^{13}C] D-glucose, up to a final glucose concentration in medium of 0.3% (w/v). 1mL samples were taken every 2-4h for growth and physiological parameter analysis (Section 3.2.6.2). Biomass samples were harvested at mid-exponential phase (OD_{600} 0.8), and centrifuged at 5000 r.p.m. for 10min at 4°C, washed in 0.9% NaCl, and hydrolyzed in a fume hood with 200μL of 6M HCl at 105°C for 24h, to break down the protein-based cell content (which amounts to almost half of the total cell content) down to its amino acids. Then, also in a fume hood, samples were dried at 85∘C. Dried amino acids in the hydrosylate were then carefully solubilized in 40μL of dymethylformamide (DMF) and derivatized with 40μL of TBDMSTFA at 85°C for 1h. During this procedure, the amino acids become volatile so that they can be later separated and quantified by Gas Chromatography Mass Spectrometry (GC-MS). After incubation, the derivatized samples were transferred to a vial insert (5183-2085, Agilent) in a GC vial with a silicon sealed cap (Agilent Technologies, Cheadle, UK), and analyzed using GC-MS.

GC-MS analysis was performed using a 7890B Series gas chromatograph equipped with a 30m $\times$ 0.25mm $\times$ 0.25μm ZB-5MSi column (Agilent Technologies). Helium flow was held at 1.5mL min^{-1}. 1μL samples were injected 3 times at a split ratio of 1:5. The temperatures for the inlet, interface, and ion source were 230°C, 250°C, and 200°C, respectively. After a 2.4min solvent delay at 80°C, the oven temperature was increased by 5°C min^{-1} up to 310°C at which it was held for 6min before dropping back to 80°C for the next cycle. At the interface between the GC and MS, electron impact ionization caused the fragmentation of the derivatized amino acids. Electron impact (70eV) mass spectra were recorded from m/z 50 to 600 at 1.4 scans sec^{-1}. The instrument was autotuned for mass calibration.

To guarantee analytical accuracy, analytical amino acid standards were tested first, as well as unlabelled biomass samples, in replicates and each injected three times. A quantification table was set up in the Enhanced Data Analysis software (Agilent GC-MS 5975, Agilent Technologies) to automate the integration of the ion chromatogram peak areas of all ion fragments of interest corresponding to the amino acid isotopes under evaluation in this study, which included all single ion traces of masses up to 3 a.m.u. heavier than a fully labeled fragment.

3.2.6.2 Growth and physiological parameter analysis

Throughout the labelling experiment, physiological parameters such as growth rate, fermentation product release, biomass yield and glucose uptake rate were measured. 1mL samples were taken every 2-4h. As an indicator for growth, cell optical density (OD) at 600nm (OD_{600}) was measured in a 1.5mL cuvette (Brandtech Scientific, Inc.) using a Tecan200P microplate reader and fresh M9 minimal medium as a blank. After optical density measurements, the 1mL samples were transferred to 1.5mL microcentrifuge tubes (VWR) and centrifuged at 13,000 r.p.m. for 5min. The supernatant was transferred into a 2mL HPLC vial (Agilent Technologies) with a silicon sealed screw cap and stored at -20°C for further High-performance Liquid Chromatography (HPLC) analysis.

An Agilent 1200 series HPLC instrument equipped with an Aminex HPX-87H column (Bio-Rad Laboratories) and a reflective index detector RID-10A was used to determine the concentrations of glucose, ethanol, lactate, formate, succinate and acetate in 100μL samples, using 5mM H_2SO_4 as the mobile phase, a flow rate of 0.6mL minˆ1 and a column temperature of 60°C. For HPLC characterization serial dilutions of glucose (Sigma Aldrich), sodium acetate (Sigma Aldrich), sodium lactate (Sigma Aldrich), sodium formate (Sigma Aldrich), sodium succinate (Fisher Scientific) and absolute ethanol (VWR) were used to create standard curves for determining the amounts of these compounds in the samples (an example of such standard curves is included in Appendix E Figure 10).

Specific growth rates were determined from log-linear regressions of time-dependent changes in optical density. (Appendix E Figure 11). Given the clearly linear relationship between time (h) and OD_{600}, 2 data points were used to calculate the growth rate (4h and 8h for WT and PCK, and 6h and 10h for ATPs), using the formula below:

$$\mu = \frac{(\log_{10} OD_{t2} - \log_{10} OD_{t1}) x 2.303}{t_2 - t_1} \tag{3.6}$$

Where; OD_{t2} = OD_{600} at timepoint 2; OD_{t1} = OD_{600} at timepoint 1; t_2 = timepoint 2; t_1 = timepoint 1

Generation time, which provides the number of generations produced in the selected timeframe, can be calculated as follows:

$$g = \frac{\log_{10} OD_{t2} - \log_{10} OD_{t1}}{\log_{10} 2} \tag{3.7}$$

The relationship between growth rate and generation time can be verified by:

$$\mu = \frac{\ln 2}{g} \tag{3.8}$$

Growth rates and generation times are shown in Appendix E Table 14.

The OD_{600} values were converted to cell dry weight concentrations (gDW L^{-1}) using a linear fit calibration curve (Appendix E Figure 12). Briefly, strains were first grown overnight in LB at 37°C before washing with M9 minimal medium and incubating in M9 at 37°C and 180 r.p.m. orbital shaking until enough biomass accumulated to prepare triplicates of biomass dilutions at OD 0.8, 0.4, 0.2, 0.1 and 0.05. Biomass samples were filtered using filter paper (47mm diameter, 0.7μm pore size) and dried at 60°C until the weight remained constant. The increase in weight of the filter paper represented the dried biomass weight of the samples, which were plotted to find a linear relationship between OD_{600} and gDW L^{-1} values.

Uptake and secretion rates were determined using two timepoints, one from the start of exponential growth and mid-exponential growth (4h and 8h for WT and PCK, and 6h and 10h for ATPs) and calculated using the following equation, adapted from Chen et al. (2011) and Sauer et al. (1999):

$$v = \frac{C_{t_2} - C_{t_1}}{gDW_{t_2} - gDW_{t_1}}\mu \tag{3.9}$$

Where, C_t and gDW_t are the extracellular metabolite concentrations and biomass dry weight at time t; and μ is the growth rate of the culture.

3.2.6.3 Metabolic Flux Analysis

From the isotope distribution patterns of the proteinogenic amino acids (Section 3.2.6.1), relative pathway contributions (metabolic flux ratios) were computed, internal metabolic fluxes were estimated using the WUFlux software, following its specifications (He et al., 2016). WUFlux includes a data correction routine to perform batch-wise comparison between all measured amino acid fragments and theoretical mass isotopic distributions. It also normalizes the integrated areas such that the sum of all mass isotopes for a given fragment is 1, and flags measurement inaccuracies. WUFlux uses a stoichiometric model including all relevant reactions in central carbon metabolism, including the Embden-Meyerhof-Parnas pathway, oxidative pentose phosphate pathway (PPP), Entner–Doudoroff pathway, the TCA cycle, Calvin cycle (only in the cyanobacterial template), the C1 (including 5-methyl-tetrahydrofolate and 5,10-methylene-tetrahydrofolate) metabolism path-

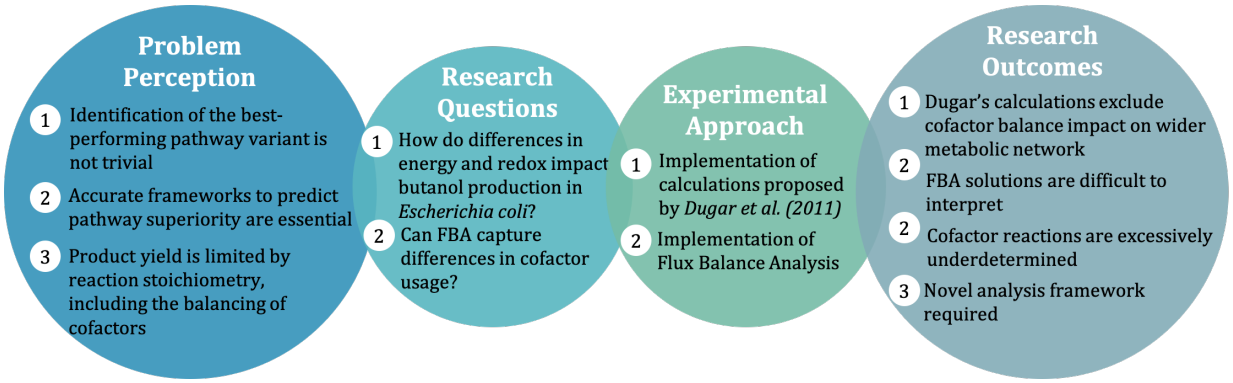
way, and anaplerotic pathway. The model used in this study is shown in Appendix F Table 15.

3.2.7 Statistical analysis

Assuming a normal distribution (Fay and Gerow, 2013), a two-sided student's t-test was used to determine statistically significant changes across independent samples, with asterisks indicating significance (* = $P \leq 0.05$; ** = $P \leq 0.01$; *** = $P \leq 0.005$). All experimental data presented in this study is shown as average ± SD from three (or four, if it is a labelling experiment) biological replicates, unless stated otherwise. All individual data points have also been included in the figures.

4

Energy and Redox Impact on Butanol Production



Notes from the Author:

This chapter builds on previously submitted work for an earlier qualifying degree (de Arroyo Garcia, 2016). Using the same case studies as a proof-of-concept, some elements of the analysis were replicated, but the modelling approach was updated, run using a different stoichiometric model and pre-existing conditions, as well as additional analytical frameworks.

Part of the work in this chapter has also been published in Paper I: **de Arroyo Garcia, L.**, Jones, P. R. (2020) "*in silico* co-factor balance estimation using constraint-based modelling informs metabolic engineering in *Escherichia coli*" . *PLoS Comput Biol* 16(8): e1008125. https://doi.org/10.1371/journal.pcbi.1008125

4.1 Introduction

The primary objective of producing industrially-attractive chemicals using metabolic engineering is to obtain the highest possible conversion of desired product from the original substrate. Here, microbes are treated as factories that synthesize industrial compounds, so it is essential to consider the capability of the cells' native metabolism to mesh with the demands of synthetic pathways, as these pathways will alter cellular homeostasis. The maximum theoretical yield will be affected, amongst other factors, by the fine balance between consumption and production of metabolic cofactors in the metabolic network (Dugar and Stephanopoulos, 2011, Varma et al., 1993). It is difficult to precisely pin-point the most important cofactors influencing the performance of a catalytic system. However, as I show in Chapter 2, it is well understood that glucose catabolism inevitably results in the creation of ATP and NAD(P)H, which are essential for the conversion of available carbon source into target chemical, biomass and by-products, and must be recycled to achieve mass balance in central metabolism.

When growing in steady state conditions, microorganisms coordinate the production of energy and redox to match consumption. Their network structures and regulatory systems will assist in performing this balancing act. It is therefore no surprise that cofactor balancing is poorly optimized for any synthetic objectives, and cofactor balancing will therefore need to be optimized if we are to effectively introduce non-native pathways that rely on native cofactor concentrations (King and Feist, 2014). Ideally, an optimal catalyst would divert most resources towards mass production of the target product, whilst retaining enough energy to guarantee cell maintenance. With this ideal scenario in mind, metabolic waste products would result when a major portion of the flux is forced through a pathway that is imbalanced with respect to ATP or NAD(P)H. Waste production would thus be symptomatic of imbalances in the metabolic network. Particularly with regards to industrial compounds, numerous pathway variants may also exist. Some pathways may be better than others with respect to the total energy content and efficiency, so it is essential that these are carefully selected to ensure high energy content and an efficient production route (Dugar and Stephanopoulos, 2011). To aid with these early stages of pathway design and strain selection, I asked whether it would be possible to harness ATP and NAD(P)H balance information to identify better-performing pathways. To what extent does cofactor usage affect target production across a range of biosynthetic pathway variants? To answer this question, 8 pathway variations for the production of butanol were selected as a case study.

4.1.1 Bio-based butanol as an alternative fuel molecule

Butanol is a four-carbon alcohol and important chemical feedstock (Saini et al., 2014). It has received particular attention because of its superior physical specifications than ethanol (the number 1 biofuel in terms of market demand, as per International Energy Agency (2018)), including higher density (29.2 vs 19.6 MJ L^{-1}), higher air-to-fuel ratio (11.1 vs 9.0), and lower hygroscopicity, which results in less corrosion of fuel-storage vessels. Moreover, it can be blended with gasoline at any concentrations and used with existing pipeline and engine infrastructures (Pasztor et al., 2015).

Butanol can be naturally produced by distinct routes present in Clostridium species of gram-positive bacteria (Lan and Liao, 2012), however, it is currently commercially produced from propylene (which is derived from petroleum) because the traditional acetone–butanol–ethanol (ABE) fermentation using *Clostridium acetobutylicum* became economically infeasible in the 1960s (Pasztor et al., 2015). Nonetheless, various novel microbial routes towards the production of butanol have now been reported (Dellomonaco et al., 2011, Menon et al., 2015, Pasztor et al., 2015) most of which are in any event derived from the native clostridial butanol-producing pathways reported in Lan and Liao (2012). Butanol is particularly interesting because in each case, the pathways can be extended to produce 1-hexanol and other higher alcohols. The CoA-dependent butanol pathway from clostridia can also be hijacked to produce other attractive chemicals such as propane, making this core pathway a very appealing biochemical route (Menon et al., 2015).
What was of particular interest about this compound in relation to this study is that the direct synthesis of n-butanol from glucose can cause a NADH/NAD^+ redox imbalance because more NADH is required in the synthetic pathway than is generated by glycolysis. The production titer can be improved by manipulating the reactions such as pyruvate dehydrogenase (PDH) in glycolysis to increase the NADH availability in the cell, or even increasing flux through PPP to produce surplus NADPH which can then be converted into NADH by enhancing transhydrogenase activity (Saini et al., 2016). Furthermore, as I will explore shortly, not only is butanol production redox limited, but it can also become ATP-limited, depending on the biosynthetic pathway used. These observations make butanol a very interesting chemical for these purposes: (i) there is a considerable number of pathway variants that can be studied systematically, and (ii) these pathways consume cofactors at different levels, making butanol an ideal starting point to systematically evaluate the impact of varying cofactor usage in the process of pathway selection.

4.1.2 Chapter overview

Using 8 different pathway variations for the production of butanol and butanol precursors, the relationship between varying demands for ATP and NAD(P)H and butanol production were evaluated using two different methods.
First, I used the energetic and stoichiometric calculations developed by Dugar and Stephanopoulos (2011), which are pathway-specific. Described in more detail in Section 2.3.3, this method adjusts the maximal theoretical pathway yield based on ATP and NAD(P)H imbalances, thereby estimating the relative, maximal theoretical potential of the synthetic pathway in question. Dugar and Stephanopoulos (2011) used this method to predict the relative potential of a number of industrial chemicals, including the original clostridial route for butanol production. In the same paper, they acknowledged that genome-scale models (GSMs) may also be used for such calculations, but did not attempt to implement this, present preliminary results or a comparative analysis.
Therefore, the second method evaluated in chapter was to follow Dugar and Stephanopoulos (2011)'s suggestions and use the COBRA methods flux balance analysis (FBA) and flux variability analysis (FVA) to assess how biosynthetic pathways integrate with the wider metabolic network, using *E.coli* as the model

system.
aim of this chapter was to implement and compare two different methods for the *a priori* assessment of how cofactor usage limits microbial performance and target yields.
The ultimate goal was to determine whether COBRA methods like FBA and FVA could enhance our understanding as to why some pathways perform better than others from a cofactor usage perspective. Ultimately, the field would greatly benefit from a similar assessment of pathway potential to that by Dugar and Stephanopoulos (2011) but using using the more flexible, transferable and easier to implement computational framework Flux Balance analysis (FBA).

4.2 Results

4.2.1 Biosynthetic pathways for butanol and butanol precursor production

8 synthetic pathways for the production of butanol and butanol precursors were selected for this study due to their distinct energy and redox requirements (Figure 4.1). These pathways correspond to the same synthetic pathways as originally proposed and experimentally introduced by Menon et al. (2015) and Pasztor et al. (2015).
Seven out of 8 pathways comprise parallel variations of the Clostridial butanol pathway, differing from one another in the conversion of acetyl-CoA to acetoacetyl-CoA (atoB vs. NphT7 routes) and butyryl-CoA to butyraldehyde (AdhE2 vs. TPC7 routes), whilst one pathway relies on fatty acid biosynthesis. Interestingly, Pasztor et al. (2015) proposed the FAS route in 2015 with the initial idea to construct a synthetic pathway for the biosynthesis of propane using an acyl-ACP thioesterase with a specificity for butyryl-ACP and an oxygen-insensitive carboxylic acid reductase (CAR) (Kallio et al., 2014). They observed that butanol accumulated as an undesirable by-product and proposed the pathway as a platform to effectively produce butanol instead.

4.2.2 Implementation of Relative Potential Calculations

The calculations presented by Dugar and Stephanopoulos (2011) were used to estimate the relative synthetic pathway potential of all previously described butanol and butanol precursor production pathways (Figure 4.1). For simplicity, I will refer to these calculations as **relative potential calculations**, or RPCs.
First, Dugar and Stephanopoulos (2011) present an energetic calculation of maximum yield (Y^E) to assess the maximum amount of product that can be produced from the substrate, measured in moles of product per mol of substrate (equation 4.1).

$$Y^E = \frac{\gamma_S}{\gamma_P} \tag{4.1}$$

Where γ_S = degree of reductance of the substrate and γ_P = degree of reductance of the product. The degree of reductance is calculated by first determining the number of electrons that each atom in a molecule can

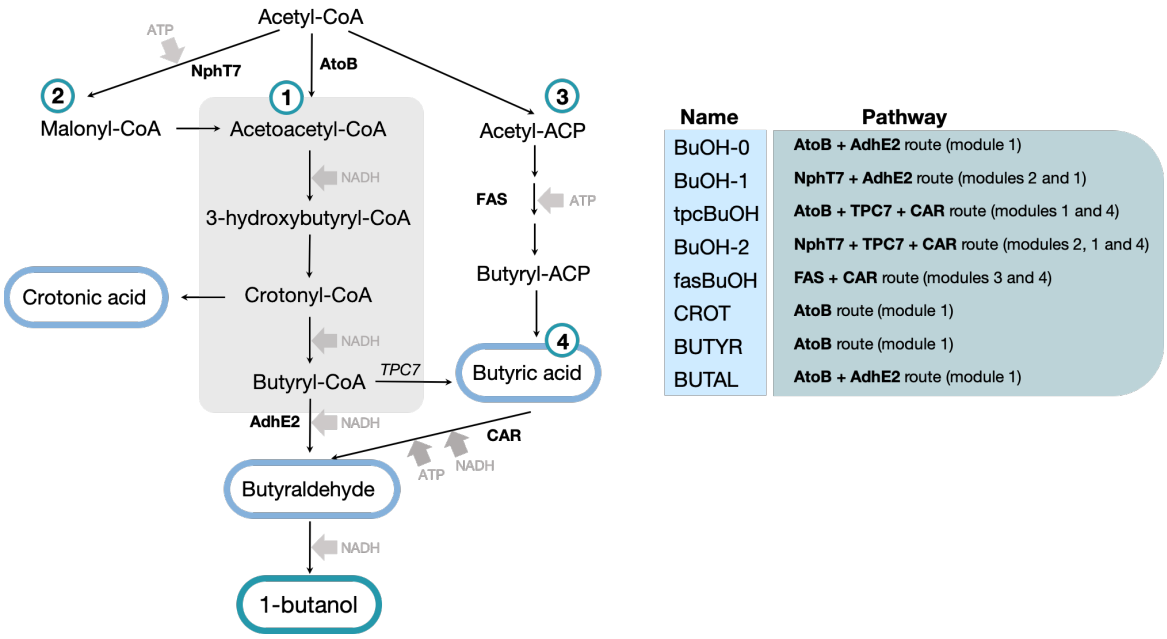


Figure 4.1: Engineered pathways used in this study and their cofactor requirements. 8 pathways that produce butanol (dark blue circle) and butanol precursors (light blue circles) were selected for this study. These pathways are based on variations of the so called "core pathway" (module 1, grey), which is redox dependent and ATP neutral. By combining modules 1 to 4, 8 unique pathway variations are possible with varying demands for energy and redox. Butanol pathways include: BuOH-0: module 1, including the AtoB and AdhE2 route; BuOH-1: modules 2 and 1, NphT7 and AdhE2 route; tpcBuOH: modules 1 and 3, AtoB and TPC7 route; BuOH-2: modules 2 and 4, NphT7 and TPC7 route; fasBuOH: modules 3 and 4, fas and CAR; Butanol precursor pathways include: CROT: module 1, AtoB route for the production of crotonic acid; BUTYR: AtoB route for the production of butyric acid; and BUTAL: AtoB route for butyraldehyde production. CP – Core Pathway; ACP – acyl carrier protein; AtoB – acetyl-CoA acetyltransferase; AdhE2 – aldehyde alcohol dehydrogenase; NphT7 – acetoacetyl-CoA synthase; TPC – acyl-ACP thioesterase; CAR – carboxylic acid reductase.

donate or take up to reach a full valence shell, and multiplying this value by the number of atoms of that type in the molecule. For example, the degree of reductance of glucose (C6H12O6) is γ_S = 4 x 6 + 1x12 -2x6 = 24.

Next, Dugar and Stephanopoulos (2011) use a system of linear equations v_1 to v_4 (biochemical equations 4.2-4.6), where stoichiometric coefficients a, b, c, d and e (NADPH requirement, product yield, ATP release, NADH production and CO_2 release, respectively), which are pathway specific, can be estimated for each catalyst. Equation 4.2 represents the stoichiometric pathway balance for product formation per carbon-mol of substrate. Equations 4.3 and 4.4 account for substrate consumption to produce NADPH (via PPP) and ATP (via oxidative fermentation). Equation 4.4 is only included in the calculations when ATP is required for product synthesis, i.e. when $c < 0$. Equation 4.5 accounts for the energy and redox requirement to

produce glycerol for cofactor regeneration, and equation 4.6 accounts for the additional ATP requirement to produce biomass. Here, a represents mol ATP required per mol carbon fixed as biomass, and ε represents the fraction of carbon lost to CO_2 in the process of biomass formation. From Dugar and Stephanopoulos (2011), it is assumed that $a = 2.42$ mol ATP/mol carbon in biomass, $\varepsilon = 0.1$; and x is calculated from Equation 4.7, where κ is the degree of reduction of biomass, $\kappa = 4.2$.

$$v_1 = -CH_2O - aNADPH + bProduct + cATP + dNADH + eCO_2 \tag{4.2}$$

$$v_2 = -CH_2O + 2NADPH + CO_2 \tag{4.3}$$

$$v_3 = -CH_2O + 4.82ATP + CO_2 \tag{4.4}$$

$$v_4 = -CH_2O - \frac{1}{3}ATP - \frac{1}{3}NADH + CH_{\frac{8}{3}}O(\text{glycerol}) \tag{4.5}$$

$$v_5 = -CH_2O - \frac{a}{(1+\varepsilon)}ATP + \frac{1}{(1+\varepsilon)}CH_{1.83}O_{0.56}N_{0.17}(\text{Biomass}) + \frac{x}{(1+\varepsilon)}NADH + \frac{a}{(1+\varepsilon)}NADH + CO_2 \tag{4.6}$$

$$x = \frac{4(1+\varepsilon-\kappa)}{2} \tag{4.7}$$

Assuming a carbon feedstock of glucose and glycolytic fermentation, these calculations consider a 1 mol glucose assimilation into 2 mol of acetyl-CoA, resulting in the release of 2 mol ATP, 2 CO_2 and 4 mol NADH prior to product formation, given the primary precursor for all pathways was acetyl-CoA. The available redox, a total 4 mol NADH, can then be directly used to produce 1 mol of butanol, or precursors thereof. After analyzing the pathways stoichiometrically from their initial building block, acetyl-CoA, until their final product, pathway specific coefficients a to e as mentioned above were calculated (Table 4.1) and normalized per glucose carbon atom (Table 4.2). It is important to note that pathway fasBuOH assumes that butyryl-ACP is produced from the preliminary steps of fatty acid synthesis from acetyl-CoA, consuming 1 additional NADPH and releasing an extra 2 mol CO_2 in the process.
Next, pathway yield (Y^P) is defined as:

$$Y^P = Y\frac{v_1}{(v_1 + v_2 + v_3)} \tag{4.8}$$

where Y is the product yield from equation 4.2. It can also be expressed as:

$$Y^P = Y\frac{1}{1 + \frac{a}{2} - (\frac{c}{4.82})|_{(\text{if c<0;else c=0})}} \tag{4.9}$$

Table 4.1: Pathway coefficients for NADPH demand, product release, ATP release, NADH release and CO_2 release of all butanol and butanol precursor pathways under investigation. Reaction-specific stoichiometric coefficients were calculated per reaction involved in the synthetic pathway, and the so called pathway coefficient retrieved as the sum across all reactions. Respiro-fermentative conditions assumed that a net release of 2 mol of acetyl-CoA yields 2 mol ATP, 2 CO_2 and 4 mol NADH per mol of carbon source assimilated (glucose, $C_6H_{12}O_6$) prior to product formation. This was accounted for in the below calculations. Pathways are labelled as per Section 4.2.1

	Pathway	Product Formula	a (NADPH)	b (product)	c (ATP)	d (NADH)	e (CO_2)
1	BuOH-0	Butanol ($C_4H_{10}O$)	0	1	2	0	2
2	BuOH-1	Butanol ($C_4H_{10}O$)	0	1	1	0	3
3	tpcBuOH	Butanol ($C_4H_{10}O$)	1	1	1	1	2
4	BuOH-2	Butanol ($C_4H_{10}O$)	1	1	0	1	3
5	fasBuOH	Butanol ($C_4H_{10}O$)	2	1	0	2	4
6	CROT	Crotonate ($C_4H_6O_2$)	0	1	2	3	2
7	BUTYR	Butyrate ($C_4H_8O_2$)	0	1	2	2	2
8	BUTAL	Butyraldehyde (C_4H_8O)	0	1	2	1	2

Where the condition is applied only when ATP is required by the pathway, i.e. when $c < 0$. Y^P was then adjusted to account for pathway inefficiencies. Firstly, assuming that cells thrive to be redox-neutral, any excess NAD(P)H is depleted using an electron sink (i.e. glycerol excretion) via equation 4.10, to yield $Y^{P,G}$.

$$Y^{P,G} = \frac{\nu_1}{(\nu_1 + \nu_2 + \nu_3 + \nu_4)} = Y\frac{1}{1 + \frac{a}{2} + (\frac{d-c}{4.82}|_{(\text{if d - c <0; else d - c=0})}) + 3d} \tag{4.10}$$

Next, $Y^{P,G}$ is then further adjusted after any excess ATP is diverted towards biomass formation, using equation 4.6 to produce $Y^{P,G,X}$.

$$Y^{P,G,X} = Y\frac{\nu_1}{\nu_1 + \nu_2 + \nu_3 + \nu_4 + \nu_5} = Y\frac{(a+x)}{(1+\frac{a}{2})(a+x) + c(3x+1+\varepsilon) + d(3a-(1+\varepsilon))} \tag{4.11}$$

Ultimately, Pathway efficiency (η) is calculated by doing the ratio of $Y^{P,G,X}$ and Y^E (equation 4.12).

$$\eta = \frac{Y^{P,G,X}}{Y^E} \tag{4.12}$$

All of the target pathways in this study to which the RPCs were applied comprise more than one chemical reaction, so these calculations were applied by calculating the net balance of ATP, redox and CO_2 for all carbon-carrying reactions from the original carbon source (glucose) to the end product.

Table 4.2: Normalized pathway coefficients a (NADPH), b (product), c (ATP), d (NADH) and e (CO_2) of all butanol and butanol precursor pathways. Stoichiometric coefficients were normalized per carbon-mol of glucose.

	Pathway	a (NADPH)	b (Product)	c (ATP)	d (NADH)	e (CO_2)
1	BuOH-0	0.000	0.167	0.333	0.000	0.333
2	BuOH-1	0.000	0.167	0.167	0.000	0.500
3	tpcBuOH	0.167	0.167	0.167	0.167	0.333
4	BuOH-2	0.167	0.167	0.000	0.167	0.500
5	fasBuOH	0.333	0.167	0.000	0.333	0.667
6	CROT	0.000	0.167	0.333	0.500	0.333
7	BUTYR	0.000	0.167	0.333	0.333	0.333
8	BUTAL	0.000	0.167	0.333	0.167	0.333

Table 4.3: Maximum yield (Y^E), pathway yield (Y^P), adjusted pathway yields ($Y^{P,G}$ and $Y^{P,G,X}$) and pathway efficiency (η) of all butanol and butanol precursor pathways..

	Pathway	Y^E	Y^{Ea}	Y^P	$Y^{P,G}$	$Y^{P,G,X}$	η
1	BuOH-0	1.000	0.411	1.000	1.000	0.884	0.884
2	BuOH-1	1.000	0.411	1.000	1.000	0.915	0.915
3	tpcBuOH	1.000	0.411	0.923	0.632	0.626	0.626
4	BuOH-2	1.000	0.411	0.923	0.618	0.665	0.665
5	fasBuOH	1.000	0.411	0.857	0.447	0.498	0.498
6	CROT	1.333	0.637	1.000	0.395	0.409	0.306
7	BUTYR	1.200	0.587	1.000	0.500	0.493	0.411
8	BUTAL	1.091	0.437	1.000	0.667	0.623	0.571

Maximum yield (Y^E), pathway yield (Y^P) and adjusted pathway yields $Y^{P,G}$ and ($Y^{P,G,X}$) are compiled in Table 4.3. The final yield value, $Y^{P,G,X}$ represents the highest possible yield achievable, assuming there are no competing pathways. Energetically and stoichiometrically balanced pathways are when $Y^E = Y^{P,G,X}$, but this was not the case for any of my selected pathways. This is because all engineered pathways included imbalances at either the redox or energy levels, if not both. Upon inspection, it was noticed that the less efficient pathways were those that released more CO_2 and reducing power. BuOH-0 and BuOH-1 were the most efficient pathways, as they were electron balanced and were only adjusted for ATP surplus. tpcBuOH, BuOH-2 and fasBuOH presented NADH surplus under the RPC assumptions, because they both included an NADPH-consuming step in order to produce the final target. As such, the need to supply NADPH into the system had to be accounted for. Across these models, it was noticed that when excess reducing equivalents were formed (i.e. NADH generated as a "product" of the reaction), a lower pathway yield was observed relative to the theoretical maximum. Butanol precursor pathways CROT, BUTYR and BUTAL had their yields penalised the most as these models are imbalanced both in terms of redox and ATP under the RPC framework.

4.2.3 Modified genome-scale models of *E.coli*

As an alternative to the RPCs, I then aimed to estimate the theoretical limits of butanol production by applying a flux balance based approach using the bacterium *E.coli* as a model system. It was expected that this approach would enable not only an exploration of how cofactor systems place limits on bioproduction, but also how pathway requirements are accommodated by the cell.

In order to enable butanol or butanol precursor production in *Escherichia coli*, the selected synthetic pathways were implemented into separate copies of the *E.coli* Core Model (Orth et al., 2010b), generating a total of 8 models with unique redox and energy features. I therefore analysed a total of 9 models: (1) iDAG85 (hereinafter known as BuOH-0), a butanol producer, includes the combination of reactions AtoB and AdhE2 along with the so-called Core Pathway (CP) shown in Figure 4.1. It comprises a total of 85 reactions and 70 metabolites and is ATP neutral; (2) iDAG87 (hereinafter BuOH-1) produces butanol via the ATP consuming reaction NphT7 (Lan and Liao, 2012). BuOH-1 includes 87 reactions and 72 metabolites; (3) iDAG86 (hereinafter tpcBuOH), a butanol-producing pathway that integrates enzymes AtoB and converts butyryl-CoA into butyraldehyde via a thioesterase and an ATP-dependent carboxylic acid reductase reaction (referred to as TPC7 in Menon et al. (2015)). This model is made up of 86 reactions and 71 metabolites; (4) iDAG88 (hereinafter BuOH-2), which incorporates reactions NphT7 and TPC7 and includes 88 reactions and 73 metabolites; (5) iDAG91 (hereinafter fasBuOH), an ACP-dependent butanol pathway that relies on Fatty Acid Synthesis (FAS), a thioesterase to release butyric acid and an ATP-dependent carboxylic acid reductase to generate the aldehyde. It is comprised of 91 reactions and 77 metabolites; (6) iDAG83 (hereinafter CROT), which produces crotonic acid via CP and is made up of 83 reactions and 68 metabolites; (7) iDAG84_butyr (hereinafter BUTYR), a butyrate producer via CP, which includes 84 reactions and 69 metabolites; (8) iDAG84_butal (hereinafter BUTAL), which yields

butyraldehyde via CP and is made up of 84 reactions and 69 metabolites; and finally (9) Wild Type, or WT, the version of the *E.coli* Core Model that excludes all reactions required for butanol production and fatty acid biosynthesis, containing a total of 77 reactions and 63 metabolites. Model features have been summarised in Table 4.4.

4.2.4 Butanol production potential shows lower efficiency upon higher ATP demand

Figure 4.2 shows the theoretical carbon yields of each pathway and ATP and NAD(P)H coefficients of each synthetic pathway. Pathway coefficients for ATP and NAD(P)H were calculated as the sum of each reaction stoichiometric coefficients, from acetyl-CoA through to the final target. For convenience, these stoichiometric coefficients are referred to as cofactor demand (negative values, indicating that the cofactor is consumed by the introduced pathway) and cofactor surplus (positive values, indicating cofactor production by the introduced pathway). For modelling purposes, I assumed that NADH and NADPH are

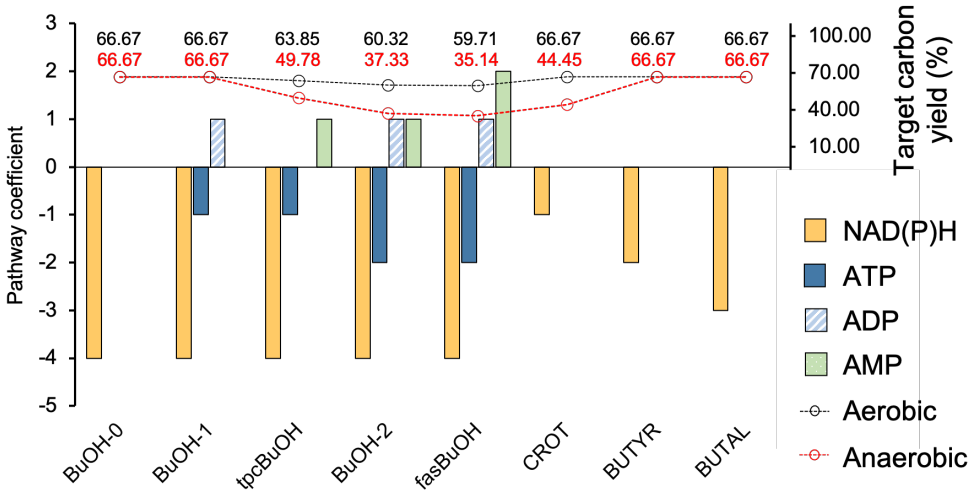


Figure 4.2: Cofactor requirements of the engineered pathways selected for this study. Cofactor requirements of all pathways introduced into the *E.coli* Core model to simulate butanol and butanol precursor production, and the aerobic (black) and anaerobic (red) carbon yields shown as a percentage of glucose carbon influx after target production maximization. Cofactor requirements are calculated as the sum of stoichiometric coefficients in all reactions starting from acetyl-CoA through to the final target molecule. Negative ATP/NAD(P)H coefficients represent cofactor demand, which refers to the consumption of a particular cofactor by the introduced pathway, i.e. ATP/NAD(P)H going into the reaction. Cofactor surplus, alternatively, is used to describe any cofactor being produced or released by a pathway. NAD(P)H surplus is indicated as positive NAD(P)H released by the pathway (subsequently from NAD(P) going into the reaction).

Table 4.4: Summary of key features of the modified *E.coli* models used in this study. I have indicated model names, final target, corresponding objective function (as per reaction ID), total number of model reactions and metabolites, degrees of freedom and also the ATP and NAD(P)H pathway coefficients, calculated as the sum of reaction stoichiometry coefficients of all introduced reactions from acetyl-CoA to the final target.

Model Name	Target	Objective Function	Metabolites	Reactions	Degrees of Freedom	ATP	NAD(P)H
WT	biomass	biomass	63	77	14		
BuOH-0	butanol	BTOH_sink	70	85	15	0	-4
BuOH-1	butanol	BTOH_sink	72	87	15	-1	-4
tpcBuOH	butanol	BTOH_sink	71	86	15	-1	-4
BuOH-2	butanol	BTOH_sink	73	88	15	-2	-4
fasBuOH	butanol	BTOH_sink	77	91	14	-2	-4
CROT	crotonic acid	CROAC_sink	68	83	15	0	-1
BUTYR	butyrate	BTAC_sink	69	84	15	0	-2
BUTAL	butyraldehyde	BTAL_sink	69	84	15	0	-3

interchangeable for modelling purposes, meaning if a hypothetical pathway consumes 1 NADPH and 1 NADH, the total redox produced is deemed to be -2 NAD(P)H, instead of each cofactor being considered separately. Under this assumption, all butanol pathways had the same redox demand (i.e. a net of 4 NAD(P)H required for butanol production) but varied in ATP demand, whilst the butanol precursor pathways had no ATP demand but instead vary in redox demand (Table 4.4).
Maximal product yield was obtained by selecting their corresponding sink reaction as the objective function and maximizing these reactions using parsimonious FBA (pFBA) (Lewis et al., 2010) (Section 3.1.3). For the wild type model, growth rate optimization was selected as the objective function. Under aerobic conditions, carbon yields ranged between 59.94-66.67%, not far from reported carbon yields in experimental studies (58.67% in (Shen et al., 2011)) and within the range of reported carbon yields for butanol production calculated using alternative methods (Dugar and Stephanopoulos, 2011) (Table 4.3). Under anaerobic conditions, the range increased to span 35.14-66.67%. It became noticeable, however, that the butanol models with highest ATP demands (tpcBuOH through to fasBuOH) had lower target production efficiencies. I suspected that these differences may stem from the need to utilise oxidative PPP to supply additional redox and the recycling of AMP and ADP, since these were not accounted for by the calculations presented in Dugar and Stephanopoulos (2011).

4.2.5 Engineered models present high flux variability in cofactor-related reactions

The solution space of all models was investigated by flux variability analysis (FVA) (Mahadevan and Schilling, 2003) (Appendix A Table 1). A single solution was found with the wild type and models tpcBuOH, BuOH-2 and fasBuOH. In contrast, the butanol producing models BuOH-0 and BuOH-1 had varying flux ranges in 17 out of 85 and 87 reactions, respectively, and butanol precursor producers CROT, BUTYR and BUTAL had varying flux ranges in 36 reactions, out of a total of 83, 84 and 84 reactions, respectively. This was not all that surprising, given that the models had been left mostly unconstrained, so some flux variability was expected. Notably, however, none of the reactions displaying multiple possible solutions were directly in the path towards butanol, i.e. the models displayed no flux variability in their target product pathways. Instead, 16 reactions common to these models were cofactor related reactions, and were often involved in futile cycling, as I will later continue to explore in Section 5.3.3 of this thesis.
From Figure 4.3, it was observed that CROT's solution had the highest flux variability ranges. CROT was also the model with the lowest cofactor demands (Figure 4.2). In contrast to this, BuOH-1 had the least amount of variation, and it was also the model with the highest cofactor demands (Figure 4.2). This seems to be suggesting that surplus cofactor is one such factor contributing to flux variability. I also noticed that the reactions that show the highest range variation were all ATP related (ADK1, ATPM, ATPS4r, FBP, PPCK, PPC, PPS, PYK) whereas redox-related reactions showed smaller ranges of variability. This was an interest observation, given that, for example, the biosynthetic pathway in CROT did not actually engage any ATP at all. Dugar and Stephanopoulos (2011) would classify this pathway as an ATP-neutral pathway,

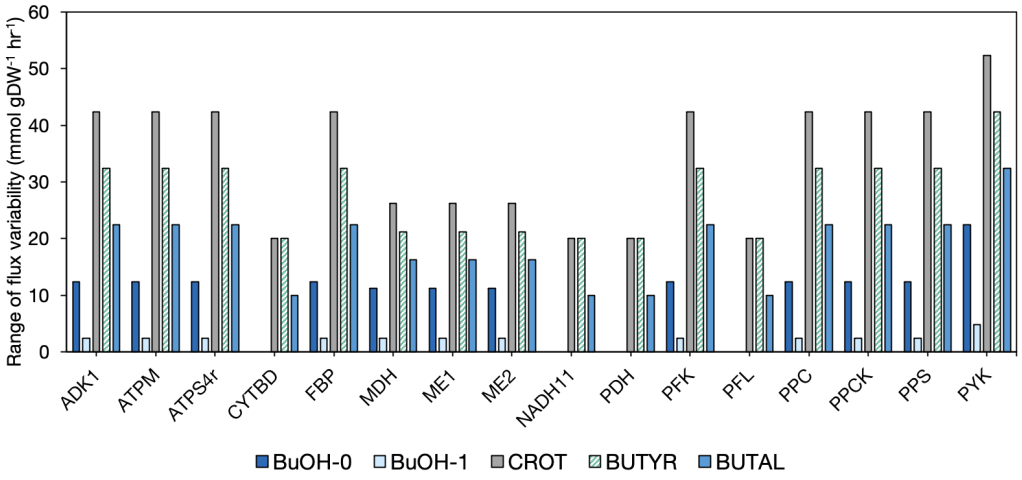


Figure 4.3: Ranges of flux variability of cofactor related reactions across models BuOH-0, BuOH-1, CROT, BUTYR and BUTAL. Flux Variability Analysis (FVA) was used to investigate the range of fluxes across cofactor related reactions, i.e. reactions that either consume or produce ATP and/or NAD(P)H across the models that showed non-unique solutions. Ranges were calculated by subtracting the minimum flux from the maximum flux recorded by FVA.

yet CROT showed a great deal of variation in ATP related reactions. This seems to be contradicting what has been captured experimentally in (Holm et al., 2010), which reported denser networks as a result of perturbations at the ATP level (i.e. ATP restoration had a more focused response) whereas redox perturbations seemed to invoke more widespread metabolic changes.
These observations hinted towards the connectivity between ATP and NAD(P)H balancing, and suggested that FBA could capture additional insights at the network level that the RPCs could not (Dugar and Stephanopoulos, 2011).

4.2.6 Butanol production capabilities of the *E.coli* metabolic network

To evaluate the metabolic responses to changes in cofactor demands during biosynthetic production in *E.coli*, flux distributions corresponding to the maximal possible butanol yield achievable were illustrated (Figure 4.4). These solutions represent the fluxes that result in the maximum carbon conversion towards the particular target from the initial carbon source.

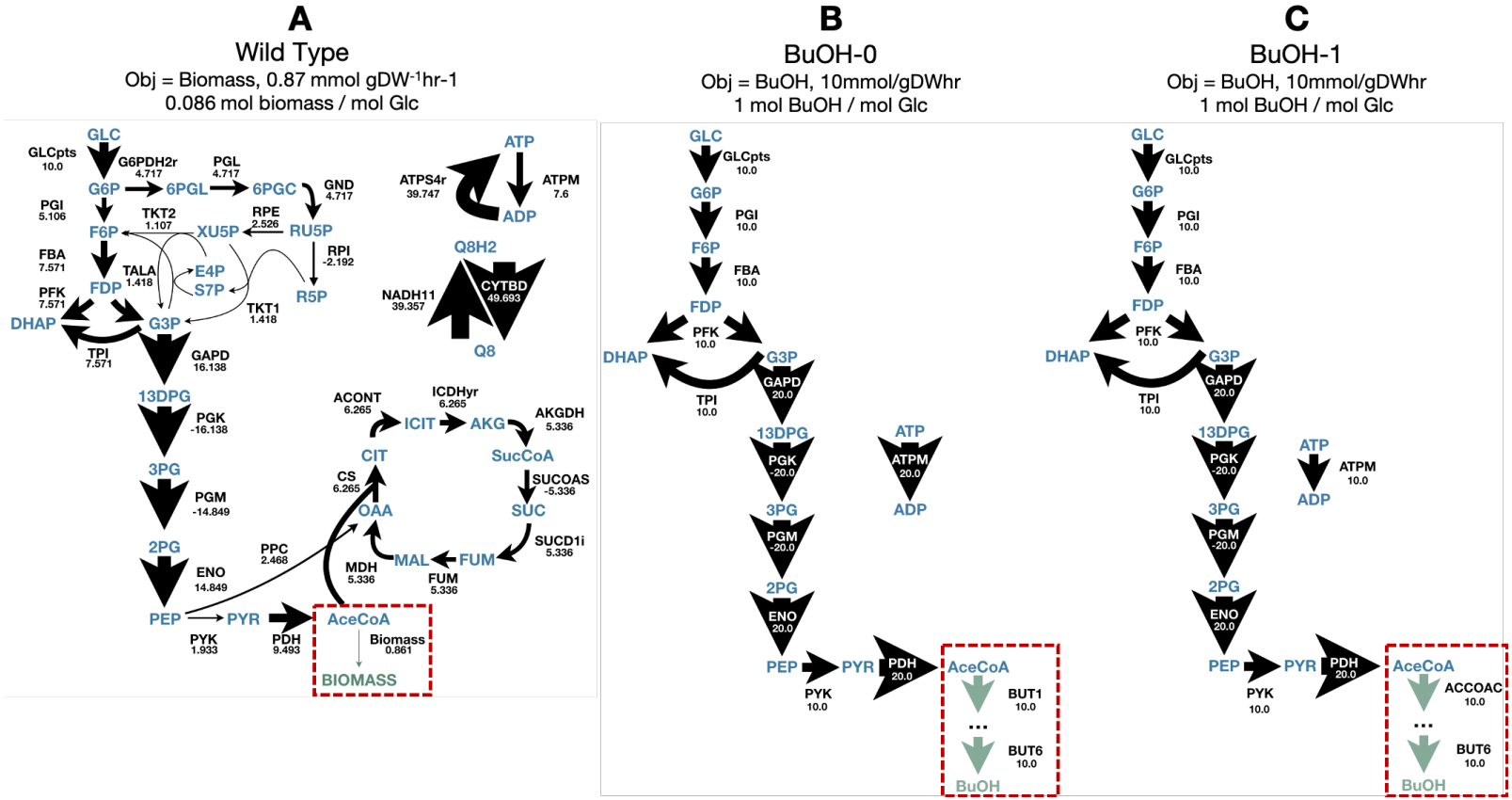
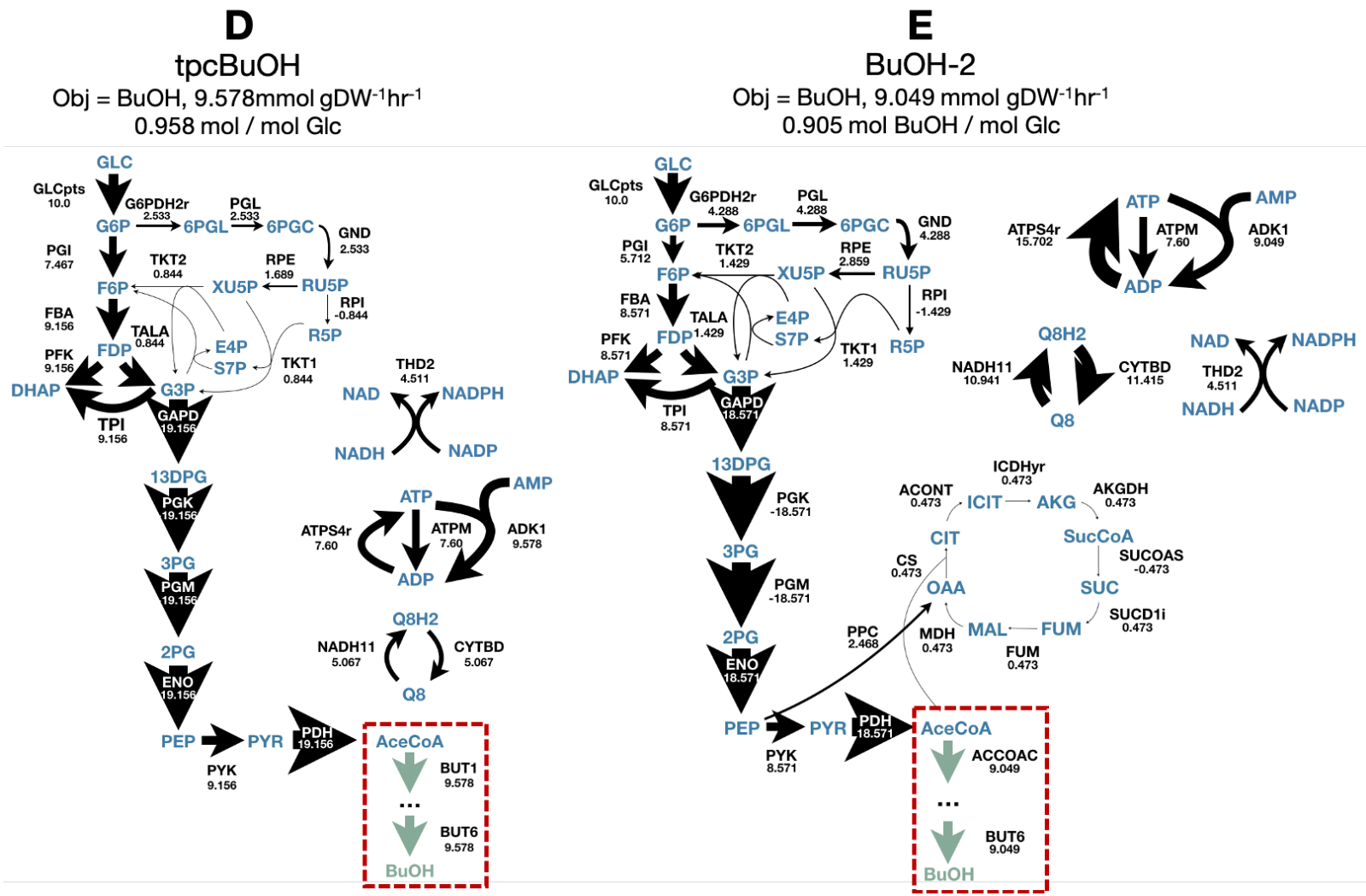
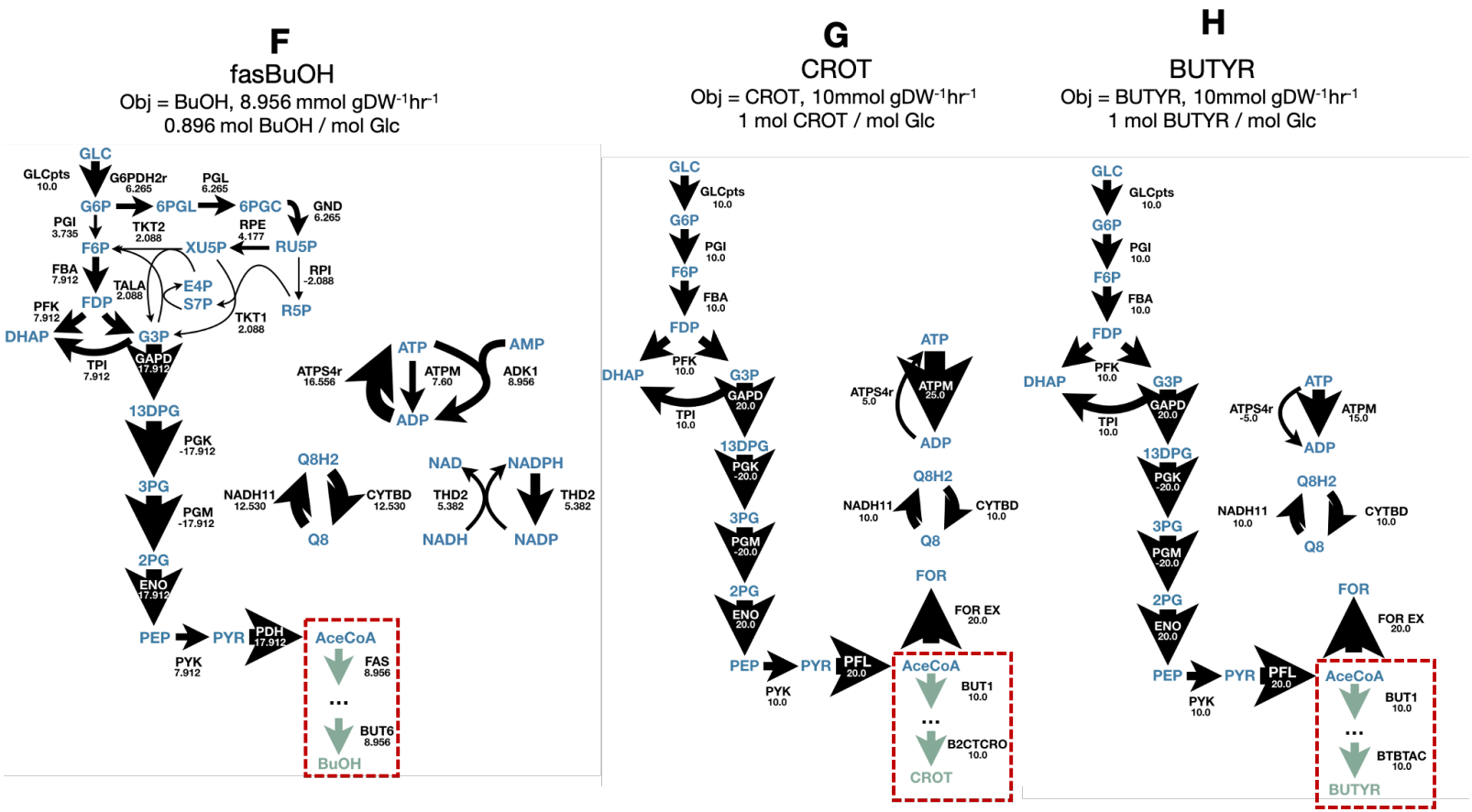


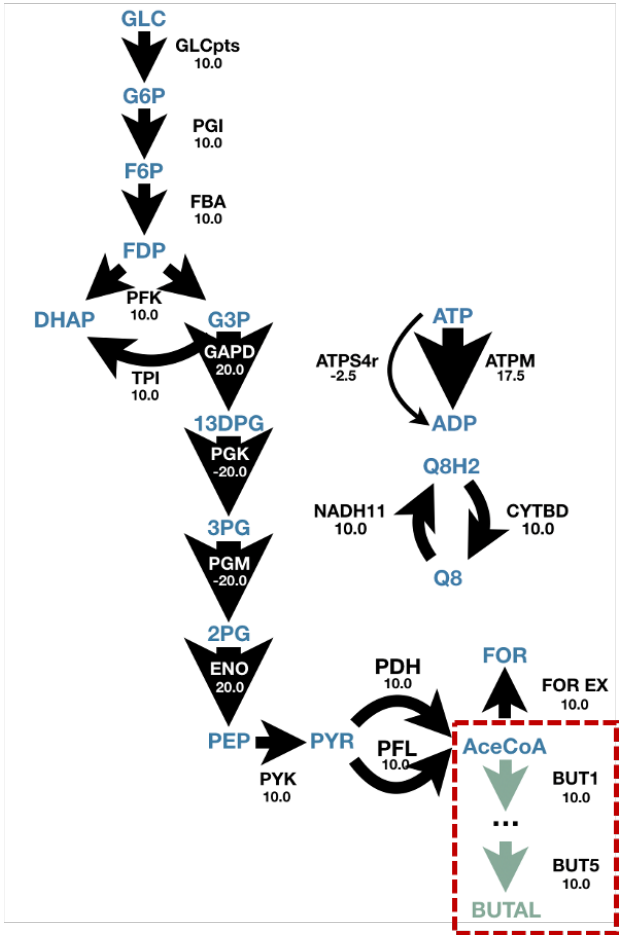
Figure 4.4: Flux maps of butanol and butanol precursor producers, using *Escherichia coli* as model system. Models were optimised using pFBA and the *Escherichia coli* Core Model, using either biomass formation or target production as the objective function, accordingly. Solutions were simulated under aerobic conditions (EX_o2_e_ = -10 mmol gDW^{-1} hr^{-1}) and were otherwise unconstrained


D
tpcBuOH
Obj = BuOH, 9.578mmol gDW-1hr-1
0.958 mol / mol Glc
GLC
GLCpts 10.0
G6P
G6PDH2r 2.533
6PGL
PGL 2.533
6PGC
GND 2.533
PGI 7.467
F6P
TKT2 0.844
XU5P
RPE 1.689
RU5P
RPI -0.844
R5P
FBA 9.156
TALA 0.844
E4P
FDP
S7P
TKT1 0.844
PFK 9.156
DHAP
G3P
TPI 9.156
GAPD 19.156
THD2 4.511
NAD
NADPH
NADH
NADP
13DPG
PGK 19.156
ATP
AMP
ATPS4r 7.60
ATPM 7.60
ADK1 9.578
ADP
3PG
PGM 19.156
Q8H2
NADH11 5.067
CYTBD 5.067
Q8
2PG
ENO 19.156
PEP
PYR
PDH 19.156
PYK 9.156
AceCoA
BUT1 9.578
...
BUT6 9.578
BuOH
E
BuOH-2
Obj = BuOH, 9.049 mmol gDW-1hr-1
0.905 mol BuOH / mol Glc
GLC
GLCpts 10.0
G6P
G6PDH2r 4.288
6PGL
PGL 4.288
6PGC
GND 4.288
PGI 5.712
F6P
TKT2 1.429
XU5P
RPE 2.859
RU5P
RPI -1.429
R5P
FBA 8.571
TALA 1.429
E4P
FDP
S7P
TKT1 1.429
PFK 8.571
DHAP
G3P
TPI 8.571
GAPD 18.571
ATP
AMP
ATPS4r 15.702
ATPM 7.60
ADK1 9.049
ADP
Q8H2
NADH11 10.941
CYTBD 11.415
Q8
NAD
NADPH
THD2 4.511
NADH
NADP
13DPG
PGK -18.571
ICDHyr 0.473
ACONT 0.473
ICIT
AKG
AKGDH 0.473
CIT
SucCoA
SUCOAS -0.473
CS 0.473
OAA
SUC
3PG
PGM -18.571
SUCD1i 0.473
PPC 2.468
MDH 0.473
MAL
FUM
FUM 0.473
2PG
ENO 18.571
PEP
PYR
PDH 18.571
PYK 8.571
AceCoA
ACCOAC 9.049
...
BUT6 9.049
BuOH


F
fasBuOH
Obj = BuOH, 8.956 mmol gDW-1hr-1
0.896 mol BuOH / mol Glc
G
CROT
Obj = CROT, 10mmol gDW-1hr-1
1 mol CROT / mol Glc
H
BUTYR
Obj = BUTYR, 10mmol gDW-1hr-1
1 mol BUTYR / mol Glc
GLC
GLCpts 10.0
G6P
G6PDH2r 6.265
PGL 6.265
6PGL
6PGC
GND 6.265
PGI 3.735
TKT2 2.088
RPE 4.177
F6P
XU5P
RU5P
FBA 7.912
TALA 2.088
E4P
S7P
RPI -2.088
R5P
PFK 7.912
FDP
TKT1 2.088
DHAP
G3P
TPI 7.912
GAPD 17.912
ATP
AMP
ATPS4r 16.556
ATPM 7.60
ADK1 8.956
ADP
13DPG
PGK -17.912
Q8H2
NADH11 12.530
CYTBD 12.530
Q8
NAD
NADPH
THD2 5.382
NADH
NADP
3PG
PGM -17.912
2PG
ENO 17.912
PEP
PYR
PYK 7.912
PDH 17.912
AceCoA
FAS 8.956
...
BUT6 8.956
BuOH
GLCpts 10.0
PGI 10.0
FBA 10.0
PFK 10.0
TPI 10.0
GAPD 20.0
PGK -20.0
PGM -20.0
ENO 20.0
PYK 10.0
ATPS4r 5.0
ATPM 25.0
NADH11 10.0
CYTBD 10.0
PFL 20.0
FOR
FOR EX 20.0
BUT1 10.0
B2CTCRO 10.0
CROT
ATPS4r -5.0
ATPM 15.0
BTBTAC 10.0
BUTYR

BUTAL

Obj = BUTAL_sink, 10mmol $gDW^{-1}hr^{-1}$
1 mol BUTAL / mol Glc



Upon a general inspection of the numerical results (Appendix B Tables 4 and 5), it became clear that all solutions relied heavily on glycolysis as the backbone carbon route between glucose (the original carbon source) and the final target. The redox demand of the butanol biosynthetic route was mainly satisfied by the interplay between glycolysis and the and the Pyruvate Dehydrogenase reaction (PDH). PDH not only decarboxylates pyruvate to produce acetyl-CoA, but also releases additional NADH, required for butanol production. Experimentally, butanol production was initially hindered by the limited NADH output from glucose catabolism, as producing 1 mol n-butanol from 1 mol glucose requires more NADH than that provided by glycolysis (Saini et al., 2016). To boost NADH availability experimentally and thus drive butanol production, increasing the activity of PDH has been used as a successful strategy (Lim et al., 2013, Saini et al., 2016, Shen et al., 2011). This is also validated by my *in silico* observations. PDH also contributes to the release of CO_2, which has a direct impact on the final carbon conversion (Varma et al., 1993).

It was also observed that, as the butanol producers (especially tpcBuOH, BuOH-2 and fasBuOH in Figures 4.4.C, 4.4D and 4.4.E) increased in energy demand, other areas of metabolism became engaged, particularly the Pentose Phosphate Pathway (PPP). Models tpcBuOH and and BuOH-2 both included an NADPH-consuming, AMP-producing step in order to produce the final target. As such, these systems needed to generate enough NADPH, supplied by PPP. PPP has two net consequences: (1) More NAD(P)H per glucose, likely the driving force to increase NADPH availability for the TPC7 route in these models, and (2) loss of carbon as CO_2.

These internal flux differences affecting the core carbon routes did not seem to have much impact on the final carbon yields (Figure 4.2). More generally however, differences in ATP were generally solved by burning excess ATP via reverse ATPS4r (ATP synthase) or reaction ATPM. Excess redox power was dealt with by either activating NADH11 (Electron Transport Chain, or ETC) and ATPS4r then burning the excess ATP, or by cycling between THD2 (NADP transhydrogenase) or NADTRHD (NAD transhydrogenase) to balance the NAD(P) redox pools. Surplus AMP was recycled by engaging ADK1. These reactions were also the same reactions that appeared to have wide flux variability ranges in Figure 4.3. The underdetermined nature of the systems (i.e. the possibility of many flux patterns to reach the same objective) was mainly driven by cofactor related reactions and was a clear issue that needed addressing.

Analogous simulations were also run under anaerobic conditions (Appendix B Table 5). In this new set of conditions, the engineered models used the introduced pathways as a major sink for electrons, aided however by the release of acetate in the case of tpcBuOH, BuOH-2 and fasBuOH in order to reach balance. Furthermore, unsurprisingly, no models had flux running through the biomass reaction. This is because, by selecting the butanol (or butanol precursor) sink reaction as the optimization criterium, FBA ignored any alternative routes that deviate resources away from the selected objective. Biomass production competes with the pathways for cellular resources because the biomass equation, in addition to the synthesis of many other metabolites, involves energy and redox, so producing any biomass at all would reduce the maximum yield of butanol, contradicting the assumption of optimal yield under pFBA. Under experimental conditions, balanced growth and product formation is essential in a bioprocess.

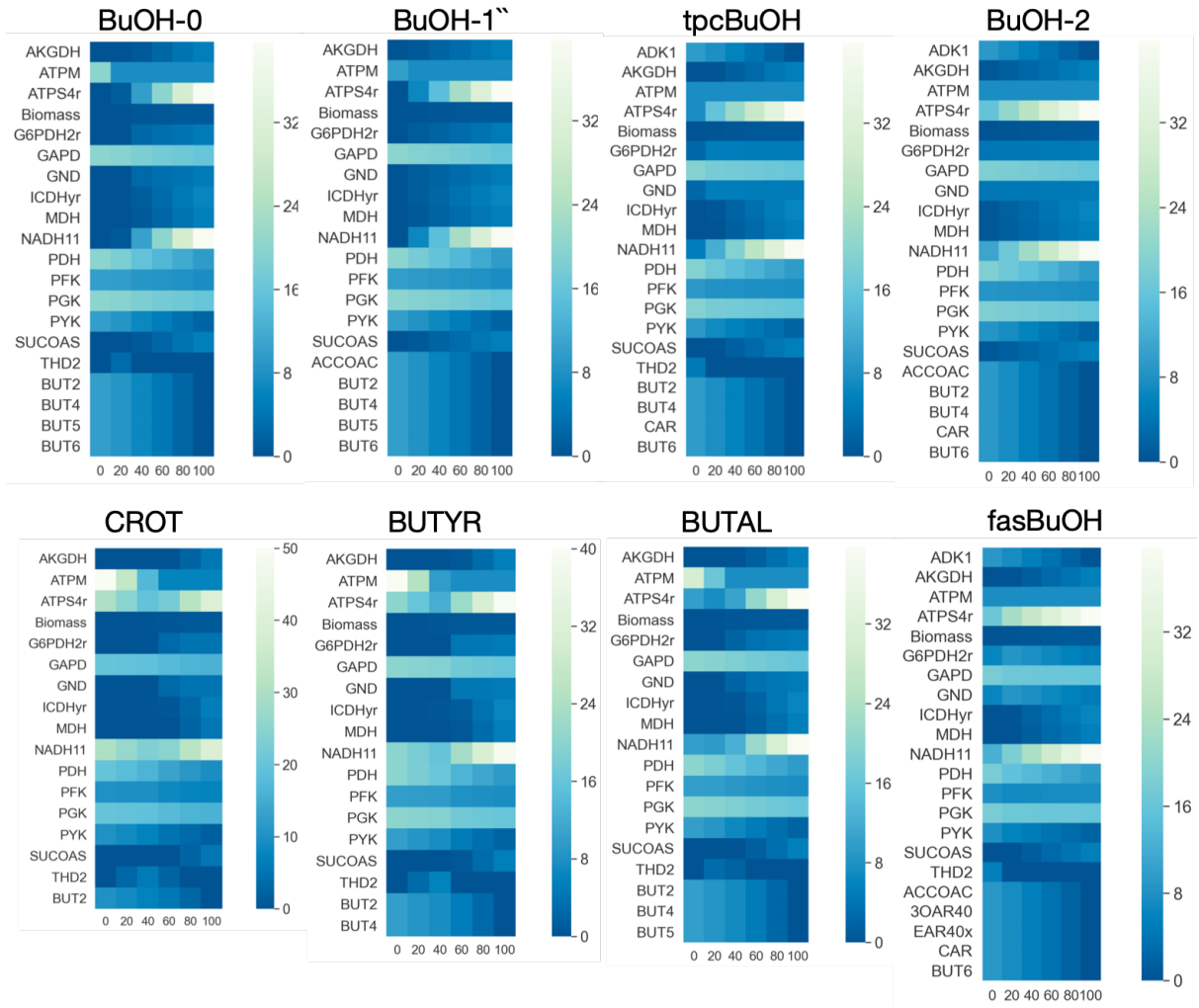


Figure 4.5: Flux maps of *Escherichia coli* CCM during butanol and butanol precursor production. Flux variation according to biomass production. The biomass reaction of all engineered models was fixed at 100%, 80%, 60%, 40%, 20% and 0% of the growth observed in the wild type. Models were then optimized for target production.

To address the industrially relevant trade-off between growth and product formation, I ran an additional test where the biomass reaction of the engineered strains was fixed to 100%, 80%, 60%, 40%, 20% and 0% of the WT biomass (Figure 4.5). As expected, butanol production was negatively correlated with the gradual increase in biomass formation, indicative of the trade-off between growth and bioproduction metabolism. Not surprisingly, there is no butanol production when biomass is fixed to 100% WT.

It was also noted how the fluxes across all ATP/NAD(P)H related reactions changed gradually with the

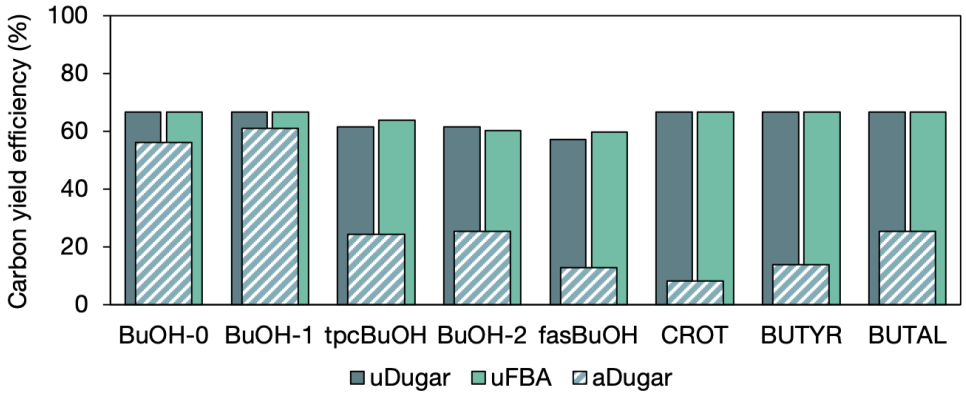


Figure 4.6: Carbon yield efficiencies calculated by FBA and by implementing the RPCs. Carbon yields of butanol and butanol precursor models were compared across the two approaches evaluated in this study: unconstrained pFBA (labelled "uFBA"), the initial estimates obtained by applying the RPCs from Dugar and Stephanopoulos (2011) (labelled unadjusted Dugar, or "uDugar") and the final estimates after adjusting the RPC for any energy and redox imbalances (labelled adjusted Dugar, or "aDugar").

gradual change in biomass formation. The results illustrate that there is an undeniable relationship between changes in biomass, cofactor metabolism, and ultimately target production.

4.2.7 Limitations of the existing methods

Figure 4.6 displays a carbon yield comparison between the butanol production estimates obtained from FBA (uFBA) and the unadjusted (uDugar) and adjusted (aDugar) pathway yield estimates after applying the RPCs from Dugar and Stephanopoulos (2011). The adjusted pathway yields were calculated after accounting for redox, CO_2 and ATP imbalances (shown as $Y^{P,G,X}$ in Section 4.2.2). Both methodologies report similar unadjusted theoretical yields for all models, and both methodologies agree on the best performing pathways. However, there are certainly a number of limitations to both methods.

The RPCs were proposed by Dugar and Stephanopoulos (2011) as an easy-to-use pathway selection tool to ensure high energy content of the final product as well as an efficient production route. They consider cofactor demand and any ATP and redox surplus generated by the target pathway(s) as potential limits on the maximal theoretical yield achievable. These parameters can be determined before pathway engineering commences and with minimal data, which is greatly beneficial if we are to minimise experimental testing. However, Dugar and Stephanopoulos (2011) allow for only one possibility to address each potential imbalance: excess ATP can only be resolved via biomass production, whilst excess NADH is consumed by a glycerol sink, which leads to quite a narrow method that only allows for testing a very restricted number of conditions. Furthermore, it is acknowledged in the paper that the yields are valid assuming glucose as the substrate, but numerous other assimilation pathways are available for other sugars that might lead to better yields for certain products, compared to glucose.

Unlike the theory suggested by Dugar and Stephanopoulos (2011), under the FBA framework, the butanol producing models with increasing ATP demands (tpcBuOH through to fasBuOH) also presented lower target production efficiencies. There are also discrepancies with carbon balance between the two methods. To illustrate this, using the RPCs, BuOH-1 and BuOH-2 both appeared to release excess CO_2, however FBA was able to account for the fact that HCO_3 is formed from CO_2 in these models so in fact the net balance is zero. FBA thus provides a more complete depiction of pathway potential and the limits they impose on the wider biological network. FBA yields the maximum amount of desirable product that can be converted from glucose, and this is calculated purely based on the assumption that the full conversion from all available glucose to product with no other products being formed can be achieved (Pasztor et al., 2015). This is similar to the calculation of pathway yield in Dugar and Stephanopoulos (2011), which looks at the maximum amount of product which can be made from the selected substrate. However, with FBA it is possible to can account for both the synthetic pathway(s) own cofactor requirements, as well as the metabolic network's overall need to balance its cofactor pools. The maximum theoretical efficiency will inherently be limited by the effective balancing of both the pathway and network-specific cofactor demands (Varma et al., 1993). Thus, the flux of cofactors and their rebalancing across different subsections of metabolism are more comprehensible captured by this method. However, FBA suffered from excessive flexibility due to the underdetermined nature of cofactor related reactions (Section 4.2.5). Related to this latter point, from the flux distribution vectors obtained by FBA I was unable to pin-point precisely where exactly in the network the cofactor imbalance is being incurred. Even accompanied by comprehensive visualization (e.g. Figures 4.4 or 4.5), it was challenging to note any inefficiencies in cofactor usage. It became evident that FBA alone would not suffice to make meaningful observations as to the effective use of cofactors in a metabolic network using COBRA methods.

4.3 Conclusion

In this chapter, maximal yields of butanol and butanol precursor production capabilities were assessed by pathway-specific energetic and stoichiometric calculations, Dugar and Stephanopoulos (2011)'s RPCs, as well as by incorporating butanol (or butanol precursor) pathways into the *E. coli* metabolic network and maximizing them using FBA. These values represent the maximal theoretical stoichiometric production capability of the metabolic network.

I show that the RPCs are built upon a set of strong assumptions, do not consider any biological or environmental cues and are not easy to extend to various pathways, carbon sources and products. This method is constrained to specific sets of conditions, and does not take into account other metabolic constraints such as pathway rigidity, competing pathways, feedback repression, kinetics, additional regulation or alternative environmental conditions, to name a few.

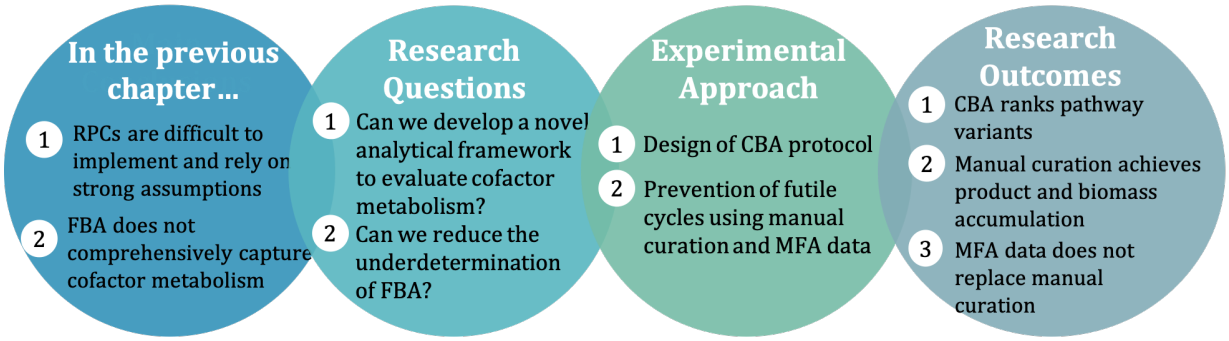
It is undeniable that FBA is a powerful alternative as it easily enables the assessment of any sort of environmental conditions, different kinds of media and carbon feedstocks, different growth and metabolic states. Despite this observation, it became clear that FBA alone did not suffice to discern cofactor usage

differences between systems. FBA also suffered from the excessive underdetermined nature of cofactor related reactions.

I thus proposed the design of a Cofactor Balance Assessment protocol (hereinafter known as CBA) to track cofactor production across the metabolic network as well as their contributions to biomass, waste, target production and metabolic maintenance. The ultimate goal was to use CBA to determine the optimal cofactor profile required for optimal yield efficiencies and facilitate understanding why some pathways perform better than others, from a cofactor usage perspective.

5

Development of a Cofactor Balance Assessment (CBA) Protocol



Notes from the Author:

Part of the work in this chapter has also been published in Paper I: **de Arroyo Garcia, L.**, Jones, P. R. (2020) "*in silico* co-factor balance estimation using constraint-based modelling informs metabolic engineering in *Escherichia coli*" . *PLoS Comput Biol* 16(8): e1008125. https://doi.org/10.1371/journal.pcbi.1008125 This work was also presented at the Data-Driven Biotechnology Conference, organized by Novo Nordisk Foundation in Hillerod, Denmark. I thank them greatly for sponsoring my attendance, which was instrumental to improve on my work.

5.1 Introduction

5.1.1 Is it possible to track cofactor balancing using FBA?

In Chapter 4, I concluded that, whilst the calculations developed by Dugar and Stephanopoulos (2011) are powerful and informative, they relied on stringent assumptions and did not easily extend to various pathways, products, or enabled the systematic assessment of varying environmental and biological cues. Although pFBA solutions elucidated butanol biosynthetic potential in *E.coli*, the source of cofactor imbalance was challenging to identify. Moreover, FVA solutions illustrated that cofactor related reactions presented non-unique fluxes that could equally satisfy the objectives, indicating the excessively underdetermined nature of the models.

Given that an understanding of cofactor balancing would very useful and informative to predict the yield superiority of synthetic pathways, but pFBA alone does not suffice, I asked, would it be possible to develop a data analysis protocol to gain insights into cofactor usage? And would it be possible to reduce the underdetermination of FBA?

Constraint-based tools developed to assist with cofactor manipulation have been reviewed in detail in Section 2.3.2 (Chapter 2). However, proposals so far have either focused on optimization of already experimentally-implemented systems or cofactor specifity swaps. The very few more comprehensive tools concentrate narrowly on NADH/NADPH balancing. The community still lacked a systematic framework that can inform the earlier stages of pathway and strain selection and which can pinpoint the sources of imbalance in particular systems. Predicting the optimal cofactor profile required for optimal yield efficiencies would facilitate the *a priori* assessment of cofactor metabolism across a selection of metabolic engineering designs before these are experimentally implemented, and provide a better understanding as to why some pathways perform better than others, from a cofactor usage perspective.

5.1.2 Chapter summary

This chapter focuses on the development of a Cofactor Balance Assessment protocol, hereinafter known as CBA, to track cofactor metabolism in metabolic engineering designs, using well-known modelling techniques, such as parsimonious FBA (Lewis et al., 2010), FVA (Mahadevan and Schilling, 2003), and MOMA (Segre et al., 2002).

Building on Chapter 4, I use the Core stoichiometric model of *E.coli* and the production of butanol and butanol precursors to evaluate how variations in ATP and redox demands contribute to yield efficiency. The study highlighted once again the underdetermined nature of some of the butanol and butanol precursor models, demonstrated by considerable dissipation of excess ATP and NAD(P)H in high-flux futile cycles. Although some futile cycling may take place naturally, I assumed that futile cycling would not vary as much as constraint–based modelling predicted due to internal regulation, insufficient enzyme quantities and/or thermodynamic constraints imposed by both the chemistry of each reaction and *in vivo* metabolite concentrations (Noor et al., 2014, Russell, 2007). Two methods to reduce the underdetermination of

these models were attempted. The first one, based on manual constraints, resulted in solutions with no apparent futile cycles and in the formation of biomass in 7 out of 8 engineered models, even when butanol production was set as the objective function for optimization. The second approach was based on the use of experimentally-derived constraints, but this approach did not remove futile cycling in its entirety. The CBA analysis helped explain why some pathways were predicted to achieve a higher yield than others.
Furthermore, the manually curated FBA solutions and the solutions obtained by implementing the RPCs in Chapter 4 reached similar theoretical yield values and agreed on the highest yielding pathway. However, they differed in the way cofactor imbalances are adjusted both at the ATP and NAD(P)H level. The CBA algorithm was further supplemented with a sensitivity analysis addressing the relationship between changing ATP and NAD(P)H demands and improvements in theoretical yield. This provided insights into the balance sweet spot and the degree of imbalance across pathway variants, as well as an additional validation framework to select the best-performing catalysts.

5.2 CBA protocol development

The CBA protocol was designed to provide insights into the energy spending of a modelled organism, by tracking the production of cytoplasmic cofactors ATP and NAD(P)H across the metabolic network and their contributions to biomass, waste, target production and metabolic maintenance (Figure 5.1). The ultimate goal was to use CBA to inform how a synthetic pathway's cofactor demands integrate with the host cell's own cofactor pools and influence yield efficiency.
CBA was written as a COBRApy-compatible Python function, so it can easily complement any COBRApy simulations run in the Python environment. Briefly, CBA sums the energy and redox synthesis fluxes of all reactions involving each of the two relevant cofactors, and splits it into four categories: (1) biomass production, (2) product production, (3) waste release and (4) cellular maintenance (Figure 5.2). For example, a strain of *E.coli* engineered to produce butanol diverts a particular amount of energy and redox to produce the chemical target, whilst the rest is distributed across reactions that lead to biomass formation, metabolic maintenance and waste release. Upon linear optimization, the CBA protocol determines the net flux through each category, and this information can be used to understand how effective an engineered system is at producing a chemical target, with respect to the resources being dissipated to achieve the optimal objective.
I detail below the method foundations, inputs, method parameters, functions and key assumptions.

5.2.1 Method foundations

The CBA protocol is built based on the following foundational concepts. Firstly, knowing that energy and redox cofactors can be either phosphorylated and reduced, or hydrolysed and oxidized, respectively, depending on the biological state of the cell, the first assumption I was able to make was that the flow of cofactors (i.e. whether they are produced or consumed) will depend not only on the stoichiometry of the

reaction involving such cofactor, but also on the reaction flux directionality (Figure 5.1).
Second, stoichiometric information can be gathered from stoichiometric models, which function as databases that specify all components involved in a particular reaction. As a result, they can be used to identify all reactions involving ATP and/or reducing equivalents in a particular metabolic network. Then, reaction fluxes can be determined using COBRApy methods, which estimate the flux distribution across the entire network. Because ATP, NADH and NADPH are highly connected cofactors and key players in carbon metabolism, by tracking these cofactors alone all major portions of central carbon metabolism are accessed.
Finally, cofactor usage can be classified according to whether cofactors produced or consumed in the process of generating biomass, producing a particular target, contributing to cellular maintenance or releasing waste, amongst other biological processes that are not considered within the scope of this study. With this, cofactor metabolism is divided into the following categories, represented graphically in Figure 5.1.B:

- **Cofactor production**, accounts for all reactions that generate a positive cofactor flux.
- **Biomass production**, any cofactor involved in the generation of additional biomass.
- **Target production**, any cofactor consumed or produced during target optimization. The target category is pathway-specific, and accounts for only those synthetic reactions introduced into the stoichiometric model. This category will be either net positive or negative according to whether the synthetic pathway produces or drains intracellular cofactors, respectively. For example, if the synthetic pathway is ATP-neutral, the net value for this category will be zero, whereas if the pathway requires the consumption of 20mmol ATP, then the net will be -20 (Figure 5.1).
- **Waste release**, accounts for any cofactor consumed or produced at the expense of waste release. This includes, at the ATP level, any ATP **produced** at the expense of acetate excretion, as well as any ATP burned in ATP-hydrolysing reactions, such as ATPM and ADK1. At the redox level, this category includes any reduced cofactors recycled through the release of fermentation by-products (such as in reactions LDH_D and AdhE2), or any oxidized cofactors reduced at the expense of CO_2 release (such as reactions GND, PDH, AKGDH and ICDHyr)
- **Cellular maintenance**, any cofactor consumed in any additional metabolic activities not considered in the above categories.

When an organism is engineered to produce a target chemical the flux dynamics of the metabolic network will change, and so will its cofactor metabolism. The above categories represent the "stress" placed on cellular metabolism during biochemical overproduction, which will vary accordingly depending on the redox and energy content of the original carbon source, its point of entry into the metabolic network, the nature of the final target, and the metabolic routes activated during biochemical overproduction. I hypothesised that, considering the categories described previously, unique cofactor profiles could be captured from whole-cell metabolic models by systematically tracking all cofactor related reactions. I specify the details of function development in Section 5.2.2.

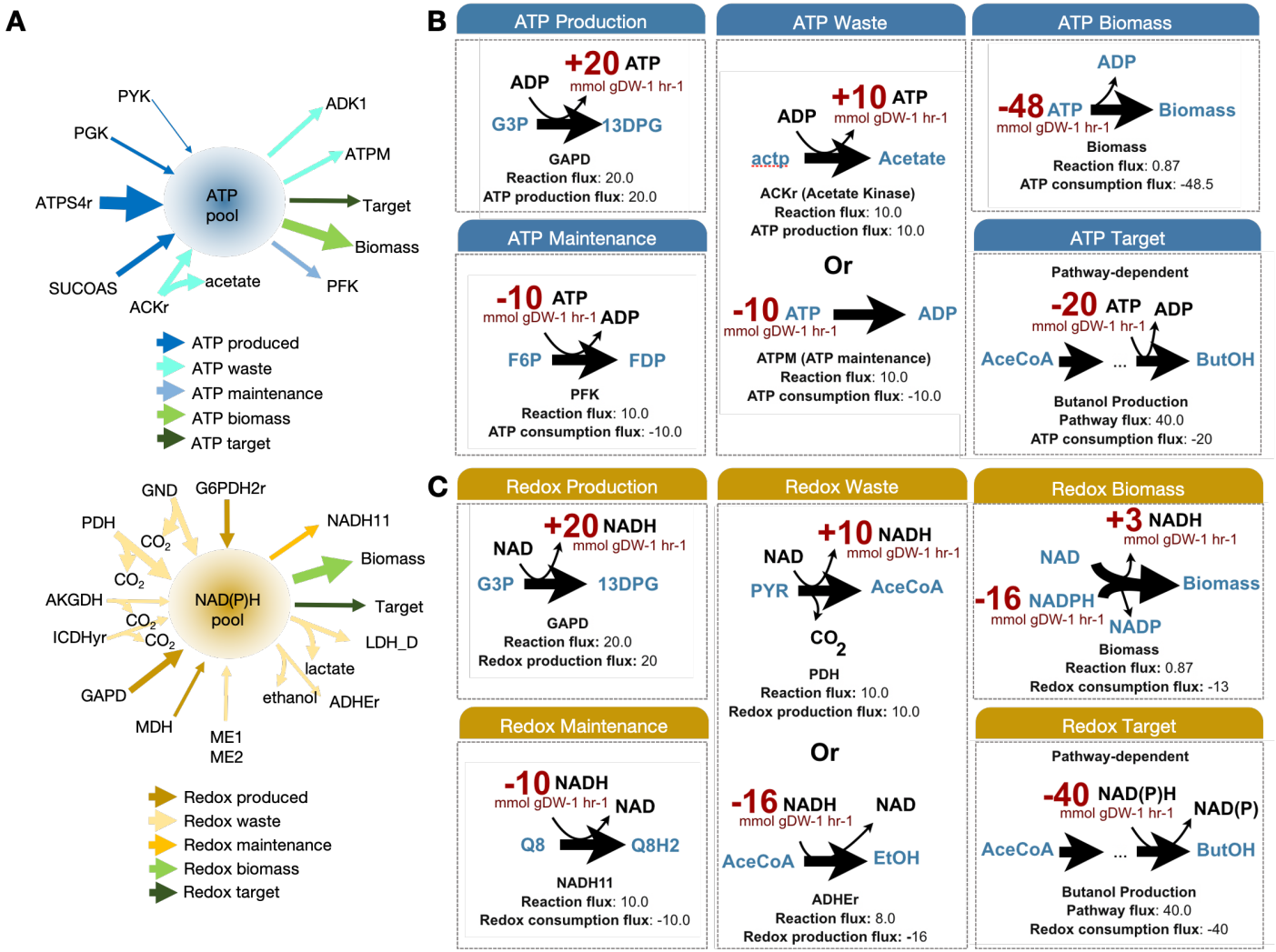


Figure 5.1: Toy illustration of ATP and NAD(P)H reactions and reaction categories accounted for by the CBA protocol. (A) All reactions in the *E.coli* Core Model that directly contribute to the intracellular levels of ATP and NAD(P)H pools (blue or yellow circle, accordingly) that are considered by CBA. Arrows pointing inwards on the left display reactions leading to ATP or NAD(P)H build-up (i.e. cofactor production), while arrows pointing outwards on the right show reactions that drain the cofactor pools (i.e. cofactor consumption). The thickness of the arrows represent the varying fluxes of these reactions. The CBA protocol distributes cofactor flux across five core categories: (1) cofactor production, (2) biomass production, (3) waste release, (4) cellular maintenance and (4) target production (this category is target product specific). (B) and (C) show a toy classification of ATP and NAD(P)H reactions, respectively, into the aforementioned categories by the CBA protocol. Cofactor fluxes are dependent on the cofactor stoichiometric coefficient and flux calculated by COBRA.

Figure 5.1.A shows how cofactor pools (shown as blue and yellow circles for ATP and NAD(P)H, respectively) are maintained by reactions that contribute to their production (arrows going into the circles representing cofactor production), while other reactions contribute to their drain or consumption (arrows pointing away from the circles). Thus, cofactor production accounts for all reactions that generate a positive cofactor flux. The waste category accounts for both any cofactor reduced during CO_2 or acetate excretion, as well as any cofactor oxidized during fermentation product release or hydrolysed in cofactor burning reactions. For instance, reactions such as GND or PDH reduce NAD^+ to produce produce NADH, but also release CO_2 in the process, which directly impacts the carbon conversion. This means that reducing equivalents can be produced (and thus have a positive value) regardless of the direction in which the carbon is flowing and shall be considered to be contributing to waste release in these circumstances. The biomass category includes any cofactor flux involved in biomass formation. The target category is pathway-specific, and accounts for only those synthetic reactions introduced into the stoichiometric model, and will lead to a positive or negative flux according to whether the synthetic pathway leads to the formation or drain of intracellular cofactors, respectively. E.g. If the synthetic pathway is ATP-neutral, the net value for this category will be zero. The maintenance category any cofactor consumed in additional metabolic activities and not considered in the aforementioned categories. For categories including both positive and negative cofactor fluxes, the net is calculated for that category.

5.2.2 CBA function development

The CBA protocol was built as a single, all-encompassing function (see Table 5.1 below) that can be called out to calculate the sum of all energy and redox synthesis fluxes, and subsequently categorise cofactor fluxes according to: (1) biomass production, (2) product production, (3) waste release and (4) cellular maintenance requirements.

CBA relies on the following inputs, all of which can be specified by the user: a stoichiometric model (normally a GSM), a matching flux distribution (obtained by FBA, pFBA, MOMA, or equivalent), and an array of so called "target reactions", in the event that the user is implementing synthetic reactions and wants to maximize for target production. Alternatively, the CBA function will accept an empty array.

First, the CBA function uses the stoichiometric model as a source of cofactor stoichiometry information. It uses a for loop to search for an identify all reactions involving ATP, NADH and NADPH, and stores the identified reactions and relevant cofactor stoichiometry (Table 5.1, 2-12). It then matches the identified reactions to their corresponding flux estimates to determine which of the identified reactions are carrying any flux at all, and then specify the level at which ATP or NAD(P)H is being either produced or consumed, known as cofactor flux score (CFS):

$$CFS_{i,j} = S_{i,j}v_j \tag{5.1}$$

where; $S_{i,j}$ = stoichiometry of cofactor i (ATP or NAD(P)H) in reaction j; and v_j = flux of reaction j (Table 5.1, 13-24).

CFS calculation is followed by the categorisation of reactions. This procedure is similar for both ATP and NAD(P)H so I will deal with them simultaneously. This is also the stage at which NADH and NAD(P)H are pooled to calculate the total redox flux in the system. To calculate the amount of cofactor flux involved in target production (i.e. either consumed or produced in the process), the CFSs of the "target reactions" specified by the user are summed to calculate the total cofactor involved (Table 5.1, 43-48 for ATP and 78-85 for NAD(P)H). The estimation of cofactor flux involved in biomass production is straightforward, i.e. it will correspond to the CFSs calculated for the biomass reaction. In the case of redox flux, the net CFS is calculated from the nadh and nadph CFSs.

In categorising cofactor waste flux, the procedure differed, because I wanted the CBA function to also account for cofactor-related reactions that released CO_2, in order to measure the amount of cofactor involved in diverting carbon away from target production. As such, the CBA function first identifies which of the ATP and NAD(P)H-related reactions also release CO_2. Once identified, their CFSs are calculated as previously described and stored (Table 5.1, 26-33 for ATP and 63-70 for NAD(P)H). Next, the CBA function also accounts for specific reactions known to either hydrolyse ATP (like ATP and ADK1) or release fermentation products (like ACKr for acetate production and LDH_D for lactate release) in this category. Once all CO_2-releasing reactions and waste reactions have had their CFSs calculated and stored, the total cofactor flux is calculated by summing up all CFSs.

Cofactor maintenance flux is calculated by adding up all negative CFSs (which indicate cofactor consumption), and adjusting the net value by subtracting the total cofactor flux involved in biomass, waste and target production if any of these categories have a net negative value (Table 5.1, 56-61 for ATP and 97-106 for NAD(P)H). In a similar manner, cofactor production is calculated by adding up all positive CFSs (which indicate cofactor generation), and adjusting the net value by subtracting the total cofactor flux involved in waste, target, and biomass release if these categories have a net positive value (Table 5.1, 50-55 for ATP and 87-96 for NAD(P)H). These net final scores consist of summed flux values that describe the overall "weight" of each category. In addition to Table 5.1, which outlines the technical procedure of how the CBA function was built, please refer to Figure 5.2 for a summary diagram of this procedure.

Table 5.1: CBA Function

```
1:  procedure CBA_calculation(model, solution, target reactions [array])
2:  Initialize dicts for storing reaction:cofactor stoich information for atp, nadh and nadph
3:  for reaction in model.reactions do
4:      for metabolite, stoich in reaction.metabolites do
5:          if metabolite is atp then
6:              append reaction:stoich pair into atp stoich dict
7:          if metabolite is nadh then
8:             append reaction:stoich pair into nadh stoich dict
9:          if metabolite is nadph then
10:             append reaction:stoich pair into nadph stoich dict
11:     end for
12: end for
13: Initialize new dicts for storing reaction:cofactor balance information for atp, nadh and nadph
14: for reaction, flux in solution.fluxes do
15:     if reaction is in atp stoich dict then
16:         CFS = flux x atp stoich of reaction
17:         append reaction:CFS pair into atp balance dict
18:     if reaction is in nadh stoich dict then
19:         CFS = flux x nadh stoich of reaction
20:         append reaction:CFS pair into nadh balance dict
21:     if reaction is in nadph stoich dict then
22:         CFS = flux x naph stoich of reaction
23:         append reaction:CFS pair into nadph balance dict
24: end for
```

atp waste flux calculation

```
25: initialize atp waste dict
26: for reaction in model.reactions do
27:     for metabolite, stoich in reaction.metabolites do
28:         if metabolite == CO2 and atp is in reaction.metabolites then
29:             temporarily store atp stoich using temp_stoich = stoich
30:             CFS = flux x temp_stoich
31:             append reaction:CFS pair into atp waste dict if CFS >0
32:     end for
33: end for
34: for reaction, CFS in atp balance dict do
35:     if reaction is ATPM then (ATPM hydrolyses atp into adp and pi)
36:         append reaction:CFS pair into atp waste dict
37:     if reaction is ADK1 then (ADK1 hydrolyses atp and uses amp to yield 2 adp)
38:         append reaction:CFS pair into atp waste dict
39:     if reaction is ACKr then (ACKr produces atp and releases acetate)
40:         append reaction:CFS pair into atp waste dict
41: end for
42: sum atp waste dict
```

atp target flux calculation

initialize atp target dict
for reaction in target reactions [user-inputted array] **do**
    if reaction is in atp balance dict **then**
        append reaction:CFS pair into atp target dict
end for
sum total atp target dict

atp biomass flux calculation

total atp biomass = CFS of biomass reaction in atp balance dict

atp production flux calculation

initialize atp production dict
for reaction, CFS in atp balance dict **do**
    if CFS >0 **then**
        append reaction:CFS pair into atp production dict
end for
sum atp production dict, **less** total atp waste **if** total atp waste >0, **less** total atp target **if** total atp target >0

atp maintenance flux calculation

initialize atp maintenance dict
for reaction, CFS in atp balance dict **do**
    if CFS <0 **then**
        append reaction:CFS pair into atp maintenance dict
end for
sum atp maintenance dict, **less** total atp biomass if total atp biomass <0, **less** total atp waste if total atp waste <0, **less** total atp target if total atp target <0

nad(p)h waste flux calculation

initialize redox waste dict
for reaction in model.reactions **do**
    for metabolite, stoich in reaction.metabolites **do**
        if metabolite == CO2 and nadh or nadph is in reaction.metabolites **then**
            temporarily store redox stoich of CO2 reaction using temp_stoich = stoich
            CFS = flux x temp_stoich of reaction
            append reaction:CFS pair into redox waste dict if CFS >0
    end for
end for
for reaction, CFS in nadh balance dict **do**
    if reaction is LDH_D **then** (LDH_D produces lactate at the expense of nadh oxidation)
        append reaction:CFS pair into redox waste dict
    if reaction is ADHEr **then** (ADHEr produces ethanol at the expense of nadh oxidation)
        append reaction:CFS pair into atp waste dict
end for
sum redox waste dict

nad(p)h target flux calculation

initialize redox target dict
for reaction in target reactions [user-inputted array] **do**
    if reaction is in nadh balance dict **then**
        append reaction:CFS pair into redox target dict
    if reaction is in nadph balance dict **do**
        append reaction:CFS pair into redox target dict
end for
sum total redox target dict

nad(p)h biomass flux calculation

total redox biomass = CFS of biomass reaction in nadh balance dict + CFS of biomass reaction in nadph balance dict

nad(p)h production flux calculation

initialize redox production dict
for reaction, CFS in nadh balance dict **do**
    if CFS >0 **then**
append reaction:CFS pair into redox production dict
end for
for reaction, CFS in nadph balance dict **do**
    if CFS >0 **do**
        append reaction:CFS pair into redox production dict
end for
sum redox production dict, **less** total redox waste **if** total redox waste >0, **less** total redox target **if** total atp target >0, **less** total redox biomass **if** total redox biomass >0

nad(p)h maintenance flux calculation

initialize redox maintenance dict
for reaction, CFS in nadh balance dict **do**
    if CFS <0 **then**
        append reaction:CFS pair into redox maintenance dict
end for
for reaction, CFS in nadph balance dict **do**
    if CFS <0 **then**
        append reaction:CFS pair into redox maintenance dict
end for
sum redox maintenance dict, **less** total redox biomass if total redox biomass <0, **less** total redox waste if total redox waste <0, **less** total redox target if total redox target <0

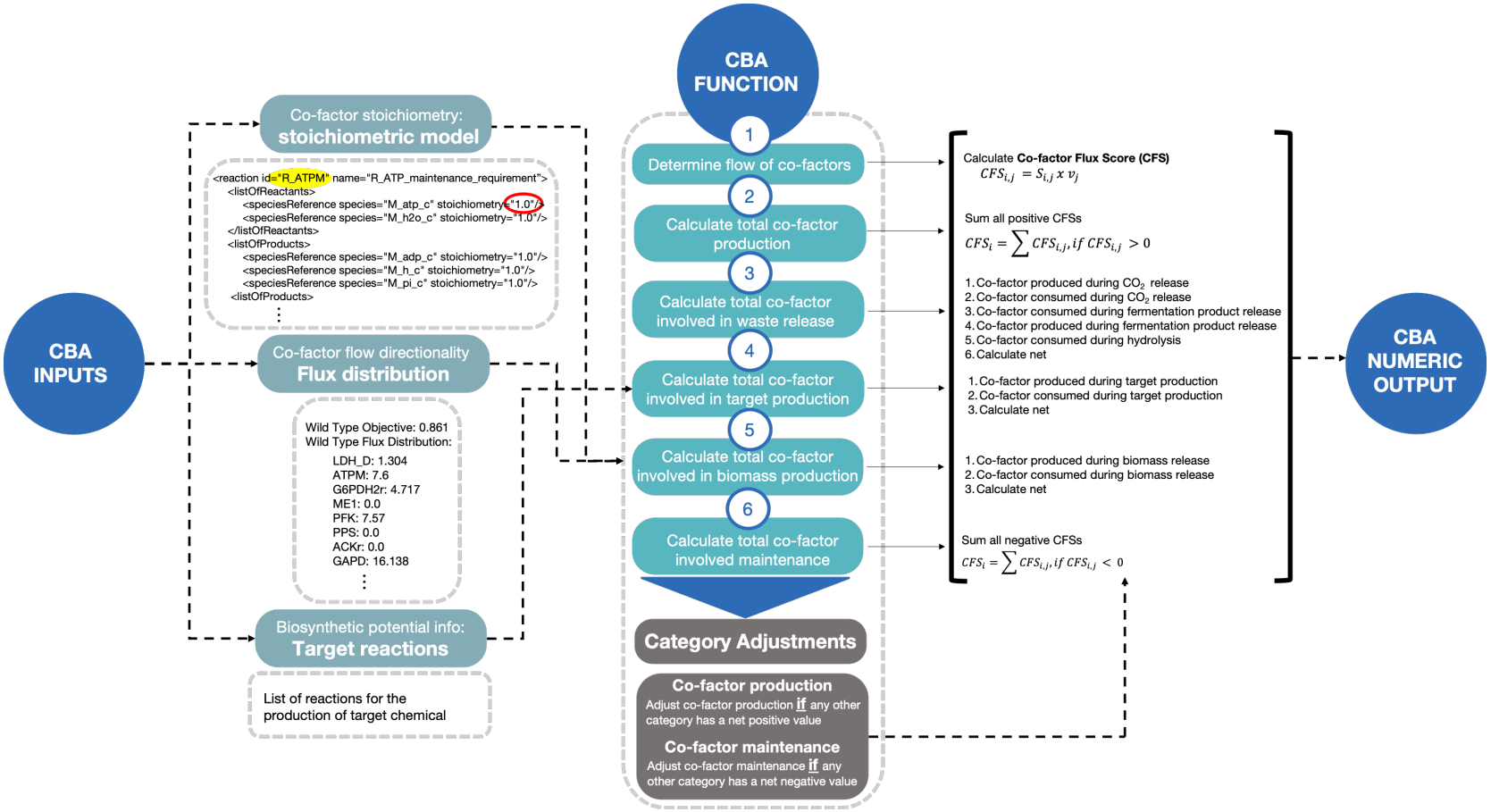


Figure 5.2: Components of the Cofactor Balance Assessment (CBA) pipeline and summarised workflow. Stoichiometric models contain reaction information, such as whether they consume or produce ATP and NAD(P). The *E.coli* Core stoichiometric model and the COBRApy package were used, and selected reactions were implemented to build the path to novel products. CBA classifies reactions in the model according to whether they are involved in the consumption or production of NAD(P)/ATP, assigns them a cofactor balance score, and groupins them into categories as represented above. Finally, the total balance per category is calculated the total sum of flux and adjusted to provide a final value for each category. The result is a profile displaying the fraction of the total cofactor produced involved in maintainenance, biomass, target and waste production.

5.2.3 CBA assumptions

The CBA protocol relies on the following assumptions:

- CBA solely tracks ATP and NAD(P)H metabolism, while all other cofactors are excluded from the analysis.
- For modelling purposes, it assumes that NADH and NADPH are interchangeable, even though in reality NADH and NADPH are not biologically equivalent
- It is assumed in this work that the above categories specified in Section 5.2.1 are the main categories contributing towards cofactor metabolism, whilst other cofactor assisted biological functions such as intracellular and extracellular transport, cell motility, cell division, stress, gene and protein expression, are excluded from analysis (or believed to be considered within the maintenance category, if at all).

5.3 Results

5.3.1 CBA produces distinct energy and redox profiles under aerobic and anaerobic conditions

The CBA function was tested on all eight models originally implemented in Section 4.2.3 (Chapter 4), which included butanol or butanol precursor production pathways (Figure 4.1) under both aerobic (Figure 5.3.A) and anaerobic (Figure 5.3.B) conditions, using the Core stoichiometric model of *E.coli* (Orth et al., 2010b). The selected biosynthetic pathways were suitable case studies due to their varying demands for ATP and redox carriers. First, flux distribution data optimised for target production was gathered as per Sections 3.1.1-3.1.3 under both aerobic and anaerobic conditions.

Under aerobic conditions, solutions for the engineered models displayed smaller magnitude fluxes for ATP synthesis and consumption than WT (Fig. 5.3.A), in line with a lower requirement for ATP by the product pathways given that the backbone route towards butanol production is ATP neutral. As well as ATP neutral, the pathway backbone, referred to as "Core Pathway" in this study (as shown in Figure 4.1), is also strongly dependent on four NAD(P)H-driven steps (Saini et al., 2016). All butanol models relied on glyceraldehyde-3-phosphate dehydrogenase and the pyruvate dehydrogenase reaction (PDH) for the supply of redox. The PDH reaction is known to provide the extra redox needed for butanol production (Lim et al., 2013, Saini et al., 2016, Shen et al., 2011). As illustrated in Figure 5.1, PDH is labelled as "waste", because NADH formation contributes to the loss of carbon through CO_2 release (hence the positive value observed for NAD(P)H waste under aerobic conditions (Figure 5.4.A, yellow). However, it simultaneously supplies the target pathway a key limiting factor, NADH, which is essential to optimize flux towards butanol production.

In the absence of O_2, however, the ATP production levels were similar for all models apart from BuOH-2 and fasBuOH, which also presented the lowest yields (Figure 5.4.A). This was an interesting observation,

especially given that the butanol producers differed specifically in ATP demands. Under both aerobic and anaerobic conditions, solutions for the engineered models showed no biomass accumulation. For BuOH-0, one of the highest yielding butanol models, more than half of the generated ATP went into the waste category.

Under the assumption that NADPH and NADH are interchangeable as explained in Section 4.2.4, all metabolic pathways with the same end-product will have the same net redox requirements unless there is a change in non-target products (e.g. fermentation products, or biomass). However, I observed considerable variation in NAD(P)H categories between models, both under aerobic and anaerobic conditions. Examining individual reactions (Appendix B Tables 4 and 5), it was noticed that the TPC route in models tpcBuOH, BuOH-2 and fasBuOH, which included a carboxylic acid reductase reaction that consumed 1 mol NADPH and produced 1 mol AMP from ATP, was causing the coupling of electron metabolism with energy metabolism. As a result, I noted (1) 18.8%, 18.2% and 23.3% of total NAD(P)H was produced by the PPP resulting in a higher yield of NADPH per glucose and (2) the activation of the ADK1 reaction to recycle AMP produced by TPC7. Even though the butanol pathways all have the same demand for electrons, they have differing requirements for ATP. Homeostatic adjustments to the different ATP requirements resulted in changes in metabolism influencing also NAD(P)H. Moreover, although the flux through PPP was lower under anaerobic conditions relative to aerobic for models BuOH-2 and fasBuOH, flux through PPP surprisingly increased for tpcBuOH under anaerobic conditions. Further differences in the CBA redox profiles may arise from the fact that flux may or may not be directed via cofactor-dependent routes, e.g. PFL, where electrons are channeled into H_2 or excreted formate under anaerobic conditions vs. PDH, where electrons are channelled back into NAD^+ (as per Appendix B Tables 4 and 5), a concept known as "degeneracy" or "genetic buffering", brought by identical reactions coded by different genes that constitute alternative yet functionally overlapping pathways (Stelling et al., 2004).

More generally, it was also observed that in order to cater to the increasing demands for ATP across the butanol pathways, the systems simply produced more net ATP, as depicted by the steady increase in ATP production along the x-axis (e.g. compare BuOH-2 with BuOH-0 on Figure 5.4A, blue). In contrast to these observations, the butanol precursor models (CROT, BUTYR and BUTAL), which did not demand ATP and only partly involved the Core Pathway, simply produced less ATP and also less NAD(P)H, decreasing the redox production by up to 2-fold. So, here I asked, is bacterial metabolism that flexible? i.e. is the range of flux solutions predicted by stoichiometric modelling greater than what is possible in reality? This question was explored in detail in the following three sections of this thesis, both through the use of manual curations (Section 5.3.3) and through the implementation of constrained informed by MFA-derived flux data (Section 5.3.4). Particularly in Section 5.3.4, where experimental data was used, it became evident that constraint-based modelling presents higher flux variability than what has so far been observed experimentally.

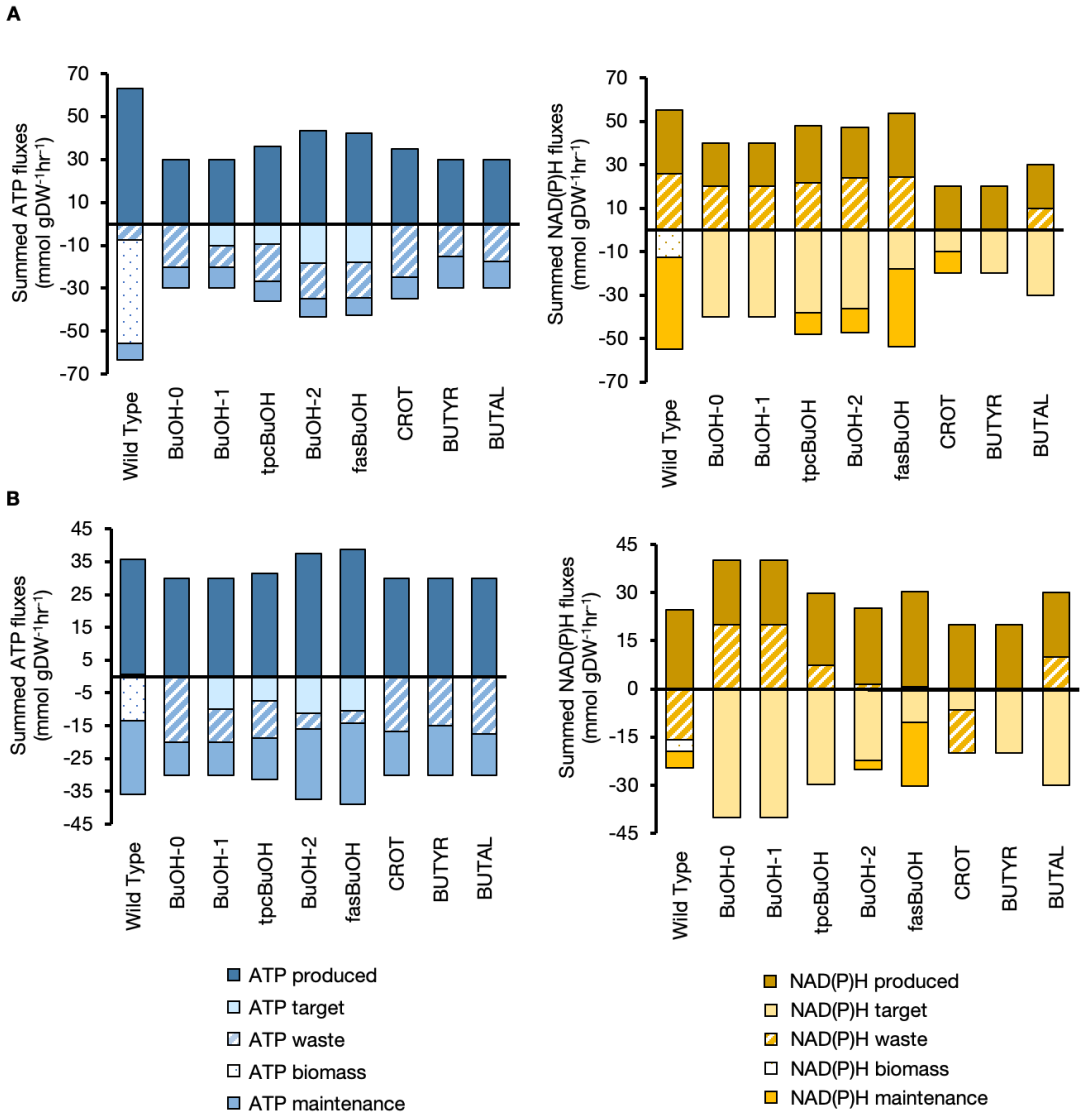


Figure 5.3: CBA-derived network cofactor usage profiles. After FBA optimization, the CBA protocol calculated the net ATP and NAD(P)H production and classified ATP and NAD(P)H-related reactions according to whether these cofactors were consumed or produced during biomass, waste, target production or cellular maintenance. All models were initially unconstrained and simulated under both aerobic and anaerobic conditions. (A) ATP and NAD(P)H profiles under aerobic conditions; (B) ATP and NAD(P)H profiles under anaerobic conditions.

5.3.2 CBA shows cofactor dissipation occurs as result of FBA's underdetermination

The general metabolic cost for non-growth-associated energy requirements, ATPM (reaction 2.4.1), is represented by an artificial reaction that breaks down ATP into ADP and Pi (equation 5.2):

$$\mathrm{ATP} \longrightarrow \mathrm{ADP} + \mathrm{Pi} \tag{5.2}$$

In contrast to the wild type (estimated at 7.6 mmol gDW^{-1} hr^{-1} by Varma and Palsson (1994)), all models displayed a flux increase through the ATPM or ADK1 reaction of up to 3-fold, ranging between 7.6-25 mmol gDW^{-1} hr^{-1} under aerobic conditions (Appendix B Table 4). Up to 71.4% of the total ATP produced was dissipated through these reactions alone, suggesting surplus energy in most models. The ATP neutrality of the Core Pathway (Figure 4.1) could be causing the net ATP excess in these systems, as ATP is generated by substrate-level phosphorylation during glycolysis in order to produce acetyl-CoA, the initial building block for target production. This results in an increased need to hydrolyse ATP in models including pathways with low or no ATP demand, such as BuOH-0 (where 66.7% of total ATP produced went into the waste category). As the ATP demand of the synthetic pathways increased, the fraction of ATP wasted also gradually dropped, (Figure 5.3.A, blue). The observation that artificially enforced ATP-hydrolysis can enhance product yield with engineered *E. coli* (Boecker et al., 2019) supports the idea that ATP availability influences the allocation between biomass and other carbon-products. Model BuOH-1, which included an ATP-consuming step to produce Malonyl-CoA from Acetyl-CoA in the butanol route, had 33.3% of the total ATP produced being wasted, with the remainder being consumed by the NphT7 enzyme, which catalyses this additional ATP-consuming step as the main route towards target optimization. Particularly for models with no ATP demands (BuOH-0, CROT, BUTYR, and BUTAL), 50-71.8% of the ATP went into waste. Even under anaerobic conditions, where most models produced similar ATP yield, the fraction of ATP wastage ranged between 13% to 66.6% (Figure 5.3.B, indicating that models with no or low ATP demands still dissipated surplus energy through ATP-burning reactions or cycles (described further in Section 5.3.3).

I also noticed that any redox imbalances in models tpcBuOH, BuOH-2 and fasBuOH were circumvented by the activity of NAD(P) transhydrogenase THD2 (reaction 2.4.2):

$$\mathrm{NADH} + \mathrm{NADP} + H^{+} \longrightarrow \mathrm{NA}D^{+} + \mathrm{NADPH} \tag{5.3}$$

The non-growth-associated dissipation of excess ATP, also referred to as "energy spilling" or ATP burning, has been proposed as a principle for cells to handle energy surplus (Russell and Cook, 1995), but the extent to which *E.coli* does so is less understood (Russell, 2007). The reversible nature of ATP synthase has also been suggested through the action of the rotational mechanism of the F1 subunit, but only under stress conditions (Rühle and Leister, 2015). In the case of redox balance, transhydrogenase activity is also known to be one of the various mechanisms to guarantee redox homeostasis (Fuhrer and Sauer, 2009). Some

fermentative bacteria can also alter their net ATP production as they change their end products (Russell and Cook, 1995). However, given the limited understanding of *E.coli*'s capabilities to dissipate surplus energy (Russell, 2007), and earlier reports suggesting the flexible nature of stoichiometric models including synthetic pathways that are cofactor imbalanced (Ghosh et al., 2011), these cofactor burning observations were suggestive of FBA having more flexibility than what would be expected in reality. In contrast, based on observations from fermentation studies, the analysis by Dugar and Stephanopoulos (2011) assumed the cell's ability to rebalance energy and redox through biomass and glycerol formation, respectively (Dugar and Stephanopoulos, 2011). Left unconstrained, FBA did not resort to such solutions in the core wild-type model.

5.3.3 Manual constraint of futile cycles leads to yield-efficient and biomass-viable solutions

During the assessment of the case studies, I asked whether constraining the underdetermination of FBA would result in more realistic flux distributions. The first possibility was to manually correct this flexibility, with the explicit assumption that the non-growth associated maintenance requirements (captured by the ATPM reaction in the *E.coli* core model) reported by Varma and Palsson (1994) already captured the natural ATP dissipation levels that can occur in the *E.coli* metabolic network. Consequently, given that the ATPM flux value observed in the wild type when optimized for biomass formation was 7.6mmol gDW^{-1} hr^{-1}, this value was treated as a "cut-off" for the ATPM reaction, so I constrained the ATPM reaction of the engineered models to a maximum flux of 7.6 mmol gDW^{-1} hr^{-1}.

When ATPM was constrained, the updated flux distributions (now ATPM-constrained) would instead divert the surplus energy through high-flux, cofactor spilling reaction pairs, also known as "futile cycles". Futile cycles are pairs of anabolic and catabolic reactions that act in an antagonistic fashion, consuming either ATP or NAD(P)H through one reaction whilst phosphorylating or reducing a particular reactant, and a complementary reaction that regenerates the initial metabolite to close the loop (Russell and Cook, 1995). Reactions like phosphoenolpyruvate carboxylase (PPC) and phosphoenolpyruvate carboxykinase (PCK) are known to combine experimentally to form a futile cycle that potentially dissipates ATP (Kim and Copley, 2007, Meza et al., 2012, Yang et al., 2003), but this is highly likely to be conditional, as observed by Yang and colleagues when varying the dilution rate (Yang et al., 2003), or lead to an increase in biomass yield due to higher ATP production rather than less ATP turnover (Chao and Liao, 1994*a*, Chao et al., 1993). Even when over-expressed, the potential antagonistic activity between pyruvate kinase (PYK) and phosphoenolpyruvate synthase in *E. coli* did not result in any significant futile cycle (Patnaik et al., 1992). It has now become apparent that futile cycles are tightly regulated to prevent energy waste (Russell, 2007, Russell and Cook, 1995). Consequently, unrealistic futile cycles were identified manually and systematically, by identifying new cyclic fluxes between reaction pairs (or more than 2 reactions in rare instances) that in all cases involved cofactor metabolism and appeared as a result of changes to the original, unmodified model. Examples of such identified high-flux, futile cycles are shown in Figure 5.4.A.

Table 5.2: Candidate reactions for manual curation

Reaction ID	Name	Gene	Models constrained
NAD(P)H metabolism			
THD2	NAD(P) transhydrogenase (periplasm)	*pnt*	tpcBuOH, BuOH-2, fasBuOH, CROT, BUTYR, BUTAL
ME1	malic enzyme (NAD)	*sfc*	BuOH-1, tpcBuOH, BuOH-2, fasBuOH, BUTYR
ME2	malic enzyme (NADP)	*mae*	BuOH-1, tpcBuOH, BuOH-2, fasBuOH, CROT, BUTYR, BUTAL
NADTRHD	solublre transhydrogenase	*sthA*	CROT
THD2	membrane-bound transhydrogenase	*pntAB*	tpcBuOH, BuOH-2, fasBuOH, CROT, BUTYR, BUTAL
SUCCt2b	succinate efflux via proton import	*dcuc*	tpcBuOH, BuOH-2, fasBuOH
PFL	pyruvate formate lyase	*pflc*	tpcBuOH, BuOH-2, fasBuOH
ATP metabolism			
ATPM	ATP maintenance reaction		BuOH-0, BuOH-1, tpcBuOH, BuOH-2, fasBuOH, CROT, BUTYR, BUTAL
FBP	fructose bi-phosphate aldolase	*fbp*	BuOH-0, BuOH-1, tpcBuOH, BuOH-2, fasBuOH, CROT, BUTYR, BUTAL
ADK1	aldenylate kinase	*adk*	CROT, BUTYR, BUTAL
PPC	phosphoenolpyruvate carboxylase	*ppc*	BuOH-0, BuOH-1, tpcBuOH, BuOH-2, fasBuOH, CROT, BUTYR, BUTAL
PPS	phosphoenolpyruvate synthase	*ppsa*	BuOH-0, BuOH-1, tpcBuOH, BuOH-2, fasBuOH
ATPS4r	ATP synthase	*atpIBEFHAGDC*	BuOH-0, BuOH-1, CROT, BUTYR, BUTAL

To decide whether such cycles were actually unrealistic futile cycles, no cut-off values were used. Instead, typically, these reaction pairs presented identical (or proportional) fluxes, and carried and obvious and significant amount flux, always exceeding 5 mmol gDW^{-1} hr^{-1}. These *in silico* cycles also involved the transhydrogenases THD2 and/or NADTRHD, and redox-driven reactions linking Glycolysis, PPP and the TCA cycles (Saini et al., 2016). In a stepwise manner, futile cycles were identified by directly comparing the flux distributions of the engineered models to that of the wild type (Figure 5.4.B). After the detection of a futile cycle, the non-cofactor-consuming reaction was capped by limiting its upper or lower bound according to the maximal flux value observed for the same reaction in the wild type (Figure 5.4.C). This also meant that if the corresponding reaction was inactive in the wild type, the flux of the same reaction in the engineered system would become zero. Like a whack-a-mole, with each constrained cycle appeared another. This iterative, manual curation was repeated, followed by optimization and flux distribution evaluation until no more futile cycles were observed (Figure 5.4.D). All reactions considered during manual constraining are included in Table 5.2.

Manually curated models without any apparent futile cycles (Figure 5.5), simulated to optimise target

production, led to single-flux solutions across all models both under aerobic and anaerobic conditions (Appendix A Tables 2 and 3) and resulted in solutions that channelled excess cofactors through the biomass equation. Under aerobic conditions (Figure 5.5.A, Appendix B Table 6), 7 out of 8 engineered models yielded solutions that led to both target and biomass accumulation. Under anaerobic conditions (Figure 5.5.B, Appendix B Table 7), three of the models yielded solutions that were fully balanced both in terms of ATP and NAD(P)H through the use of fermentation product release only, and without any biomass production. Here, biomass production appears not to be contradicting the optimisation principle, but instead placing an upper limit on the maximum yield achievable, which is in line with suggestions made by Dugar and Stephanopoulos (2011) that biomass is used as a sink for energy surplus to achieve cofactor balance, at the expense of product yield. From a biotechnological perspective, an engineered organism having no ATP or redox flux consumed in a biomass reaction would be considered the most balanced, as all the key resources (ATP and redox) are utilised by the target biosynthetic pathway. In this respect, BuOH-1 appears to have the most balanced pathway after applying manual constraints as it has the highest carbon yield (64.43%) and lowest biomass yield. In reality, some biomass accumulation is clearly essential in all strains in order to synthesize the biocatalyst in the first place.

tpcBuOH, BuOH-2 and fasBuOH, which included butanol pathways with the highest ATP demands (Table 4.4), displayed an AMP imbalance in the TPC7 pathway, which had an impact on yield, dropping their yield to 56.3%, 61.3% and 51.07%, respectively. In these cases, 21%, 22.6% and 16.6% of the total ATP was set aside for AMP recycling through ADK1 (Appendix A Table 1 and Appendix C Table 10), so the metabolic network flux distribution required further readjustment to produce enough ATP for target production and maintenance through CCM, further altering the redox balance along the way. These results illustrate that the more intertwined the imbalance of a synthetic pathway is, the more the host needs to put in a larger proportion of its energy budget trying to reach balance at the expense of product yield, as previously suggested by Weusthuis et al. (2010), who coined this concept as "incomplete oxidation". The methodology in Dugar and Stephanopoulos (2011), which states that ATP-neutral or requiring pathways are more efficient is only confined to pathway potential (i.e. it excludes a network-wide analysis) and does not account for the adjustment of additional cofactor imbalances throughout the network. In contrast, the CBA protocol illustrates that a higher ATP (or redox) demand by the synthetic pathways may not always translate into higher productivity, because the host network may need to accommodate the increased ATP demand by the synthetic pathway, with with subsequent knock-on effects. For example, an imbalance in ATP homeostasis is typically solved by changes in the flux of pathways that involve electron transfer (Varma et al., 1993, Weusthuis et al., 2010). In contrast to previous studies that have studied the behaviour of cofactors in isolation (Chen et al., 2011, Dugar and Stephanopoulos, 2011, Garcia Martin et al., (2015, Ghosh et al., 2011), the CBA protocol strongly suggests that ATP and NAD(P)H balancing cannot be assessed in isolation from each other, or even from the balance of additional cofactors, such as AMP and ADP.

The manually curated models were also compared against the RPC estimates calculated in Section 4.2.2 of Chapter 4, based on Dugar and Stephanopoulos (2011). Figure 5.5.C displays a comparison of the

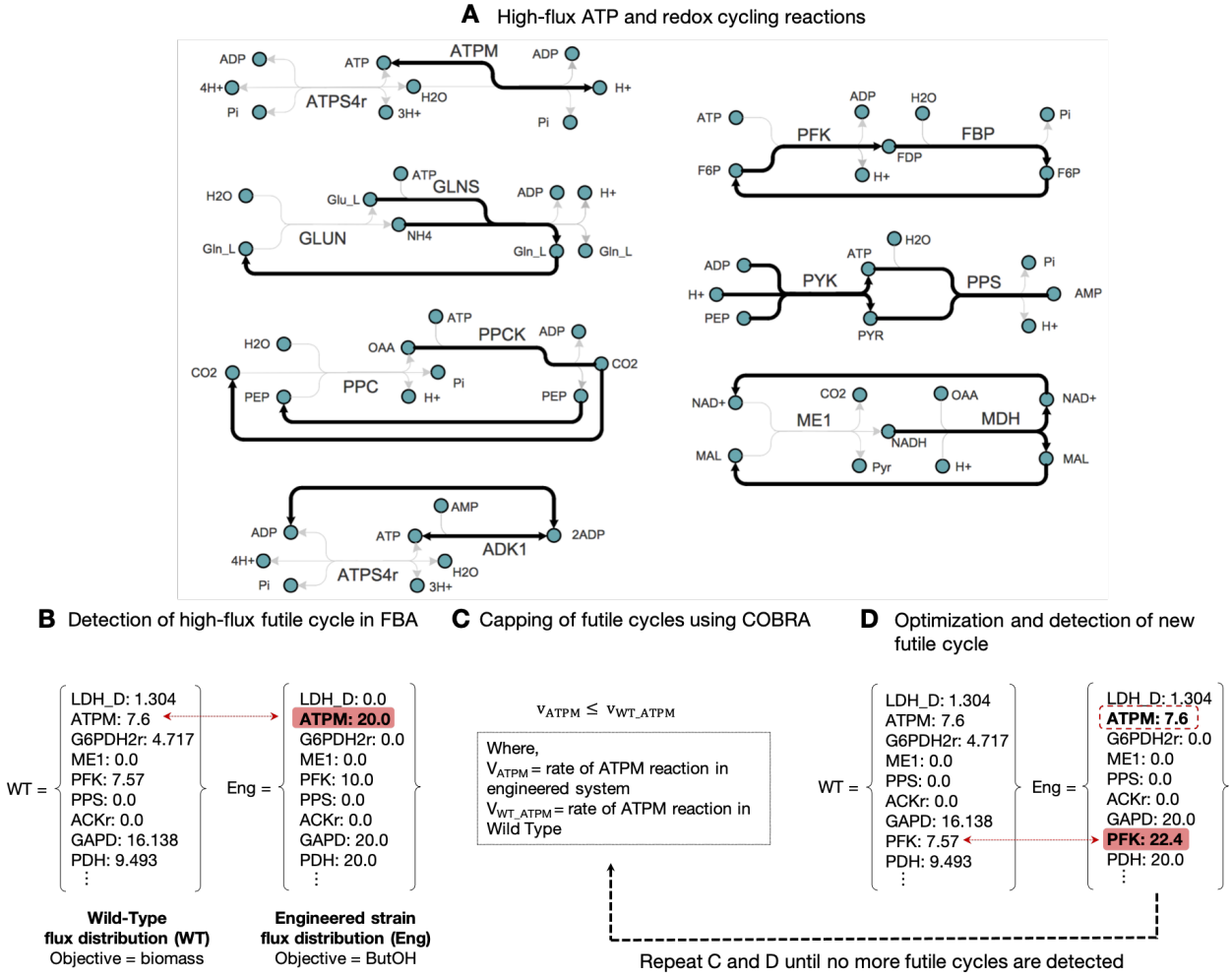


Figure 5.4: Identification and removal of futile cycles. (A) examples of cofactor futile cycles identified in this study - pairs of cycling reactions in which ATP is consumed through one reaction and the original metabolites are recycled through the pair reaction; (B) ATP-burning and high-flux futile cycles were identified by directly comparing the engineered strain and the wild-type flux distributions; (C) The identified ATP-burning reaction or futile cycle was constrained by limiting the upper bound to the maximal flux observed for the equivalent reaction in the wild type; (D) After optimization, the flux distributions of the wild type and engineered system were compared again and the next high-flux futile cycle was detected and constrained accordingly (as per C). Steps (C) and (D) were repeated until no more futile cycles were detected.

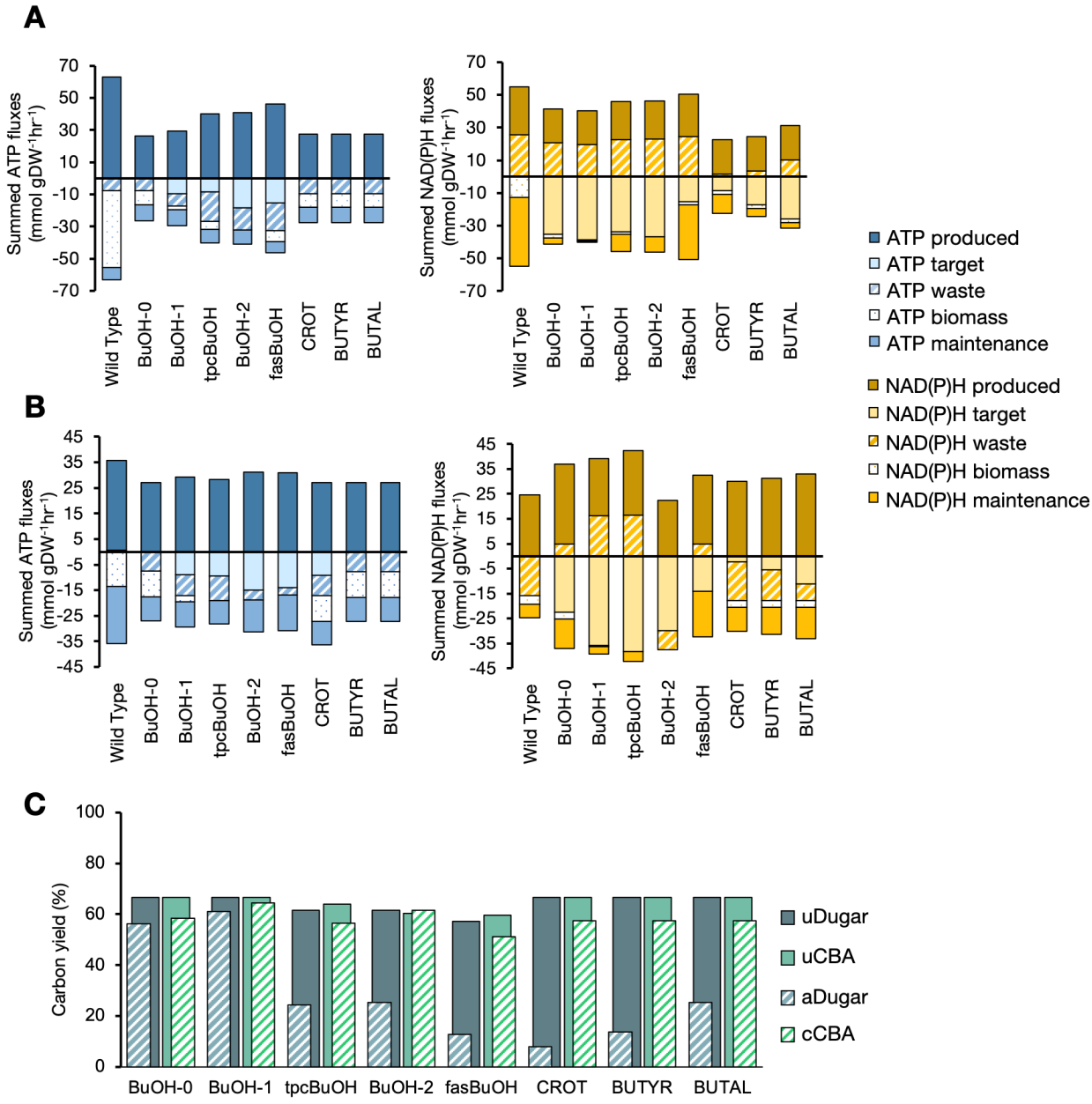


Figure 5.5: CBA-derived cofactor usage profiles after manual curation of the models to minimize futile cycling and carbon yield comparison. The engineered models were manually constrained to minimize high-flux ATP futile cycles as described in Section 5.3.3, and led to (A) curated ATP and NAD(P)H CBA profiles under aerobic conditions and (B) ATP and NAD(P)H CBA profiles under anaerobic conditions. (C) Carbon yields estimated by pFBA and using the RPCs. uDugar – unadjusted RPCs ; uFBA – unconstrained pFBA (prior to manual curation); aDugar – adjusted RPCs after accounting for ATP and NAD(P)H imbalances; cFBA – curated pFBA carbon yields estimates, obtained after manually constraining high-flux futile cycles.

carbon yields between the unadjusted pathway yield estimates, known as Y^P in Dugar and Stephanopoulos (2011), but here labelled as uDugar, and adjusted estimates after accounting for redox, CO_2 and ATP imbalances, known as $Y^{P,G,X}$ in Dugar and Stephanopoulos (2011) but here labelled aDugar, as well as the pFBA-derived estimates before and after manual curation of high-flux futile cycles. To remind the reader briefly, yield and efficiency estimates using the RPCs are pathway-specific and depend on the pathways' NADPH demand, product release, and ATP, NADH and CO_2 surplus coefficients between their initial building block (i.e. acetyl-CoA), and their final target. For further information on how these estimates are calculated, please refer to Section 4.2.2 of Chapter 4.

Manually curated solutions presented lower carbon efficiencies and more closely approximated the adjusted RPC estimates of BuOH-0 and BuOH-1. Both CBA and RPC frameworks agreed that BuOH-1 is the most yield efficient solution. In line with their theory (Dugar and Stephanopoulos, 2011), this solution includes a synthetic pathway that is both ATP-requiring and puts in the least amount of cofactor towards biomass production (Figure 5.5.A), yielding the highest butanol production while diverting the least amount of energy towards waste and/or futile cycles.

Obvious discrepancies still remain though. Under the RPC framework, tpcBuOH, BuOH-2 and fasBuOH, which are NAD(P)H-dependant, exhibited an NADH surplus of up to 2 mol, so the maximal theoretical yield is in fact penalized twice: it is first adjusted for redox and CO_2 imbalances, then adjusted once more to account for any ATP imbalance. Similar consequences arise for CROT, BUTYR, and BUTAL, which are both redox and energy imbalanced under this framework. This explains the considerable drop in carbon yield for the adjusted dugar estimates. Alternatively, with stoichiometric modelling, both redox and ATP imbalances can be systematically addressed, as both cofactors are needed for the biomass reaction to carry any flux at all. Furthermore, the need for these solutions to recycle AMP and address NADPH demand are handled by fine tuning the wider metabolic network to render both the pathway and the entire system at balance, so the impact on the final theoretical yield is lower.

I concluded that, although both methodologies reported similar unadjusted theoretical yields for all models, and both methodologies agreed on the best performing pathway, the CBA protocol provided a more complete depiction on the metabolic potential of pathways and the limits they may pose on a biological network upon implementation. Recycling of cofactors by by-product formation, cofactor maintenance reactions and the tight coordination between different subsections of metabolism are just a few examples of the very interesting observations that can be made with a network-centric protocol that would otherwise not be possible to account for with alternative methods.

5.3.4 Evaluation of alternative methods to reduce high-flux futile cycles

pFBA solutions inherently rely on the objective function used to calculate the optimal flux distribution (García Sánchez and Torres Sáez, 2014, Schuetz et al., 2007). Changing the objective function to the production of a particular excreted molecule determines the maximal production potential of the metabolic network, i.e. the maximum amount of product that can be converted from glucose. This widely used

optimization criterion (Bonarius et al., 1996, 1997, Savinell and Palsson, 1992, Varma et al., 1993) is applied purely based on the assumption that we can achieve the full conversion from all available glucose to product with no by-products being formed (Pasztor et al., 2015), which is arguably not a realistic assumption, but does still represent examples of ideal catalysts. As the objective of my study was to implement a more versatile approach to the RPC analysis by Dugar and Stephanopoulos (2011), I adhered to single linear optimizations based on product maximization, but results, as far as cofactor metabolism is concerned, were mostly underdetermined. In Section 5.3.3, it was shown that high-flux futile cycles could be addressed through manual curation of carefully selected cofactor reactions (Figure 5.4, Table 5.2). From this, it became evident that the more constraints from known biochemical principles were added to the model, the narrower the range of phenotypes would be. In this section, I asked whether a similar outcome could be obtained by other approaches, including (1) "loopless" FBA (Schellenberger et al., 2011), (2) constraining the model with flux data from ^{13}C-MFA, and (3) using 'Minimization Of Metabolic Adjustments' (MOMA) for predicting gene knock-out phenotypes (Segre et al., 2002).

I first implemented the COBRApy "loopless_solution" function (Ebrahim et al., 2013), however, this did not eliminate the futile cofactor cycles (excel file available upon request). Alternatively, I turned to flux data. First, I wanted to understand how much larger the flux ranges of the *E.coli* metabolic network were compared to observations captured experimentally. The following landmark "omics" studies have gained insights into the physiological and metabolic responses of the *E.coli* metabolic network following environmental and/or genetic perturbations (Haverkorn Van Rijsewijk et al., 2011, Ishii et al., 2007, Long and Antoniewicz, 2019, Long et al., 2016). One approach is to truncate reactions within central carbon metabolism (Ishii et al., 2007), as CCM carries the bulk of the carbon flux and is also responsible for the generation of energy, cofactors and most other cellular functions. Regulators of central carbon metabolism are also of great interest (Haverkorn Van Rijsewijk et al., 2011), as well as modifications related to energy and redox metabolism (Holm et al., 2010), given that cofactor availability is fundamental to manipulate and rewire fluxes to increase metabolic yields. Ishii et al. (2007) studied metabolic robustness responses to both genetic and environmental changes, using 24 single-gene knockout strains of *E.coli* from the Keio Collection (Baba et al., 2006) bearing single-gene disruptions from glycolysis and PPP (Ishii et al., 2007). In Haverkorn Van Rijsewijk et al. (2011), the metabolic responses to knockouts of 91 transcriptional regulators from CCM were evaluated, in an attempt to discern the key regulatory processes governing central metabolism (Haverkorn Van Rijsewijk et al., 2011). More recently, Long and Antoniewicz (2019) carried out a comprehensive metabolic characterization of wild-type *E. coli* and 20 knockout strains of CCM under identical growth conditions and using current best practices in ^{13}C-MFA. As such, three flux datasets were initially selected (Haverkorn Van Rijsewijk et al., 2011, Ishii et al., 2007, Long and Antoniewicz, 2019) because they were specific to *E.coli* MG1655, the stoichiometric models used for fitting the data were similar to the *E.coli* core model, they evaluated a large number of targets and implemented a holistic omics approach to the generation, curation and analysis of the data.

Assuming that, for each particular reaction estimated by MFA, the flux estimated across all mutant strains for that reaction represented the available catalytic range of that particular reaction, minimal and maximal

flux ranges were derived for 19 cofactor related reactions that were both experimentally measured and present in the *E.coli* Core model (Figure 5.6). In Ishii et al. (2007), ranges between 0.01-9.8 were observed, with an average of 4.6 ±3.4. In Haverkorn Van Rijsewijk et al. (2011), fluxes spanned ranges between 1.3 to 15.3, with the average range being 5 ± 3.8, similar to Ishii et al. (2007). The Long dataset presented the most flexible ranges between 0.06 to 30, with an average flux range of 8.17 ± 8.8, 1.7-fold higher than the previous two datasets. Interestingly, this dataset also presented considerably high ranges through the THD2 (NAD(P)H transhydrogenase) and ATPS4r (oxidative phosphorylation) reactions (10.2 and 18.5, respectively). In the case of THD2, fluxes spanned both positive and negative flux values, indicative of a high reaction flexibility able to manipulate redox availability according to changes in demand. Also from the Long dataset, I noticed that all TCA reactions had very narrow flux ranges, apart from ICDHyr. This reaction also controls flux entry into the glyoxylate shunt, so the wide flux range was a surprising observation.

All datasets showed discrepancies across important nodes in metabolism. While reactions like GND and G6PDH2r, which propagate flux down PPP, appeared to have wider fluxes in Ishii et al. (2007) and Haverkorn Van Rijsewijk et al. (2011), they presented fluxes up to 3.8 fold narrower in Long and Antoniewicz (2019). Similarly, PYK, a key reaction enabling acetyl-CoA and fermentation product release from pyruvate after glycolysis, was 2-fold narrower in Long and Antoniewicz (2019) when compared to Ishii et al. (2007) and Haverkorn Van Rijsewijk et al. (2011). In the contrary, ICDHyr, a key node reaction in the TCA cycle, presented a small flux range in both Ishii and Haverkorn datasets (of 8.7 and 3.7, respectively) whilst the range was 8 fold more flexible in the Long dataset. Furthermore, all datasets also disagreed on the flux ranges within glycolysis. While the Long and Haverkorn datasets presented similar flux ranges for the GAPD reaction, the Ishii dataset observed a 2.5-fold narrower range; and while Long et al. reported a very wide catalytic range across PFK, Ishii et al. (2007) and Haverkorn Van Rijsewijk et al. (2011) did not.

Discrepancies across the flux range profiles may arise not only from the fact that the goals and purpose of the selected studies were different (each of these studies explored specific parts of the metabolic and regulatory network), but also from the wide differences in the experimental setup and the MFA predictive methods used. It is well understood that the experimental and culturing conditions dramatically affect the flux estimates (Long and Antoniewicz, 2014). For example, aerobic vs. anaerobic conditions, as well as substrate-limited vs. substrate-rich conditions lead to cellular responses and metabolic states that are significantly different. While Ishii et al. (2007) carried out their evaluation in glucose-limited, chemostat conditions at several different dilution rates and using bioreactors with a 1L total working volume, both Haverkorn Van Rijsewijk et al. (2011) and Long and Antoniewicz (2019) used an aerobic batch format, albeit with different working volumes (35mL working volume vs 10 mL, respectively). Media conditions also differed across these studies, with Ishii et al. (2007) using their own synthetic medium recipe, whilst Haverkorn Van Rijsewijk et al. (2011) and Long and Antoniewicz (2019) used M9 Medium. Differences in the experimental setup are further exacerbated by differences in ^{13}C-MFA methodology employed, including tracer selection, data correction, software usage and metabolic map selection for internal flux estimation, which varied across all three studies. The conclusion I gather from this is that we have to be extremely

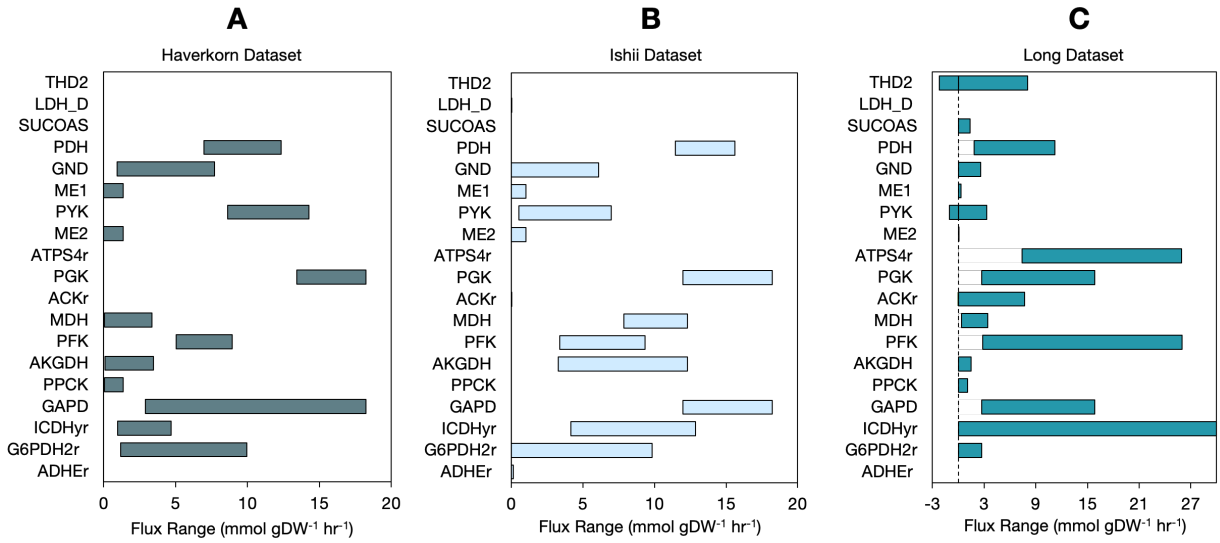


Figure 5.6: Flux ranges gathered from Haverkorn Van Rijsewijk et al. (2011), Ishii et al. (2007) and Long and Antoniewicz (2019). Flux ranges were derived from flux datasets in the original studies, assuming that the fluxes estimated for the same reaction across all mutant strains comprised the full catalytic range of such reaction. Only the ranges of cofactor-related reactions were kept and presented here. (A) flux ranges of cofactor-related reactions from Haverkorn Van Rijsewijk et al. (2011), which presented average ranges between 1.3-15.3; (B) flux ranges from Ishii et al. (2007), which captured ranges between 0.01-9.8; (C) flux ranges from Long and Antoniewicz (2019), which spanned flux ranges between 0.06-30.

prudent when selecting and handling flux datasets, as it is often difficult to deconvolute the effects of knockout mutations and environmental (experimental) changes.

Because I wanted to examine the differences across the range of flexibility of *in silico* and experimental *E.coli* using the aforementioned datasets, the initial intention was to compare these datasets to FVA data. However, this would potentially lead to an unfair comparison between ^{13}C-MFA and FVA variability. There are various reasons for this. First, knocking-out a single gene (as was the case in the three selected MFA studies) would certainly not push the reactions of the network to their optimal range, as is the case in FVA. Secondly, FBA predicts a long-term, balanced optimal flux distribution, and because the ^{13}C-MFA datasets were gathered using knockout strains, often this evolutionary objective is not reached. I concluded that these should instead be better compared against MOMA (Segre et al., 2002).

MOMA provides a flux distribution corresponding to a knockout strain, and approximates this to a reference model. However, it does not produce a range of fluxes. To create a flux range for comparison with the ranges derived from MFA data (Figure 5.6), I needed to simulate the knockout strains in the flux dataset. The Haverkorn dataset implemented knockouts of transcriptional factors, and then measured the effect of these on central metabolism with MFA. This meant that it was not possible to implement these knockouts with MOMA as suggested, because the *E.coli* Core model did not capture transcriptional information.

Finally, I settled on the most recent dataset provided by Long and Antoniewicz (2019). As far as I know, this is the highest quality dataset available which also can be modelled. This study evaluated the effects of 20 reactions in central metabolism, 15 of which were in the model used in this study. The flux ranges derived from this dataset are shown in Table 5.3.

Table 5.3: Upper and lower bound constraints derived from ^{13}C-MFA. Reactions displayed are the cofactor related reactions measured in Long and Antoniewicz (2019). Ranges were calculated by recording all of the reaction fluxes across all mutant strains evaluated in Long and Antoniewicz (2019). The maximal and minimal fluxes observed were used to derive upper and lower bound constraints, respectively, assuming that the full range comprised the catalytic range of the reaction in question. Units shown in mmol $gDW^{-1}hr^{-1}$

Reaction Name	Lower Bound	Upper Bound	Range
PFK	2.843	25.962	23.119
GAPD	2.721	15.845	13.124
PGK	2.721	15.845	13.124
PYK	-0.273	3.326	3.599
G6PDH2r	0.019	2.695	2.676
GND	0	2.605	2.605
PDH	1.846	11.2	9.354
ICDHyr	0	29	29
AKGDH	0	1.512	1.512
SUCOAS	-0.061	1.366	1.427
MDH	0.356	3.445	3.089
ME2	0	0.044	0.044
ME1	0	0.302	0.302
PPCK	0	1.058	1.058
ACKr	-0.076	7.69	7.766
ATPS4r	7.417	25.821	18.404
THD2	-0.984	8.058	9.042

Applying each of these knockouts as a separate MOMA simulation,I again assumed that the resulting flux for each measured reaction in Long and Antoniewicz (2019) across all MOMA solutions comprised the full catalytic range of the reaction. These are hereinafter referred to as "MOMA" ranges, and are displayed in Table 5.4. Finally, and for the sake of completion and comparison, the *E.coli* FVA range was also included, as it helps to provide context, given that pFBA is used in the CBA calculation. This time, however, FVA ranges were simulated at 97% of the optimum to capture the deviation in growth rate found in Long and Antoniewicz (2019). The FVA ranges are shown in Table 5.5.

The flux variability ranges captured in Long and Antoniewicz (2019) were directly compared against the MOMA ranges (Figure 5.7.A) and the FVA ranges (Figure 5.7.B). 10 out of 17 and 12 out of 17 reactions displayed a greater variability range *in silico* (FVA and MOMA, respectively) compared to that measured with ^{13}C-MFA (Tables 5.3-5.5). The MFA data indicated that some reactions were more "plastic" (i.e. more

Table 5.4: Upper and lower bound constraints derived from MOMA (Mahadevan and Schilling, 2003). Using MOMA, each of the knockout strains evaluated in Long and Antoniewicz (2019) was simulated, and the flux running through the cofactor related reactions displayed here was recorded. Ranges were calculated by finding the maximal and minimal flux values across all mutant strains, assuming that they captured the full catalytic range of the reaction. Units shown in mmol $gDW^{-1}hr^{-1}$

Reaction Name	Lower Bound	Upper Bound	Range
GAPD	0	17.257	17.257
PGK	0	17.257	17.257
PYK	0.329	4.453	4.124
G6PDH2r	0	21.332	21.332
GND	0	21.332	21.332
PDH	0	10.906	10.906
ICDHyr	0	7.82	7.82
AKGDH	0	6.932	6.932
SUCOAS	0	6.932	0
MDH	0	5.35	5.35
ME2	0	1.582	1.582
ME1	0	0.795	0.795
PPCK	0	50.85	50.85
ACKr	0	0	0
ATPS4r	0	5.573	43.269
THD2	0	5.604	5.604

flexible, able to change flux more widely according to changes in demand), such as PFK, GAPD and PGK in glycolysis, whilst other reactions were more "rigid" (i.e. showing no or very little change in flux, such as ME1, ME2 and PPCK (ranges of 1.3). Interestingly, some of the reactions that were rigid in reality were predicted to display wide flux ranges using both FVA (Table 5.5) and MOMA (Table 5.4). These rigid reactions were also commonly involved in high-flux futile cycling in the unconstrained engineered models, very likely stemming from the underdetermination of the unconstrained models.

The MFA flux ranges derived from Long and Antoniewicz (2019) were also implemented as upper and lower bound constraints for the same reactions in the engineered butanol models followed by optimization using the CBA algorithm (Fig. 5.8.A), in an attempt to evaluate whether this could replace manual capping of futile cofactor cycles. The use of MFA flux constraints was not sufficient to eliminate all cofactor dissipation in the engineered models, and the use of MFA constraints did not result in any biomass formation. ATP burning through ATPM was still present in all solutions, with the exception of BuOH-2 and fasBuOH (Appendix B Table 8). Furthermore, redox balancing through the THD2 and NADTRHD reactions was active in all models except CROT. Interestingly, all models except BuOH-2 and fasBuOH presented fluxes through PFL (Appendix B Table 8). This could be explained by the narrower flux range of the neighbouring PDH reaction (Table 5.3). If experimental constraints were also combined with additional capping of the

Table 5.5: Upper and lower bound constraints derived from Flux Variability Analysis. Flux variability analysis of the *E.coli* core model calculated at 0.972 of the optimum

Reaction Name	Lower Bound	Upper Bound	Range
PFK	4.981	11.906	6.925
GAPD	13.588	17.774	4.186
PGK	13.588	17.774	4.186
PYK	0	7.626	7.626
G6PDH2r	0	12.559	12.559
GND	0	12.559	12.559
PDH	6.575	15.254	8.679
ICDHyr	0.903	8.178	7.275
AKGDH	0	7.274	7.274
SUCOAS	0	7.274	7.274
MDH	2.431	10.7	8.269
ME2	0	5.007	5.007
ME1	0	3.641	3.641
PPCK	0	4.155	4.155
ACKr	-1.093	0	1.093
ATPS4r	34.67	43.043	8.373
THD2	0	12.559	12.559

ATPM maintenance reaction and FBP (and in the case of tpcBuOH also PPCK) following the manual curation procedure outlined earlier (Figure 5.4, Table 5.2), the predicted maximum target product yield was markedly reduced and in most models except for BuOH-0 and BuOH-1 they were similar to that predicted through the RPCs (Figure 5.8.B, Appendix B Table 9).

The results obtained from the CBA protocol, in conjunction with a comparison of the carbon yields across unconstrained and constrained models are shown in Figure 5.8. The protocol now captured lower waste and biomass metrics for the butanol producers, and higher waste metrics for the butanol precursor models, more in line with the findings presented by Dugar and Stephanopoulos (2011). The CBA protocol confirmed again the superiority of the BuOH-1 solution at channeling cofactors towards optimal butanol secretion. Interestingly, engineered models presented objective values that were on average 36% lower than when manually curated, resulting in theoretical yields closer to the adjusted estimates reported by the calculations of Dugar and Stephanopoulos (2011). This suggests that the use of experimental constraints, rather than manual curation, may be a more suitable approach to obtaining reasonable flux estimates, as they capture competing pathways (such as succinate release, as per Appendix B Table 8) and additional intracellular metabolic complexity that may impact the calculation of biosynthetic capabilities. However, MFA-derived constraints alone did not achieve this outcome, whack-a-mole manual curation of futile cycles was still needed to minimize cofactor dissipation. I thus concluded that MFA-based constraints, based

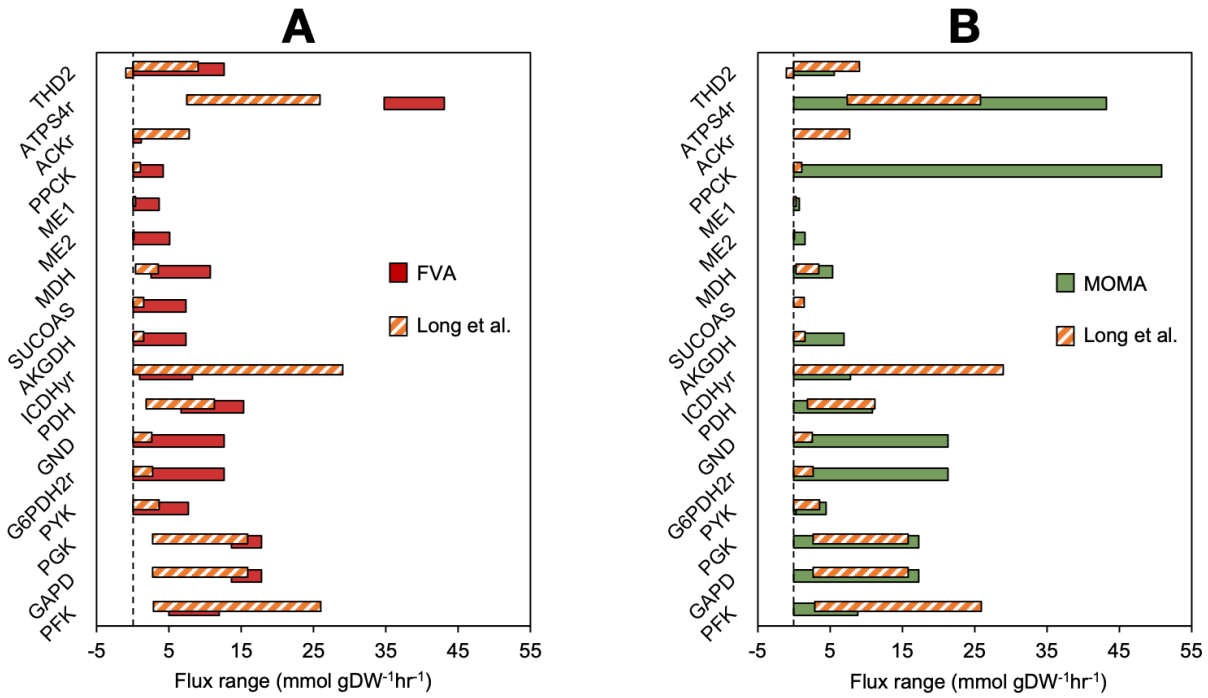


Figure 5.7: MFA-derived estimates compared to FVA and MOMA estimates. Flux ranges were extracted from a pre-existing MFA flux dataset (Long and Antoniewicz, 2019), using a Python algorithm to select the minimal and maximal flux ranges and assuming that the fluxes estimated for the same reaction across all mutant strains used in the study comprised the full catalytic range of such reaction. Only the ranges of cofactor-related reactions were kept and presented here. These flux ranges are indicated as "Long et al." in the figure, and were compared against (A) FVA flux ranges (orange stripes), and (B) MOMA-derived flux ranges (green). MOMA ranges were estimated using the wild type solution as a reference and sequentially implementing the single-gene knockouts in Long and Antoniewicz (2019), with biomass formation as the objective function.

on the Long and Antoniewicz (2019) dataset, although they assisted in minimizing manual curation, were therefore not an ideal replacement for manual whack-a-mole futile cycle capping.

5.3.5 Sensitivity analysis identifies optimal cofactor balance

The CBA analysis seemed to suggest that models having the least amount or no flux towards biomass are the most balanced ATP and redox-wise. For example, the manually curated BuOH-1 model had the least amount of futile cycling and ATP burning, the highest butanol yield and lowest biomass yield (Section 5.3.3). Is this the optimum or could the catalytic system be improved even further? Here, I define the optimal balance as the cofactor profile that results in maximal theoretical target yield. It is assumed that the optimal solution would minimise carbon-containing waste release, and energy and redox waste

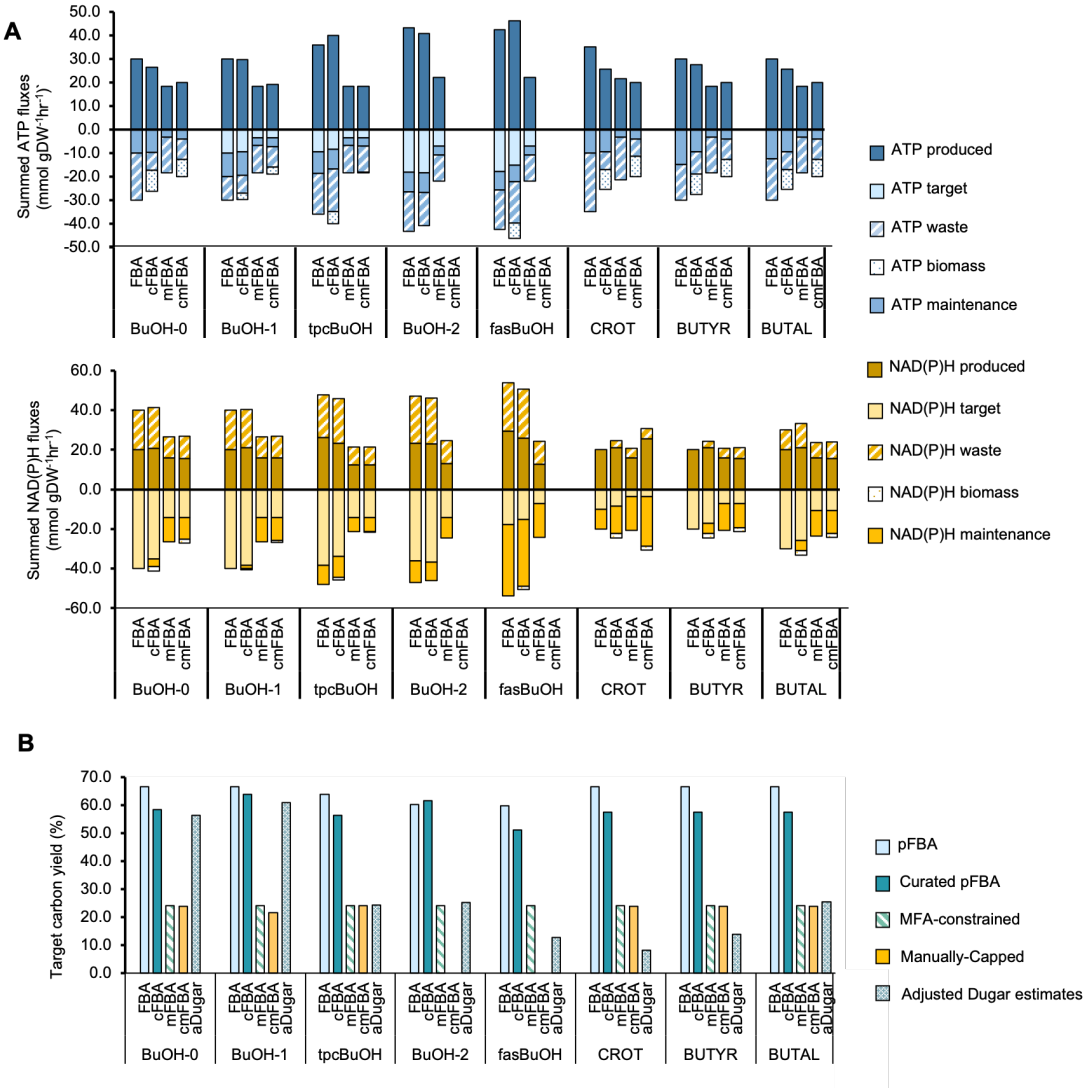


Figure 5.8: CBA cofactor profile comparison across unconstrained, manually curated and experimentally constrained solutions and carbon yield comparison. (A) ATP (blue) and NAD(P)H (yellow) CBA-derived cofactor usage profiles compared across all approaches evaluated in this study; (B) Carbon yields of butanol and butanol precursor models, compared across all approaches evaluated in this study: unconstrained pFBA (labelled "FBA"); manually curated pFBA solutions with minimized high-flux futile cycling (labelled "cFBA"); experimentally-constrained solutions using MFA-derived flux data (labelled "mFBA"); experimentally-constrained solutions using MFA-derived flux data with further capping in cofactor cycling reactions (labelled "cmCBA").

streams. Under aerobic conditions, any excess is generally in the form of CO_2 release, whereas under anaerobic conditions, fermentation products should also be carefully monitored. To better understand the effect of cofactor balancing on product yield, a sensitivity analysis was designed to assess the relationship between changing ATP and NAD(P)H demands and improvements in theoretical yield. Such an approach could also potentially be used to generate a metric indicative of the cofactor imbalance in each model. The cofactor sensitivity analysis was applied to the manually-curated, aerobic, butanol models BuOH-0, BuOH-1, tpcBuOH and fasBuOH, by introducing an artificial NADH and ATP sink/generator that modifies the pathway's ATP and NADH stoichiometric coefficients, to simulate pathways with both cofactor surplus (excess ATP/NADH produced by the target pathway) and cofactor demand (ATP/NADH going into the pathway). I anticipated that this evaluation may be particularly relevant to pathway variants producing the same target, as theoretical yield differences could be triggered by cofactor usage differences across synthetic pathways.

To develop the cofactor sensitivity analysis, the demand for ATP and NAD(P)H was artificially varied to assess any improvements in theoretical yield, by introducing into the existing stoichiometry of the aldehyde reductase reaction (Equation 5.4), the final step in the butanol production chain, an artificial ATP production step (Equation 5.5). By looping through this reaction's A(X)P and NAD(P)H stoichiometries in a stepwise manner, from cofactor production (Equation 5.6), through to cofactor consumption (Equation 5.7), I built a landscape plot using the full range of possible metabolic scenarios. I evaluated a range of -10 to +10 (consumption of -10 ATP/NAD(P)H to a surplus of +10). With each step, every time the stoichiometry changed, I optimised for the selected objective (butanol production) and stored the butanol and biomass yields for further evaluation.

$$butyraldehyde + NADH + H^+ \rightarrow n-butanol + NAD^+ \tag{5.4}$$

$$butyraldehyde + NADH + H^+ + ADP + Pi \rightarrow n-butanol + NAD^+ + ATP \tag{5.5}$$

$$butyraldehyde + XNAD^+ + YADP + YPi \rightarrow n-butanol + XNADH + XH^+ + YATP \tag{5.6}$$

$$butyraldehyde + XNADH + XH^+ + YATP \rightarrow n-butanol + XNAD^+ + YADP + YPi \tag{5.7}$$

The resulting 3D landscapes describe the impact of changes in cofactor demand on product yield under aerobic conditions (Figure 5.9) and anaerobic conditions (Figure 5.10). Under aerobic conditions, models BuOH-0 and BuOH-1, which result from the implementation of pathway combinations AtoB + AdhE2 and NphT7 + AdhE2 (Figure 4.1), accordingly, could withstand growing pathway ATP demands and redox surplus by forming a plateau at a theoretical carbon yield of 66.67%, without displaying biosynthetic deficiency (Figures 5.9.A and 5.9.B), a behaviour associated with more robust systems (Wu et al., 2016). These pathways exhibit high glycolytic flux under aerobic conditions and the increasing demand for ATP is

satisfied by higher respiration thanks to the increasing redox availability . The original synthetic pathways in these two engineered systems reached carbon yields of up to 58.53% and 64.47% (Figure 4.2), but artificial cofactor dissipation boosted the yield by 8.14% and 2.2%, accordingly. The highest biomass rates were recorded in the presence of both growing ATP and NADH surplus, at the expense of butanol production (Figures 5.9.A and 5.9.B).

The tpcBuOH and fasBuOH models displayed limited capability to accommodate any change in cofactor demand, thus forming a cliff and causing a drop in product yield (Figures 5.9.C and 5.9.D). Both of these CAR-dependent models have high ATP demand, as AMP needs to be recycled via an ATP-consuming ADK1 reaction, consuming further ATP in the process. They also have a high requirement for NADPH, resulting in increased flux through the Pentose Phosphate Pathway. As so referred in (Wu et al., 2016), these more unstable models (tpcBuOH and fasBuOH) were highly sensitive to changes in cofactor demand and manifested sharp cliffs upon changes in cofactor usage. They achieved carbon yields of 56.43% and 51.1%, respectively, without the artificial cofactor variation, but by manipulating their cofactor demands the maximal carbon yields increased up to 61.3% and 56.67%, accordingly. Under anaerobic conditions, all models displayed more unstable behaviours, lower butanol carbon yields and only BuOH-0 and BuOH-1 produced biomass. These observations suggest that if the sweet spot for optimal balance between the introduced pathway and host metabolism is small, it reduces the chances that a high-yielding integrated combination of pathway and host cell metabolic network can materialise. It doesn't exclude the possibility of high yield, but it makes it less likely and prone to higher metabolic instability.

These results suggest that it would be possible to determine a stoichiometrically "optimal" ATP/NAD(P)H ratio for each pathway in relation to the rest of the network, providing insights into the yield and cofactor profiles of optimally balanced strains. This analysis helps define both the optimal balance and the level of imbalance in a particular metabolic engineering design, which is indicated by the distance between the optimal solution and the current solution. The smaller this distance is, the closer the current design is to the optimum achievable by the biosynthetic pathway. Knowing this, pathway engineering can be theoretically guided both in terms of selecting the optimal host strain background and by indicating the cofactor robustness of the pathway and likelihood that high yield is achieved. This information opens up a new horizon for further metabolic engineering adjustments that can potentially lead to more rapid implementation of high-yielding production strains.

5.4 Chapter summary

This chapter presents the first attempt at developing a stoichiometric-modeling-based cofactor Balance Assessment (CBA) protocol to monitor cofactor usage and its system-wide impact on cell behaviour and the design of optimal catalysts. With this, metabolic engineering designs can be selected based not only on the highest yield achievable but on knowledge of the stress imposed by cofactor metabolism. The CBA protocol describes cofactor (im)balance as the fraction of cofactor diverted to biomass, maintenance or waste, instead of target production, and captures how organisms may be limited by pre-existing redox

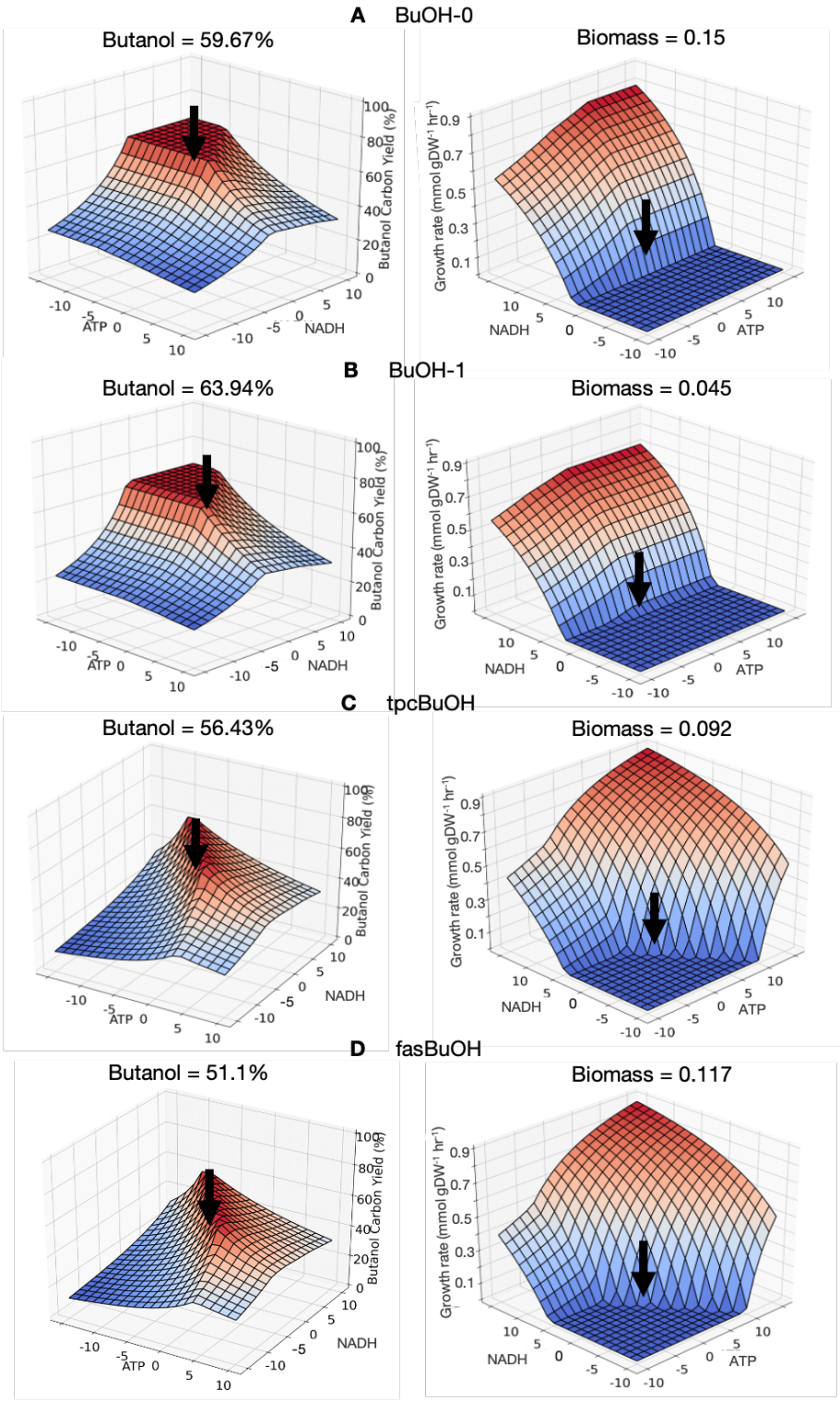


Figure 5.9: aerobic butanol rates (mmol gDW^{-1}hr^{-1}) of engineered *E.coli* strains in response to changes in ATP and NAD(P) demands. Each model represents a unique pathway variant for butanol production, which has previously been manually-curated and optimized for the selected objective under aerobic conditions. (A) BuOH-0 model, comprised of route AtoB + AdhE2; (B) BuOH-1, including reactions NphT7 + AdhE2; (C) tpcBuOH made up of AtoB + TPC7; (D) BuOH-2, comprising reactions NphT7 + TPC7.

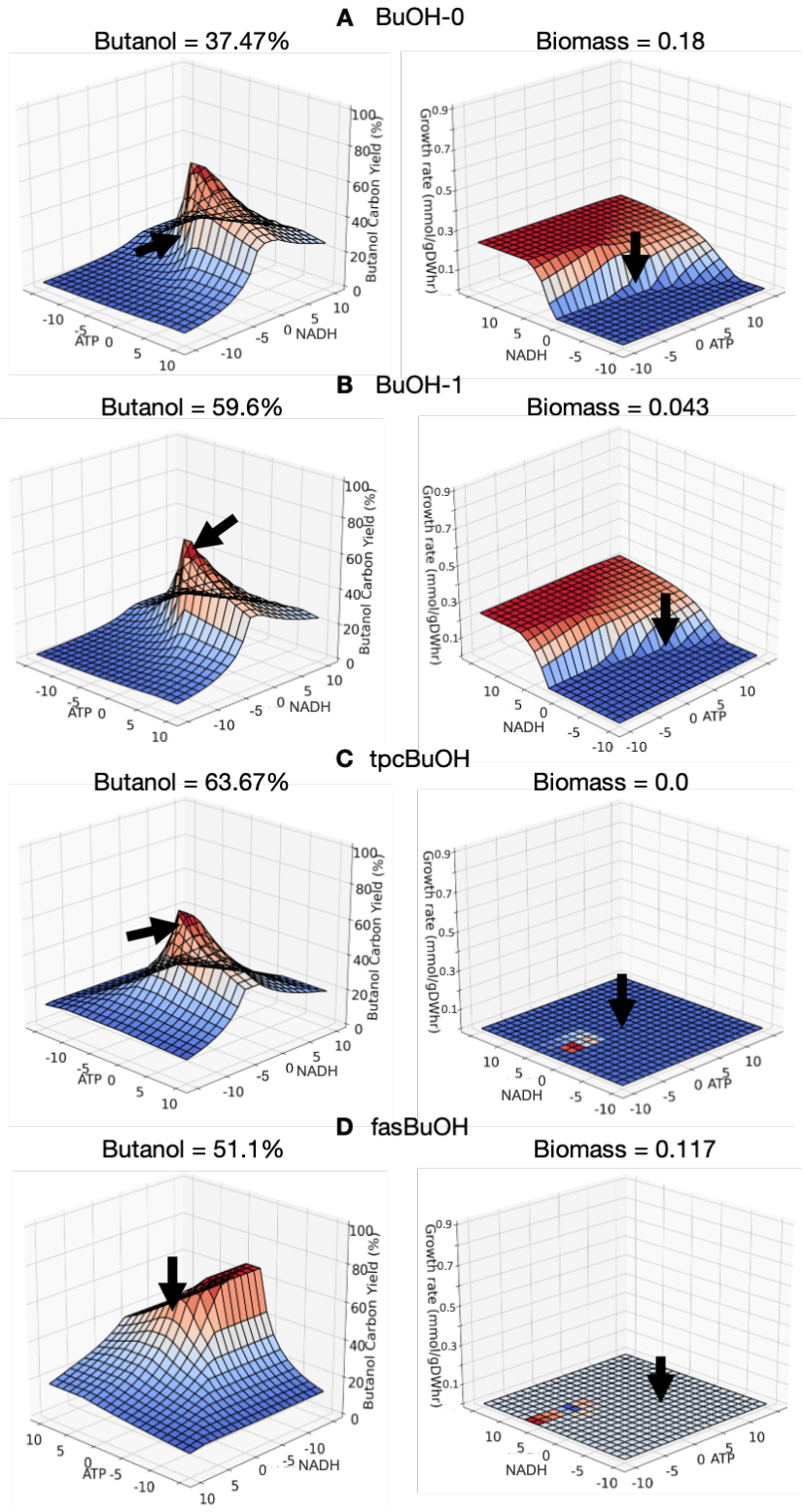


Figure 5.10: anaerobic butanol rates (mmol gDW^{-1}hr^{-1}) of engineered *E.coli* strains in response to changes in ATP and NAD(P) demands. Each model represents a unique pathway variant for butanol production, which has previously been manually-curated and optimized for the selected objective under aerobic conditions. (A) BuOH-0 model, comprised of route AtoB + AdhE2; (B) BuOH-1, including reactions NphT7 + AdhE2; (C) tpcBuOH made up of AtoB + TPC7; (D) BuOH-2, comprising reactions NphT7 + TPC7.

and energy constraints. Only three elements are used as inputs in this modelling framework: (1) the stoichiometry of cofactor related metabolic reactions, (2) measured intracellular fluxes and (3) a set of reactions leading to target optimization. The contributions of ATP and NAD(P)H to biomass, target production, waste release or maintenance can then be estimated as the sum of fluxes of all cofactor related reactions consuming and/or producing such cofactors.

The CBA function hinges on the following assumptions. First, it relies on ATP and NAD(P)H tracking, whilst all other cofactors are not considered. Selecting ATP and NAD(P)H amongst all available cofactors was a reasonable approach because, after all, they are three of the most widely used and widely connected cofactors, and therefore their availability (or lack therefore) will have a considerable impact on bioproduction. By only considering ATP and NAD(P)H it was possible to simplify the development, testing and interpretation of results. Nonetheless, the flexibility of the COBRApy environment allows for the CBA protocol to be modified to consider alternative cofactors or to be amplified to incorporate further cofactors into the analysis in the future. Second, CBA also assumes that NADH and NADPH are interchangeable. In reality, NADH and NADPH are chemically and biologically distinct, but CBA accounts for the consumption and and production of these two cofactors in an overall "redox" category, instead of having two separate groups. This approach eased the redox classification, especially given that the transhydrogenase system is allowed to freely channel electrons between the two cofactor pools when needed. Computationally, however, this classification had no impact on the behaviour of the system itself, i.e. the models and their corresponding solutions do not change as a result of this joint categorisation, it only affects the way in which results are presented. Finally, it is also assumed that only biomass, cell maintenance, target and waste release contribute to cofactor usage. For biotechnological purposes, it is reasonable to assume that these categories are generally the key areas of focus for cofactor flux evaluation, enabling a systematic comparison between the amount of cofactor flux being put towards target production, compared to everything else. Tracking the waste and biomass profiles is particularly relevant, as it enables us to discern the extent of cofactor excess that is currently imbalanced by the system. Nonetheless, this particular categorisation approach is not fixed, and it would accept modifications and expansion both on the nature and number of categories, particularly when regulatory or transcriptional data is incorporated into the constraint-based modelling approach, or when moving into more complex systems such as yeast, which may require a categorisation scheme that incorporates compartmentalisation.

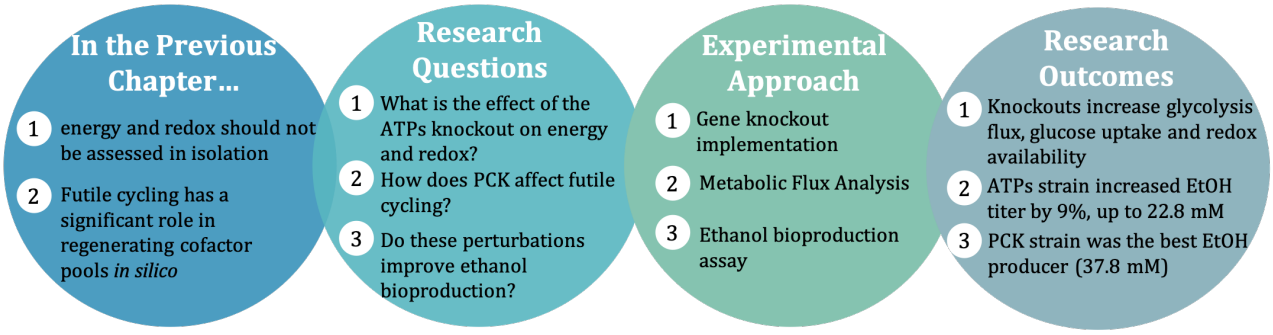
These results suggest that introducing cofactor-balanced pathways reduces the burden placed on the rest of the metabolic network. They also indicate that ATP and NAD(P)H balance cannot be assessed in isolation from each other, or from the balance of additional cofactors, such as AMP and ADP. A manual curation method was also evaluated whereby solutions were constrained to reduce their flexibility. I showed that subject to appropriate constraints, solutions had no apparent futile cycles, but it is vital that we do not forget FBA's flexible nature and optimistic estimates when using this *in silico* approach.

In this chapter, I also present an approach whereby the "optimal", artificial cofactor stoichiometry can be directly compared to the stoichiometry of the existing design to gain an understanding of the level of imbalance presented by the design under evaluation, providing insights into the yield and cofactor balance

sweet spot of engineered strains. Identifying the optimal balance is especially relevant when narrowing down pathway variants, as identification of the most efficient and robust pathway for target production becomes an essential step in the design approach. I believe it is very important to understand how far an existing synthetically modified organism's cofactor profile may be from the optimal, as diverging away from the optimal profile represents non-optimally balanced solutions. These results indicate that it is possible to substantially increase target production if we modify the ATP and redox demands of the introduced pathway(s). This information opens up a new horizon for further metabolic engineering adjustments that can lead to better yields, whereby strategies reviewed in Chapter 2, such a synthetic build-up of NAD(P)H/ATP or a synthetic sink for NAD(P)H/ATP, could be introduced into the system to better cater to the optimal network-wide cofactor profile.

6

Energy and Redox Perturbations for Enhanced Bioproduction



Chapter Acknowledgements

I kindly thank the Bacterial Cell Factory Optimisation Laboratory from the Technical University of Denmark (DTU), for providing the SIJ488 *E.coli* strains needed for this work.
I would also like to thank Dr Lian He for the extensive and reliable assistance setting up and optimising my WUFlux simulations for the precise estimation of internal metabolic fluxes.
Finally, I am endlessly grateful to my lab colleagues for their particular contributions: Jessica Rollit (then a PhD student), for preparing and providing electrocompetent SIJ488, and for her assistance with knockout

optimisation, GC-MS set-up and troubleshooting, HPLC standard preparation and HPLC set-up; Pachara Sattayawat (then a PhD student) for helping me with HPLC calibration and assembling sequence runs, and for providing me with the plasmids necessary for this study; Mathieu Bousquet (then a PhD student), for his advice on labelling experiments, MFA sample preparation optimzation, and data correction and analysis; and finally my undergraduate student Lidia Ripoll-Sánchez for implementing the PCK knockout strain during her final year Biochemistry project (Ripoll Sanchez, 2018), which was extensively used throughout this study.

6.1 Introduction

In Chapter 5, I presented the CBA protocol, a COBRA-based modelling framework for the study of cofactor balancing at the network-scale. CBA elucidates how a synthetic pathway's cofactor demands integrate with the wider metabolic network. Thus, CBA informs the earlier stages of pathway selection by distinguishing cofactor-efficient pathways from more wasteful pathways prior to experimental implementation.
The results from the CBA protocol brought about two significant outcomes warranting further investigation. First, it emphasised the importance of not considering energy and redox in isolation from each other, or even from the roles of additional cofactors, such as AMP and ADP. It is widely understood that the biochemical roles of ATP and NAD(P)H are tightly connected through selected metabolic nodes, for example, through ATP generation from oxidative phosphorylation, which is fueled by NADH from the ETC. Furthermore, both cofactors are needed for biomass production. Despite their inherent interdependence, the field has continued to study ATP and NAD(P)H as isolated entities, providing in the process only a partial insight into their regulatory potential (Please refer to Chapter 2 for a detailed review of the literature). Secondly, the CBA results were also characterised by the dissipation of surplus redox and energy through futile cycling. Both under aerobic and anaerobic conditions, models responded to surplus ATP and NAD(P)H by activating reaction pairs that could consume any excess cofactor (Figure 5.4). One such commonly-appearing futile cycle was that involving anaplerotic reactions PCK and PPC, which has been shown to dissipate excess ATP when overexpressed experimentally (Chao and Liao, 1994*b*), but whose native components seem to be tightly regulated in reality (Yang et al., 2003).
Therefore, in this chapter, I set out to answer some of these questions from the *in silico* analysis using wet lab experiments. Specifically, what is the effect of the ATP synthase (ATPs) knockout on cellular energy and redox? How does the presence or absence of PCK affect futile cycling? Is there even any futile cycling in *E.coli*? And, can these cofactor-related perturbations be harnessed to improve ethanol bioproduction in *E.coli*?

6.1.1 Chapter overview

In Chapter 6, I first investigated the coordinated roles of ATP and NADH, and the existence of futile cycling in *E.coli* by evaluating the physiological and metabolic responses of *E.coli* to cofactor perturba-

tions implemented by knocking out: (i) the ATP synthase operon (ATPs), the node linking redox and energy metabolism; and (ii) PCK, which is involved in the PCK-PPC anaplerotic futile cycle (Figure 6.1). To characterise the strains, a growth analysis was performed. Furthermore, the absolute central carbon metabolism fluxes of the wild type and mutant strains were estimated relative to their extracellular glucose uptake rates using a labelling experiment and metabolic flux analysis (MFA) using the software WUFlux (He et al., 2016). Next, these cofactor manipulations were further studied in combination with an ethanol-producing plasmid system to assay their effect on ethanol bioproduction.
With this study, I thus proposed that, in order to achieve any desired metabolic changes in an engineered system, it is crucial to first understand the effect of cofactor-related modifications on the cell's native system, before using them as optimisation strategies for bioproduction of target products.

6.2 Results

6.2.1 Target energy and redox manipulations

To answer the above research questions, I selected the ATPs operon and PCK genes as knockout targets to trigger perturbations at the protein (and cofactor) level (Figure 6.1).
The ATPs enzyme catalyses ATP formation using the free energy conserved as a proton motive force across the bacterial inner membrane by the electron transport chain (simplified as Equations 6.1 and 6.2).

$$ATPase : NADH + ADP + P_i \rightarrow 3ATP \tag{6.1}$$

$$ATPase : FADH_2 + ADP + P_i \rightarrow 2ATP \tag{6.2}$$

Unlike Holm et al. (2010), where an evaluation of the regulatory effect of energy and redox was performed through the heterologous overexpression of water-forming NADH oxidase and the soluble F1 component of the ATPs enzyme, in this study I wanted to assess the regulatory effect of truncating the connecting node between energy and redox on cofactor availability. In particular, what is the regulatory effect of ATP availability (or lack thereof) on central carbon metabolism? Enforced ATP dissipation has been previously reported to stimulate substrate-level phosphorylation in other areas of metabolism to rebalance the intracellular ATP supply, namely by glycolysis and ATP-producing acetate formation (Hara and Kondo, 2015, Holm et al., 2010, Kihira et al., 2012, Koebmann et al., 2002).
Alternatively, a knockout of the PCK gene was also implemented (Equation 6.3). PCK is a crucial component in anaplerotic metabolism, substantially studied previously in the context futile cycling, which involved the heterologous overexpression of PCK in conjunction with PPC (Chao and Liao, 1994*b*, Patnaik et al., 1992). Apart from the Keio collection knockout (Baba et al., 2006) and the study of its regulated activity in Yang et al. (2003), the physiological and metabolic responses of this knockout have not been further studied. I would have presumed that, if PCK futile cycling existed in wild type *E.coli* under normal

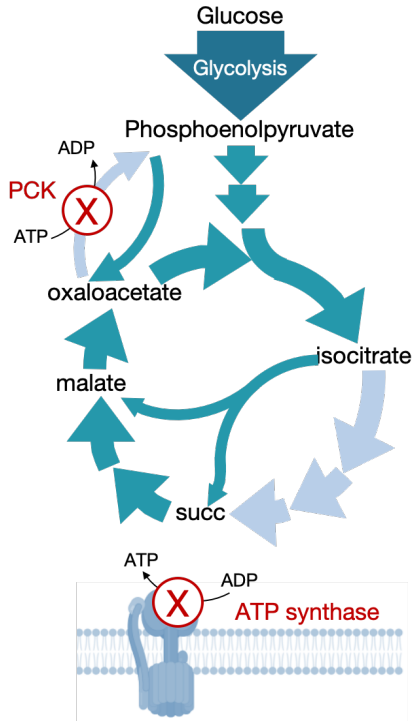


Figure 6.1: Knockout targets selected in this study. (1) The PCK knockout eliminates the scope of a potential PCK-PPC ATP futile cycle to rebalance any ATP surplus; and (2) A full knockout of the ATP synthase operon prevents ATP production through oxidative phosphorylation, interferes with the main link between ATP and NADH and also potentially redirects flux through glycolysis, as a means to produce ATP via substrate-level phosphorylation instead.

conditions, the PCK knockout would instead activate futile cycling elsewhere to dissipate the excess ATP. This is because if there is no PCK activity, the system would need to find an alternative means to burn the surplus energy. This presumption arose from my *in silico* observations, in which excess energy and redox led to a whack-a-mole behaviour upon constraining high-flux futile cycles, as discussed in detail in Section 5.3.3.

$$PCK : oxaloacetate + ATP \rightarrow phosphoenolpyruvate + ADP + CO_2 \tag{6.3}$$

The two knockout strains PCK and ATPs were constructed using the method specified by Jensen et al. (2015), and validated as suggested by Datsenko and Wanner (2000), following the method outlined in figure 6.2. For more information, please refer to Section 3.2.5 (Chapter 3). Figure 6.3 shows successful Kan^R cassette amplification (A), gene deletion (B) and cassette removal (C) for the ATPs strain. The PCK strain had been previously generated in an earlier study (Ripoll Sanchez, 2018) and was thus only subject

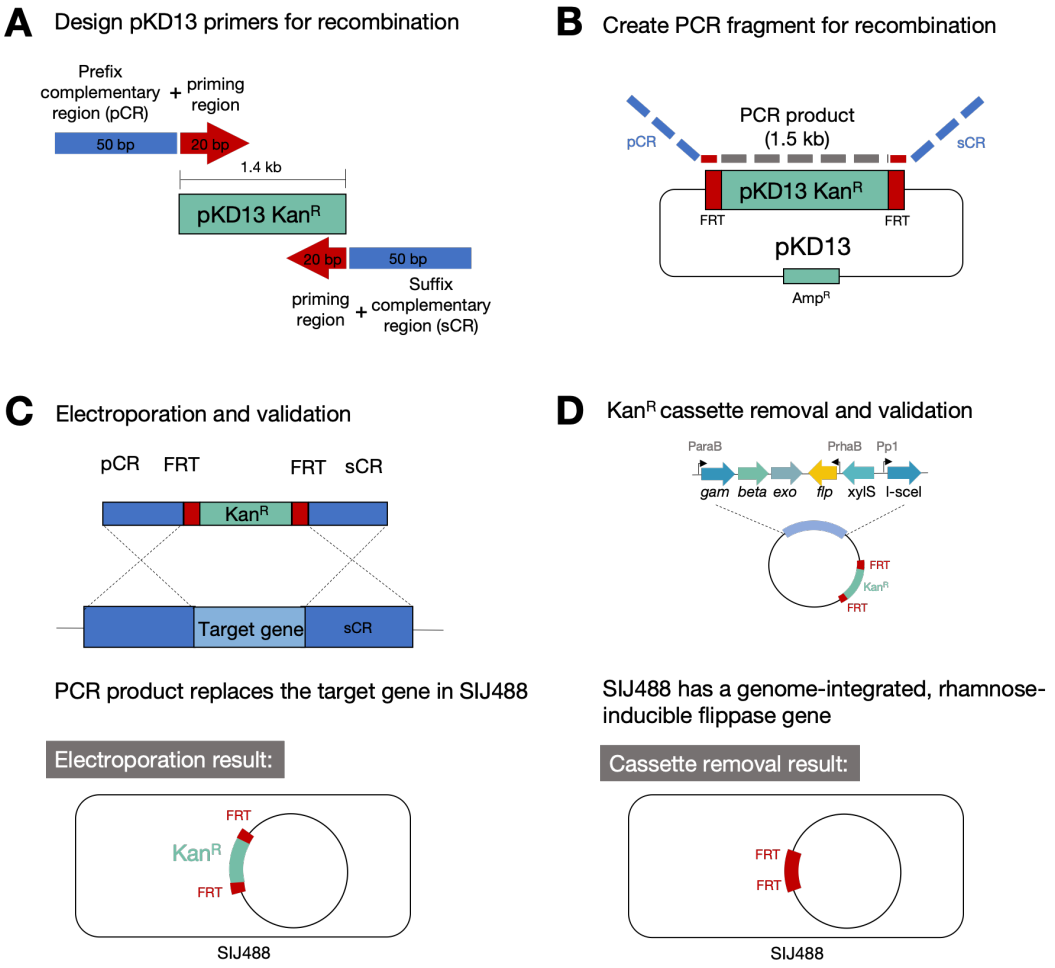


Figure 6.2: Gene deletion method, following (Jensen et al., 2015) and (Datsenko and Wanner, 2000). (A) recombination primers are designed to include (i) a 20bp priming region which binds the FRT-flanked ends of the Kan^R cassette in plasmid pKD13, as well as (ii) a 50bp homology region which binds upstream/downstream of the target gene(s) to be knocked out; (B) PCR product is created by polymerase chain reaction (C) gene deletion is performed by electroporation of the PCR products, which then integrate by recombination of the 50bp homology arms with the target gene(s) (D) Kan^R cassette removal is performed through the induction of a fippase gene that recognizes the FRT regions of the cassette.

to cassette removal validation with Colony PCR and sequencing analysis (Figure 6.2.D). Please refer to Appendix D Figures 1-6 for sequencing validation results.

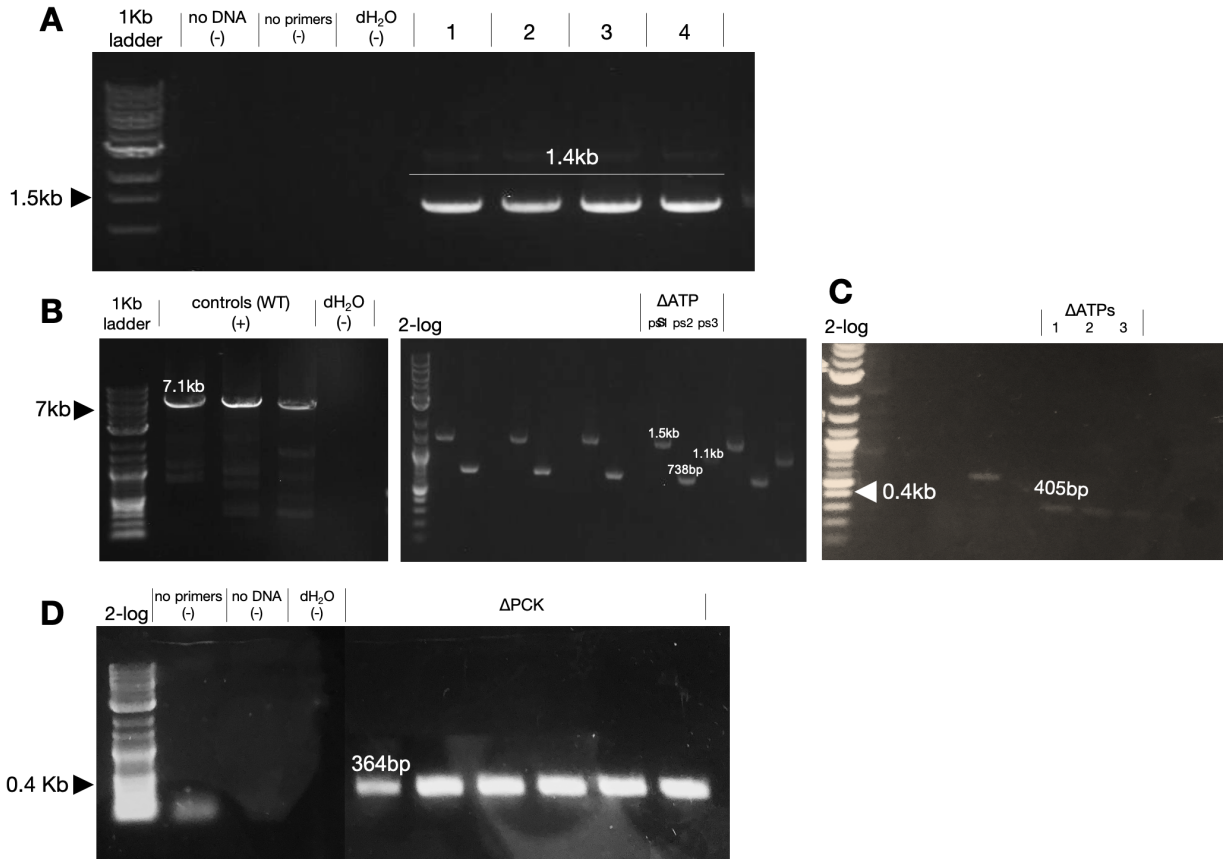


Figure 6.3: Gene Delection Implementation. (A) PCR amplification of pKD13's Kan^R casette with ATPs-complementary arms (100% amplification efficiency, n=4); (B) Gene deletion of the ATPs operon by electroporation of Kan^R PCR product using 250ng PCR product for electroporation. Three primer combinations (labelled PS1, PS2 and PS3 here) were used to confirm amplification. 40% of the colonies checked using this procedure contained the correct knockout mutation (n=5); (C) PCR validation of ATPs cassette removal. 80% of rhamnose-induced colonies got the antibiotic casette successfully removed (n=5); (D) PCK cassette removal validation.75% of colonies tested were positive.

6.2.2 Growth and secretion product profiles show physiological responses to cofactor perturbations

To assess the physiological responses to changes in cofactor metabolism, the wild type and mutant strains were subject to a 10h growth assessment in M9 minimal medium (Section 3.2.6.2, Chapter 3). Figure 6.4.A shows the growth profile, as OD_{600} measurements over the 10h incubation period, of the WT and knockout strains ATPs and PCK. As expected, the WT strain was the best growing strain, followed by PCK, and finally by ATPs. The specific growth rate of the ATPs strain decreased by 11% compared to the

Table 6.1: Physiological parameters and external fluxes (shown as mmol gDW^{-1} hr^{-1}) of WT and mutant strains ATPs and PCK. Timepoints used for calculations were 4h and 8h for WT and PCK and 6h and 10h for ATPs. μ is the growth rate (h^{-1}). Glucose (glc), acetate (ac), formate (for), succinate (suc), lactate (lac) and ethanol (Etoh) concentrations were measured using high-performance liquid chromatography (HPLC). Values are given as the mean SD n= 4. n.d.: Not detected.

	Wild Type	ATPs	PCK
μ (h^{-1})	0.376 ± 0.017	0.333 ± 0.015	0.375 ± 0.015
V_{glc}	8.325 ± 1.548	21.047 ± 1.479	9.824 ± 0.380
V_{ac}	2.934 ± 0.594	11.102 ± 0.133	1.755 ± 0.418
V_{etoh}	0.424 ± 0.100	1.096 ± 0.068	n.d.
V_{lac}	0.100 ± 0.116	n.d.	0.105 ± 0.122
V_{for}	n.d.	n.d.	0.270 ± 0.113
V_{suc}	n.d.	n.d.	0.051 ± 0.010

reference, which is a smaller decrease than previously reported in a separate study that overexpressed a heterologous soluble F_1-ATPase to decrease the intracellular cellular ATP pool (Holm et al., 2010). These results suggested there could potentially have been issues with oxygenation during the growth evaluation, which may have negatively impacted the growth of the WT strain. The PCK mutation appeared not to have a no major effect on growth rate compared to WT.

Figure 6.4.B shows that both mutants increased their glucose consumption rates. More specifically, ATPs presented a considerable increase glucose uptake rate of 82%, compared to the WT. Certainly, one of the effects of lower ATP availability could be the increase in the specific substrate uptake rate. In principle, the cell might counteract the loss of ATP completely by increasing the substrate uptake rate by the amount required to synthesize the lost ATP (so that the growth rate is not impaired) (Hädicke and Klamt, 2015). However, the observed increased substrate uptake rates upon ATPase shutdown could not completely restore the original growth rate (Figure 6.4.A). Similarly, faster glucose consumption in the PCK strain did not seem to be channeled towards higher growth. What was most striking about the ATPs strain was the level of acetate overflow (Figure 6.4.C), which almost doubled with respect to the WT, whereas it decreased by 31.5% in the PCK strain. This seems to be an example of overflow metabolism. Small amounts of other acids were also measured (Figure 6.4.D-G).

6.2.3 Metabolic flux redistributions of the PCK and ATPs strains

To elucidate the impact of the knockouts in the internal flux redistribution, I estimated the internal fluxes of both the wild type and mutant strains using a tracer experiment and ^{13}C metabolic flux analysis (^{13}C-

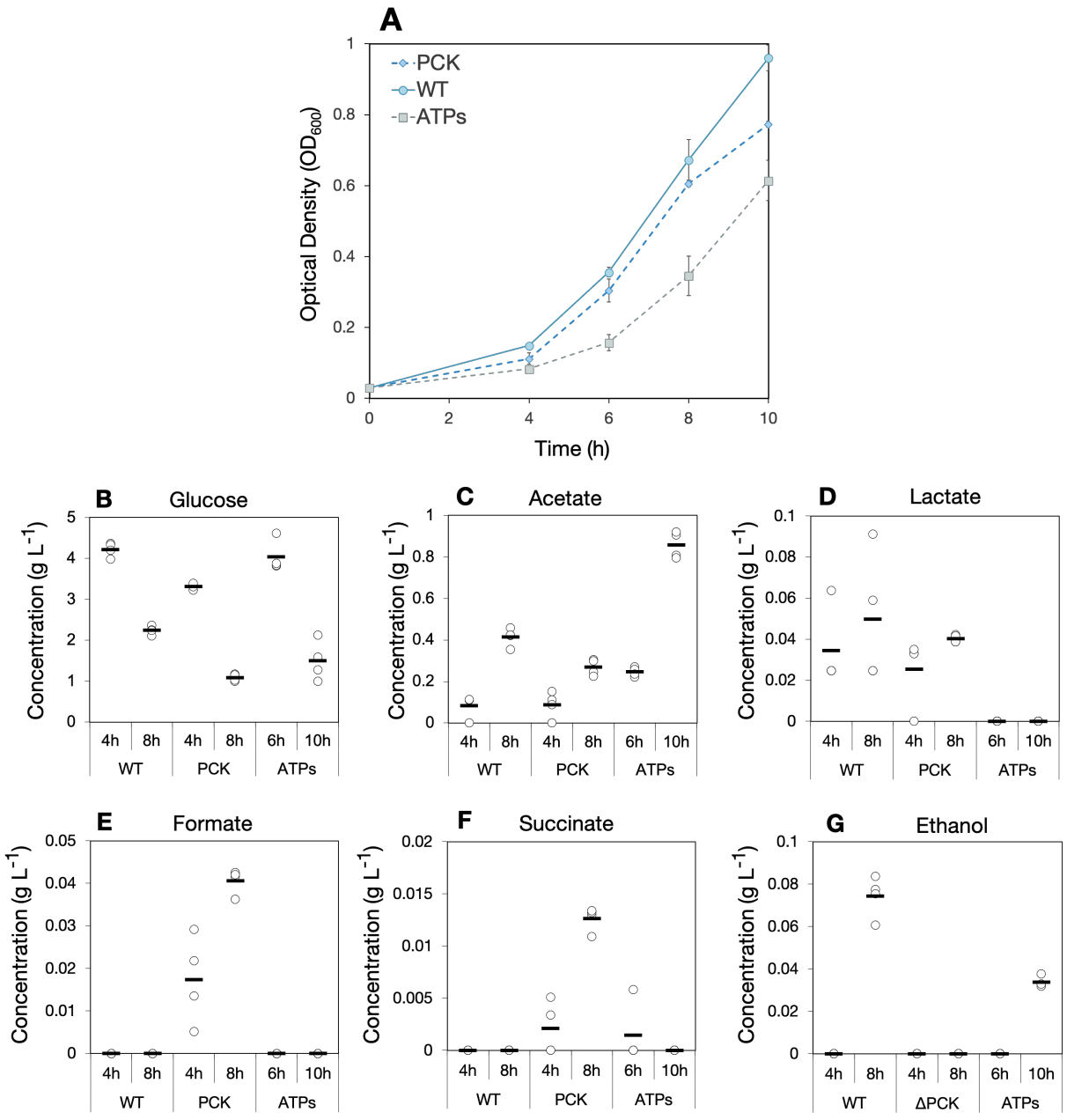


Figure 6.4: Growth curve and uptake and secretion measurements of WT and knockout strains ATPs and PCK. Data was gathered over a 10h labeling experiment, with samples taken every 2h-4h. (A) Growth analysis shown as averaged OD_{600} measurements (n=4) over time.

MFA, or simply MFA). The experimental procedure, performed following the widely recognised protocol by Zamboni et al. (2009) summarised in Figure 6.5, can be widely subdivided into the following three steps. First, a growth experiment using [1–^{13}C] and [U–^{13}C] labelled D-glucose was performed to label the cell's protein and detect the ^{13}C-labelled amino acids using GC-MS. Amino acid fragment patterns were quantified, corrected for natural abundance of isotopes, and fed into the MFA software WUFlux (He et al., 2016) for further flux analysis. Metabolic fluxes (Appendix F Table 16) were estimated using best-fitting techniques, relative to the measured glucose uptake rates, fermentation product secretion profiles and growth rates (Table 6.1). The best fit and standard deviations are shown in Figure 6.6. Please refer to Section 3.2.6.1 (Chapter 3) for a more detailed outline of this procedure.

The WT flux map (Figure 6.6.A) was characterised by significant glycolytic flux during the exponential growth phase, and flux diversions into PPP, the TCA and glyoxylate cycles. About 26% of glucose went into the PPP, while flux through the citrate synthase (CS) and aconitase (ACONT) reactions in TCA accounted for 23% of glucose consumption. This was moderately higher than reported in earlier studies (Chen et al., 2011, Fischer et al., 2004, Nicolas et al., 2007). Interestingly, no flux was recorded through the PCK reaction, suggesting that, under this particular set of conditions, there was no evidence of futile cycling between PCK and PPC. From the literature, it appears that the role of conditions is relevant in futile cycling. For example, environmental or even genotypic and strain-specific conditions are likely to have a major impact on the experimental setup, subsequent observations, and therefore ultimately also the conclusions. While I sampled in batch conditions and during exponential growth (i.e. at pseudo-steady state for flux analysis purposes), Yang et al. (2003) reported considerable futile cycling rates using across fed-batch cultures with increasing dilution rates (Yang et al., 2003). It is very possible that there will be a range of conditions where no futile cycling is observed, and some conditions were it is. But, how could this be approached experimentally? One possibility could be to gradually upregulate ATP production under the control of an inducible promoter and evaluate the effect of this on central and overflow metabolism, and similarly gradually reduce ATP production, either using the soluble F1-ATPase which hydrolyses ATP (Holm et al., 2010), or through the supply of partial or absolute ATPase inhibitors (Dadi et al., 2009, Nakanishi-Matsui et al., 2016).

Another interesting observation from the WT was that there was no or very little flux flowing between metabolites α-ketoglutarate (AKG) and succinate (SUC), indicating that the tricarboxylic acid (TCA) cycle was incomplete. Chen et al. (2011) reported similar observations, while Nicolas et al. (2007) and Fischer et al. (2004) observed complete TCA cycles. To support their findings, Chen et al. (2011) performed an additional CO_2/HCO_3 labelling experiment and FBA modelling blocking flux through AKGDH and SUCOAS to zero, which showed no major effects on growth rate. Although I was unable to perform additional validation, their findings support the plausibility of this study's results. Furthermore, the fact that TCA reactions AKGDH and SUCOAS produce both ATP and NADH (which aerobically can be converted into additional ATP), raised the question of whether the lack of complete TCA activity may be reflecting the lack of oxygenation, noted previously due to the lower WT growth than expected (Section 6.2.2). When I measured the growth rate of the WT, it was found to be around 37% lower than the rate

predicted by pFBA, if this was simulated at the same glucose uptake rate as experimentally measured (Table 6.1). As Chen et al. (2011) also point out, lower oxygenation could be decreasing the rates through oxidative phosphorylation, which would reduce the regeneration of NAD^+ through the electron transport chain and subsequently lower production of additional NADH, hence the modest TCA cycle fluxes seen in the flux map through ATP and NADH-producing reactions.

However, these observations would require further validation through additional labelling experiments with increased oxygenation, measurements of gas exchange, and even transcriptional analysis as suggested by Chen et al. (2011).

Moving on to the PCK strain in Figure 6.6.B, metabolic flux analysis revealed that abolishing PCK activity increased the glucose uptake rate by 75% to 10.204 mmol gDW^{-1} hr^{-1}, but this did not translate into higher growth rate. The PCK knockout was estimated to produce lower biomass per glucose consumption, and presented increased glycolytic and PPP flux, though fluxes remained proportional to the change in glucose uptake, with upper glycolysis processing about 72% of total glucose, lower glycolysis constant at 175% total glucose, and PPP at 26-27% glucose, in line with WT. Previous observations from fed-batch cultures recorded PCK-PPC futile cycling of up to 8% of the total ATP produced in the WT, and a consequent decrease in flux through PPC and an increase in glyoxylate activity as a result of the PCK knockout (Yang et al., 2003). In contrast to these findings, no futile cycling between PCK and PPC was experimentally estimated in the WT, either by Chen et al. (2011) or in this work, neither was futile cycling captured by pFBA simulations in the WT in aerobic conditions. I also notice that the PPC flux doubled, while the TCA and glyoxylate activity decreased with respect to the WT.

The PCK knockout appeared to trigger higher PPC fluxes into the TCA cycle and upregulation of the malic reaction flux driving the flux back out of TCA towards pyruvate formation (PYR). Even though it was previously highlighted that futile-cycling could be condition-specific, this distinct cyclic behaviour, not captured in the WT strain, would simply not occur as a result of knocking out a gene allegedly inactive in the wild type. Therefore, although inconclusive, this observation suggested that there was also a possibility that the labeling experiment was not resolving PCK activity, resulting in the MFA data prep and analysis algorithms not being detecting the reaction in the WT. The use of alternative MFA softwares, such as FiatFlux (Zamboni et al., 2005), could help answer this question. The cyclic behaviour between phosphoenolpyruvate (PEP), oxaloacetate (OAA) and malate (MAL) also seemed to suggest that PCK activity is in fact more intricate and complex beyond mere futile cycling to waste ATP. After all, PCK is one of the key players controlling the concentrations of PEP and OAA. PEP controls glucose uptake by regulating the activity of the phosphotransferase (PTS) system. It also regulates the reactions catalyzed by phosphofructokinase (PFK), pyruvate kinase (PYK) and PPC. High PEP downregulates glucose uptake and upregulates pyruvate kinase, which produces ATP. Thanks to PCK, PEP can be regenerated gluconeogenically from the TCA cycle when its concentration drops. This is supported by the fact that PCK overexpression showed lower glycolysis activity (Kwon et al., 2014). Abolishment of PCK activity could be leading to a drainage of PEP (hence the higher glucose uptake, which is ATP-consuming) and an inability of PEP carboxylase to fulfil the anaplerotic balancing between PEP and OAA, thus

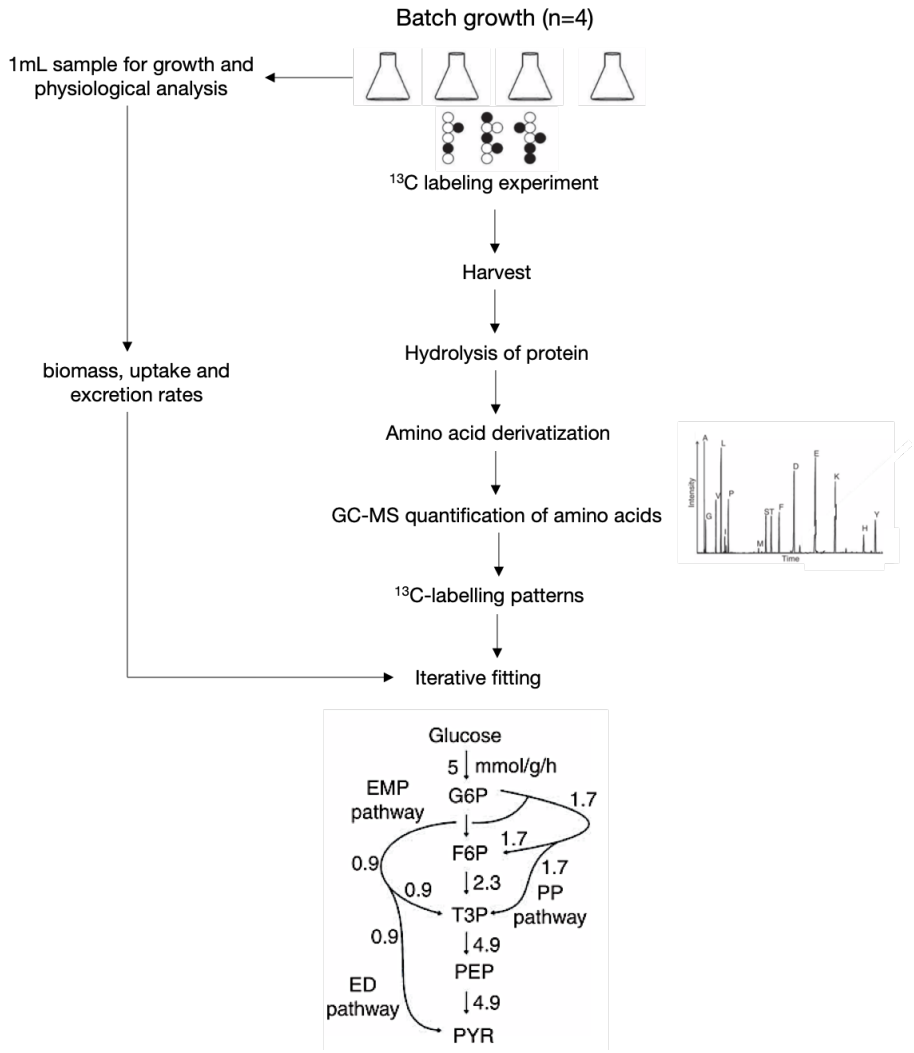


Figure 6.5: Workflow for ^{13}C MFA. [1–^{13}C] and [U–^{13}C] labelled D-glucose were selected as tracers for a 10h parallel labelling experiment. Meanwhile, 1mL samples were taken every 2h for growth and secretion product evaluation using high-performance liquid chromatography (HPLC). Labelled cultures were prepared by hydrolysing the protein and derivatizing the amino acids. Labelled amino acid fragments were detected using GC-MS and quantified. Finally, data was corrected and fed into WUFlux (He et al., 2016) for metabolic flux estimation, using the measured growth rate, glucose uptake and metabolite secretion rates as constraints. Goodness-of-fit measurements as well as confidence intervals for the internal fluxes were obtained (figure adapted from Zamboni et al. (2009)).

engaging the malic enzyme to restore metabolite concentrations via a non-ATP-producing route towards pyruvate.

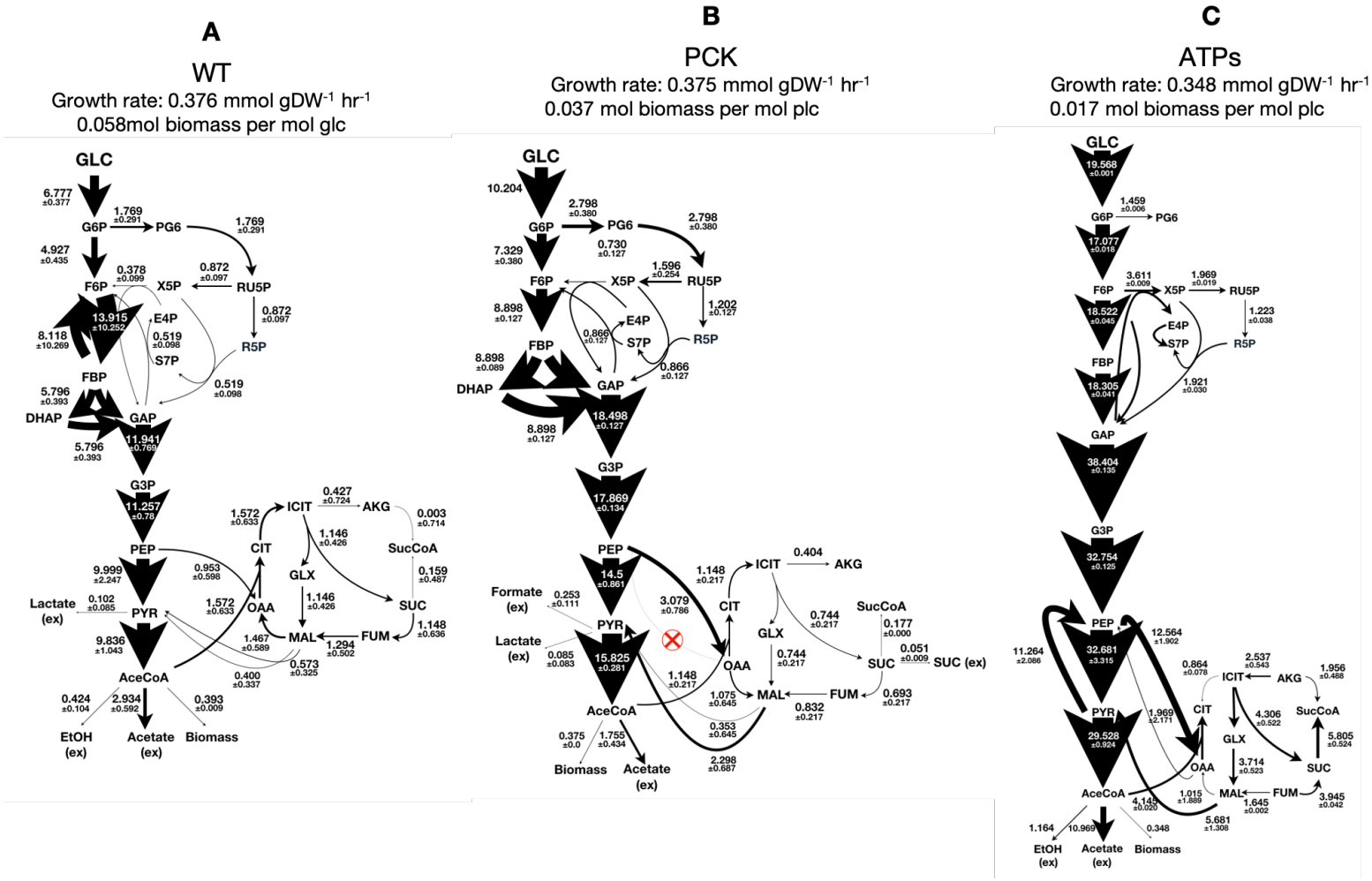


Figure 6.6: **^{13}C-flux maps of central carbon metabolism**. Determined from best fitting with 76 reactions and measurements of glucose uptake, byproduct secretion and growth rates from ^{13}C labelling experiments (n=4) as per Table 6.1. The WUFlux software was used for fitting (He et al., 2016). Values represent the best fit and standard deviation. The width of the black arrows is proportional to the estimated flux values. (A) Wild type; (B) PCK mutant, in which the PCK reaction was knocked out; (C) ATPs mutant, in which the ATP synthase reaction was knocked out. Abbreviations: ACCOA, acetate; AKG, a-ketoglutarate; CIT, citrate; ETOH, ethanol; E4P, erythrose-4-phosphate; CO2, carbon dioxide/formate; FUM, fumarate; GLX, glyoxylate; G3P, glyceraldehyde-3-phosphate; GLYC, glycerol; HIS, histidine; H6P, hexose-6-phosphate; Lac, lactate; MAL, malate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PYR, pyruvate; R5P, pentose-phosphates; SUCC, succinate; S7P, sedoheptulose-7-phosphate; 6PGC, 6-Phospho-D-gluconate.

This three-way cycle between PEP-OAA-MAL is suggestive of two propositions. The first proposition is that the ATP-consuming role of PCK is somewhat restored through the uptake of additional glucose (which is ATP-consuming), and the diversion of flux through OAA and MAL, with a consequent reduction in enolase activity (which is ATP-producing). Second, PCK appears to have a fundamental regulatory role that is not simply to waste ATP via futile cycling, but instead to regulate metabolite pools and anaplerosis. Indeed, these results suggest that there is an inherent interplay between these reactions as a result of the PCK knockout, suggesting a more significant role beyond mere regulation of energy pools. Yang et al. (2003) also suggested that PCK's role in balancing metabolite concentrations may be more significant than its role in regulating of the ATP/ADP ratio. Nonetheless, the inferences made from the present flux study cannot be concluded upon until they are further supported with measurements of metabolite concentrations and enzyme activity.

Finally, Figure 6.6.C displays the flux map of the ATPs mutant. This strain was characterised by a 3-fold increase in glucose consumption with respect to the wild type, and high glycolytic flux. Upper and lower glycolysis were about 15% and 20% higher in ATPs than WT, respectively, consistent with less efficient use of carbon for biomass production (growth). Furthermore, no or very little flux was recorded through PPP and the relative TCA cycle activity decreased by 15% in ATPs relative to the WT, indicating that the higher glycolytic flux in ATPs was mostly being released as acetate in ATP, instead of being redirected into the TCA cycle. Interestingly, nonetheless, the flux through the glyoxylate cycle was 22% of glucose uptake in the ATPs strain, compared to 17% in the WT. Acetate overflow was 17.2% of the total glucose consumed in the PCK strain with respect to 43.2% in the WT, a 40% decrease between these two strains, while it was 56% of the total glucose in the ATPs strain.

The faster operation of glycolysis in the ATPs strain also led to flux redistributions around key branch points. At the phosphoenolpyruvate and pyruvate branch points, which produce pyruvate from phosphoenolpyruvate via pyruvate kinase (PFK) and acetyl-CoA from pyruvate via the pyruvate dehydrogenase reaction (also known as PDH), the flux increased by 107% and 103%, respectively in the ATPs strain. This prominent increase in flux through these two branchpoints, which produce key metabolic precursors and building blocks for many bioproduction pathways was an exciting finding from a metabolic engineering perspective, suggesting that this strain's flux redirection could prompt an increase in flux for any pathways relying on these precursors as a starting point.

These observations are supported by the findings in Holm et al. (2010), which showed that hijacking ATP production by overexpressing the soluble F1-unit of the ATP synthase which constantly consumes ATP without pumping protons was shown to increase substrate uptake, glycolytic flux and decrease biomass production under aerobic conditions. Furthermore, it is somewhat not surprising that the flux map of ATPs somewhat resembles that of Chen et al. (2011)'s anaerobic *E.coli*. In the absence of oxidative phosphorylation, the ATPs strain effectively works as a pseudo-anaerobe, an *E.coli* strain that can be grown in oxic conditions but behaves like an anaerobic organism, with most of the pyruvate consumed as acetate to restore the energy balance through ATP production, byproduct release to regenerate the NAD^+ to allow central metabolism to continue (Ingram et al., 1987), and a remarkable increase in the

transhydrogenase flux to restore the NADH and NADPH ratios (Fuhrer and Sauer, 2009) (Appendix F Table 16).

6.2.4 ATPs and PCK as platforms for ethanol production in *E.coli*

Both the ATPs and PCK strains were characterised by higher glucose uptake rates, higher glycolytic activity, metabolic readjustments geared towards reduced ATP burning, and what appeared like a flux redirection to regenerate metabolite pools across key metabolic branch points. We know from Dugar and Stephanopoulos (2011) that pathways that are ATP-neutral or ATP-producing are deemed inefficient from a yield perspective because they generate excess ATP either through the pathways themselves or through surplus across the glycofermentative network, and any excess energy must be consumed elsewhere, often through either the activation of futile pathways or excess biomass production. I therefore wondered whether the ATPs and PCK strains could serve as attractive platforms for enhanced bioproduction of ATP-neutral and/or ATP-producing pathways for the production of targets that are more reduced than pyruvate, in order to create redox and energy coupling between target production and cellular growth.

To answer these questions, I turned to ethanol production as a proof-of-concept. Ethanol is the number one biofuel, used prominently in the transport industry as an environmentally friendly renewable additive (and potential alternative) to gasoline (Larsen et al., 2009). It is produced mostly from fermentation of conventional feedstocks, and its production is expected to reach 104bn liters in 2020 (International Energy Agency, 2018). Ethanol demand increased drastically between 1975 and 2003, and because of its potential to adversely affect feedstock availability, ethanol production soon became another landmark biotechnological reality (Ingram et al., 1999, 1987). Today, ethanol production is the main biotechnological application of yeast (Semkiv et al., 2016), but many other organisms have been tested for production at high productivity and yields (Zhou et al., 2008), so its pathway biochemistry is very well understood. For these reasons, it seemed like a sound starting point to gain further understanding of the behvaiour, growth profile and dynamics of the ATPs and PCK strains in a bio-production setting.

Therefore, in this section, a study was conducted to assess ethanol bioproduction capabilities in four different strains of *E.coli*: the WT strain (with no gene deletions), the ATPs strain (full knockout of the ATPase operon), the PCK strain (full knockout of the PCK gene) and a ATPs-PCK strain (a double knockout strain containing full gene knockouts of the ATPase operon and PCK gene). Because the final ethanol yield could be adversely affected by the release of alternative fermentation products, the excretion of organic acids formate, lactate, acetate and succinate was also monitored as a function of glucose consumption (Figure 6.7.A). Enhanced ethanol production was achieved by introducing the pyruvate decarboxylase gene from *Zymomonas mobilis* as an IPTG-inducible vector system, which was evaluated against an empty-vector control (Figure 6.7.B). In other words, two strategies were simultaneously evaluated to enhance ethanol overproduction, namely targeted gene knockouts of select genes and transformation of a heterologous plasmid system.

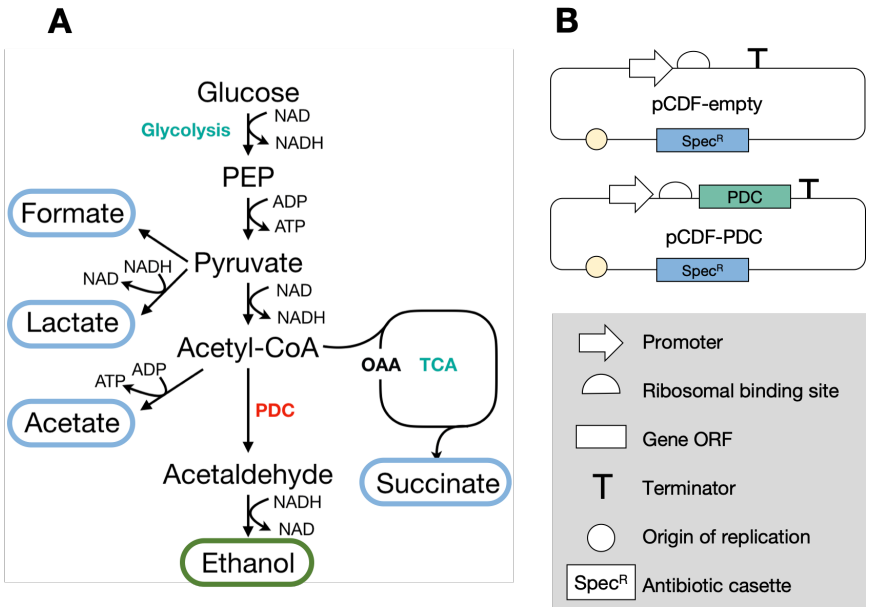


Figure 6.7: Biosynthetic pathway of ethanol in engineered *E. coli*. (A) Relevant metabolic routes for *in vivo* production of ethanol in *E.coli*, and relevant fermentation products monitored in this study. (B) Plasmid construct used for ethanol production (hereinafter known as PDC), and the empty vector control (hereinafter known as pCDF).

6.2.5 ATPs marginally improves ethanol bioproduction in *E.coli*

I began the ethanol bioproduction assay by first comparing ethanol production across the WT and ATPs strains. Briefly, to enable ethanol overproduction in *E.coli*, I electroporated the pCDF empty vector and PDC expression vector into the WT and ATPs mutant strains, as per Section 3.2.5. First, I wanted to understand the differences across the WT and ATPs strains of varying protein stability and its impact on ethanol production. I thus evaluated growth and ethanol production at varying levels of protein expression. Briefly, 3 biological replicates were grown in 25mL of M9 minimal medium with 2% (w/v) glucose at 37°C and 180 r.p.m. for 4h before inducing with 0.25mM, 0.5mM or 0.75mM IPTG and incubating for 48h at 30°C and 150 r.p.m.

Figure 6.8 shows the growth profiles (as OD_{600} over time), glucose consumption and ethanol production of the WT and ATPs harbouring the pCDF and PDC plasmids after protein expression with three different levels of inducer. The growth of the WT and ATPs strains did not seem to be significantly affected by the IPTG concentration. It was expected that the growth of the ATPs strain would be severely impacted during bioproduction, due its inherent mutation as well as the enhanced expression of a heterologous system. However, similarly to observations made during the labelling experiment previously (see Section 6.2.2), the difference in growth between the WT and ATPs was not as sharp as would have been expected, suggesting

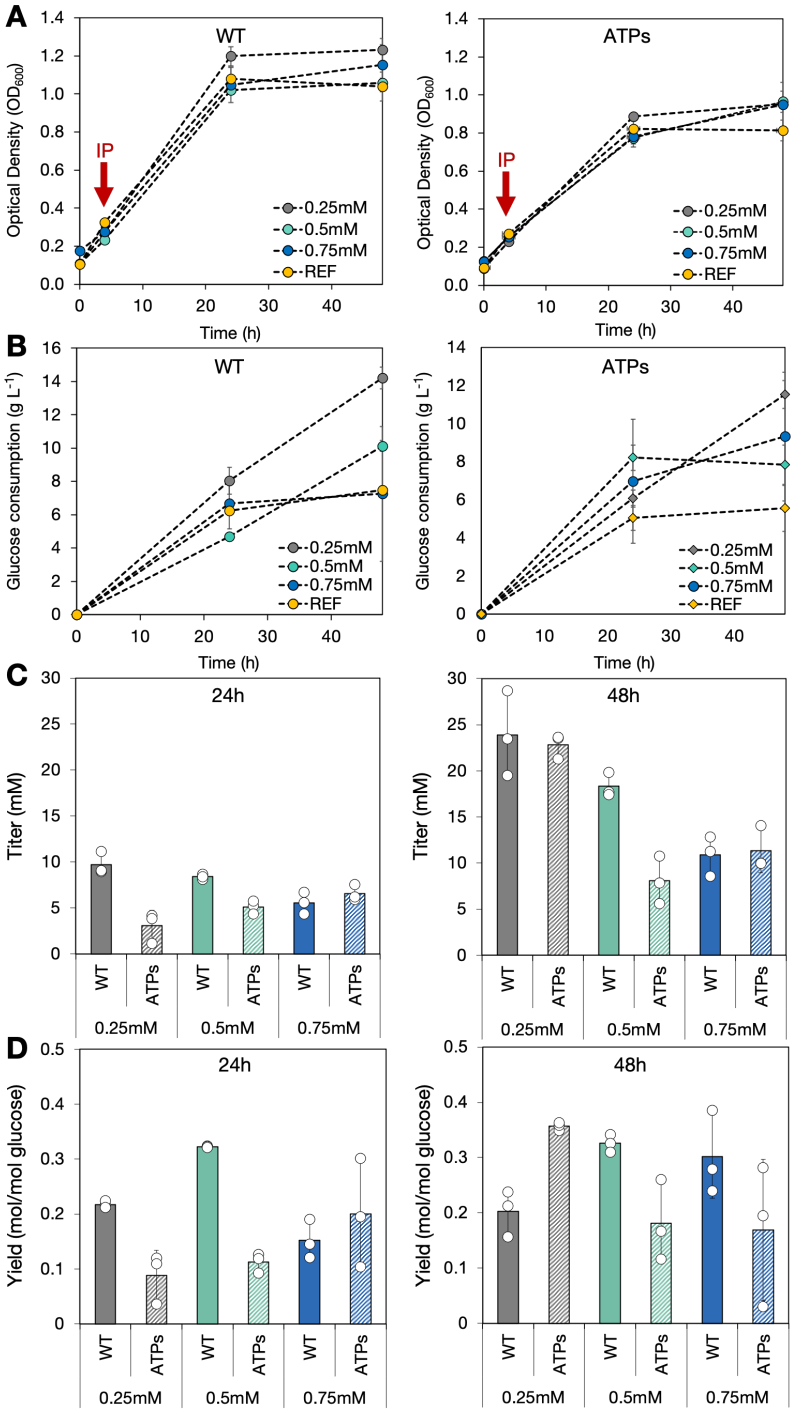


Figure 6.8: Protein expression evaluation of ethanol bioproduction strains WT and ATPs. Ethanol production was assayed by inducing 3 biological replicates with 0.25mM, 0.5mM or 0.75mM IPTG and cultivating them in M9 minimal medium + 2% (w/v) glucose for 48h. Samples collected 24h and 48h after induction. (A) growth profiles of WT and ATPs strains over the 48h incubation period. The red arrow indicates the Induction Point (IP), at the 4h timepoint. Empty-vector controls were also made but not included in the figure; (B) glucose consumption for strains in A over the 48h incubation period. To calculate glucose consumption, the glucose in the medium was measured at 24h and 48h, an subtracted from the initial glucose concentration in order to determine how much glucose was consumed at the relevant time points; (C) Ethanol titer (in mM) for strains in A at 24h and 48h timepoints. Titer measures the amount of ethanol in the medium; (D) Ethanol yield (as mol/mol glucose) for strains in A at the 24h and 48h timepoints. Ethanol yield indicates mol of ethanol in the medium per mol glucose consumed by the culture.

once again the possibility that the experimental setup could have suffered from a lack of oxygenation, which should be addressed in future studies. With respect to glucose consumption (Figure 6.8.B), it was somewhat not surprising to record considerable increases in glucose consumption upon expression of the PDC systems with respect to the reference strains, both in the context of the WT and ATPs strains, given that PDC overexpression would require the influx of additional carbon.

Ethanol excretion is shown as titer (mM in the medium, Figure 6.8.C) and yield (mol ethanol per mol of glucose consumed, Figure 6.8.D). The best performing ATPs strain was grown with 0.25mM IPTG at 48h, both in terms of titer (22.825mM) and yield (0.357mol/mol glucose). In contrast, the WT strain that produced the highest titer at 48h after induction with 0.25mM IPTG, but when the ethanol in the medium was normalized per mol of glucose consumed, the best performing strain in terms of yield was in fact the strain incubated with 0.5mM IPTG at 48h, so because this achieved a higher ethanol production per unit of carbon, I decided to proceed for further analysis using 0.5mM IPTG as the selected inducer concentration. It is also noteworthy that there were also clear physiological differences across the strains. This is illustrated by differences in growth in relation to ethanol production. The yields for the WT remaining fairly constant across the 24h and 48h timepoints, while the ATPs strain seemed to perform better (both in terms of yield and titer) at the 48h timepoint. This is likely stemming from the fact that ATPs grows considerably slower, so ethanol is more likely to accumulate over longer periods of incubation that the wild type.

Once the optimal inducer concentration was selected, I analysed ethanol production across the best producing WT and ATPs strains, in comparison to the reference and empty-vector strains. Figure 6.9.A illustrates the growth curve of the reference and empty-vector pCDF strains, as well as the WT and ATPs strains bearing the PDC system. Interestingly, the ATPs strains harbouring the plasmid systems grew better than the reference ATPs strain to some degree. Figure 6.9.B shows the glucose consumption of these strains over the 48h incubation period. Both the WT and ATPs strains harbouring the PDC system consumed considerably more glucose than the controls, as expected.

All strains of *E.coli* including the PDC system resulted in enhanced ethanol titer and yield with respect to the reference (no plasmid) strains and empty-vector controls (Figure 6.9). PDC-driven ethanol production in the WT strain reached a yield of 0.326 ± 0.016 mol/mol glucose and titer of 18.316 ± 1.310 mM after 48h, with respect to a yield of 0.130 ± 0.009 mol/mol glucose and titer of 5.401 ± 0.411 mM for the reference strain ($P \leq 0.005$). Expression of the PDC vector system into the ATPs strain boosted the yield and titer up to 0.357 mol/mol glucose and 22.825 ± 1.314 mM, respectively, with respect to the reference strain ($P \leq 0.005$). Ethanol yield in the ATPs strain was 9% higher than in the WT strain overexpressing PDC ($P \leq 0.01$) and reached 70% of the theoretical maximum (Yang et al., 2014). This illustrates that the ATPs strain can increase the product yield further than the wild type, and although results were statistically significant, improvements were small. Previous work by Yang et al. (2014) on the overexpression of PDC and ADHb for homoethanol expression in *E.coli* showed that while TCA and ethanol production proteins were dramatically upregulated in their PDC strain with respect to the wild type, glycolytic proteins were down-regulated, which was an interesting observation given that glycolysis is the main route towards ethanol

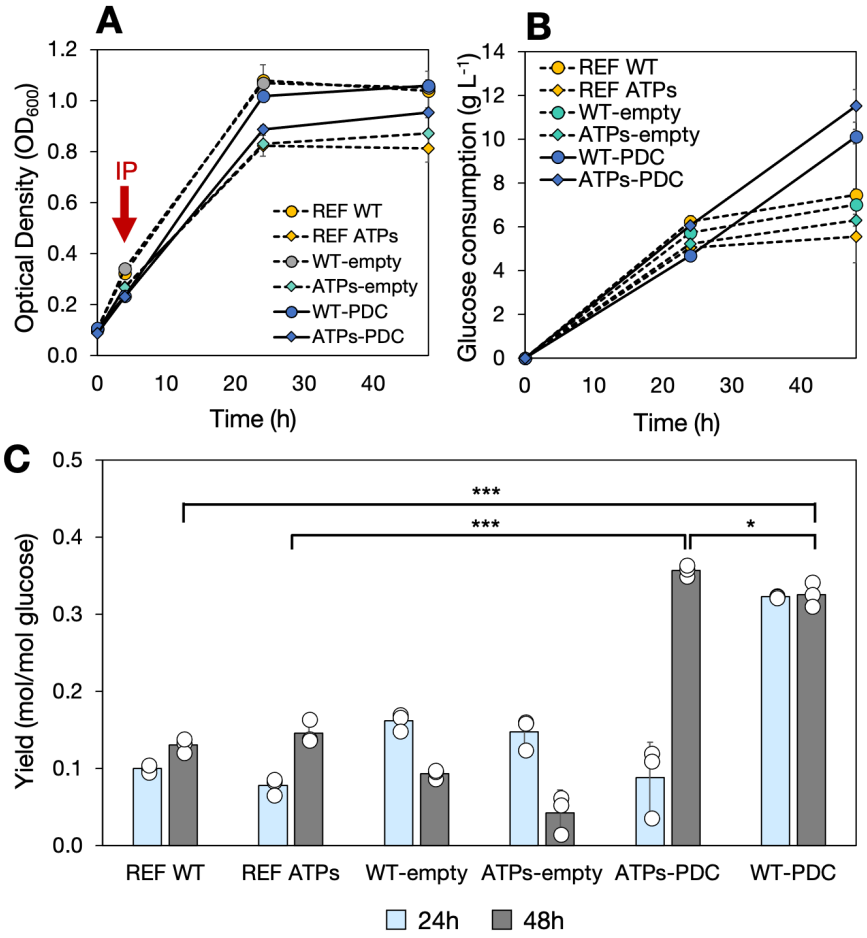


Figure 6.9: Ethanol bioproduction evaluation of best producing WT and ATPs strains. The WT strain bearing the PDC system was induced with 0.5mM IPTG, while the ATPs system was induced with 0.25mM IPTG at the Induction Point (IP) indicated by a red arrow at the 4h timepoint. (A) growth profiles, as OD_{600} over time (h), of WT-PDC, ATPS-PDC and their corresponding empty-vector and reference (no plasmid) controls. Non-induced controls were excluded. Strains were cultivated for a total period of 48h.; (B) glucose consumption for strains in A; (C) Ethanol yield, as mol ethanol per mol glucose, for strains in A. Samples were taken 24 and 48h after induction with the selected inducer concentrations. Asterisks indicate significant difference between independent samples (* $P \leq 0.05$; *** $P \leq 0.005$). Data is illustrated as the average from 3 biological replicates and the error bars represent the standard deviation across biological replicates of the same sample.

production. Similar observations were reported in yeast, where higher ethanol concentrations produced anaerobically from pyruvate inhibited glycolytic activity (Ghiaci et al., 2013). Nonetheless, Yang et al.

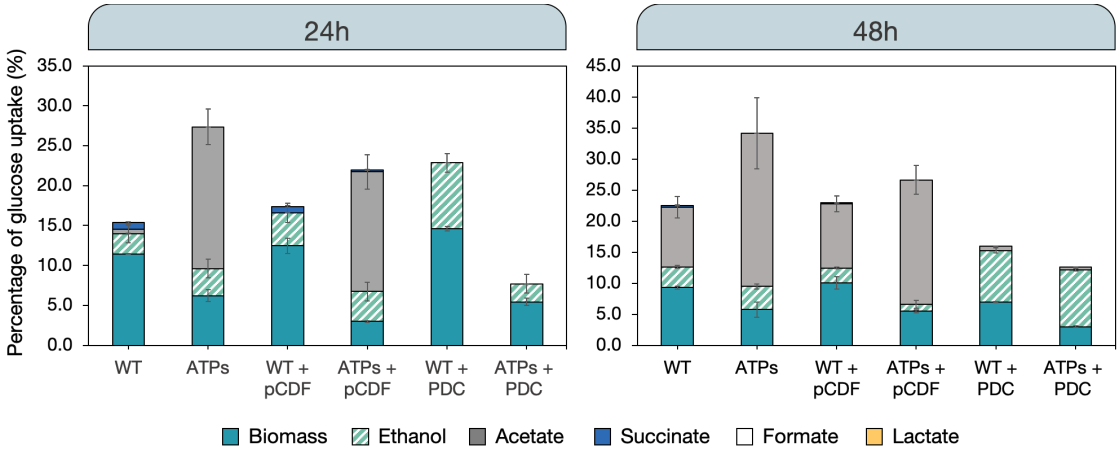


Figure 6.10: Carbon profile of best producing WT and ATPs strains. The WT strain bearing the PDC system was induced with 0.5mM IPTG, while the ATPs system was induced with 0.25mM IPTG. Secretion profiles were assessed by collecting samples every 24h and performing High-performance liquid chromatography to detect the most common fermentation products in *E.coli*: ethanol, acetate, succinate, formate and lactate. (A) product profile at 24h; (B) product profile at 48h.

(2014) achieved a 0.449 g/g glucose yield (88% theoretical yield), and propose to focus on the overexpression of glycolytic proteins to enhance ethanol production further. To our benefit, the ATPs knockout naturally boosts glycolytic activity without the need for further modifications. Nonetheless, I show in this study that the higher glycolytic flux in the ATPs strain only yielded marginal yield improvements. Yang et al. (2014) report 99 significant changes in protein expression during PDC-drive ethanol overproduction, including an upregulation of TCA activity as well as the coordination of changes needed for the cells to adapt to internal and external environments with high ethanol content. It is very likely that these changes are posing an added pressure on the ATPs strain, which already must handle its stringent energy metabolism requirements as a result of the ATPs knockout, which could be translating into marginal improvements in ethanol production.

I also wondered whether additional carbon was being dissipated by the ATPs strain, so I examined the mass balance i.e. products in the medium as a function of the glucose consumption (Figure 6.10). The first important observation was that while the ATPs control strains produced considerable acetate, as was expected from the earlier flux analysis, engineered ATPs strains overproducing ethanol did not, which was unexpected but not unreasonable. It is possible that the ethanol-producing ATPs did not produce any acetate because it simply did not need it, as overproduction of ethanol is ATP-neutral and enough ATP is possibly being generated by the higher glycolytic flux which guarantees biomass maintenance and renewal of relevant catalytic functions (Hädicke and Klamt, 2015). Furthermore, applying the theory from De Kok et al. (2012), the low but positive yields of ATP are sufficient to maintain cellular performance but not to allow for unrestricted growth, and this is illustrated by the smaller accumulation of biomass and higher

ethanol yield in the ATPs strain, compared to the larger biomass accumulation and lower ethanol yield of the WT. This indicates the effectiveness of the ATPs strain as a platform strain for overproduction of ATP-neutral pathways benefitting from reduced ATP turnover.
Most importantly, however, after performing a carbon balance based on the glucose consumed by the cultures, I noticed that only up to 30% of the carbon was being recovered (Figure 6.10). A number of suggestions can be made from these results. Firstly, it is common knowledge that some ethanol can be lost by evaporation during incubation, but that would certainly not make up for over 70% of the total carbon. CO_2 is one additional dissipation, but it would be surprising if these loses would explain the difference, especially in the ATPs mutant which did not present much oxidation through the TCA cycle. It was therefore very possible that something else was being lost and not being experimentally measured. In fact, Kihira et al. (2012) evaluated the effects on glucose metabolism of genetic mutations affecting the respiratory chain in *E.coli* and they showed that all of the strains excreted 2-oxoglutarate as a product, as well as pyruvate and acetate. Interestingly, the strain with the highest number of mutations also excreted glutamic acid. This study provides a number of useful clues as to the unique secretion profile we might expect from the ATPs strain. Noting from Figure 6.6.C that the ATPs mutation results in only partial TCA and PPP activity, the secretion of acids such as fumarate, citrate, and oxaloacetate might also be crucial to track either by HPLC or LC-MS techniques in the future. If the ATPs strain, which effectively already behaves like an anaerobe, can effectively perform bioproduction under anaerobic conditions, an additional alternative could be to perform the fermentation in serum bottles, which would eliminate the chances of loss of any products, and also allow for the CO_2 to be quantified.
Albeit with a considerable number of areas for troubleshooting and avenues for future work, these results are exciting nonetheless. It was possible to conclude that deletion of the ATPs operon significantly increases target production, and although only small improvements were achieved within the context of ethanol, this strain may be an even better platform for ATP-producing biosynthetic pathways, which may couple target production to additional ATP release for enhanced cellular maintenance. Although unexplored in this study, acetate overproduction could be one such target to benefit from this strategy, given that the ATPs strain already naturally diverts a vast amount of carbon towards the production of this compound.

6.2.6 The PCK knockout significantly improves ethanol production in *E.coli*

Given the small improvement in ethanol yield observed with the ATPs strain, I then asked whether the deletion of the ATP-consuming PCK reaction would have an even more significant effect on ethanol bioproduction. To date, the role of PCK on cofactor manipulation has been studied within the context of futile cycling only, by overexpressing PCK in conjunction with PPC. This particular futile cycle has previously shown to increase biofuel production in cyanobacteria (Erdrich et al., 2014). Similar observations were made using *Saccharomyces cerevisiae*, which accumulated more ethanol compared to the wild-type strain during alcoholic fermentation on glucose (Semkiv et al., 2016). With the aim to study the metabolic effects of knocking out this gene, this PhD study found that the PCK knockout appeared to also decrease the

net ATP flux by increasing ATP-consuming glucose uptake and reducing flux through ENO. I therefore asked, could the knockout of the PCK reaction positively influence ethanol bioproduction? To answer this question, I used the PCK strain and also knocked out the PCK gene in the ATPs strain using already outlined methodologies (Figure 6.2, and Section 3.2.5 for more information), which resulted in a double mutant hereinafter known as ATPs-PCK (Figure 6.11).

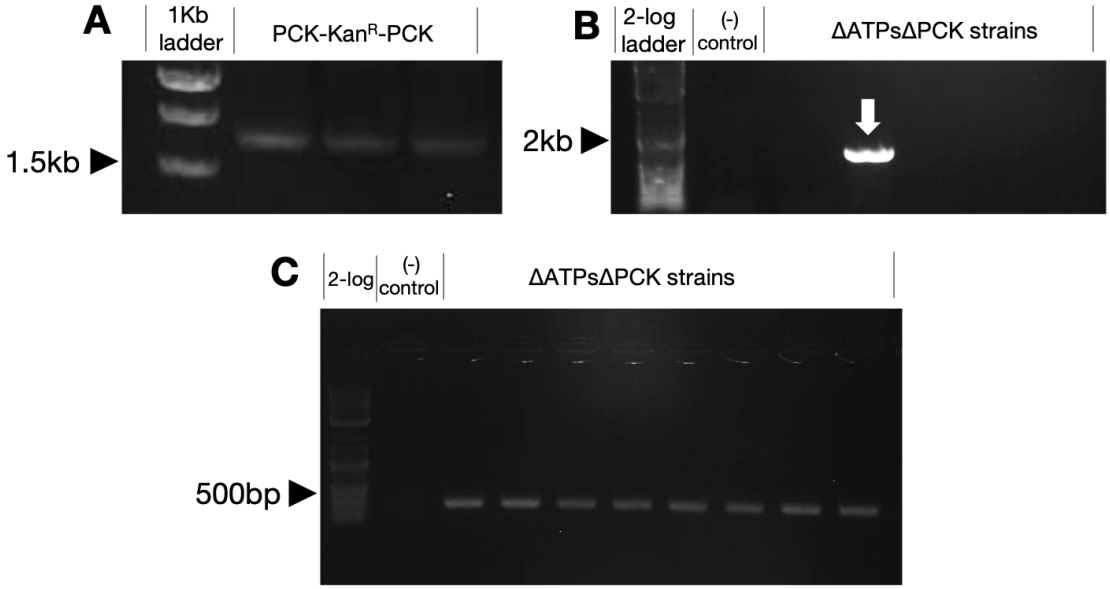


Figure 6.11: ATPs-PCK double mutant implementation. (A) Amplification of pKD13's Kan^R cassette using recombination primers that include a 50-bp homology arm complementary to the 3' and 5' ends of the PCK gene (100% amplification efficiency, n=3); (B) after electroporation of the PCR products in A into *E.coli*, dreamtaq amplification was used to check that the Kan^R cassette was successfully inserted to replace the PCK gene (20% electroporation efficiency, n=10); (C) Q5 PCR amplification of the PCK knockout after removal of the Kan^R cassette. Positive bands (300bp) found in all identified hits (100% efficiency, n = 8). PCR amplification and gel electrophoresis were followed by gel extraction for further DNA sequencing (Appendix D Figures 7-9).

In an analogous experimental procedure as previously outlined, mutant strains (PCK and ATPs-PCK) bearing the PDC plasmid and their negative controls were cultivated in in M9 minimal medium with 2% (wt/vol) glucose, and induced with three different concentrations of IPTG inducer to assess the effect of protein expression on bioproduction. Results are illustrated in Figure 6.12.

A number of observations can be made. Firstly, cellular growth of the PCK strain appeared to be enhanced by the addition of 0.75mM IPTG (Figure 6.12.A, PCK panel), while all other PCK strains, as well as the ATPs-PCK double mutants did not appear to be significantly affected by the IPTG concentration, as far as growth was concerned (Figure 6.12.A , ATPs-PCK panel). Similarly, glucose consumption of the PCK strain grown in 0.75mM IPTG also increased considerably (Figure 6.12.B). Ethanol titers (mM) are shown in Figure 6.12.C. Judging by the titer, as well as growth and glucose consumption, it appeared

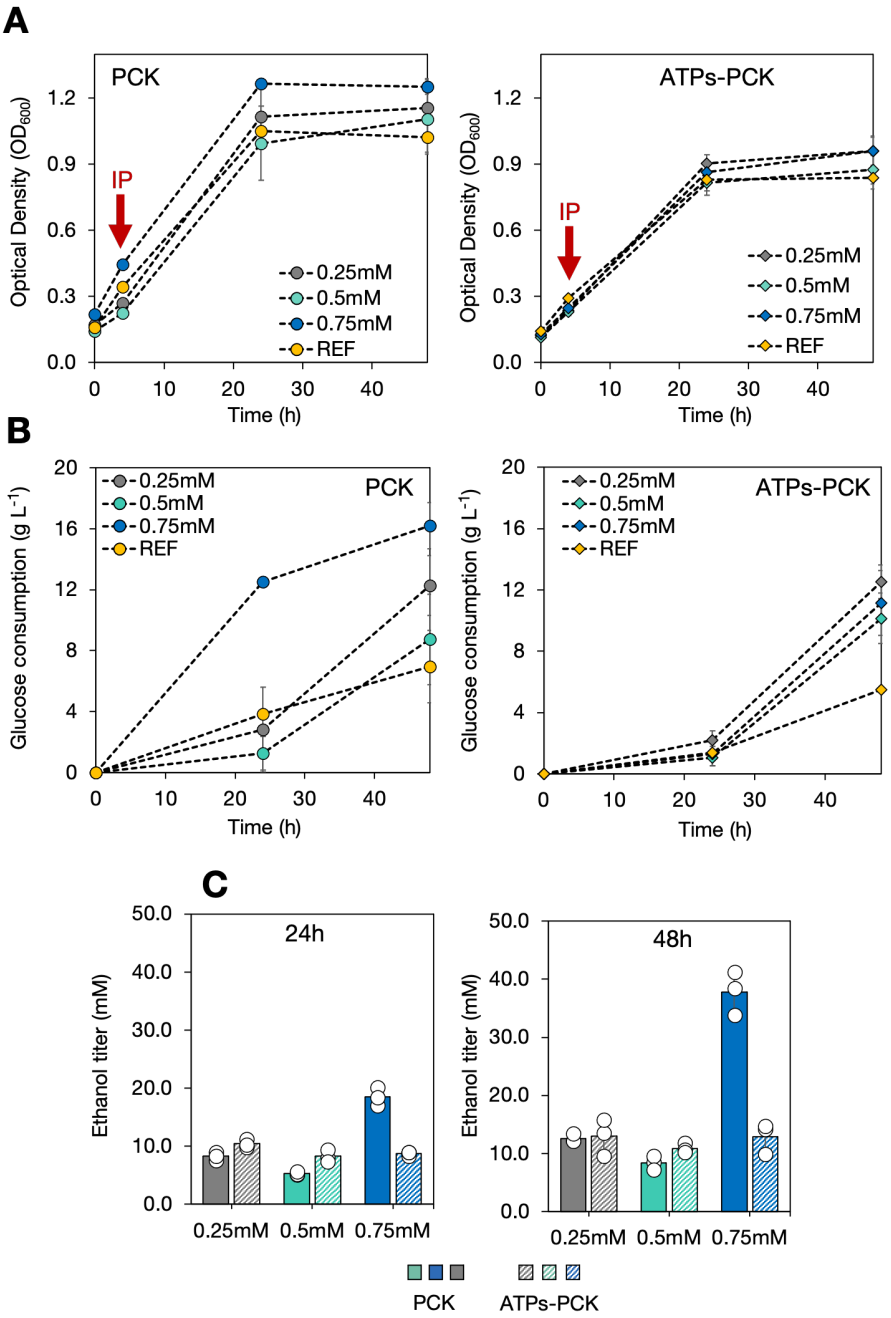


Figure 6.12: Protein expression evaluation of ethanol bioproduction in the PCK and ATPs-PCK strains. (A) growth profiles of PCK (single-knockout strain) and ATPs-PCK (double-knockout strain) induced with 0.25mM, 0.5mM or 0.75mM IPTG. The red arrow indicates the Induction Point (IP), at the 4h timepoint. Empty-plasmid controls were made but not included in the figure; (B) glucose consumption for strains in A; (C) Ethanol production by introducing enzyme PDC into the PCK and ATPs-PCK strains and overexpressing this at with 0.25mM, 0.5mM and 0.75mM IPTG. Measurements were performed at 24h and 48h.

that 0.75mM was the optimal concentration tested for ethanol production in the PCK strain. In contrast to this, the three inducer concentrations tested in the ATPs-PCK double mutant strains did not have a significant impact on ethanol production, so I selected the smallest concentration (0.25mM) to proceed with the analysis.

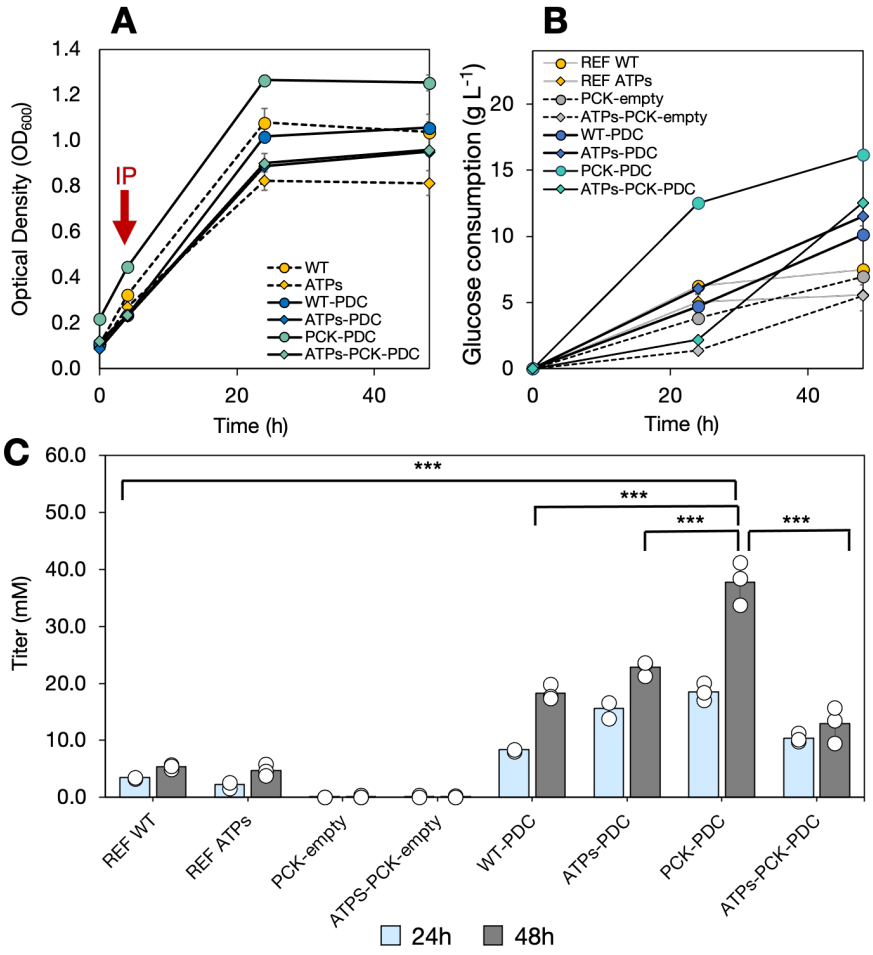


Figure 6.13: Ethanol bioproduction evaluation of best producing PCK and ATPs-PCK strains. The PCK strain was induced with 0.75mM IPTG, while the ATPs-PCK system was induced with 0.25mM IPTG. The PCK and ATPs-PCK strains bearing the PDC system were compared against their empty-vector and reference (no plasmid) controls. Non-induced controls were not made. All strains were cultivated in M9 minimal medium and 2% (w/v) glucose for a total incubation period of 48h. Samples were taken 24 and 48h after induction. (A) Growth profiles, as OD_{600} over time (h). (B) glucose consumption for strains in A; (C) Ethanol titer (mM) for strains in A. Asterisks indicate significant difference between independent samples (* $P \leq 0.05$; *** $P \leq 0.005$). Data is illustrated as the average from 3 biological replicates and the error bars represent the standard deviation across biological replicates of the same sample.

After selecting for the optimal IPTG concentrations, I analysed the growth profile, glucose consumption and ethanol production across the best ethanol producers (Figure 6.13). I noticed that the 24h glucose consumption timepoints for some of the PCK measurements were generally too low across some of the samples, which led to the identification of technical errors with the HPLC glucose standard during one of the HPLC runs. Hence, to avoid any mismeasurements, I analysed these results as titer measurements only (in mM). Figure 6.13 shows the growth profile, glucose consumption and ethanol titers of the PCK and ATPs-PCK ethanol-producing strains, compared to their reference (no plasmid) strains and empty-vector controls. These results were also compared to the ethanol producers from Section 6.2.5 which lacked the PCK knockout. Once again, all strains harbouring the PDC system resulted in enhanced ethanol titer with respect to the reference (no plasmid) strains and empty-vector controls. However, the PCK knockout significantly improved ethanol production, amounting to a titler of 37.824 ± 3.764 after 48h, compared to 18.316 ± 1.310 mM in the WT strain and 22.825 ± 1.314 mM in the ATPs strain ($P \leq 0.005$). In contrast to this, the ATPs-PDC double mutant adversely affected ethanol production, recording the lowest titer out of all engineered strains, specifically 12.949 ± 3.156.

These results were exciting, because although Zhou et al. (2008) reported 44.5 mM ethanol, 38 mM were produced in this study using a strain bearing a single gene mutation and a single plasmid for PDC over-expression, with no further modifications to the strain's fermentation product profile. The above assay supports earlier suggestions of the role of the PCK reaction in *E.coli*'s central metabolism being more complex than mere regulation of futile cycling. Interestingly, referring back to Figure 6.6.B, flux redirection in the PCK strain not only reduces ATP availability by increasing glucose consumption and diverting flux away from glycolysis into the TCA and malic reactions, it also appears to be making additional redox readily available for bioproduction, which is a limiting factor in ethanol production. This additional redox is likely to be driving ethanol production forward. Although inconclusive until further metabolite assessments are performed, it is possible to infer from the flux analysis results that the PCK knockout may also be causing changes in key metabolite pools, such as PEP, OAA and pyruvate, in a way which is favourable for ethanol production. It is undeniable that PCK and potentially other anaplerotic reactions such as PPC and NADH and NADPH-specific malic reactions have a fundamental role in regulating metabolite pools, and may potentially respond very flexibly not only to changes in cofactor demands but also to the availability of key cellular metabolites.

6.3 Chapter summary

In this study, I began by evaluating the coordinated roles of ATP and NADH and the presence of futile cycling in *E.coli* by evaluating the physiological and metabolic flux responses to two targeted genetic manipulations affecting *E.coli*'s native cofactor pools: the ATP synthase operon and the PCK gene.

While futile cycling between PCK and PPC was not recoded in the wild type, the predominant phenotypes of *E.coli* harbouring both the ATPs and PCK knockouts were characterised by high glycolytic flux at the expense of biomass production. The ATPs strain increased its acetate overflow to restore energy

balance, in line with previously reported studies (Hara and Kondo, 2015, Holm et al., 2010, Kihira et al., 2012, Koebmann et al., 2002). The PCK strain reduced its ATP flux and made additional redox readily available. Furthermore the cyclic behaviour between phosphoenol pyruvate, oxaloacetate, malate and pyruvate suggested that PCK activity is more intricate and complex than mere ATP modulation through futile cycling.

These cofactor manipulation strategies were evaluated within the context of ethanol bioproduction as a proof-of-concept. All strains including the PDC system for enhanced ethanol production resulted in enhanced ethanol titer and yield. Ethanol yield in the ATPs strain was 9% higher than in the WT strain overexpressing PDC ($P \leq 0.01$) and reach 70% of the theoretical maximum (Yang et al., 2014). As suggested by (De Kok et al., 2012), the low but positive yields of ATP in this strain were sufficient to maintain cellular performance but not to allow for unrestricted growth, illustrated by the smaller accumulation of biomass and higher ethanol yield, compared to the larger biomass accumulation and lower ethanol yield of the WT. The strain harbouring the PCK knockout significantly improved ethanol production up to 37.8 mM, likely benefiting from the flux redirection away from glycolysis and into the TCA and malic reactions, which provides the additional redox necessary for ethanol formation. Compared to 44.5mM ethanol recorded in the landmark study by Zhou et al. (2008), the PCK strain bearing only one genetic mutation and the introduction of one inducible plasmid for overexpression of PDC,

This study is supported using readily available data and demonstrates that significant improvements can be made with very few, targeted modifications. Certainly though, mainly question remain unresolved, namely, do NADH and ATP share regulatory control? Does oxygen availability regulate metabolism beyond influencing the presence of ATP? How and why is the ATPs strain dissipating so much carbon? Nonetheless, and albeit through small improvements, these results show the effectiveness of the ATPs strain as a platform strain for overproduction of ATP-neutral pathways benefitting from reduced ATP turnover, and supports, as per Hädicke and Klamt (2015), that even better yields may be achieved if the product synthesis pathways are ATP-producing, since a higher fraction of the substrate must be directed to the product to generate the ATP needed for biomass synthesis.

7

Conclusions and Future Work

7.1 Summary

How can we better harness cofactor metabolism to improve bioproduction?
Metabolic engineering represents one of the cornerstones of sustainable industrial production since its inception in the final decades of the 20^{th} century, and for many more decades to come. Traditionally, the focus has been on the development of engineering strategies to maximise the carbon flux towards the chemical target. However, carbon metabolism alone cannot achieve efficient bioproduction, because it requires the appropriate redox and energy supply as well as the effective regeneration of such cofactors. Using computational and experimental strategies, butanol and ethanol as case studies, and *E.coli* as a model system, this study proposes assessing the impact and limitations of cofactor usage on bioproduction at an early stage, and uses selected cofactor manipulations to develop platform strains to improve bioproduction. In Chapter 4, I began by evaluating how two existing frameworks enable the study of ATP and NAD(P)H metabolism and the limits these cofactors impose on bioproduction. The analytical calculations developed by Dugar and Stephanopoulos (2011) were used to adjust the final yields of eight different butanol and butanol precursor pathways based on their sources of cofactor imbalance. Secondly, using the same biosynthetic pathways, I used pFBA to maximise butanol production in *E.coli*. pFBA yields represented the maximal production capabilities of the *E.coli* metabolic network, after accounting for any limits imposed by cofactors on the rest of the network. The method developed by Dugar and Stephanopoulos (2011), although effective for pinpointing the source of imbalance, was pathway-specific, not easily implemented, and unable to account for environmental cues or host selection. Alternatively, pFBA, although an easy-to-implement and powerful alternative, did not alone suffice to easily discern cofactor usage differences across

biosynthetic designs.

Consequently, in Chapter 5, I designed the Cofactor Balance Assessment (CBA) protocol, a COBRA-compatible framework that tracks ATP and NAD(P)H metabolism across metabolic engineering designs. Building on Chapter 4, CBA was tested using butanol and butanol precursor production and the *E.coli* Core model to evaluate how variations in energy and redox demands affect butanol production. This study highlighted the excessively underdetermined nature of the *E.coli* metabolic network, characterised by significant futile cycling involving cofactor-consuming reaction pairs. As a result, I evaluated three curation methods. First, I manually capped unrealistic futile cycling. Manually-curated results confirmed that pathways that divert the least amount to biomass are more cofactor-balanced, as these pathways reduce the burden placed on the rest of the metabolic network. After unsuccessfully running loopless FBA, I then used fluxomic data to apply biologically plausible flux ranges as constraints to reduce the solution space of the models. Experimentally-constrained solutions more closely resembled the yield estimates calculated by the method in Dugar and Stephanopoulos (2011), suggesting that the use of experimental constraints, in combination with some manual capping if any futile cycles remain, may be a quicker and more suitable approach to obtain flux estimates that account for competing pathways, intracellular complexity and cofactor demands across the wider network. Finally, in this chapter I also present a complementary tool that evaluates the three-dimensional relationship between target yield, ATP and NADH usage for more accurate pathway ranking. This framework identifies the cofactor-to-yield sweet spot and yields insights into its stability. This analysis helps metabolic engineers identify how far away from the optimum their selected target pathway is, and the chances that a high-yielding system can materialise.

As well as the relevant role of futile cycling in restoring cofactor balance, CBA also highlighted that ATP and NAD(P)H balance cannot be assessed in isolation from each other. Thus, in order to investigate the existence of futile cycling and the energy-redox connectivity in *E.coli*, in Chapter 6 I performed a fully experimental analysis in which I first evaluated the physiological and metabolic effects of knocking out the ATP synthase and PCK genes on E.coli's central metabolism. Although wild type *E.coli* did not show any futile cycling between PCK and PPC, both knockout strains were characterised by higher glucose uptake rates, high glycolytic flux at the expense of biomass production, lower ATP fluxes and considerable rewiring of metabolic reactions across key metabolite pools. Assuming that higher glycolytic flux, lower ATP production and readily available redox would make up a suitable phenotype for the production of ATP-neutral chemical targets, I evaluated the biotechnological performance of these strains using ethanol bioproduction as proof-of-concept. All strains incorporating a plasmid system including the IPTG-inducible PDC gene resulted in enhanced ethanol titers and yields. Although the ATPs strain increased ethanol production by 9% with respect to the wild-type and reached 70% of the theoretical maximum, the strain harbouring the PCK knockout produced significantly more ethanol. The lower but positive yields of ATP in both of these strains were sufficient to maintain cellular performance but not to allow for unrestricted growth, illustrated by the smaller accumulation of biomass and higher ethanol yield, compared to the larger biomass accumulation and lower ethanol yield of the WT.

7.2 What new contributions to knowledge have been achieved?

Combining pathway and host engineering has had a great impact on the generation of effective biocatalysts, but it is undeniable that improving energy and redox chemistry through cofactor manipulation has the potential to considerably improve bioproduction. Fundamental studies of pathway fluxes and global responses to cofactor manipulations have provided further essential information for better understanding the regulatory roles of cofactors, and advances in systems engineering on a larger and more precise scale have additionally contributed to improving the availability of the desired cofactors necessary improved bioproduction.

Undoubtedly, the CBA protocol remains the striking novelty and primary contribution of this work. The main objective was to integrate the need to consider cofactor-related information with the identification of the most efficient pathway(s) for chemical production at the early stages of any metabolic engineering project. With CBA, I propose an exploratory method that can aid with pathway selection by quantifying the extent of cofactor imbalance and discerning how organisms handle their pre-existing redox and energy imbalances. Additional tools are facilitated to help users understand how to improve a system's cofactor usage profile. The method's foundations and results were initially presented in May 2017 at the Data-Driven Biotechnology international conference in Hillerod, Denmark and later published on a peer-reviewed journal (see Publications Section at the start of this thesis).

Earlier work had explored the effects of cofactor swapping on the global network using COBRA-based modelling (Ghosh et al., 2011, King and Feist, 2014). Further studies were aimed at depicting cofactor balance profiles to explain and interpret experimental data (Chen et al., 2011, Garcia Martin et al., (2015, He et al., 2016). Instead, the overarching aim of CBA is to elucidate, prior to experimental implementation, a system's cofactor contributions to biomass, target, maintenance and waste release, and pinpoint potential areas for engineering in response to particular sources of imbalance before adventuring into the wet laboratory. Pathway and host selection is thus based on knowledge of cofactor balances of the best possible pathway solution, not just of the highest possible yield. Only three elements are used in this modelling framework: (i) the stoichiometry of cofactor-related metabolic reactions, (ii) measured intracellular fluxes, and (iii) an optimisation principle. Cofactor balance can then be expressed as summed fluxes of consumed and produced ATP and NAD(P)H contributing to biomass, waste, maintenance or target formation. This method also hinges on a very simple assumption: organisms strive to be redox and energy neutral. This assumption is not only at par with existing modelling frameworks (King et al., 2015, Orth et al., 2010, Wiechert, 2001), which are based on the assumption of steady-state, but it is also widely supported and accepted in the community (De Kok et al., 2012, Dugar and Stephanopoulos, 2011, Varma et al., 1993).

CBA also emphasised the need to consider the coordination between ATP and NAD(P)H, thus pointing towards the need to close the gap in our understanding of both energy and redox in combination, and not as isolated entities, as they have been mostly studied so far. Furthermore, CBA also highlighted the relevant role of futile cycling (at least under *in silico* conditions) to regenerate any surplus cofactors. I took these two features and evaluated them in parallel in a fully experimental study which, although presented at its

early stage in this PhD study, is worthy of consideration because it brings some interesting contributions. Firstly, I concluded that, under the pseudo steady state conditions tested, PCK and PPC did not appear to form a futile cycle in *E.coli*, confirming earlier literature that this ATP-burning cycle might be tightly regulated and condition-specific (Russell, 2007, Russell and Cook, 1995, Yang et al., 2003). Nonetheless, the distinct phenotype of the PCK knockout strain, characterised by the rewiring of reactions controlling key metabolite pools, highlighted the inherently complex role of the PCK gene, which seems to expand beyond mere futile cycling, and certainly warrants further investigation. Finally, I report that the ATPs strain increases ethanol production yields by 9% with respect to the wild type engineered to overproduce ethanol, and reaches 70% of the theoretical optimum, with no further genetic modifications added. PCK considerably improved the ethanol titer compared to ATPs, and was the best producer in this study. Although the characterisation of these strains was confined solely to physiological and flux analyses and they were only tested for the production of ethanol within the scope of this study, these results indicate the potential of these strains as platforms for the production of cofactor-neutral or cofactor-surplus biosynthetic pathways.

7.3 Reflection on the methods developed

Although this study presents a number of exciting contributions, which have been highlighted in Section 7.2, many limitations still exist.

7.3.1 Limitations of the CBA method

The CBA protocol is currently centred on the use of flux balance analysis (and variations thereof, such as pFBA, FVA MOMA), which is based on linear optimisations. As I show in this PhD study, FVA was performed to test the size of the systems' solution space (Appendix A Table 1), and even though theoretically many solutions may exist, as it turns out, regardless of what the objective function was, no or very little variation amongst the individual reactions was found with most of the models. Two of the models displayed variation in 15% of the reactions (14/99 or 18/101) but none of these reactions were directly in the pathway towards butanol, and all were directly involved in futile cofactor cycling. The limited solution space in these models is most likely due to the fact that a minimal core model was used, and because the focus was placed on the optimal solutions (not suboptimal solution spaces). Hence, for the purpose of this work, the single solution provided by FBA was sufficiently representative and simpler to use, and once constraints from MFA or manual elimination of futile cycling was applied, there only ever was a single solution. However, although I highlight the strength of the CBA approach and its usefulness in integrating with any metabolic model, it may become increasingly difficult for estimates to be informative as models increase in size and complexity. This is not an issue when applying the analysis using a core model, as I already show, but it can certainly become an issue when dealing with models that do not result in unique solutions. In these alternative scenarios, it would be better for CBA to be compatible with

FVA solutions instead. However, in its current form, CBA does not allow for this as it is only capable of processing a flux vector, and not a range of fluxes.

Although the methodology is focused on the use of flux balance analysis, solutions were subject to extensive manual and systematic curation of reaction constraints. Arguably, the manual capping procedure to eliminate futile cycles may be perceived as somewhat problematic, because it is capping the fluxes with values from the wild-type flux distribution, which was obtained by maximizing biomass production, which would narrow the solution space to be more akin to the predicted fluxes under growth metabolism. This approach also potentially excludes other reactions that could be used but would not be active the condition of maximum biomass production. Indeed, no curation strategy will ever be perfect, as it will be based on human judgement. However, this approach represents a reasonable compromise as it still allows for biomass synthesis (essential to synthesise the biocatalyst), and indeed when the artificial cofactor cycles were minimised, new cycles appeared, and eventually biomass was the only final sink for excess ATP, thereby connecting the present study with earlier theories (De Kok et al., 2012, Dugar and Stephanopoulos, 2011) where biomass was assumed to act as a sink for excess ATP. In any event, in reality, biomass and bioproduction compete for resources, so there will always be a trade-off between growth and bioproduction metabolism, and this may entail somewhat sacrificing mathematical accuracies to reach theoretical outcomes which represent a fair compromise of what would be expected also in experimental conditions.

As an alternative to manual curation, I also explored the use of experimentally-derived flux data to constrain the FBA solutions to more biologically plausible regimes. In this ambit, arguably the reliance on MFA data from multiple mutant strains could make the approach only available for *E. coli* or yeast as host strains, and may instead limit the approach to other organisms that may not have been as well characterised, or may not be as straightforward to cultivate and be subject to labelling a flux analysis studies. In this respect, the aim of the present study was mainly to highlight some of the underdetermination issues with COBRA in the absence of constraints, in particular the appearance of futile cycles which cannot be overcome with alternative tools such as loopless FBA. The use of MFA data was useful to get an idea of how a synthetic pathway could be limited in different hosts, but in the case that MFA data is not available, manual curations or MOMA could be considered instead. The ultimate goal is that other researchers can carry out CBA analyses with or without MFA data as constraints, or alternatively plug in any alternative curation methods which they consider suitable.

A further limitation of the CBA protocol could be that NADPH and NADH are analysed interchangeably. For modelling purposes, CBA pools NADH and NADPH fluxes to present an overall "redox" category. In reality, however, NADH and NADPH are biologically and chemically distinct. Biologically speaking, NADH and NADPH are used in very distinct areas of metabolism and under distinct physiological states. Chemically, they are produced via different means, and they are used up by different enzymes, meaning a reaction that is NADH-specific will not be active under high NADPH concentrations unless its specification is fine-tuned with enzyme engineering. Therefore, the distinct nature of these cofactors will have a direct impact on metabolism. Computationally, however, their usage within the model will still continue to be very much dependent on which cofactor is used by which reaction, i.e. NADPH generated by the PPP will

be captured by CBA, so in fact, the models and their respective solutions are in fact not affected by this joint categorisation of redox metbolism, but it is of course, something that the user should always keep in mind. Similarly, apart from ATP and NAD(P)H, no further cofactors are considered in this study, which does limit its wider application, if alternative cofactors are of interest to the user.

Finally, it is important to bear in mind that the current analysis is a theoretical analysis, performed in a theoretical manner which relies on the usage of stoichiometries recorded in the GSMs. As a result, many model parameters, such as the accuracy of biomass equations, reaction directionality, cofactor specificities of enzymes, and any further constraints applied during the implementation and modelling stages will significantly influence the internal flux distributions, and thus the cofactor profiles of the networks under analysis. It is therefore vital that the accuracy of the input model(s) and parameters is questioned and evaluated prior to modelling. In this particular study, for instance, it was concluded that the curated *E. coli* core model is a very well-established and simple model, and one that has been used by many others previously. Arguably, it is possibly one of the most accurate models available to query central metabolism, given that increasingly complex models shift the focus away from central metabolism and are more likely to also incorporate reactions that do not reflect reality. Indeed, the *E. coli* core model lacks some of the metabolic features in other systems, but this does not reduce its value or the value of the principles evaluated with it, which should be applicable also to other larger and more complex models. Whenever a new concept is evaluated, I would argue that it makes most sense to commence with the simplest, most well-understood system available, which will likely have been questioned, curated and informed by a large number of players in the community. In order to clarify any possible ambiguities, it is possible to always elaborate further by incorporating new reactions or additional constraints from ^{13}C-labelling experiments and carrying out a more in-depth comparison. However, it is important that the field of computational modelling remains prudent, and that we do not forget to always consider the limitations of GSMs and COBRA as a querying tool more generally. Although FBA-based modeling has previously exhibited significant success (Simeonidis and Price, 2015, Yim et al., 2011), certainly not every FBA prediction can be trusted and FBA may not always be the most suitable predictive method (Stanford et al., 2015, Zomorrodi et al., 2012).

7.3.2 Limitations of the experimental approach

The analysis of the experimental results was complex and not easy to interpret. One of the main reasons for this may be that, although the cofactor modifications evaluated in this study were triggered by the implementation of a single gene knockout, the fact that ATP is a highly connected cofactor, together with its inherent interconnectivity with NADH, led to complex downstream metabolic ramifications which were difficult to detect and interpret just from the existing flux data. The systems clearly need further investigation, and although I mention throughout the relevant sections of Chapter 6 and also in Section 7.4.3 specific experiments that could be pursued to supplement and improve the analysis, I focus on a number of areas below.

On the question of whether PPC-PCK futile cycling exists in *E.coli*, it is still not known with certainty

whether this is actually the case, and if so, when exactly it happens *in vivo*. As far as this study is concerned, I could only conclude that the wild type strain showed no flux running through the PCK reaction under the particular pseudo steady-state conditions tested and suboptimal oxygenation, which contrasts from the observations made in Yang et al. (2003). However, the strain including the PCK knockout had a distinct phenotype to the wild type, characterised by higher glucose uptake and glycolytic rates and considerable flux redirection across its anaplerotic reactions and some TCA reactions. If these consequences were brought about as a result of the PCK knockout, PCK activity in the wild type should indeed be investigated further, because either the conditions tested did not capture PCK activity in the wild type, or the flux analysis setup was not able to appropriately estimate the flux of this reaction. I elaborate on the limitations of the MFA approach below more specifically. With respect to the phenotype of the PCK strain, although there is apparent flux rewiring across reactions linking PEP, OAA, MAL and PYR, any observations made about the dynamics of these metabolites were inferred from the flux data itself, and further characterisation of metabolite pools, transcription and protein levels is needed for these points to be concluded upon (see Section 7.4.3).

Similarly, on the characterisation of the ATPs strain, what was evident from this study's flux data was that the ATPase knockout led to the redirection of flux across most of central metabolism for the restoration of both energy and redox, so it was clear that such an energy perturbation triggers adjustments across redox reactions. However, whether energy and redox are jointly regulated, or whether the regulation of redox reactions occurs as a result of *E.coli*'s attempt to restore its energy state could not be answered from a metabolic flux analysis alone. Holm et al. (2010) elaborate nicely on the question of regulation, especially regarding ATP and NADH. In this particular study, they integrated metabolic flux, protein interaction and transcriptional regulation data and observed that ATP-related reactions had a higher degree of enzymatic and transcriptional regulation and higher robustness against perturbations. Reducing ATP availability upregulated proton-translocating mechanisms, prompting a more widespread response, whereas lower NADH was narrowly addressed by activating reactions that produced more NADH. Furthermore, Holm et al. (2010) concluded that some transcriptional factors are cofactor specific, while others regulate both ATP and NADH. Thus, there appears to be a distinct hierarchical organization governing transcriptional responses to energy and redox perturbations, with a global regulatory shared response between energy and redox, and a local response that is cofactor-specific. In any event, there is an inherent difficulty in experimentally measuring energy and redox balancing more generally, and further linking this to regulatory data. One option could be to infer cofactor balancing from the MFA reaction fluxes using stoichiometry and flux information from the estimated reactions (Chen et al., 2011, Garcia Martin et al., (2015, He et al., 2016), akin to the foundations upon which CBA was developed and feeding this into the global network developed by Holm et al. (2010), but until these evaluations are performed, this area of research remains open for further investigation.

More specifically on the MFA front, medium composition was an important consideration in flux analysis. Most ^{13}C flux analysis studies published to date involved experiments in minimal medium with a single carbon source, because this minimises the number of degrees of freedom and the possibility of dy-

namic transition between carbon sources during data fitting. This is the approach that was pursued in this study. However, the reality is that almost none of the high-producing strains reported are actually optimal in minimal medium. Instead, generally, media conditions will be supplemented to allow the cells to be more productive. So, here the question becomes, how much are we willing to sacrifice data fitting accuracy in order to use environmental conditions that more closely resemble what we would expect in real biotechnological settings? From the flux analysis perspective, the goal should be to produce the highest possible flux resolution. Every system subject to flux analysis should thus be optimized for the selection of optimal tracers and design of parallel experiments (using two or more different types of isotopes), which have been shown to dramatically improve flux measurement precision (Long and Antoniewicz, 2014). Significant efforts have been directed towards proper reporting, error analysis, transparency of methods and reproducibility. However this all comes at a cost, not only in the literal (economic) sense, which is extremely relevant and a key limiting factor in science, but also concerns the investment of human capital and resources, because this takes time and effort, and it is an area of research in itself.

Finally, it is worth mentioning that an important limitation of MFA is that it only generally applies to central metabolism and amino acid metabolism, and is fitted using core metabolic models. This is because of the assumption that a small number of reactions in central metabolism will carry the bulk of the flux, so there is no or very little need to track fluxes in other parts of metabolism. This is not very useful if we are working with biotechnological systems requiring flux redirections through alternative sub-systems, such as lipid or nitrogen metabolism, as these will not be included as part of the final flux map unless these pathways' isotope patterns and intermediates are resolved, which requires considerable adaptation of the existing protocols. Another shortcoming of this analytical method is that the results will depend on the metabolic network selected to fit the labelling data. Flux analysis at a genome-scale level has not been attempted yet (Saha et al., 2014). This also limits the analysis of biotechnological strains which may not have suitable metabolic networks available. Although this analytical technique was suitable for the purposes of this particular study, which was performed at small-scale, under fairly standard conditions and using a discrete number of strains, for the purposes of making the analysis of cofactor balancing widespread across the board, including a variety of hosts and pathways, it is probably best to stay away from metabolic flux analysis.

7.4 Future research opportunities

Indeed, although cofactor engineering research is fast developing, a considerable number of gaps in our knowledge still remain. A number of avenues for future research have been mentioned throughout this thesis, but I compile the most exciting opportunities for future development below, based on a number of areas that could be thoroughly extended directly from this study.

7.4.1 Extension of the CBA protocol

I have previously highlighted that during this PhD study I focused primarily on the generation of the method's foundations, including hypothesis selection, algorithm development and preliminary testing. Although the butanol case study was very powerful and insightful, assessing CBA's functionality within the context of a number of additional pathways and models would be an interesting, and potentially greatly beneficial addition to this work. To illustrate this, one of the key strengths of the work published by Dugar and Stephanopoulos (2011) was their ability to apply their calculations to a range of single pathways, as well as pinpointing the nature of the imbalances in every particular case. Assessing CBA's robustness using the case studies presented in Dugar and Stephanopoulos (2011) would be a good starting point for comparison.

Transitioning into larger, genome-scale models, such as iJO1366, as well as venturing into the study of alternative hosts, such as yeast or cyanobacteria, would make CBA transferrable across a wide range of model sizes and organisms, which would further support the impact of CBA's contribution to the future of predictive metabolic engineering.

Concerning the python algorithm itself, the protocol could be adapted to also account for additional cofactors, such as $FADH_2$. I would expect that including $FADH_2$ into the analysis would lead to expansion of the redox profile of the system under evaluation, especially if this is also pooled together with NADH and NADPH. Importantly, however, it would likely yield a better approximation of the redox balance in proton translocation, for example. It would also enable us to track the release of other metabolically relevant molecules such as succinate and fumarate. What is interesting about this flexibility however, is the fact that the user could tailor the analysis to their specific needs, beyond simply ATP, NADH or even FADH2. Furthermore, additional functions could be developed to enable an assessment of additional, and more complex, optimisation principles. Currently, the CBA function accepts a single flux distribution vector, which means that it is compatible with FBA, pFBA and MOMA frameworks, although the latter was not attempted during this study. As mentioned in Section 7.3.1, CBA is not currently compatible with FVA, given that FVA's solutions are in the form of a flux range. Nonetheless, adapting CBA to analyse FVA solutions would be an extremely powerful addition, especially when working with larger, more complex models that do not often lead to unique solutions.

7.4.2 Expanding the scope and applications of CBA: the MetEOr web server

One of the original motivations for developing CBA was that it was often very difficult to track the flow of cofactors within COBRApy, especially with the increasing size and complexity of some metabolic networks. How to best analyse and visualise results was an ongoing challenge. Furthermore, performing COBRA analyses, and even running CBA, required complex software installations and familiarity with programming. As a result of these challenges, making CBA, and FBA more generally, available as an interactive web-based tool, became an ongoing ambition. Building a web-based tool would make metabolic modelling more easily interpretable, and would make cofactor balancing assessments accessible to the wider

metabolic engineering community.

During this study, I had the opportunity to collaborate with the Bioinformatics group at the Life Sciences Department of Imperial College London to develop a web server, known as MetEOr (Metabolic Balance Calculator), now available on www.meteorfba.com (Davidson et al., 2021). MetEOr is an interactive web-based tool that allows users to edit SBML models and dynamically and interactively perform FBA. Furthermore, it enables users to quantify net ATP and NAD(P)H production for any particular host, calculate cofactor profiles (i.e. the fraction of these cofactors contributing to biomass formation or waste release, as opposed to cell maintenance and target production) and visualise the models' flux distributions (Davidson et al., 2019). Since CBA had already been developed using python programming and the COBRA package for python, it was possible to adapt the code for it to run within a back-end server, and deploy results through a web interface. Basic functions such as uploading and modifying models, and retrieving FBA and CBA results have already been deployed (a demo is shown in Figure 7.1). Next steps will include expanding the range of functions to also include FVA and building advanced visualisation options. In addition to these, a comparative tool would enable users to contrast results (e.g. compare yields from two models including two or more different pathways, or using two or more different hosts) and including the sensitivity analysis presented in Section 5.3.5 (Chapter 5), in order to rank pathways according to optimal cofactor balances, would also be very powerful to the MetEOr users. I thank Mira Davidson, Gregory Leeman, Ferran Cardoso Rodriguez, Melpi Kasapi and Virginia Fiarclough greatly for their efforts getting this work up to its current stage.

7.4.3 Experimental strategies: development of platform strains

A number of suggestions are proposed below that could elaborate on some of the questions and results presented in Chapter 6. These suggestions include both specific experiments to cross-check the existing analyses, as well as exciting areas of future development which I did not have the time to explore within the scope of this study.

A number of additions to the 13-C metabolic flux analysis protocol could be pursued first to validate the existing flux results. The protocol used, described in detail by Zamboni et al. (2009), takes advantage of the fact that different pathways within metabolism cleave and scramble the molecules (and subsequently incorporate the labelled carbon) in different ways, so this could be harnessed to quantify fluxes by different means. First, it could be interesting to test out alternative softwares to WUFlux, to see whether the fitting algorithms carry any weight in resolving absolute and relative fluxes. An additional supplement could be to calculate flux ratios using softwares like SUMOFLUX (Kogadeeva and Zamboni, 2016) and FiatFlux (Zamboni et al., 2005). Flux ratio analysis and metabolic flux analysis are complementary methods, and could be used to cross-validate each other. Detection of isotopic patterns can also be carried out using NMR. An alternative areas of focus could involve running MFA under anaerobic conditions, which would be particularly interesting for the ATPs strain and was strongly considered during this study but not performed due to time constraints.

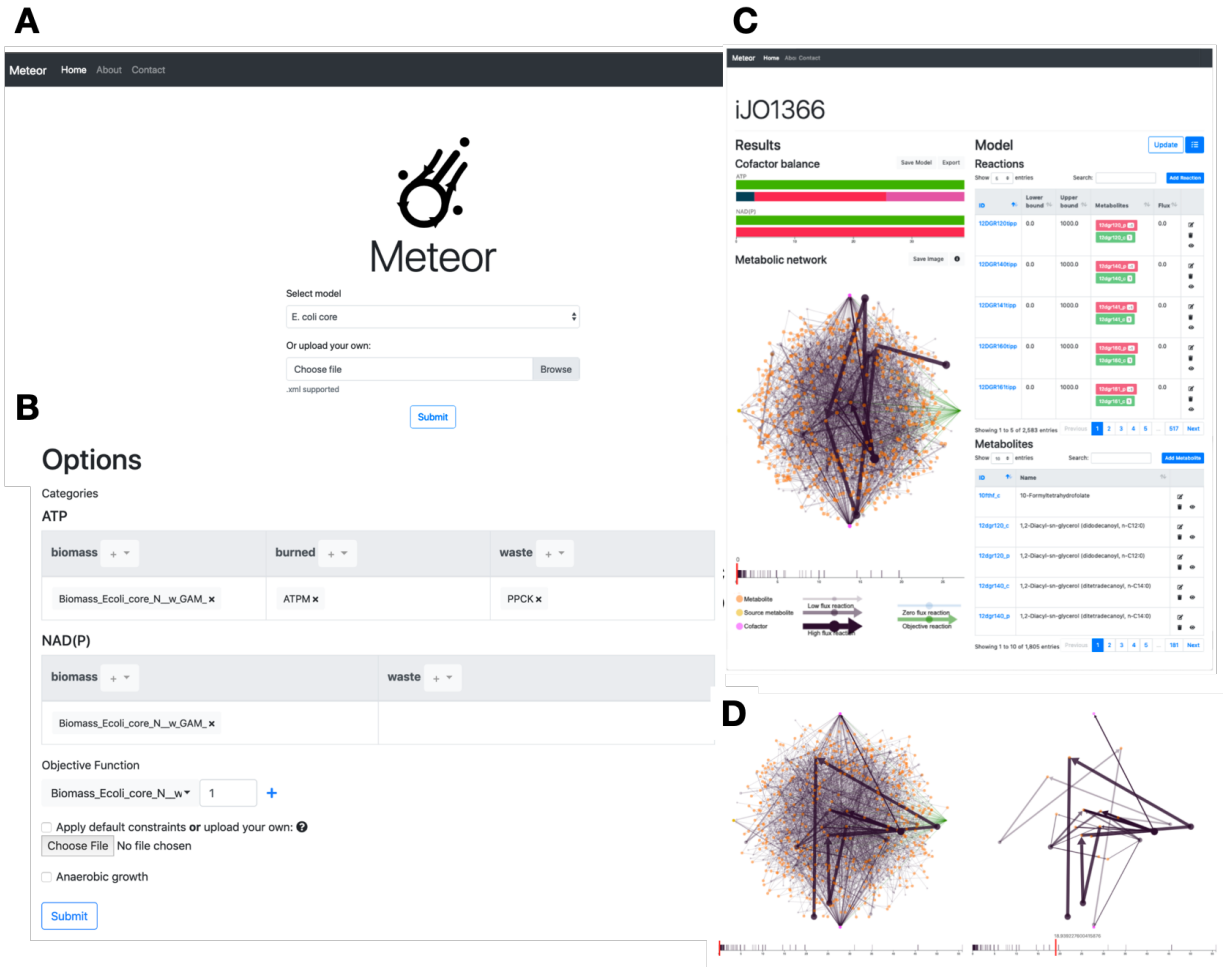


Figure 7.1: MetEOr demonstration (taken from Davidson et al. (2019)). (A) Home page, where users can select one of the provided standard models or upload their own SBML file (format-compliant) to start running tests. (B) Options form. Users can add, edit or remove selected reactions, modify FBA settings (inc.applying constraints and changing the objective function) and specify how to categorise some cofactor related reactions. (C) Results page, including bargraphs to show the CBA profiles, network visualisation and editing functions and flux display. (D) slider function filters the network by flux.

Shifting away from MFA, strain characterisation will be an essential part of any future developments in this area. Although the ATPs and PCK strains improved ethanol production, it is evident that these systems need further investigation and optimization. Firstly, the addition of uninduced controls would be of particular interest to include in future experiments, particularly in the ATPs strain, which showed marginal (although statistically significant) improvements in ethanol accumulation in the medium. Such a control would help determine, for instance, whether there is low expression in the induced control and leaky

expression in the uninduced control. Transcriptomic and metabolite measurements should be considered, especially in order to validate the distinct rewiring that appears to be taking place in the PCK strain. To further test the existence and extent of futile cycling in *E.coli*, ATP could be supplied into the medium at various concentrations, while tracking PCK and PPC activity.

In spite of the lack of characterisation and understanding as to how the systems operate, some interesting insights were made, particularly when the mutant strains were engineered to produce ethanol. Presumably, the lower biomass and ATP flux in these strains led to the increase in ethanol yield and titer. If this is in fact the case, I would suggest to evaluate these strains in the presence of alternative ATP-neutral or even ATP-producing pathways, because the distinct cofactor profiles of these strains could be what is driving the increased production of chemical targets with complementary cofactor characteristics. In the future, it would be interesting to perform a wider analysis considering additional pathways, as well as additional strains suitable for ATP-consuming pathways instead. This would entail the development of a strain with an ATP-surplus type of phenotype, in order to supply the system with the extra energy that the ATP-demanding pathway(s) will need. Complemented with a CBA analysis, this pragmatic approach to the building of "platform" strains with ideal cofactor usage profiles opens up an exciting avenue of research whereby synthetic pathways with specific cofactor characteristics can be plugged into strains with complementary cofactor usage profiles, potentially speeding up engineering by reducing the number of manipulations to be performed and more quickly reaching the desired yields.

7.5 Conclusion: What can be gained from understanding and engineering cofactor metabolism?

With this study, I aimed to emphasise that consolidated experimental and computational frameworks to track and optimise the use of cofactors is crucial to understand cell behaviour, and could have a great impact on the design of optimal catalysts.

CBA evaluates the concept of cofactor balancing using a relatively simple method built using a widely-used programming language and a modelling package that is well consolidated in the metabolic engineering field. While this PhD study only evaluated butanol and butanol precursor production as proof-of-concept, the CBA protocol thrives on COBRApy's flexibility to evaluate different stoichiometric models, target products, pathway routes, strains, carbon sources and environmental and genetic conditions, with minimal manipulations to the python function that performs the analysis. Most importantly, although the results from the CBA analysis are very simple, it is a powerful way to help metabolic engineers select optimal strain designs. It is also an effective tool to evaluate host-pathway interactions, given that one particular organism may be more suitable than another at accommodating the cofactor demands of a pathway of interest.

Constraint-based modelling remains a powerful tool for predicting system-wide behavioural changes in response to various conditions and provides a comprehensive prediction of entire flux solutions for optimal productivity. This modelling framework can help streamline experimental approaches, and support strain

and experimental design. When used in conjunction with the CBA protocol developed during this PhD study, it then becomes possible to track and monitor cofactor usage and its system-wide implications. As a result, metabolic engineering designs can be selected based not only on the highest yield achievable (which often does not translate to reality experimentally), but based also on the stress that cofactors can impose on the wider metabolic network. Nevertheless, it is vital that we do not forget that although COBRA-based modelling approaches have previously exhibited significant success (Simeonidis and Price, 2015, Yim et al., 2011), we should remain prudent and continue to question any *in silico* estimates, as COBRA modelling may not always be the most suitable predictive method (Chen et al., 2011).

This study also shows that metabolic responses to cofactor driven perturbations can lead to distinct phenotypes resulting in enhanced biochemical yields. The lower but positive yields of ATP present in the strains developed in this study were sufficient to maintain cellular performance but not to allow for unrestricted growth, illustrated by the smaller accumulation of biomass and higher ethanol yield in the engineered strains, compared to the larger biomass accumulation and lower ethanol yield of the WT. These results suggest that significant yield improvements can be achieved experimentally with very few, but highly targeted cofactor-driven modifications that reduce the waste of cofactors, showing the potential of these strains as platforms to improve bioproduction of cofactor-neutral or cofactor-surplus synthetic pathways.

To conclude, the *in silico* and *in vivo* strategies developed herein are a powerful means to achieve optimal strain designs more effectively and systematically, particularly if a large number of pathway variants are under evaluation. Having knowledge of the extent of cofactor imbalance can more accurately discriminate catalysts that are more balanced (and thus more productive) and pinpoint potential areas for engineering in response to particular sources of imbalance. If selected pathways are then combined with strains harbouring compatible cofactor profiles, the metabolic engineering sector will be able to benefit from greater productivities with a minimal number of genetic and metabolic manipulations leading to the design of more efficient strains.

References

Ahn, J., Chung, B. K. S., Lee, D. Y., Park, M., Karimi, I. A., Jung, J. K. and Lee, H. (2011), 'NADPH-dependent pgi-gene knockout *Escherichia coli* metabolism producing shikimate on different carbon sources', *FEMS Microbiology Letters* **324**, 10–16.
URL: *https://academic.oup.com/femsle/article/324/1/10/613959*

Akhtar, K. M., Jones, P. R., Carbonell, P. and Pedraza, J. M. (2014), 'Cofactor engineering for enhancing the flux of metabolic pathways', *Biotechnology and Bioengineering* **2**.
URL: *www.frontiersin.org*

Anaconda Software Distribution (2020), 'Anaconda: v4.7.12'.
URL: *http://www.anaconda.com*

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., Mori, H. and Mori, H. (2006), 'Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection', *Molecular Systems Biology* .
URL: *http://ecoli.aist-nara.ac.jp/*

Bailey, J. E. (1991), 'Toward a Science of Metabolic Engineering', *Science* **252**(5013), 1668—-1675.
URL: *http://science.sciencemag.org/*

BBSRC (2019), Uk research and innovation delivery plan 2019, Technical report.
URL: *https://www.ukri.org/files/about/dps/bbsrc-dp-2019/*

Boecker, S., Zahoor, A., Schramm, T., Link, H. and Klamt, S. (2019), 'Broadening the scope of enforced atp wasting as a tool for metabolic engineering in *Escherichia coli*', *Biotechnology Journal* **14**(1800438), 1–9.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/biot.201800438*

Bonarius, H. P., Hatzimanikatis, V., Meesters, K. P. H., de Gooijer, C. D., Schmid, G. and Tramper, J. (1996), 'Metabolic flux analysis of hybridoma cells in different culture media using mass balances', *Biotechnology and Bioengineering* **50**, 299–318.

Bonarius, H. P. J., Schmid, G. and Tramper, J. (1997), 'Flux analysis of underdetermined metabolic networks: the quest for the missing constraints', *Trends in Biotechnology* **15**(8), 308–314.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0167779997010676*

Bordbar, A., Monk, J. M., King, Z. A. and Palsson, B. O. (2014), 'Constraint-based models predict metabolic and associated cellular functions', *Nature Reviews Genetics* **15**, 107–120.
URL: *www.nature.com/reviews/genetics*

BP (2019), Full report – BP Statistical Review of World Energy 2019, Technical report, BP.
URL: *https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2019-full-report.pdf*

Broderick, J. B. (2001), 'Coenzymes and Cofactors', *Encyclopedia of Life Sciences* .

Burdette, L. A., Leach, S. A., Wong, H. T. and Tullman-Ercek, D. (2018), 'Developing Gram-negative bacteria for the secretion of heterologous proteins', *Microbial Cell Factories* .
URL: *https://doi.org/10.1186/s12934-018-1041-5*

Carlson, R. (2016), Estimating the biotech sector's contribution to the US economy, Technical report.
URL: *http://www.*

Celton, M., Goelzer, A., Camarasa, C., Fromion, V. and Dequin, S. (2012), 'A constraint-based model analysis of the metabolic consequences of increased NADPH oxidation in *Saccharomyces cerevisiae*', *Metabolic Engineering* **14**(4), 366–379.

Chao, Y. and Liao, J. C. (1994*a*), 'Metabolic responses to substrate futile cycling in *Escherichia coli*', **269**(7), 5122–5126.
URL: *http://www.jbc.org/content/269/7/5122.full.pdf*

Chao, Y. P. and Liao, J. C. (1994*b*), Metabolic responses to substrate futile cycling in *Escherichia coli*, Technical Report 7.
URL: *http://www.jbc.org/content/269/7/5122.full.pdf*

Chao, Y. P., Patnaik, R., Roof, W. D., Young, R. F. and Liao, J. C. (1993), 'Control of gluconeogenic growth by pps and pck in *Escherichia coli*.', *Journal of bacteriology* **175**(21), 6939–44.

Chemler, J. A., Fowler, Z. L., McHugh, K. P. and Koffas, M. A. (2010), 'Improving NADPH availability for natural product biosynthesis in *Escherichia coli* by metabolic engineering', *Metabolic Engineering* **12**(2), 96–104.

Chen, X., Alonso, A. P., Allen, D. K., Reed, J. L. and Shachar-Hill, Y. (2011), 'Synergy between 13-c metabolic flux analysis and flux balance analysis for understanding metabolic adaptation to anaerobiosis in *E. coli*', *Metabolic Engineering* **13**, 38–48.

Chen, X., Li, S. and Liu, L. (2014), 'Engineering redox balance through cofactor systems', *Trends in Biotechnology* **32**, 337–343.

Chen, X., Zhou, L., Tian, K., Kumar, A., Singh, S., Prior, B. A. and Wang, Z. (2013), 'Metabolic engineering of *Escherichia coli*: A sustainable industrial platform for bio-based chemical production', *Biotechnology Advances* **31**(8), 1200–1223.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0734975013000487*

Chin, J. W., Khankal, R., Monroe, C. A., Maranas, C. D. and Cirino, P. C. (2009), 'Analysis of nadph supply during xylitol production by engineered *Escherichia coli*', *Biotechnology and Bioengineering* **102**(1), 209–220.

Clomburg, J. M., Crumbley, A. M. and Gonzalez, R. (2017), 'Industrial biomanufacturing: The future of chemical production', *Science* **355**.
URL: *http://science.sciencemag.org/content/sci/355/6320/aag0804.full.pdf*

Copeland, W. B., Bartley, B. A., Chandran, D., Galdzicki, M., Kim, K. H., Sleight, S. C., Maranas, C. D. and Sauro, H. M. (2012), 'Computational tools for metabolic engineering', *Metabolic Engineering* **14**(3), 270–280.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3361690/pdf/nihms-365030.pdf*

Crown, S. B. and Antoniewicz, M. R. (2013), 'Parallel labeling experiments and metabolic flux analysis : Past , present and future methodologies', *Metabolic Engineering* **16**, 21–32.
URL: *http://dx.doi.org/10.1016/j.ymben.2012.11.010*

Dadi, P. K., Ahmad, M. and Ahmad, Z. (2009), 'Inhibition of atpase activity of *Escherichia coli* atp synthase by polyphenols', *International Journal of Biological Macromolecules* **45**(1), 72–79.

Dai, Z., Nielsen, J., Betenbaugh, M. J. and Bentley, W. E. (2015), 'Advancing metabolic engineering through systems biology of industrial microorganisms', *Current Opinion in Biotechnology* **36**, 8–15.
URL: *http://dx.doi.org/10.1016/j.copbio.2015.08.006*

Datsenko, K. A. and Wanner, B. L. (2000), One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products, Technical Report 12.
URL: *www.pnas.orgcgidoi10.1073pnas.120163297*

Davidson, M., Cardoso Rodriguez, F., Kasapi, M., Leeman, G., Fairclough, V. and de Arroyo Garcia, L. (2019), Meteor: An interactive web application for cofactor balancing analysis, Master's thesis, Imperial College London, South Kensington. MSc Theoretical Systems Biology and Bioinformatics.

Davidson, M., Leeman, G., Cardoso Rodriguez, F., de Arroyo Garcia, L., Fairclough, V. and Jones, P. R. (2021). (last accessed 5 February 2021).
URL: *https://www.meteorfba.com/*

de Arroyo Garcia, L. (2016), Computational analysis of metabolic pathway balance using flux balance analysis and the ins142 stoichiometric model of *Escherichia coli*, Master's thesis, Imperial College London, South Kensington. Master's Thesis.

De Kok, S., Kozak, B. U., Pronk, J. T. and Van Maris, A. J. A. (2012), 'Energy coupling in *Saccharomyces cerevisiae*: Selected opportunities for metabolic engineering', *FEMS Yeast Research* **12**(4), 387–397.

Dellomonaco, C., Clomburg, J. M., Miller, E. N. and Gonzalez, R. (2011), 'Engineered reversal of the b-oxidation cycle for the synthesis of fuels and chemicals', *Nature* .
URL: *https://www.nature.com/articles/nature10333.pdf*

Dugar, D. and Stephanopoulos, G. (2011), 'Relative potential of biosynthetic pathways for biofuels and bio-based products.', *Nature biotechnology* **29**(12), 1074–8.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/22158355*

Ebrahim, A., Lerman, J. A., Palsson, B. O. and Hyduke, D. R. (2013), 'COBRApy: COnstraints-Based Reconstruction and Analysis for Python.', *BMC systems biology* **7**(1), 74.
URL: *http://bmcsystbiol.biomedcentral.com/articles/10.1186/1752-0509-7-74*

Eisenstein, M. (2016), 'Living factories of the future', *Nature* **531**, 401–403.
URL: *http://dx.doi.org/10.1021/*

El-Chichakli, B. (2016), 'Five cornerstones of a global bioeconomy', *Nature* **535**.

Erb, T. J., Jones, P. R. and Bar-Even, A. (2017), 'Synthetic metabolism: metabolic engineering meets enzyme design', *Current Opinion in Chemical Biology* **37**, 56–62.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S1367593116302071*

Erdrich, P., Knoop, H., Steuer, R. and Klamt, S. (2014), 'Cyanobacterial biofuels: new insights and strain design strategies revealed by computational modeling', *Microbial Cell Factories* **13**, 128.
URL: *http://www.microbialcellfactories.com/content/13/128*

Escobar, H. (2019), 'Amazon fires clearly linked to deforestation, scientists say', *Science* **365**(6456), 853.
URL: *http://science.sciencemag.org/content/365/6456/853*

Fay, D. S. and Gerow, K. (2013), 'A biologist's guide to statistical thinking and analysis'.

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

Feist, A. M., Herrgård, M. J., Thiele, I. and Reed, J. L .and Ø Palsson, B. (2008), 'Reconstruction of biochemical networks in microorganisms', *Nature Reviews Microbiology* .
URL: *www.nature.com/reviews/micro*

Fernández-Castané, A., Fehér, T., Carbonell, P., Pauthenier, C. and Faulon, J. L. (2014), 'Computer-aided design for metabolic engineering', *Journal of Biotechnology* **192**, 302–313.
URL: *http://dx.doi.org/10.1016/j.jbiotec.2014.03.029*

Fischer, E., Zamboni, N. and Sauer, U. (2004), 'High-throughput metabolic flux analysis based on gas chromatography–mass spectrometry derived 13C constraints', *Analytical Biochemistry* **325**(2), 308–316.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S0003269703007528*

Fischer, J. D., Holliday, G. L., Rahman, S. A. and Thornton, J. M. (2010), 'The Structures and Physicochemical Properties of Organic Cofactors in Biocatalysis', *Journal of Molecular Biology* **403**(5), 803–824.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0022283610009940*

Fuhrer, T. and Sauer, U. (2009), 'Different Biochemical Mechanisms Ensure Network-Wide Balancing of Reducing Equivalents in Microbial Metabolism', *JOURNAL OF BACTERIOLOGY* **191**(7), 2112–2121.
URL: *http://jb.asm.org/*

Garcia Martin, H., Kumar, V. S., Weaver, D., Ghosh, A., Chubukov, V., Mukhopadhyay, A., Arkin, A. and Keasling, J. D. ((2015)), 'A method to constrain genome-scale models with 13-c labeling data', *PLOS Computational Biology* **11**.

García Sánchez, C. E. and Torres Sáez, R. G. (2014), 'Comparison and analysis of objective functions in flux balance analysis', *Biotechnology Progress* **30**(5), 985–991.

Ghiaci, P., Norbeck, J. and Larsson, C. (2013), Physiological adaptations of Saccharomyces cerevisiae evolved for improved butanol tolerance, Technical report.
URL: *http://www.biotechnologyforbiofuels.com/content/6/1/101*

Ghosh, A., Zhao, H. and Price, N. D. (2011), 'Genome-Scale Consequences of Cofactor Balancing in Engineered Pentose Utilization Pathways in Saccharomyces cerevisiae'.

Gianchandani, E. P., Chavali, A. K. and Papin, J. A. (2010), 'The application of flux balance analysis in systems biology', *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **2**(3), 372–382.

Gurobi Optimization, Inc. (2016), 'Gurobi optimizer reference manual'.
URL: *http://www.gurobi.com*

Hädicke, O., Bettenbrock, K. and Klamt, S. (2015), 'Enforced atp futile cycling increases specific productivity and yield of anaerobic lactate production in *Escherichia coli*', *Biotechnology and Bioengineering* **112**(10), 2195–2199.
URL: *http://doi.wiley.com/10.1002/bit.25623*

Hädicke, O. and Klamt, S. (2015), 'Manipulation of the ATP pool as a tool for metabolic engineering.', *Biochemical Society transactions* **43**(6), 1140–5.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/26614651*

Hara, K. Y. and Kondo, A. (2015), 'ATP regulation in bioproduction', *Microbial Cell Factories* **14**.

Haverkorn Van Rijsewijk, B. R., Nanchen, A., Nallet, S., Kleijn, R. J. and Sauer, U. (2011), 'Large-scale 13 c-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in *Escherichia coli*', *Molecular Systems Biology* **7**(109).

He, L., Wu, S. G., Zhang, M., Chen, Y. and Tang, Y. J. (2016), 'WUFlux: an open-source platform for 13-C metabolic flux analysis of bacterial metabolism'.
URL: *http://www.13cmfa.org/or http://tang.eece.wustl.edu/ToolDevelopment.htm*

HM Government UK (2018), Growing the Bioeconomy: A national bioeconomy strategy to 2030, Technical report, HM Government UK, London, UK.

Holm, A. K., Blank, L. M., Oldiges, M., Schmid, A., Solem, C., Jensen, P. R. and Vemuri, G. N. (2010), 'Metabolic and transcriptional response to cofactor perturbations in *Escherichia coli*', *The Journal of Biological Chemistry* **285**(23), 17498–17506.
URL: *http://www.jbc.org/*

Hunter, J. D. (2007), 'Matplotlib: A 2d graphics environment', *Computing in Science & Engineering* **9**(3), 90–95.

IEA (2020), 'Oil market report'. [Last Accessed: June 26th, 2020].
URL: *https://www.iea.org/reports/oil-market-report-april-2020*

Ingram, L., Aldrich, H., Borges, A., Causey, T., Martinez, A., Morales, F., Saleh, A., Underwood, S., Yomano, L., York, S., Zaldivar, J. and Zhou, S. (1999), 'Enteric Bacterial Catalysts for Fuel Ethanol Production', *Biotechnology Progress* **15**(5), 855–866.
URL: *http://doi.wiley.com/10.1021/bp9901062*

Ingram, L. O., Conway, T., Clark, D. P., Sewell, G. W. and Preston, J. F. (1987), 'Genetic engineering of ethanol production in *Escherichia coli*', *Applied and Environmental Microbiology* **53**(10), 2420–2425.

International Energy Agency (2018), Iea renewables 2018 market analysis, Technical report.

IPCC (2018), Global warming of 1.5°C, Technical report, Intergovernmental Panel on Climate Change, Switzerland.
URL: *https://report.ipcc.ch/sr15/pdf/sr15_spm_final.pdf*

Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., Hirasawa, T., Naba, M., Hirai, K., Hoque, A., Ho, P. Y., Kakazu, Y., Sugawara, K., Igarashi, S., Harada, S., Masuda, T., Sugiyama, N., Togashi, Takashi Hasegawa, M., Takai, Y., Yugi, K., Arakawa, K., Iwata, N., Toya, Y., Nakayama, Y., Nishioka, T., Shimizu, K., Mori, H. and Tomita, M. (2007), 'Multiple High-Throughput Analyses Monitor the Response of *E. coli* to Perturbations', *Science* **316**, 593–596.

Jensen, S. I., Lennen, R. M., Herrgård, M. J. and Nielsen, A. T. (2015), 'Seven gene deletions in seven days: Fast generation of *Escherichia coli* strains tolerant to acetate and osmotic stress', *Scientific Reports* **5**, 1–10.

Kai, B., Chung, S. and Lee, D.-Y. (2009), 'Flux-sum analysis: a metabolite-centric approach for understanding the metabolic network', *BMC Systems Biology* **3**, 117.
URL: *http://www.biomedcentral.com/1752-0509/3/117*

Kai-Sheng Chung, B., Lakshmanan, M., Klement, M., Mohanty, B. and Lee, D. Y. (2013), 'Genome-scale *in silico* modeling and analysis for designing synthetic terpenoid-producing microbial cell factories'.
URL: *http://dx.doi.org/10.1016/j.ces.2012.09.006*

Kallio, P., Pásztor, A., Akhtar, M. K. and Jones, P. R. (2014), 'Renewable jet fuel.', *Current opinion in biotechnology* **26C**, 50–55.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/24679258*

Kauffman, K. J., Prakash, P. and Edwards, J. S. (2003), 'Advances in flux balance analysis', *Current Opinion in Biotechnology* **14**(5), 491–496.

Kihira, C., Hayashi, Y., Azuma, N., Noda, S., Maeda, S., Fukiya, S., Wada, M., Matsushita, K. and Yokota, A. (2012), 'Alterations of glucose metabolism in *Escherichia coli* mutants defective in respiratory-chain enzymes', *Journal of Biotechnology* **158**(4), 215–223.

Kim, H. J., Deok, Y., Sang, K., Lee, Y. and Kim, P. (2012), 'An engineered *Escherichia coli* having a high intracellular level of atp and enhanced recombinant protein production', *Applied Microbial and Cell Physiology* **94**, 1079–1086.
URL: *http://rsbweb.nih.gov/ij*

Kim, J. and Copley, S. D. (2007), 'Why Metabolic Enzymes Are Essential or Nonessential for Growth of *Escherichia coli* K12 on Glucose'.
URL: *https://pubs.acs.org/sharingguidelines*

King, Z. A. and Feist, A. M. (2013), 'Optimizing Cofactor Specificity of Oxidoreductase Enzymes for the Generation of Microbial Production Strains-OptSwap', *Industrial Biotechnology* **9**, 236–246.
URL: *www.liebertpub.com*

King, Z. A. and Feist, A. M. (2014), 'Optimal cofactor swapping can increase the theoretical yield for chemical production in *Escherichia coli* and *Saccharomyces cerevisiae*', *Metabolic Engineering* **24**, 117–128.
URL: *http://dx.doi.org/10.1016/j.ymben.2014.05.009*

King, Z. A., Lloyd, C. J., Feist, A. M. and Palsson, B. O. (2015), 'Next-generation genome-scale models for metabolic engineering', *Current Opinion in Biotechnology* **35**, 23–29.
URL: *http://dx.doi.org/10.1016/j.copbio.2014.12.016*

Kitney, R., Baldwin, G., Bayer, T., Freemont, P. S., Dickinson, R., Ellis, T., Polizzi, K. and Stan, G. B. (2012), *A Primer for Synthetic Biology*, first edition edn, Imperial College Press, London, UK.

Koebmann, B. J., Westerhoff, H. V., Snoep, J. L., Nilsson, D. and Jensen, P. R. (2002), 'The Glycolytic Flux in *Escherichia coli* is Controlled by the Demand for ATP', *JOURNAL OF BACTERIOLOGY* **184**(14), 3909–3916.
URL: *http://jb.asm.org/*

Kogadeeva, M. and Zamboni, N. (2016), 'SUMOFLUX: A Generalized Method for Targeted 13 C Metabolic Flux Ratio Analysis', *PLOS Computational Biology* **12**(9), 1005109.
URL: *http://www.imsb.ethz.ch/research/zamboni/resources.html*

Kwon, Y. D., Lee, S. Y. and Kim, P. (2014), 'A physiology study of *Escherichia coli* overexpressing phosphoenolpyruvate carboxykinase', *Bioscience, Biotechnology, and Biochemistry* **72**(4), 1138–1141.
URL: *https://www.tandfonline.com/action/journalInformation?journalCode=tbbb20*

Lakshmanan, M., Yu, K., Koduru, L. and Lee, D. Y. (2015), '*In silico* model-driven cofactor engineering strategies for improving the overall NADP(H) turnover in microbial cell factories', *J Ind Microbiol Biotechnol* **42**, 1401–1414.
URL: *https://link.springer.com/content/pdf/10.1007%2Fs10295-015-1663-0.pdf*

Lan, E. I. and Liao, J. C. (2012), 'Atp drives direct photosynthetic production of 1-butanol in cyanobacteria.', *Proceedings of the National Academy of Sciences of the United States of America* **109**(16), 601–23.

Larsen, U., Johansen, T. and Schramm, J. (2009), Efp06 iea: Implementing agreement on advanced motor fuels ethanol as a fuel for road transportation main report, Technical report.

Lee, W., Kim, J., Park, E., Ham, N. S., Kim, M. D. and Seo, J. H. (2013), 'Effects of nadh kinase on nadph-dependent biotransformation processes in *Escherichia coli*', *Appl Microbiol Biotechnol* **97**.

Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Polpitiya, A. D., Adkins, J. N., Schramm, G., Purvine, S. O., Lopez-Ferrer, D., Weitz, K. K., Eils, R., Kö Nig, R., Smith, R. D. and Palsson, B. Ø. (2010), 'Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models', *Molecular Systems Biology* p. 390.

Liao, J. C., Mi, L., Pontrelli, S. and Luo, S. (2016), 'Fuelling the future: microbial engineering for the production of sustainable biofuels', *Nature Publishing Group* **14**.
URL: *https://www.nature.com/nrmicro/journal/v14/n5/pdf/nrmicro.2016.32.pdf*

Lim, J. H., Seo, S. W., Kim, Y. and Jung, Y. (2013), 'Model-driven rebalancing of the intracellular redox state for optimization of a heterologous n-butanol pathway in *Escherichia coli*'.
URL: *http://dx.doi.org/10.1016/j.ymben.2013.09.003*

Liu, J., Qi, H., Wang, C. and Wen, J. (2015), 'Model-driven intracellular redox status modulation for increasing isobutanol production in *Escherichia coli*', *Biotechnology for Biofuels* **8**, 108.
URL: *https://biotechnologyforbiofuels.biomedcentral.com/track/pdf/10.1186/s13068-015-0291-2*

Long, C. P. and Antoniewicz, M. R. (2014), 'Metabolic flux analysis of *Escherichia coli* knockouts: lessons from the Keio collection and future outlook', *Current Opinion in Biotechnology* **28**, 127–133.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0958166914000378*

Long, C. P. and Antoniewicz, M. R. (2019), 'Metabolic flux responses to deletion of 20 core enzymes reveal flexibility and limits of *E. coli* metabolism', *Metabolic Engineering* pp. 249–257.

Long, C. P., Gonzalez, J. E., Sandoval, N. R. and Antoniewicz, M. R. (2016), 'Characterization of physiological responses to 22 gene knockouts in *Escherichia coli* central carbon metabolism'.

Mahadevan, R. and Schilling, C. H. (2003), 'The effects of alternate optimal solutions in constraint-based genome-scale metabolic models', *Metabolic Engineering* **5**, 264–276.

Maia, P., Rocha, M. and Rocha, I. (2016), '*In Silico* constraint-based strain optimization methods: the quest for optimal cell factories', **80**(1), 45–67.

Martínez, I., Zhu, J., Lin, H., Bennett, G. N. and San, K. Y. (2008), 'Replacing *Escherichia coli* NAD-dependent glyceraldehyde 3-phosphate dehydrogenase (GAPDH) with a NADP-dependent enzyme from *Clostridium acetobutylicum* facilitates NADPH dependent pathways', *Metabolic Engineering* **10**(6), 352–359.

Menon, N., Pásztor, A., Menon, B. R. K., Kallio, P., Fisher, K., Akhtar, M. K., Leys, D., Jones, P. R. and Scrutton, N. S. (2015), 'A microbial platform for renewable propane synthesis based on a fermentative butanol pathway', *Biotechnology for Biofuels* **8**, 61.
URL: *http://www.biotechnologyforbiofuels.com/content/8/1/61*

Meza, E., Becker, J., Bolivar, F., Gosset, G. and Wittmann, C. (2012), Consequences of phosphoenolpyruvate:sugar phosphotranferase system and pyruvate kinase isozymes inactivation in central carbon metabolism flux distribution in *Escherichia coli*, Technical report.
URL: *http://www.microbialcellfactories.com/content/11/1/127*

Nakanishi-Matsui, M., Sekiya, M. and Futai, M. (2016), 'Atp synthase from *Escherichia coli*: Mechanism of rotational catalysis, and inhibition with the ε subunit and phytopolyphenols', *Biochimica et Biophysica Acta* **1857**, 129–140.
URL: *http://dx.doi.org/10.1016/j.bbabio.2015.11.005*

Nicolas, C., Kiefer, P., Letisse, F., Krömer, J., Massou, S., Soucaille, P., Wittmann, C., Lindley, N. D. and Portais, J. C. (2007), 'Response of the central metabolism of *Escherichia coli* to modified expression of the gene encoding the glucose-6-phosphate dehydrogenase', *FEBS Letters* **581**(20), 3771–3776.
URL: *http://doi.wiley.com/10.1016/j.febslet.2007.06.066*

Nielsen, J. and Keasling, J. D. (2016), 'Engineering Cellular Metabolism', *Cell* **164**(6), 1185–1197.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0092867416300708*

NOAA (2019), State of the climate: Global climate report for july 2019, Technical report, National Centers for Environmental Information (NOAA).
URL: *https://www.ncdc.noaa.gov/sotc/global/201907*

Noor, E., Bar-Even, A., Flamholz, A., Reznik, E., Liebermeister, W. and Milo, R. (2014), 'Pathway Thermodynamics Highlights Kinetic Obstacles in Central Metabolism', *PLoS Computational Biology* **10**(2).

O 'brien, E. J., Monk, J. M. and Palsson, B. O. (2015), 'Using genome-scale models to predict biological capabilities', *Cell* **161**, 971–987.
URL: *http://dx.doi.org/10.1016/j.cell.2015.05.019*

OECD (2018), Meeting Policy Challenges for a Sustainable Bioeconomy, Technical Report May, Organization for Economic Cooperation and Development, Paris, France.

Orth, J. D., Fleming, R. M. and Palsson, B. Ø. (2010b), 'Reconstruction and use of microbial metabolic networks: the core *Escherichia coli* metabolic model as an educational guide', *EcoSal Plus* **4**(1).
URL: *http://msb.embopress.org/content/msb/7/1/535.full.pdf*

Orth, J. D., Thiele, I. and Palsson, B. O. (2010), 'What is flux balance analysis?', *Nature Biotechnology* **28**(3), 245–248.
URL: *http://dx.doi.org/10.1038/nbt.1614*

Pasztor, A., Kallio, P., Malatinszky, D., Akhtar, M. K. and Jones, P. R. (2015), 'A synthetic o2-tolerant butanol pathway exploiting native fatty acid biosynthesis in *Escherichia coli*', *Biotechnology and Bioengineering* **112**(1), 120–128.

Patnaik, R., Roof, W. D., Young, R. F. and Liao', J. C. (1992), 'Stimulation of Glucose Catabolism in *Escherichia coli* by a Potential Futile Cycle', **174**(23), 7527–7532.
URL: *http://jb.asm.org/*

Pontrelli, S., Chiu, T.-Y., Lan, E. I., Chen, F. Y.-H., Chang, P. and Liao, J. C. (2018), '*Escherichia coli* as a host for metabolic engineering', *Metabolic Engineering* **50**, 16–46.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S1096717618300740*

Price, N. D., Reed, J. L. and Palsson, B. Ø. (2004), 'Genome-scale models of microbial cells: Evaluating the consequences of constraints', *Nature Reviews Microbiology* **2**, 886—-897.
URL: *www.nature.com/reviews/micro*

Ripoll Sanchez, L. (2018), ATP flexibility in *Escherichia coli* central carbon metabolism, B.sc. dissertation, Imperial College London, South Kensington. Final year Biochemistry Dissertation.

Rühle, T. and Leister, D. (2015), 'Assembly of F1 F0-ATP synthases', *BBA - Bioenergetics* **1847**, 849–860.
URL: *http://dx.doi.org/10.1016/j.bbabio.2015.02.005*

Russell, J. B. (2007), 'The energy spilling reactions of bacteria and other organisms', *Journal of Molecular Microbiology and Biotechnology* **13**(1-3), 1–11.

Russell, J. B. and Cook, G. M. (1995), 'Energetics of Bacterial Growth: Balance of Anabolic and Catabolic Reactions', *MICROBIOLOGICAL REVIEWS* **59**(1), 48–62.
URL: *http://mmbr.asm.org/content/59/1/48.full.pdf*

Saha, R., Chowdhury, A. and Maranas, C. D. (2014), 'Recent advances in the reconstruction of metabolic models and integration of omics data', *Current Opinion in Biotechnology* **29**, 39–45.
URL: *http://dx.doi.org/10.1016/j.copbio.2014.02.011*

Saini, M., Chen, M. H., Chiang, C.-J. and Chao, Y.-P. (2014), 'Potential production platform of n-butanol in *Escherichia coli*', *Metabolic Engineering* **27**, 76–82.
URL: *http://dx.doi.org/10.1016/j.ymben.2014.11.001*

Saini, M., Li, S.-Y., Wang, Z. W., Chiang, C.-J. and Chao, Y.-P. (2016), 'Systematic engineering of the central metabolism in *Escherichia coli* for effective production of n-butanol.', *Biotechnology for biofuels* **9**(1), 69.
URL: *http://biotechnologyforbiofuels.biomedcentral.com/articles/10.1186/s13068-016-0467-4*

Sauer, U., Lasko, D. R., Fiaux, J., Hochuli, M., Glaser, R., Szyperski, T., Wuthrich, K. and Bailey, J. (1999), 'Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism', **181**(21), 6679–6688.

Savinell, J. M. and Palsson, B. O. (1992), 'Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism', *Journal of Theoretical Biology* **154**, 421–454.

Schellenberger, J., Lewis, N. E. and Palsson, B. Ø. (2011), 'Elimination of Thermodynamically Infeasible Loops in Steady-State Metabolic Models', *Biophysical Journal* **100**, 544–553.
URL: *https://www.cell.com/action/showPdf?pii=S0006-3495%2810%2905225-2*

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. and Palsson, B. Ø. (2011), 'Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0.', *Nature protocols* **6**(9), 1290–1307.

Schuetz, R., Kuepfer, L. and Sauer, U. (2007), 'Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*', *Molecular systems biology* **3**(119), 119.

Segre, D., Vitkup, D. and Church, G. M. (2002), '(moma) analysis of optimality in natural and perturbed metabolic networks', *PNAS* **99**(23), 15112–15117.
URL: *http://www.pnas.org/content/99/23/15112.full.pdf*

Semkiv, M. V., Dmytruk, K. V., Abbas, C. A. and Sibirny, A. A. (2016), 'Activation of futile cycles as an approach to increase ethanol yield during glucose fermentation in *Saccharomyces cerevisiae*', *Bioengineered* **7**(2), 106–111.
URL: *http://dx.doi.org/10.1080/21655979.2016.1148223*

Shen, C. R., Lan, E. I., Dekishima, Y., Baez, A., Cho, K. M. and Liao, J. C. (2011), 'Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*', *Applied and Environmental Microbiology* **77**(9), 2905–2915.

Siedler, S., Bringer, S. and Bott, M. (2011), 'Increased nadph availability in *Escherichia coli*: improvement of the product per glucose ratio in reductive whole-cell biotransformation', *Applied Microbiology and Biotechnology* **92**, 929–937.
URL: *https://link.springer.com/content/pdf/10.1007%2Fs00253-011-3374-4.pdf*

Simeonidis, E. and Price, N. D. (2015), 'Genome-scale modeling for metabolic engineering', *Journal of Industrial Microbiology and Biotechnology* **42**, 327–338.
URL: *https://link.springer.com/content/pdf/10.1007%2Fs10295-014-1576-3.pdf*

Singh, A., Soh, K. C., Hatzimanikatis, V. and Gill, R. T. (2011), 'Manipulating redox and atp balancing for improved production of succinate in *E. coli*', *Metabolic Engineering* **13**(1), 76–81.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S1096717610000972*

Stanford, N. J., Millard, P. and Swainston, N. (2015), 'RobOKoD: microbial strain design for (over)production of target compounds', *Frontiers in Cell and Developmental Biology* **3**(March), 17.
URL: *http://journal.frontiersin.org/article/10.3389/fcell.2015.00017*

Stelling, J., Sauer, U., Szallasi, Z., Doyle, F. J. and Doyle, J. (2004), 'Robustness of cellular functions', *Cell* **118**(6), 675–685.

Stephanopoulos, G. (2012), 'Synthetic biology and metabolic engineering'.
URL: *https://pubs.acs.org/sharingguidelines*

Toya, Y., Nakahigashi, K., Tomita, M. and Shimizu, K. (2012), 'Metabolic regulation analysis of wild-type and arca mutant *Escherichia coli* under nitrate conditions using different levels of omics data', *This journal is c The Royal Society of Chemistry* **8**, 2593–2604.
URL: *www.rsc.org/molecularbiosystems*

Trinh, C. T., Wlaschin, A. and Srienc, F. (2009), 'Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism', *Applied Microbiology and Biotechnology* **81**, 813–826.

UNFCC (2015), Adoption of the Paris Agreement, Technical report, UN.
URL: *https://unfccc.int/sites/default/files/english_paris_agreement.pdf*

United Nations (2015), Transforming our world: The 2030 agenda for sustainable development, Technical report, UN.
URL: *https://sustainabledevelopment.un.org/content/documents/21252030 Agenda for Sustainable Development web.pdf*

United Nations (2017), World population prospects: The 2017 revision, Technical report.
URL: *https://population.un.org/wpp/Publications/Files/WPP2017_KeyFindings.pdf*

Varma, A., Boesch, B. W. and Paisson, B. O. (1993), Biochemical production capabilities of *Escherichia coIi*, Technical report.

Varma, A. and Palsson, B. O. (1994), 'Stoichiometric Flux Balance Models Quantitatively Predict Growth and Metabolic By-Product Secretion in Wild-Type *Escherichia coli* W3110', *Applied and Environmental Microbiology* **60**(10), 3724–3731.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC201879/pdf/aem00027-0254.pdf*

Vemuri, G. N. and Aristidou, A. A. (2005), 'Metabolic Engineering in the-omics Era: Elucidating and Modulating Regulatory Networks', *Microbiology and Molecular Biology Reviews* **69**(2), 197–216.
URL: *http://mmbr.asm.org/*

Wang, M., Chen, B., Fang, Y. and Tan, T. (2017), 'Cofactor engineering for more efficient production of chemicals and biofuels', *Biotechnology Advances* **35**(8), 1032–1039.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0734975017301192*

Wang, Y., San, K. Y. and Bennett, G. N. (2013), 'Cofactor engineering for advancing chemical biotechnology', *Current Opinion in Biotechnology* **24**, 994–999.
URL: *http://dx.doi.org/10.1016/j.copbio.2013.03.022*

Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Lukauskas, S., Gemperline, D. C., Augspurger, T., Halchenko, Y., Cole, J. B., Warmenhoven, J., de Ruiter, J., Pye, C., Hoyer, S., Vanderplas, J., Villalba, S., Kunter, G., Quintero, E., Bachant, P Martin, M., Meyer, K., A., M., Ram, Y., Yarkoni, T., Williams, M. L., Evans, Fitzgerald, C., Fonnesbeck, C., Lee, A. and Qalieh, A. (2020), 'mwaskom/seaborn: v0.9.0'.
URL: *https://doi.org/10.5281/zenodo.883859*

Weusthuis, R. A., Lamot, I., Van Der Oost, J. and Sanders, J. P. M. (2010), 'Microbial production of bulk chemicals: development of anaerobic processes', *Trends in Biotechnology* **29**, 153–158.

Wiechert, W. (2001), 'Minireview: Metabolic Flux Analysis', *Metabolic Engineering* **3**, 1955206.
URL: *www.idealibrary.com*

WMO (2020), 'Reported new record temperature of 38°c north of arctic circle'. [Last Accessed: June 26th, 2020].
URL: *https://public.wmo.int/en/media/news/reported-new-record-temperature-of-38%C2%B0c-north-of-arctic-circle*

Wu, G., Yan, Q., Jones, J. A., Tang, Y. J., Fong, S. S. and Koffas, M. A. G. (2016), 'Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications'.

WWF (2019), 'July 29: Earth Overshoot Day is the earliest ever'. (last accessed 18 September 2019).
URL: *http://wwf.panda.org/?350491/Earth-Overshoot-Day-2019*

Yang, C., Hua, Q., Baba, T., Mori, H. and Shimizu, K. (2003), 'Analysis of *Escherichia coli* anaplerotic metabolism and its regulation mechanisms from the metabolic responses to altered dilution rates and phosphoenolpyruvate carboxykinase knockout', *Wiley InterScience (www.interscience.wiley.com). Biotechnol Bioeng* **84**, 129–144.
URL: *www.interscience.wiley.com*

Yang, M., Li, X., Chunya, B., Wang, H., Shi, G., Yang, X., Hu, Y. and Wang, X. (2014), 'Pyruvate decarboxylase and alcohol dehydrogenase overexpression in *Escherichia coli* resulted in high ethanol production and rewired metabolic enzyme networks', *World J Microbiol Biotechnol* **30**, 2871–2883.

Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J., Osterhout, R., Stephen, R., Estadilla, J., Teisan, S., schreyer, H. B., Andrae, S., Hoong Yang, T., Yup Lee, S., Burk, M. J. and Van Dien, S. (2011), 'Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol', *Nature Chemical Biology* **7**.
URL: *www.nature.com/naturechemicalbiology*

Zamboni, N., Fendt, S. M., Rühl, M. and Sauer, U. (2009), ' ^{13}C-based metabolic flux analysis', *Nature Protocols* **4**, 878–892.
URL: *https://www.nature.com/articles/nprot.2009.58.pdf*

Zamboni, N., Fischer, E. and Sauer, U. (2005), 'FiatFlux: a software for metabolic flux analysis from ^{13}C-glucose experiments.', *BMC bioinformatics* **6**, 209.

Zhang, X., Jantama, K., Shanmugam, K. T. and Ingram, L. O. (2009), 'Reengineering *Escherichia coli* for succinate production in mineral salts medium', *Applied and Environmental Microbiology* **75**(24), 7807–7813.

Zhao, H. and Van Der Donk, W. A. (2003), 'Regeneration of cofactors for use in biocatalysis', *Current Opinion in Biotechnology* **14**(6), 583–589.

Zhou, S., Iverson, A. G. and Grayburn, W. S. (2008), 'Engineering a native homoethanol pathway in *Escherichia coli B* for ethanol production', *Biotechnology Letters* **30**(2), 335–342.
URL: *http://link.springer.com/10.1007/s10529-007-9544-x*

Zomorrodi, A. R., Suthers, P. F., Ranganathan, S. and Maranas, C. D. (2012), 'Mathematical optimization applications in metabolic networks', *Metabolic Engineering* **14**(6), 672–686.
URL: *http://dx.doi.org/10.1016/j.ymben.2012.09.005*

Appendices

A Flux Variability Analysis (FVA) estimates

Table 1. **Aerobic FVA estimates of the wild type and engineered models under aerobic conditions.** Optimized for biomass formation (WT) or target production, accordingly, using the *E.coli* Core model. Minimal and maximal range units are in mmol gDW^{-1} hr^{-1}. Fraction of optimum = 100%. Reactions are shown in no particular order.
Table 2. **Aerobic FVA estimates of the butanol and butanol precursor models after manual curation of high-flux futile cycles.** Optimized for butanol or butanol precursor production, accordingly, using the *Escherichia coli* Core Model. Minimal and maximal range units are in mmol gDW^{-1} hr^{-1}. Fraction of optimum = 100%. Reactions are shown in no particular order.
Table 3. **Anaerobic FVA estimates of the butanol and butanol precursor models after manual curation of high-flux futile cycles.** Optimized for butanol or butanol precursor production, accordingly, using the *Escherichia coli* Core Model. Minimal and maximal range units are in mmol gDW^{-1} hr^{-1}. Fraction of optimum = 100%. Reactions are shown in no particular order.

B pFBA flux distributions

Table 4. **pFBA flux distributions of wild type, butanol and butanol precursor models under aerobic conditions.** Used the *Escherichia coli* Core Model and pFBA for optimization, using either biomass formation or target production as the objective function, accordingly. Solutions were simulated under aerobic conditions (EX_o2_e_ = -10 mmol gDW^{-1} hr^{-1}) and were otherwise unconstrained. Reactions are shown in no particular order.
Table 5. **pFBA flux distributions of wild type, butanol and butanol precursor models under anaerobic conditions.** Used the *Escherichia coli* Core Model and pFBA for optimization, using either biomass formation or target production as the objective function, accordingly. Solutions were simulated under aerobic conditions (EX_o2_e_ = 0 mmol gDW^{-1} hr^{-1}) and were otherwise unconstrained. Reactions are shown in no particular order.
Table 6. **pFBA flux distributions of the butanol and butanol precursor models under aerobic conditions after manual curation of high-flux futile cycles.** Used the *Escherichia coli* Core Model and pFBA for optimization, using target production as the objective function, accordingly. Solutions were simulated under aerobic conditions (EX_o2_e_ = -10 mmol gDW^{-1} hr^{-1}). Reactions shown in no particular order.
Table 7. **pFBA flux distributions of the butanol and butanol precursor models under anaerobic conditions after manual curation of high-flux futile cycles.** Used the *Escherichia coli* Core Model and pFBA for optimization, using target production as the objective function, accordingly. Solutions were simulated under anaerobic conditions (EX_o2_e_ = 0 mmol gDW^{-1} hr^{-1}). Reactions shown in no particular order.
Table 8. **pFBA flux distributions of wild type and engineered models constrained using ^{13}C-MFA data.** Used the *Escherichia coli* Core Model and pFBA for optimization, using either biomass formation or target production as the objective function, accordingly. Solutions were simulated under aerobic conditions (EX_o2_e_ = -10 mmol gDW^{-1} hr^{-1}), constrained using ^{13}C-MFA data from Long and Antoniewicz (2019). Reactions are shown in no particular order.
Table 9. **pFBA flux distributions of butanol and butanol precursor models constrained using ^{13}C-MFA data and additional manual curation of high-flux futile reactions.** Used the *Escherichia*

coli Core Model and pFBA for optimization, using target production as the objective function. Solutions were simulated under aerobic conditions (EX_o2_e_ = -10 mmol gDW^{-1} hr^{-1}), constrained using ^{13}C-MFA data from Long and Antoniewicz (2019). High-flux futile cycles capped are shown in the first row. Reaction fluxes (with reaction IDs shown in the first column) are shown in no particular order.

C CBA estimates

Table 10. **CBA parameters and outputs of unconstrained models under aerobic conditions.** Reaction IDs, relevant co-factors and their stoichiometric coefficients, flux values, balance values and assigned balance category are included.
Table 11. **CBA parameters and outputs of unconstrained models under anaerobic conditions.** Reaction IDs, relevant co-factors and their stoichiometric coefficients, flux values, balance values and assigned balance category are included.
Table 12. **CBA parameters and outputs of manually curated models under aerobic conditions.** Reaction IDs, relevant co-factors and their stoichiometric coefficients, flux values, balance values and assigned balance category are included.
Table 13. **CBA parameters and outputs of manually curated models under anaerobic conditions.** Reaction IDs, relevant co-factors and their stoichiometric coefficients, flux values, balance values and assigned balance category are included.

D Sequencing Results

Figure 1. **Snapgene view of ATPs sequencing results.**
Figure 2. **ATPs sequencing with Matcher.** Local alignment between forward and reverse-complement sequencing results using the Matcher software, which identifies local similarities between two sequences using a rigorous algorithm based on the LALIGN application.
Figure 3. **ATPs sequencing with Water.** Local alignment between forward and reverse-complement sequencing results using the Water software, which uses the Smith-Waterman algorithm (modified for speed enhancements) to calculate the local alignment of two sequences.
Figure 4. **Snapgene view of PCK sequencing results.**
Figure 5.**PCK sequencing with Matcher.** Local alignment between forward and reverse-complement sequencing results using the Matcher software.
Figure 6. **PCK sequencing with Water.** Local alignment between forward and reverse-complement sequencing results using the Water software.
Figure 7. **Snapgene view of ATPs-PCK sequencing results.**
Figure 8.**ATPs-PCK sequencing with Matcher.** Local alignment between forward and reverse-complement sequencing results using the Matcher software.
Figure 9. **ATPs-PCK sequencing with Water.** Local alignment between forward and reverse-complement sequencing results using the Water software.

E Growth Rates and Physiological Parameters

Table 14. **Growth rates and generation times of the wild type, PCK and ATPs mutants.**
Figure 10. **Secretion product standard curves.** Glucose, acetate, succinate, lactate formate and ethanol standard curves to quantify these compounds by HPLC.

Figure 11. **OD_{600} against time (h) in logarithmic scale**. Timepoints selected to calculate growth rates are: (A) wild type: 4h and 8h; (B) PCK: 0h and 8h; and (C): ATPs: 4h and 10h.
Figure 12. **Calibration curves.** Linear functions convert optical density (OD_{600}) to cell dry weight (gDW L^{-1}) for strains WT and ATPs

F ^{13}C Metabolic Flux Analysis

Table 15. **Reactions included in the *E.coli* model available on WUFlux**
Table 16. **Absolute fluxes (mmol gDW^{-1} hr^{-1}) for the WT, ATPs and PCK strains**

Table 1: Aerobic FVA estimates of the wild type and engineered models under aerobic conditions. Optimized for biomass formation (WT) or target production, accordingly, using the *E.coli* Core model. Minimal and maximal range units are in mmol gDW^{-1} hr^{-1}. Fraction of optimum = 100%. Reactions are shown in no particular order.

	WT		BuOH-0		BuOH-1		tpcBuOH		BuOH-2	
Reaction ID	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
NPHT7					10.0	10.0			9.049	9.049
G6PDH2r	4.72	4.72					2.53	2.53	4.288	4.288
PGM	-14.85	-14.85	-20.0	-20.0	-20.0	-20.0	-19.16	-19.16	-18.571	-18.571
FBP			0.0	12.40		2.40				
FUM	5.34	5.34							0.473	0.473
Biomass	0.86	0.86								
PGI	5.11	5.11	10.0	10.0	10.0	10.0	7.47	7.47	5.712	5.712
GND	4.72	4.72	10.0	10.0			2.53	2.53	4.288	4.288
NADH11	39.36	39.36					5.07	5.07	10.941	10.941
PPC	2.47	2.47		12.40		2.40				
MDH	5.34	5.34	-11.20		-2.40				0.473	0.473
PPCK				12.40		2.40				
PPS				12.40		2.40				
ME1				11.2		2.40				
ME2				11.2		2.40				
AKGDH	5.34	5.34							0.473	0.473
PGK	-16.14	-16.14	-20.0	-20.0	-20.0	-20.0	-19.16	-19.16	-18.571	-18.571
ADK1				12.40		2.40	9.58	9.58	9.049	9.049
EX_glc_e	-10.0	-10.0	-10.0	-10.0	-10.0	-10.0	-10.0	-10.0	-10.0	-10.0
ATPM	7.60	7.60	7.60	20.0	7.60	10.0	7.60	7.60	7.600	7.60
RPI	-2.19	-2.19					-0.84	-0.84	-1.429	-1.429
RPE	2.53	2.53					1.69	1.69	2.859	2.859
TALA	1.42	1.42					0.84	0.84	1.429	1.429
SUCD4	5.34	5.34							0.473	0.473
ICDHyr	6.26	6.26							0.473	0.473
GAPD	16.14	16.14	20.0	20.0	20.0	20.0	19.16	19.16	18.571	18.571
CYTBD	44.69	44.69					5.07	5.07	11.415	11.415
SUCOAS	-5.34	-5.34							-0.473	-0.473
ENO	14.85	14.85	20.0	20.0	20.0	20.0	19.16	19.16	18.571	18.571
PDH	9.49	9.49	20.0	20.0	20.0	20.0	19.16	19.16	18.571	18.571
CS	6.26	6.26							0.473	0.473
ATPS4r	39.75	39.75	-12.40		-2.40		7.60	7.60	15.702	15.702
TKT2	1.11	1.11					0.84	0.84	1.429	1.429
TKT1	1.42	1.42					0.84	0.84	1.429	1.429
PGL	4.72	4.72					2.53	2.53	4.288	4.288
PFK	7.57	7.57	10.00	22.40	10.0	12.40	9.16	9.16	8.571	8.571
FBA	7.57	7.57	10.0	10.0	10.0	10.0	9.16	9.16	8.571	8.571
PYK	1.93	1.93		22.40	7.60	12.40	9.16	9.16		
ACONT	6.26	6.26	10.0	10.0					0.473	0.473
TPI	7.57	7.57			10.0	10.0	9.16	9.16	8.571	8.571
NADTRHD				24.8		4.8				
THD2				24.8		4.8	4.51	4.51		
HCO3E					10.0	10.0			9.049	9.049
ACCOAC					10.0	10.0			9.049	9.049
BUT1			10.0	10.0			9.58	9.58		
BUT2			10.0	10.0	10.0	10.0	9.58	9.58	9.049	9.049
BUT3			10.0	10.0	10.0	10.0	9.58	9.58	9.049	9.049
BUT4			10.0	10.0	10.0	10.0	9.58	9.58	9.049	9.049
BUT5			10.0	10.0	10.0	10.0				
BUT6			10.0	10.0	10.0	10.0	9.58	9.58	9.049	9.049
CAR							9.58	9.58	9.049	9.049
BTBTAC							9.58	9.58	9.049	9.049
BTOH_tr			10.0	10.0	10.0	10.0	9.58	9.58	9.049	9.049
BTOH_sink			10.0	10.0	10.0	10.0	9.58	9.58	9.049	9.049

	fasBuOH		CROT		BUTYR		BUTAL	
Reaction ID	Min	Max	Min	Max	Min	Max	Min	Max
G6PDH2r	6.265	6.265						
PGM	-17.912	-17.912	-20.0	-20.0	-20.0	-20.0	-20.0	-20.0
PFL				20.0		20.0		10.0
PGI	3.735	3.735	10.0	10.0	10.0	10.0	10.0	10.0
GND	6.265	6.265						
FBP				42.40		32.40		22.40
NADH11	12.530	12.530	10.0	30.0		20.00		10.0
RPI	-2.088	-2.088						
RPE	4.177	4.177						
TALA 2.088	2.088							
MDH			-26.20		-21.20		-16.20	
ME1				26.20		21.20		16.20
ME2				26.20		21.20		16.20
PGK	-17.912	-17.912	-20.0	-20.0	-20.0	-20.0	-20.00	-20.0
PPC				42.4		32.40		22.40
PPCK				42.4		32.40		22.40
ADK1	8.956	8.956		42.40		32.40		22.40
EX_glc_e	-10.0	-10.0	-10.0	-10.0	-10.0	-10.0	-10.0	-10.0
ATPM	7.600	7.60	7.60	50.0	7.60	40.0	7.60	30.0
GAPD	17.912	17.912	20.0	20.0	20.0	20.0	20.0	20.0
CYTBD	12.530	12.530	10.0	30.0		20.0		10.0
ENO	17.912	17.912	20.0	20.0	20.0	20.0	20.0	20.0
PDH	17.912	17.912		20.0		20.0	10.0	20.0
PPS				42.40		32.40		22.40
PYK	7.912	7.912		52.40		42.40		32.40
ATPS4r	16.556	16.556	-12.40	30.0	-12.40	20.0	-12.40	10.0
TKT2	2.088	2.088						
TKT1	2.088	2.088						
PGL	6.265	6.265						
PFK	7.912	7.912	10.0	52.40	10.0	42.40	10.0	32.40
FBA	7.912	7.912	10.0	10.0	10.0	10.0	10.0	10.0
TPI	7.912	7.912	10.0	10.0	10.0	10.0	10.0	10.0
NADTRHD				84.8		64.8		44.8
THD2	5.382	5.382		84.80		64.8		44.8
HCO3E	8.956	8.956						
ACCOAC	8.956	8.956						
BUT1			10.0	10.0	10.0	10.0	10.0	10.0
BUT2			10.0	10.0	10.0	10.0	10.0	10.0
BUT3			10.0	10.0	10.0	10.0	10.0	10.0
BUT4			10.0	10.0	10.0	10.0		
BUT6	8.956	8.956						
CAR	8.956	8.956						
BTBTAC	8.956	8.956						
BTOH_tr	8.956	8.956						
BTOH_sink	8.956	8.956						
BUT5							10.0	10.0
BTBTAC							10.0	10.0
B2CTCRO			10.0	10.0				
CROAC_tr			10.0	10.0				
CROAC_sink			10.0	10.0				
BTAC_tr					10.0	10.0		
BTAC_sink					10.0	10.0		
BTAL_tr							10.0	10.0
BTAL_sink							10.0	10.0
3HAD40	8.956	8.956						
MCOATA	8.956	8.956						
3OAR40	8.956	8.956						
EAR40x	8.956	8.956						
KAS15	8.956	8.956						
5_BUT1	8.956	8.956						

Table 2: Aerobic FVA estimates of the butanol and butanol precursor models after manual curation of high-flux futile cycles. Optimized for butanol or butanol precursor production, accordingly, using the *Escherichia coli* Core Model. Minimal and maximal range units are in mmol gDW^{-1} hr^{-1}. Fraction of optimum = 100%. Reactions are shown in no particular order.

Reaction ID	BuOH-0		BuOH-1		tpcBuOH		BuOH-2		fasBuOH	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
ACONT	0.17	0.17	0.18	0.18	0.10	0.10	0.00	0.00	0.13	0.13
ADK1	0.00	0.00	0.00	0.00	8.46	8.46	9.23	9.23	7.67	7.67
ATPM	7.60	7.60	7.60	7.60	7.60	7.60	2.15	4.62	7.60	7.60
ATPS4r	0.00	0.00	0.00	0.00	14.38	14.38	13.85	13.85	23.22	23.22
Biomass	0.16	0.16	0.05	0.05	0.09	0.09	0.00	0.00	0.12	0.12
CS	0.17	0.17	0.18	0.18	0.10	0.10	0.00	0.00	0.13	0.13
CYTBD	0.73	0.73	0.65	0.95	10.39	10.39	9.23	9.23	17.78	17.78
ENO	19.32	19.32	19.81	19.81	17.94	17.94	18.46	18.46	16.62	16.62
FBA	9.82	9.82	9.96	9.96	8.24	8.24	8.46	8.46	6.99	6.99
FUM	0.00	0.00	0.13	0.13	0.00	0.00	0.00	0.00	0.00	0.00
G6PDH2r	0.06	0.06	0.00	0.00	5.02	5.02	4.62	4.62	8.67	8.67
GAPD	19.56	19.56	19.88	19.88	18.08	18.08	18.46	18.46	16.80	16.80
GLCpts	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
GND	0.06	0.06	0.00	0.00	5.02	5.02	4.62	4.62	8.67	8.67
ICDHyr	0.17	0.17	0.05	0.05	0.10	0.10	0.00	0.00	0.13	0.13
ICL	0.00	0.00	0.13	0.13	0.00	0.00	0.00	0.00	0.00	0.00
MALS	0.00	0.00	0.13	0.13	0.00	0.00	0.00	0.00	0.00	0.00
MDH	-2.01	-2.01	0.26	0.26	0.00	0.00	0.00	0.00	0.00	0.00
ME1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NADH11	0.73	0.73	0.52	0.82	10.39	10.39	9.23	9.23	17.78	17.78
NADTRHD	0.00	0.00	0.00	0.75	0.00	0.00	0.00	0.00	0.00	0.00
PDH	18.32	18.32	19.36	19.66	17.37	17.37	18.46	18.46	15.90	15.90
PFK	9.82	9.82	9.96	9.96	8.24	8.24	8.46	8.46	6.99	6.99
PFL	0.00	0.00	0.00	0.30	0.00	0.00	0.00	0.00	0.00	0.00
PGI	9.91	9.91	9.99	9.99	4.96	4.96	5.39	5.39	1.31	1.31
PGK	-19.56	-19.56	-19.88	-19.88	-18.08	-18.08	-18.46	-18.46	-16.80	-16.80
PGL	0.06	0.06	0.00	0.00	5.02	5.02	4.62	4.62	8.67	8.67
PGM	-19.32	-19.32	-19.81	-19.81	-17.94	-17.94	-18.46	-18.46	-16.62	-16.62
PPC	2.47	2.47	0.00	0.00	2.47	2.47	0.00	2.47	2.47	2.47
PPCK	0.00	0.00	0.00	0.00	2.21	2.21	0.00	2.47	2.13	2.13
PYK	6.76	6.76	9.79	9.79	7.63	7.63	8.46	8.46	6.23	6.23
RPE	-0.07	-0.07	-0.03	-0.03	3.28	3.28	3.08	3.08	5.70	5.70
RPI	-0.14	-0.14	-0.03	-0.03	-1.74	-1.74	-1.54	-1.54	-2.97	-2.97
TALA	-0.01	-0.01	-0.01	-0.01	1.66	1.66	1.54	1.54	2.87	2.87
THD2	0.61	0.61	0.78	1.53	0.00	0.00	0.00	0.00	0.00	0.00
TKT1	-0.01	-0.01	-0.01	-0.01	1.66	1.66	1.54	1.54	2.87	2.87
TKT2	-0.07	-0.07	-0.02	-0.02	1.62	1.62	1.54	1.54	2.83	2.83
TPI	9.82	9.82	9.96	9.96	8.24	8.24	8.46	8.46	6.99	6.99
BUT1	8.78	8.78			8.46	8.46				
BUT2	8.78	8.78	9.59	9.59	8.46	8.46	9.23	9.23		
BUT3	8.78	8.78	9.59	9.59	8.46	8.46	9.23	9.23		
BUT4	8.78	8.78	9.59	9.59	8.46	8.46	9.23	9.23		
BUT5	8.78	8.78	9.59	9.59						
BUT6	8.78	8.78	9.59	9.59	8.46	8.46	9.23	9.23	7.67	7.67
BTOH_tr	8.78	8.78	9.59	9.59	8.46	8.46	9.23	9.23	7.67	7.67
BTOH_sink	8.78	8.78	9.59	9.59	8.46	8.46	9.23	9.23	7.67	7.67
HCO3E			9.59	9.59			9.23	9.23	7.67	7.67
ACCOAC			9.59	9.59			9.23	9.23	7.67	7.67
NPHT7			9.59	9.59			9.23	9.23		
BTBTAC					8.46	8.46	9.23	9.23		
CAR					8.46	8.46	9.23	9.23	7.67	7.67
ACPS1									0.00	0.00
BPNT									0.00	0.00
MCOATA									7.67	7.67
KAS15									7.67	7.67
3OAR40									7.67	7.67
3HAD40									7.67	7.67
EAR40x									7.67	7.67
5_BUT1									7.67	7.67
BTBTAC	0.17	0.17	0.18	0.18	0.10	0.10	0.00	0.00	0.13	0.13
BTAC_tr	0.00	0.00	0.00	0.00	8.46	8.46	9.23	9.23	7.67	7.67
BTAC_sink	7.60	7.60	7.60	7.60	7.60	7.60	2.15	4.62	7.60	7.60
BTAL_tr	0.00	0.00	0.00	0.00	14.38	14.38	13.85	13.85	23.22	23.22
BTAL_sink	0.16	0.16	0.05	0.05	0.09	0.09	0.00	0.00	0.12	0.12

Reaction ID	CROT		BUTYR		Butal	
	Min	Max	Min	Max	Min	Max
ACONT	0.165	0.165	0.165	0.165	0.165	0.165
ADK1	0	0	0	0	0	0
ATPM	7.6	7.6	7.6	7.6	7.6	7.6
ATPS4r	0	0	0	0	0	0
Biomass	0.153	0.153	0.153	0.153	0.153	0.153
CS	0.165	0.165	0.165	0.165	0.165	0.165
CYTBD	11.08	29.056	4.411	20.437	2.688	11.819
ENO	18.927	18.927	18.927	18.927	18.927	18.927
FBA	9.41	9.41	9.41	9.41	9.41	9.41
FUM	0	0	0	0	0	0
G6PDH2r	1.312	1.312	1.312	1.312	1.312	1.312
GAPD	19.156	19.156	19.156	19.156	19.156	19.156
GLCpts	10	10	10	10	10	10
GND	1.312	1.312	1.312	1.312	1.312	1.312
ICDHyr	0.165	0.165	0.165	0.165	0.165	0.165
ICL	0	0	0	0	0	0
MALS	0	0	0	0	0	0
MDH	-2.03	0	0	0	-2.03	0
ME1	0	2.03	0	0	0	2.03
NADH11	11.08	29.056	4.411	20.437	2.688	11.819
NADTRHD	0	0	0	0	0	0
PDH	0	17.975	1.949	17.975	8.844	17.975
PFK	9.41	9.41	9.41	9.41	9.41	9.41
PFL	0	17.975	0	16.026	0	9.131
PGI	8.657	8.657	8.657	8.657	8.657	8.657
PGK	-19.156	-19.156	-19.156	-19.156	-19.156	-19.156
PGL	1.312	1.312	1.312	1.312	1.312	1.312
PGM	-18.927	-18.927	-18.927	-18.927	-18.927	-18.927
PPC	2.468	2.468	2.468	2.468	2.468	2.468
PPCK	0	2.03	2.03	2.03	0	2.03
PYK	6.379	8.409	8.409	8.409	6.379	8.409
RPE	0.765	0.765	0.765	0.765	0.765	0.765
RPI	-0.547	-0.547	-0.547	-0.547	-0.547	-0.547
SUCD4	0	0	0	0	0	0
TALA	0.41	0.41	0.41	0.41	0.41	0.41
THD2	0	0	0	0	0	0
TKT1	0.41	0.41	0.41	0.41	0.41	0.41
TKT2	0.355	0.355	0.355	0.355	0.355	0.355
TPI	9.41	9.41	9.41	9.41	9.41	9.41
BUT1	8.618	8.618	8.618	8.618	8.618	8.618
BUT2	8.618	8.618	8.618	8.618	8.618	8.618
BUT3	8.618	8.618	8.618	8.618	8.618	8.618
BUT4			8.618	8.618	8.618	8.618
BUT5					8.618	8.618
B2CTCRO	8.618	8.618				
CROAC_tr	8.618	8.618				
CROT_sink	8.618	8.618				
BTBTAC			8.618	8.618		
BTAC_tr			8.618	8.618		
BTAC_sink			8.618	8.618		
BTAL_tr					8.618	8.618
BTAL_sink					8.618	8.618

Table 3: Anaerobic FVA estimates of the butanol and butanol precursor models after manual curation of high-flux futile cycles. Optimized for butanol or butanol precursor production, accordingly, using the *Escherichia coli* Core Model. Minimal and maximal range units are in mmol gDW^{-1} hr^{-1}. Fraction of optimum = 100%. Reactions are shown in no particular order.

Reaction ID	BuOH-0		BuOH-1		tpcBuOH		BuOH-2		fasBuOH	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
ACKr	0.00	0.00	0.00	0.00	0.00	0.00	-3.75	-3.75	-4.18	-4.18
ACONT	0.20	0.20	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00
ADHEr	5.53	5.53	1.34	1.34	0.00	0.00	0.00	0.00	0.05	0.05
ADK1	0.00	0.00	0.00	0.00	9.55	9.55	7.50	7.50	7.06	7.06
ATPM	7.60	7.60	7.60	7.60	0.00	0.00	0.00	0.00	0.00	0.00
ATPS4r	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-1.35	-1.35
Biomass	0.18	0.18	0.04	0.04	0.00	0.00	0.00	0.00	0.00	0.00
CS	0.20	0.20	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00
ENO	18.76	18.76	19.70	19.70	19.10	19.10	18.75	18.75	18.34	18.34
FBA	9.33	9.33	9.84	9.84	9.10	9.10	8.75	8.75	8.34	8.34
G6PDH2r	1.47	1.47	0.37	0.37	2.71	2.71	3.75	3.75	4.99	4.99
GAPD	19.03	19.03	19.77	19.77	19.10	19.10	18.75	18.75	18.34	18.34
GLCpts	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
GND	1.47	1.47	0.37	0.37	2.71	2.71	3.75	3.75	4.99	4.99
ICDHyr	0.20	0.20	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00
MDH	-0.17	-0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ME2	0.17	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PDH	14.03	14.03	18.55	18.55	13.68	13.68	3.75	3.75	0.00	0.00
PFK	9.33	9.33	9.84	9.84	9.10	9.10	8.75	8.75	8.34	8.34
PFL	3.61	3.61	0.89	0.89	5.42	5.42	15.00	15.00	18.34	18.34
PGI	8.50	8.50	9.63	9.63	7.29	7.29	6.25	6.25	5.01	5.01
PGK	-19.03	-19.03	-19.77	-19.77	-19.10	-19.10	-18.75	-18.75	-18.34	-18.34
PGL	1.47	1.47	0.37	0.37	2.71	2.71	3.75	3.75	4.99	4.99
PGM	-18.76	-18.76	-19.70	-19.70	-19.10	-19.10	-18.75	-18.75	-18.34	-18.34
PPC	0.69	0.69	0.69	0.69	0.00	0.00	0.00	0.00	0.00	0.00
PPCK	0.00	0.00	0.57	0.57	0.00	0.00	0.00	0.00	0.00	0.00
PYK	7.98	7.98	9.56	9.56	9.10	9.10	8.75	8.75	8.34	8.34
RPE	0.85	0.85	0.21	0.21	1.81	1.81	2.50	2.50	3.33	3.33
RPI	-0.62	-0.62	-0.15	-0.15	-0.90	-0.90	-1.25	-1.25	-1.66	-1.66
TALA	0.46	0.46	0.11	0.11	0.90	0.90	1.25	1.25	1.66	1.66
THD2	0.00	0.00	0.00	0.00	4.13	4.13	0.00	0.00	4.13	4.13
TKT1	0.46	0.46	0.11	0.11	0.90	0.90	1.25	1.25	1.66	1.66
TKT2	0.39	0.39	0.10	0.10	0.90	0.90	1.25	1.25	1.66	1.66
TPI	9.33	9.33	9.84	9.84	9.10	9.10	8.75	8.75	8.34	8.34
BUT1	5.62	5.62			9.55	9.55				
BUT2	5.62	5.62	8.94	8.94	9.55	9.55	7.50	7.50		
BUT3	5.62	5.62	8.94	8.94	9.55	9.55	7.50	7.50		
BUT4	5.62	5.62	8.94	8.94	9.55	9.55	7.50	7.50		
BUT5	5.62	5.62	8.94	8.94						
BUT6	5.62	5.62	8.94	8.94	9.55	9.55	7.50	7.50	7.06	7.06
BTOH_tr	5.62	5.62	8.94	8.94	9.55	9.55	7.50	7.50	7.06	7.06
BTOH_sink	5.62	5.61	8.94	8.94	9.55	9.55	7.50	7.50	7.06	7.06
HCO3E			8.94	8.94			7.50	7.50	7.06	7.06
ACCOAC			8.94	8.94			7.50	7.50	7.06	7.06
NPHT7			8.94	8.94			7.50	7.50	7.06	7.06
BTBTAC					9.55	9.55	7.50	7.50		
CAR					9.55	9.55	7.50	7.50		
MCOATA									7.06	7.06
KAS15									7.06	7.06
3OAR40									7.06	7.06
3HAD40									7.06	7.06
EAR40x									7.06	7.06
5_BUT1									7.06	7.06
ACKr	0.00	0.00	0.00	0.00	0.00	0.00	-3.75	-3.75	-4.18	-4.18
ACONT	0.20	0.20	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00
ADHEr	5.53	5.53	1.34	1.34	0.00	0.00	0.00	0.00	0.05	0.05
ADK1	0.00	0.00	0.00	0.00	9.55	9.55	7.50	7.50	7.06	7.06
ATPM	7.60	7.60	7.60	7.60	0.00	0.00	0.00	0.00	0.00	0.00

Reaction ID	CROT		BUTYR		BUTAL	
	Min	Max	Min	Max	Min	Max
ACONT	0.195	0.195	0.195	0.195	0.195	0.195
ADHEr	12.318	12.318	11.212	11.212	9.369	9.369
ATPM	7.6	7.6	7.6	7.6	7.6	7.6
Biomass	0.18	0.18	0.18	0.18	0.18	0.18
CS	0.195	0.195	0.195	0.195	0.195	0.195
ENO	18.735	18.735	18.735	18.735	18.735	18.735
FBA	9.305	9.305	9.305	9.305	9.305	9.305
G6PDH2r	1.547	1.547	1.547	1.547	1.547	1.547
GAPD	19.005	19.005	19.005	19.005	19.005	19.005
GLCpts	10	10	10	10	10	10
GND	1.547	1.547	1.547	1.547	1.547	1.547
H2Ot	1.348	1.348	1.348	1.348	-2.339	-2.339
ICDHyr	0.195	0.195	0.195	0.195	0.195	0.195
PDH	7.204	7.204	8.31	8.31	10.153	10.153
PFK	9.305	9.305	9.305	9.305	9.305	9.305
PFL	10.409	10.409	9.303	9.303	7.46	7.46
PGI	8.416	8.416	8.416	8.416	8.416	8.416
PGK	-19.005	-19.005	-19.005	-19.005	-19.005	-19.005
PGL	1.547	1.547	1.547	1.547	1.547	1.547
PGM	-18.735	-18.735	-18.735	-18.735	-18.735	-18.735
PIt	-0.664	-0.664	-0.664	-0.664	-0.664	-0.664
PPC	0.69	0.69	0.69	0.69	0.69	0.69
PPCK	0.173	0.173	0.173	0.173	0.173	0.173
PYK	8.124	8.124	8.124	8.124	8.124	8.124
RPE	0.902	0.902	0.902	0.902	0.902	0.902
RPI	-0.645	-0.645	-0.645	-0.645	-0.645	-0.645
TALA	0.483	0.483	0.483	0.483	0.483	0.483
TKT1	0.483	0.483	0.483	0.483	0.483	0.483
TKT2	0.418	0.418	0.418	0.418	0.418	0.418
TPI	9.305	9.305	9.305	9.305	9.305	9.305
BUT1	2.212	2.212	2.765	2.765	3.687	3.687
BUT2	2.212	2.212	2.765	2.765	3.687	3.687
BUT3	2.212	2.212	2.765	2.765	3.687	3.687
BUT4			2.765	2.765	3.687	3.687
BUT5					3.687	3.687
B2CTCRO	2.212	2.212				
CROAC_tr	2.212	2.212				
CROT_sink	2.212	2.212				
BTBTAC			2.765	2.765		
BTAC_tr			2.765	2.765		
BTAC_sink			2.765	2.765		
BTAL_tr					3.687	3.687
BTAL_sink					3.687	3.687

Table 4: pFBA flux distributions of the wild type, butanol and butanol precursor models under aerobic conditions. Used the *Escherichia coli* Core Model and pFBA for optimization, using either biomass formation or target production as the objective function, accordingly. Solutions were simulated under aerobic conditions (EX_o2_e_ = -10 mmol gDW^{-1} hr^{-1}) and otherwise unconstrained. Reactions shown in no particular order.

Reaction ID	WT	BUOH-0	BuOH-1	tpcBuOH	BuOH-2	fasBuOH	CROT	BUTYR	BUTAL
ADK1				9.58	9.05	8.96			
ACONT	6.27				0.47				
AKGDH	5.34				0.47				
ATPM	7.60	20	10	7.60	7.60	7.60	25	15	17.5
ATPS4r	39.75			7.60	15.70	16.56	5	-5	-2.5
Biomass	0.86								
CS	6.27				0.47				
CYTBD	44.69			5.07	11.42	12.53	10		
ENO	14.85	20	20	19.16	18.57	17.91	20	20	20
FBA	7.57	10	10	9.16	8.57	7.91	10	10	10
FUM	5.34				0.47				
G6PDH2r	4.72			2.53	4.29	6.27			
GAPD	16.14	20	20	19.16	18.57	17.91	20	20	20
GLCpts	10.00	10	10	10.00	10.00	10.00	10	10	10
GND	4.72			2.53	4.29	6.27			
ICDHyr	6.27				0.47				
MDH	5.34				0.47				
NADH11	39.36			5.07	10.94	12.53	10		
PDH	9.49	20	20	19.16	18.57	17.91			10
PFK	7.57	10	10	9.16	8.57	7.91	10	10	10
PFL							20	20	10
PGI	5.11	10	10	7.47	5.71	3.74	10	10	10
PGK	-16.14	-20	-20	-19.16	-18.57	-17.91	-20	-20	-20
PGL	4.72			2.53	4.29	6.27			
PGM	-14.85	-20	-20	-19.16	-18.57	-17.91	-20	-20	-20
PPC	2.47								
PYK	1.93	10	10	9.16	8.57	7.91	10	10	10
RPE	2.53			1.69	2.86	4.18			
RPI	-2.19			-0.84	-1.43	-2.09			
TALA	1.42			0.84	1.43	2.09			
THD2				4.51		5.38			
TKT1	1.42			0.84	1.43	2.09			
TKT2	1.11			0.84	1.43	2.09			
TPI	7.57	10	10	9.16	8.57	7.91	10	10	10
BUT1		10		9.58			10	10	10
BUT2		10	10	9.58	9.05		10	10	10
BUT3		10	10	9.58	9.05		10	10	10
BUT4		10	10	9.58	9.05			10	10
BUT5		10	10						10
BUT6		10	10	9.58	9.05	8.96			
BTOH_tr		10	10	9.58	9.05	8.96			
BTOH_sink		10	10	9.58	9.05	8.96			
HCO3E			10		9.05	8.96			
ACCOAC			10		9.05	8.96			
NPHT7			10		9.05				
BTBTAC				9.58	9.05			10	
CAR				9.58	9.05	8.96			
MCOATA						8.96			
KAS15						8.96			
3OAR40						8.96			
3HAD40						8.96			
EAR40x						8.96			
5_BUT1						8.96			
B2CTCRO							10		
CROAC_tr							10		
CROT_sink							10		
BTAC_tr								10	
BTAC_sink								10	
BTAL_tr									10
BTAL_sink									10

Table 5: pFBA flux distributions of the wild type, butanol and butanol precursor models under anaerobic conditions. Used the *Escherichia coli* Core Model and pFBA for optimization, using either biomass formation or target production as the objective function, accordingly. Solutions were simulated under anaerobic conditions (EX_o2_e_ = 0 mmol gDW^{-1} hr^{-1}) and otherwise unconstrained. Reactions shown in no particular order.

	WT	BuOH-0	BuOH-1	tpcBuOH	BuOH-2	fasBuOH	CROT	BUTYR	BUTAL
Biomass	0.241								
G6PDH2r				3.733	1.440	0.649			
ADK1				7.467	5.600	5.271			
NPHT7			10.0		5.600				
ATPM	7.600	20.0	10.0	7.60	7.60	7.600	16.667	15.0	17.5
RPE	-0.173			2.489	0.960	0.433			
GAPD	19.360	20.0	20.0	18.756	19.520	19.784	20.0	20.0	20.0
PPC	0.690								
TPI	9.760	10.0	10.0	8.756	9.520	9.784	10.0	10.0	10.0
TKT2	-0.130			1.244	0.480	0.216			
TKT1	-0.043			1.244	0.480	0.216			
PYK	8.184	10.0	10.0	8.756	9.520	9.784	10.0	10.0	10.0
BTBTAC				7.467	5.600			10.0	
ENO	19.0	20.0	20.0	18.756	19.520	19.784	20.0	20.0	20.0
FBA	9.760	10.0	10.0	8.756	9.520	9.784	10.0	10.0	10.0
ACONT	0.260								
HCO3E			10.0		5.60	5.271			
TALA	-0.043			1.244	0.480	0.216			
THD2	4.129				2.720	9.242			
ACKr	-8.297			-3.822	-8.320	-9.242			
BTOH_sink		10.0	10.0	7.467	5.60	5.271			
ICDHyr	0.260								
PDH		20.0	20.0	3.644					10.0
CS	0.260								
ATPS4r	-5.066			-0.044	-3.440	-5.614	-3.333	-5.0	-2.5.0
BUT3		10.0	10.0	7.467	5.60		6.667	10.0	10.0
ADHEr	8.042						6.667		
BUT2		10.0	10.0	7.467	5.60		6.667	10.0	10.0
BUT1		10.0		7.467			6.667	10.0	
BTOH_tr		10.0	10.0	7.467	5.60	5.271			
BUT6		10.0	10.0	7.467	5.600	5.271			
BUT5		10.0	10.0						10.0
BUT4		10.0	10.0	7.467	5.60			10.	10.0
CAR				7.467	5.60	5.271			
GLCpts	10.0	10.0	10.0	10.0	10.0	10.000	10.0	10.0	10.0
GND				3.733	1.440	0.649			
PGL				3.733	1.440	0.649			
PGM	-19.0	-20.0	-20.0	-18.756	-19.520	-19.784	-20.0	-20.0	-20.0
PGK	-19.360	-20.0	-20.0	-18.756	-19.520	-19.784	-20.0	-20.0	-20.0
PGI	9.951	10.0	10.0	6.267	8.560	9.351	10.0	10.0	10.0
RPI	-0.173			-1.244	-0.480	-0.216			
PFK	9.760	10.0	10.0	8.756	9.520	9.784	10.0	10.0	10.0
PFL	17.502			15.111	19.520	19.784	20.0	20.0	10.0
ACCOAC			10.0		5.60				
MCOATA						5.271			
3HAD40						5.271			
ACCOAC						5.271			
KAS15						5.271			
EAR40x						5.271			
5_BUT1						5.271			
3OAR40						5.271			
CROAC_tr							6.667		
CROAC_sink							6.667		
B2CTCRO							6.667		
BTAC_sink								10.0	
BTAC_tr								10.0	
BTAL_sink									10.0
BTAL_tr									10.0

Table 6: pFBA flux distributions of the butanol and butanol precursor models under aerobic conditions after manual curation of high-flux futile cycles. Used the *Escherichia coli* Core Model and pFBA for optimization, using target production as the objective function, accordingly. Solutions were simulated under aerobic conditions (EX_o2_e_ = -10 mmol gDW^{-1} hr^{-1}). Reactions shown in no particular order.

	BuOH-0	BuOH-1	tpcBuOH	BuOH-2	fasBuOH	CROT	BUTYR	BUTAL
EX_co2_b	18.099	19.408	22.229	23.077	24.356	1.039	2.988	9.883
EX_for_b		0.301				17.975	16.026	9.131
EX_glc_D_b	-10	-10	-10	-10	-10	-10	-10	-10
EX_h2o_b	9.616	9.9	30.862	32.308	32.235	9.937	3.268	10.163
EX_h_b	1.69	0.78	-15.956	-18.462	-21.757	19.594	17.645	10.75
EX_o2_b	-0.363	-0.324	-5.194	-4.615	-8.89	-5.54	-2.206	-1.344
EX_pi_b	-0.588	-0.167	-0.338		-0.431	-0.563	-0.563	-0.563
ACONT	0.172	0.179	0.099		0.126	0.165	0.165	0.165
ADK1			8.464	9.231	7.665			
ATPM	7.6	7.6	7.6	4.615	7.6	7.6	7.6	7.6
ATPS4r			14.377	13.846	23.219			
Biomass	0.16	0.045	0.092		0.117	0.153	0.153	0.153
CO2t	-18.099	-19.408	-22.229	-23.077	-24.356	-1.039	-2.988	-9.883
CS	0.172	0.179	0.099		0.126	0.165	0.165	0.165
CYTBD	0.725	0.648	10.388	9.231	17.78	11.08	4.411	2.688
ENO	19.316	19.812	17.944	18.462	16.624	18.927	18.927	18.927
FBA	9.821	9.955	8.235	8.462	6.994	9.41	9.41	9.41
FORt		-0.301				-17.975	-16.026	-9.131
FUM		0.13						
G6PDH2r	0.062		5.021	4.615	8.669	1.312	1.312	1.312
GAPD	19.555	19.88	18.082	18.462	16.799	19.156	19.156	19.156
GLCpts	10	10	10	10	10	10	10	10
GND	0.062		5.021	4.615	8.669	1.312	1.312	1.312
H2Ot	-9.616	-9.9	-30.862	-32.308	-32.235	-9.937	-3.268	-10.163
ICDHyr	0.172	0.049	0.099		0.126	0.165	0.165	0.165
ICL		0.13						
MALS		0.13						
MDH	-2.011	0.259				-2.03		-2.03
ME1						2.03		2.03
ME2	2.011							
NADH11	0.725	0.518	10.388	9.231	17.78	11.08	4.411	2.688
O2t	0.363	0.324	5.194	4.615	8.89	5.54	2.206	1.344
PDH	18.323	19.359	17.372	18.462	15.896		1.949	8.844
PFK	9.821	9.955	8.235	8.462	6.994	9.41	9.41	9.41
PFL		0.301				17.975	16.026	9.131
PGI	9.906	9.991	4.96	5.385	1.307	8.657	8.657	8.657
PGK	-19.555	-19.88	-18.082	-18.462	-16.799	-19.156	-19.156	-19.156
PGL	0.062		5.021	4.615	8.669	1.312	1.312	1.312
PGM	-19.316	-19.812	-17.944	-18.462	-16.624	-18.927	-18.927	-18.927
PIt	-0.588	-0.167	-0.338		-0.431	-0.563	-0.563	-0.563
PPC	2.468		2.468		2.468	2.468	2.468	2.468
PPCK			2.205		2.133		2.03	
PYK	6.764	9.788	7.633	8.462	6.227	6.379	8.409	6.379
RPE	-0.074	-0.033	3.281	3.077	5.695	0.765	0.765	0.765
RPI	-0.135	-0.033	-1.74	-1.538	-2.974	-0.547	-0.547	-0.547
22_2						24.727		
SUCCt2b						24.727		
SUCD1i		0.13						
SUCD4		0.13						
TALA	-0.008	-0.008	1.657	1.538	2.869	0.41	0.41	0.41
THD2	0.605	0.776						
TKT1	-0.008	-0.008	1.657	1.538	2.869	0.41	0.41	0.41
TKT2	-0.066	-0.024	1.624	1.538	2.827	0.355	0.355	0.355
TPI	9.821	9.955	8.235	8.462	6.994	9.41	9.41	9.41
HCO3E		9.591		9.231	7.665			
ACCOAC		9.591		9.231	7.665			
NPHT7		9.591		9.231				

	BuOH-0	BuOH-1	tpcBuOH	BuOH-2	fasBuOH	CROT	BUTYR	BUTAL
MCOATA					7.665			
KAS15					7.665			
3OAR40					7.665			
3HAD40					7.665			
EAR40x					7.665			
5_BUT1					7.665			
BUT1	8.776		8.464			8.618	8.618	8.618
BUT2	8.776	9.591	8.464	9.231		8.618	8.618	8.618
BUT3	8.776	9.591	8.464	9.231		8.618	8.618	8.618
B2CTCRO						8.618		
CROAC_tr						8.618		
CROT_sink						8.618		
BUT4	8.776	9.591	8.464	9.231			8.618	8.618
BTBTAC			8.464	9.231			8.618	
CAR			8.464	9.231	7.665			
BUT5	8.776	9.591						8.618
BUT6	8.776	9.591	8.464	9.231	7.665			
BTOH_tr	8.776	9.591	8.464	9.231	7.665		8.618	
BTOH_sink	8.776	9.591	8.464	9.231	7.665		8.618	
BTAL_tr								8.618
BTAL_sink								8.618

Table 7: pFBA flux distributions of the butanol and butanol precursor models under anaerobic conditions after manual curation of high-flux futile cycles. Used the *Escherichia coli* Core Model and pFBA for optimization, using target production as the objective function, accordingly. Solutions were simulated under anaerobic conditions (EX_o2_e_ = 0 mmol gDW^{-1} hr^{-1}). Reactions shown in no particular order.

	BuOH-0	BuOH-1	tpcBuOH	BuOH-2	fasBuOH	CROT	BUTYR	BUTAL
EX_ac_b				3.75	4.177			
EX_co2_b	15.166	18.834	16.387	7.5	4.991	8.428	9.534	11.377
EX_etoh_b	5.525	1.342			0.047	12.318	11.212	9.369
EX_for_b	3.611	0.891	5.419	15	18.336	10.409	9.303	7.46
EX_glc_D_b	-10	-10	-10	-10	-10	-10	-10	-10
EX_h2o_b	4.354	8.626	25.936	15	12	-1.348	-1.348	2.339
EX_h_b	5.525	1.342	-13.678	3.75	1.345	12.318	11.212	9.369
EX_pi_b	-0.665	-0.157				-0.664	-0.664	-0.664
ACKr				-3.75	-4.177			
ACt2r				-3.75	-4.177			
ADK1			9.548	7.5	7.056			
ACONT	0.195	0.046				0.195	0.195	0.195
ADHEr	5.525	1.342			0.047	12.318	11.212	9.369
ATPM	7.6	7.6				7.6	7.6	7.6
ATPS4r					-1.345			
Biomass	0.181	0.043				0.18	0.18	0.18
CO2t	-15.166	-18.834	-16.387	-7.5	-4.991	-8.428	-9.534	-11.377
CS	0.195	0.046				0.195	0.195	0.195
ENO	18.76	19.701	19.097	18.75	18.336	18.735	18.735	18.735
ETOHt2r	-5.525	-1.342			-0.047	-12.318	-11.212	-9.369

	BuOH-0	BuOH-1	tpcBuOH	BuOH-2	fasBuOH	CROT	BUTYR	BUTAL
FBA	9.332	9.836	9.097	8.75	8.336	9.305	9.305	9.305
FORt	-3.611	-0.891	-5.419	-15	-18.336	-10.409	-9.303	-7.46
G6PDH2r	1.465	0.365	2.71	3.75	4.991	1.547	1.547	1.547
GAPD	19.031	19.765	19.097	18.75	18.336	19.005	19.005	19.005
GLCpts	10	10	10	10	10	10	10	10
GND	1.465	0.365	2.71	3.75	4.991	1.547	1.547	1.547
H2Ot	-4.354	-8.626	-25.936	-15	-12	1.348	1.348	-2.339
ICDHyr	0.195	0.046				0.195	0.195	0.195
MDH	-0.172							
ME2	0.172							
PDH	14.025	18.545	13.678	3.75		7.204	8.31	10.153
PFK	9.332	9.836	9.097	8.75	8.336	9.305	9.305	9.305
PFL	3.611	0.891	5.419	15	18.336	10.409	9.303	7.46
PGI	8.498	9.626	7.29	6.25	5.009	8.416	8.416	8.416
PGK	-19.031	-19.765	-19.097	-18.75	-18.336	-19.005	-19.005	-19.005
PGL	1.465	0.365	2.71	3.75	4.991	1.547	1.547	1.547
PGM	-18.76	-19.701	-19.097	-18.75	-18.336	-18.735	-18.735	-18.735
PIt	-0.665	-0.157				-0.664	-0.664	-0.664
PPC	0.69	0.69				0.69	0.69	0.69
PPCK		0.568				0.173	0.173	0.173
PTAr				3.75	4.177			
PYK	7.976	9.557	9.097	8.75	8.336	8.124	8.124	8.124
RPE	0.847	0.213	1.806	2.5	3.328	0.902	0.902	0.902
RPI	-0.618	-0.152	-0.903	-1.25	-1.664	-0.645	-0.645	-0.645
22_2			5.419					
SUCCt2b			5.419					
TALA	0.456	0.114	0.903	1.25	1.664	0.483	0.483	0.483
THD2			4.129		4.129			
TKT1	0.456	0.114	0.903	1.25	1.664	0.483	0.483	0.483
TKT2	0.391	0.099	0.903	1.25	1.664	0.418	0.418	0.418
TPI	9.332	9.836	9.097	8.75	8.336	9.305	9.305	9.305
HCO3E		8.944		7.5	7.056			
ACCOAC		8.944		7.5	7.056			
MCOATA					7.056			
KAS15					7.056			
3OAR40					7.056			
3HAD40					7.056			
EAR40x					7.056			
5_BUT1					7.056			
NPHT7		8.944		7.5				
BUT1	5.619		9.548			2.212	2.765	3.687
BUT2	5.619	8.944	9.548	7.5		2.212	2.765	3.687
BUT3	5.619	8.944	9.548	7.5		2.212	2.765	3.687
BUT4	5.619	8.944	9.548	7.5			2.765	3.687
BTBTAC			9.548	7.5			2.765	
CAR			9.548	7.5	7.056			
BUT5	5.619	8.944						3.687
BUT6	5.619	8.944	9.548	7.5	7.056			
BTOH_tr	5.619	8.944	9.548	7.5	7.056			
BTOH_sink	5.619	8.944	9.548	7.5	7.056			
B2CTCRO						2.212		
CROAC_tr						2.212		
CROT_sink						2.212		
BUTYR_tr							2.765	
BUTYR_sink							2.765	
BTAL_tr								3.687
BTAL_sink								3.687

Table 8: pFBA flux distributions of the wild type, butanol and butanol precursor models constrained using ^{13}C-MFA data. Used the *Escherichia coli* Core Model and pFBA for optimization, using either biomass formation or target production as the objective function, accordingly. Solutions were simulated under aerobic conditions (EX_o2_e_ = -10 mmol gDW^{-1} hr^{-1}) and constrained using ^{13}C-MFA data from Long and Antoniewicz (2019). Reactions shown in no particular order.

	WT	BuOH-0	BuOH-1	tpcBuOH	BuOH-2	fasBuOH	CROT	BUTYR	BUTAL
ACONT	4.623	0.382	0.381	0.382	0.63	0.707	0.381	0.382	0.382
ADK1				3.595	3.595	3.595			
AKGDH	0.061	0.061	0.061	0.061	0.061	0.061	0.061	0.061	0.061
AKGt2r	-2.132				-0.248	-0.326			
ATPM	7.6	15.024	11.429	7.834	7.6	7.6	18.143	15.024	15.024
ATPS4r	25.821	7.417	7.417	7.417	10.529	10.452	10.536	7.417	7.417
Biomass	0.524								
CS	4.623	0.382	0.381	0.382	0.63	0.707	0.381	0.382	0.382
CYTBD	29.476	7.192	7.192	5.754	8.493	8.803	12.238	8.643	7.911
ENO	14.113	7.571	7.571	7.571	8.068	8.223	7.571	7.571	7.571
FBA	7.19	3.352	3.351	3.352	3.6	3.677	3.352	3.352	3.352
FUM	1.926	0.382	0.381	0.382	0.381	0.382	0.381	0.382	0.382
G6PDH2r	2.605	2.605	2.605	2.605	2.605	2.605	2.605	2.605	2.605
GAPD	14.898	7.571	7.571	7.571	8.068	8.223	7.571	7.571	7.571
GLCpts	8.58	4.22	4.22	4.22	4.468	4.546	4.22	4.22	4.22
GND	2.605	2.605	2.605	2.605	2.605	2.605	2.605	2.605	2.605
ICDHyr	2.758	0.061	0.061	0.061	0.309	0.387	0.061	0.061	0.061
ICL	1.865	0.32	0.321	0.32	0.321	0.321	0.321	0.32	0.32
MALS	1.865	0.32	0.321	0.32	0.321	0.321	0.321	0.32	0.32
MDH	3.445	0.356	0.356	0.356	0.356	0.356	0.356	0.356	0.356
ME1	0.302	0.302	0.302	0.302	0.302	0.302	0.302	0.302	0.302
ME2	0.044	0.044	0.044	0.044	0.044	0.044	0.044	0.044	0.044
NADH11	27.55	6.811	6.811	5.373	8.111	8.421	11.856	8.262	7.53
NADTRHD		4.331	4.331	0.736	0.985		5.315	4.363	4.331
PDH	8.53	7.585	7.585	6.147	8.14	8.218	1.846	1.846	4.709
PFK	7.19	3.352	3.351	3.352	3.6	3.677	3.352	3.352	3.352
PFL		0.307	0.307	1.745			6.046	6.046	3.183
PGI	5.867	1.615	1.615	1.615	1.863	1.941	1.615	1.615	1.615
PGK	-14.898	-7.571	-7.571	-7.571	-8.068	-8.223	-7.571	-7.571	-7.571
PGL	2.605	2.605	2.605	2.605	2.605	2.605	2.605	2.605	2.605
PGM	-14.113	-7.571	-7.571	-7.571	-8.068	-8.223	-7.571	-7.571	-7.571
PPC	2.115	0.026	0.025	0.026	0.274	0.351	0.025	0.026	0.026
PYK	3.146	3.326	3.326	3.326	3.326	3.326	3.326	3.326	3.326
RPE	1.36	1.737	1.737	1.737	1.737	1.737	1.737	1.737	1.737
RPI	-1.245	-0.868	-0.868	-0.868	-0.868	-0.868	-0.868	-0.868	-0.868
TALA	0.775	0.868	0.868	0.868	0.868	0.868	0.868	0.868	0.868
THD2	1.545	-0.984	-0.984	-0.984	-0.984	1.549		-0.952	-0.984
TKT1	0.775	0.868	0.868	0.868	0.868	0.868	0.868	0.868	0.868
TKT2	0.585	0.868	0.868	0.868	0.868	0.868	0.868	0.868	0.868
TPI	7.19	3.352	3.351	3.352	3.6	3.677	3.351	3.352	3.352
HCO3E			3.595		3.595	3.595			
ACCOAC			3.595		3.595	3.595			
NPHT7			3.595		3.595				
BUT1		3.595		3.595			3.595	3.595	3.595
BUT2		3.595	3.595	3.595	3.595		3.595	3.595	3.595
BUT3		3.595	3.595	3.595	3.595		3.595	3.595	3.595
BUT4		3.595	3.595	3.595	3.595			3.595	3.595
BTBTAC				3.595	3.595			3.595	
CAR				3.595	3.595	3.595			
BUT5		3.595	3.595						3.595
BUT6		3.595	3.595	3.595	3.595	3.595			
BTOH_tr		3.595	3.595	3.595	3.595	3.595			
BTOH_sink		3.595	3.595	3.595	3.595	3.595			
MCOATA						3.595			
KAS15						3.595			
3OAR40						3.595			
3HAD40						3.595			
EAR40x						3.595			
5_BUT1						3.595			
B2CTCRO							3.595		
CROAC_tr							3.595		
CROT_sink							3.595		
BTAC_tr								3.595	
BTAC_sink								3.595	
BTAL_tr									3.595
BTAL_sink									3.595

Table 9: pFBA flux distributions of butanol and butanol precursor models constrained using ^{13}C-MFA data and additional manual curation of high-flux futile reactions. Used the *Escherichia coli* Core Model and pFBA for optimization, using target production as the objective function. Solutions were simulated under aerobic conditions (EX_o2_e_ = -10 mmol gDW^{-1} hr^{-1}), constrained using ^{13}C-MFA data from Long and Antoniewicz (2019). High-flux futile cycles capped are shown in the first row. Reaction fluxes (with reaction IDs shown in the first column) are shown in no particular order.

Reactions capped	BuOH-0 ATPM, FBP	BuOH-1 ATPM, FBP	tpcBuOH ATPM, PPCK, FBP	CROT ATPM, FBP	BUTYR ATPM, FBP	BUTAL ATPM, FBP
ACONT	0.519	0.442	0.387	0.519	0.519	0.519
ADK1			3.595			
AKGDH	0.061	0.061	0.061	0.061	0.061	0.061
ATPM	7.6	7.6	7.6	7.6	7.6	7.6
ATPS4r	7.417	7.417	7.417	7.417	7.417	7.417
Biomass	0.128	0.056	0.005	0.128	0.128	0.128
CS	0.519	0.442	0.387	0.519	0.519	0.519
CYTBD	7.559	7.352	5.768	12.095	8.994	8.276
ENO	8.967	8.18	7.623	8.967	8.967	8.967
FBA	4.188	3.716	3.382	4.188	4.188	4.188
FUM	0.382	0.382	0.382	0.381	0.382	0.382
G6PDH2r	2.605	2.605	2.605	2.605	2.605	2.605
GAPD	9.158	8.263	7.63	9.158	9.158	9.158
GLCpts	5.183	4.64	4.255	5.183	5.183	5.183
GND	2.605	2.605	2.605	2.605	2.605	2.605
ICDHyr	0.199	0.121	0.066	0.199	0.199	0.199
ICL	0.32	0.32	0.32	0.321	0.32	0.32
MALS	0.32	0.32	0.32	0.321	0.32	0.32
MDH	0.356	0.356	0.356	0.356	0.356	0.356
ME1	0.302	0.302	0.302	0.302	0.302	0.302
ME2	0.044	0.044	0.044	0.044	0.044	0.044
NADH11	7.177	6.971	5.387	11.713	8.612	7.895
NADTRHD	2.14	3.376	0.656	8.1	2.14	2.14
PDH	8.072	7.797	6.165	1.846	2.332	5.202
PFK	4.188	3.716	3.382	4.188	4.188	4.188
PFL	0.421	0.357	1.749	6.647	6.161	3.291
PGI	2.552	2.023	1.649	2.552	2.552	2.552
PGK	-9.158	-8.263	-7.63	-9.158	-9.158	-9.158
PGL	2.605	2.605	2.605	2.605	2.605	2.605
PGM	-8.967	-8.18	-7.623	-8.967	-8.967	-8.967
PPC	1.45	1.243	0.039	1.45	1.45	1.45
PPCK	1.058	1.058		1.058	1.058	1.058
PYK	3.326	3.326	3.326	3.326	3.326	3.326
RPE	1.645	1.697	1.733	1.645	1.645	1.645
RPI	-0.96	-0.908	-0.872	-0.96	-0.96	-0.96
TALA	0.845	0.858	0.867	0.845	0.845	0.845
THD2	-0.984	-0.984	-0.984	4.975	-0.984	-0.984
TKT1	0.845	0.858	0.867	0.845	0.845	0.845
TKT2	0.799	0.838	0.866	0.799	0.799	0.799
TPI	4.188	3.716	3.382	4.188	4.188	4.188
HCO3E		3.592				
ACCOAC		3.592				
NPHT7		3.592				
BUT1	3.587		3.595	3.587	3.587	3.587
BUT2	3.587	3.592	3.595	3.587	3.587	3.587
BUT3	3.587	3.592	3.595	3.587	3.587	3.587
BUT4	3.587	3.592	3.595		3.587	3.587
BTBTAC			3.595		3.587	3.587
CAR			3.595			
BUT5	3.587	3.592				
BUT6	3.587	3.592	3.595			
BTOH_tr	3.587	3.592	3.595			
BTOH_sink	3.587	3.592	3.595			
B2CTCRO				3.587		
CROAC_tr				3.587		
CROT_sink				3.587		
BTAC_tr					3.587	
BTAC_sink					3.587	
BTAL_tr						3.587
BTAL_sink						3.587

Table 10: CBA parameters and outputs of unconstrained butanol and butanol precursor models under aerobic conditions. Reaction IDs, relevant co-factors and their stoichiometric coefficients, flux values, balance values and assigned balance category are included.

Reaction ID	Co-factor	Stoich. coefficient	Flux Distribution	Balance Value	Balance Category
Wild Type (WT)					
ATPM	ATP	-1	7.6	-7.6	Waste
Biomass	ATP	-55.703	0.87	-47.983	Biomass
SUCOAS	ATP	-1	-5.336	5.336	Production
PYK	ATP	1	1.935	1.935	Production
ATPS4r	ATP	1	39.747	39.747	Production
PGK	ATP	-1	-16.138	16.138	Production
PFK	ATP	-1	7.571	-7.571	Maintenance
G6PDH2r	NADPH	1	4.717	4.717	Waste
ICDHyr	NADPH	1	6.265	6.265	Waste
AKGDH	NADH	1	5.336	5.336	Waste
GND	NADPH	1	4.717	4.717	Waste
PDH	NADH	1	9.493	9.493	Waste
GAPD	NADH	1	16.138	16.138	Production
Biomass	NADH/NADPH	3.547/-18.225	0.87	-15.699	Biomass
BuOH-0					
PFK	ATP	-1	10	-10	Maintenance
PGK	ATP	-1	-20	20	Production
PYK	ATP	1	10	10	Production
ATPM	ATP	-1	20	-20	Waste
GAPD	NADH	1	20	20	Production
BUT2	NADH	-1	10	-10	Target
BUT4	NADH	-1	10	-10	Target
BUT5	NADH	-1	10	-10	Target
BUT6	NADH	-1	10	-10	Target
PDH	NADH	1	20	20	Waste
BuOH-1					
PFK	ATP	-1	10	-10	Maintenance
PGK	ATP	-1	-20	20	Production
PYK	ATP	1	10	10	Production
ATPM	ATP	-1	10	-10	Waste
ACCOAC	ATP	-1	10	-10	Target
GAPD	NADH	1	20	20	Production
BUT2	NADH	-1	10	-10	Target
BUT4	NADH	-1	10	-10	Target
BUT5	NADH	-1	10	-10	Target
BUT6	NADH	-1	10	-10	Target
PDH	NADH	1	20	20	Waste
tpcBuOH					
PYK	ATP	1	9.156	9.156	Production
ATPS4r	ATP	1	7.6	7.6	Production
PGK	ATP	-1	-19.156	19.156	Production
CAR	ATP	-1	9.578	-9.578	Target
PFK	ATP	-1	9.156	-9.156	Maintenance
ATPM	ATP	-1	9.578	-9.578	Waste
ADK1	ATP	-1	9.578	-9.578	Waste
GAPD	NADH	1	19.156	19.156	Production
THD2	NADPH	1	4.511	4.511	Production
G6PDH2r	NADPH	1	2.533	2.533	Production
CAR	NADPH	-1	9.578	-9.578	Target
BUT2	NADH	-1	9.578	-9.578	Target
BUT4	NADH	-1	9.578	-9.578	Target
BUT6	NADH	-1	9.578	-9.578	Target
GND	NADPH	1	2.533	2.533	Waste
PDH	NADH	1	19.156	19.156	Waste
NADH11	NADH	-1	5.067	-5.067	Maintenance
THD2	NADH	-1	4.511	-4.511	Maintenance
BuOH-2					
SUCOAS	ATP	1	0.473	0.473	Production
PYK	ATP	1	8.571	8.571	Production
ATPS4r	ATP	1	15.702	15.702	Production
PGK	ATP	-1	-18.571	18.571	Production

Reaction ID	Co-factor	Stoich. coefficient	Flux Distribution	Balance Value	Balance Category
BuOH-2					
CAR	ATP	-1	9.049	-9.049	Target
ACCOAC	ATP	-1	9.049	-9.049	Target
PFK	ATP	-1	8.571	-8.571	Maintenance
ATPM	ATP	-1	7.6	-7.6	Waste
ADK1	ATP	-1	9.049	-9.049	Waste
MDH	NADH	1	0.473	0.473	Production
GAPD	NADH	1	18.570	18.570	Production
G6PDH2r	NADPH	1	4.288	4.288	Production
CAR	NADPH	-1	9.049	-9.049	Target
BUT2	NADH	-1	9.049	-9.049	Target
BUT4	NADH	-1	9.049	-9.049	Target
BUT6	NADH	-1	9.049	-9.049	Target
NADH11	NADH	-1	10.941	-10.941	Maintenance
ICDHyr	NADH	1	0.473	0.473	Waste
AKGDH	NADH	1	0.473	0.473	Waste
GND	NADPH	1	4.288	4.288	Waste
PDH	NADH	1	18.571	18.571	Waste
fasBuOH					
PYK	ATP	1	7.912	7.912	Production
ATPS4r	ATP	1	16.556	16.556	Production
PGK	ATP	-1	-17.912	17.912	Production
CAR	ATP	-1	8.956	-8.956	Target
ACCOAC	ATP	-1	8.956	-8.956	Target
PFK	ATP	-1	7.912	-7.912	Maintenance
ADK1	ATP	-1	8.956	-8.956	Waste
ATPM	ATP	-1	7.6	-7.6	Waste
G6PDH2r	NADPH	1	6.265	6.265	Production
GAPD	NADH	1	17.912	17.912	Production
THD2	NADPH	1	5.381	5.381	Production
CAR	NADPH	-1	8.956	-8.956	Target
BUT6	NADH	-1	8.956	-8.956	Target
30AR40	NADPH	-1	8.956	-8.956	Maintenance
EAR40x	NADH	-1	8.956	-8.956	Maintenance
NADH11	NADH	-1	12.530	-12.530	Maintenance
THD2	NADH	-1	5.381	-5.381	Maintenance
GND	NADPH	1	6.265	6.265	Waste
PDH	NADH	1	17.912	17.912	Waste
CROT					
PYK	ATP	1	10	10	Production
ATPS4r	ATP	1	5	5	Production
PGK	ATP	-1	-20	20	Production
PFK	ATP	-1	10	-10	Maintenance
ATPM	ATP	-1	25	-25	Waste
GAPD	NADH	1	20	20	Production
BUT2	NADH	-1	10	-10	Target
NADH11	NADH	-1	10	-10	Maintenance
BUTYR					
PYK	ATP	1	10	10	Production
PGK	ATP	-1	-20	20	Production
ATPS4r	ATP	1	-5	-5	Maintenance
PFK	ATP	-1	10	-10	Maintenance
ATPM	ATP	-1	15	-15	Waste
GAPD	NADH	1	20	20	Production
BUT2	NADH	-1	10	-10	Target
BUT4	NADH	-1	10	-10	Target
BUTAL					
PYK	ATP	1	10	10	Production
PGK	ATP	-1	-20	20	Production
ATPS4r	ATP	1	-2.5	-2.5	Maintenance
PFK	ATP	-1	10	-10	Maintenance
ATPM	ATP	-1	17.5	-17.5	Waste
GAPD	NADH	1	20	20	Production
PDH	NADH	1	20	20	Waste
BUT2	NADH	-1	10	-10	Target
BUT4	NADH	-1	10	-10	Target
BUT5	NADH	-1	10	-10	Target

Table 11: CBA parameters and outputs of unconstrained butanol and butanol precursor models under anaerobic conditions. Reaction IDs, relevant co-factors and their stoichiometric coefficients, flux values, balance values and assigned balance category are included.

Reaction ID	Co-factor	Stoich. coefficient	Flux Distribution	Balance Value	Balance Category
Wild Type (WT)					
ATPM	ATP	1	7.6	-7.6	Waste
ACKr	ATP	-1	-8.297	8.297	Waste
Biomass	ATP	-55.703	0.241	-13.415	Biomass
PYK	ATP	1	8.184	8.184	Production
ATPS4r	ATP	1	-5.067	-5.067	Maintenance
PGK	ATP	-1	-19.360	19.360	Production
PFK	ATP	-1	9.761	-9.761	Maintenance
THD2	NADPH	1	4.129	4.129	Production
ICDHyr	NADPH	1	0.260	0.260	Waste
THD2	NADH	-1	4.129	-4.129	Maintenance
ADHEr	NADH	-2	8.042	-16.085	Waste
GAPD	NADH	1	19.340	19.360	Production
Biomass	NADH/NADPH	3.547/-18.225	0.241	-3.535	Biomass
BuOH-0					
PFK	ATP	-1	10	-10	Maintenance
PGK	ATP	-1	-20	20	Production
PYK	ATP	1	10	10	Production
ATPM	ATP	-1	20	-20	Waste
GAPD	NADH	1	20	20	Production
BUT2	NADH	-1	10	-10	Target
BUT4	NADH	-1	10	-10	Target
BUT5	NADH	-1	10	-10	Target
BUT6	NADH	-1	10	-10	Target
PDH	NADH	1	20	20	Waste
BuOH-1					
PFK	ATP	-1	10	-10	Maintenance
PGK	ATP	-1	-20	20	Production
PYK	ATP	1	10	10	Production
ATPM	ATP	-1	10	-10	Waste
ACCOAC	ATP	-1	10	-10	Target
GAPD	NADH	1	20	20	Production
BUT2	NADH	-1	10	-10	Target
BUT4	NADH	-1	10	-10	Target
BUT5	NADH	-1	10	-10	Target
BUT6	NADH	-1	10	-10	Target
PDH	NADH	1	20	20	Waste
tpcBuOH					
PYK	ATP	1	8.756	8.756	Production
ATPS4r	ATP	1	7.6	7.6	Production
PGK	ATP	-1	-18.756	18.756	Production
CAR	ATP	-1	7.467	-7.467	Target
ATPS4r	ATP	1	-0.045	-0.045	Maintenance
PFK	ATP	-1	8.756	-8.756	Maintenance
ATPM	ATP	-1	7.6	-7.6	Waste
ADK1	ATP	-1	7.467	-7.467	Waste
ACKr	ATP	-1	-3.822	3.822	Waste
G6PDH2r	NADPH	1	3.733	3.733	Production
GAPD	NADH	1	18.756	18.756	Production
CAR	NADPH	-1	7.467	-7.467	Target
BUT2	NADH	-1	7.467	-7.467	Target
BUT4	NADH	-1	7.467	-7.467	Target
BUT6	NADH	-1	7.467	-7.467	Target
GND	NADPH	1	3.733	3.733	Waste
PDH	NADH	1	3.644	3.644	Waste
BuOH-2					
PYK	ATP	1	9.520	9.520	Production
ATPS4r	ATP	1	-3.440	-3.440	Maintenance
PGK	ATP	-1	-19.520	19.520	Production
CAR	ATP	-1	5.6	-5.6	Target
ACCOAC	ATP	-1	5.6	-5.6	Target

Reaction ID	Co-factor	Stoich. coefficient	Flux Distribution	Balance Value	Balance Category
BuOH-2					
PFK	ATP	-1	9.52	-9.52	Maintenance
ATPM	ATP	-1	7.6	-7.6	Waste
ADK1	ATP	-1	5.6	-5.6	Waste
ACKr	ATP	-1	-8.32	8.32	Waste
GAPD	NADH	1	19.520	19.520	Production
G6PDH2r	NADPH	1	1.440	1.440	Production
THD2	NADPH	1	2.720	2.720	Production
CAR	NADPH	-1	5.6	-5.6	Target
BUT2	NADH	-1	5.6	-5.6	Target
BUT4	NADH	-1	5.6	-5.6	Target
BUT6	NADH	-1	5.6	-5.6	Target
THD2	NADH	-1	2.720	-2.720	Maintenance
GND	NADPH	1	1.439	1.439	Waste
fasBuOH					
PYK	ATP	1	9.784	9.784	Production
ATPS4r	ATP	1	-5.614	-5.614	Maintenance
PGK	ATP	-1	-19.784	19.784	Production
CAR	ATP	-1	5.271	-5.271	Target
ACCOAC	ATP	-1	5.271	-5.271	Target
PFK	ATP	-1	9.784	-9.784	Maintenance
ADK1	ATP	-1	5.271	-5.271	Waste
ATPM	ATP	-1	7.6	-7.6	Waste
ACKr	ATP	-1	-9.242	9.242	Waste
G6PDH2r	NADPH	1	0.649	0.649	Production
GAPD	NADH	1	19.784	19.784	Production
THD2	NADPH	1	9.242	9.242	Production
CAR	NADPH	-1	5.271	-5.271	Target
BUT6	NADH	-1	5.271	-5.271	Target
30AR40	NADPH	-1	5.271	-5.271	Maintenance
EAR40x	NADH	-1	5.271	-5.271	Maintenance
THD2	NADH	-1	9.242	-9.242	Maintenance
GND	NADPH	1	0.649	0.649	Waste
CROT					
PYK	ATP	1	10	10	Production
ATPS4r	ATP	1	-3.333	-3.333	Maintenance
PGK	ATP	-1	-20	20	Production
PFK	ATP	-1	10	-10	Maintenance
ATPM	ATP	-1	16.667	-16.667	Waste
GAPD	NADH	1	20	20	Production
BUT2	NADH	-1	6.667	-6.667	Target
ADHEr	NADH	-1	13.333	-13.333	Waste
BUTYR					
PYK	ATP	1	10	10	Production
PGK	ATP	-1	-20	20	Production
ATPS4r	ATP	1	-5	-5	Maintenance
PFK	ATP	-1	10	-10	Maintenance
ATPM	ATP	-1	15	-15	Waste
GAPD	NADH	1	20	20	Production
BUT2	NADH	-1	10	-10	Target
BUT4	NADH	-1	10	-10	Target
BUTAL					
PYK	ATP	1	10	10	Production
PGK	ATP	-1	-20	20	Production
ATPS4r	ATP	1	-2.5	-2.5	Maintenance
PFK	ATP	-1	10	-10	Maintenance
ATPM	ATP	-1	17.5	-17.5	Waste
GAPD	NADH	1	20	20	Production
PDH	NADH	1	10	10	Waste
BUT2	NADH	-1	10	-10	Target
BUT4	NADH	-1	10	-10	Target
BUT5	NADH	-1	10	-10	Target

Table 12: CBA parameters and outputs of manually curated butanol and butanol precursor models under aerobic conditions. Reaction IDs, relevant co-factors and their stoichiometric coefficients, flux values, balance values and assigned balance category are included.

Reaction ID	Co-factor	Stoich. coefficient	Flux Distribution	Balance Value	Balance Category
BuOH-0					
PFK	ATP	-1	9.821	-9.821	Maintenance
PGK	ATP	-1	-19.555	19.555	Production
PYK	ATP	1	6.764	6.764	Production
ATPM	ATP	-1	7.60	-7.60	Waste
Biomass	ATP	-55.703	0.160	-8.899	Biomass
GAPD	NADH	1	19.555	19.555	Production
G6PDH2r	NADPH	1	0.062	0.062	Production
THD2	NADPH	1	0.605	0.605	Production
BUT2	NADH	-1	8.776	-8.776	Target
BUT4	NADH	-1	8.776	-8.776	Target
BUT5	NADH	-1	8.776	-8.776	Target
BUT6	NADH	-1	8.776	-8.776	Target
MDH	NADH	1	-2.011	-2.011	Maintenance
NADH11	NADH	-1	0.725	-0.725	Maintenance
THD2	NADH	-1	0.605	-0.605	Maintenance
PDH	NADH	1	18.323	18.323	Waste
ICDHyr	NADH	1	0.172	0.172	Waste
ME2	NADH	1	2.011	2.011	Waste
GND	NADH	1	0.062	0.062	Waste
Biomass	NADH/NADPH	3.547/-18.225	0.160	-2.345	Biomass
BuOH-1					
PFK	ATP	-1	9.924	-9.924	Maintenance
PGK	ATP	-1	-19.855	19.855	Production
PYK	ATP	1	9.652	9.652	Production
ATPM	ATP	-1	7.60	-7.60	Waste
ACCOAC	ATP	-1	9.667	-9.667	Target
Biomass	ATP	-55.703	0.045	-2.316	Biomass
GAPD	NADH	1	19.855	19.855	Production
G6PDH2r	NADPH	1	0.103	0.103	Production
THD2	NADPH	1	0.506	0.506	Production
BUT2	NADH	-1	9.667	-9.667	Target
BUT4	NADH	-1	9.667	-9.667	Target
BUT5	NADH	-1	9.667	-9.667	Target
BUT6	NADH	-1	9.667	-9.667	Target
NADH11	NADH	-1	0.363	-0.363	Maintenance
THD2	NADH	-1	0.506	-0.506	Maintenance
ICDHyr	NADH	1	0.049	0.049	Waste
GND	NADPH	1	0.103	0.103	Waste
PDH	NADH	1	19.534	19.534	Waste
Biomass	NADH/NADPH	3.457/ -18.225	0.045	-0.610	Biomass
tpcBuOH					
PYK	ATP	1	7.633	7.633	Production
ATPS4r	ATP	1	14.377	14.377	Production
PGK	ATP	-1	-18.082	18.082	Production
CAR	ATP	-1	8.464	-8.464	Target
PFK	ATP	-1	8.235	-8.235	Maintenance
ATPM	ATP	-1	7.60	-7.60	Waste
ADK1	ATP	-1	8.464	-8.464	Waste
PPCK	ATP	-1	2.205	-2.205	Waste
Biomass	ATP	-55.703	0.092	-5.123	Biomass
GAPD	NADH	1	18.082	18.082	Production
G6PDH2r	NADPH	1	5.021	5.021	Production

Reaction ID	Co-factor	Stoich. coefficient	Flux Distribution	Balance Value	Balance Category
tpcBuOH					
CAR	NADPH	-1	8.464	-8.464	Target
BUT2	NADH	-1	8.464	-8.464	Target
BUT4	NADH	-1	8.464	-8.464	Target
BUT6	NADH	-1	8.464	-8.464	Target
GND	NADPH	1	5.021	5.021	Waste
PDH	NADH	1	17.373	17.373	Waste
ICDHyr	NADH	1	0.099	0.099	Waste
NADH11	NADH	-1	10.388	-10.388	Maintenance
Biomass	NADH/NADPH	3.457/ -18.225	0.092	-1.350	Biomass
BuOH-2					
PYK	ATP	1	8.462	8.462	Production
ATPS4r	ATP	1	13.846	13.846	Production
PGK	ATP	-1	-18.462	18.462	Production
CAR	ATP	-1	9.231	-9.231	Target
ACCOAC	ATP	-1	9.231	-9.231	Target
PFK	ATP	-1	8.462	-8.462	Maintenance
ATPM	ATP	-1	4.615	-4.615	Waste
ADK1	ATP	-1	9.231	-9.231	Waste
GAPD	NADH	1	18.462	18.462	Production
G6PDH2r	NADPH	1	4.615	4.615	Production
CAR	NADPH	-1	9.231	-9.049	Target
BUT2	NADH	-1	9.231	-9.231	Target
BUT4	NADH	-1	9.231	-9.231	Target
BUT6	NADH	-1	9.231	-9.231	Target
NADH11	NADH	-1	9.231	-9.231	Maintenance
GND	NADPH	1	4.615	4.615	Waste
PDH	NADH	1	18.462	18.462	Waste
fasBuOH					
PYK	ATP	1	6.228	6.228	Production
ATPS4r	ATP	1	23.219	23.219	Production
PGK	ATP	-1	-16.799	16.799	Production
CAR	ATP	-1	7.665	-7.665	Target
ACCOAC	ATP	-1	7.665	-7.665	Target
PFK	ATP	-1	6.994	-6.994	Maintenance
ADK1	ATP	-1	7.665	-7.665	Waste
ATPM	ATP	-1	7.60	-7.60	Waste
PPCK	ATP	-1	2.133	-2.133	Waste
Biomass	ATP	-55.703	0.117	-6.523	Biomass
G6PDH2r	NADPH	1	8.669	8.669	Production
GAPD	NADH	1	16.799	16.799	Production
CAR	NADPH	-1	7.665	-7.665	Target
BUT6	NADH	-1	7.665	-7.665	Target
30AR40	NADPH	-1	7.665	-7.665	Maintenance
EAR40x	NADH	-1	7.665	-7.665	Maintenance
NADH11	NADH	-1	17.779	-17.779	Maintenance
GND	NADPH	1	8.669	8.669	Waste
PDH	NADH	1	15.896	15.896	Waste
ICDHyr	NADH	1	0.126	0.126	Waste
Biomass	NADH/NADPH	3.457/ -18.225	0.117	-1.719	Biomass
CROT					
PYK	ATP	1	8.410	8.410	Production
PGK	ATP	-1	19.156	19.156	Production
PFK	ATP	-1	9.410	-9.410	Maintenance
ATPM	ATP	-1	7.60	-7.60	Waste
PPCK	ATP	-1	2.030	-2.030	Waste
Biomass	ATP	-55.703	0.153	-8.525	Biomass
GAPD	NADH	1	19.156	19.156	Production
G6PDH2r	NADPH	1	1.312	1.312	Production

Reaction ID	Co-factor	Stoich. coefficient	Flux Distribution	Balance Value	Balance Category
CROT					
BUT2	NADH	-1	8.618	-8.618	Target
NADH11	NADH	-1	11.080	-11.080	Maintenance
ICDHyr	NADH	1	0.165	0.165	Waste
GND	NADPH	1	1.312	1.312	Waste
Biomass	NADH/NADPH	3.457/ -18.225	0.153	-2.246	Biomass
BUTYR					
PYK	ATP	1	8.409	8.409	Production
PGK	ATP	-1	-19.156	19.156	Production
PFK	ATP	-1	9.410	-9.410	Maintenance
ATPM	ATP	-1	7.60	-7.60	Waste
PPCK	ATP	-1	2.030	-2.030	Waste
Biomass	ATP	-55.703	0.153	-8.525	Biomass
GAPD	NADH	1	19.156	19.156	Production
G6PDH2r	NADPH	1	1.312	1.312	Production
BUT2	NADH	-1	8.618	-8.618	Target
BUT4	NADH	-1	8.618	-8.618	Target
NADH11	NADH	-1	4.411	-4.411	Maintenance
ICDHyr	NADH	1	0.165	0.165	Waste
GND	NADPH	1	1.312	1.312	Waste
PDH	NADH	1	1.949	1.949	Waste
Biomass	NADH/NADPH	3.457/ -18.225	0.153	-2.246	Biomass
BUTAL					
PYK	ATP	1	8.409	8.409	Production
PGK	ATP	-1	-19.156	19.156	Production
ATPS4r	ATP	1	-2.5	-2.5	Maintenance
PFK	ATP	-1	9.410	-9.410	Maintenance
ATPM	ATP	-1	7.60	-7.60	Waste
PPCK	ATP	-1	2.030	-2.030	Waste
Biomass	ATP	-55.703	0.153	-8.525	Biomass
GAPD	NADH	1	19.156	19.156	Production
G6PDH2r	NADPH	1	1.312	1.312	Production
NADH11	NADH	-1	2.688	-2.688	Maintenance
PDH	NADH	1	8.844	8.844	Waste
ICDHyr	NADH	1	0.165	0.165	Waste
GND	NADPH	1	1.312	1.312	Waste
BUT2	NADH	-1	8.618	-8.618	Target
BUT4	NADH	-1	8.618	-8.618	Target
BUT5	NADH	-1	8.618	-8.618	Target
Biomass	NADH/NADPH	3.457/ -18.225	0.153	-2.246	Biomass

Table 13: CBA parameters and outputs of manually curated butanol and butanol precursor models under anaerobic conditions. Reaction IDs, relevant co-factors and their stoichiometric coefficients, flux values, balance values and assigned balance category are included.

Reaction ID	Co-factor	Stoich. coefficient	Flux Distribution	Balance Value	Balance Category
BuOH-0					
PFK	ATP	-1	9.821	-9.821	Maintenance
PGK	ATP	-1	-19.555	19.555	Production
PYK	ATP	1	6.764	6.764	Production
ATPM	ATP	-1	7.60	-7.60	Waste
Biomass	ATP	-55.703	0.160	-8.899	Biomass
GAPD	NADH	1	19.555	19.555	Production
G6PDH2r	NADPH	1	0.062	0.062	Production
THD2	NADPH	1	0.605	0.605	Production
BUT2	NADH	-1	8.776	-8.776	Target
BUT4	NADH	-1	8.776	-8.776	Target
BUT5	NADH	-1	8.776	-8.776	Target
BUT6	NADH	-1	8.776	-8.776	Target
MDH	NADH	1	-2.011	-2.011	Maintenance
NADH11	NADH	-1	0.725	-0.725	Maintenance
THD2	NADH	-1	0.605	-0.605	Maintenance
PDH	NADH	1	18.323	18.323	Waste
ICDHyr	NADH	1	0.172	0.172	Waste
ME2	NADH	1	2.011	2.011	Waste
GND	NADH	1	0.062	0.062	Waste
Biomass	NADH/NADPH	3.547/-18.225	0.160	-2.345	Biomass
BuOH-1					
PFK	ATP	-1	9.924	-9.924	Maintenance
PGK	ATP	-1	-19.855	19.855	Production
PYK	ATP	1	9.652	9.652	Production
ATPM	ATP	-1	7.60	-7.60	Waste
ACCOAC	ATP	-1	9.667	-9.667	Target
Biomass	ATP	-55.703	0.045	-2.316	Biomass
GAPD	NADH	1	19.855	19.855	Production
G6PDH2r	NADPH	1	0.103	0.103	Production
THD2	NADPH	1	0.506	0.506	Production
BUT2	NADH	-1	9.667	-9.667	Target
BUT4	NADH	-1	9.667	-9.667	Target
BUT5	NADH	-1	9.667	-9.667	Target
BUT6	NADH	-1	9.667	-9.667	Target
NADH11	NADH	-1	0.363	-0.363	Maintenance
THD2	NADH	-1	0.506	-0.506	Maintenance
ICDHyr	NADH	1	0.049	0.049	Waste
GND	NADPH	1	0.103	0.103	Waste
PDH	NADH	1	19.534	19.534	Waste
Biomass	NADH/NADPH	3.457/ -18.225	0.045	-0.610	Biomass
tpcBuOH					
PYK	ATP	1	7.633	7.633	Production
ATPS4r	ATP	1	14.377	14.377	Production
PGK	ATP	-1	-18.082	18.082	Production
CAR	ATP	-1	8.464	-8.464	Target
PFK	ATP	-1	8.235	-8.235	Maintenance
ATPM	ATP	-1	7.60	-7.60	Waste
ADK1	ATP	-1	8.464	-8.464	Waste

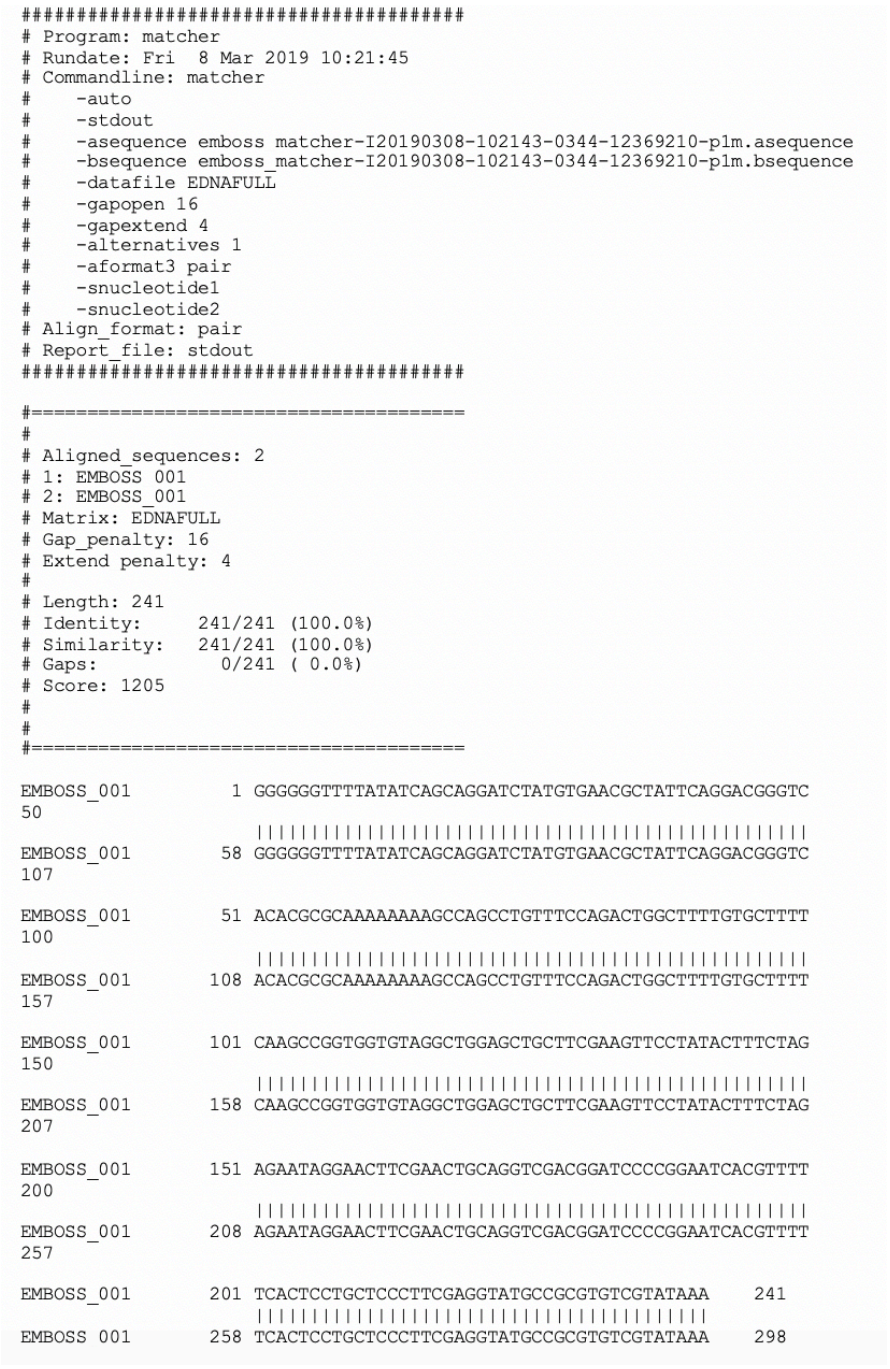
Reaction ID	Co-factor	Stoich. coefficient	Flux Distribution	Balance Value	Balance Category
tpcBuOH					
PPCK	ATP	-1	2.205	-2.205	Waste
Biomass	ATP	-55.703	0.092	-5.123	Biomass
GAPD	NADH	1	18.082	18.082	Production
G6PDH2r	NADPH	1	5.021	5.021	Production
CAR	NADPH	-1	8.464	-8.464	Target
BUT2	NADH	-1	8.464	-8.464	Target
BUT4	NADH	-1	8.464	-8.464	Target
BUT6	NADH	-1	8.464	-8.464	Target
GND	NADPH	1	5.021	5.021	Waste
PDH	NADH	1	17.373	17.373	Waste
ICDHyr	NADH	1	0.099	0.099	Waste
NADH11	NADH	-1	10.388	-10.388	Maintenance
Biomass	NADH/NADPH	3.457/ -18.225	0.092	-1.350	Biomass
BuOH-2					
PYK	ATP	1	8.462	8.462	Production
ATPS4r	ATP	1	13.846	13.846	Production
PGK	ATP	-1	-18.462	18.462	Production
CAR	ATP	-1	9.231	-9.231	Target
ACCOAC	ATP	-1	9.231	-9.231	Target
PFK	ATP	-1	8.462	-8.462	Maintenance
ATPM	ATP	-1	4.615	-4.615	Waste
ADK1	ATP	-1	9.231	-9.231	Waste
GAPD	NADH	1	18.462	18.462	Production
G6PDH2r	NADPH	1	4.615	4.615	Production
CAR	NADPH	-1	9.231	-9.049	Target
BUT2	NADH	-1	9.231	-9.231	Target
BUT4	NADH	-1	9.231	-9.231	Target
BUT6	NADH	-1	9.231	-9.231	Target
NADH11	NADH	-1	9.231	-9.231	Maintenance
GND	NADPH	1	4.615	4.615	Waste
PDH	NADH	1	18.462	18.462	Waste
fasBuOH					
PYK	ATP	1	6.228	6.228	Production
ATPS4r	ATP	1	23.219	23.219	Production
PGK	ATP	-1	-16.799	16.799	Production
CAR	ATP	-1	7.665	-7.665	Target
ACCOAC	ATP	-1	7.665	-7.665	Target
PFK	ATP	-1	6.994	-6.994	Maintenance
ADK1	ATP	-1	7.665	-7.665	Waste
ATPM	ATP	-1	7.60	-7.60	Waste
PPCK	ATP	-1	2.133	-2.133	Waste
Biomass	ATP	-55.703	0.117	-6.523	Biomass
G6PDH2r	NADPH	1	8.669	8.669	Production
GAPD	NADH	1	16.799	16.799	Production
CAR	NADPH	-1	7.665	-7.665	Target
BUT6	NADH	-1	7.665	-7.665	Target
30AR40	NADPH	-1	7.665	-7.665	Maintenance
EAR40x	NADH	-1	7.665	-7.665	Maintenance
NADH11	NADH	-1	17.779	-17.779	Maintenance
GND	NADPH	1	8.669	8.669	Waste
PDH	NADH	1	15.896	15.896	Waste
ICDHyr	NADH	1	0.126	0.126	Waste
Biomass	NADH/NADPH	3.457/ -18.225	0.117	-1.719	Biomass

Reaction ID	Co-factor	Stoich. coefficient	Flux Distribution	Balance Value	Balance Category
CROT					
PYK	ATP	1	8.410	8.410	Production
PGK	ATP	-1	19.156	19.156	Production
PFK	ATP	-1	9.410	-9.410	Maintenance
ATPM	ATP	-1	7.60	-7.60	Waste
PPCK	ATP	-1	2.030	-2.030	Waste
Biomass	ATP	-55.703	0.153	-8.525	Biomass
GAPD	NADH	1	19.156	19.156	Production
G6PDH2r	NADPH	1	1.312	1.312	Production
BUT2	NADH	-1	8.618	-8.618	Target
NADH11	NADH	-1	11.080	-11.080	Maintenance
ICDHyr	NADH	1	0.165	0.165	Waste
GND	NADPH	1	1.312	1.312	Waste
Biomass	NADH/NADPH	3.457/ -18.225	0.153	-2.246	Biomass
BUTYR					
PYK	ATP	1	8.409	8.409	Production
PGK	ATP	-1	-19.156	19.156	Production
PFK	ATP	-1	9.410	-9.410	Maintenance
ATPM	ATP	-1	7.60	-7.60	Waste
PPCK	ATP	-1	2.030	-2.030	Waste
Biomass	ATP	-55.703	0.153	-8.525	Biomass
GAPD	NADH	1	19.156	19.156	Production
G6PDH2r	NADPH	1	1.312	1.312	Production
BUT2	NADH	-1	8.618	-8.618	Target
BUT4	NADH	-1	8.618	-8.618	Target
NADH11	NADH	-1	4.411	-4.411	Maintenance
ICDHyr	NADH	1	0.165	0.165	Waste
GND	NADPH	1	1.312	1.312	Waste
PDH	NADH	1	1.949	1.949	Waste
Biomass	NADH/NADPH	3.457/ -18.225	0.153	-2.246	Biomass
BUTAL					
PYK	ATP	1	8.409	8.409	Production
PGK	ATP	-1	-19.156	19.156	Production
ATPS4r	ATP	1	-2.5	-2.5	Maintenance
PFK	ATP	-1	9.410	-9.410	Maintenance
ATPM	ATP	-1	7.60	-7.60	Waste
PPCK	ATP	-1	2.030	-2.030	Waste
Biomass	ATP	-55.703	0.153	-8.525	Biomass
GAPD	NADH	1	19.156	19.156	Production
G6PDH2r	NADPH	1	1.312	1.312	Production
NADH11	NADH	-1	2.688	-2.688	Maintenance
PDH	NADH	1	8.844	8.844	Waste
ICDHyr	NADH	1	0.165	0.165	Waste
GND	NADPH	1	1.312	1.312	Waste
BUT2	NADH	-1	8.618	-8.618	Target
BUT4	NADH	-1	8.618	-8.618	Target
BUT5	NADH	-1	8.618	-8.618	Target
Biomass	NADH/NADPH	3.457/ -18.225	0.153	-2.246	Biomass

Table 14: Growth rates and generation times of the wild type, PCK and ATPs mutants.

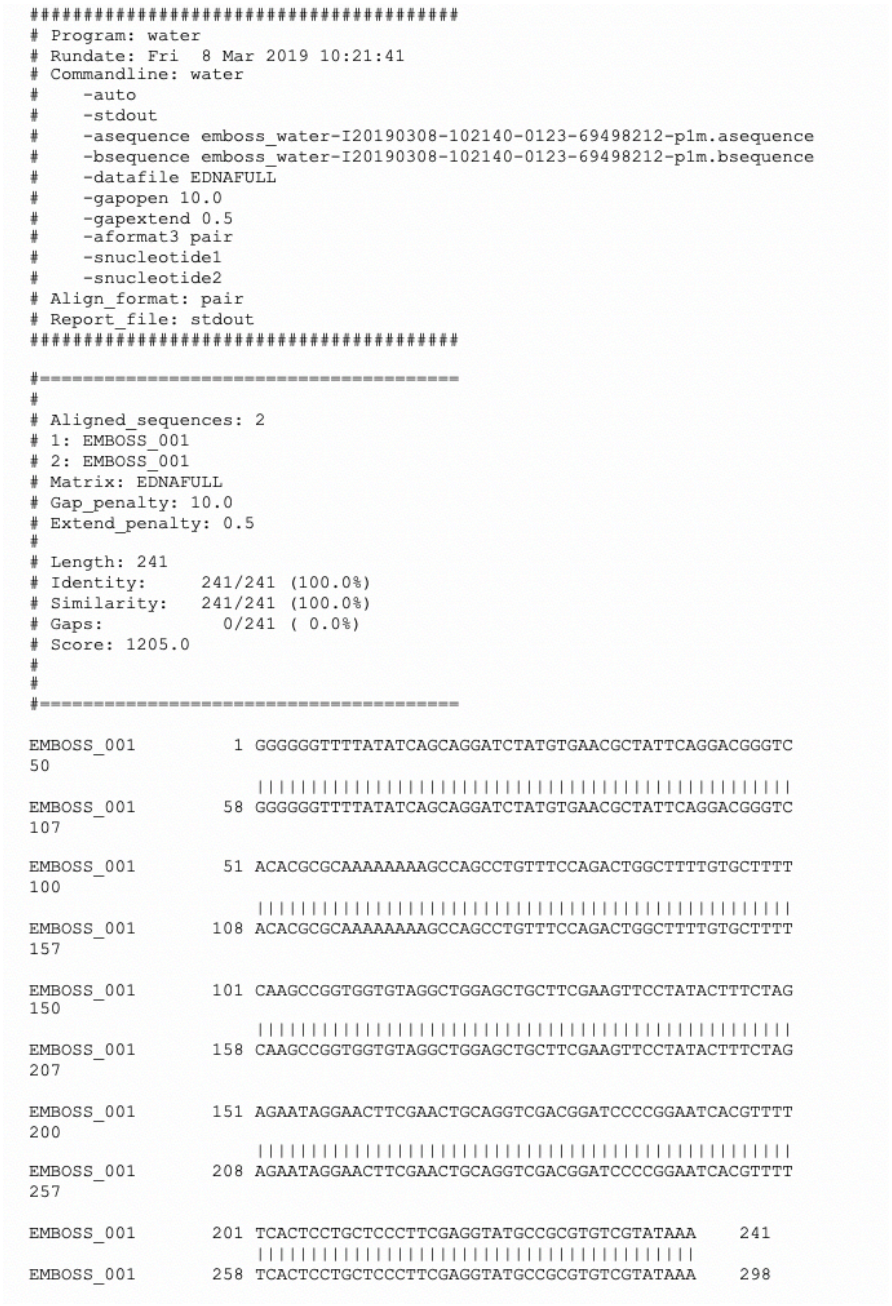
	WT	PCK	ATPs
growth rate, μ (as per Equation 3.6)	0.377	0.376	0.333
generation time, g (as per Equation 3.7)	1.839	1.841	2.079
growth rate, μ (as per Equation 3.8)	0.377	0.376	0.333

Figure 1: Snapgene view of ATPs sequencing results. Orange: ATPs complementary region; pink: 20-bp pKD13 priming region; Green: flipase specific FRT regions



```
########################################
# Program: matcher
# Rundate: Fri  8 Mar 2019 10:21:45
# Commandline: matcher
#    -auto
#    -stdout
#    -asequence emboss matcher-I20190308-102143-0344-12369210-p1m.asequence
#    -bsequence emboss_matcher-I20190308-102143-0344-12369210-p1m.bsequence
#    -datafile EDNAFULL
#    -gapopen 16
#    -gapextend 4
#    -alternatives 1
#    -aformat3 pair
#    -snucleotide1
#    -snucleotide2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS 001
# 2: EMBOSS_001
# Matrix: EDNAFULL
# Gap_penalty: 16
# Extend penalty: 4
#
# Length: 241
# Identity:     241/241 (100.0%)
# Similarity:   241/241 (100.0%)
# Gaps:           0/241 ( 0.0%)
# Score: 1205
#
#
#=======================================

EMBOSS_001         1 GGGGGGTTTTATATCAGCAGGATCTATGTGAACGCTATTCAGGACGGGTC
50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        58 GGGGGGTTTTATATCAGCAGGATCTATGTGAACGCTATTCAGGACGGGTC
107

EMBOSS_001        51 ACACGCGCAAAAAAAAGCCAGCCTGTTTCCAGACTGGCTTTTGTGCTTTT
100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       108 ACACGCGCAAAAAAAAGCCAGCCTGTTTCCAGACTGGCTTTTGTGCTTTT
157

EMBOSS_001       101 CAAGCCGGTGGTGTAGGCTGGAGCTGCTTCGAAGTTCCTATACTTTCTAG
150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       158 CAAGCCGGTGGTGTAGGCTGGAGCTGCTTCGAAGTTCCTATACTTTCTAG
207

EMBOSS_001       151 AGAATAGGAACTTCGAACTGCAGGTCGACGGATCCCCGGAATCACGTTTT
200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       208 AGAATAGGAACTTCGAACTGCAGGTCGACGGATCCCCGGAATCACGTTTT
257

EMBOSS_001       201 TCACTCCTGCTCCCTTCGAGGTATGCCGCGTGTCGTATAAA    241
                     |||||||||||||||||||||||||||||||||||||||||
EMBOSS 001       258 TCACTCCTGCTCCCTTCGAGGTATGCCGCGTGTCGTATAAA    298
```

Figure 2: ATPs sequencing: Local alignment between forward and reverse-complement sequencing results using the Matcher software, which identifies local similarities between two sequences using a rigorous algorithm based on the LALIGN application.



```
########################################
# Program: water
# Rundate: Fri  8 Mar 2019 10:21:41
# Commandline: water
#    -auto
#    -stdout
#    -asequence emboss_water-I20190308-102140-0123-69498212-p1m.asequence
#    -bsequence emboss_water-I20190308-102140-0123-69498212-p1m.bsequence
#    -datafile EDNAFULL
#    -gapopen 10.0
#    -gapextend 0.5
#    -aformat3 pair
#    -snucleotide1
#    -snucleotide2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 241
# Identity:     241/241 (100.0%)
# Similarity:   241/241 (100.0%)
# Gaps:           0/241 ( 0.0%)
# Score: 1205.0
#
#
#=======================================

EMBOSS_001         1 GGGGGGTTTTATATCAGCAGGATCTATGTGAACGCTATTCAGGACGGGTC
50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        58 GGGGGGTTTTATATCAGCAGGATCTATGTGAACGCTATTCAGGACGGGTC
107

EMBOSS_001        51 ACACGCGCAAAAAAAAGCCAGCCTGTTTCCAGACTGGCTTTTGTGCTTTT
100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       108 ACACGCGCAAAAAAAAGCCAGCCTGTTTCCAGACTGGCTTTTGTGCTTTT
157

EMBOSS_001       101 CAAGCCGGTGGTGTAGGCTGGAGCTGCTTCGAAGTTCCTATACTTTCTAG
150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       158 CAAGCCGGTGGTGTAGGCTGGAGCTGCTTCGAAGTTCCTATACTTTCTAG
207

EMBOSS_001       151 AGAATAGGAACTTCGAACTGCAGGTCGACGGATCCCCGGAATCACGTTTT
200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       208 AGAATAGGAACTTCGAACTGCAGGTCGACGGATCCCCGGAATCACGTTTT
257

EMBOSS_001       201 TCACTCCTGCTCCCTTCGAGGTATGCCGCGTGTCGTATAAA    241
                     |||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       258 TCACTCCTGCTCCCTTCGAGGTATGCCGCGTGTCGTATAAA    298
```

Figure 3: ATPs sequencing: Local alignment between forward and reverse-complement sequencing results using the Water software, which uses the Smith-Waterman algorithm (modified for speed enhancements) to calculate the local alignment of two sequences.

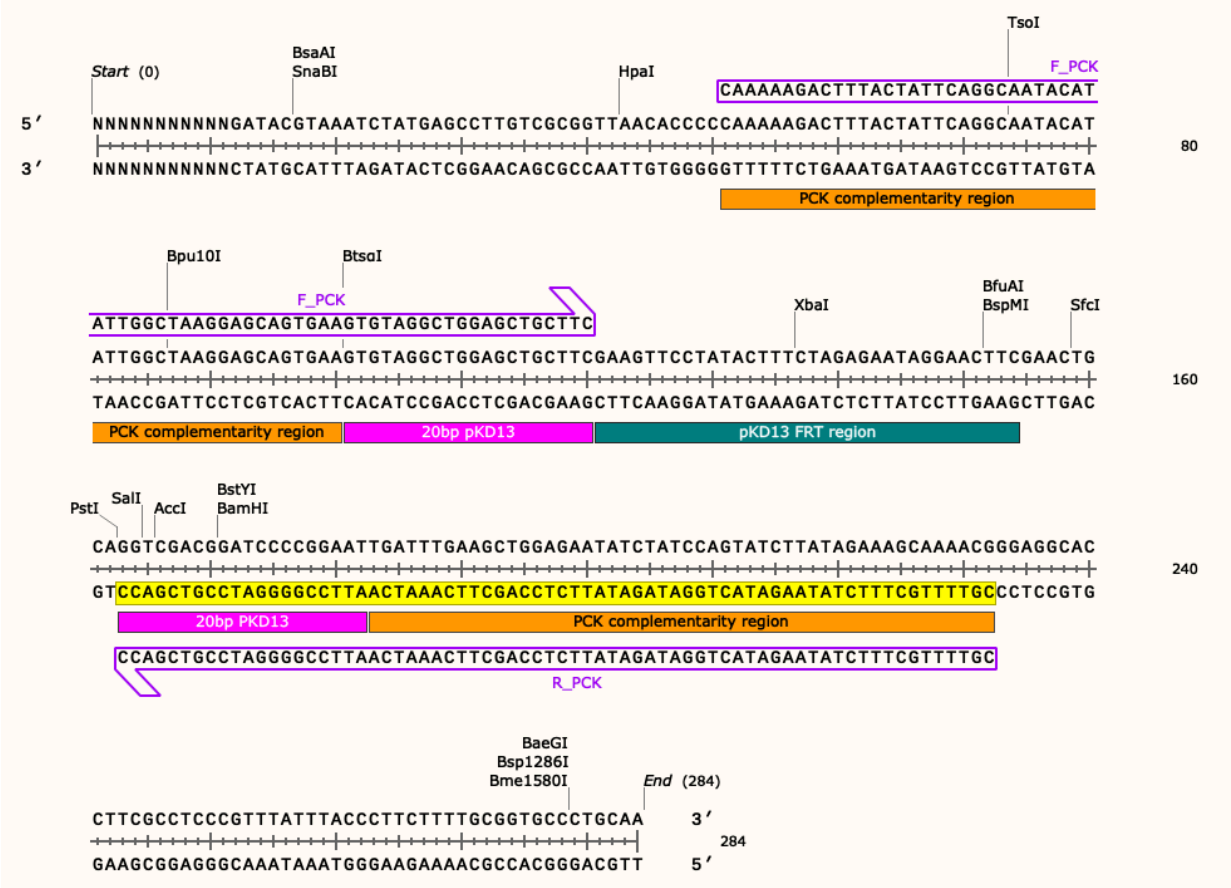
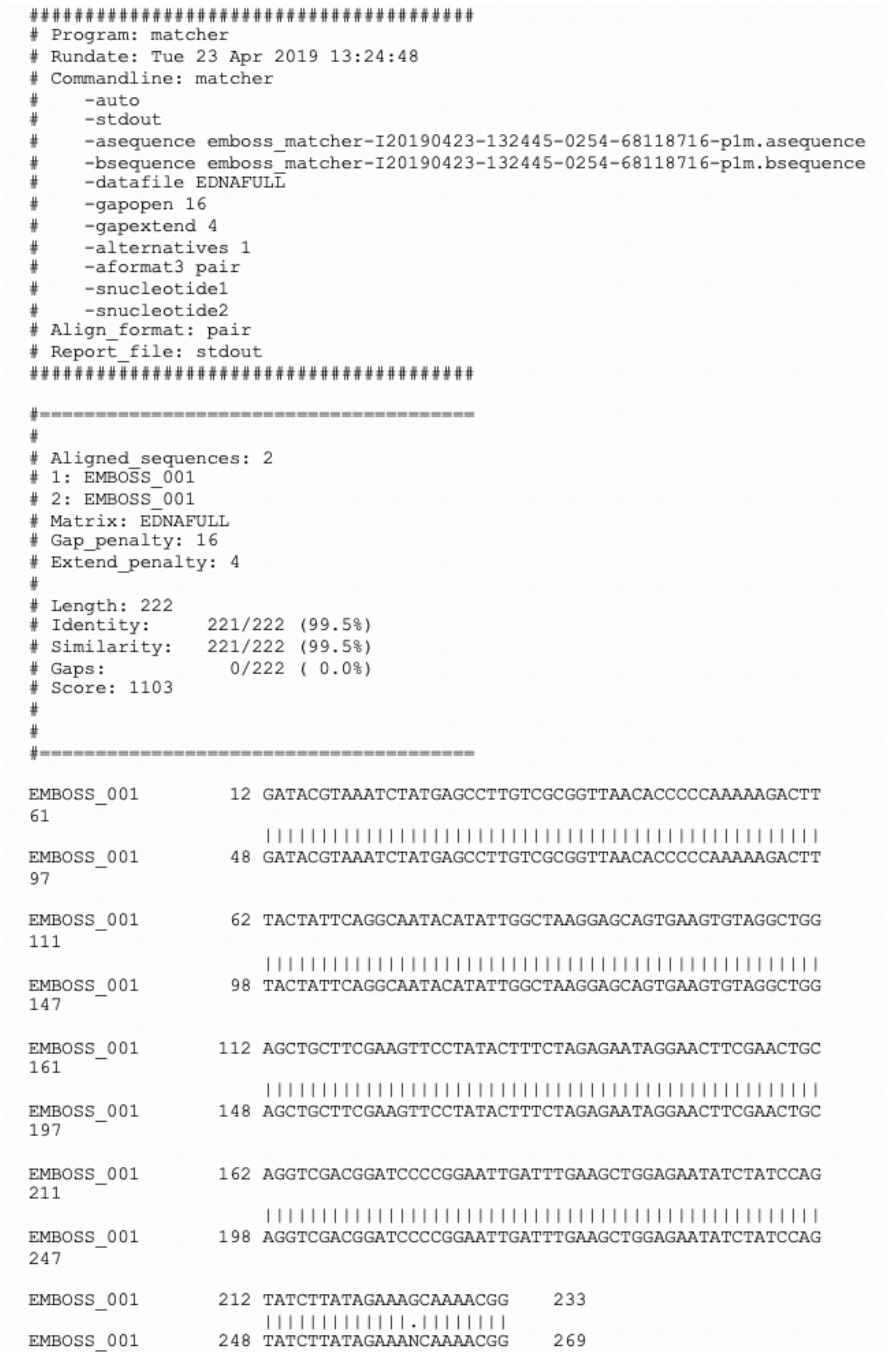
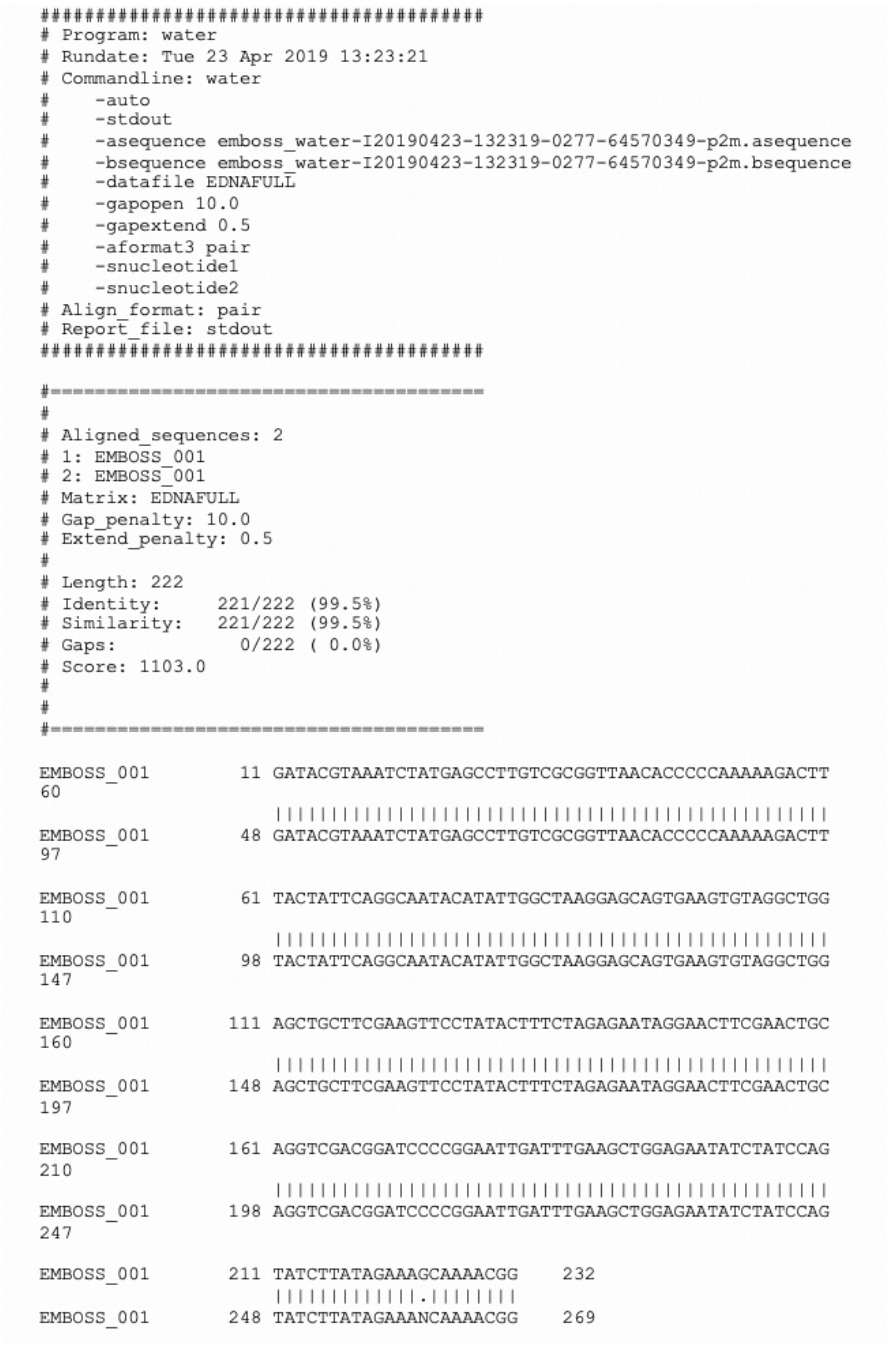


Figure 4: Snapgene view of PCK sequencing results. Orange: PCK complementary region; pink: 20-bp pKD13 priming region; Green: flipase specific FRT regions



```
########################################
# Program: matcher
# Rundate: Tue 23 Apr 2019 13:24:48
# Commandline: matcher
#    -auto
#    -stdout
#    -asequence emboss_matcher-I20190423-132445-0254-68118716-p1m.asequence
#    -bsequence emboss_matcher-I20190423-132445-0254-68118716-p1m.bsequence
#    -datafile EDNAFULL
#    -gapopen 16
#    -gapextend 4
#    -alternatives 1
#    -aformat3 pair
#    -snucleotide1
#    -snucleotide2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EDNAFULL
# Gap_penalty: 16
# Extend_penalty: 4
#
# Length: 222
# Identity:     221/222 (99.5%)
# Similarity:   221/222 (99.5%)
# Gaps:           0/222 ( 0.0%)
# Score: 1103
#
#
#=======================================

EMBOSS_001        12 GATACGTAAATCTATGAGCCTTGTCGCGGTTAACACCCCCAAAAAGACTT
61
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        48 GATACGTAAATCTATGAGCCTTGTCGCGGTTAACACCCCCAAAAAGACTT
97

EMBOSS_001        62 TACTATTCAGGCAATACATATTGGCTAAGGAGCAGTGAAGTGTAGGCTGG
111
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        98 TACTATTCAGGCAATACATATTGGCTAAGGAGCAGTGAAGTGTAGGCTGG
147

EMBOSS_001       112 AGCTGCTTCGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGAACTGC
161
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       148 AGCTGCTTCGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGAACTGC
197

EMBOSS_001       162 AGGTCGACGGATCCCCGGAATTGATTTGAAGCTGGAGAATATCTATCCAG
211
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       198 AGGTCGACGGATCCCCGGAATTGATTTGAAGCTGGAGAATATCTATCCAG
247

EMBOSS_001       212 TATCTTATAGAAAGCAAAACGG    233
                     |||||||||||||.||||||||
EMBOSS_001       248 TATCTTATAGAAANCAAAACGG    269
```

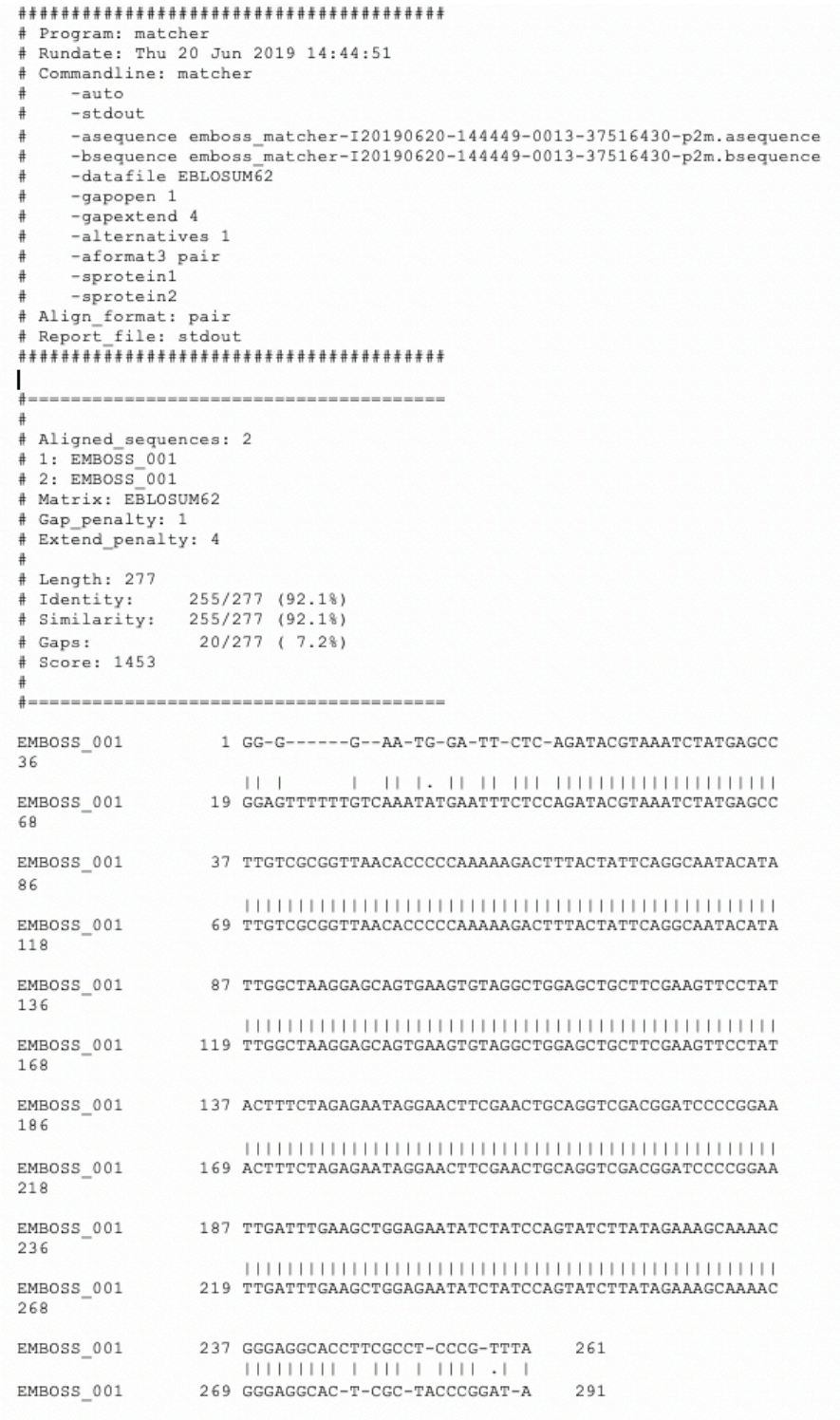
Figure 5: PCK sequencing: Local alignment between forward and reverse-complement sequencing results using the Matcher software.



```
########################################
# Program: water
# Rundate: Tue 23 Apr 2019 13:23:21
# Commandline: water
#    -auto
#    -stdout
#    -asequence emboss_water-I20190423-132319-0277-64570349-p2m.asequence
#    -bsequence emboss_water-I20190423-132319-0277-64570349-p2m.bsequence
#    -datafile EDNAFULL
#    -gapopen 10.0
#    -gapextend 0.5
#    -aformat3 pair
#    -snucleotide1
#    -snucleotide2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 222
# Identity:     221/222 (99.5%)
# Similarity:   221/222 (99.5%)
# Gaps:           0/222 ( 0.0%)
# Score: 1103.0
#
#
#=======================================

EMBOSS_001        11 GATACGTAAATCTATGAGCCTTGTCGCGGTTAACACCCCCAAAAAGACTT
60
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        48 GATACGTAAATCTATGAGCCTTGTCGCGGTTAACACCCCCAAAAAGACTT
97

EMBOSS_001        61 TACTATTCAGGCAATACATATTGGCTAAGGAGCAGTGAAGTGTAGGCTGG
110
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        98 TACTATTCAGGCAATACATATTGGCTAAGGAGCAGTGAAGTGTAGGCTGG
147

EMBOSS_001       111 AGCTGCTTCGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGAACTGC
160
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       148 AGCTGCTTCGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGAACTGC
197

EMBOSS_001       161 AGGTCGACGGATCCCCGGAATTGATTTGAAGCTGGAGAATATCTATCCAG
210
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       198 AGGTCGACGGATCCCCGGAATTGATTTGAAGCTGGAGAATATCTATCCAG
247

EMBOSS_001       211 TATCTTATAGAAAGCAAAACGG    232
                     |||||||||||||.||||||||
EMBOSS_001       248 TATCTTATAGAAANCAAAACGG    269
```

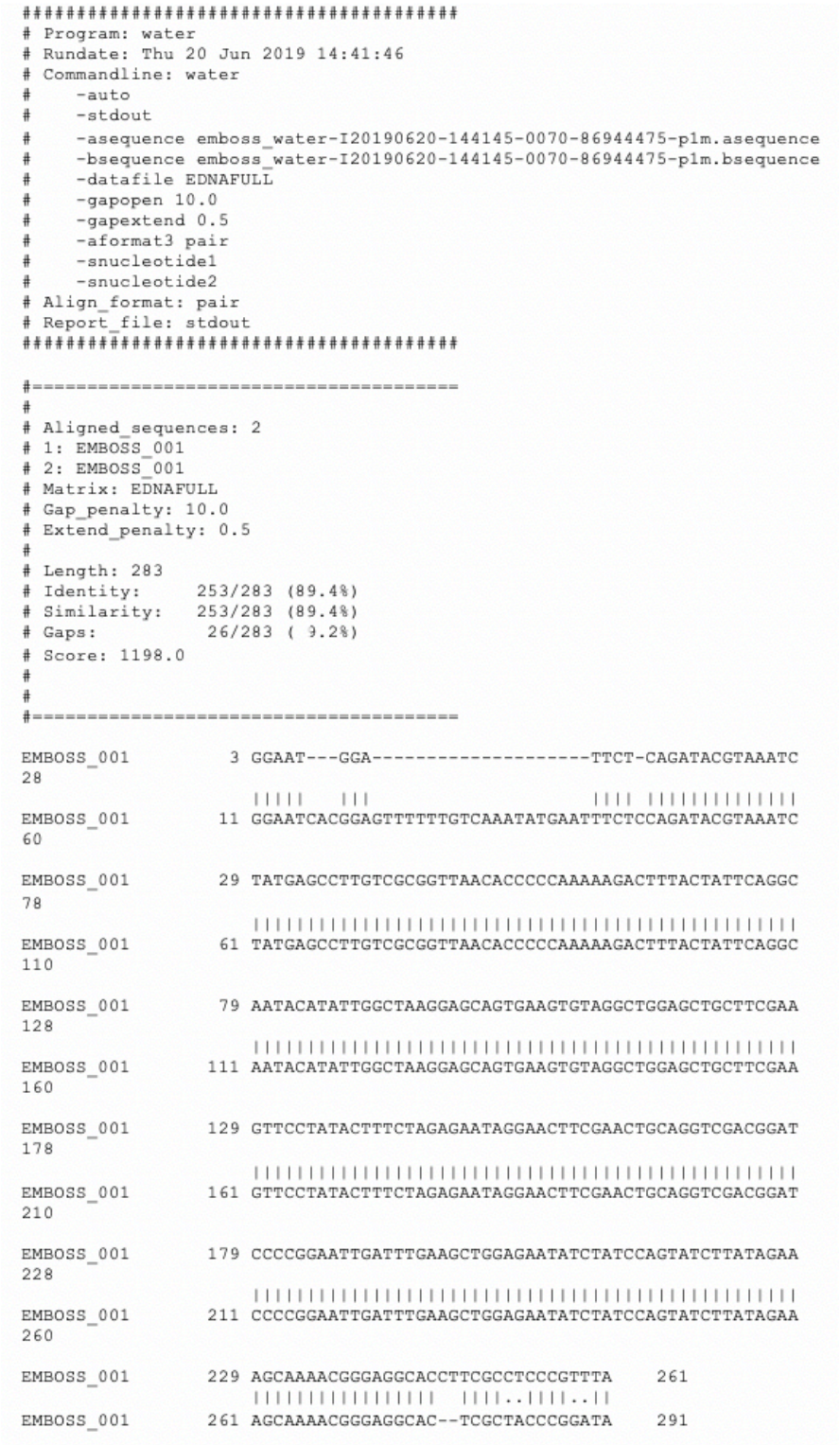
Figure 6: PCK sequencing: Local alignment between forward and reverse-complement sequencing results using the Water software.

Figure 7: Snapgene view of ATPs-PCK double mutant sequencing results. Orange: PCK complementary region; pink: 20-bp pKD13 priming region; Green: flipase specific FRT regions.



```
########################################
# Program: matcher
# Rundate: Thu 20 Jun 2019 14:44:51
# Commandline: matcher
#    -auto
#    -stdout
#    -asequence emboss_matcher-I20190620-144449-0013-37516430-p2m.asequence
#    -bsequence emboss_matcher-I20190620-144449-0013-37516430-p2m.bsequence
#    -datafile EBLOSUM62
#    -gapopen 1
#    -gapextend 4
#    -alternatives 1
#    -aformat3 pair
#    -sprotein1
#    -sprotein2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 1
# Extend_penalty: 4
#
# Length: 277
# Identity:     255/277 (92.1%)
# Similarity:   255/277 (92.1%)
# Gaps:          20/277 ( 7.2%)
# Score: 1453
#
#=======================================

EMBOSS_001         1 GG-G------G--AA-TG-GA-TT-CTC-AGATACGTAAATCTATGAGCC
36
                     || |      |  || |. || || ||| ||||||||||||||||||||
EMBOSS_001        19 GGAGTTTTTTGTCAAATATGAATTTCTCCAGATACGTAAATCTATGAGCC
68

EMBOSS_001        37 TTGTCGCGGTTAACACCCCCAAAAAGACTTTACTATTCAGGCAATACATA
86
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        69 TTGTCGCGGTTAACACCCCCAAAAAGACTTTACTATTCAGGCAATACATA
118

EMBOSS_001        87 TTGGCTAAGGAGCAGTGAAGTGTAGGCTGGAGCTGCTTCGAAGTTCCTAT
136
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       119 TTGGCTAAGGAGCAGTGAAGTGTAGGCTGGAGCTGCTTCGAAGTTCCTAT
168

EMBOSS_001       137 ACTTTCTAGAGAATAGGAACTTCGAACTGCAGGTCGACGGATCCCCGGAA
186
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       169 ACTTTCTAGAGAATAGGAACTTCGAACTGCAGGTCGACGGATCCCCGGAA
218

EMBOSS_001       187 TTGATTTGAAGCTGGAGAATATCTATCCAGTATCTTATAGAAAGCAAAAC
236
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       219 TTGATTTGAAGCTGGAGAATATCTATCCAGTATCTTATAGAAAGCAAAAC
268

EMBOSS_001       237 GGGAGGCACCTTCGCCT-CCCG-TTTA    261
                     ||||||||| | ||| | |||| .| |
EMBOSS_001       269 GGGAGGCAC-T-CGC-TACCCGGAT-A    291
```

Figure 8: ATPs-PCK double mutant sequencing: Local alignment between forward and reverse-complement sequencing results using the Matcher software.



```
########################################
# Program: water
# Rundate: Thu 20 Jun 2019 14:41:46
# Commandline: water
#    -auto
#    -stdout
#    -asequence emboss_water-I20190620-144145-0070-86944475-p1m.asequence
#    -bsequence emboss_water-I20190620-144145-0070-86944475-p1m.bsequence
#    -datafile EDNAFULL
#    -gapopen 10.0
#    -gapextend 0.5
#    -aformat3 pair
#    -snucleotide1
#    -snucleotide2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 283
# Identity:     253/283 (89.4%)
# Similarity:   253/283 (89.4%)
# Gaps:          26/283 ( 9.2%)
# Score: 1198.0
#
#
#=======================================

EMBOSS_001         3 GGAAT---GGA--------------------TTCT-CAGATACGTAAATC
28
                     |||||   |||                    |||| ||||||||||||||
EMBOSS_001        11 GGAATCACGGAGTTTTTTGTCAAATATGAATTTCTCCAGATACGTAAATC
60

EMBOSS_001        29 TATGAGCCTTGTCGCGGTTAACACCCCCAAAAAGACTTTACTATTCAGGC
78
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        61 TATGAGCCTTGTCGCGGTTAACACCCCCAAAAAGACTTTACTATTCAGGC
110

EMBOSS_001        79 AATACATATTGGCTAAGGAGCAGTGAAGTGTAGGCTGGAGCTGCTTCGAA
128
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       111 AATACATATTGGCTAAGGAGCAGTGAAGTGTAGGCTGGAGCTGCTTCGAA
160

EMBOSS_001       129 GTTCCTATACTTTCTAGAGAATAGGAACTTCGAACTGCAGGTCGACGGAT
178
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       161 GTTCCTATACTTTCTAGAGAATAGGAACTTCGAACTGCAGGTCGACGGAT
210

EMBOSS_001       179 CCCCGGAATTGATTTGAAGCTGGAGAATATCTATCCAGTATCTTATAGAA
228
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       211 CCCCGGAATTGATTTGAAGCTGGAGAATATCTATCCAGTATCTTATAGAA
260

EMBOSS_001       229 AGCAAAACGGGAGGCACCTTCGCCTCCCGTTTA    261
                     |||||||||||||||||  ||||..||||..||
EMBOSS_001       261 AGCAAAACGGGAGGCAC--TCGCTACCCGGATA    291
```

Figure 9: ATPs-PCK double mutant sequencing: Local alignment between forward and reverse-complement sequencing results using the Water software.

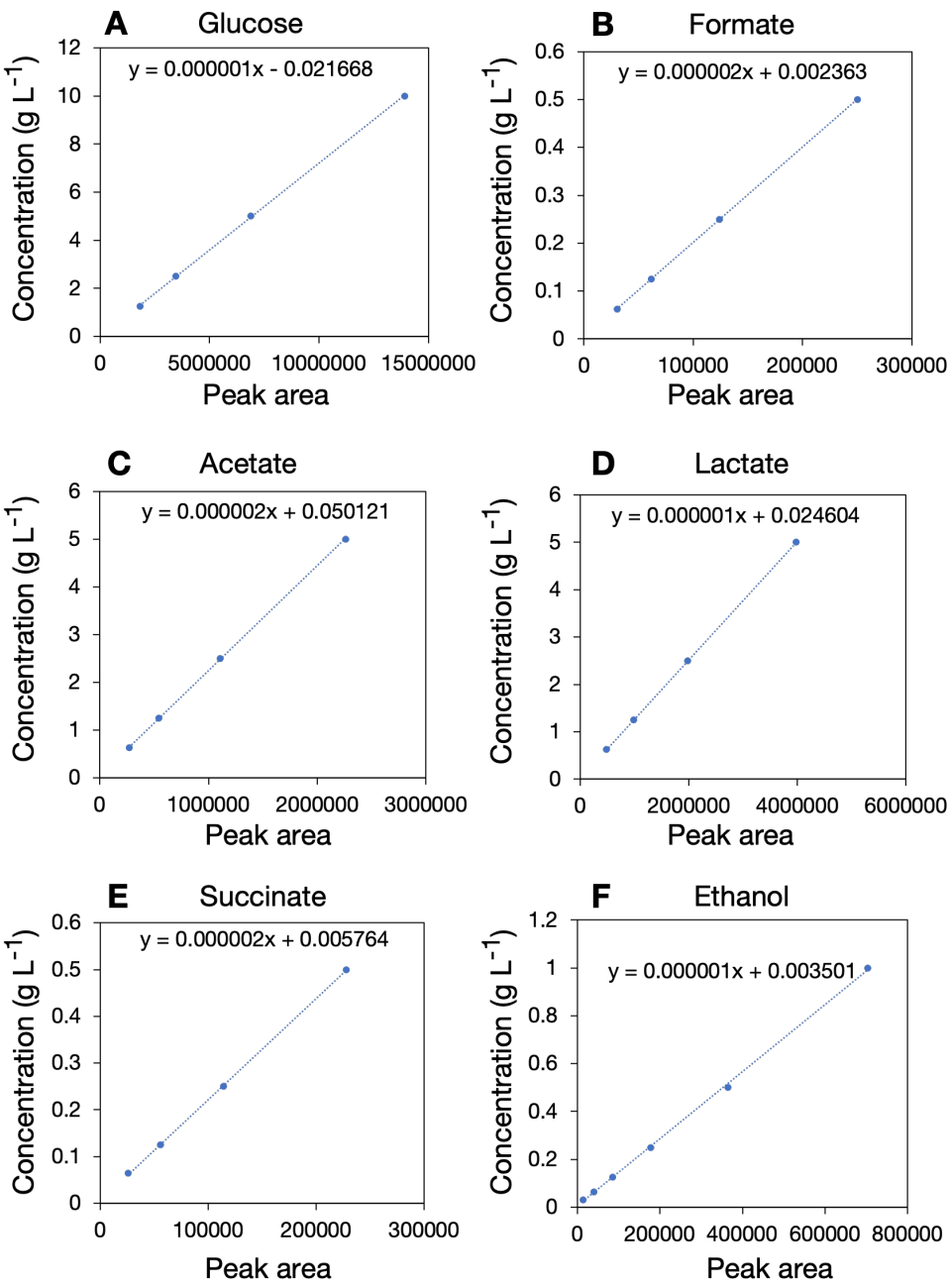


Figure 10: Glucose, acetate, succinate, lactate formate and ethanol standard curves used to quantify these compounds by HPLC

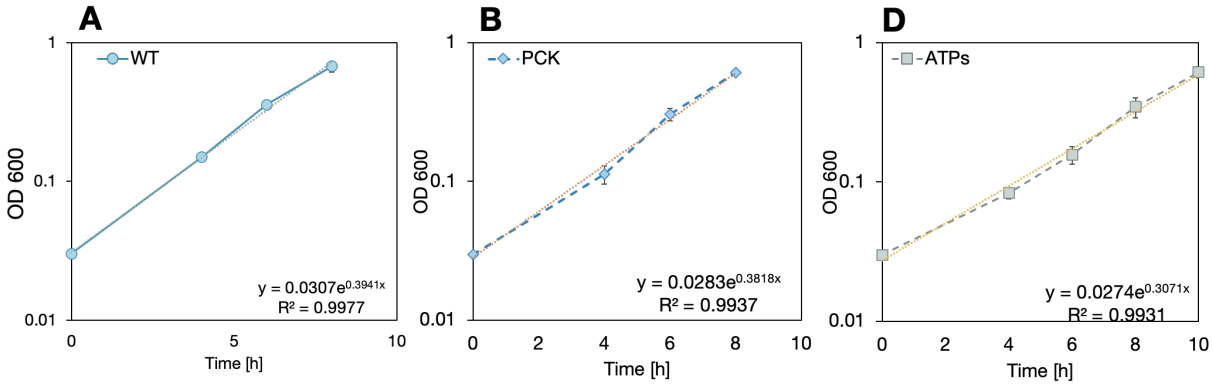


Figure 11: OD_{600} against time (h) in logarithmic scale. Timepoints selected to calculate growth rates are: (A) wild type: 4h and 8h; (B) PCK: 0h and 8h; and (C): ATPs: 4h and 10h.

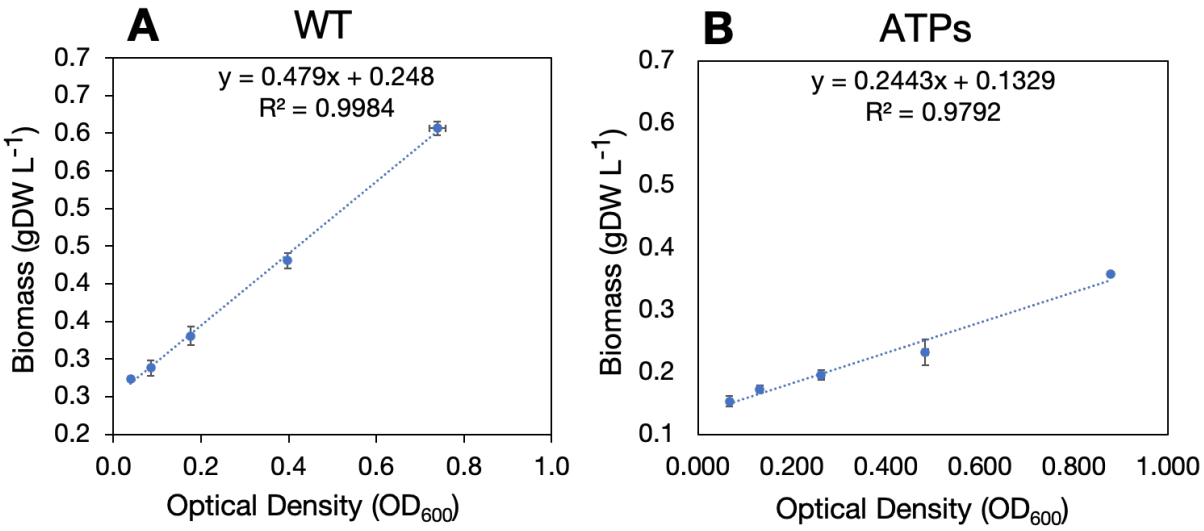


Figure 12: Calibration curves to convert optical density (OD_{600}) to cell dry weight (gDW L^{-1}) for strains WT and ATPs.

Table 15: Reactions included in the *E.coli* model available on WUFlux.

Glycolysis
Glucose(substrate) + ATP == G6P
G6P == F6P
F6P + ATP == FBP
FBP == F6P
FBP == DHAP + GAP
DHAP == GAP
GAP == G3P + ATP + NADH
G3P == PEP
PEP == PYR + ATP
PYR + 2*ATP == PEP

Citric Acid Cycle
PYR == AceCoA + CO2 + NADH
AceCoA + OAA == CIT
CIT == ICIT
ICIT == AKG + CO2 + NADPH
AKG == SucCoA + CO2 + NADH
SucCoA == SUC + ATP
SUC == FUM + FADH2
FUM == MAL
MAL == OAA + NADH

Glyoxylate Shunt
ICIT == GLX + SUC
GLX + AceCoA == MAL

Amphibolic reactions
MAL == PYR + CO2 + NADH
MAL == PYR + CO2 + NADPH
PEP + CO2 == OAA
OAA + ATP == PEP + CO2
PYR + ATP + CO2 == OAA

Pentose Phosphate Pathway
G6P == PG6 + NADPH
PG6 == CO2 + Ru5P + NADPH
Ru5P == X5P
Ru5P == R5P
X5P + R5P == GAP + S7P
GAP + S7P == E4P + F6P
X5P + E4P == GAP + F6P
PG6 == PYR + GAP

Waste release
AceCoA == Ac + ATP

Amino Acid Biosynthesis
AKG + NADPH == GLU
GLU + ATP == GLN
GLU + ATP + 2*NADPH == PRO
GLU + GLN + CO2 + ASP + AceCoA + 5*ATP + NADPH == ARG + AKG + FUM + Ac
OAA + GLU == ASP + AKG
ASP + 2*ATP == ASN
PYR + GLU == ALA + AKG
G3P + GLU == SER + AKG + NADH
SER == GLY + Methylene_THF
GLY == Methylene_THF + CO2 + NADH
Methylene_THF + NADH == Methyl_THF
Methylene_THF == Formyl_THF + NADPH
ASP + 2*ATP + 2*NADPH == THR
THR == GLY + AceCoA + NADH
SER + AceCoA + 3*ATP + 4*NADPH == CYS + Ac
ASP + PYR + GLU + SucCoA + ATP + 2*NADPH == LYS + CO2 + AKG + SUC

Amino Acid Biosynthesis
ASP + Methyl_THF + CYS + SucCoA + ATP + 2*NADPH == MET + PYR + SUC
GLU + NADPH + 2*PYR == VAL + AKG + CO2
AceCoA + 2*PYR + GLU + NADPH == LEU + AKG + NADH + 2*CO2
THR + PYR + GLU + NADPH == ILE + AKG + CO2
E4P + 2*PEP + GLU + ATP + NADPH == PHE + AKG + CO2
E4P + 2*PEP + GLU + ATP + NADPH == TYR + AKG + NADH + CO2
SER + R5P + 2*PEP + E4P + GLN + 3*ATP + NADPH == TRP + GAP + PYR + GLU + CO2
R5P + Formyl_THF + GLN + ASP + 5*ATP == HIS + AKG + FUM + 2*NADH

Transhydrogenation
NADH == NADPH

Oxidative Phosphorylation
NADH == 3*ATP
FADH2 == 2*ATP

ATP Hydrolysis
ATP == ATP_maintenance

Transport
Ac == Acetate_ex
CO2 == CO2_ex
CO2_air + CO2 == CO2 + CO2_ex

Fatty Acid Synthesis
AceCoA + 1.75*NADPH + 0.875*ATP == fatty_acid_C2

Biomass Synthesis
0.488*ALA + (...) == 39.68*Biomass + (...)

Additional reactions
AceCoA + 2*NADH == etoh_ex
PYR + NADH == lactate_ex

Table 16: Absolute fluxes (mmol gDW^{-1} hr^{-1}) for the WT, ATPs and PCK strains.

	WT		PCK		ATPs	
reactions	best fit	stdev	best fit	stdev	best fit	stdev
Glucose(substrate) + ATP == G6P	6.777	0.377	10.204	0.000	19.568	0.001
G6P == F6P	4.928	0.435	7.329	0.380	17.077	0.018
F6P + ATP == FBP	13.915	10.252	8.898	0.127	18.522	0.045
FBP == F6P	8.118	10.269				
FBP == DHAP + GAP	5.797	0.393	8.898	0.127	18.305	0.041
DHAP == GAP	5.797	0.393	8.898	0.127	22.819	0.091
GAP == G3P + ATP + NADH	11.941	0.769	18.498	0.127	38.404	0.135
G3P == PEP	11.257	0.780	17.869	0.134	32.754	0.125
PEP == PYR + ATP	10.000	2.247	14.500	0.861	32.681	3.315
PYR + 2*ATP == PEP	0.000	2.022	0.000	0.005	11.264	2.086
PYR == AceCoA + CO2 + NADH	9.836	1.043	15.825	0.281	29.528	0.924
AceCoA + OAA == CIT	1.572	0.633	1.148	0.217	4.145	0.020
CIT == ICIT	1.572	0.633	1.148	0.217	-0.864	0.078
ICIT == AKG + CO2 + NADPH	0.427	0.724	0.404	0.000	-2.537	0.543
AKG == SucCoA + CO2 + NADH	0.003	0.714	0.000	0.000	1.956	0.488
SucCoA == SUC + ATP	-0.183	0.714	-0.177	0.000	-5.805	0.524
SUC == FUM + FADH2	1.148	0.636	0.693	0.217	-3.945	0.042
FUM == MAL	1.294	0.636	0.832	0.217	1.645	0.002
MAL == OAA + NADH	1.467	0.589	-1.075	0.615	1.015	1.889
MAL == PYR + CO2 + NADH	0.573	0.325	0.353	0.645	1.152	1.195
MAL == PYR + CO2 + NADPH	0.400	0.337	2.298	0.687	5.681	1.308
PEP + CO2 == OAA	0.953	0.598	3.079	0.786	12.564	1.902
OAA + ATP == PEP + CO2	0.000	0.383	0.000	0.000	1.969	2.171
ICIT == GLX + SUC	1.146	0.426	0.744	0.217	4.306	0.522
GLX + AceCoA == MAL	1.146	0.426	0.744	0.217	3.714	0.523
G6P == PG6 + NADPH	1.769	0.291	2.798	0.380	1.459	0.006
PG6 == CO2 + Ru5P + NADPH	1.769	0.291	2.798	0.380	0.000	0.000
Ru5P == X5P	0.897	0.196	1.596	0.254	-1.969	0.019
Ru5P == R5P	0.872	0.097	1.202	0.127	1.223	0.038
X5P + R5P == GAP + S7P	0.519	0.098	0.866	0.127	1.921	0.030
GAP + S7P == E4P + F6P	0.519	0.098	0.866	0.127	-3.624	0.035
X5P + E4P == GAP + F6P	0.378	0.099	0.730	0.127	-3.611	0.009
AceCoA == Ac + ATP	2.732	0.592	1.562	0.434	10.099	0.021
DHAP + NADPH == PYR + 2*NADH	0.000	0.000	0.000	0.000	0.000	0.000
DHAP + NADH == PYR + 2*NADH	0.000	0.000			0.000	0.000
NADH == NADPH	6.829	2.651	18.0572	1.8088	87.321	1.350
NADH == 3*ATP	17.306	3.690	16.713	1.678	0.000	0.000
FADH2 == 2*ATP	1.148	0.636	0.6932	0.2166	0.000	0.000
ATP == ATP_maintenance	40.359	17.782	42.0882	5.923	7.375	2.292
Ac == Acetate_ex	2.934	0.592	1.755	0.4344	10.969	0.000
AceCoA + 1.75*NADPH + 0.875*ATP == fatty_acid_C2	2.606	1.230	11.3758	0.5314	23.371	0.839
CO2 == CO2_ex	12.836	1.871	19.3681	0.5751	56.058	0.980
CO2_air + CO2 == CO2 + CO2_ex	0.000	0.000	0	0	0.000	0.000
0.488*ALA + (...) ==39.68*Biomass+ (...)	0.393	0.009	0.375	0	0.348	0.000
PYR == Fornate_ex + AceCoA			0.253	0.1111		
SUC == Succinate_ex			0.0509	0.0099		
AceCoA + 2*NADH == etoh_ex	0.424	0.104			1.164	0.001
PYR + NADH == lactate_ex	0.102	0.085	0.0851	0.0829		