Metabolic engineering of *E. coli* for growth-coupled alkane biosynthesis



Jochem R. Nielsen

Worcester College

University of Oxford

A thesis submitted in partial fulfilment for the degree

Doctor of Philosophy

Trinity Term 2024

Abstract

The production of sustainable fuels represents an essential technology enabling the transition away from fossil fuel use. Over the past two decades several biological routes for the production of chemically identical molecules to those in fossil fuels ('drop-in' fuels) have been discovered and characterised. These have held promise as breakthrough technologies but have yet to reach commercial viability. Industrial scale bioprocess development for drop-in fuel production has been held back by high feedstock prices and inadequate process efficiencies.

Currently, bioethanol and biodiesel are produced using crop-derived sugars and oils, so called '1st generation' biofuels. Although their use positively impacts net CO_2 emissions, their derivatization from food crops means they compete for arable land. To avoid this 'food vs. fuel' issue, alternative carbon sources for bioproduction must be explored. In this thesis the use of two non-sugar substrates, acetate and ethanol, were explored as carbon sources for hydrocarbon biosynthesis. Either carbon source enabled the production of these compounds at titres and product/substrate yields similar to that of glucose.

The rate and efficiency of substrate conversion to product largely governs the viability of a bioprocess. These properties are in part defined by the enzymes that make up the relevant metabolic pathways. One of these pathways is the cyanobacterial alkane biosynthesis pathway discovered in 2010, consisting of acyl-ACP reductase (AAR) and aldehyde deformylating oxygenase (ADO). The terminal ADO enzyme is known for its low catalytic efficiency, and as a result many enzyme engineering efforts have targeted this enzyme.

This thesis describes the design of growth-coupled approaches for engineering ADO. Growth-coupling involves linking enzyme activity to cell fitness and taking advantage of the principles of directed evolution to select for improved enzyme variants. Several growth-coupling approaches were assessed in this work through a combination of *in silico* flux balance analysis and experimental work, mostly focussing on the byproduct of ADO – formate. These efforts resulted in functional growth-coupling of ADO activity via the reductive glycine pathway.

Another approach to growth-couple ADO activity was based on redox cofactor auxotrophy. An *E. coli* strain lacking all central metabolic pathways for NAD^+ reduction was constructed, dubbed NADHaux, and its growth could be tightly linked to NADH generation by formate oxidation. Taken together, this thesis provides insights into the extremes of *E. coli* redox metabolism and presents hopeful results regarding the use of non-sugar carbon sources for bioproduction. In addition, it describes valuable novel methodologies for the evolutionary engineering of alkane biosynthesis.

Acknowledgements

The completion of this work marks the end of a chapter. I count myself profoundly lucky to have shared my time here with such a supportive group of friends and family. This short acknowledgement section is beyond insufficient to express my gratitude to you all.

I would firstly like to thank my supervisor Prof. **Wei Huang** for the opportunity to complete my DPhil in his lab, and for his enthusiasm, optimism, and support over the years.

The time spent at my lab bench was always made better by my wonderful colleagues at Begbroke. I'd like to thank **Simona** for the fantastic memories made sharing in our conference adventure, and for your help, uplifting positive attitude, and friendship. To **Yutong**, whom I pity deeply for being seated next to me, thank you for your endless patience while listening to my (scientific) ramblings and for your advice and friendship over the years. To **Laura**, **Joseph**, **Yun, Ben, Morgan, Weiming**, **Nick**, **Reza**, **Nichola** thank you all for the many lunches, formals, drinks, long (work) chats, driving me to Begbroke and back, many coffees, and the supportive professional environment you all helped create.

My time in Oxford would've been wasted if not for the fantastic friends I've been lucky to make, including those outside of work. I can't stress enough how **Jacques, Jack, Jai, Jeremie, Nicolas** and **Andrew** have positively shaped my time here. Thank you for the unforgettable times had during our years of sharing houses including during COVID lockdown, and for our evenings at the pub, long work chats, dangerous doses of self-deprecating humour and for the constant encouragement you all offered. I could not have wished for a more genuine group of friends. Of course, I cannot imagine my time in Oxford without **Anya, Lisa, Ricky, George, Amy,** and **Phoebe**: thank you for your friendship, listening ear, and for the memorable moments shared together. To **Thijs** and **Julia**, thank you for discovering Oxford with me venue-by-venue and for your continuous encouragement and kindness, you were my Dutch pied-à-terre in Oxford. **Andrew** and **Lauren**, I appreciate you both for staying in touch ever

since our first weeks at Worcester College, I remember our MCR times fondly. Thank you for encouraging and supporting me over the years.

I’m extremely lucky to have had many good friends from across the Channel who continued to positively impact my life even from abroad. Firstly, I would like to thank **Aucke, Emiel, Mark, Xavier, Hugo, Maarten,** and **Richard** for their continued friendship and support, with special appreciation for **Willem, Bart** and **Bas** who went out of their way to visit me. Our memorable trip to Brazil gave me the motivation I needed to finalise this work. I also want to recognise more good friends from my time in Leiden, namely **Daniel, Noud, Bart** and **Maurits**. Thank you all for your support and for coming over to Oxford to visit me. To **Danice, Arne, Anouschka** and **Jordy,** thank you for the great days and evenings we spent together both in Oxford and across the Channel, you made my trips home especially enjoyable. Finally, I’m deeply grateful to **Alison** for her selfless encouragement and support over the years.

To my brother **Hans**, **Mercedes,** and little **Ines**, thank you for always filling my visits home with joyful highlights. I’m looking forward to carrying out my duties as uncle (reading and Duplo) more frequently in the future!

Finally, I am endlessly grateful to my parents, **Fenna** and **Frants**, for their unwavering encouragement and constant presence. During the inevitable low points encountered during this degree, and all those before it, you have been a beaming source of warmth, humour, and support. Consider this work yours.

Declaration of Authorship

I hereby declare that, unless otherwise stated, the contents of this thesis is entirely my own work. Where applicable, the rightful authors involved in any collaborative work are duly acknowledged. I confirm that no part of this work has been submitted as part of any other degree at this or any other University.



Jochem R. Nielsen

26th September 2024

Table of Contents

Abstract **2**
Acknowledgements **3**
Declaration of Authorship **5**
List of Figures **10**
List of Tables **12**
Abbreviations **13**
1. General introduction **14**
1.1 A warming world 14
1.2 The global energy diet and losing our fossil appetite 15
1.3 Thesis scope and objectives 16
1.4 Thesis layout 16
1.5 Key research output 18
2. Sustainable hydrocarbons **20**
2.1 Chemical hydrocarbon synthesis 21
2.2 Biofuels: generations and limitations 22
1st generation biofuels 22
2nd generation (advanced) biofuels 23
3rd and 4th generation biofuels 24
2.3 Enzymes and pathways for hydrocarbon biosynthesis 25
2.4 Development of *Escherichia coli* as a cell factory for alkane biosynthesis 28
2.4.1 Microbial chassis for alkane biosynthesis 28
2.4.2 Metabolic engineering strategies to improve hydrocarbon biosynthesis in *E. coli* 31
2.4.2.1 Push strategies 32
2.4.2.2 Pull strategies 33
2.4.2.3 Block strategies 34
2.5 Substrate selection for alkane biosynthesis 36
2.5.1 Glycolytic carbon sources 36
2.5.2 C_2 carbon sources 37
2.5.3 C_1 carbon sources 38
2.6 Enzyme engineering to enhance alkane biosynthesis 39
2.7 Harnessing evolution for engineering microbial hydrocarbon biosynthesis 44
2.7.1 Opportunities for the directed evolution of hydrocarbon-producing enzymes 44
2.7.2 Growth-coupled enzyme engineering through manipulation of redox cofactor regeneration 71
3. Construction and characterization of MoClo-compatible vectors for modular exploration of the protein expression space in *E. coli* **111**

3.1 Abstract 111
3.2 Introduction 111
3.3 Materials and methods 114
3.3.1 Strains, plasmids, media, and growth conditions 114
3.3.2 Assembly of the medium copy MoClo destination vectors 114
3.3.3 Routine Golden Gate assemblies 115
3.3.4 Gene deletion of *adhE* in *E. coli* BW25113 116
3.3.5 Domestication of *adhE* and randomised *adhE** construct assembly 117
3.3.6 Selection of *adhE** expression constructs 117
3.3.7 Microtiter plate assays 118
3.4 Results 119
3.4.1 Expression level and burden assessment of the vector kit 119
3.4.2 Growth-coupled optimisation of *adhE* expression by randomised GGA 124
3.5 Discussion 126
3.6 Conclusion 128
3.7 References 129
3.8 Supplementary data 132
4. Hydrocarbon biosynthesis in *E. coli* using non-sugar carbon sources 138
4.1 Abstract 138
4.2 Introduction 138
4.3 Materials and method 141
4.3.1 Strains, plasmids, and routine growth conditions 141
4.3.2 Cloning procedures 143
4.3.3 Gene deletions 145
4.3.4 Microtiter plate growth assays 145
4.3.5 Hydrocarbon production via AAR and AAR-ADO 146
4.3.6 Gas chromatography (GC-FID) 147
4.3.7 High performance liquid chromatography (HPLC) 147
4.4 Results 148
4.4.1 Optimization of acyl-ACP reductase (AAR) activity with glycolytic carbon sources 148
4.4.2 Alkane biosynthesis with glucose as a carbon source 155
4.4.3 Characterization of *E. coli* growth using acetate and formate 157
4.4.4 Oxygen tolerance of *adhE* and characterization of *E. coli* growth using ethanol 159
4.4.5 Hydrocarbon biosynthesis using acetate and formate or ethanol 163
4.4.6 Alkane biosynthesis using ethanol as a carbon source 169
4.4.7 Growth-decoupling using C_2 carbon sources by regulatory control 171
4.5 Discussion and conclusion 176

4.6 References 181
4.7 Supplementary data 184
5. Novel metabolic strategies for growth-coupling alkane biosynthesis in *E. coli* 187
5.1 Statement of collaborative contributions 187
5.2 Abstract 187
5.3 Introduction 188
5.4 Materials and methods 192
5.4.1 Strains, plasmids, media and routine growth 192
5.4.2 Constructing Δpgi ΔsthA 193
5.4.3 Growth-rescue of ΔPS 194
5.4.4 Growth characterization and rescue of C_1 auxotrophs 194
5.4.5 Flux balance analysis 195
5.4.6 Microtiter plate growth assays 195
5.5 Results 196
5.5.1 $NADP^+$ auxotrophy for growth-coupled alkane biosynthesis 196
5.5.2 Growth-coupled alkane biosynthesis via NADH auxotrophy 204
5.5.3 Growth-coupled alkane biosynthesis via the reductive glycine pathway 211
5.6 Discussion 219
5.7 Conclusion 223
5.8 References 224
5.9 Supplementary Data 227
6. A sensitive NADH-auxotrophic *E. coli* for use as an enzyme engineering platform 228
6.1 Abstract 228
6.2 Introduction 228
6.3 Methods and materials 231
6.3.1 Strains, plasmids, media and routine growth 231
6.3.2 Flux balance analysis 233
6.3.3 Plasmid construction 234
6.3.4 Microtiter plate growth assays 234
6.3.5 Gene deletions 235
6.3.6 Formate dehydrogenase expression strength selection 236
6.3.7 Whole genome sequencing 237
6.4 Results 237
6.4.1 Design and metabolic modelling to achieve NADH auxotrophy 237
6.4.2 Confirming Δ*lpd* phenotype and rescuing its growth 242
6.4.3 Deletion of central metabolic NADH-regenerating enzymes 244
6.4.4 Characterization of mutants carrying an $NADP^+$-dependent GAPDH 247

6.4.5 Growth of NADHaux using a glycolytic NADH source reveals $NADP^+$ auxotrophy 252
6.4.6 Growth rescue of NADHaux via beta oxidation 255
6.4.7 Growth rescue of NADHaux by formate oxidation 256
6.4.8 An NADHaux 'cheater' carries mutations in NADH:ubiquinone oxidoreductase 259
6.5 Discussion 261
6.6 Conclusion 265
6.7 References 265
6.8 Supplementary data 269
7. Construction of an acetyl-CoA auxotroph reveals unexpected flexibility in aerobic metabolism connecting glycolysis to the tricarboxylic acid cycle 272
7.1 Abstract 272
7.2 Introduction 272
7.3 Materials and methods 274
7.3.1 Strains, plasmids, media, and growth conditions 274
7.3.2 Gene deletions 275
7.3.3 Microtiter plate growth assays 276
7.4 Results 277
7.4.1 Growth of *E. coli* BW25113 without pyruvate dehydrogenase- and oxidase activity 277
7.4.2 Pyruvate-formate lyase and pyruvate-flavodoxin oxidoreductase contribute to aerobic growth 280
7.4.3 The activity of 2-amino-3-ketobutyrate ligase contributes to aerobic growth 282
7.4.4 Other metabolic routes towards AcCoA 284
7.5 Discussion & conclusion 285
7.6 References 289
7.7 Supplementary data 291
8. General discussion 292
9. Concluding remarks 298
10. References 299

List of Figures

Chapter 1

Figure 1.1. Schematic overview of topics covered in this thesis……..**19**

Chapter 2

Figure 2.1. Enzymatic pathways for bacterial hydrocarbon synthesis……..…….. **26**
Figure 2.2. Screening methods for hydrocarbon biosynthesis enzymes……..…….. **54**
Figure 2.3. Selection methods for hydrocarbon-producing enzymes……..…… **58**
Figure 2.4. Schematic representations of redox cofactor regeneration as a growth-coupling strategy, and several ways this can be employed for enzyme engineering……**79**
Figure 2.5. *E. coli* central metabolic reactions involving redox cofactors…….**86**

Chapter 3

Figure 3.1. Constituent parts of the destination vectors generated in this study……**119**
Figure 3.2. Fluorescence readings of MoClo-constructed GFP constructs……..……..**120**
Figure 3.3. Burden assessment of MoClo-compatible destination vectors with different origins of replication…….…… **123**
Figure 3.4. Growth-coupled selection for optimal *adhE** expression levels using MoClo-mediated randomised construct assembly……..…**125**

Chapter 4

Figure 4.1. Hydrocarbon biosynthesis in *E. coli* BW25113 expressing AAR ……..…...**149**
Figure 4.2. Effects of deleting *tesC* on hydrocarbon biosynthesis via AAR …… **151**
Figure 4.3. Effects of fatty acid degradation regulator (*fadR*) expression on growth and hydrocarbon biosynthesis using acyl-ACP reductase (AAR) in *E. coli ΔtesC*……**153**
Figure 4.4. Alkane biosynthesis in *E. coli* using glucose……...**156**
Figure 4.5. Characterising *E. coli* tolerance and growth using acetate and formate…….**158**
Figure 4.6. Oxygen tolerance, metabolite biosynthesis and expression optimization for AdhE-mediated growth on ethanol…… **160**
Figure 4.7. Hydrocarbon biosynthesis using acetate and formate…….. **164**
Figure 4.8. Hydrocarbon biosynthesis using ethanol as a carbon source……..…**166**
Figure 4.9. Comparison of bioproduction metrics for AAR-mediated hydrocarbon synthesis…….…… **168**
Figure 4.10. Alkane biosynthesis using ethanol as a carbon source…….…… **170**
Figure 4.11. The glyoxylate shunt and the genomic organisation of the *aceBAK* operon **172**
Figure 4.12. Growth inhibition on C_2 carbon sources by regulatory control of the glyoxylate shunt…….…… **174**
Figure S4.1. Tolerance of *E. coli* BW25113 to potassium chloride……..……**186**

Chapter 5

Figure 5.1. Schematic overview of *E. coli* central metabolism and fatty acid biosynthesis, focussing on NADP(H)-utilising enzymes…… **197**
Figure 5.2. Growth rescue of *E. coli* BW25113 *Δpgi ΔsthA* through aldehyde reductase activity…… **199**
Figure 5.3. Growth rescue of *E. coli* BW25113 *Δpgi ΔsthA* through induction of

fatty acid biosynthesis…… **200**
Figure 5.4. Flux balance analysis of ΔPS for growth-coupled alkane biosynthesis……... **202**
Figure 5.5. Dependence of biomass flux on formate availability in NADH-deficient *E. coli* models…….. **205**
Figure 5.6. Biomass- and alkane fluxes of in silico mutants with FDH and ADO…….. **208**
Figure 5.7. Self-sufficient in silico alkane and formate biosynthesis via fatty acid synthesis and AAR/ADO activity……... **210**
Figure 5.8. Schematic of the metabolic architecture of formate dependent $C_1(G)_{aux}$ and $C_1(G)S_{aux}$ strains…….. **212**
Figure 5.9. Growth dependence of $C_1(G)_{aux}$ and $C_1(G)S_{aux}$ on formate …….. **214**
Figure 5.10. Proposed metabolic architecture for growth-coupling (AAR-)ADO activity via the reductive glycine pathway …… **214**
Figure 5.11. Formate-based growth coupling of the AAR-ADO alkane biosynthesis pathway …… **217**
Figure S5.1. Growth response of $C_1(G)S_{aux}$ to supplements…… **227**

Chapter 6

Figure 6.1. Simplified schematic of *E. coli* central metabolism focusing on reactions involving redox cofactors…… **239**
Figure 6.2. Flux balance analysis results of NADH-deficient *E. coli* models…… **241**
Figure 6.3. Characterization of *E. coli Δlpd* (ΔL) and rescuing growth by formate oxidation…… **243**
Figure 6.4. Growth characterization of *E. coli* strains lacking NADH-generating central metabolic enzymes……... **246**
Figure 6.5. Characterization of ΔLS_GN growth……..…**248**
Figure 6.6. Growth characterization of ΔLSPMA_GC…… **251**
Figure 6.7. Growth characterization of NADHaux reveals auxotrophy for $NADP^+$…… **254**
Figure 6.8. Growth rescue of NADHaux via beta oxidation…….. **255**
Figure 6.9. Formate-dependent growth of NADHaux……**258**
Figure 6.10. Growth characterization of *nuoF*-mutant M1…… **260**
Figure S6.1. Ethanol growth-rescue of the *Δlpd* mutant…… **269**

Chapter 7

Figure 7.1. Schematic of *E. coli* central metabolism and pathways leading towards acetyl-CoA formation…….. **278**
Figure 7.2. Growth characterization of Δ*lpd* and Δ*lpd* Δ*poxB* mutants…….. **279**
Figure 7.3. Growth characterization of strains lacking acetyl-CoA generating enzymes.... **281**
Figure 7.4. The effect of a 2-amino-3-ketobutyrate ligase (*kbl*) deletion on growth of ΔLXYP…… **283**

Chapter 8

Figure 8.1. The dynamic ranges of $C1(G)S_{aux}$ and NADHaux ensure a selective pressure for formate biosynthesis over three orders of magnitude…….. **296**

List of Tables

Chapter 2

Table 2.1. Properties of hydrocarbons in the context of directed evolution……... **51**
Table 2.2. Transcription factor-based reporter systems for aliphatic hydrocarbon detection……..…….. **53**
Table 2.3. Comparison of several widely used screening/selection techniques for enzyme engineering……**75**
Table 2.4. Engineered strains displaying cofactor regeneration deficiencies and any enzymes engineered using these platforms…….**100**

Chapter 3

Table 3.1. Strains used in this work……..…… **114**
Table S3.1. Plasmids used and constructed in this study …….. **132**
Table S3.2. DNA sequences of primers used in this study ……..…….. **134**
Table S3.3. Synthesised DNA fragments used in this study……..……. **135**
Table S3.4. Growth parameters of *E. coli* BW25113 transformed with various GFP expression constructs……..…….. **137**

Chapter 4

Table 4.1. Strains used and engineered in this study…… **141**
Table 4.2. Plasmids used and constructed in this work……**142**
Table S4.1. Primers used in this study……..….. **184**
Table S4.2. Synthesised DNA fragments used in this study……..….. **185**

Chapter 5

Table 5.1. Strains used and constructed in this study…….. **192**
Table 5.2. Plasmids used in this work…… **193**
Table S5.1. Primers used in this study…… **227**

Chapter 6

Table 6.1. Strains used and constructed in this work…… **232**
Table 6.2. Plasmids used and constructed in this work…… **232**
Table S6.1. Primers used in this study…… **269**
Table S6.2. Synthesised DNA fragments used in this work…… **270**

Chapter 7

Table 7.1. Plasmids used in this study…… **275**
Table 7.2. Strains used and constructed in this study…….. **275**
Table S7.1. Primers used in this study…… **291**

Abbreviations

αKG	α-ketoglutarate / 2-oxoglutarate
AAR	Acyl-ACP reductase
ADO	Aldehyde deformylating oxygenase
ATP	Adenosine triphosphate
6PG	6-phosphogluconate
ACP	Acyl carrier protein
A.U.	Arbitrary units
C_1	One carbon
C_2	Two carbon
C_3	Three carbon
CoA	Coenzyme A
CDS	Coding DNA sequence
DE	Directed evolution
DNA	Deoxyribonucleic acid
NAD^+	Nicotinamide adenine dinucleotide
$NADP^+$	Nicotinamide adenine dinucleotide phosphate
NCD^+	Nicotinamide cytosine dinucleotide
NMN^+	Nicotinamide mononucleotide
F6P	Fructose-6-phosphate
FAD	Flavin adenine dinucleotide
FAS	Fatty acid synthesis
R5P	Ribulose-5-phosphate
G6P	Glucose-6-phosphate
GA3P	Glyceraldehyde-3-phosphate
GC-FID	Gas chromatography – flame ionisation detector
GFP	Green fluorescent protein
HPLC	High performance liquid chromatography
LB	Luria-Bertani medium
OD	Optical density
Ori	Origin of replication
RBS	Ribosome binding site
WT	Wildtype

1. General introduction

1.1 A warming world

Already since the 1970's the emission of CO_2 and other greenhouse gasses have been linked to increases in global temperatures (Broecker, 1975; Cooper, 1978). Over the following decades, climate scientists have amassed incontrovertible evidence that global warming is occurring. The most recent report published by the Intergovernmental Panel on Climate Change (IPCC) in 2023 has indisputably linked anthropogenic CO_2 emissions to increasing global surface temperature, which has now warmed 1.1^{o}C on average (2011-2020) compared to the pre-industrial period (1850-1900) (IPCC, 2023). In addition to a rise in global surface temperature, ocean acidification and rising ocean levels are extremely likely and virtually certain, respectively, to be caused by human activity and are both poised to worsen over time, according to the latest IPCC report. The major contributor to anthropological climate change is the exploitation of fossil fuels, a collective term for carbon-containing substances including coal, natural gas, and oil, formed by the deposition and decay of organic material over millions of years. Their combustion releases the carbon they contained back into the atmosphere in the form of carbon emissions, mostly as CO_2. The timescales required for fossil fuel formation makes their use inherently unsustainable, since timely replacement of these deposits is not possible.

The findings presented by climate researchers urge us to take action to reduce carbon emissions for the sake of alleviating global warming. Combating climate change will undoubtedly prove to be the single most urgent and impactful global challenge presented to us in the 21st century. The effects of climate change are already being felt, with increases in the frequency of extreme weather events such as heatwaves virtually certainly being caused by human activity (IPCC, 2023). These effects are still modest compared to what is expected if carbon emissions are not tapered soon; global flooding events, climate change-induced

refugee crises, vast losses of biodiversity and the emergence of areas inhospitable for human life belonging to the more severe consequences of unchecked climate change (IPCC, 2023; Mazhin et al., 2020; O'Neill et al., 2017). To ensure the maintenance of biodiversity and habitable living conditions worldwide, cleaner sources of energy must be deployed at a global scale within the next several decades.

1.2 The global energy diet and losing our fossil appetite

The world's heavy dependence on fossil fuels is reflected in our energy consumption profile, as more than 80% of the global energy supply was generated using fossil fuels in 2021 (IEA, 2021). Our appetite for energy has also ballooned; global energy consumption has increased from 230 EJ to 606 EJ between 1971 to 2019, which means energy consumption per capita has grown by 22% when corrected for the relevant global population sizes. The wide geographical availability of fossil fuels, and the global infrastructure built around their use, has made them an unmissable staple in our energy diet.

Besides the energy and industrial sectors, global transportation -including shipping, aviation, and land vehicles- heavily relies on the use of fossil-derived petroleum fuels. In 2022, around 20-25% of global CO_2 emissions originated from the transport sector and more than 90% of all transport-related energy consumption stemmed from fossil fuel use (IEA, 2023). Although there are differences in the exact chemical composition of fuels dependent on their purpose, they all contain paraffins (straight-chain alkanes) of carbon chain length between C_5-C_{20}. These compounds make up the largest proportion of all fuels by weight, and therefore represent an important class of target molecules for improving the sustainability of the transportation sector (Zhou et al., 2018).

Because liquid hydrocarbon fuel usage is ubiquitous worldwide, developing low-carbon alternatives that can leverage the existing global supply chains would have the potential to be deployed rapidly, resulting in high CO_2 emission reductions. Sustainable fossil fuel alternatives that can be used in existing infrastructure with minimal adaptation are coined 'drop-in' fuels.

These can be produced chemically or through biotechnological means, the latter of which is of special interest due to the relatively low energy input required to operate bioprocesses in contrast to chemical synthesis, and the ability of microbes to convert a wide variety of substrates into desirable compounds. However, to compete with fossil fuels and chemically synthesised drop-in fuels, process performance and costs of these biotechnological processes must be made more attractive.

1.3 Thesis scope and objectives

The broad aim of the research presented in this thesis is to contribute towards the development of microbial alkane biosynthesis as a sustainable alternative to fossil fuels. This field has garnered much attention over the past decade with the promise of delivering commercially viable drop-in biofuel production as a new addition to the global energy portfolio. However, despite significant advances, the field has yet to mature towards the industrial scale, hindered by inefficient production metrics and high costs. The work presented in this thesis aims to address several of these challenges, specifically:

i. By demonstrating that the biosynthesis of hydrocarbons in *E. coli* is possible using alternative carbon sources to glucose.
ii. By designing novel strategies to enable the growth-coupling of alkane biosynthesis enzymes, strictly linking fuel production to microbial growth.
iii. By engineering a microbial enzyme engineering strain capable of selecting for the activity of an alkane-producing enzyme.
iv. By providing fundamental insights into the metabolic plasticity of *E. coli*.

1.4 Thesis layout

The research presented in this thesis is conceptually divided into two sections, one focussing on tool development and hydrocarbon biosynthesis (“Alkanes”), while the second section revolves around the construction of auxotrophic strains capable of selecting for specific

enzyme activities (“Auxotrophy”). To connect these overarching themes there is a bridging chapter, bringing together both topics (“Alkanes & auxotrophy”) (see Figure 1.1).

Literature review – Progress in hydrocarbon biosynthesis

- **Chapter 2** encompasses an in-depth review of the field, covering sustainable hydrocarbon production, microbial chassis development, metabolic engineering for bioproduction and the use of non-sugar carbon sources. Finally, progress in the field of enzyme engineering techniques for hydrocarbon-producing enzymes is reported.
- This chapter contains a literature review published in Biotechnology Advances: “*Nielsen, J. R., Weusthuis, R. A., & Huang, W. E. (2023). Growth-coupled enzyme engineering through manipulation of redox cofactor regeneration”*

Alkanes – Establishing hydrocarbon biosynthesis in *E. coli*

- **Chapter 3** describes the development and characterization of a MoClo-compatible plasmid toolkit, consisting of backbone plasmids with low- and medium copy numbers to enable expression levels better suited to biosynthetic pathways.
- **Chapter 4** covers the various steps taken to optimise hydrocarbon biosynthesis in *E. coli*. Besides presenting several novel optimization approaches, this chapter also focusses on the use of C_2 carbon sources for the production of hydrocarbons.

Alkanes & auxotrophy – Coupling microbial growth to alkane biosynthesis

- **Chapter 5** presents several novel approaches to metabolically link the biosynthesis of alkanes to growth, including via redox cofactor auxotrophy and the reductive glycine pathway. A combination of theoretical analyses and experimental work is presented, culminating in successful growth-coupling of an alkane biosynthesis pathway.

Auxotrophy – Development of widely useable enzyme engineering strains

- **Chapter 6** details the construction of an NADH auxotrophic *E. coli* lacking all the canonical central pathways used for NAD^+ reduction. The strain is extensively

characterised, and both endogenous as well as exogenous metabolic pathways are used to rescue its growth.

- **Chapter 7** investigates the essentiality of the metabolic routes connecting pyruvate to acetyl-CoA. The results indicate that oxygen sensitive enzymes can contribute to acetyl-CoA formation even in aerobic environments, further highlighting the flexibility of *E. coli* central metabolism.

1.5 Key research output

This thesis comprised a combination of fundamental and applied synthetic biology research. Some of the main research findings have been summarised below:

- A plasmid toolkit containing MoClo-compatible plasmids will be made publicly available to others to accelerate their synthetic biology research. Besides a broadening of the available cloning methods, this thesis demonstrated how this toolkit could be used to optimise expression of growth-coupled enzymes.
- The C_2 compounds ethanol and acetate were successfully used as sole carbon sources for the biosynthesis of fatty aldehydes, alcohols, and alkanes in *E. coli*. This represents the first report describing the use of C_2 carbon sources for hydrocarbon biosynthesis in *E. coli* using the AAR-(ADO) enzymes.
- A growth-coupling approach was used to link the activity of the AAR-ADO pathway to the fitness of a formate-dependent strain via the reductive glycine pathway. This represents a novel approach for potential growth-based engineering of the AAR-ADO alkane biosynthesis pathway.
- A novel *E. coli* mutant 'M1' was discovered that was engineered to lack all central NADH-generating pathways but was found to carry an additional mutation in the NADH-dehydrogenase I (NDH-I) complex (*nuoF*^{E183A}). This mutation allows oxidative phosphorylation to operate using NADPH, as opposed to just NADH, representing a significant redesign of *E. coli* redox metabolism.

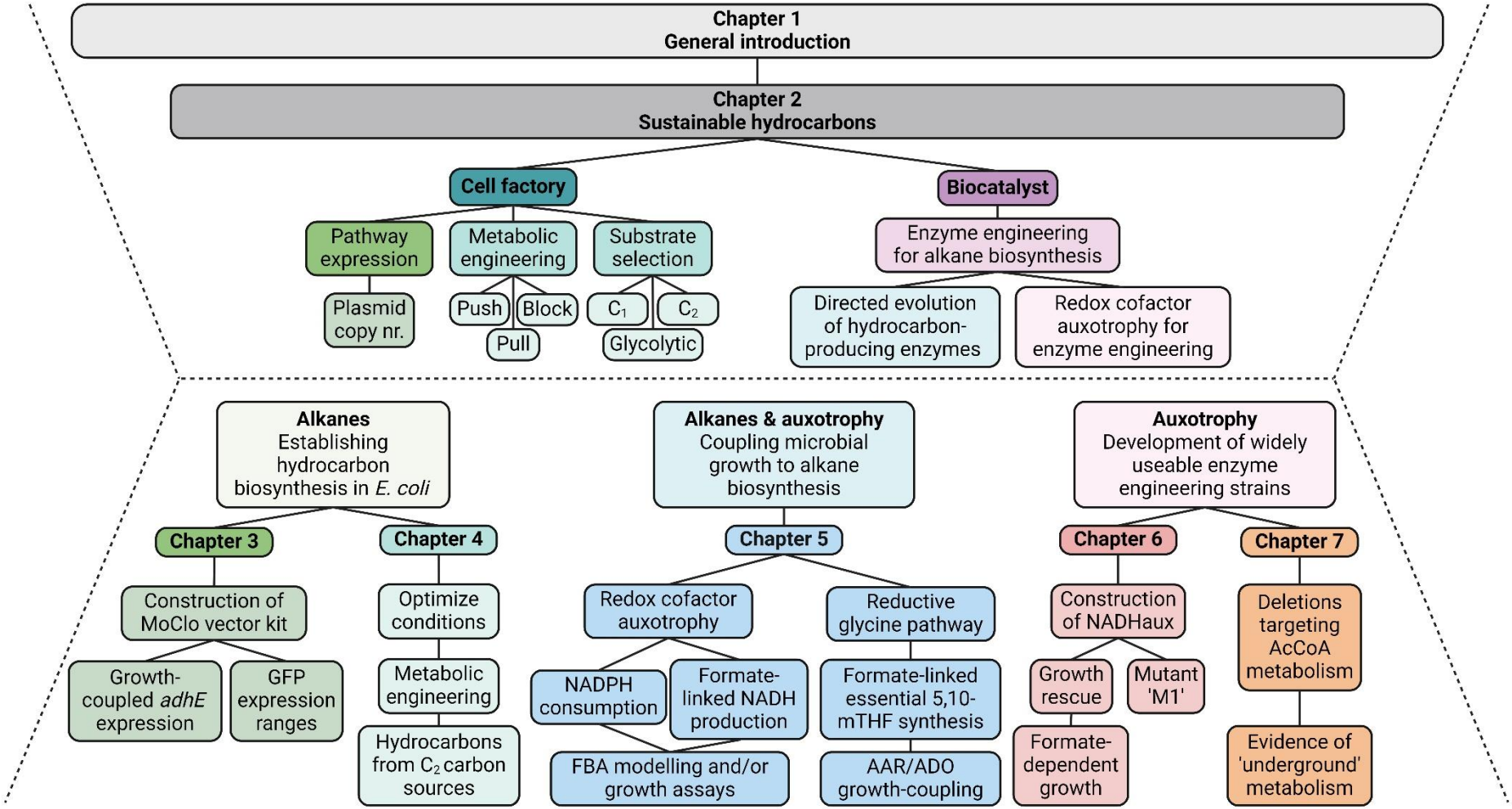


Figure 1.1. Schematic overview of topics covered in this thesis. The work was conceptually divided into three related research tracks: 'Alkanes', 'Alkanes & auxotrophy' and 'Auxotrophy', with the central aim of contributing to the development of microbial alkane biosynthesis.

2. Sustainable hydrocarbons

The transition away from fossil fuels will take the form of a gradual shift, as opposed to a sudden switch, towards a different energy infrastructure. This change will be marked by the use of so called ‘drop in’ fuels exhibiting more palatable carbon footprints than fossil fuels, while still being compatible with the deeply embedded global oil & gas infrastructure.

The production of sustainable hydrocarbons represents an important step towards a full decarbonization of the transport sector, especially in shipping and aviation. Several chemical methods exist to upgrade simple carbon molecules into fuel-length hydrocarbons, some of which are reaching commercial maturity. In contrast, the commercial exploitation of biological processes for drop-in hydrocarbon production have mostly been unsuccessful, exemplified by biological fuel producers transitioning into higher-value specialty chemical production.

The advantages of using biological over chemical production methods include the relatively mild process conditions required for biological catalysis, the possibility of upgrading a wide variety of substrates, and lower amounts of chemical waste production (S. Y. Lee et al., 2019). However, this often comes at the expense of lower process performance, as efficient biological catalysis is challenging to establish, difficult to manipulate once in use and can deviate from process requirements due to the stochastic properties of biological systems. Several challenges therefore stand in the way of biological hydrocarbon production:

i) Discovering and engineering efficient hydrocarbon production enzymes;

ii) Maximizing product synthesis by metabolic- and process engineering;

iii) Establishing the use of renewable and affordable carbon sources as substrates for hydrocarbon biosynthesis.

In the following sections we elaborate on these challenges and provide perspectives for their future development.

2.1 Chemical hydrocarbon synthesis

Several methods exist to (electro)chemically synthesise hydrocarbon molecules of varying chain lengths from basic carbon building blocks. These technologies include Fischer-Tropsch (FT) synthesis, hydrotreated esters and fatty acids (HEFA) and Alcohol-to-Jet (AtJ) processes.

Briefly, FT synthesis requires synthesis gas (syngas), consisting of H_2 and CO or CH_4, and requires high pressures, temperatures, and the presence of metal catalysts to polymerise the gasses into longer chain hydrocarbons (Meurer & Kern, 2021). The FT method can generate a wide spectrum of products ranging from short-chain hydrocarbons to waxes, the latter of which would need to be cracked to achieve suitable chain lengths for fuel use (Shafer et al., 2019). The FT process can be used to produce biokerosene (also called sustainable aviation fuel, SAF), for example from gasified organic- or non-organic waste. An advantage of this method is the broad availability of substrate, since syngas can be derived from a variety of (non) organic sources. However, the requirement for a clean source of H_2 represents a significant bottleneck, as sources of green H_2 are limited and in high demand worldwide. In addition, the high temperatures and pressures required for FT synthesis mean that this method is energy intensive, reducing its efficiency in terms of the final product energy balance.

Another notable method to produce drop-in biofuels is via hydrotreatment of esters and fatty acids (HEFA). This process utilises the already existing fatty acid (esters), for example from used cooking oils, and simply deoxygenates them to generate drop-in fuels. Occasionally further cracking of the resulting products is required to achieve the desired carbon chain lengths. A variety of hydrocarbon end products can be generated using HEFA, for example a process for SAF production using this method has been approved by the American Society for Testing and Materials (ASTM), dubbed HEFA-Synthetic Paraffinic Kerosene (HEFA-SPK) (M. Song et al., 2023). In contrast to FT synthesis, an attractive feature of this technology is that polymerization of the carbon molecules is not required. Therefore, HEFA represents one of the few commercial-scale technologies available for drop-in fuel production, which has led

to high demand, substrate scarcity, and high prices. Besides this, the requirement for a source of green H_2 further introduces a substrate scarcity.

Finally, Alcohol-to-Jet (AtJ) is a technology approved by the ASTM as a method to produce SAF from ethanol and isobutanol. This process first dehydrates the alcohol, then oligomerises the remaining hydrocarbons, followed by a hydrogenation and fractionation step to form (liquid) fuels of varying chain lengths (Xie et al., 2024). This technology is currently being deployed on a commercial scale by several parties such as LanzaJet. Since pure alcohol fuels cannot usually be blended with traditional fuels at high percentages, a clear advantage of upgrading alcohols to longer chain hydrocarbons means that they circumvent this 'blend-wall' (Valdivia et al., 2016). By upgrading alcohol products, the resulting fuels are much more widely useable. Similarly to all other chemical synthesis methods, AtJ also requires a source of green H_2.

The technologies described above provide a variety of methods to upgrade simple carbon molecules to more complex drop-in fuels compatible with petrochemical infrastructure. A clear advantage of all these processes is their scalability. In contrast, the requirement for an abundance of green H_2, the high energy demands, and scarce substrates (HEFA) currently limit their ability to scale in a truly sustainable way. In the next sections, both traditional and more advanced biofuels will be reviewed as sustainable sources of energy.

2.2 Biofuels: generations and limitations

1st generation biofuels

At the time of writing the most commonly used biofuels belong to the '1st generation' category, defined by their derivatization from edible crops. This category is made up almost exclusively out of biodiesel and bioethanol. The former is produced through a transesterification reaction using mostly canola- or soybean oil, while the latter is produced through yeast fermentation of sugars from sugarcane, sugar beet or maize to ethanol (Cavelius et al., 2023).

Since atmospheric CO_2 is fixed by the crop plant to form biomass, the release of CO_2 by the combustion of biofuel results in a closed carbon loop making them carbon-neutral in theory. In practice however, the cradle-to-grave greenhouse gas (GHG) emissions related to bioethanol use are only roughly 50% that of fossil fuels, driven mostly by the high GHG emissions involved in crop cultivation (Muñoz et al., 2014). Bioethanol fuel can be blended with petroleum and used in unmodified internal combustion engines (ICEs). Most gasoline in Europe and the United States is currently blended at percentages ranging between 5-10% v/v. In Brazil however, up to 100% ethanol is used in vehicles with ICEs specifically made for ethanol combustion (Jeswani et al., 2020). Using unsuitable blend percentages in non-adapted ICEs may result in damage or loss of reliability in cold conditions due to some of the chemical properties of biodiesel and ethanol, such as their elevated hygroscopicity. This results in a so-called 'blend-wall' which limits their large-scale adoption as carbon emission-mitigating fuel alternatives (Valdivia et al., 2016).

A more significant problem with the use of 1st generation biofuels however arises from the food vs. fuel debate. Vast areas of arable land are required to produce biofuel crops, for example approximately 25% of U.S. agricultural land was used for production of crops for bioethanol fermentations in 2015 (Mumm et al., 2015). This land use competes directly for food production while arable land is in high demand. In addition, densely populated countries, or countries with climates unsuitable for biofuel crop production will not be able to dedicate arable land to the production of biofuels. The significant drawbacks involved in using arable land for fuel production have spurred research into new, food-independent, biofuels.

2nd generation (advanced) biofuels

The second category of biofuels is referred to as '2nd generation', or 'advanced', characterised by the use of non-edible biomass for the production of fuels. This involves the use of non-edible parts of crops or waste from agriculture and forestry, requiring the separation of sugars from lignocellulosic material (Cavelius et al., 2023). Since lignocellulosic material is exceptionally recalcitrant and also contains non-sugar lignin, the sugar-containing

(hemi)cellulose fractions must be efficiently separated to ensure maximum yields. These fractions must subsequently be hydrolysed to release sugars for catabolism by microbes, a process called saccharification, requiring either exogenous enzyme supplementation or the endogenous production of hydrolytic enzymes (Ambaye et al., 2021). While this 2nd generation technology is demonstrably functional on a lab scale, achieving robust and efficient conversion of lignocellulosic material at scale has proven exceptionally challenging. Indeed, many attempts at establishing commercial 2nd generation biofuel processes have ultimately been unsuccessful.

Several new technologies may yet provide opportunities for this sector. For example, cost- and chemical-efficient separation of (hemi)cellulose and lignin fractions are under development. In addition, more efficient saccharification of (hemi)cellulose fractions may reduce costs and increase process viability. Finally, the use of (engineered) microbes able to digest both C_5 and C_6 sugars simultaneously will enable higher yields of fuel product per unit biomass.

3rd and 4th generation biofuels

The final two 'generations' of biofuels differ from the previous generations in that they do not involve arable land use and carbon is directly derived from CO_2. The use of cyanobacteria, micro- and macroalgae is envisaged for 3rd generation fuels specifically. These organisms are light-harvesting and fix CO_2 at higher rates than terrestrial plants, improving energy efficiency of fuel production (Cavelius et al., 2023). Some species are known to accumulate high levels of lipids and other carbon storage compounds, further enabling their use as a fuel source. Since CO_2 is required for biomass- and product formation, the integration of CO_2-fixing processes with those that emit large amounts of CO_2 can directly ensure that emission are converted to useful bioproducts. However, fuel production using light-harvesting organisms has been attempted for decades and has yet to reach commercial scales. The most severe bottleneck is the cultivation of these organisms. To reduce costs this can be done in open-air ponds, which exposes the process to interfering variables and requires high surface areas to

harvest enough sunlight. Alternatively, cultivation can be performed in photobioreactors, which enable control over the process conditions but come at high operating costs making them less suitable for low-cost products such as fuels.

Finally, the term '4th generation biofuels' is relatively loosely defined but includes the direct fixation of CO_2 and the introduction of genetic modifications to enhance microbes for fuel production. Some interpretations include only algae and cyanobacteria in this definition, while others involve all engineered microbes. We include all organisms in this term, since considerable steps have been taken to produce sustainable fuels using non-light harvesting microbes. For example, gas fermentation technology utilised by LanzaTech has reached commercial maturity for the production of ethanol from concentrated CO_2 streams (Köpke et al., 2010), as well as the pilot-scale production of acetone and isopropanol (Liew et al., 2022). The commercial exploitation of cyanobacteria and algae for lipids and fuel production has been less successful; currently no process is operated at commercial scale. The requirements for ideal light intensity and CO_2 availability, the high capital- and operating costs, and high energetic input required for product extraction have resulted in a stagnation in commercial interest in algal and cyanobacterial fuels (Hoang et al., 2022; Singh et al., 2024).

None of the above technologies involve the direct production of 'drop-in' fuels. An important shift in the field of biofuels will be the biosynthesis of molecules chemically identical to fossil fuels, which is covered in the remainder of this literature review.

2.3 Enzymes and pathways for hydrocarbon biosynthesis

This section is intended to serve as a brief introduction to the aliphatic hydrocarbon biosynthesis pathways currently known to exist in nature (Figure 2.1). We focus specifically on bacterial biosynthesis pathways, as existing eukaryotic pathways are relatively understudied in terms of microbial bioproduction. A more in-depth description of each pathway, and any engineering efforts directed at them, has been compiled in section 2.7.1.

In the space of only several years, a host of hydrocarbon synthesis enzymes were discovered and characterised, sparking a surge in advanced biofuel research. Isolation of the first long-chain alkene biosynthesis pathway occurred in 2010, with the characterization of head-to-head fatty acid condensing enzymes encoded by *oleA, oleBC and oleD* in *Micrococcus luteus* (Beller et al., 2010). This was quickly followed by a study from Schirmer and coworkers, who confirmed that cyanobacterial acyl-ACP reductase (AAR) and aldehyde deformylating oxygenase (ADO) operated in tandem to generate n_{-1} alkanes from fatty acyl-ACP precursors (Schirmer et al., 2010). In contrast to *oleABCD* activity, AAR-ADO activity directly produced diesel-length fuels, hence this pathway received much attention in the following years.

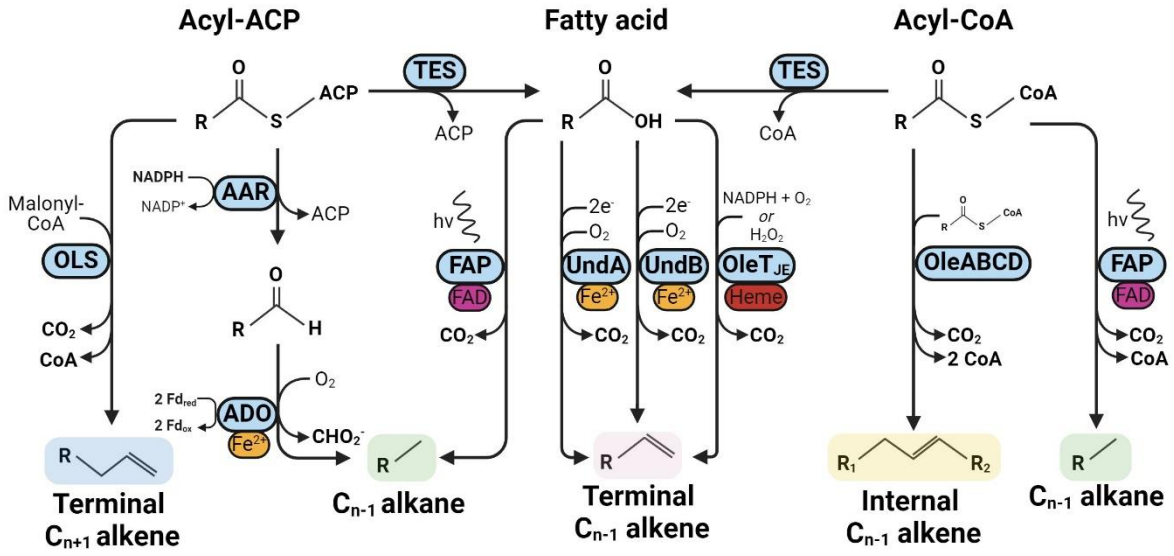


Figure 2.1. Enzymatic pathways for bacterial hydrocarbon synthesis. All known biosynthetic pathways for alkene and alkane biosynthesis, including: cyanobacterial olefin synthase (OLS), acyl-ACP reductase (AAR), aldehyde deformylating oxygenase (ADO), fatty acid photodecarboxylase (FAP), α-olefin synthases UndA/UndB/OleT$_{JE}$ and head-to-head fatty acid condensation by OleABCD.

Only 8 months later Rude and coworkers published a study using a similar reverse genetics approach as used by Shirmer *et al.,* where they found a P450 fatty acid decarboxylase responsible for terminal alkene synthesis in *Jeotgalicoccus,* dubbed OleT$_{JE}$ (Rude et al., 2011). When expressed heterologously in *E. coli*, OleT$_{JE}$ was capable of producing C_{17} and C_{15} terminal alkenes. More work on cyanobacterial hydrocarbon synthesis revealed terminal

alkene production by olefin synthase (OLS), a multidomain protein with homology to type I polyketide synthases (Mendez-Perez et al., 2011). This protein complex was found to be responsible for $C_{19:1}$ alkene synthesis, originating from a carbon C_{n+1} stoichiometry differing from that of C_{n-1} found in $OleT_{JE}$.

In 2014, a nonheme iron oxidase was isolated that was responsible for 1-undecene biosynthesis in several *Psuedomonas* species, named *undA* (Rui et al., 2014). The authors screened 6000 fosmid libraries in *E. coli* to identify isolates producing 1-undecene. The production of medium-chain aliphatic hydrocarbons was of considerable interest, due to their compatibility with kerosene. Closely related to *undA* was the discovery of *undB* by the same authors a year later, who used a similar genetic library screening approach to identify another 1-undecene biosynthesis enzyme (Rui et al., 2015). In contrast to cytosolic *undA*, the *undB* enzyme was a membrane-bound desaturase-type enzyme, capable of decarboxylating C_6-C_{18} fatty acid substrates but with a clear preference for C_{10}-C_{14}.

The most recently discovered aliphatic hydrocarbon producer uses an entirely different catalytic mechanism, namely photodecarboxylation. The fatty acid photodecarboxylase (FAP) was isolated from *Chlorella variabilis* (therefore often referred to as *Cv*FAP) and carries an FAD prosthetic group, requiring blue light to catalyse its activity (Sorigué et al., 2017). This enzyme was recently found to preferentially use acyl-CoA substrate as opposed to free fatty acids (Li et al., 2020). In contrast to many other hydrocarbon synthesis enzymes, but similar to ADO, FAP does not produce terminal alkenes but instead produces a mixture of alkanes and internal alkenes depending on the degree of saturation of the substrate.

Further detail on these pathways and relevant enzyme engineering efforts is covered in section 2.7.1.

2.4 Development of *Escherichia coli* as a cell factory for alkane biosynthesis

2.4.1 Microbial chassis for alkane biosynthesis

Successful deployment of bioproduction processes rely on many factors, one of which is the chosen microbial chassis. This choice is in turn dependent on several aspects, including the intended target product, the production conditions, and the desired substrate(s). For example, some microbes may natively produce the desired compound(s) albeit at low levels, potentially simplifying subsequent engineering steps. Others may natively be able to use a certain desirable substrate, removing the need for heterologous expression of substrate catabolism pathways. Finally, if the production of a toxic compound is required, using a host that is natively tolerant to said compound may be more prudent than engineering tolerance in a sensitive host.

In general, lab-scale bioproduction studies benefit most from using a well-characterised host chassis that can be genetically engineered with ease, even if this may not be the ideal chassis in terms of product accumulation, product tolerance, substrate uptake or the intended final growth conditions for larger scale production. The facultative anaerobe *Escherichia coli* has developed into the standard workhorse for molecular- and synthetic biology. It is fast-growing, genetically tractable, and due to its popularity represents one of the most extensively characterised microbes. As such, *E. coli* is frequently chosen as the initial bioproduction host, providing a testing platform for characterization and optimization of the desired process.

The first demonstration of a heterologous alkane biosynthesis pathway was performed in 2010, using *E. coli* MG1655. As detailed in Chapter 2.1, the expressed cyanobacterial pathway consisted of acyl-ACP reductase (AAR) and aldehyde deformylating oxygenase (ADO), generating a C_n aldehyde intermediate from C_n acyl-ACP, followed by a C_{n-1} alkane and formate. Schirmer and coworkers noted that around 80% of all hydrocarbon products were detected extracellularly (Schirmer et al., 2010). Since product separation often constitutes a large fraction of total bioproduction costs, this was promising evidence that *E. coli* may be a

suitable host for heterologous alkane biosynthesis. In later years, *E. coli* has been the target of extensive metabolic engineering to improve alkane production metrics, which we expand on in later sections of this chapter.

Besides *E. coli*, only a handful of other hosts have been assessed for their capacity to biosynthesise alkanes. An engineered strain of the oleaginous bacterium *Rhodococcus opacus* was used to produce long-chain hydrocarbons (LCHCs) from glucose using the AAR-ADO pathway (H. M. Kim et al., 2019). This bacterium is capable of accumulating triacylglycerols (TAGs) to 76% of its dry cell weight (Kurosawa et al., 2010). The introduced genetic modifications included deletion of an alkane monooxygenase *alk-1*, deletion of acyl-CoA dehydrogenase *fadE*, overexpression of acyl-CoA synthase *fadD*, overexpression of acyl-CoA- and acyl-ACP reductases (ACR and AAR, respectively), and finally overexpression of ADO. Fed-batch fermentation with glucose using this strain led to LCHC titres of ~5.2 g/L with pentadecane as the main product, representing the highest alkane titre achieved to date across all microbial hosts. This work demonstrated how a host capable of natively generating high levels of relevant intermediate compounds can be exploited to produce alkanes as the final product.

Two other works described alkane biosynthesis using *Acinetobacter baylyi* and *Cupriavidus necator*, both representing proof-of-concept studies for using substrates other than glucose for alkane biosynthesis. Firstly, alkane production using *A. baylyi* was achieved with acetate as a carbon source, which in turn was produced by *Acinetobacter woodii* growing on H_2 and CO_2. The *A. baylyi* strain expressed the AAR-ADO pathway, resulting in a titre of around 74 µg/L after a 9h fed-batch cultivation (Lehtinen et al., 2018). The second study used *C. necator*, an autotrophic PHB-producing microbe, to produce alkanes from CO_2 as a sole carbon source. The authors deleted the PHB biosynthesis pathway to redirect acetyl-CoA flux towards fatty acid biosynthesis and overexpressed the AAR and ADO genes to stimulate alkane production. An alkane titre of 4.4 mg/L was achieved using CO_2 as a sole carbon source during production, given that fructose was first used for biomass accumulation (Crepin et al., 2016).

Besides bacteria also yeasts have been the subject of hydrocarbon production studies. Both *Saccharomyces cerevisiae* and *Yarrowia lipolytica* have been exploited as alkane biosynthesis hosts using the AAR-ADO pathway, although *Y. lipolytica* has also been endowed with the *Cv*FAP pathway (see Section 2.3 or 2.7.1 for more information on the *Cv*FAP enzyme). The production of alkanes in *S. cerevisiae* was achieved by co-expression of AAR-ADO with the *E. coli* ferredoxin/ferredoxin reductase system, resulting only in heptadecane biosynthesis. The authors found that additional deletion of hexadecanal dehydrogenase Hfd1 made the production of tri- and pentadecane possible, even without expression of acyl-ACP reductase (AAR). Additional expression of AAR improved performance slightly, reaching a final titre of ~22 µg/gDW (alkanes per cell dry weight) (Buijs et al., 2015). In the case of *Y. lipolytica,* the overexpression of a carboxylic acid reductase along with ADO was found to result in an alkane titre of around 23 mg/L, mostly consisting of heptadecane, heptadecene and pentadecane. The authors also found that the alk(a/e)ne product profile could be altered considerably based on the intracellular localization of an acyl-CoA reductase together with ADO, demonstrating how engineering eukaryotic hosts for alkane biosynthesis can add a level of complexity (Xu et al., 2016). In other work exploring *Y. lipolytica* as an alkane production host a fatty acid overproducing strain was used for *Cv*FAP-mediated alkane biosynthesis, resulting in a maximum titre of ~59 mg/L alkanes (Bruder et al., 2019). A report published shortly after identified acyl-CoAs as the preferred substrate for *Cv*FAP, instead of solely fatty acids. The authors therefore decided to use a *Y. lipolytica* strain that did not accumulate fatty acids and they expressed multiple copies of *Cv*FAP to stimulate product-forming flux. This resulted in the accumulation of 1.47 g/L alkanes after a fed-batch fermentation using glucose, and approximately 300 mg/L alkanes when using acetate as a carbon source (Li et al., 2020). The mixed successes achieved using yeast hosts for alkane biosynthesis demonstrated how fundamental knowledge of the preferred enzyme substrates, the effects of enzyme localization, and influence of potentially competing pathways is highly important for optimizing bioproduction metrics.

The studies using hosts other than *E. coli* demonstrated that alkane biosynthesis can be achieved in a variety of hosts using a variety of substrates. In addition, the high titres achieved using *R. opacus* (AAR-ADO) and *Y. lipolytica* (*Cv*FAP) pointed towards the advantage of using natively oleaginous hosts. However, the vast majority of alkane biosynthesis work has been performed using *E. coli* as chassis microbe, including seminal works initially characterizing the AAR-ADO pathway and many of the studies exploring metabolic engineering strategies to improve alkane biosynthesis. For this reason, we will expand further on the use of specifically *E. coli* as an alkane production host.

2.4.2 Metabolic engineering strategies to improve hydrocarbon biosynthesis in *E. coli*

As genetic engineering techniques have improved over the past decades, the manipulation of cell metabolism to enhance bioproduction has become commonplace. Using a combination of techniques including gene knockouts, flux redirection, cofactor balancing, and enzyme overexpression, bioproduction can be enhanced in terms of titre, yields and productivities. The production of (long-chain) hydrocarbons, such as alkanes, fatty alcohols, and fatty acids, often involve many enzymatic reactions due to the required carbon chain elongation steps. Fatty acid synthesis is the main biosynthetic pathway responsible for chain elongation in *E. coli*. As a result, most metabolic engineering strategies have focused on improving substrate flux towards this pathway, and removing competing pathways that may siphon off carbon and reducing power in the process.

These approaches can roughly be categorised as ‘push’, ‘pull’ and ‘block’ strategies, referring to the effect they have on the respective metabolic pathway(s) and the resulting thermodynamic balance between substrate(s), intermediate(s), and product(s) (Mehrer et al., 2017). ‘Push’ strategies aim to increase product synthesis by redirecting and/or increasing flux towards the desired biosynthetic pathway or increasing the availability of suitable cofactors for product synthesis. ‘Pull’ strategies aim to increase flux through terminal enzymes of the production pathway, often through overexpression or engineering of the terminal pathway enzymes. Besides directly targeting enzyme activity and expression, pull strategies can also

involve enhancing product secretion and separation from the fermentation broth (e.g. by phase separation), as this shifts the thermodynamic balance towards product formation. Finally, 'block' strategies involve the deletion or inhibition of competing pathways that may consume carbon or cofactors destined for product synthesis. These 'blocks' can be implemented at any metabolic node along the pathway from substrate to product. Block strategies are popular since suitable targets can be identified easily, and usually only a handful of gene deletions are required to achieve significant gains in bioproduction metrics. Besides directly affecting enzyme levels (e.g. overexpression/deletion), these strategies can also involve alterations in gene regulation. In the following sections each strategy will be separately investigated in more depth, with a focus on their use in enhancing hydrocarbon biosynthesis.

2.4.2.1 Push strategies

The synthesis of hydrocarbon products in *E. coli* utilises the native fatty acid synthesis (FAS) pathway, named fatty acid synthase II (FAS II). Therefore, understanding FAS in *E. coli* can guide engineering approaches to improve this pathway's flux. Firstly, initiation and elongation of fatty acids involves several cytosolic enzymes: *accABCD, fabA, fabB, fabD, fabF, fabG, fabI, fabH* and *fabZ* (Cronan, 2021; Janßen & Steinbüchel, 2014). The first committed step of FAS is the carboxylation of acetyl-CoA to malonyl-CoA by the AccABCD complex. In a 'push' attempt, Davis and coworkers attempted to overexpress *acc* enzymes separately and in combinations, resulting in up to 6-fold improvements in FAS rates but only with additional overexpression of a thioesterase (Davis et al., 2000). Instead of directly overexpressing structural FAS enzymes, global upregulation of FAS was achieved through the expression of *fadR*, the fatty acid degradation regulator, leading to 7.5-fold increases in free fatty acid titre (Zhang et al., 2012). This approach was also found to lead to increases in alkane titres when expressing the AAR-ADO pathway (X. Song et al., 2015).

The FAS pathway consumes large amounts of NADPH cofactors; 14 mol NADPH are required for elongating acetyl-CoA to 1 mol palmitoyl-ACP. When using the AAR-ADO alkane biosynthesis pathway, an acyl-ACP must subsequently be reduced and deformylated to form

pentadecane at the cost of another 3 mol NADPH (Fatma et al., 2018). Two widely used methods to regenerate NADPH include supplementation of the medium with formate or phosphite and heterologously expressing the respective dehydrogenase enzyme (Calzadiaz-Ramirez et al., 2020; Jaroensuk et al., 2019; Johannes et al., 2007). Another option is to redirect carbon flux through metabolic pathways that yield (more) NADPH. By knocking out the glucose-6-phosphate isomerase (*pgi*) gene in *E. coli*, carbon flux from glucose is forced through glucose-6-phosphate dehydrogenase and subsequently the Entner-Doudoroff (ED) or oxidative pentose phosphate pathway (PPP) pathways. This results in the generation of one or two mol NADPH per mol glucose, respectively, before entering lower glycolysis and the TCA cycle. Several studies have introduced perturbations in glycolytic metabolism to enhance the yield of NADPH-dependent products (Bartek et al., 2010; Chemler et al., 2010; Chin & Cirino, 2011; W.-H. Lee et al., 2013). The study reporting the highest alkane titre in *E. coli* to date used an overexpression of *zwf* (glucose-6-phosphate dehydrogenase) and a deletion of *edd* to funnel carbon into the PPP with the goal of increasing NADPH availability (Fatma et al., 2018). Another study showed that the expression of an $NAD(P)^+$-dependent formate dehydrogenase along with acyl-CoA reductase and ADO resulted in a 50% alkane yield (mol/mol) from exogenous fatty acids, by recycling the formate produced by ADO and preventing acidification of the medium (Jaroensuk et al., 2019). These studies demonstrated that cofactor balancing can contribute to ensuring high alkane biosynthesis titres and yields.

2.4.2.2 Pull strategies

While ‘push’ strategies tend to force a higher flux towards the starting point of product synthesis, ‘pull’ strategies focus on improving flux through the terminal product-yielding reactions. These also include bioprocess-related changes aimed at increasing product secretion from the cell and/or separation from the culture broth.

Improving the activity of terminal enzyme flux can be achieved through optimization of enzyme expression or culture conditions, however the most impactful changes are often achieved through enzyme engineering. Terminal enzyme activity is often the bottleneck in hydrocarbon

biosynthesis, making engineering of the relevant enzymes a crucial ‘pull’ strategy to improve process performance. We elaborate further on enzyme engineering efforts for improving alkane biosynthesis in sections 2.6 and 2.7.1.

Other ‘pull’ strategies don’t target the enzymatic activity itself, but rather target the secretion and separation of product from the cell and culture broth, respectively. Lower concentrations of product can improve the thermodynamic balance towards product synthesis, as opposed to a situation where product accumulates. Since fuel-length hydrocarbons are often insoluble in aqueous environments, or gaseous at short chain-lengths, product accumulation hindering enzyme activity is not expected to be a considerable issue. Indeed, the first study reporting AAR-ADO pathway activity noted that ~80% of all hydrocarbon products were found extracellularly (Schirmer et al., 2010). One of the main perks of hydrocarbon biosynthesis is the tendency of these molecules to separate from the fermentation broth naturally due to their hydrophobicity and lower density than water, greatly simplifying downstream processing. Still, several studies have found that overlaying aqueous culture broth with hydrocarbons, such as dodecane, can significantly increase product synthesis (Fatma et al., 2016, 2018; Youngquist et al., 2013). These studies did not investigate whether this is simply due to a more efficient extraction of product from the cells and culture broth, or if the actual flux through the biosynthetic pathway was improved. Either way, hydrocarbon overlaying can instigate more efficient product separation directly into the non-aqueous phase, highlighting it as a useful ‘pull’ strategy for hydrocarbon biosynthesis.

2.4.2.3 Block strategies

While ‘push’ and ‘pull’ strategies usually involve increasing flux through enzymatic reactions, ‘block’ strategies aim to remove certain reactions entirely from cell metabolism to avoid them carrying flux. This technique is most frequently used to stop unwanted diversion of carbon towards side product formation.

In *E. coli,* popular ‘blockage’ points involve metabolic pathways producing native fermentation products of *E. coli*, including lactate (*ldhA*), ethanol (*adhE*), formate (*pflB*) and acetate (*pta/ackA*). Indeed, the highest alkane titre and yield reported in the literature so far was achieved in *E. coli* strains with a Δ*ldhA* Δ*pta* Δ*pflB* genotype, among other changes (Fatma et al., 2018).

The AAR-ADO pathway results in the formation of a C_n aldehyde intermediate, which can be reduced to an alcohol by one of *E. coli*’s many native aldehyde reductases (ALRs). This decreases alkane yields, and as such these ALRs have frequently been targeted for deletion to avoid carbon flux towards alcohol formation. The substrate specificity and activity of these ALRs have been characterised extensively over the last decade. These efforts have resulted in the construction of strains with deletions in ALRs, leading to improved production of aromatic aldehydes (Kunjapur et al., 2014), medium chain aldehydes (Rodriguez & Atsumi, 2014), long chain alcohols (Fatma et al., 2016) and long chain alkanes (X. Song et al., 2015). The substrate specificities of most ALRs were found to be broad, meaning that widespread deletion of ALRs was required to construct strains with low ALR activity, as exemplified in work by Kunjapur *et al* (7 deletions made) and Rodriguez *et al* (13 deletions made) (Kunjapur et al., 2014; Rodriguez & Atsumi, 2014). Even so, several ALRs were identified that strongly competed for ADO’s native substrate (C_{14}-C_{18} aldehydes), including *ybbO*, *yqhD* and *yahK*. The deletion of *yqhD* was shown to improve alkane biosynthesis 2-fold (X. Song et al., 2015). More recently, aldehyde oxidation to fatty acids in resting cells was also found to be a significant limitation for aldehyde accumulation in *E. coli*, warranting the identification and deletion of the responsible dehydrogenases (Butler et al., 2023).

In terms of improving alkane biosynthesis in *E. coli* via ‘block’ approaches, the deletion of native fermentative enzymes, aldehyde reductases, aldehyde dehydrogenases and NADPH sinks would have to be performed to engineer *E. coli* as a high-performance bioproduction host.

2.5 Substrate selection for alkane biosynthesis

Any bioproduction process requires choosing a suitable carbon source (C-source) to be used for biosynthesis of the desired compound. The choice of C-source largely shapes the ensuing metabolic engineering strategy, since it defines what catabolic pathway will be used, the redox balance involved, the theoretical maximum yields, any metabolic bottlenecks that may be present and any regulatory effects from carbon catabolism that may affect product synthesis. As such, the choice of C-source is one of the first and most impactful decisions defining later bioprocess performance.

2.5.1 Glycolytic carbon sources

Glucose is the most frequently used C-source for lab-scale bioproduction studies using *E. coli*. Glucose confers *E. coli* with the fastest growth rate compared to any other C-source in standard minimal medium conditions and glucose catabolism is extremely well characterised, making it an attractive initial starting point for bioproduction studies. However, glucose is expensive at industrial scale and is inefficiently used in terms of carbon yield, since catabolism of one glucose molecule by the native glycolytic pathway results in the loss of two CO_2 in forming two acetyl-CoA, a carbon loss of 33%. For the biosynthesis of low value, high volume commodities such as hydrocarbons, the use of glucose represents a poor choice in terms carbon retention and the resulting bioprocess economics. At commercial scale, profitable biological alkane production requires yields exceeding 90% of the theoretical yield on glucose to compete with existing bioethanol production processes (Kang & Nielsen, 2016). Although native glycolysis is not carbon efficient, modified glycolytic pathways could improve carbon yields (Bogorad et al., 2013; Iacometti et al., 2022; P. P. Lin et al., 2018). Perhaps a more pressing matter is the sourcing of glucose. Crop-derived glucose is used for commercial bioethanol production, raising concerns regarding arable land use as discussed earlier.

Glycerol represents another potential glycolytic feedstock. This C_3 compound is produced in large amounts as a byproduct from biodiesel production (~10% w/w), and therefore represents

a waste feedstock (Z. Chen & Liu, 2016). A significant advantage of using glycerol is its high degree of reduction per carbon due to the three alcohol groups it carries. This allows the release of more reducing power per catabolised carbon, which offers advantages when attempting to produce highly reduced compounds such as hydrocarbons. However, similarly to glucose, glycerol (C_3) is decarboxylated upon its catabolism to acetyl-CoA (C_2). Glycerol can be catabolised natively by popular biorefinery platform strains including *E. coli* and *S. cerevisiae*, although metabolic engineering efforts are underway to improve its uptake and subsequent conversion to bioproducts (Blankschien et al., 2010; Xiberras et al., 2019).

2.5.2 C_2 carbon sources

There is a keen interest in using carbon sources that avoid the 'food vs. fuel' issue and are derived from sustainable sources. These alternative C-sources can be obtained from e.g. food-, plant- or agricultural waste, or derived electrocatalytically from CO_2.

Of particular interest is acetate, a C_2 carboxylic acid that can be electrocatalytically produced and is also the terminal product of acetogenic fermentation. This is a form of anaerobic digestion (AD) taking place when microbial consortia convert biological matter into short chain carboxylic acids such as acetate (C_2), propionate (C_3) and butyrate (C_4) (Meegoda et al., 2018). A benefit of AD is the flexibility in feedstock input since various types of biological material can be digested to a relatively homogenous product stream. Acetate has been explored as a C-source for various bioproducts, such as malate, succinate, pyruvate, and 3-hydroxypropionate (Gong et al., 2022). Commonly used cell factories such as *E. coli* can also readily metabolise acetate, removing the need for engineering substrate uptake pathways.

Another potential non-sugar C-source includes ethanol which is traditionally produced using yeast fermentation. The resulting bioethanol is currently used in fuel blends but upgrading it to aliphatic hydrocarbons would improve its compatibility with current fuel infrastructures. Compared to acetate, ethanol is a more reduced carbon source, which results in a higher availability of reducing power during its catabolism. Previous work has shown that both acetate

and ethanol can be used as carbon sources for polyhydroxybutyrate (PHB) synthesis, with ethanol outperforming glucose in terms of titre and yield (Sun et al., 2020). The microbial upgrading of ethanol could also be linked to gas fermentation processes utilising *Clostridium*, which are currently used to produce ethanol from waste CO_2 at industrial scale (Köpke et al., 2010). A benefit of using C_2 carbon sources is that their catabolism towards acetyl-CoA often involves relatively few enzymatic reactions. Since acetyl-CoA is the starting point for fatty acid synthesis in *E. coli,* the use of these C-sources enables greater control over the redirection of carbon flux towards fatty acid-derived products compared to using highly branched glycolytic pathways.

2.5.3 C_1 carbon sources

Although there are clear advantages to using C_2 carbon sources over pure sugars in terms of sustainability and carbon efficiency, they must still be derived from primary feedstock or processed before they can be used as C-sources for biological processes. This increases the complexity of both obtaining sufficient C-source and upgrading this to product while retaining commercial viability. An alternative strategy is to directly use C_1 carbon sources including methane, methanol, formaldehyde, formate and of course, CO_2 itself. The former compounds can be derived from CO_2 using electrocatalytic means, producing a relatively homogeneous feedstock stream which simplifies process development. This approach does require an abundant supply of green electricity, which is currently in high demand globally and competes with chemical drop-in fuel synthesis.

The direct use of CO_2 is challenging due to the low concentration of atmospheric CO_2, the resulting low efficiency of biological CO_2-fixing modules, and the high costs involved in concentrating CO_2. Even so, advances in the metabolic engineering of autotrophic organisms have resulted in improvements of growth on C_1 compounds (Della Valle et al., 2024). Other methods have also been explored to circumvent these issues, including the use of concentrated streams of waste CO_2. Most notably, gas fermentation technology using *Clostridium* uses CO_2-rich off-gas from steel mills to produce ethanol (Köpke et al., 2010).

Other work using this gas fermentation technology has demonstrated that ABE (acetone, butanol, ethanol) fermentation is possible by using concentrated CO_2 streams (Liew et al., 2022).

An alternative to direct fixation of CO_2 is its conversion into formate and methanol by electrocatalysis, and subsequent use of these compounds as C_1 carbon sources. Formate is a highly soluble carboxylic acid, while methanol is liquid at room temperature and pressure, considerably simplifying storage and transport compared to CO_2. The concept of a 'formate bioeconomy' has been explored, highlighting the logistical benefits of using formate as a C_1 carbon source instead of CO_2 directly (Yishai et al., 2016). Formate and methanol can be utilised for biomass formation in *E. coli* through the reductive glycine pathway (Claassens et al., 2022; S. Kim et al., 2020). Alternatively, acetate and formate can be used in combination, with formate acting as a source of reducing power, or as a carbon source via reversed pyruvate formate-lyase activity (Zelcbuch et al., 2016).

2.6 Enzyme engineering to enhance alkane biosynthesis

As covered in section 2.4, the importance of highly specific, efficient, and robust enzymes at the terminal nodes of metabolic pathways cannot be understated. These final product-forming enzymes dictate bioprocess performance in terms of the final distribution of synthesised products, as well as improving process thermodynamic by consuming the substrates produced in prior metabolic reactions at higher rates. Enzyme engineering methods can roughly be grouped into either rational- or random methods. The former involves the targeted mutation of one or several residues based on structural- or phylogenetic information, resulting in relatively small pools of variants. Random protein engineering methods introduce diversity at random in chosen regions of the enzyme or over the entire sequence. Due to the much higher number of variants created using this approach, screening or selection procedures must be established to identify desirable variants.

In section 2.3 and 2.7.1 we expand on each hydrocarbon production pathway and their properties. However, in terms of enzyme engineering ADO represents the most comprehensively studied and engineered enzyme involved in alkane biosynthesis. The crystal structure of ADO from *Prochlorococcus marinus* MIT9313 was already solved by the Joint Centre for Structural Genomics (JCSG) before an enzymatic function was even assigned to. Many ADO structures have been determined since, opening the door for rational protein engineering attempts (Buer et al., 2014; Jia et al., 2014; Khara et al., 2013).

Most studies rationally mutating ADO used the available structural and phylogenetic information to probe interesting regions of the enzyme, such as (un)conserved regions, the substrate binding channel, and surface residues (Bao et al., 2016; Khara et al., 2013; Kudo et al., 2019; Shakeel et al., 2018; Wang et al., 2017). As is often the case with rational engineering approaches, the structural information alone is not enough to predict whether mutations will be beneficial. As such, most point mutations introduced into ADO have resulted in wildtype-level or poorer performance.

Several highly beneficial point mutations have been uncovered using rational engineering. In terms of improving k_{cat}^{app} and product yield, some of the most productive mutations were W178R in *S. elongatus* PCC7942 ADO, triple mutant Q49H/N123H/F150Y in *S. elongatus* PCC7942 ADO and L148R in *Synechocystis sp.* PCC6803 ADO. It is also clear from these studies that mutating ADO (surface) residues can result in improvements in enzyme solubility and intracellular protein abundance, implying that there is significant room for improvement in terms of optimizing ADO solubility and stability. In fact, a study comparing enzyme characteristics of ADO homologs found that ADO from *Gloeobacter violaceus* PCC7421 was 5-25 times more abundant intracellularly compared to a variety of homologs, although its low specific activity compensated for its abundance (Kudo et al., 2019). Using the available sequence- and catalytic information of ADO homologs to rationally improve desired characteristics is still a relatively unexplored area for ADO engineering. No published works

have attempted to consolidate single beneficial point mutations from different studies into one ADO variant, an approach which may yield more improved variants.

A critical characteristic for alkane biosynthesis enzymes is their substrate specificity. By introducing amino acids with bulky sidechains into the substrate channel, the specificity toward C_{14}, C_{16} and C_{18} aldehydes could be strongly reduced, whereas specificity for shorter chain C_4 to C_{12} aldehydes was improved (Khara et al., 2013). Unfortunately, shifting the substate specificity towards shorter chain lengths also decreased overall activity considerably. There is evidence that short chain molecules may partially inhibit ADO's activity, since C_8-C_{12} aldehydes were reported to inhibit ADO activity above certain concentrations (Andre et al., 2013). Considering the natural substrate specificity of most cyanobacterial ADOs and the concomitant structural features associated with this specificity, it could be that even with considerable protein engineering efforts, high activity and specificity for short chain length substrates will not be possible.

Besides directly engineering the ADO protein, several other techniques have been used to improve alkane biosynthesis including designing different protein fusion constructs and protein co-localization by scaffolding. As described in a previous section, ADO has been shown to interact with AAR presumably to shuttle fatty aldehyde substrate into ADO's substrate binding channel (Chang et al., 2020; Gao et al., 2020; Warui et al., 2015). A protein fusion construct of ADO with C-terminal linker and AAR resulted in a 4.8-fold enhanced alkane titre compared to expressing the enzymes individually, which further supports the hypothesis that interaction of these two enzymes is important for alkane biosynthesis. Even more effective however was the colocalization of AAR and ADO using a zinc-finger DNA scaffolding system. A 3:1 ratio of ADO:AAR was found to be most effective for alkane biosynthesis, approximately doubling the achieved titre of the ADO-AAR fusion protein (Rahman et al., 2014).

Besides efficient substrate delivery, ADO activity requires reducing power which is thought to be delivered by an endogenous ferredoxin and ferredoxin reductase (Fd/FNR) pair (Rajakovich et al., 2015). By fusing ADO from *S. elongatus* PCC7942 with a FNR/Fd pair,

k_{cat}/K_m values could be improved approximately 3-fold compared to the reaction with the enzymes in solution separately (Wang et al., 2014).

Finally, H_2O_2 is known to be a potent inhibitor of ADO activity, and its removal from *in vitro* reactions through the addition of catalase has been shown to significantly improve ADO turnover number. In addition, an ADO-catalase fusion construct expressed in *E. coli* together with AAR was shown to improve alkane yield per ADO protein approximately 4-fold. This provided evidence that intracellular inhibition of ADO was occurring due to H_2O_2 presence (Andre et al., 2013).

Many of these studies have fused proteins of interest to the C-terminus of ADO (Rahman et al., 2014; Wang et al., 2014), or have included 6-His (histidine) tags on the C-terminus of ADO (Andre et al., 2013). A recent study has found that the C-terminus of *S. elongatus* PCC7942 ADO is a target for native *E. coli* proteases, and the (partial) truncation of this region or the addition of 6-His tag C-terminal to this region was found to significantly reduce ADO protein degradation *in vivo* (Liu et al., 2020). This 'protecting' effect of a C-terminal 6-His tag may also extend to other C-terminally fused constructs. This may have implications regarding whether the observed *in vivo* improvements of fusion constructs were due to the constructs themselves conferring activity benefits, or simply due to reduced ADO degradation by endogenous proteases.

Besides ADO, the stability of AAR has also been probed in recent works. One of these reported a comprehensive biophysical study of AAR where the authors found that AAR exhibited only marginal stability. Certain regions could be stabilised by addition of KCl, especially those regions known to interact with ADO (Gao et al., 2020; Sharma et al., 2021). This stabilizing effect of KCl had already been observed in much earlier work by Lin & coworkers, where they found that the addition of KCl, but not NaCl, significantly improved octadecanal consumption rates by AAR (F. Lin et al., 2013). The addition of an N-terminal 12-His tag to AAR was also recently found to considerably improve overall alkane production of an AAR-ADO pathway by improving its solubility (Han et al., 2023). Taken together, these

results suggested that limited AAR stability may also be affecting overall alkane biosynthesis pathway performance, and that the addition of positively charged residues (12-His tag) or ions (KCl) can benefit alkane biosynthesis performance.

Soon after the discovery and characterisation of the ADO enzyme from *S. elongatus* PCC7942 its crystal structure was generated, providing researchers the opportunity to design rationally guided mutant enzymes, a process which has now been repeated for many ADO homologs. These attempts have led to significant but modest (~5 times) improvements in solubility, stability, and specific catalytic activities of ADO over the past 14 years. It has become apparent that traditional rational protein engineering efforts have not sufficed in generating ADO variants with vastly improved activities. When faced with unknowns regarding the sequence – structure – function relationship of enzymes, rational engineering efforts become inefficient and unlikely to yield beneficial mutations. A preferred method in this case is to apply random protein engineering and directed evolution to blindly follow an evolutionary path towards the desired enzyme activity (Arnold, 1998). We expand on this technique and its application for hydrocarbon biosynthesis enzymes in the following section.

2.7 Harnessing evolution for engineering microbial hydrocarbon biosynthesis

2.7.1 Opportunities for the directed evolution of hydrocarbon-producing enzymes

This manuscript is in preparation for submission.

Opportunities for the directed evolution of hydrocarbon-producing enzymes

Abstract

Enzymes capable of catalysing the production of hydrocarbons hold promise for sustainable fuel synthesis. However, the native activities of these enzymes are insufficient for their exploitation in commercial bioprocesses. Enzyme engineering approaches such as directed evolution (DE) can be used to significantly improve the properties of enzymes towards desirable standards. The efficacy of DE depends largely on the method's capacity to detect target enzyme activity accurately and sensitively, preferably at a high throughput. While DE is a widely used and proven approach, it has remained underutilised for the engineering of hydrocarbon-producing enzymes. This can be partially attributed to the physical- and chemical properties of the target products, aliphatic hydrocarbons, which can be insoluble, gaseous, and are chemically relatively inert in physiological conditions. Detection of these molecules *in vivo* presents several unique challenges, as does coupling their presence to cell fitness. In this review, we summarise the available directed evolution methods for engineering hydrocarbon-producing enzymes, including both screening- and selection procedures. Finally, we provide future perspectives on the development of the field.

Introduction

The rapidly changing climate demands global attention for the reduction of greenhouse gas emissions. Around 20-25% of these emissions originate from the unsustainable combustion of hydrocarbon fuels in the personal- and commercial transport sectors (IEA, 2023). Over the past decades, transportation fuels have been replaced to some degree with bioethanol and biodiesel. Although these positively contribute to reducing net emissions, the chemical

composition of these fuels is not identical to fossil hydrocarbons. They must therefore be blended with conventional fuels before use in most internal combustion engines (ICEs) which reduces their scalability, resulting in the so-called “blend wall”. In addition, concerns about arable land requirements for biofuel production reduce their attractiveness as fossil fuel alternatives.

The biological production of aliphatic hydrocarbons presents a promising solution to some of the longstanding issues in the biofuel sector. Firstly, the production of chemically identical hydrocarbons (‘drop-in’ fuels) alleviates the blend wall issue, allowing either neat usage or significantly increased blend percentages. Secondly, microbial cell factories can, either natively or through genetic engineering, consume low-cost substrates including various waste streams, CO_2, or substrates derived from CO_2 such as acetate and formate. This can shift bio-based processes from the current mostly crop-derived substrates towards truly sustainable production (Scown & Keasling, 2022).

Although promising, microbes rarely accumulate hydrocarbons at desired titres, rates, and yields (TRY) natively (Kang & Nielsen, 2016). In fact, most widely studied microbes don’t natively produce any useable fuels, requiring at least some initial metabolic engineering to establish a biosynthesis pathway for the desired compound. Some product formation can often be achieved this way, but the resulting TRY metrics are usually far below those required for commercial exploitation. Fortunately, the modern synthetic biologist has a wide array of tools available to them for targeted alteration of cell metabolism, allowing them to apply engineering principles to cells to improve production. Although some standard techniques can be used to improve TRY, e.g. reducing side product formation, redirecting carbon flux towards product, and balancing cofactor availability, sometimes the (terminal) enzymes involved in product synthesis remain limiting.

In these cases, enzyme engineering approaches can be used to alter enzyme solubility, activity, (thermo)stability or substrate specificity, leading to improvements in production. Rational or semi-rational approaches can be used to direct mutagenesis towards certain

residues, but this requires a comprehensive understanding of the enzyme's catalytic mechanism, usually achieved by determining a crystal structure of the enzyme in question. Even with this information, rational enzyme engineering can lead to disappointing results when designed alterations have unexpected effects, such as a radical loss in stability or activity due to unforeseen interactions between residues. In these cases, a preferred approach is directed evolution (DE), a method using iterative rounds of mutagenesis and screening or selection to identify enzymes with desirable traits (Arnold, 1998).

To use a DE approach, the desired activity of the enzyme must be measurable (screening) or contribute to fitness (selection) within a population of cells carrying enzyme variants (Packer & Liu, 2015). This requires the development of a robust method to screen or select for the desirable activity, which in itself is the most challenging aspect of DE. However, compared to other classes of enzyme products, establishing a DE method for hydrocarbon production faces some unique challenges that we discuss in this work.

In this review, we summarise the development and use of DE approaches for hydrocarbon-producing pathways and enzymes and provide perspectives and suggestions for their further development. This summary may help accelerate the use of DE methods to establish commercially viable sustainable fuel production.

Hydrocarbon biosynthesis

Various hydrocarbon-producing enzymes have been identified in nature. Already since the 1960's waxes, alkanes and alkenes have been detected in insect-, plant- and cyanobacterial extracts (Baker et al., 1963; Han et al., 1968; Oró et al., 1965). These compounds have disparate functions depending on the respective producer, ranging from protection against desiccation in insects and plants to promoting cell wall flexibility and salt tolerance in cyanobacteria (Berla et al., 2015; Holze et al., 2020; Jetter & Kunst, 2008; Lea-Smith et al., 2016; Yamamori et al., 2018). Research efforts over the last two decades have led to the initial identification, characterization, and application of several of the metabolic pathways

responsible for hydrocarbon biosynthesis. For the purposes of this review, we focus on bacterial hydrocarbon synthesis enzymes since many of these have successfully been heterologously expressed and their catalytic activity characterised. In addition, we use the term 'hydrocarbons' to refer to aliphatic hydrocarbons (alkanes/alkenes), since these compounds are abundant in natural gas, gasoline, kerosene, and diesel, hence they have significant potential as drop-in fuels.

Terminal alkene biosynthesis

The first isolated enzyme responsible for terminal alkene biosynthesis was found to be a cytochrome P450 from the bacterium *Jeotgalicoccus* sp., dubbed $OleT_{JE}$, catalysing the decarboxylation of fatty acids (Rude et al., 2011). Its substrate specificity was found to be extremely broad, enabling the production of short-, medium- and long chain C_{n-1} terminal alkenes ($C_{4:1}$-$C_{20:1}$) from the respective fatty acids with concomitant release of CO_2 (Dennig et al., 2015). The $OleT_{JE}$ enzyme also displayed flexibility in terms of its electron source, which can either be from H_2O_2 without the need for oxygen or via a reduced electron carrier in the presence of oxygen (Y. Liu et al., 2014).

Another terminal alkene biosynthetic pathway was discovered in *Synechococcus* sp. PCC7002, comprising a multidomain protein with homology to type I polyketide synthase named olefin synthase (*ols*) (Mendez-Perez et al., 2011). Characterization of Ols activity uncovered that it was responsible for C_{19} terminal alkene synthesis. Since this cyanobacterium did not accumulate fatty acids over C_{18} in length, the synthesis of C_{19} compounds was expected to occur through additional acyltransferase activity adding two carbons from malonyl-CoA, with a final decarboxylation step to yield C_{19} product.

A reverse genetics approach was used to identify the enzymes responsible for 1-undecene biosynthesis in *Pseudomonas*. This led to the discovery of an iron oxidase UndA, which utilised O_2 and a reducing agent to decarboxylate fatty acids to corresponding terminal alkenes (Rui et al., 2014). Further study later revealed it to constitute a new class of diiron

decarboxylase (Manley et al., 2019). The UndA enzyme could convert C_{10} – C_{14} fatty acids to their corresponding C_{n-1} terminal alkenes but was unable to utilise substrates with chain lengths ≤C_8 and ≥C_{16}, indicating a narrow substrate specificity range (Rui et al., 2014). Upon further study another 1-undecene producing enzyme was found by the same group, dubbed UndB, which also catalysed the oxidative decarboxylation of fatty acids. This enzyme was found to have a broader substrate specificity than UndA, ranging from C_6 – C_{17} fatty acids, although with a preference for C_{10}-C_{14} (Rui et al., 2015). In contrast to UndA, the UndB enzyme is membrane-bound, which has hindered the biochemical analysis of its catalytic mode of action. However, it displays homology to other diiron desaturases, making it likely that UndB also contains a diiron core. The biosynthesis of 1-alkenes in yeast was achieved using UndB (Zhou et al., 2018), and it was found capable of catalysing the decarboxylation of dicarboxylic acids (S. Wu et al., 2019).

Compared to $OleT_{JE}$ and UndA, the membrane-bound UndB enzyme results in higher production of 1-alkenes when expressed *in vivo* both in *E. coli* and *S. cerevisiae* (K. Liu & Li, 2020). Most engineering efforts have focussed on redox partner optimization, especially for $OleT_{JE}$ (Dennig et al., 2015; Fang et al., 2017; Y. Liu et al., 2014). One recently published pre-print describes the use of DE to evolve UndB (Ward et al., 2023). The authors described using *in silico* guided residue selection and alanine scanning to identify UndB variants with improved substrate specificity towards C_9 – C_{14} dicarboxylic acids, achieving approximately 5-fold improvements in conversion of C_{10} dicarboxylic acid to 1,7-octadiene. Another study devised a novel screening method for the directed evolution of $OleT_{JE}$, which we have detailed in a later section (H. Xu et al., 2020). Apart from these works, no other studies report using DE for the engineering of $OleT_{JE}$, Ols, UndA or UndB.

Internal alkene biosynthesis

The biosynthesis of internal alkenes occurs via the head-to-head condensation of fatty acids. The *oleABCD* genes from *Micrococcus luteus* were found to be responsible for the biosynthesis of $C_{27:3}$ and $C_{29:3}$ alkenes, through the consumption of two fatty acyl-CoA and the

release of one CO_2 (Beller et al., 2010). Soon after, more *ole*-type operons were discovered in a variety of bacteria. Some of these were characterised revealing that a wide array of olefin products can be generated through this pathway, ranging from C_{24} – C_{31} in length and with up to 9 unsaturated bonds (Sukovich et al., 2010). The *oleABCD* enzymes have not yet been the subject of enzyme engineering efforts by directed evolution.

Alkane biosynthesis

Saturated alkanes are biosynthesised in plants, most notably by ECERIFERUM 1 and 3 (CER1/CER3) in *Arabidopsis thaliana* (Aarts et al., 1995; Bernard et al., 2012; Bourdenx et al., 2011). In insects, an NADPH-dependent cytochrome P450 from the CYP4G family was shown to be a highly conserved enzyme involved in alkane biosynthesis (Y et al., 2012). However, both plant- and insect alkane biosynthesis enzymes have not been used extensively by researchers interested in biofuel production. Owing to their native eukaryotic environments, plant- and insect enzymes are often difficult to express in widely used model organisms such as *E. coli*. Additionally, the carbon chain length specificity of plant- and insect alkane biosynthesis enzymes (C_{20}-C_{34}) exceeds the range required for drop-in kerosene- or petroleum fuels, although some products may be suitable as diesel replacements. We expect that studies looking to use these enzymes for biobased production will first have to address the issues involved in establishing sufficient soluble and functional enzyme. Instead, the vast majority of research for alkane biosynthesis has focussed on using and engineering cyanobacterial alkane biosynthesis enzymes.

In 2010, a landmark study by Schirmer *et al*. reported the identification and characterization of enzymes responsible for alkane biosynthesis in the cyanobacterium *Synechococcus elongatus* PCC7942 (Schirmer et al., 2010). The proposed pathway consisted of an Acyl-ACP Reductase (AAR) and a terminal enzyme initially designated an aldehyde decarbonylase, owing to its supposed side-product being carbon monoxide. This was corrected in 2011, when the side-product of its reaction was found to be formate rather than CO, resulting in a renaming of the enzyme to aldehyde deformylating oxygenase (ADO) (N. Li et al., 2012; Warui et al.,

2011). The AAR enzyme produces free fatty aldehydes using NADPH as an electron donor and fatty acyl-ACP as a preferred substrate, although it can also utilise acyl-CoA. The resulting fatty aldehyde product is then deformylated by ADO to produce a C_{n-1} alkane and formate as a by-product. The terminal ADO enzyme requires input of four electrons to perform the deformylation reaction, which are preferentially delivered by a ferredoxin (Fd)/ferredoxin reductase system (FNR) (J. Zhang et al., 2013). A Fd/FNR pair is therefore hypothesised to be the endogenous reducing system for ADO in cyanobacteria. Native ADO displays a wide substrate specificity ranging from C_4 – C_{18} aldehydes, and several studies have successfully improved its specific activity towards short-chain substrates (Amer et al., 2020; Khara et al., 2013). Having been a target of metabolic- and enzyme engineers for over a decade, maximum titres of the heterologously expressed AAR/ADO pathway have currently reached the g/L level in *E. coli* (Cao et al., 2016; Fatma et al., 2018). However, production rates and yields still remain below commercial requirements, and most studies report ADO to be significantly rate limiting, making it an attractive enzyme engineering target. Only a handful of studies have used DE approaches to engineer ADO, which are described in later sections.

A more recently discovered alkane- and alkene-forming enzyme catalyses the light-driven decarboxylation of fatty acids or fatty acyl-CoAs (J. Li et al., 2020; Sorigué et al., 2016, 2017). This activity was exhibited by the *Chlorella variabilis* fatty acid photodecarboxylase (*cv*FAP), which displays a substrate specificity of C_{12}-C_{18} fatty acids with a preference for C_{16} and C_{17} and requires blue light (~400-500 nm) for its activity. Rational engineering of the enzyme has improved its specificity towards short-chain product formation, allowing the light-driven generation of alkanes as short as ethane (W. Xu et al., 2022). Only recently has the *cv*FAP enzyme first been engineered via DE, resulting in a comprehensive study where the enzyme was repurposed to enable highly selective decarboxylative radical cyclization (Ju et al., 2024). We note that although a novel in-house screening pipeline was devised and used successfully in this work, product identification and quantification occurred via analytical techniques such

as GC-MS or HPLC. As a result, the throughput of such an approach would still be constrained and require relatively small variant library sizes.

Challenges facing directed evolution of hydrocarbon-producing enzymes

Establishing a DE protocol varies in difficulty, depending mostly on the physical- and chemical properties of the target product, and the extent to which the target enzyme has been characterised in terms of its activity and expression in the host cell. The detection and accurate quantification of a target product for DE is subject to several requirements (Table 2.1). Hydrocarbons satisfy few or none of these requirements, depending on their specific chemical composition. Hence, establishing a DE protocol for hydrocarbon synthesis is distinctly challenging.

Table 2.1. Properties of hydrocarbons in the context of directed evolution

Compound property	Ideal compound for DE*	Hydrocarbons
Solubility	**High** Product is ideally highly soluble to enable interactions with proteins and/or other chemicals.	**Low** Often insoluble in aqueous environments, can be gaseous or solid depending on chain length, tends to localise to the cell membrane or aggregate due to hydrophobicity.
Compartmentalization	**High** No extracellular leakage of the product. In case of cell lysis, released product cannot be taken up by living cells.	**Low** Short chain hydrocarbons are volatile, medium- and long chain hydrocarbons can passively diffuse through cell membranes.
Interaction between compound and host metabolism	**Depending** For screening purposes, interaction with host metabolism should be low. For selection, coupling product formation to host metabolism can be a useful property.	**Low** For most model organisms used as screening or selection platforms hydrocarbons are not a native metabolite.
Toxicity	**Low** Ideally, the target product should not be toxic.	**Medium** Short and medium-chain hydrocarbons can be toxic. Longer chain hydrocarbons exhibit low toxicity.

**Assuming a DE system is used where intracellular product concentration correlates with product detection response (screening) or growth (rate) (selection).*

With these challenges in mind, attempts have been made to establish screening and selection protocols for hydrocarbon synthesis. The majority of these focus on detection and reporting of hydrocarbon presence through a biosensor, although approaches involving selection instead

of screening are starting to be adopted. We summarise the development of both these approaches for the DE of hydrocarbon biosynthesis enzymes in the following sections.

Hydrocarbon screening methods

Transcription factor-based reporter systems

Many DE approaches are based on measuring the abundance of a product of interest via a reporter system. By pinpointing genes that play a role in either catabolism or anabolism of the desired product, genetic elements involved in transcriptional responses to its presence can be identified and characterised. Following on, these genetic elements can be used in synthetic constructs to drive reporter gene expression. For example, such a construct could consist of a hydrocarbon-sensitive transcription factor (TF) and a green fluorescent protein (GFP) of which the expression is controlled by the same TF (Figure 2.2A).

Hydrocarbon-degrading microbes have been an abundant source of useful hydrocarbon-sensitive TFs. For example, *Pseudomonas oleovorans* carries the alkS/P_{alkB} system, while *Acinetobacter baylyi* regulates alkane degradation through the alkR/P_{alkM} expression system (Ratajczak et al., 1998; van Beilen et al., 1994). Both of these systems have been extensively characterised and repurposed for hydrocarbon detection. These systems have been used to detect hydrocarbon contamination in water using fluorescent- or luminescent outputs (Sticher et al., 1997; D. Zhang et al., 2012), and to detect the uptake of alkanes (Grant et al., 2014). As the biosynthetic pathways for hydrocarbon production were discovered and used for bioproduction, these hydrocarbon biosensors were further developed specifically with the aim of quantifying hydrocarbon production *in vivo* (see Table 2.2).

A recent study reported the first use of alkane-sensitive biosensors to engineer ADO (Chen et al., 2023). This work included the engineering of the alkane-responsive TF AlkS to increase its sensitivity to longer-chain alkanes. Four rounds of random mutagenesis were performed using bidirectional FACS screening, where cells displaying low fluorescence in the absence of alkanes were isolated first, followed by the addition of alkanes and isolation of the top 1%

of fluorescent cells. After optimization of the biosensor, the authors randomly mutagenised the AAR and ADO genes over four rounds and identified isolates displaying a 13-fold improvement in alkane titre. This work highlighted the importance of engineering the hydrocarbon-detecting TF before attempting a DE approach. The successful engineering of AAR-ADO reported in this study highlights the potential of this biosensor approach.

Table 2.2. Transcription factor-based reporter systems for aliphatic hydrocarbon detection

Detector ─ Reporter	Specificity	Sensitivity	Study findings	Reference
P_{alkS}/alkR ─ P_{alkM}-GFP	C_{13} – C_{17}	1-3 mg/g CDW	Improved sensor performance, combined two alkane response elements	(W. Wu et al., 2015)
P_{alkB}/alkS* ─ P_{alkS}-eGFP	C_{10} – C_{17}	50-500 mg/L	Engineered AlkS for improved response range, DE of AAR and ADO	(Chen et al., 2023)
P_Lλ/alkS* ─ P_{alkS}-GFP	C_5 – C_{10}	Not indicated	Engineered AlkS for short-chain alkane sensitivity	(Reed et al., 2012)
P_{alkRM}/alkR ─ P_{alkM}-GFP *and* P-luxAB	C_{12} – C_{18}	40 mM alkanes, no range given	Monitoring aldehyde and alkane flux simultaneously	(Lehtinen et al., 2017, 2018)

Chemical screening methods

An alternative to TF-based hydrocarbon screening is the use of chemicals that detect the activity of the target enzyme (Figure 2.2B). One such approach was used in the DE of $OleT_{JE}$ using a colourimetric screen for H_2O_2 consumption (H. Xu et al., 2020). The $OleT_{JE}$ enzyme can use H_2O_2 as sole electron- and oxygen source to catalyse fatty acid decarboxylation, and the authors exploited this principle by expressing $OleT_{JE}$ mutant libraries in catalase-deficient *E. coli* to identify H_2O_2-consuming variants. This approach revealed several variants with improved activity, chemoselectivity, and altered substrate specificities.

Advantages and limitations of screening methods

An advantage of TF-based detection is that a response is generated based on direct interaction with the target molecule. It is also possible to swap out one system for another, for example to engineer specificity towards a certain chain length. Indeed, theoretically hydrocarbon-responsive TFs can be engineered towards specificity for a wide range of chain lengths. It would therefore be possible to construct a library of variants with certain specificities, enabling the swapping of detector modules to suit the needs of the DE campaign.

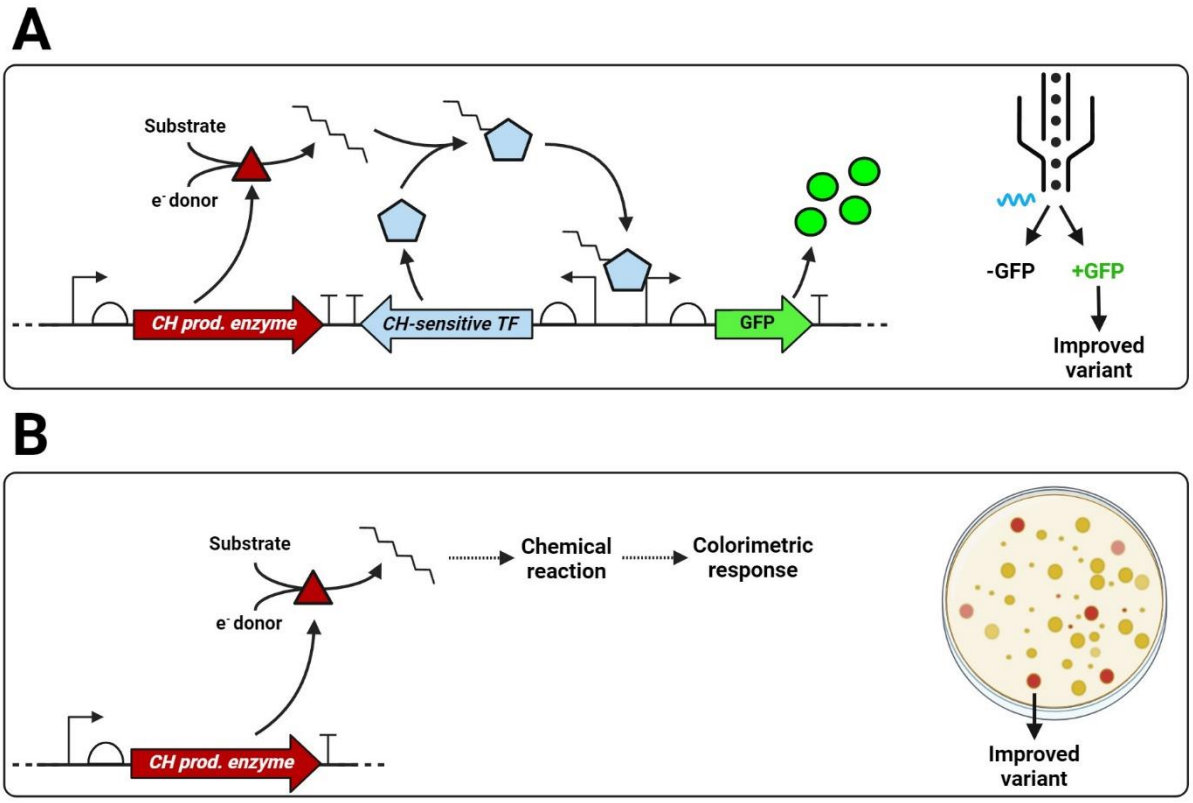


Figure 2.2. Screening methods for hydrocarbon biosynthesis enzymes. (A) Schematic representation of a synthetic genetic construct consisting of a hydrocarbon production gene (CH prod. enzyme), hydrocarbon-detecting transcription factor (CH-sensitive TF) and a TF-responsive element, in this case a GFP reporter gene. The expression of GFP is linked to hydrocarbon presence, and GFP-expressing cells can be sorted by a FACS procedure resulting in enrichment for improved hydrocarbon biosynthesis enzymes. **(B)** A chemical reaction with either the product (hydrocarbon) or substrate(s) can be coupled to a colourimetric response to identify isolates with (improved) activity. Created with Biorender.com.

However, this approach also has several limitations. Firstly, although interaction with the target compound is direct, any chosen TF will not give an equal transcriptional response to all compounds. The specificity and sensitivity of the TF therefore defines the output of the DE

campaign. It is likely that TFs will have to be engineered to display the required response prior to their use in DE studies, as exemplified in recent work (Chen et al., 2023). In addition, as already discussed before, most hydrocarbons can freely diffuse into and out of the cell. This presents an issue for hydrocarbon detection-based screening methods especially when using longer cultivation times, as non-producing cells can take up exogenous hydrocarbons, leading to increased false positive rates. This would require either a) spatial separation of variants, reducing throughput, b) rapid detection of improved variants before product leakage occurs or c), the use of counterselection methods to identify and remove cheating cell populations.

As for chemical screening methods, a considerable disadvantage is its low throughput since variants must usually be physically separated to discern differences in staining. This limits many of these approaches to agar plate-based methods with a maximum throughput of ~10^3-10^4 variants/day, which could still be useful for screening smaller libraries (e.g. smart libraries or site saturation mutagenesis libraries). In general however, we expect that DE approaches will move away from agar plate-based screening and towards other high-throughput methods.

The use of the aforementioned biosensors depends on the availability of specialist equipment to perform high-throughput single-cell quantification of fluorescence and sorting. If specialist equipment is not available, screening throughput has to be lowered several orders of magnitude to allow manual microtiter- or agar plate screening. Increasing the throughput of screening methods therefore also increases procedure complexity and costs, lowering accessibility. Since neutral or detrimental mutations are also far more common than desirable ones, enzyme screening campaigns inevitably result in the observation of mostly isolates of low interest. Whilst this could be valuable for gathering information about the enzyme's sequence-structure-function relationships, especially in light of AI-based rational mutagenesis approaches (Yang et al., 2019), it still means that most resources are spent on observing non- or poorly functioning variants.

Growth-coupled selection methods

Another DE strategy involves coupling the activity of the target enzyme to cell fitness. Enzymes exhibiting increased activity are then selected for within a population, exploiting the principles of natural selection to enrich a population with cells carrying desirable enzyme variants. Selection methods are often based on the target enzyme activity rescuing a natural- or engineered metabolic deficiency. When growth is dependent on the activity of an enzyme the enzyme is deemed to be growth-coupled (for a comprehensive overview of growth coupling principles see (Alter & Ebert, 2019). Growth-coupled approaches offer a low-cost and procedurally simple method for enzyme engineering. As growth (rate) is the output metric, populations of variants can be enriched for those conferring growth benefits using selective serial passages of the population, or by selective cultivation in turbidostats. The throughput of such approaches is usually only limited by the transformation efficiency of the host, which can be in the range of 10^8 – 10^{10} CFU/µg DNA for *E. coli,* or by the target's mutation rate in the case of *in vivo* mutagenesis. Since all successfully transformed hosts are simultaneously assessed for their fitness in selective conditions, the throughput of growth-coupled selection is amongst the highest of all existing enzyme engineering approaches, whilst only requiring minimal resources. A caveat to this is that the success rate of identifying a beneficial variant from a pool of competing variants can be limited by clonal interference (De Visser & Rozen, 2006). Besides competition between lineages, successful use of growth-coupled selection requires that growth benefits conferred by a certain enzyme variant are substantial enough to detect on the chosen experimental timescale. Finally, the composition of populations expressing enzyme variants can be tracked by (deep) sequencing, allowing researchers to determine whether certain variants are increasingly represented over time, but comprehensive detection of these changes in large variant pools requires a higher sequencing depth than with smaller libraries.

In addition, DE by selection is only as good as the imposed selection pressure and its stringency (Arnold, 1998). It is therefore imperative that the selection pressure for such an

approach is chosen carefully and extensively characterised. When growth is coupled to product synthesis, it is important to establish the relationship between output growth (rate) and the target product biosynthesis rate to establish a dynamic range of the selection platform. This information will be useful in avoiding either a maximization of growth rate with unevolved enzyme activity, leaving no selective pressure for further improvement of the enzyme, or a situation where enzyme activity is too low to rescue growth to begin with. To avoid the first issue, concentrations of the growth-coupled enzyme can be reduced by swapping constitutive promoters to lower-strength variants, reducing induction levels of inducible promoters, or by its fusion to a protein degradation tag. However, in most cases growth-coupled schemes are bottlenecked by the initial step of coupling growth solely to enzyme activity.

Hydrocarbon-producing enzymes vary in their activity, but most have catalytic far below those of strongly growth-coupled central metabolic enzymes (Bar-Even et al., 2011). For example, ADO is infamous for its sluggish activity, displaying turnover numbers of ~1 min^{-1} *in vitro* (Marsh & Waugh, 2013). The light-driven fatty acid photodecarboxylase *Cv*FAP was shown to be more efficient but still below the median turnover number of central- and secondary metabolic enzymes, achieving a k_{cat} of ~0.9 s^{-1} (Sorigué et al., 2017). Ensuring these relatively inefficient enzymes can rescue growth therefore requires their coupling to sensitive growth-rescue mechanisms, such as essential but low-flux metabolic pathways. Growth can also be made co-dependent instead of fully reliant on target enzyme activity, by supplying small amounts of the target product to help first establish growth. As the enzyme evolves towards higher activity this supplementation can be reduced until cell growth is fully dependent on the target activity. This approach of ‘weaning’ the cell off a rescue metabolite requires a comprehensive characterization of the relevant dynamic range.

As discussed earlier, hydrocarbons are a challenging class of molecules to target for DE approaches. Engineering cells to be dependent on hydrocarbon-producing enzymes therefore requires some creativity when designing appropriate selection methods, which we have compiled below.

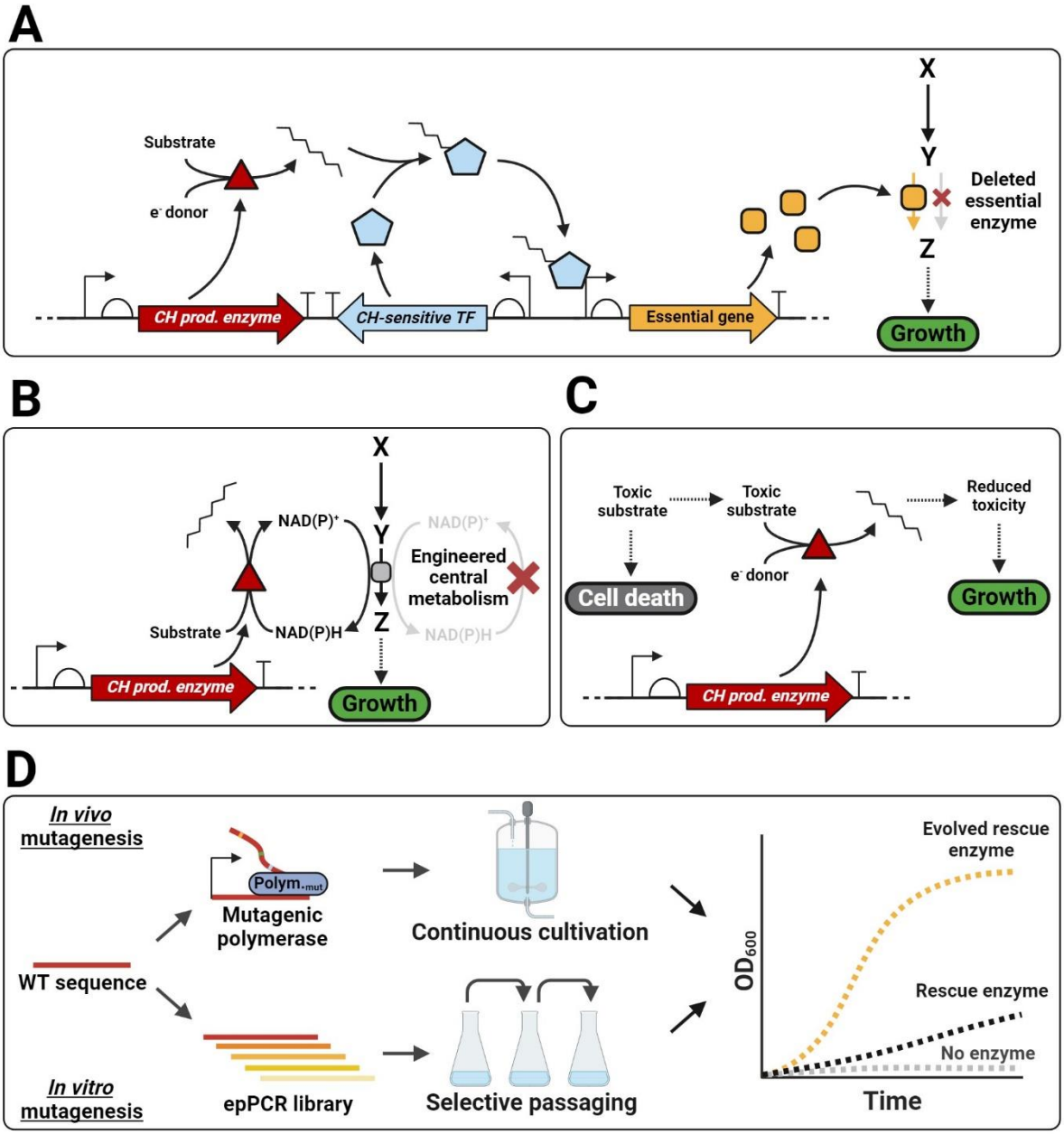


Figure 2.3. Selection methods for hydrocarbon-producing enzymes. (A) Schematic representation of a synthetic genetic construct consisting of a hydrocarbon production gene (CH prod. enzyme), hydrocarbon-detecting transcription factor (CH-sensitive TF) and a TF-responsive element. The responsive element consists of an essential gene to enable hydrocarbon-dependent essential gene expression. **(B)** Schematic of the principle of cofactor-based growth coupling. Microbial selection platform can be engineered to exhibit a deficiency in the oxidation or reduction of redox cofactors. These microbes are unable to grow unless a rescue reaction is introduced restoring redox cofactor cycling, in this example through the NAD(P)H-dependent reduction of substrate to hydrocarbon. **(C)** A toxic concentration of substrate can be used arrest growth or induce cell death unless the substrate is converted to the target hydrocarbon compound, resulting in growth and selection for enzyme variants with higher activity. **(D)** Evolution using these systems can be performed *in vivo* using continuous mutagenesis tools and continuous cultivation (top). Alternatively, more traditional methods such as error-prone PCR (epPCR) can be used with selection by serial passaging (bottom). Target enzyme activity is directly coupled to growth, so evolved enzymes confer a higher growth rate to the selection strain. Created with Biorender.com.

Transcription factor-based growth rescue

Hydrocarbon-sensitive transcription factors (TFs) have been widely used in the context of enzyme screening, as well as engineered to improve their performance. These hydrocarbon-sensitive TFs can be viewed as modular hydrocarbon detection units. They do not necessarily have to be linked to expression of a reporter molecule, but instead can also be employed to drive expression of genes essential for growth. This way, the functionality of these TF systems can be transformed from that of a screening method to a selection method (Figure 2.3A).

A prior study detailed the use of TF-based detection of small molecules to establish growth-coupled systems (Dietrich et al., 2013). The authors coupled a TF sensitive to 1-butanol to the expression of tetracycline transporter gene *tetA*. The resulting strain could grow in tetracycline-containing medium only when supplemented with 1-butanol, or when 1-butanol was biosynthesised *in vivo*. This system was then used to engineer the RBS sequences of a mixed alcohol biosynthesis pathway. Although this work did not target aliphatic hydrocarbon synthesis, the underlying principle of the work demonstrates its viability.

In terms of alkane/alkene biosynthesis, one such TF-based selection approach has been established in *Acinetobacter baylyi*, using its native alkane-sensitive alkR/P_{alkM} transcriptional regulator and promoter system (Kennerley, 2022). The *A. baylyi* strain was first made auxotrophic for the essential amino acid lysine by knocking out 4-hydroxy-tetrahydropicolinate reductase (*dapB*). This gene was then re-integrated into the chromosome but under control of the alkane-sensitive alkR/P_{alkM} system. Growth of this strain could be rescued either by direct supplementation of lysine in the medium, or by addition of mineral oil, inducing expression of the essential gene and thereby complementing growth. Using this system, a mixed population of *A. baylyi* strains transformed with low, medium, and high efficiency alkane biosynthesis plasmids were shown to be enriched for the high alkane production plasmids over time. When assessed separately, strains with high-efficiency alkane biosynthesis plasmids displayed increased growth rates and decreased lag phases compared to those transformed with lower efficiency alkane production plasmids.

This study demonstrated that TF-based hydrocarbon detection could also be used to establish a selection system for their production, as opposed to a screening method. Due to the growth-based selection principle only a selective cultivation of the strain would be required to achieve population enrichment of positive clones, considerably simplifying the process compared to a screening method. This approach does still suffer from some of the same limitations as TF-based screening approaches, including potential cross-feeding of product in the population and the dependence on TF specificity and genetic circuit integrity.

Cofactor-based growth coupling

A relatively new concept in the field of growth-coupled enzyme engineering is to use redox cofactor auxotrophy as a selection method (Nielsen et al., 2023). In this approach, strains unable to grow due to an imposed deficiency in the reduction or oxidation of a certain cofactor pair (e.g. NAD^+/NADH or $NADP^+$/NADPH) can be used to select for enzymes that alleviate this redox imbalance. This method offers a way to couple growth to redox cofactor-requiring enzymes or pathways, without having to use the target product as the selective molecule directly (Figure 2.3B). Since redox cofactors are the selected molecules in this approach, issues caused by low solubility and poor compartmentalization of hydrocarbon products can be avoided.

The majority of bio-based hydrocarbons can be biosynthesised through carbon chain-elongating steps in metabolism, such as fatty acid biosynthesis or reverse beta oxidation. These pathways require high cofactor input in the form of NAD(P)H, making them viable targets for cofactor-based growth-coupling using strains deficient in the cycling of NAD(P)H back to their oxidised counterparts $NADP^+$ or NAD^+. Several hydrocarbon biosynthesis enzymes also require cofactors, such as AAR (NADPH, directly), ADO (NAD(P)H, ferredoxin-mediated) and $OleT_{JE}$ (NAD(P)H, cytochrome P450 reductase). Redox cofactor auxotrophs have already been used to engineer a 3-hydroxy-acyl-CoA reductase (Machado et al., 2012), carboxylic acid reductase (Kramer et al., 2020), P450 cytochrome (Maxel et al., 2020) and a nitroreductase (Sellés Vidal et al., 2021), to name several. The breadth of enzyme activity

targetable by this DE approach makes it a promising technology for engineering hydrocarbon-producing enzymes.

Although this approach has not been used to engineer hydrocarbon production so far, we expect cofactor-based growth coupling to become a more widely used tool in the DE repertoire due to its unique product-independent property.

Substrate toxicity alleviation

Instead of selecting for the biosynthesis of a target product, the consumption of a substrate can also be linked to cell fitness if the substrate is toxic and the product less so. Consumption of the toxic compound by the target enzyme alleviates its toxic effects, either reducing the growth lag phase of the host cell, improving its growth rate, or a combination of the two (Figure 2.3C).

Although this approach has a limited scope, since the target enzyme must have a toxic substrate and a less toxic product, it provides a direct way to link enzyme activity to growth. One study has successfully employed this approach to engineer the aldehyde deformylating oxygenase (ADO). The aldehyde substrate for ADO is toxic to *E. coli* depending on concentration and carbon chain length. The authors exploited this property by adding inhibitory concentrations of aldehydes to *E. coli* cultures, where they found that ADO-expressing cells exhibited improved growth (Y. Liu et al., 2020). After a round of random mutagenesis, the authors identified an *E. coli* protease degradation target on the C-terminus of the protein, which when removed or protected by a His-tag significantly improved the concentration of soluble ADO protein *in vivo*. The authors used an engineered strain deficient in several native aldehyde reductases to ensure aldehyde consumption by the host was limited. This approach therefore lends itself well to toxic substrates that are poorly, or not at all, degraded by the host but can be catalysed to non-toxic products by the target enzyme.

A second relevant study, although not focussing on aliphatic hydrocarbon production, demonstrated that the detoxification of medium-chain fatty acids can be used as a selective

principle (Hu et al., 2020). In this work, the authors engineered carboxylic acid reductases for improved activity against medium-chain fatty acids (MCFA), to generate the concomitant alcohols. One aspect of the work echoes the previous ADO study exploiting aldehyde toxicity, namely that endogenous pathways capable of detoxifying the substrate were removed. For example, deletions of a MCFA exporter and an aldehyde dehydrogenase were performed to ensure accumulation of alcohol product. Considering several hydrocarbon-forming enzymes use fatty acids directly as a substrate, this approach may be worthwhile to explore.

For many hydrocarbon-producing enzymes, their substrate specificity in terms of carbon chain length is a major focal point for enzyme engineering. Substrate toxicity alleviation provides an attractive method to customise selective pressure towards specificity for a certain carbon chain length. Besides its limitation for toxic substrates, a downside of this method is that it can only be used to directly target the enzyme of interest, preventing whole-cell engineering.

Advantages and limitations of hydrocarbon selection methods

The use of selection methods for engineering hydrocarbon enzymes offers a way to perform ultra-high throughput DE with minimal resource requirements or procedural difficulties. These methods are therefore highly attractive for applying selection to large combinatorial libraries, performing continuous mutagenesis, and evolving whole-cell systems towards improved product synthesis (Figure 2.3D).

While the advantages are considerable, growth-coupled selection methods are challenging to establish. Some level of metabolic engineering is often required, and extensive characterization of the resulting metabolic deficiency must be performed before a DE campaign can be initiated. In addition, as mentioned before, the catalytic rate of hydrocarbon biosynthesis enzymes is lower on average than essential enzymes in central metabolism (Bar-Even et al., 2011). Establishing growth-coupling solely to target enzyme activity therefore requires careful planning regarding which essential metabolic node to target, since this defines the dynamic range within which selective pressure will be exerted. Establishing selection

based on essential gene expression may therefore be a useful alternative to the direct metabolic coupling of enzyme activity, since expression strength can also be modulated through other means such as RBS swapping or using multi-copy constructs.

Besides establishing a selection method, ensuring the required selective pressures remain in place during DE procedures can be challenging as growth is often severely inhibited. Escape mutants that disrupt the DE process can therefore arise, requiring either a stringent selection method or a counterselection method to combat cheating populations.

Future perspectives

Directed evolution methods for hydrocarbon-producing enzymes are challenging to establish, often requiring preliminary genetic- and protein engineering to establish high-throughput screening and selection procedures. Much ground has been covered in recent years, but several research areas remain unapproached.

The sensitivity, specificity and robustness of hydrocarbon biosensors has steadily improved over the past decade, with recent work demonstrating its potential for engineering the AAR-ADO pathway (Chen et al., 2023). Despite these improvements, the majority of studies still apply them solely for fluorescent reporter gene expression, even though this requires specialised equipment and cannot easily be used in continuous evolution systems. We envision these biosensors being used in growth-coupled designs, where the detection of hydrocarbons is coupled to the expression of gene (circuits) linked to growth. Although this would require initial strain engineering and characterization, these growth-coupled selection methods have increased throughput and lower resource requirements, potentially accelerating the discovery of improved enzyme/pathway variants.

The use of substrate toxicity as a selection method for aliphatic hydrocarbon production has only been used to engineer the ADO (Y. Liu et al., 2020), but other potential targets for this approach are fatty acid decarboxylating enzymes such as $OleT_{JE}$, UndA, UndB and *Cv*FAP. Short- and medium chain fatty acids are toxic to *E. coli* and *S. cerevisiae* above certain

concentrations depending on their chain length (Hu et al., 2020; Royce et al., 2013), meaning that enzymes catalysing their conversion to less toxic alkanes or alkenes could potentially be engineered using this approach. Efficient production of gaseous hydrocarbons such as propane and butane is also highly sought after, but performing DE using the gasses themselves as the target enzyme product is challenging due to a lack of compartmentalization (assuming that their intracellular concentration determines the DE system's response). Using the alleviation of short-chain fatty acid toxicity as the selective principle may allow engineering of these enzymes with more ease. When using *E. coli* as a host for enzyme selection, exogenous fatty acids can also be easily excluded from cell metabolism by deletion of fatty acid CoA ligase (*fadD*), avoiding their activation to acyl-CoA and subsequent degradation.

The studies focussed on engineering hydrocarbon biosynthesis have mostly reported either host-level metabolic engineering (e.g. carbon flux improvements, redox balancing) or rational mutagenesis and screening specifically targeting only the terminal hydrocarbon synthesis enzymes. However, when hydrocarbons are synthesised from a sole carbon source (i.e. the substrate of the terminal enzyme is not supplemented), some of the DE methods set out in this review could be used to evolve entire metabolic pathways and other cellular processes towards improved hydrocarbon synthesis. Besides improving bioprocess performance using common substrates such as glucose, growth-coupled designs enable the simultaneous engineering of product biosynthesis and substrate catabolism. This can be exploited to evolve cell factories towards optimal assimilation and conversion of novel CO_2-derived carbon sources. *In vivo* continuous targeted hypermutation (Molina et al., 2022) or whole-genome mutagenesis tools (Wang et al., 2009) represent suitable technologies for performing this type of evolutionary engineering towards improved product synthesis. As of yet, *in vivo* mutagenesis coupled with a high-throughput DE approach has not been performed in the context of hydrocarbon biosynthesis.

Conclusion

The field of drop-in hydrocarbon biosynthesis has struggled to establish processes with high titres, rates, and yields. Even though several hydrocarbon-producing pathways have been identified and extensively characterised over the past decade and a half, the activities of the constituent enzymes often remain unsuitable for commercial use. A considerable bottleneck to the improvement of hydrocarbon-producing enzymes and strains is a lack of appropriate screening or selection methods to identify improved variants in a high-throughput manner. Designing directed evolution methods for hydrocarbon synthesis is challenging and requires the creative implementation of multiple facets of synthetic biology. In this review, we have summarised the development of directed evolutions methods for hydrocarbon production, including their advantages and limitations.

The most widely reported directed evolution system is a screening method based on a hydrocarbon-sensitive genetic construct driving expression of a fluorescent reporter protein. Although these systems initially displayed low selectivity and sensitivity, considerable improvements have been made over the years. These efforts have resulted in screening systems that can now detect alkanes with broader chain length specificity, while also exhibiting improved dynamic range and lower leakiness. As a result, recent studies have reported the application of these tools for the directed evolution of ADO, achieving promising improvements in activity.

A relatively unexplored method for engineering hydrocarbon production is directed evolution by selection. Many of the limitations of screening approaches can be circumvented using selection methods, but these may require more extensive genetic engineering and characterization to establish. The selection methods summarised in this work include selection by substrate toxicity alleviation, cofactor-based growth coupling and hydrocarbon sensitive transcription-factors driving expression of essential genes. Besides their use as enzyme engineering platforms, strains displaying a growth dependence on hydrocarbon biosynthesis may be valuable assets in the context of bioprocess stability and antibiotic-free

biomanufacturing. Despite these attractive features, the body of work encompassing growth-based selection for hydrocarbon biosynthesis is minimal and remains an area with high potential.

In summary, the field of drop-in hydrocarbon biosynthesis is steadily moving from the initial characterizations and metabolic engineering efforts towards using the principles of directed evolution to accelerate the design-build-test-learn cycle. By fully utilising the continuously expanding toolbox of synthetic biology, screening and selection of hydrocarbon-producing enzymes and organisms will be made increasingly accessible. Hopefully, this will accelerate the necessary discovery of novel enzyme variants and strain modifications to help drive sustainable hydrocarbon synthesis towards commercial viability.

Acknowledgements

This work was supported by the BBSRC (grant BB/M011224/1) and the Oxford Interdisciplinary Bioscience Doctoral Training Partnership.

References

Aarts, M. G. M., Keijzer, C. J., Stiekema, W. J., & Pereira, A. (1995). Molecular Characterization of the CER1 Gene of Arabidopsis Involved in Epicuticular Wax Biosynthesis and Pollen Fertility. *The Plant Cell*, *7*(12), 2115. https://doi.org/10.2307/3870155

Alter, T. B., & Ebert, B. E. (2019). Determination of growth-coupling strategies and their underlying principles. *BMC Bioinformatics 2019 20:1*, *20*(1), 1–17. https://doi.org/10.1186/S12859-019-2946-7

Amer, M., Toogood, H., & Scrutton, N. S. (2020). Engineering nature for gaseous hydrocarbon production. *Microbial Cell Factories*, *19*(1), 1–14. https://doi.org/10.1186/S12934-020-01470-6

Arnold, F. H. (1998). Design by Directed Evolution. *Accounts of Chemical Research*, 31(3), 125–131. https://doi.org/10.1021/AR960017F

Baker, G. L., Vroman, H. E., & Padmore, J. (1963). Hydrocarbons of the American cockroach. *Biochemical and Biophysical Research Communications*, *13*(5), 360–365. https://doi.org/10.1016/0006-291X(63)90348-6

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., & Milo, R. (2011). The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, *50*(21), 4402–4410. https://doi.org/10.1021/BI2002289

Beller, H. R., Goh, E. B., & Keasling, J. D. (2010). Genes involved in long-chain alkene biosynthesis in micrococcus Luteus. *Applied and Environmental Microbiology*, *76*(4), 1212–1223. https://doi.org/10.1128/AEM.02312-09

Berla, B. M., Saha, R., Maranas, C. D., & Pakrasi, H. B. (2015). Cyanobacterial Alkanes Modulate Photosynthetic Cyclic Electron Flow to Assist Growth under Cold Stress. *Scientific Reports 2015 5:1*, *5*(1), 1–12. https://doi.org/10.1038/srep14894

Bernard, A., Domergue, F., Pascal, S., Jetter, R., Renne, C., Faure, J.-D., Haslam, R. P., Napier, J. A., Lessire, R., & Joubès, J. (2012). Reconstitution of Plant Alkane Biosynthesis in Yeast Demonstrates That Arabidopsis ECERIFERUM1 and ECERIFERUM3 Are Core Components of a Very-Long-Chain Alkane Synthesis Complex. *The Plant Cell*, *24*(7), 3106–3118. https://doi.org/10.1105/TPC.112.099796

Bourdenx, B., Bernard, A., Domergue, F., Pascal, S., Léger, A., Roby, D., Pervent, M., Vile, D., Haslam, R. P., Napier, J. A., Lessire, R., & Joubès, J. (2011). Overexpression of Arabidopsis ECERIFERUM1 Promotes Wax Very-Long-Chain Alkane Biosynthesis and Influences Plant Response to Biotic and Abiotic Stresses. *Plant Physiology*, *156*(1), 29–45. https://doi.org/10.1104/PP.111.172320

Cao, Y. X., Xiao, W. H., Zhang, J. L., Xie, Z. X., Ding, M. Z., & Yuan, Y. J. (2016). Heterologous biosynthesis and manipulation of alkanes in Escherichia coli. *Metabolic Engineering*, *38*, 19–28. https://doi.org/10.1016/J.YMBEN.2016.06.002

Chen, D., Xu, S., Li, S., Tao, S., Li, L., Chen, S., & Wu, L. (2023). Directly Evolved AlkS-Based Biosensor Platform for Monitoring and High-Throughput Screening of Alkane Production. *ACS Synthetic Biology*, *12*(3), 832–841. https://doi.org/10.1021/ACSSYNBIO.2C00620

CO2 Emissions in 2022 – Analysis - IEA. (2023). https://www.iea.org/reports/co2-emissions-in-2022

De Visser, J. A. G. M., & Rozen, D. E. (2006). Clonal Interference and the Periodic Selection of New Beneficial Mutations in Escherichia coli. *Genetics*, 172(4), 2093. https://doi.org/10.1534/GENETICS.105.052373

Dennig, A., Kuhn, M., Tassoti, S., Thiessenhusen, A., Gilch, S., Bülter, T., Haas, T., Hall, M., & Faber, K. (2015). Oxidative Decarboxylation of Short-Chain Fatty Acids to 1-Alkenes. *Angewandte Chemie International Edition*, *54*(30), 8819–8822. https://doi.org/10.1002/ANIE.201502925

Fang, B., Xu, H., Liu, Y., Qi, F., Zhang, W., Chen, H., Wang, C., Wang, Y., Yang, W., & Li, S. (2017). Mutagenesis and redox partners analysis of the P450 fatty acid decarboxylase OleTJE. *Scientific Reports 2017 7:1*, *7*(1), 1–7. https://doi.org/10.1038/srep44258

Fatma, Z., Hartman, H., Poolman, M. G., Fell, D. A., Srivastava, S., Shakeel, T., & Yazdani, S. S. (2018). Model-assisted metabolic engineering of Escherichia coli for long chain alkane and alcohol production. *Metabolic Engineering*, *46*, 1–12. https://doi.org/10.1016/j.ymben.2018.01.002

Grant, C., Deszcz, D., Wei, Y. C., Martínez-Torres, R. J., Morris, P., Folliard, T., Sreenivasan, R., Ward, J., Dalby, P., Woodley, J. M., & Baganz, F. (2014). Identification and use of an alkane transporter plug-in for applications in biocatalysis and whole-cell biosensing of alkanes. *Scientific Reports 2014 4:1*, *4*(1), 1–9. https://doi.org/10.1038/srep05844

Han, J., McCarthy, E. D., Calvin, M., & Benn, M. H. (1968). Hydrocarbon constituents of the blue-green algae Nostoc muscorum, Anacystis nidulans, Phormidium luridium and Chlorogloea fritschii. *Journal of the Chemical Society C: Organic*, *0*, 2785–2791. https://doi.org/10.1039/J39680002785

Holze, H., Schrader, L., & Buellesbach, J. (2020). Advances in deciphering the genetic basis of insect cuticular hydrocarbon biosynthesis and variation. *Heredity 2020 126:2*, *126*(2), 219–234. https://doi.org/10.1038/s41437-020-00380-y

Hu, Y., Zhu, Z., Gradischnig, D., Winkler, M., Nielsen, J., & Siewers, V. (2020). Engineering carboxylic acid reductase for selective synthesis of medium-chain fatty alcohols in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(37), 22974–22983. https://doi.org/10.1073/PNAS.2010521117

Jetter, R., & Kunst, L. (2008). Plant surface lipid biosynthetic pathways and their utility for metabolic engineering of waxes and hydrocarbon biofuels. *The Plant Journal*, *54*(4), 670–683. https://doi.org/10.1111/J.1365-313X.2008.03467.X

Ju, S., Li, D., Mai, B. K., Liu, X., Vallota-Eastman, A., Wu, J., Valentine, D. L., Liu, P., & Yang, Y. (2024). Stereodivergent photobiocatalytic radical cyclization through the repurposing and directed evolution of fatty acid photodecarboxylases. *Nature Chemistry 2024*, 1–9. https://doi.org/10.1038/s41557-024-01494-0

Kang, M.-K., & Nielsen, J. (2016). Biobased production of alkanes and alkenes through metabolic engineering of microorganisms. *Journal of Industrial Microbiology & Biotechnology 2016 44:4*, *44*(4), 613–622. https://doi.org/10.1007/S10295-016-1814-Y

Kennerley, J. (2022). *an Evolving Synthetic Biological Platform for Alkane Production* [University of Oxford]. https://doi.org/10.13039/501100000266

Khara, B., Menon, N., Levy, C., Mansell, D., Das, D., Marsh, E. N. G., Leys, D., & Scrutton, N. S. (2013). Production of Propane and Other Short-Chain Alkanes by Structure-Based Engineering of Ligand Specificity in Aldehyde-Deformylating Oxygenase. *ChemBioChem*, *14*(10), 1204–1208. https://doi.org/10.1002/cbic.201300307

Kramer, L., Le, X., Rodriguez, M., Wilson, M. A., Guo, J., & Niu, W. (2020). Engineering Carboxylic Acid Reductase (CAR) through a Whole-Cell Growth-Coupled NADPH Recycling Strategy. *ACS Synthetic Biology*, *9*(7), 1632–1637. https://doi.org/10.1021/ACSSYNBIO.0C00290

Lea-Smith, D. J., Ortiz-Suarez, M. L., Lenn, T., Nürnberg, D. J., Baers, L. L., Davey, M. P., Parolini, L., Huber, R. G., Cotton, C. A. R., Mastroianni, G., Bombelli, P., Ungerer, P., Stevens, T. J., Smith, A. G., Bond, P. J., Mullineaux, C. W., & Howe, C. J. (2016). Hydrocarbons Are Essential for Optimal Cell Size, Division, and Growth of Cyanobacteria. *Plant Physiology*, *172*(3), 1928–1940. https://doi.org/10.1104/PP.16.01205

Lehtinen, T., Santala, V., & Santala, S. (2017). Twin-layer biosensor for real-time monitoring of alkane metabolism. *FEMS Microbiology Letters*, *364*(6), 53. https://doi.org/10.1093/femsle/fnx053

Lehtinen, T., Virtanen, H., Santala, S., & Santala, V. (2018). Production of alkanes from CO2 by engineered bacteria. *Biotechnology for Biofuels*, *11*(1), 228. https://doi.org/10.1186/s13068-018-1229-2

Li, J., Ma, Y., Liu, N., Eser, B. E., Guo, Z., Jensen, P. R., & Stephanopoulos, G. (2020). Synthesis of high-titer alka(e)nes in Yarrowia lipolytica is enabled by a discovered mechanism. *Nature Communications 2020 11:1*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-19995-0

Li, N., Chang, W. C., Warui, D. M., Booker, S. J., Krebs, C., & Bollinger, J. M. (2012). Evidence for only oxygenative cleavage of aldehydes to alk(a/e)nes and formate by cyanobacterial aldehyde decarbonylases. *Biochemistry*, *51*(40), 7908–7916. https://doi.org/10.1021/bi300912n

Liu, K., & Li, S. (2020). Biosynthesis of fatty acid-derived hydrocarbons: perspectives on enzymology and enzyme engineering. *Current Opinion in Biotechnology*, *62*, 7–14. https://doi.org/10.1016/J.COPBIO.2019.07.005

Liu, Y., Chen, J., Khusnutdinova, A. N., Correia, K., Diep, P., Batyrova, K. A., Nemr, K., Flick, R., Stogios, P., Yakunin, A. F., & Mahadevan, R. (2020). A novel C-terminal degron identified in bacterial aldehyde decarbonylases using directed evolution. *Biotechnology for Biofuels*, *13*(1), 114. https://doi.org/10.1186/s13068-020-01753-5

Liu, Y., Wang, C., Yan, J., Zhang, W., Guan, W., Lu, X., & Li, S. (2014). Hydrogen peroxide-independent production of α-alkenes by OleT JE P450 fatty acid decarboxylase. *Biotechnology for Biofuels*, *7*(1), 1–12. https://doi.org/10.1186/1754-6834-7-28

Machado, H. B., Dekishima, Y., Luo, H., Lan, E. I., & Liao, J. C. (2012). A selection platform for carbon chain elongation using the CoA-dependent pathway to produce linear higher alcohols. *Metabolic Engineering*, *14*(5), 504–511. https://doi.org/10.1016/J.YMBEN.2012.07.002

Manley, O. M., Fan, R., Guo, Y., & Makris, T. M. (2019). Oxidative Decarboxylase UndA Utilizes a Dinuclear Iron Cofactor. *Journal of the American Chemical Society*, *141*(22), 8684–8688. https://doi.org/10.1021/JACS.9B02545

Marsh, E. N. G., & Waugh, M. W. (2013). Aldehyde decarbonylases: Enigmatic enzymes of hydrocarbon biosynthesis. In *ACS Catalysis* (Vol. 3, Issue 11, pp. 2515–2521). American Chemical Society. https://doi.org/10.1021/cs400637t

Maxel, S., King, E., Zhang, Y., Luo, R., & Li, H. (2020). Leveraging Oxidative Stress to Regulate Redox Balance-Based, In Vivo Growth Selections for Oxygenase Engineering. *Cite This: ACS Synth. Biol*, *2020*, 3133. https://doi.org/10.1021/acssynbio.0c00380

Mendez-Perez, D., Begemann, M. B., & Pfleger, B. F. (2011). Modular synthase-encoding gene involved in α-olefin biosynthesis in Synechococcus sp. strain pcC 7002. *Applied and Environmental Microbiology*, *77*(12), 4264–4267. https://doi.org/10.1128/AEM.00467-11

Molina, R. S., Rix, G., Mengiste, A. A., Álvarez, B., Seo, D., Chen, H., Hurtado, J. E., Zhang, Q., García-García, J. D., Heins, Z. J., Almhjell, P. J., Arnold, F. H., Khalil, A. S., Hanson, A. D., Dueber, J. E., Schaffer, D. V., Chen, F., Kim, S., Fernández, L. Á., … Liu, C. C. (2022). In vivo hypermutation and continuous evolution. *Nature Reviews Methods Primers 2022 2:1*, *2*(1), 1–22. https://doi.org/10.1038/s43586-022-00119-5

Nielsen, J. R., Weusthuis, R. A., & Huang, W. E. (2023). Growth-coupled enzyme engineering through manipulation of redox cofactor regeneration. *Biotechnology Advances*, 108102. https://doi.org/10.1016/J.BIOTECHADV.2023.108102

Oró, J., Nooner, D. W., & Wikström, S. A. (1965). Paraffinic Hydrocarbons in Pasture Plants. *Science*, *147*(3660), 870–873. https://doi.org/10.1126/SCIENCE.147.3660.870

Packer, M. S., & Liu, D. R. (2015). Methods for the directed evolution of proteins. *Nature Reviews Genetics 2015 16:7*, *16*(7), 379–394. https://doi.org/10.1038/nrg3927

Ratajczak, A., Geißdörfer, W., & Hillen, W. (1998). Expression of alkane hydroxylase from Acinetobacter sp. strain ADP1 is induced by a broad range of n-alkanes and requires the transcriptional activator AlkR. *Journal of Bacteriology*, *180*(22), 5822–5827. https://doi.org/10.1128/JB.180.22.5822-5827.1998

Reed, B., Blazeck, J., & Alper, H. (2012). Evolution of an alkane-inducible biosensor for increased responsiveness to short-chain alkanes. *Journal of Biotechnology*, *158*(3), 75–79. https://doi.org/10.1016/J.JBIOTEC.2012.01.028

Royce, L. A., Liu, P., Stebbins, M. J., Hanson, B. C., & Jarboe, L. R. (2013). The damaging effects of short chain fatty acids on Escherichia coli membranes. *Applied Microbiology and Biotechnology*, *97*(18), 8317. https://doi.org/10.1007/S00253-013-5113-5

Rude, M. A., Baron, T. S., Brubaker, S., Alibhai, M., Del Cardayre, S. B., & Schirmer, A. (2011). Terminal olefin (1-alkene) biosynthesis by a novel P450 fatty acid decarboxylase from Jeotgalicoccus species. *Applied and Environmental Microbiology*, *77*(5), 1718–1727. https://doi.org/10.1128/AEM.02580-10

Rui, Z., Harris, N. C., Zhu, X., Huang, W., & Zhang, W. (2015). Discovery of a Family of Desaturase-Like Enzymes for 1-Alkene Biosynthesis. *ACS Catalysis*, *5*(12), 7091–7094. https://doi.org/10.1021/ACSCATAL.5B01842

Rui, Z., Li, X., Zhu, X., Liu, J., Domigan, B., Barr, I., Cate, J. H. D., & Zhang, W. (2014). Microbial biosynthesis of medium-chain 1-alkenes by a nonheme iron oxidase. *Proceedings of the National Academy of Sciences*, *111*(51), 18237–18242. https://doi.org/10.1073/PNAS.1419701112

Schirmer, A., Rude, M. A., Li, X., Popova, E., & Del Cardayre, S. B. (2010). Microbial biosynthesis of alkanes. *Science*, *329*(5991), 559–562. https://doi.org/10.1126/science.1187936

Scown, C. D., & Keasling, J. D. (2022). Sustainable manufacturing with synthetic biology. *Nature Biotechnology 2022 40:3*, *40*(3), 304–307. https://doi.org/10.1038/s41587-022-01248-8

Sellés Vidal, L., Murray, J. W., & Heap, J. T. (2021). Versatile selective evolutionary pressure using synthetic defect in universal metabolism. *Nature Communications 2021 12:1*, *12*(1), 1–15. https://doi.org/10.1038/s41467-021-27266-9

Sorigué, D., Légeret, B., Cuiné, S., Blangy, S., Moulin, S., Billon, E., Richaud, P., Brugière, S., Couté, Y., Nurizzo, D., Müller, P., Brettel, K., Pignol, D., Arnoux, P., Li-Beisson, Y., Peltier, G., & Beisson, F. (2017). An algal photoenzyme converts fatty acids to hydrocarbons. *Science*, *357*(6354), 903–907. https://doi.org/10.1126/science.aan6349

Sorigué, D., Légeret, B., Cuiné, S., Morales, P., Mirabella, B., Guédeney, G., Li-Beisson, Y., Jetter, R., Peltier, G., & Beisson, F. (2016). Microalgae Synthesize Hydrocarbons from Long-Chain Fatty Acids via a Light-Dependent Pathway. *Plant Physiology*, *171*(4), 2393–2405. https://doi.org/10.1104/PP.16.00462

Sticher, P., Jaspers, M. C. M., Stemmler, K., Harms, H., Zehnder, A. J. B., & Van Der Meer, J. R. (1997). Development and characterization of a whole-cell bioluminescent sensor for bioavailable middle-chain alkanes in contaminated groundwater samples. *Applied and Environmental Microbiology*, *63*(10), 4053–4060. https://doi.org/10.1128/AEM.63.10.4053-4060.1997

Sukovich, D. J., Seffernick, J. L., Richman, J. E., Gralnick, J. A., & Wackett, L. P. (2010). Widespread head-to-head hydrocarbon biosynthesis in bacteria and role of OleA. *Applied and Environmental Microbiology*, *76*(12), 3850–3862. https://doi.org/10.1128/AEM.00436-10

van Beilen, J. B., Wubbolts, M. G., & Witholt, B. (1994). Genetics of alkane oxidation by Pseudomonas oleovorans. *Biodegradation*, *5*(3–4), 161–174. https://doi.org/10.1007/BF00696457

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., & Church, G. M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature 2009 460:7257*, *460*(7257), 894–898. https://doi.org/10.1038/nature08187

Ward, T., Zou, Z., Wu, S., Gerngross D-Bsse, D., Zürich, E., Lozhkin, B., & Chen, D. (2023). *A Microbial Cell Factory Combining an Artificial Metathase with a Fatty Acid Decarboxylase for New-to-Nature Synthesis of Cycloalkenes*. https://doi.org/10.21203/RS.3.RS-3706252/V1

Warui, D. M., Li, N., Nørgaard, H., Krebs, C., Bollinger, J. M., & Booker, S. J. (2011). Detection of formate, rather than carbon monoxide, as the stoichiometric coproduct in conversion of fatty aldehydes to alkanes by a cyanobacterial aldehyde decarbonylase. *Journal of the American Chemical Society*, *133*(10), 3316–3319. https://doi.org/10.1021/ja111607x

Wu, S., Zhou, Y., Gerngross, D., Jeschek, M., & Ward, T. R. (2019). Chemo-enzymatic cascades to produce cycloalkenes from bio-based resources. *Nature Communications 2019 10:1*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-13071-y

Wu, W., Zhang, L., Yao, L., Tan, X., Liu, X., & Lu, X. (2015). Genetically assembled fluorescent biosensor for in situ detection of bio-synthesized alkanes. *Nature Publishing Group*, *5*, 10907. https://doi.org/10.1038/srep10907

Xu, H., Liang, W., Ning, L., Jiang, Y., Yang, W., Wang, C., Qi, F., Ma, L., Du, L., Fourage, L., Zhou, Y. J., & Li, S. (2020). Directed Evolution of P450 Fatty Acid Decarboxylases via High-Throughput Screening towards Improved Catalytic Activity. *ChemCatChem*, *12*(1), 80–84. https://doi.org/10.1002/CCTC.201901347

Xu, W., Chen, Y., Li, D., Wang, Z., Xu, J., & Wu, Q. (2022). Rational design of fatty acid photodecarboxylase enables the efficient decarboxylation of medium- and short-chain fatty acids for the production of gasoline bio-alkanes. *Molecular Catalysis*, *524*, 112261. https://doi.org/10.1016/J.MCAT.2022.112261

Y, Q., C, T., C, W.-T., G, L. G., S, Y., E, W., T, F., N, T., GJ, B., & R, F. (2012). An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(37), 14858–14863. https://doi.org/10.1073/PNAS.1208650109

Yamamori, T., Kageyama, H., Tanaka, Y., & Takabe, T. (2018). Requirement of alkanes for salt tolerance of Cyanobacteria: characterization of alkane synthesis genes from salt-sensitive Synechococcus elongatus PCC7942 and salt-tolerant Aphanothece halophytica. *Letters in Applied Microbiology*, *67*(3), 299–305. https://doi.org/10.1111/LAM.13038

Yang, K. K., Wu, Z., & Arnold, F. H. (2019). Machine-learning-guided directed evolution for protein engineering. *Nature Methods 2019 16:8*, *16*(8), 687–694. https://doi.org/10.1038/s41592-019-0496-6

Zhang, D., He, Y., Wang, Y., Wang, H., Wu, L., Aries, E., & Huang, W. E. (2012). Whole-cell bacterial bioreporter for actively searching and sensing of alkanes and oil spills. *Microbial Biotechnology*, *5*(1), 87–97. https://doi.org/10.1111/J.1751-7915.2011.00301.X

Zhang, J., Lu, X., & Li, J. J. (2013). Conversion of fatty aldehydes into alk (a/e)nes by in vitro reconstituted cyanobacterial aldehyde-deformylating oxygenase with the cognate electron transfer system. *Biotechnology for Biofuels*, *6*(1), 86. https://doi.org/10.1186/1754-6834-6-86

Zhou, Y. J., Hu, Y., Zhu, Z., Siewers, V., & Nielsen, J. (2018). Engineering 1-Alkene Biosynthesis and Secretion by Dynamic Regulation in Yeast. *ACS Synthetic Biology*, *7*(2), 584–590. https://doi.org/10.1021/ACSSYNBIO.7B00338

2.7.2 Growth-coupled enzyme engineering through manipulation of redox cofactor regeneration

The following article has been published in Biotechnology Advances:

Nielsen, J. R., Weusthuis, R. A., & Huang, W. E. (2023). Growth-coupled enzyme engineering through manipulation of redox cofactor regeneration. Biotechnology Advances, 108102. https://doi.org/10.1016/J.BIOTECHADV.2023.108102

Growth-coupled enzyme engineering through manipulation of redox cofactor regeneration

Jochem R. Nielsen[a] (jochem.nielsen@worc.ox.ac.uk), Ruud A. Weusthuis[b] (ruud.weusthuis@wur.nl), Wei E. Huang[a] (wei.huang@eng.ox.ac.uk)

[a]Department of Engineering Science, University of Oxford, Oxford, OX1 3PJ, United Kingdom
[b]Department of Bioprocess Engineering, Wageningen University & Research, Wageningen, 6700AA, The Netherlands

Corresponding author: wei.huang@eng.ox.ac.uk

Abstract

Enzymes need to be efficient, robust, and highly specific for their effective use in commercial bioproduction. These properties can be introduced using various enzyme engineering techniques, with random mutagenesis and directed evolution (DE) often being chosen when there is a lack of structural information -or mechanistic understanding- of the enzyme. The screening or selection step of DE is the limiting part of this process, since it must ideally be (ultra)-high throughput, specifically target the catalytic activity of the enzyme and have an accurately quantifiable metric for said activity. Growth-coupling selection strategies involve coupling a desired enzyme activity to cellular metabolism and therefore growth, where growth (rate) becomes the output metric. Redox cofactors (NAD^+/NADH and $NADP^+$/NADPH) have recently been identified as promising target molecules for growth coupling, owing to their essentiality for cellular metabolism and ubiquitous nature. Redox cofactor oxidation or reduction can be disrupted through metabolic engineering and the use of specific culturing conditions, rendering the cell inviable unless a ‘rescue’ reaction complements the imposed metabolic deficiency. Using this principle, enzyme variants displaying improved cofactor oxidation or reduction rates can be selected for through an increased growth rate of the cell.

In recent years, several *E. coli* strains have been developed that are deficient in the oxidation or reduction of NAD^+/NADH and $NADP^+$/NADPH pairs, and of non-canonical redox cofactor pairs NMN^+/NMNH and NCD^+/NCDH, which provides researchers with a versatile toolbox of enzyme engineering platforms. A range of redox cofactor dependent enzymes have since been engineered using a variety of these strains, demonstrating the power of using this growth-coupling technique for enzyme engineering. This review aims to summarise the metabolic engineering involved in creating strains auxotrophic for the reduced or oxidised state of redox cofactors, and the resulting successes in using them for enzyme engineering. Perspectives on the unique features and potential future applications of this technique are also presented.

Keywords

Growth-coupling, metabolic engineering, redox cofactor, enzyme engineering, redox cofactor auxotrophy, non-canonical redox cofactor

Abbreviations

-

Introduction

An increasingly wide range of chemicals are now being produced from renewable feedstocks through biotechnological means (Lee et al., 2019). Bioprocesses rely heavily on enzymatic catalysis for the efficient production of these compounds. These reactions can take place *in vitro* using purified enzymes, or through whole-cell *in vivo* catalysis (Claassens et al., 2019; Rasor et al., 2021). Ensuring that enzymes perform optimally in their required environment is therefore of high interest for sustainable production (Scown & Keasling, 2022).

When potentially useful enzymes are discovered, their native functionality and characteristics often align poorly with the requirements for their use in commercial bioprocesses. Properties such as substrate specificity, catalytic rate and (thermo)stability are among several critical factors that must be optimized for efficient enzyme-driven bioprocesses (Bornscheuer et al.,

2012). This optimization can be time consuming, costly, and challenging, and therefore effective and cost-efficient ways for engineering enzymes with desirable characteristics are highly sought after.

To address these challenges, a wide variety of protein engineering approaches have been developed, which includes methodologies for identifying improved enzyme variants through screening and/or selection.

1.1 Protein engineering

Protein engineering methods are now widely used to improve and optimize enzymatic properties. These methods can roughly be divided into rational-, semi-rational- and random approaches. The first requires full structural knowledge of the enzyme in question, allowing for structure-guided site-specific engineering (Brannigan & Wilkinson, 2002). Semi-rational approaches cover a spectrum of protein engineering methods, neither fully rational nor random. These methods usually involve predictions of enzyme structures, sometimes in combination with other data such as phylogenetic analyses or molecular dynamics simulations, to infer regions and/or single residues that may be suitable targets for engineering (Lutz, 2010). Computational predictions of protein structures, such as those made by AlphaFold, are further driving the use of (semi)-rational protein engineering approaches (Jumper et al., 2021). Recently, DeepMind has announced that their AlphaFold tool has now made computational predictions of protein structures for almost every known protein – some 200 million (Callaway, 2022). Protein structure prediction tools can also support *de novo* protein design, paving the way towards made-for-purpose proteins (Huang et al., 2016).

The final approach is collectively named random protein engineering, where diversity is introduced at random into the target sequence and vast populations of mutated enzymes are subsequently assessed through screening and/or selection methods (Packer & Liu, 2015). (Semi)-rational engineering approaches require at least some information about the relationship between sequence, structure and the contribution of specific amino acid residues

to the function of interest of the target protein. This input information is not needed using random engineering approaches, making it a more accessible technique when faced with limited understanding of the relationship between protein sequence and performance (Arnold, 1998b).

1.2 Random mutagenesis methods

Several methods have been developed to introduce genetic diversity specifically into target sequences. A popular technique is *in vitro* mutagenesis via error-prone PCR, where a region of DNA is amplified using a mutagenic polymerase and/or under amplification conditions that give rise to a higher mutation rate (Mccullum et al., 2010)**.** Other *in vitro* approaches include gene shuffling and (iterative) site-saturation mutagenesis (Christians et al., 1999; Coco et al., 2001; Reetz & Carballeira, 2007; Zheng et al., 2004). More recently developed continuous *in vivo* mutagenesis tools, such as EvolvR (Halperin et al., 2018), OrthoRep (Ravikumar et al., 2018), T7-DIVA (Álvarez et al., 2020) and eMuta (Park & Kim, 2021), allow researchers to introduce diversity into targeted sequences *in vivo*. Another popular technique is multiplex automated genome engineering (MAGE), which allows continuous *in vivo* mutagenesis of multiple targets simultaneously, making it possible to generate highly diverse combinatorial libraries (Wang et al., 2009). Using these and other methods (see Simon et al., 2019 for a comprehensive overview), diversity can now be introduced in a highly targeted -and often continuous- fashion. These continuous mutagenesis approaches also allow researchers to dynamically alter selection pressures by perturbing the conditions of the evolution process while it is ongoing.

1.3 Directed evolution

One of the most effective methods for exploiting random protein engineering is directed evolution (DE), where a chosen selective pressure is imposed on an enzyme mutant library to identify improved variants (Arnold, 1998a). Although DE has been successfully used to engineer a multitude of enzymes, in part due to advances mentioned in section 1.2, its efficacy is determined and limited by the screening or selection step. To fully harness the power of DE,

this final step must be (ultra-)high throughput to ensure maximum sampling of the sequence space. In addition, the chosen method must be able to determine the activity of the target enzyme accurately, with limited false positives. This has been approached in a variety of ways since the inception of DE, ranging from petri dish or microtiter plate-based screening (10^4 – 10^6 variants per day) to fluorescence-activated cell-sorting (FACS) which can screen up to 10^7 variants per day (see Table) (Longwell et al., 2017). These methods can be used to measure production of the target product or consumption of the substrate, often via colorimetric or fluorescence detection methods.

Table 2.3. Comparison of several widely used screening/selection techniques for enzyme engineering

Technique	Throughput (variants/day)	Benefits	Limitations	References
In vitro				
Microchamber arrays	~10^6	Flexible platform, quantitative, measurements over time	Non-native cellular environment, expensive equipment, reporter-dependent	Chen et al., 2015; Longwell et al., 2017
Compartmentalization (e.g. oil/water) + FACS	~10^7	Flexible platform, quantitative	Non-native cellular environment, expensive equipment, reporter-dependent, endpoint measurement	Becker et al., 2004; Colin et al., 2015; Tawfik & Griffiths, 1998; Xiao et al., 2015
In vivo				
Agar plate-based (fluorescence/colorimetric)	10^4 - 10^5	Inexpensive, native cellular environment	Low throughput, qualitative, reporter-dependent, laborious	Longwell et al., 2017; Packer & Liu, 2015
Microtiter plate-based (fluorescence/colorimetric)	10^5 - 10^6	Quantitative, native cellular environment, measurements over time	Reporter-dependent, laborious	Dörr et al., 2016; Packer & Liu, 2015
Cell-as-compartment + FACS	~10^7	Quantitative, native cellular environment, iterative rounds of sorting possible	Expensive equipment, reporter-dependent	Aharoni et al., 2006; Longwell et al., 2017; Yang & Withers, 2009
Metabolite auxotrophy complementation	10^8 - 10^{10} (limited by transformation efficiency)	Inexpensive, growth-based, native cellular environment, continuous evolution possible	Requires metabolic engineering, semi-quantitative, platform limited in scope	Chen, J. et al., 2022; MacBeath et al., 1998; Orsi et al., 2021
Redox cofactor auxotrophy complementation	10^8 - 10^{10} (limited by transformation efficiency)	Inexpensive, growth-based, substrate/product independent, native cellular environment, versatile platform, continuous evolution possible	Requires metabolic engineering, reaction must involve cofactor, semi-quantitative, enzyme products not directly selected for	*Literature cited in this review*

1.4 Growth-coupled selection

Screening methods require the discrete observation of each enzyme variant in a library, whereas selection methods directly eliminate undesirable enzyme variants (Xiao et al., 2015). One form of enzyme selection is through growth complementation, which provides a cheap and intuitive method of selecting improved enzyme variants. When the catalytic activity of an enzyme is associated with the growth of a cell, it is said to be growth-coupled. The extent of growth-coupling can vary on a spectrum from weak to strong. Weak growth coupling is observed when a target compound is only produced in conditions where growth rates are above a certain threshold, an example being fermentative production of ethanol and acetate in *E. coli* and *S. cerevisiae* under anaerobic conditions, or during overflow metabolism (Alter & Ebert, 2019). Strong growth coupling refers to a state where the production of the target compound is essential for carbon metabolism and therefore growth. Examples of strong growth coupling are the essential secretion of oxidation byproduct CO_2, or lactate and acetate formation by acetogenic bacteria during growth on H_2 and CO_2 (Alter & Ebert, 2019).

When an enzyme's activity is strongly coupled to growth, cells with more active, stable, or highly expressed variants of this enzyme are expected to display an increased growth rate (given the growth rate is not yet at its maximum). Growth-coupling can be exploited as an ultra-high throughput form of selection since throughput is only limited by the transformation efficiency of the cell (Xiao et al., 2015).

Growth coupling can generally be achieved through two methods. The first of these is by ensuring growth is dependent on product synthesis by the target enzyme, for example by disrupting an essential biosynthetic pathway and allowing the target enzyme's product to 'rescue' the auxotrophic state caused by this disruption. This option lends itself well to the engineering of enzymes involved in native metabolite biosynthesis, as the production of almost any metabolite can theoretically be coupled to growth in several model organisms (Von Kamp & Klamt, 2017). An example of this approach includes the disruption of TCA cycle genes

involved in succinate biosynthesis, rendering the cell inviable unless complemented by an alternative pathway that produces this essential metabolite (Theodosiou et al., 2022).

Although some growth coupling strategies may be designed holistically, the complexity of cell metabolism means that designing stringent growth coupling strategies often requires considerable knowledge of the cell's metabolic networks and their interactions. Untangling this metabolic complexity to achieve growth coupling has become increasingly accessible through the development of metabolic modelling tools designed for this specific purpose. For example, programs such as GCopt and GCfront allow users to specify a substrate, a desired reaction they wish to couple to growth and a maximum amount of gene deletions (Alter & Ebert, 2019; Legon et al., 2022). An algorithm then predicts a set of genetic modifications required to growth-couple the desired reaction. The GCfront program also outputs solutions with varying growth-coupling strengths, offering flexibility regarding the desired selection pressure stringency. If the target enzyme produces a non-native product, coupling growth to essential metabolite synthesis becomes more challenging, since it would require introducing additional reactions which link non-native product synthesis to host cell metabolism.

A second method of growth coupling involves linking enzyme activity to the global energy state of the cell, for example through ATP production and/or ensuring a cellular redox state conducive to growth (Alter & Ebert, 2019). This approach for growth-coupling may be more broadly applicable than complementing metabolite auxotrophy, as ATP availability and redox balancing are ubiquitous requirements for functional cell metabolism. In recent years, redox cofactors have been explored as growth-coupling targets for enzyme engineering strategies. The regeneration of these cofactors is deeply linked to the energy state of the cell and therefore to growth, making the coupling of this process to a target enzyme a viable selection strategy for stringent growth-coupling (Alter & Ebert, 2019).

Recent reviews have covered several benefits and applications of using synthetic biology to create strains suitable for growth-coupled selection (Chen, J. et al., 2022; Orsi et al., 2021). This work aims to provide perspectives and summarize in greater detail the recent advances

specifically in the engineering of strains deficient in the oxidized or reduced states of redox cofactor pairs, including the subsequent use of these strains as enzyme engineering platforms. The more widespread use of these *in vivo* enzyme engineering platforms may accelerate the development of improved biocatalysts and bioprocesses.

Redox cofactor oxidation or reduction rate as an enzyme selection strategy

2.1 Redox cofactors

Redox cofactors are essential redox couples that facilitate electron transfer for many biosynthetic and catabolic reactions (Wang et al., 2013). These molecules are ubiquitous to life and participate in around 25% of all metabolic reactions catalogued in the KEGG database (Opitz & Heiland, 2015). The most common cofactors are nicotinamide adenine dinucleotide (NAD^+), nicotinamide adenine dinucleotide phosphate ($NADP^+$) and their reduced counterparts NADH and NADPH. In general, the NAD^+/NADH cofactor pair is involved in oxidation reactions in both catabolism and anabolism, whilst $NADP^+$/NADPH is required for anabolic reduction reactions in e.g. DNA and fatty acid biosynthesis (Lin & Guarente, 2003; Spaans et al., 2015). The intracellular ratios of [NAD^+]/[NADH] and [$NADP^+$]/[NADPH] reflect their respective roles, as [NAD^+]/[NADH] ratios are usually high whilst [$NADP^+$]/[NADPH] ratios are low, although these ratios can vary significantly depending on growth state, environmental pressures, and nutrient availability (Bennett et al., 2009; De Graef et al., 1999; Spaans et al., 2015; Zhou et al., 2011). Overall, the intracellular concentration and ratio of the [NAD^+]/[NADH] redox cofactor couple thermodynamically favors the oxidation of substrates. In contrast, the ratio of [$NADP^+$]/[NADPH] reflects its role in providing reducing power for anabolic reactions required for cell growth and maintenance.

2.2 Redox cofactor auxotrophy

For the purposes of this review, the term ‘cofactor auxotrophy’ is used to refer to a cellular state where the redox state ratio of a cofactor pair is virtually entirely either reduced or oxidized. In this sense, the cell could be considered auxotrophic for a specific redox state of a

certain cofactor pair. It is acknowledged that cells are not in fact auxotrophic for the biosynthesis of redox cofactors, however, an extreme redox state ratio of a cofactor pair manifests itself similarly to auxotrophy for any other essential metabolite; growth ceases until the extreme redox state ratio of the cofactor pair is relieved. Hence, 'cofactor auxotrophy' was considered an appropriate term in the context of growth-coupled selection.

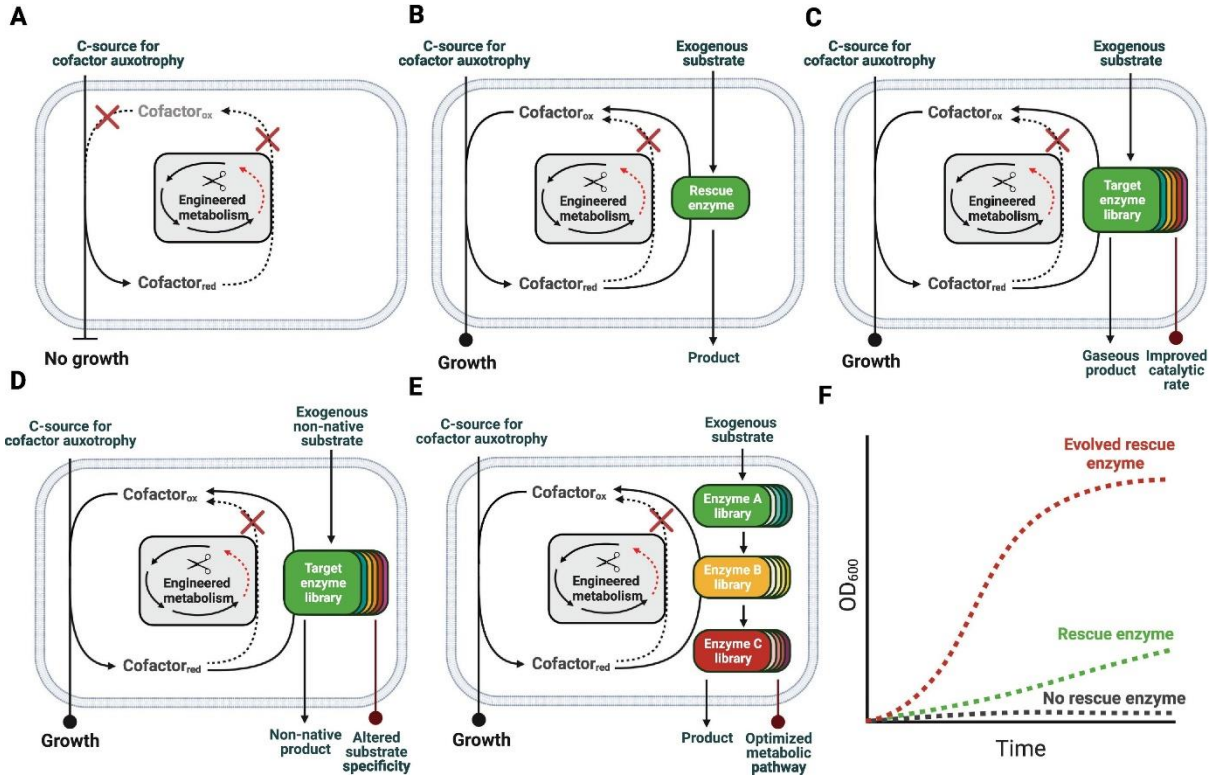


Figure 2.4. Schematic representations of redox cofactor regeneration as a growth-coupling strategy, and several ways this can be employed for enzyme engineering. These figures use the example of a cell deficient in the oxidised state of a cofactor pair, for example NAD^+ or $NADP^+$, when grown on a certain carbon source. **A**) Engineered cells displaying redox cofactor auxotrophy, unable to grow without a functional cofactor regeneration pathway. **B**) Introduction of a 'rescue' enzyme and substrate to complement the deficient phenotype allows for cycling of the relevant cofactor and growth. **C**) Introduction of a randomly mutated enzyme library and a relevant substrate to select for highly active variants via the cofactor regeneration rate, even if the target product is gaseous and therefore not amenable for intracellular detection. **D**) Engineering enzyme substrate specificity by supplementing engineered cells with a non-native substrate for the target enzyme library. **E**) Multi-enzyme metabolic pathway targeted for mutagenesis, where more effective enzyme variant combinations will be selected for via the cofactor oxidation rate of the pathway, leading to selection for an optimal set of enzyme variants and an increased product synthesis. **F**) Schematic graph of the output metric -growth- in the absence and presence of a wildtype rescue enzyme, and a variant from a random mutagenesis library selected for using the redox cofactor growth-coupling approach. Created with BioRender.com.

In theory, the activity of virtually any reaction requiring redox cofactors can be coupled to growth by first inducing cofactor auxotrophy for either the oxidized or reduced state of a specific redox cofactor pair. When these mutated strains are subsequently complemented with a desired reaction that ‘rescues’ the imbalanced redox state ratio, the growth rate of the cell can be linked to enzyme activity through the cofactor reduction or oxidation rate (see Figure 2.4A and B). Since redox cofactors are common reactants for a variety of enzymatic reactions, a single cofactor auxotrophic strain can be used to engineer a multitude of enzymes that are able to rescue the deficiency. This contrasts with other growth-coupling strategies where growth is coupled to the synthesis of a specific product, making these platforms less broadly employable.

2.3 Cofactor specificity engineering

In theory, the activity of virtually any reaction requiring redox cofactors can be coupled to growth by first inducing cofactor auxotrophy for either the oxidized or reduced state of a specific redox cofactor pair. When these mutated strains are subsequently complemented with a desired reaction that ‘rescues’ the imbalanced redox state ratio, the growth rate of the cell can be linked to enzyme activity through the cofactor reduction or oxidation rate (see Figure 2.4A and B). Since redox cofactors are common reactants for a variety of enzymatic reactions, a single cofactor auxotrophic strain can be used to engineer a multitude of enzymes that are able to rescue the deficiency. This contrasts with other growth-coupling strategies where growth is coupled to the synthesis of a specific product, making these platforms less broadly employable.

One possible application of redox cofactor growth-coupling is to engineer the cofactor specificity of enzymes, which is especially relevant from a metabolic engineering perspective. For example, altering cofactor specificities of catabolic (e.g. glycolytic) enzymes from NAD^+/NADH to $NADP^+$/NADPH can result in higher yields of NADPH-dependent biosynthetic pathways. Conversely, engineering $NADP^+$/NADPH utilising enzymes involved in substrate catabolism towards NAD^+/NADH dependence can increase yields of fermentation pathways

(Aslan et al., 2017; Bengtsson et al., 2009; Liu et al., 2018; Shen et al., 2011; Tamakawa et al., 2011). Adapting the cofactor specificity of enzymes has become more accessible with tools such as CSR-SALAD (Cofactor Specificity Reversal–Structural Analysis and Library Design), which can predict amino acid residues likely to interact with NAD/NADP cofactors and suggests mutational libraries to screen for altering this specificity (Cahn et al., 2017).

2.4 Benefits of growth coupling via cofactor auxotrophy

Engineering cofactor specificity using cofactor auxotrophic strains directly selects for the desired specificity via the reduction or oxidation rate of the target cofactor pair. However, for many enzymes that require cofactors, other substrates or product(s) of the enzyme are of greater interest than the cofactor itself. In this case, when using cofactor auxotrophic strains, the cofactor oxidation or reduction rate acts as a proxy for the biosynthesis rate of the product of interest. This represents a unique benefit of using cofactor auxotrophy as a selection method: selection acts independently of substrates or products of the enzyme of interest. It is therefore particularly useful for engineering enzymes which require cofactors but catalyse the formation of desirable products that are otherwise difficult to detect, quantify or couple to growth *in vivo*. This could be due to these products being non-native to the cell which makes growth-coupling more challenging, or perhaps because the physical properties of the products make them inherently inaccessible *in vivo*. For example, coupling growth to products that are insoluble, gaseous, or rapidly secreted makes it challenging to identify an improved enzyme variant from a library due to a lack of compartmentalization and/or solubility of the product. This limits the engineering of enzymes involved in the biosynthesis of lipids, biofuels, gasses, organic solvents, or polymeric compounds, many of which are chemicals of significant commercial interest in the context of sustainable bioproduction (Jang et al., 2012; Liao et al., 2016; Zhou et al., 2018).

The substrate- and product independence of selection when using cofactor auxotrophs distinguishes this approach from other growth coupling strategies, offering several unique

benefits. Firstly, by using the cofactor reduction or oxidation rate as a selection pressure, the desired product does not have to be amenable for detection, quantification, or chemical alteration after it has been produced (Figure 2.4C). Depending on the product, this may represent a significant benefit over using other high-throughput approaches requiring direct detection of the product for screening and selection. In addition, this property also makes redox cofactor growth-coupling a powerful method for engineering substrate specificity, since one platform strain can be used to engineer an enzyme towards specificity for a variety of substrates (Figure 2.4D). Increased flux through metabolic pathways can also be selected for, as long as the pathway is cofactor dependent and the redox reaction is not rate-limiting (Figure 2.4E). Growth coupling approaches also offer an easy readout for detecting improved enzymatic activity, although this is not unique to cofactor-based growth coupling approaches (Figure 2.4F).

A final benefit of cofactor auxotrophs as growth coupling platforms stems from the ubiquitous nature of redox cofactors in microbial metabolism. Cofactor reduction and oxidation mainly occurs via central metabolic pathways, such as glycolysis (NAD^+-reducing), the pentose phosphate pathway ($NADP^+$-reducing), fermentation pathways (NADH-oxidizing) and the TCA cycle (NAD^+- and $NADP^+$-reducing) (De Graef et al., 1999; Spaans et al., 2015). Many of these metabolic pathways are central to microbial life, meaning that genetic engineering strategies for disrupting redox cofactor reduction or oxidation should mostly be laterally transferable to other microbes of interest. This is especially relevant for engineering enzymes for *in vivo* use, for example as part of a metabolic pathway. If engineering occurs *in vitro* or in a model organism not destined to be the final host of the enzyme, the environment the enzyme was engineered in may not adequately represent the cellular context it will ultimately be used in. Although various microbial hosts and expression systems have been developed specifically for heterologous protein expression, for example strains with decreased protease activity, several issues may still crop up when attempting to express non-native proteins (Kaur et al., 2018; Overton, 2014). These may include codon bias, erroneous disulfide bond formation and

chaperone incompatibility (Rosano & Ceccarelli, 2014). In addition, eukaryotic protein expression in prokaryotic hosts may be problematic due to a lack of glycosylating activity, although developments are underway to establish glycosylating pathways in *E. coli* (Gupta & Shukla, 2015). If any of these issues occur, it may be necessary to optimize the enzyme again for *in vivo* use in the final host, costing time and resources (Golynskiy et al., 2013). Using strains modified for growth coupling, enzymes can be directly engineered within the environment of the desired microbial host, given that the host is genetically accessible.

3. Development of NAD(P)$^+$/NAD(P)H growth-coupled selection platforms

Several studies have engineered and characterized strains of *E. coli, Pseudomonas putida* and *Corynebacterium glutamicum* unable to reduce or oxidize cofactors in certain conditions. These strains were created through a combination of rewiring central metabolism, supplying specific carbon sources that result in cofactor auxotrophy when consumed and through cultivation in either aerobic or anaerobic conditions (for elaboration and references see sections 3.1 – 3.4). These cofactor auxotrophic strains now exist for oxidized and reduced states of the NAD$^+$/NADH and NADP$^+$/NADPH redox cofactor pairs, and their use as enzyme engineering platforms has also been demonstrated.

3.1 Mutants deficient in NADH oxidation

The oxidation of NADH is achieved through two main routes in *E. coli*, depending on oxygen availability. In aerobic conditions, NADH is mainly oxidized by respiration, generating ATP. In the absence of oxygen, NADH can be oxidized through various fermentation pathways, generating products such as lactate and ethanol (De Graef et al., 1999). Blocking these mixed fermentation pathways in *E. coli* during anaerobic growth causes a deficiency in NADH oxidation, along with an accumulation of intermediate metabolites such as pyruvate and acetyl-CoA when grown on glycolytic substrates. This principle has been exploited to drive NADH oxidizing pathways for the anaerobic synthesis of 2-methylpropan-1-ol, 2,3-butanediol, 1-butanol and L-alanine (Bastian et al., 2011; Liang & Shen, 2017; Machado et al., 2012; Shen

et al., 2011; Zhang et al., 2007). These studies focused on enhancing bioproduction of the mentioned chemicals but not on using the mutant strains for enzyme engineering.

One of the earliest examples of redox cofactor auxotrophic strains being used for enzyme engineering relied on exploiting NAD^+ auxotrophy as a selection pressure. Machado *et al*. used a strain previously engineered by Shen *et al.*, comprising an *E. coli* with several deletions in mixed fermentation pathway genes, including lactate dehydrogenase (*ldhA*), alcohol dehydrogenase (*adhE*) and fumarate reductase (*frdBC*). This strain was unable to grown on glucose in anaerobic conditions, unless an NADH-oxidizing reaction was available. To achieve production of C6- and C8 alcohols, an NADH-dependent 3-hydroxy-acyl-CoA reductase from *Ralstonia eutropha* (PaaH1) was introduced. Expression of PaaH1 allowed anaerobic growth on glucose and improved hexanoic acid production. The enzyme was subsequently targeted for random mutagenesis by epPCR, and improved variants were identified by their faster anaerobic growth on glucose. This resulted in the discovery of several improved variants, finally allowing up to 67% higher n-hexanol titres (see Table 1.4) (Machado et al., 2012).

Other studies have engineered enzymes by exploiting similar redox principles. Sellés Vidal *et al*. also generated a strain unable to oxidize NADH anaerobically when supplying glucose as a carbon source, by deletion of *ldhA* and *adhE* (Sellés Vidal et al., 2021). The authors noted that the resulting *ΔldhA ΔadhE* strain was not completely unable to oxidize NADH when grown anaerobically, as growth was also observed when an NADPH-dependent alcohol dehydrogenase was expressed with supplementation of acetone. The authors hypothesized that transhydrogenase activity was responsible for this observed phenotype, and after additional deletion of both the soluble transhydrogenase (*sthA*) and the proton-driven pyridine nucleotide transhydrogenase (*pntAB*), growth of this strain in the same conditions was abolished. This highlighted the importance of separating the NAD^+/NADH and $NADP^+$/NADPH cofactor pools by removing transhydrogenase activity. Growth of strains depending on anaerobic conditions for selection can also be complemented by simply growing the mutant in aerobic conditions, providing a straightforward method for non-selective growth.

Using this anaerobic NAD^+-auxotrophic platform strain, the authors were able to revert the cofactor specificity of several enzymes, including two alcohol dehydrogenases, an imine reductase and a nitroreductase (see Table 1.4). Finally, the *ΔldhA ΔadhE* strain was used to engineer an isopropanol production pathway, where flux through the pathway alleviated the imposed cofactor redox state imbalance through the final enzymatic reaction which produced $NADP^+$ (transhydrogenase activity was left intact to also allow growth complementation through $NADP^+$ production). The promoters and RBSs of the isopropanol production pathway genes were randomized, and growth of cells transformed with this library were assessed in selective conditions. This resulted in a variant displaying the highest yield and titer of isopropanol from gluconate substrate to date. This strain may be applicable as a platform for engineering other NAD(P)H-dependent enzymes and pathways, especially those that are oxygen sensitive.

An NAD^+-auxotrophic *E. coli* strain for use in aerobic conditions required additional modifications since NAD^+ is regenerated through respiration. In addition to disrupting several fermentation pathways (*adhE, ldhA, frdBC*) and *pntAB*, Maxel *et al.* knocked out several respiration-associated genes including NADH-quinone oxidoreductase subunit (*nuoF*), type II NADH:quinone oxidoreductase (*ndh*) and chorismite pyruvate-lyase (*ubiC*) (Figure 2.5). This strain displayed a greatly diminished capacity for NADH-oxidation, verified by observations of growth-rescue in cells expressing NADH-dependent oxidases, but not NADPH-dependent oxidases (Maxel et al., 2021). This platform was used to engineer a ~2900-fold relative cofactor specificity switch (NADPH to NADH) for a cyclohexanone monooxygenase (CHMO), through use of a site-saturation mutagenesis library followed by one round of random mutagenesis and selection. This platform strain may be particularly useful for engineering other NADH-dependent oxygenases.

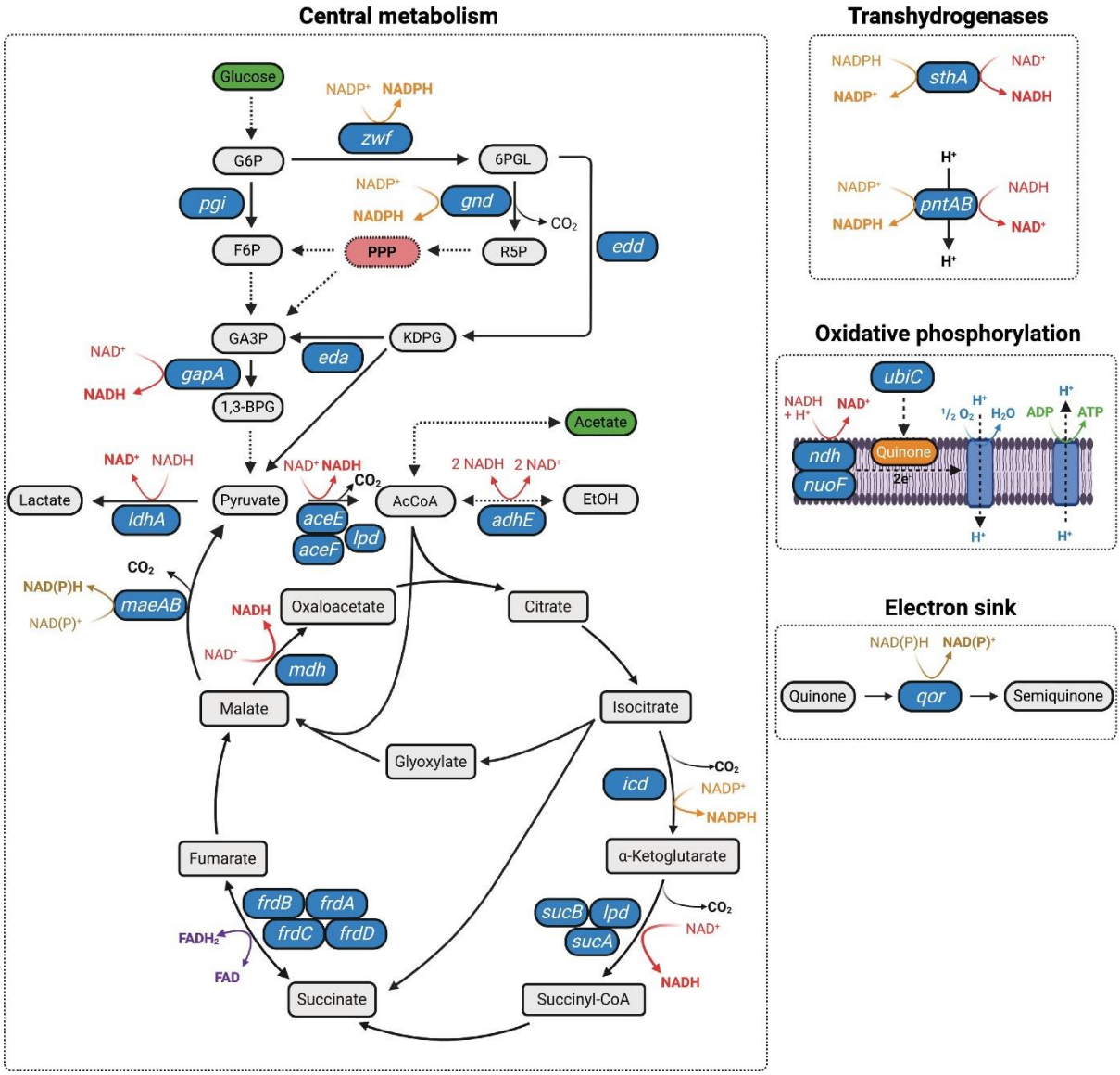


Figure 1.5. *E. coli* central metabolic reactions involving redox cofactors. *E. coli* gene names (blue boxes) represent enzyme(s) (complexes) that have been targeted in various studies to disrupt redox cofactor regeneration. These reactions may involve NAD^+/NADH (red), $NADP^+$/NADPH (orange), FAD/$FADH_2$ (purple) or either NAD(P)(H) cofactor pair (gold). Abbreviations: *aceE* (pyruvate dehydrogenase); *aceF* (dihydrolipoamide acetyltransferase); *adhE* (alcohol dehydrogenase); *eda* (2-keto-3-deoxy-6-phosphogluconate aldolase); *edd* (phosphogluconate dehydratase); *frdA* (fumarate reductase subunit A); *frdB* (fumarate reductase subunit B); *frdC* (fumarate reductase subunit C); *frdD* (fumarate reductase subunit D); *gapA* (glyceraldehyde-3-phosphate dehydrogenase); *gnd* (6-phosphogluconate dehydrogenase); *icd* (isocitrate dehydrogenase); *ldhA* (lactate dehydrogenase); *lpd* (dihydrolipoyl dehydrogenase); *maeA* (NAD^+-dependent malic enzyme); *maeB* ($NADP^+$-dependent malic enzyme); *mdh* (malate dehydrogenase); *ndh* (NADH:quinone oxidoreductase); *nuoF* (NADH:quinone oxidoreductase); *pgi* (glucose-6-phosphate isomerase); *pntAB* (pyridine nucleotide transhydrogenase); *qor* (quinone oxidoreductase); *sthA* (soluble transhydrogenase); *sucA* (2-oxoglutarate dehydrogenase); *sucB* (dihydrolipoyllysine succinyltransferase); *ubiC* (chorismite pyruvate-lyase); *zwf* (glucose-6-phosphate 1-dehydrogenase). Created with BioRender.com.

3.2 Mutants deficient in NAD^+ reduction

Through a single gene deletion, Wenk *et al*. engineered an *E. coli* strain that displayed auxotrophy for reducing power (NADH and NADPH) when aerobically grown on acetate. Dihydrolipoyl dehydrogenase (*lpd*) is a cofactor-binding subunit required for oxidative regeneration of lipoic acid, essential for the functionality of the pyruvate- and 2-ketoglutarate dehydrogenase complexes (PDHc and αKGDHc, respectively) and the glycine cleavage system (Wenk et al., 2020). Through deletion of *lpd*, carbon flux from glycolysis to the TCA cycle was blocked due to the absence of pyruvate dehydrogenase activity, although this is not the case for strains displaying pyruvate oxidase activity (Li et al., 2006). The *Δlpd* strain was not able to grow on acetate as a sole carbon source due to a lack of reducing power generated through acetate metabolism. Acetate can still be activated to acetyl-CoA and assimilated via the glyoxylate shunt, however absence of αKGDHc activity makes generating reducing power from acetate impossible. In addition, the authors showed that growth on pyruvate or succinate as sole carbon sources, or when supplemented with acetate, was also not possible as generating reducing power from these substrates requires PDHc and/or αKGDHc activity. When acetate was supplemented with upper glycolytic substrates such as glucose or glycerol, growth was rescued due to the reducing power made available through glycolysis. Growth rescue of the *Δlpd* strain grown on acetate was also observed after expression of NAD^+-dependent formate-, ethanol- or methanol dehydrogenases, when the respective substrates for these enzymes were supplemented. This work demonstrated how redox cofactor auxotrophy could be made dependent on the supplied carbon source, providing an uncomplicated way to cultivate the auxotrophic strains simply by adding rescuing substrates until selection was required. The strain has not yet been used for enzyme engineering.

The *Δlpd* strain was only assessed for redox cofactor auxotrophy under aerobic conditions. A strain displaying NADH auxotrophy in anaerobic conditions would provide a useful platform for the engineering of oxygen-sensitive NADH-consuming enzymes.

3.3 Mutants deficient in NADPH oxidation

Two strategies have been described to induce $NADP^+$ auxotrophy, both relying on engineering the glycolytic pathway to overproduce NADPH but used in different conditions (anaerobic versus aerobic). Zhang *et al*. focused on the *E. coli gapA* gene (GAPDH: glyceraldehyde-3-phosphate dehydrogenase) to generate an $NADP^+$ auxotroph suitable for use as a selection platform for NADPH-dependent fermentative enzymes. The *gapA* enzyme is essential for glycolytic growth of *E. coli* and natively NAD^+-dependent (Seta et al., 1997; Yun et al., 2000). By expressing a heterologous phosphorylating or non-phosphorylating $NADP^+$-dependent GAPDH (*gapC* from *Clostridium acetobutylicum* or *gapN* from *Streptococcus mutans,* respectively) and knocking out the native *gapA* gene, glycolysis until the pyruvate node was engineered to be entirely NADPH-generating when grown on glucose. These GAPDH enzymes have previously been heterologously expressed in *E. coli* to increase the NADPH yield of glycolysis (Centeno-Leija et al., 2013; Martínez et al., 2008; Takeno et al., 2016). With glucose as the sole carbon source growth was not observed in *ΔgapA* strains, unless complemented with expression of either *gapC* or *gapN*. To prove NADP(H)-specific cofactor cycling was taking place, the authors grew the *gapN*- and *gapC*-complemented *ΔgapA* strains in anaerobic conditions in LB supplemented with glucose, with the strain additionally harboring either an NADH-dependent alcohol dehydrogenase (ADH) from *Klebsiella pneumoniae* (*budC*) or an NADPH-dependent ADH from *Clostridium beijerinckii* (*adh*). Interestingly, the *ΔgapA* strain complemented with *gapC* displayed growth with expression of either ADH, and even in the absence of the ADH substrate acetoin, which the authors attributed to the less stringent cofactor specificity of *gapC* and NAD^+-reducing activity of native fermentation pathways. In contrast, growth of the *gapN*-complemented *ΔgapA* strain was only observed upon addition of acetoin and when the NADPH-dependent *adh* was expressed, but not with the NADH-dependent *budC*, indicating NADPH-specific cofactor cycling was taking place. The authors did not abolish transhydrogenase activity in this strain since no background growth was observed with acetoin supplementation and expression of *budC*, speculating that

expression levels of *sthA* and *pntAB* were not sufficient to supply the required $NADP^+$ from transhydrogenase activity. This contrasts with other studies finding that removal of transhydrogenase activity was essential for ensuring specific cofactor auxotrophy and to avoid background growth (Lindner et al., 2018; Maxel, Aspacio, et al., 2020; Sellés Vidal et al., 2021). Regardless, the authors were able to use the *ΔgapA +gapN* strain to engineer a D-lactate dehydrogenase towards NADPH-specificity, improving the k_{cat}/K_M 470-fold in a single round of mutagenesis (Table 1.4) (Zhang et al., 2018).

Another study induced $NADP^+$ auxotrophy in *E. coli* and demonstrated its use in an aerobic environment. Glycolysis was also engineered, however in this work carbon flux was redirected through the pentose phosphate pathway (PPP) to force NADPH generation. This was achieved by knocking out the glucose-6-phosphate isomerase gene (*pgi*), an approach also used in several bioproduction studies aiming to increase NADPH production from glycolysis (Chemler et al., 2010; Chin & Cirino, 2011; Sekar et al., 2017). To ensure maximum NADPH was produced from the carbon flux, the Entner-Doudoroff pathway was also knocked-out by deletion of the *edd* gene. Finally, two NADPH sinks including the soluble transhydrogenase (*sthA*) and quinone oxidoreductase (*qor*) were knocked-out to avoid background NADPH oxidation. The resulting strain (Δ*pgi* Δ*edd* Δ*sthA* Δ*qor*) was unable to grow on glucose aerobically but displayed growth with glycerol substrate, since glycerol catabolism bypasses the glucose-6-phosphate isomerase and doesn't involve the pentose phosphate pathway. In addition, growth on glucose could also be rescued by expression of a heterologous NADPH-dependent oxidase, but not an NADH-dependent variant (Maxel, Aspacio, et al., 2020). This strain was subsequently used for engineering several enzyme properties in separate studies, including improving the catalytic activity of a 4-hydroxybenzoate hydroxylase (Maxel, Aspacio, et al., 2020), the substrate specificity of a P450 cytochrome (Maxel, King, et al., 2020) and the thermostability of a cyclohexanone monooxygenase (Maxel, Zhang, et al., 2020) (see Table 1.4). This latter work represents the only study so far where thermostability was explicitly targeted for improvement using a cofactor auxotrophic strain. In these studies, enzyme

diversity was generated using either site-saturation mutagenesis or random mutagenesis (Maxel, Aspacio, et al., 2020; Maxel, King, et al., 2020; Maxel, Zhang, et al., 2020). The engineering successes achieved using this single platform strain showcased the flexibility of cofactor auxotrophs for engineering various properties of a range of enzymes.

A simpler version of the above strain has also been described, where the *edd* and *qor* genes were not deleted (Kramer et al., 2020). This mutant (*E. coli* BW25113 Δ*pgi* Δ*sthA*) cannot strictly be classified as an $NADP^+$-auxotrophic strain when grown on glucose, since after a ~24h lag-phase in selective conditions the culture was found to resume growth. Nevertheless, this strain was used to engineer an NADPH-dependent carboxylic acid reductase (CAR) from *Mycobacterium avium*. The CAR enzyme was mutagenized by error-prone PCR and expressed with the supplementation of either 2-methoxybenzoate or adipate to the culture medium. In a single round of selection, variants with a 4.7-fold improvement in k_{cat}/K_M for 2-methoxybenzoate and a 17-fold improvement in k_{cat}/K_M for adipate were identified (Table 1.4).

3.4 Mutants deficient in $NADP^+$ reduction

The *E. coli* NADPH-auxotroph engineered by Lindner *et al*. was one of the first redox cofactor auxotrophs designed with its application as an enzyme engineering platform in mind (Lindner et al., 2018). Several central metabolic enzymes were knocked-out to avoid $NADP^+$ reduction when glucose was provided as a carbon source. Glycolysis and TCA cycle knockouts included glucose-6-phosphate dehydrogenase (*zwf*), isocitrate dehydrogenase (*icd*) and $NADP^+$-dependent malic enzyme (*maeB*) (Figure 2.5). Strains with a deletion of *icd* required addition of 2-ketoglutarate to the medium as a precursor for glutamate and downstream amino acids biosynthesis. Upon deletion of the proton-driven transhydrogenase (*pntAB*) in addition to *zwf*, *icd* and *maeB*, growth was unexpectedly abolished on all carbon sources. This included gluconate, which is metabolized via NADPH-generating 6-phosphogluconate dehydrogenase (*gnd*) and was intended as a ‘rescue’ carbon source to allow growth in the auxotrophic strain. Upon additional deletion of the soluble transhydrogenase (*sthA*), growth on gluconate was restored, leading the authors to conclude that *sthA* activity was sufficient to deplete the

NADPH pool generated through gluconate assimilation. This implied that the *sthA* and *pntAB* transhydrogenase enzymes were depleting and regenerating cofactor pools in what seemed to be a futile cycle. As has been explored further in other work, this dual transhydrogenase activity may be beneficial for adapting to changes in substrate availability (Sauer et al., 2004). This NADPH-auxotrophic strain was used with the goal of engineering the cofactor specificity of an NAD^+-specific formate dehydrogenase (FDH) from *Pseudomonas sp.* 101 towards $NADP^+$-dependence. Using combinatorial saturation mutagenesis followed by selection using the NADPH-auxotrophic strain, a single round of mutagenesis yielded the most efficient and specific $NADP^+$-dependent FDH to date (Calzadiaz-Ramirez et al., 2020).

Recently, the redox metabolism of *Pseudomonas putida* was probed in work by Volke *et al*., leading to the construction of an NADPH-auxotrophic *P. putida* strain (Volke et al., 2022). The contribution of central metabolic enzymes to the maintenance of a redox balance in *P. putida* is less well understood than in *E. coli,* hence the authors decided to disrupt sets of target genes sequentially, to gain an understanding of their involvement in redox metabolism. CRISPR/nCas9-assisted engineering was used to disrupt genes of NADP(H)-dependent enzymes in modules based on their location in central metabolism, namely those involved in 'upper' metabolism (G6PDH and 6PGDH), 'lower' metabolism (Icd, MaeB and 2-ketoaldonate/hydroxypyruvate/glyoxylate reductase GhrB) and the two transhydrogenases (SthA and PntAB). The authors found that separate disruption of these modules had limited effects on the growth rate of strains using pyruvate as the sole carbon source. However, the combinatorial disruption of the upper metabolic enzymes (G6PDH and 6PGDH) and the transhydrogenases (SthA and PntAB), reduced the growth rate on pyruvate significantly. Upon additional disruption of the genes for the three lower metabolic enzymes, generating a strain designated 'NADPH-depleted *P. putida*', growth was not observed with pyruvate as the sole carbon source. The authors demonstrated that growth of this mutant could be rescued by supplementation with glutamate, the assimilation of which generated NADPH via glutamate dehydrogenase activity, or via supplementation with formate in conjunction with expression of

an $NADP^+$-dependent formate dehydrogenase engineered in other work (Calzadiaz-Ramirez et al., 2020). Although no enzymes have been engineered using this strain yet, this NADPH-depleted *P. putida* represents the first *P. putida* strain that could be used for growth-coupled enzyme engineering based on cofactor auxotrophy.

Besides *E. coli* and *P. putida, Corynebacterium glutamicum* has recently also been engineered to exhibit NADPH-auxotrophy (Chen, S. L. et al., 2022). Chen *et al.* applied a strategy similar to that which Lindner *et al.* used to create an *E. coli* NADPH-auxotroph, as described earlier in section 3.4 (Lindner et al., 2018). Briefly, glucose assimilation through the NADPH-generating pentose phosphate pathway was blocked by deletion of *zwf*, in addition to deletion of the $NADP^+$-dependent malic enzyme *malE* and replacement of the $NADP^+$-dependent *icd* with an NAD^+-dependent *icd* from *Streptococcus mutans.* These changes yielded a *C. glutamicum* strain unable to grow on glucose, acetate, succinate, or pyruvate. Growth of this strain on gluconate was possible however, due to the activity of NADPH-generating 6-phosphogluconate dehydrogenase (*gnd*). In contrast to *E. coli*, *C. glutamicum* has not been found to possess either soluble or membrane-bound transhydrogenases, hence their activity did not have to be taken into account (Kabus et al., 2007; Yamauchi et al., 2014). The authors used this NADPH-auxotrophic strain as a starting point for optimizing NADPH/NADH ratios for L-leucine biosynthesis. First, the native NAD^+-dependent GAPDH was replaced with $NADP^+$-dependent GAPDH from *C. acetobutylicum*. Subsequently, *zwf* and the native NAD^+-dependent *icd* were both re-introduced with mutated promoters to alter their expression levels, and these strains were screened for improved L-leucine production. Using this approach, a strain with improved growth, L-leucine titre and productivity was identified. Although the authors did not use their strain as a selection platform for enzyme engineering, it demonstrated how cofactor auxotrophic strains can be used as a starting point for redox balance optimization in cell factories. In future work, the NADPH-auxotrophic version of this strain could be used as an enzyme selection platform for engineering $NADP^+$-dependent enzymes within the industrially relevant *C. glutamicum* chassis.

As described in section 3.2, the *E. coli* MG1655 *Δlpd* strain described by Wenk *et al*. displayed auxotrophy for reducing power when grown on acetate. Hence, this strain could also be used for engineering $NADP^+$-dependent enzymes, although growth rescue using $NADP^+$-dependent enzymes was not demonstrated explicitly in their initial study (Wenk et al., 2020). The NADPH-auxotrophic strains highlighted in this section were assessed in aerobic conditions, and an NADPH-auxotroph specifically for anaerobic use has yet to be described.

4. Development of non-canonical redox cofactor selection platforms

A relatively unexplored area of redox cofactor engineering is the use of 'non-canonical' redox cofactors. These molecules retain the nicotinamide ring found in NAD(P)-based cofactors but have alternative functional groups, for example nicotinamide mononucleotide (NMN^+/NMNH) and nicotinamide cytosine dinucleotide pairs (NCD^+/NCDH) (King et al., 2020). These compounds are amenable for use as redox cofactors, but their alternative structures usually cause them to display low interactions with native NAD(P)(H)-dependent enzymes. Even so, several enzymes have been engineered to accept non-canonical redox cofactors, as summarized in other work (King et al., 2020). Previous work has described the biosynthesis of NCD^+ (Wang et al., 2021) and NMN^+ (Black et al., 2020) cofactors *in vivo*, which allowed non-canonical cofactor cycling without exogenous supplementation of the cofactor. Besides their use *in vivo*, methods for employing non-canonical cofactor cycling in cell-free systems have also been developed (Black & Li, 2022). One of the potential benefits of applying these alternative cofactors is that it can offer greater control over the fate of electrons derived from substrates. This is due to limited cross-talk between native and non-canonical redox cofactor usage in cell metabolism, which may be exploited to improve product yields of fermentation processes (Weusthuis et al., 2020). Several strains have recently been engineered which display (partial) growth dependence on non-canonical cofactor cycling.

4.1 Mutants deficient in NMNH oxidation

One study developed an NMN^+-dependent glucose dehydrogenase ortholog from *Bacillus subtilis* (BsGDH Ortho), and used this enzyme to couple NMN^+/NMNH cofactor pair cycling to growth via the Entner-Doudoroff pathway (Black et al., 2019). To achieve this, knockouts of alternative glucose utilisation routes (*Δzwf Δpgi Δgnd*) and potential NMN^+ degradation pathways (*ΔpncC ΔnadR*) were made. An NMNH-utilising enoate reductase *xenA* from *P. putida* was overexpressed, in addition to glucose facilitator *glf* (*Zymomonas mobilis*), gluconate kinase *gntK* (*Ralstonia eutropha*) and the NMN^+-biosynthesis genes *nadE* and *nadV* (both from *Francisella tularensis*). The resulting strain only displayed growth when expressing both NMN^+-dependent BsGDH Ortho and NMNH-dependent XenA, implying that glucose utilisation could only occur in strains displaying NMN^+/NMNH cofactor cycling. This strain, when lacking the NMNH-dependent XenA enzyme, could be used to engineer enzymes towards NMNH-dependency.

A second strain was engineered that utilised the NMN^+/NMNH cofactor pair in central metabolism (King et al., 2022). In this work, the $NADP^+$-dependent non-phosphorylating GAPDH from *Streptococcus mutans* (*gapN*) was engineered towards NMN^+ dependence through multiple rounds of rational mutagenesis, guided by *in silico* Rosetta modelling and molecular dynamics simulations. Two variants, GapN Penta and GapN Ortho, were highlighted in this study. The former variant displayed a 45-fold improved k_{cat}/K_M for NMN^+ compared to the WT *gapN in vitro*, although this efficiency was still considerably lower than the WT enzyme for $NADP^+$. To assess the GapN variants' functionality *in vivo*, an *E. coli* BW25113 derivative (*ΔgapA ΔmgsA ΔpncC*) was used in conjunction with expression of the previously mentioned XenA enzyme, to allow cofactor cycling by oxidation of NMNH. In the presence of GapN Penta, the strain still displayed growth on glucose without NMN^+ supplementation, indicating non-specific ($NAD(P)^+$) cofactor reduction. The other variant, GapN Ortho, displayed significantly improved cofactor specificity *in vitro*, attaining a 3.4×10^6-fold cofactor specificity switch from $NADP^+$ to NMN^+. This was also demonstrated *in vivo*, as

the GapN Ortho enzyme only supported minimal growth on glucose without NMN^+ supplementation (when supplemented with NMN^+, both GapN variants supported rapid growth). This strain may provide another platform for engineering NMNH-dependent enzymes, by replacing the currently used XenA enoate reductase with a target library. This work also reiterated the efficacy of computational approaches for rationally engineering cofactor specificity.

4.2 Mutants deficient in NMN^+ reduction

In recent work, a growth-coupled selection system was engineered based on a deficiency in NMN^+ reduction (Zhang et al., 2022). The production of reduced glutathione (GSH) was chosen for establishing coupling between cofactor cycling and growth. The *E. coli* SHuffle strain carries deletions in two genes involved in the production of GSH in *E. coli*, glutathione reductase (*gor*) and thioredoxin reductase (*trxB*) (Lobstein et al., 2012). Deletion of these genes resulted in a strain sensitive to oxidative stress, and supplementing the growth medium with 0.5 mM diamide was sufficient to inhibit cell growth. When native NADPH-dependent glutathione reductase was expressed on a plasmid in these conditions, growth was restored. To link this process to NMN^+/NMNH cycling, an NMNH-dependent glutathione reductase was developed in this work through several rounds of rational mutagenesis. This resulted in a variant (Gor Ortho) featuring a 6×10^4-fold cofactor specificity switch from NADPH to NMNH. As a proof-of-principle, an NMN^+-dependent glucose dehydrogenase, previously developed by the same group (BsGDH Ortho, see section 4.1), was used to support NMN^+/NMNH cofactor cycling and therefore growth in the SHuffle strain harbouring Gor Ortho. Supplementing the medium with NMN^+ was required to allow growth of strains dependent on NMN^+/NMNH-cycling. Growth was not observed when expressing NADPH-dependent BsGDH in the same conditions, demonstrating the stringency of the system.

This study also applied non-canonical redox cofactor auxotrophy for growth-coupled enzyme engineering for the first time (Zhang et al., 2022). A thermostable phosphite dehydrogenase

from *Pseudomonas stutzeri* (TS-PTDH) (Johannes et al., 2005) was first rationally engineered to accept the NMN^+/NMNH cofactor, resulting in a variant with ~3-fold higher k_{cat}/K_M for NMN^+ compared to the wildtype (TS-PTDH A155N). This variant was used as the starting point for constructing a site-saturation mutagenesis library, which was subsequently transformed (~2.5 x 10^6 transformants) into the SHuffle strain harbouring Gor Ortho. After two rounds of growth-coupled selection, the authors identified a variant (LY-13) displaying 110- and 44-fold improved catalytic efficiency compared to TS-PTDH and TS-PTDH A155N, respectively. The evolved LY-13 enzyme also performed remarkably well regarding temporal stability *in vitro*, reaching total turnover numbers (TTN) of ~45000 when coupled with NMNH-dependent XenA (see section 4.1), to produce levodione from ketoisophorone using phosphite as electron donor.

The work by Zhang *et al*. provided another useful strain for growth-coupled enzyme engineering, dependent on NMN^+/NMNH cycling. It represents the first platform strain auxotrophic for the reduced state of a non-canonical redox cofactor. The flexibility of this platform also allows it to be used for engineering other enzymes towards NMN^+-dependency.

4.3 Mutants deficient in NCDH oxidation

Another recent study established a growth-coupled NCD^+/NCDH cofactor cycling system, also involving phosphite dehydrogenases (Liu et al., 2022). However, their study revolved around rescuing phosphate auxotrophy in *E. coli* via NCD^+-dependent phosphite oxidation. The enzyme responsible for this, phosphite dehydrogenase (PDH, not to be confused with pyruvate dehydrogenase complex PDHc), is natively NAD^+-dependent, but the authors had previously engineered an NCD^+-dependent version (PDH*), which was used in this work (Liu et al., 2019). Further previous work described the development of a strain autotrophic for NCD^+, through reprogramming the nicotinic acid mononucleotide adenylyltransferase (NadD) to allow *in vivo* NCD^+ synthesis (Wang et al., 2021). These previously established elements were combined in a strain unable to metabolize phosphite (*E. coli* BW14329), due to inactivity

of *phoA* and deletions in the *phn* operon (Metcalf et al., 1991). This resulted in a strain displaying growth defects in medium containing only phosphite as the phosphorus source, and use of the PDH* enzyme ensured a dependence on NCD^+/NCDH cycling in this context. On medium containing glycerol and phosphite, faster growth was observed when the NCD^+-dependent PDH* enzyme was expressed in conjunction with an NCDH-specific malic enzyme (ME*), in contrast to the native NADH-dependent variant (ME), implying that growth was (partially) dependent on NCD^+/NCDH cycling. This growth difference was not observed when using glucose instead of glycerol. The lack of full cofactor redox state auxotrophy observed in this system is likely due to low, but appreciable, specificity of PDH* and ME* towards the native NAD^+/NADH cofactor pair. Although there is room for improving the stringency of the system, this work represents a novel approach for growth-coupling NCDH-dependent enzymes in an aerobic environment.

The above studies highlight the readiness of non-canonical redox cofactor cycling to be used for growth-coupled enzyme engineering. With the recent surge in interest in non-canonical redox cofactor cycling systems, it is likely that more industrially relevant enzymes will soon be engineered towards, and optimized for, non-canonical redox cofactor use. This could allow stricter control over electron flux from substrate to product leading to better yields in industrial processes.

5. Limitations of growth coupling via cofactor auxotrophy

As is the case with all high-throughput enzyme engineering strategies, one must carefully consider the first law of random mutagenesis; "*you get what you screen for"* (Arnold, 1998a), which also applies to selection. If using the cofactor reduction or oxidation rate as a proxy for the synthesis of an enzyme product, synthesis of the desired product is in fact not being directly selected for. Furthermore, in all strategies using cofactor reduction or oxidation as a selection method there is evolutionary pressure not just on the target enzyme, but on all enzymes present in the cell reasonably capable of evolving towards alleviating the cofactor

redox state deficiency. The implications of this problem were recently experimentally demonstrated by cultivation of an *E. coli* strain unable to reduce $NADP^+$ (Lindner et al., 2018). The strain was continuously cultured for 500-1100 generations in selective conditions, after which it had evolved to circumvent the imposed limitation through reversion of the cofactor specificity of either malic enzyme (*maeA*) or dihydrolipoyl dehydrogenase (*lpd*) towards $NADP^+$ (Bouzon et al., 2021). Naturally, the diversity introduced into an error-prone PCR library or via *in vivo* targeted hypermutagenesis should far exceed the diversity generation rate through natural evolution, but long-term continuous evolution experiments as described above may result in 'cheating' cells. To avoid this situation, it may be necessary to monitor the population's genotype or to develop a counter-selection method which can remove cheating cells from the population.

The evolution of known enzymes or pathways towards complementing metabolic deficiencies such as cofactor auxotrophy could be monitored with relative ease. More challenging however is taking into account the contribution of 'dark matter' metabolism in complementing engineered auxotrophies. These 'dark' metabolic pathways represent parts of cellular metabolism that are poorly characterized, for example 25% of proteins in *E. coli* have uncharacterized functions despite being one of the best characterized model organisms (MohammadiPeyhani et al., 2022). In the context of growth coupled selection, these metabolic 'blind spots' could prove problematic as they could alleviate imposed auxotrophies, either directly or via evolutionary adaptation as described earlier. This limitation may be more pronounced when attempting to introduce metabolic auxotrophies into poorly characterized organisms, as opposed to using model organisms with generally well-understood metabolisms.

Besides cells potentially circumventing the imposed cofactor auxotrophy via evolutionary adaptation or 'dark matter' metabolic pathways, there are also practical limitations to using growth coupling approaches. Most of the cofactor auxotrophic systems described in section 4 rely on supplementation of the cell medium with the substrate for the enzyme of interest, for

example formate (Calzadiaz-Ramirez et al., 2020), phosphite (Liu et al., 2022), cyclohexanone (Maxel, Zhang, et al., 2020), carboxylic acids (Kramer et al., 2020), acetone or 2-methyl-1-pyrroline (Sellés Vidal et al., 2021). The efficacy of enzyme selection will depend in part on how well this substrate is taken up by the cell and whether or not it is toxic, as low uptake may affect enzyme kinetics and the concomitant growth rate of the cell, while toxicity will similarly affect cell growth. Limited substrate uptake can be circumvented by using a native metabolite as the substrate for the enzyme of interest, but in cases where exogenous addition of substrate is required the uptake and toxicity of the substrate will both have to be assessed.

Table 1.4. Engineered strains displaying cofactor regeneration deficiencies and any enzymes engineered using these platforms

Regeneration deficiency	Strain and any genetic modifications	Carbon source(s) for auxotrophy	(An)aerobic cultivation	Enzyme target	Result	Reference
NADH → NAD^+	*E. coli* BW25113 *ΔldhA ΔadhE ΔfrdBC* +*Treponema denticola* and *Euglena gracilis* trans-enoyl-CoA reductases, +*Ralstonia eutropha* thiolase	Glucose	Anaerobic	*Ralstonia eutropha* 3-hydroxy-acyl-CoA reductase (NAD^+-dependent)	67% higher hexanoic acid titre	Machado et al., 2012
	E. coli BW25113 *ΔldhA ΔadhE*	Glucose	Anaerobic	*Clostridium beijerinckii* alcohol dehydrogenase ($NADP^+$-dependent)	Cofactor specificity reversal for NAD^+ (k_{cat}/K_M = 19 min^{-1} mM^{-1})	Sellés Vidal et al., 2021
	E. coli BW25113 *ΔldhA ΔadhE ΔpntAB ΔsthA*	Glucose	Anaerobic	*Thermoanaerobacter brockii* alcohol dehydrogenase ($NADP^+$-dependent)	Cofactor specificity reversal and 7.1-fold increased k_{cat}/K_M for oxidation of isopropanol using NAD^+ compared to $NADP^+$	
				Myxococcus stipitatus imine reductase ($NADP^+$-dependent)	Cofactor specificity reversal for NAD^+ (k_{cat}/K_M = 1.06 min^{-1} μM^{-1})	
				Enterobacter cloacae nitroreductase NfsB (NAD^+/$NADP^+$-dependent)	Evolved substrate specificity towards non-native substrate 2-nitrobenzoic acid. Improved efficiency for 4-nitrobenzyl alcohol	
NADH → NAD^+	*E. coli* BW25113 *ΔldhA ΔadhE ΔfrdBC ΔnuoF Δndh ΔubiC ΔpntAB*	Glucose	Aerobic	*Acinetobacter sp.* cyclohexanone monooxygenase (NADPH-dependent)	Cofactor specificity reversal, 2900-fold increased k_{cat}/K_M for NADH	Maxel et al., 2021
NAD^+ → NADH $NADP^+$ → NADPH	*E. coli* MG1655 *Δlpd*	Acetate	Aerobic	-	-	Wenk et al., 2020
NADPH → $NADP^+$	*E. coli* K12 (W3110) *ΔgapA* + *Streptococcus mutans gapN* ($NADP^+$-dependent)	Glucose	Anaerobic	*L. delbrueckii* lactate dehydrogenase (NADH-dependent)	Cofactor specificity reversal, 470-fold increased k_{cat}/K_M for NADPH	Zhang et al., 2018
NADPH → $NADP^+$	*E. coli* BW25113 *Δpgi Δedd ΔsthA Δqor*	Glucose	Aerobic	*Pseudomonas aeruginosa* 4-hydroxybenzoate hydroxylase (NADPH-dependent)	8-fold increased k_{cat}/K_M for 3,4-dihydroxybenzoic acid	Maxel, Aspacio, et al., 2020
				Bacillus megaterium P450 BM3 (NADPH-dependent)	2.5-fold improved total activity against acenaphthene substrate	Maxel, King, et al., 2020
				Acinetobacter sp. Cyclohexanone monooxygenase (NADPH-dependent)	Improved thermostability, 4.4-fold increased residual activity at 30°C	Maxel, Zhang, et al., 2020
NADPH → $NADP^+$	*E. coli* BW25113 *Δpgi ΔsthA*	Glucose	Aerobic	*Mycobacterium avium* carboxylic acid reductase (NADPH-dependent)	4.7-fold increase in k_{cat}/K_M for 2-methoxybenzoate, 17-fold increase in k_{cat}/K_M for adipate	Kramer et al., 2020

$NADP^+ \rightarrow NADPH$	*E. coli* MG1655 *Δzwf ΔmaeB ΔsthA ΔpntAB Δicd*	Glucose, glycerol, pyruvate, acetate, xylose, succinate	Aerobic	*Pseudomonas sp.* 101 formate dehydrogenase (NADH-dependent)	Cofactor specificity reversal, 510-fold improved specificity ratio for NADPH	Calzadiaz-Ramirez et al., 2020; Lindner et al., 2018
$NADP^+ \rightarrow NADPH$	*P. putida* KT2440 *ΔsthA, pntA^{W238*}, pntB^{Q117*}, pgi-I$^{Q129*-W229*}$, pgi-II^{Q129*}, ghrB^{W124*}, maeB^{Q314*}, icd^{Q78*}*	Pyruvate	Aerobic	-	-	Volke et al., 2022
$NADP^+ \rightarrow NADPH$	*C. glutamicum* XL-1 *Δzwf ΔmalE Δicd$_{Cg}$::icd$_{Sm}$*	Glucose, acetate, succinate, pyruvate	Aerobic	-	-	Chen, S. L. et al., 2022
$NMNH \rightarrow NMN^+$	*E. coli* BW25113 *Δpgi Δzwf Δgnd ΔpncC ΔnadR +B. subtilis gdh* (BsGDH Ortho, NMN^+-dependent), *+Z. mobilis glf. +R. eutropha gntK. +F. tularensis nadEV*	Glucose	Aerobic	-	-	Black et al., 2019
$NMNH \rightarrow NMN^+$	*E. coli* BW25113 *ΔgapA ΔmgsA ΔpncC +S. mutans gapN* (GapN Ortho, NMN^+-dependent)	Glucose	Aerobic	-	-	King et al., 2022
$NMN^+ \rightarrow NMNH$	*E. coli* T7 Shuffle Express (BL21 derivative harbouring *Δgor ΔtrxB*) *+E. coli gor* (Gor Ortho, NMNH-dependent)	Glucose, yeast extract	Aerobic	Thermostable *Pseudomonas stutzeri* phosphite dehydrogenase (TS-PTDH, NAD^+-dependent)	110-fold improved k_{cat}/K_M for NMN^+ compared to wildtype TS-PTDH	Zhang et al., 2022
$NCDH \rightarrow NCD^+$	*E. coli* BW14329 *+E. coli* NcdS (engineered *nadD* for NCD^+ synthesis) *+Ralstonia sp.* 4506 PDH* (NCD^+-dependent)	Glycerol (+phosphite)	Aerobic	-	-	Liu et al., 2022

6. Perspectives and outlook

Despite the considerable successes achieved so far, several avenues of research for redox cofactor-based growth-coupling remain unexplored. One of these includes the more widespread engineering of cofactor auxotrophic strains of organisms besides *E. coli*. The generation of NADPH-auxotrophic *Corynebacterium glutamicum* and *Pseudomonas putida* strains has recently been reported, which represent the first cofactor auxotrophic strains of these species. Although they have not yet been used for enzyme engineering, these strains can be employed for this purpose, opening up the possibility for growth-coupled enzyme engineering within two widely used chassis. However, other industrially relevant hosts, such as *Ralstonia eutropha, Clostridium sp., Yarrowia lipolytica* and *Saccharomyces cerevisiae* could be suitable candidates for adapting into similar enzyme engineering platforms. Considering that cofactor auxotrophy can already be induced through introduction of a single knockout in *E. coli* MG1655 (Δ*lpd*) with growth on acetate (Wenk et al., 2020), the construction of orthogonal strains from different species may be relatively easily achieved. In addition, one functional platform strain can be used to engineer a multitude of enzymes, which may make development of these platforms a worthwhile investment for researchers.

Growth-coupling via cofactor oxidation or reduction rates represents a major opportunity for engineering the biocatalysis of products which can't satisfy the requirements for other high-throughput enzyme engineering approaches. This includes the need for compartmentalization of products and their fast and accurate quantification, usually essential conditions to ensure that the sequence of an enzyme variant from a library can be linked to its performance. Considerable interest exists for the sustainable production of gasses, aliphatic hydrocarbons, and organic solvents, particularly in the context of decarbonisation and the production of sustainable platform chemicals. However, most of these products do not satisfy the above requirements, due to their inherent physical properties or simply a lack of suitable available methods. Cofactor-based growth coupling can remove significant roadblocks for engineering metabolic pathways or enzymes involved in producing these chemicals, by providing *in vivo*

ultra-high throughput screening capacity where it may not have been possible using other techniques. Application of these growth coupling systems is not limited to engineering these specific types of enzymes however. Any cofactor-utilising enzyme or pathway can in theory be targeted for engineering using these platform strains, which could provide an accelerated pathway towards altering the properties of a wide variety of enzymes. Applications of these growth coupling platforms may include altering the cofactor specificity of relevant enzymes to optimize cofactor availability in cell factories, screening cofactor-utilising enzymes with libraries of substrates to uncover novel catalytic capabilities, and optimizing the expression levels of heterologous metabolic pathway enzymes by coupling pathway flux to growth.

Besides their use as enzyme engineering platforms, strains displaying growth dependence may prove useful as bioproduction hosts. For example, the use of antibiotics in bioproduction can be costly and lead to antibiotic resistant strains. Strains dependent on enzyme activity and/or exogenous substrate addition for growth will be selected for without requiring antibiotic selection pressure. In addition, the use of cofactor auxotrophic hosts could alleviate biosafety concerns, if growth were made dependent on specific culturing conditions and/or the supplementation of substrates only present in the bioproduction environment.

Finally, the use of non-canonical redox cofactors could enhance the efficient biosynthesis of products and remains a largely unexplored area of metabolic engineering. Yields and productivities of bioprocesses are often insufficient for commercial exploitation, in part because cell factories tend to divert considerable amounts of energy and carbon from substrates towards growth. High yield bioproduction may be achievable through tighter control over electron flux from substrate to product, which is challenging when using $NAD(P)^+$/NAD(P)H cofactors involved in many reactions. The development of non-canonical redox cofactor cycling within cell factories could allow for a closed-loop and controllable system of electron transfer (King et al., 2020, 2022; Weusthuis et al., 2020). Engineering relevant bioproduction enzymes towards non-canonical redox cofactor utilisation can be

accelerated through using strains auxotrophic for these cofactors, in conjunction with growth-coupling approaches.

7. Conclusion

The recent flurry of interest in redox cofactor auxotrophic strains showcases their broad applicability and potential for enzyme engineering. Growth-coupling is a powerful selection method, since it is ultra-high throughput, provides an affordable and simple method of detecting improved enzyme variants and allows for enzyme engineering directly within desired cellular environments. At least one auxotrophic *E. coli* strain exists for every redox state of the canonical $NAD(P)^+$/NAD(P)H cofactor pairs, and considerable progress has been made in the generation of strains auxotrophic for redox states of non-canonical redox cofactor pairs NMN^+/NMNH and NCD^+/NCDH. Taken together, these represent a set of highly versatile platforms for engineering redox cofactor-utilising enzymes. Several strains have been used for engineering a range of redox cofactor-dependent enzymes, often with considerable success after only one round of mutagenesis. Substrate specificity, catalytic rate and (thermo)stability are important properties for enzyme use in commercial applications, and redox cofactor auxotrophic strains have now been successfully used to enhance all three of these properties in a variety of enzymes. Further development of (non-canonical) redox cofactor auxotrophic strains is of great interest for accelerating the engineering of robust and efficient biocatalysts. This in turn is important for the overarching goal of developing highly efficient cell factories which can achieve commercially viable titres, yields and productivities for sustainable bioproduction.

Acknowledgements

This work was supported by funding from the Biotechnology and Biological Sciences Research Council (BBSRC) [grant number BB/M011224/1] and the Oxford Interdisciplinary Doctoral Training Partnership. We also thank the Engineering and Physical Sciences Research Council (EPSRC) for their support [grant number EP/M002403/1 and EP/N009746/1].

References

Aharoni, A., Thieme, K., Chiu, C. P. C., Buchini, S., Lairson, L. L., Chen, H., Strynadka, N. C. J., Wakarchuk, W. W., & Withers, S. G. (2006). High-throughput screening methodology for the directed evolution of glycosyltransferases. *Nature Methods 2006 3:8*, *3*(8), 609–614. https://doi.org/10.1038/nmeth899

Alter, T. B., & Ebert, B. E. (2019). Determination of growth-coupling strategies and their underlying principles. *BMC Bioinformatics*, *20*(1), 447. https://doi.org/10.1186/s12859-019-2946-7

Álvarez, B., Mencía, M., de Lorenzo, V., & Fernández, L. Á. (2020). In vivo diversification of target genomic sites using processive base deaminase fusions blocked by dCas9. *Nature Communications 2020 11:1*, *11*(1), 1–14. https://doi.org/10.1038/s41467-020-20230-z

Arnold, F. H. (1998a). Design by Directed Evolution. *Accounts of Chemical Research*, 31(3), 125–131. https://doi.org/10.1021/AR960017F

Arnold, F. H. (1998b). When blind is better: Protein design by evolution. *Nature Biotechnology* 1998 16:7, 16(7), 617–618. https://doi.org/10.1038/nbt0798-617

Aslan, S., Noor, E., & Bar-Even, A. (2017). Holistic bioengineering: Rewiring central metabolism for enhanced bioproduction. In *Biochemical Journal* (Vol. 474, Issue 23, pp. 3935–3950). Portland Press Ltd. https://doi.org/10.1042/BCJ20170377

Bastian, S., Liu, X., Meyerowitz, J. T., Snow, C. D., Chen, M. M. Y., & Arnold, F. H. (2011). Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in Escherichia coli. *Metabolic Engineering*, *13*(3), 345–352. https://doi.org/10.1016/J.YMBEN.2011.02.004

Becker, S., Schmoldt, H. U., Adams, T. M., Wilhelm, S., & Kolmar, H. (2004). Ultra-high-throughput screening based on cell-surface display and fluorescence-activated cell sorting for the identification of novel biocatalysts. *Current Opinion in Biotechnology*, *15*(4), 323–329. https://doi.org/10.1016/J.COPBIO.2004.06.001

Bengtsson, O., Hahn-Hägerdal, B., & Gorwa-Grauslund, M. F. (2009). Xylose reductase from Pichia stipitis with altered coenzyme preference improves ethanolic xylose fermentation by recombinant Saccharomyces cerevisiae. *Biotechnology for Biofuels*, *2*(1), 1–10. https://doi.org/10.1186/1754-6834-2-9

Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., & Rabinowitz, J. D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nature Chemical Biology 2009 5:8*, *5*(8), 593–599. https://doi.org/10.1038/nchembio.186

Black, W. B., Aspacio, D., Bever, D., King, E., Zhang, L., & Li, H. (2020). Metabolic engineering of Escherichia coli for optimized biosynthesis of nicotinamide mononucleotide, a noncanonical redox cofactor. *Microbial Cell Factories*, *19*(1), 1–10. https://doi.org/10.1186/S12934-020-01415-Z

Black, W. B., & Li, H. (2022). Cell-Free Noncanonical Redox Cofactor Systems. *Methods in Molecular Biology*, *2433*, 185–198. https://doi.org/10.1007/978-1-0716-1998-8_11

Black, W. B., Zhang, L., Mak, W. S., Maxel, S., Cui, Y., King, E., Fong, B., Sanchez Martinez, A., Siegel, J. B., & Li, H. (2019). Engineering a nicotinamide mononucleotide redox cofactor system for biocatalysis. *Nature Chemical Biology 2019 16:1*, *16*(1), 87–94. https://doi.org/10.1038/s41589-019-0402-7

Bornscheuer, U. T., Huisman, G. W., Kazlauskas, R. J., Lutz, S., Moore, J. C., & Robins, K. (2012). Engineering the third wave of biocatalysis. *Nature 2012 485:7397*, *485*(7397), 185–194. https://doi.org/10.1038/nature11117

Bouzon, M., Döring, V., Dubois, I., Berger, A., Stoffel, G. M. M., Ramirez, L. C., Meyer, S. N., Fouré, M., Roche, D., Perret, A., Erb, T. J., Bar-Even, A., & Lindner, S. N. (2021). Change in cofactor specificity of oxidoreductases by adaptive evolution of an escherichia coli nadph-auxotrophic strain. *MBio*, *12*(4). https://doi.org/10.1128/MBIO.00329-21

Brannigan, J. A., & Wilkinson, A. J. (2002). Protein engineering 20 years on. *Nature Reviews Molecular Cell Biology 2002 3:12*, *3*(12), 964–970. https://doi.org/10.1038/nrm975

Cahn, J. K. B., Werlang, C. A., Baumschlager, A., Brinkmann-Chen, S., Mayo, S. L., & Arnold, F. H. (2017). A General Tool for Engineering the NAD/NADP Cofactor Preference of Oxidoreductases. *ACS Synthetic Biology*, *6*(2), 326–333. https://doi.org/10.1021/acssynbio.6b00188

Callaway, E. (2022). 'The entire protein universe': AI predicts shape of nearly every known protein. *Nature*, *608*(7921), 15–16. https://doi.org/10.1038/D41586-022-02083-2

Calzadiaz-Ramirez, L., Calvó-Tusell, C., Stoffel, G. M. M., Lindner, S. N., Osuna, S., Erb, T. J., Garcia-Borràs, M., Bar-Even, A., & Acevedo-Rocha, C. G. (2020). In Vivo Selection for Formate Dehydrogenases with High Efficiency and Specificity toward NADP + . *ACS Catalysis*, *10*(14), 7512–7525. https://doi.org/10.1021/acscatal.0c01487

Centeno-Leija, S., Utrilla, J., Flores, N., Rodriguez, A., Gosset, G., & Martinez, A. (2013). Metabolic and transcriptional response of Escherichia coli with a NADP(+)-dependent glyceraldehyde 3-phosphate dehydrogenase from Streptococcus mutans. *Antonie van Leeuwenhoek*, *104*(6), 913–924. https://doi.org/10.1007/S10482-013-0010-6

Chemler, J. A., Fowler, Z. L., McHugh, K. P., & Koffas, M. A. G. (2010). Improving NADPH availability for natural product biosynthesis in Escherichia coli by metabolic engineering. *Metabolic Engineering*, *12*(2), 96–104. https://doi.org/10.1016/J.YMBEN.2009.07.003

Chen, B., Lim, S., Kannan, A., Alford, S. C., Sunden, F., Herschlag, D., Dimov, I. K., Baer, T. M., & Cochran, J. R. (2015). High-throughput analysis and protein engineering using microcapillary arrays. *Nature Chemical Biology 2015 12:2*, *12*(2), 76–81. https://doi.org/10.1038/nchembio.1978

Chen, J., Wang, Y., Zheng, P., & Sun, J. (2022). Engineering synthetic auxotrophs for growth-coupled directed protein evolution. *Trends in Biotechnology*. https://doi.org/10.1016/J.TIBTECH.2022.01.010

Chen, S. L., Liu, T. S., Zhang, W. G., & Xu, J. Z. (2022). An NADPH-auxotrophic Corynebacterium glutamicum recombinant strain and used it to construct L-leucine high-yielding strain. *International Microbiology*, 1–14. https://doi.org/10.1007/S10123-022-00270-9

Chin, J. W., & Cirino, P. C. (2011). Improved NADPH supply for xylitol production by engineered Escherichia coli with glycolytic mutations. *Biotechnology Progress*, *27*(2), 333–341. https://doi.org/10.1002/BTPR.559

Christians, F. C., Scapozza, L., Crameri, A., Folkers, G., & Stemmer, W. P. C. (1999). Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling. *Nature Biotechnology 1999 17:3*, *17*(3), 259–264. https://doi.org/10.1038/7003

Claassens, N. J., Burgener, S., Vögeli, B., Erb, T. J., & Bar-Even, A. (2019). A critical comparison of cellular and cell-free bioproduction systems. *Current Opinion in Biotechnology*, *60*, 221–229. https://doi.org/10.1016/J.COPBIO.2019.05.003

Coco, W. M., Levinson, W. E., Crist, M. J., Hektor, H. J., Darzins, A., Pienkos, P. T., Squires, C. H., & Monticello, D. J. (2001). DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nature Biotechnology 2001 19:4*, *19*(4), 354–359. https://doi.org/10.1038/86744

Colin, P. Y., Zinchenko, A., & Hollfelder, F. (2015). Enzyme engineering in biomimetic compartments. *Current Opinion in Structural Biology*, *33*, 42–51. https://doi.org/10.1016/J.SBI.2015.06.001

De Graef, M. R., Alexeeva, S., Snoep, J. L., & Teixeira De Mattos, M. J. (1999). The steady-state internal redox state (NADH/NAD) reflects the external redox state and is correlated with catabolic adaptation in Escherichia coli. *Journal of Bacteriology*, *181*(8), 2351–2357. https://doi.org/10.1128/JB.181.8.2351-2357.1999

Dörr, M., Fibinger, M. P. C., Last, D., Schmidt, S., Santos-Aberturas, J., Böttcher, D., Hummel, A., Vickers, C., Voss, M., & Bornscheuer, U. T. (2016). Fully automatized high-throughput enzyme library screening using a robotic platform. *Biotechnology and Bioengineering*, *113*(7), 1421–1432. https://doi.org/10.1002/BIT.25925

Golynskiy, M. V., Haugner, J. C., Morelli, A., Morrone, D., & Seelig, B. (2013). In vitro evolution of enzymes. *Methods in Molecular Biology (Clifton, N.J.)*, *978*, 73. https://doi.org/10.1007/978-1-62703-293-3_6

Gupta, S. K., & Shukla, P. (2015). Advanced technologies for improved expression of recombinant proteins in bacteria: perspectives and applications. *Https://Doi.Org/10.3109/07388551.2015.1084264*, *36*(6), 1089–1098. https://doi.org/10.3109/07388551.2015.1084264

Halperin, S. O., Tou, C. J., Wong, E. B., Modavi, C., Schaffer, D. V., & Dueber, J. E. (2018). CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature 2018 560:7717*, *560*(7717), 248–252. https://doi.org/10.1038/s41586-018-0384-8

Huang, P. S., Boyken, S. E., & Baker, D. (2016). The coming of age of de novo protein design. *Nature 2016 537:7620*, *537*(7620), 320–327. https://doi.org/10.1038/nature19946

Jang, Y. S., Kim, B., Shin, J. H., Choi, Y. J., Choi, S., Song, C. W., Lee, J., Park, H. G., & Lee, S. Y. (2012). Bio-based production of C2-C6 platform chemicals. *Biotechnology and Bioengineering*, *109*(10), 2437–2459. https://doi.org/10.1002/BIT.24599

Johannes, T. W., Woodyer, R. D., & Zhao, H. (2005). Directed evolution of a thermostable phosphite dehydrogenase for NAD(P)H regeneration. *Applied and Environmental Microbiology*, *71*(10), 5728–5734. https://doi.org/10.1128/AEM.71.10.5728-5734.2005

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature 2021 596:7873*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kabus, A., Georgi, T., Wendisch, V. F., & Bott, M. (2007). Expression of the Escherichia coli pntAB genes

encoding a membrane-bound transhydrogenase in Corynebacterium glutamicum improves L-lysine formation. *Applied Microbiology and Biotechnology*, *75*(1), 47–53. https://doi.org/10.1007/S00253-006-0804-9

Kaur, J., Kumar, A., & Kaur, J. (2018). Strategies for optimization of heterologous protein expression in *E. coli*: Roadblocks and reinforcements. *International Journal of Biological Macromolecules*, *106*, 803–822. https://doi.org/10.1016/J.IJBIOMAC.2017.08.080

King, E., Cui, Y., Aspacio, D., Nicklen, F., Zhang, L., Maxel, S., Luo, R., Siegel, J. B., Aitchison, E., & Li, H. (2022). Engineering Embden–Meyerhof–Parnas Glycolysis to Generate Noncanonical Reducing Power. *ACS Catalysis*, 8582–8592. https://doi.org/10.1021/ACSCATAL.2C01837

King, E., Maxel, S., & Li, H. (2020). Engineering natural and noncanonical nicotinamide cofactor-dependent enzymes: design principles and technology development. *Current Opinion in Biotechnology*, *66*, 217–226. https://doi.org/10.1016/J.COPBIO.2020.08.005

Kramer, L., Le, X., Rodriguez, M., Wilson, M. A., Guo, J., & Niu, W. (2020). Engineering Carboxylic Acid Reductase (CAR) through a Whole-Cell Growth-Coupled NADPH Recycling Strategy. *ACS Synthetic Biology*, *9*(7), 1632–1637. https://doi.org/10.1021/ACSSYNBIO.0C00290

Lee, S. Y., Kim, H. U., Chae, T. U., Cho, J. S., Kim, J. W., Shin, J. H., Kim, D. I., Ko, Y. S., Jang, W. D., & Jang, Y. S. (2019). A comprehensive metabolic map for production of bio-based chemicals. *Nature Catalysis 2019 2:1*, *2*(1), 18–33. https://doi.org/10.1038/s41929-018-0212-4

Legon, L., Corre, C., Bates, D. G., & Mannan, A. A. (2022). gcFront: a tool for determining a Pareto front of growth-coupled cell factory designs. *Bioinformatics*. https://doi.org/10.1093/BIOINFORMATICS/BTAC376

Li, M., Ho, P. Y., Yao, S., & Shimizu, K. (2006). Effect of lpdA gene knockout on the metabolism in Escherichia coli based on enzyme activities, intracellular metabolite concentrations and metabolic flux analysis by 13C-labeling experiments. *Journal of Biotechnology*, *122*(2), 254–266. https://doi.org/10.1016/J.JBIOTEC.2005.09.016

Liang, K., & Shen, C. R. (2017). Selection of an endogenous 2,3-butanediol pathway in Escherichia coli by fermentative redox balance. *Metabolic Engineering*, *39*, 181–191. https://doi.org/10.1016/J.YMBEN.2016.11.012

Liao, J. C., Mi, L., Pontrelli, S., & Luo, S. (2016). Fuelling the future: microbial engineering for the production of sustainable biofuels. *Nature Reviews Microbiology 2016 14:5*, *14*(5), 288–304. https://doi.org/10.1038/nrmicro.2016.32

Lin, S.-J., & Guarente, L. (2003). Nicotinamide adenine dinucleotide, a metabolic regulator of transcription, longevity and disease. *Current Opinion in Cell Biology*, *15*, 241–246. https://doi.org/10.1016/S0955-0674(03)00006-1

Lindner, S. N., Calzadiaz Ramirez, L., Krü, J. L., Yishai, O., Belkhelfa, S., He, H., Bouzon, M., Dö, V., & Bar-Even, A. (2018). *NADPH-Auxotrophic E. coli: A Sensor Strain for Testing in Vivo Regeneration of NADPH*. https://doi.org/10.1021/acssynbio.8b00313

Liu, J., Li, H., Zhao, G., Caiyin, Q., & Qiao, J. (2018). Redox cofactor engineering in industrial microorganisms: strategies, recent applications and future directions. *Journal of Industrial Microbiology and Biotechnology*, *45*(5), 313–327. https://doi.org/10.1007/S10295-018-2031-7

Liu, Y., Feng, Y., Wang, L., Guo, X., Liu, W., Li, Q., Wang, X., Xue, S., & Zhao, Z. K. (2019). Structural Insights into Phosphite Dehydrogenase Variants Favoring a Non-natural Redox Cofactor. *ACS Catalysis*, *9*(3), 1883–1887. https://doi.org/10.1021/ACSCATAL.8B04822

Liu, Y., Li, Z., Guo, X., Wang, X., Zhao, Z. K., Chloramphenicol, C., & Kanamycin, K. (2022). A phosphite-based screening platform for identification of enzymes favoring nonnatural cofactors. *Scientific Reports 2022 12:1*, *12*(1), 1–8. https://doi.org/10.1038/s41598-022-16599-0

Lobstein, J., Emrich, C. A., Jeans, C., Faulkner, M., Riggs, P., & Berkmen, M. (2012). SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories*, *11*(1), 1–16. https://doi.org/10.1186/1475-2859-11-56

Longwell, C. K., Labanieh, L., & Cochran, J. R. (2017). High-throughput screening technologies for enzyme engineering. *Current Opinion in Biotechnology*, *48*, 196–202. https://doi.org/10.1016/J.COPBIO.2017.05.012

Lutz, S. (2010). Beyond directed evolution - semi-rational protein engineering and design. *Current Opinion in Biotechnology*, *21*(6), 734. https://doi.org/10.1016/J.COPBIO.2010.08.011

MacBeath, G., Kast, P., & Hilvert, D. (1998). Redesigning enzyme topology by directed evolution. *Science*, *279*(5358), 1958–1961. https://doi.org/10.1126/SCIENCE.279.5358.1958

Machado, H. B., Dekishima, Y., Luo, H., Lan, E. I., & Liao, J. C. (2012). A selection platform for carbon chain elongation using the CoA-dependent pathway to produce linear higher alcohols. *Metabolic Engineering*, *14*(5), 504–511. https://doi.org/10.1016/J.YMBEN.2012.07.002

Martínez, I., Zhu, J., Lin, H., Bennett, G. N., & 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. https://doi.org/10.1016/j.ymben.2008.09.001

Maxel, S., Aspacio, D., King, E., Zhang, L., Acosta, A. P., & Li, H. (2020). A Growth-Based, High-Throughput Selection Platform Enables Remodeling of 4-Hydroxybenzoate Hydroxylase Active Site. *ACS Catalysis*, *10*(12), 6969–6974. https://doi.org/10.1021/ACSCATAL.0C01892

Maxel, S., King, E., Zhang, Y., Luo, R., & Li, H. (2020). Leveraging Oxidative Stress to Regulate Redox Balance-Based, In Vivo Growth Selections for Oxygenase Engineering. *Cite This: ACS Synth. Biol*, *2020*, 3133. https://doi.org/10.1021/acssynbio.0c00380

Maxel, S., Saleh, S., King, E., Aspacio, D., Zhang, L., Luo, R., & Li, H. (2021). Growth-Based, High-Throughput Selection for NADH Preference in an Oxygen-Dependent Biocatalyst. *ACS Synthetic Biology*, *10*. https://doi.org/10.1021/ACSSYNBIO.1C00258

Maxel, S., Zhang, L., King, E., Acosta, A. P., Luo, R., & Li, H. (2020). In Vivo, High-Throughput Selection of Thermostable Cyclohexanone Monooxygenase (CHMO). *Catalysts*, *10*(8), 935. https://doi.org/10.3390/catal10080935

Mccullum, E. O., Williams, B. a R., Zhang, J., & Chaput, J. C. (2010). Error Based PCR Mutagenesis Protocols. *Methods in Molecular Biology*, *634*(3), 103–109. http://link.springer.com/10.1007/978-1-60761-652-8

Metcalf, W. W., Wanner, B. L., Chen, C.-M., Ye, Q.-Z., Zhu, Z., Wanner, B. L., Walsh, C. T., & Biol, J. (1991). Involvement of the Escherichia coli phn (psiD) Gene Cluster in Assimilation of Phosphorus in the Form of Phosphonates, Phosphite, Pi Esters, and Pi. *JOURNAL OF BACTERIOLOGY*, *173*(2), 587–600.

MohammadiPeyhani, H., Hafner, J., Sveshnikova, A., Viterbo, V., & Hatzimanikatis, V. (2022). Expanding biochemical knowledge and illuminating metabolic dark matter with ATLASx. *Nature Communications 2022 13:1*, *13*(1), 1–12. https://doi.org/10.1038/s41467-022-29238-z

Opitz, C. A., & Heiland, I. (2015). Dynamics of NAD-metabolism: everything but constant. *Biochemical Society Transactions*, *43*(6), 1127–1132. https://doi.org/10.1042/BST20150133

Orsi, E., Claassens, N. J., Nikel, P. I., & Lindner, S. N. (2021). Growth-coupled selection of synthetic modules to accelerate cell factory development. *Nature Communications 2021 12:1*, *12*(1), 1–5. https://doi.org/10.1038/s41467-021-25665-6

Overton, T. W. (2014). Recombinant protein production in bacterial hosts. *Drug Discovery Today*, *19*(5), 590–601. https://doi.org/10.1016/J.DRUDIS.2013.11.008

Packer, M. S., & Liu, D. R. (2015). Methods for the directed evolution of proteins. *Nature Reviews Genetics 2015 16:7*, *16*(7), 379–394. https://doi.org/10.1038/nrg3927

Park, H., & Kim, S. (2021). Gene-specific mutagenesis enables rapid continuous evolution of enzymes in vivo. *Nucleic Acids Research*, *49*(6), e32–e32. https://doi.org/10.1093/NAR/GKAA1231

Rasor, B. J., Vögeli, B., Landwehr, G. M., Bogart, J. W., Karim, A. S., & Jewett, M. C. (2021). Toward sustainable, cell-free biomanufacturing. *Current Opinion in Biotechnology*, *69*, 136–144. https://doi.org/10.1016/J.COPBIO.2020.12.012

Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A., & Liu, C. C. (2018). Scalable, continuous evolution of genes at mutation rates above genomic error thresholds. *Cell*, *175*(7), 1946. https://doi.org/10.1016/J.CELL.2018.10.021

Reetz, M. T., & Carballeira, J. D. (2007). Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nature Protocols 2007 2:4*, *2*(4), 891–903. https://doi.org/10.1038/nprot.2007.72

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology*, *5*(APR), 172. https://doi.org/10.3389/FMICB.2014.00172

Sauer, U., Canonaco, F., Heri, S., Perrenoud, A., & Fischer, E. (2004). The Soluble and Membrane-bound Transhydrogenases UdhA and PntAB Have Divergent Functions in NADPH Metabolism of Escherichia coli*. *Journal of Biological Chemistry*, *279*(8), 6613–6619. https://doi.org/10.1074/JBC.M311657200

Scown, C. D., & Keasling, J. D. (2022). Sustainable manufacturing with synthetic biology. *Nature Biotechnology 2022 40:3*, *40*(3), 304–307. https://doi.org/10.1038/s41587-022-01248-8

Sekar, B. S. , Seol, E., & Park, S. (2017). Co-production of hydrogen and ethanol from glucose in Escherichia coli by activation of pentose-phosphate pathway through deletion of phosphoglucose isomerase (pgi) and overexpression of glucose-6-phosphate dehydrogenase (zwf) and 6-phosphogluconate dehyd. *Biotechnology for Biofuels*, *10*(1), 1–12. https://doi.org/10.1186/S13068-017-0768-2

Sellés Vidal, L., Murray, J. W., & Heap, J. T. (2021). Versatile selective evolutionary pressure using synthetic defect in universal metabolism. *Nature Communications 2021 12:1*, *12*(1), 1–15. https://doi.org/10.1038/s41467-021-27266-9

Seta, F. Della, Boschi-Muller, S., Vignais, M. L., & Branlant, G. (1997). Characterization of Escherichia coli Strains with gapA and gapB Genes Deleted. In *JOURNAL OF BACTERIOLOGY* (Vol. 179, Issue 16).

Shen, C. R., Lan, E. I., Dekishima, Y., Baez, A., Cho, K. M., & Liao, J. C. (2011). Driving forces enable high-titer anaerobic 1-butanol synthesis in Escherichia coli. *Applied and Environmental Microbiology*, *77*(9), 2905–2915. https://doi.org/10.1128/AEM.03034-10

Simon, A. J., d'Oelsnitz, S., & Ellington, A. D. (2019). Synthetic evolution. *Nature Biotechnology 2019 37:7*, *37*(7), 730–743. https://doi.org/10.1038/s41587-019-0157-4

Spaans, S. K., Weusthuis, R. A., van der Oost, J., & Kengen, S. W. M. (2015). NADPH-generating systems in bacteria and archaea. In *Frontiers in Microbiology* (Vol. 6, Issue JUL, p. 742). Frontiers Research Foundation. https://doi.org/10.3389/fmicb.2015.00742

Takeno, S., Hori, K., Ohtani, S., Mimura, A., Mitsuhashi, S., & Ikeda, M. (2016). l-Lysine production independent of the oxidative pentose phosphate pathway by Corynebacterium glutamicum with the Streptococcus mutans gapN gene. *Metabolic Engineering*, *37*, 1–10. https://doi.org/10.1016/J.YMBEN.2016.03.007

Tamakawa, H., Ikushima, S., & Yoshida, S. (2011). Ethanol Production from Xylose by a Recombinant Candida utilis Strain Expressing Protein-Engineered Xylose Reductase and Xylitol Dehydrogenase. *Bioscience, Biotechnology, and Biochemistry*, *75*(10), 1994–2000. https://doi.org/10.1271/BBB.110426

Tawfik, D. S., & Griffiths, A. D. (1998). Man-made cell-like compartments for molecular evolution. *Nature Biotechnology*, *16*. http://www.nature.com/naturebiotechnology

Theodosiou, E., Tüllinghoff, A., Toepel, J., & Bühler, B. (2022). Exploitation of Hetero- and Phototrophic Metabolic Modules for Redox-Intensive Whole-Cell Biocatalysis. *Frontiers in Bioengineering and Biotechnology*, *10*, 462. https://doi.org/10.3389/FBIOE.2022.855715

Volke, D. C., Martino, R. A., Kozaeva, E., Smania, A. M., & Nikel, P. I. (2022). Modular (de)construction of complex bacterial phenotypes by CRISPR/nCas9-assisted, multiplex cytidine base-editing. *Nature Communications 2022 13:1*, *13*(1), 1–14. https://doi.org/10.1038/s41467-022-30780-z

Von Kamp, A., & Klamt, S. (2017). Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms. *Nature Communications 2017 8:1*, *8*(1), 1–10. https://doi.org/10.1038/ncomms15956

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., & Church, G. M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature 2009 460:7257*, *460*(7257), 894–898. https://doi.org/10.1038/nature08187

Wang, X., Feng, Y., Guo, X., Wang, Q., Ning, S., Li, Q., Wang, J., Wang, L., & Zhao, Z. K. (2021). Creating enzymes and self-sufficient cells for biosynthesis of the non-natural cofactor nicotinamide cytosine dinucleotide. *Nature Communications 2021 12:1*, *12*(1), 1–9. https://doi.org/10.1038/s41467-021-22357-z

Wang, Y., San, K. Y., & Bennett, G. N. (2013). Cofactor engineering for advancing chemical biotechnology. *Current Opinion in Biotechnology*, *24*(6), 994–999. https://doi.org/10.1016/J.COPBIO.2013.03.022

Wenk, S., Schann, K., He, H., Rainaldi, V., Kim, S., Lindner, S. N., & Bar-Even, A. (2020). An "energy-auxotroph" *Escherichia coli* provides an in vivo platform for assessing NADH regeneration systems. *Biotechnology and Bioengineering*, *117*(11), 3422–3434. https://doi.org/10.1002/bit.27490

Weusthuis, R. A., Folch, P. L., Pozo-Rodríguez, A., & Paul, C. E. (2020). Applying Non-canonical Redox Cofactors in Fermentation Processes. In *iScience* (Vol. 23, Issue 9, p. 101471). Elsevier Inc. https://doi.org/10.1016/j.isci.2020.101471

Xiao, H., Bao, Z., & Zhao, H. (2015). High throughput screening and selection methods for directed enzyme evolution. *Industrial and Engineering Chemistry Research*, *54*(16), 4011–4020. https://doi.org/10.1021/IE503060A

Yamauchi, Y., Hirasawa, T., Nishii, M., Furusawa, C., & Shimizu, H. (2014). Enhanced acetic acid and succinic acid production under microaerobic conditions by Corynebacterium glutamicum harboring Escherichia coli transhydrogenase gene pntAB. *The Journal of General and Applied Microbiology*, *60*(3), 112–118. https://doi.org/10.2323/JGAM.60.112

Yang, G., & Withers, S. G. (2009). Ultrahigh-throughput FACS-based screening for directed enzyme evolution. *ChemBioChem*, *10*(17), 2704–2715. https://doi.org/10.1002/cbic.200900384

Yun, M., Park, C. G., Kim, J. Y., & Park, H. W. (2000). Structural Analysis of Glyceraldehyde 3-Phosphate Dehydrogenase from Escherichia coli: Direct Evidence of Substrate Binding and Cofactor-Induced Conformational Changes†,‡. *Biochemistry*, *39*(35), 10702–10710. https://doi.org/10.1021/BI9927080

Zhang, L., King, E., Black, W. B., Heckmann, C. M., Wolder, A., Cui, Y., Nicklen, F., Siegel, J. B., Luo, R., Paul, C. E., & Li, H. (2022). Directed evolution of phosphite dehydrogenase to cycle noncanonical redox cofactors via universal growth selection platform. *Nature Communications 2022 13:1*, *13*(1), 1–12. https://doi.org/10.1038/s41467-022-32727-w

Zhang, L., King, E., Luo, R., & Li, H. (2018). Development of a High-Throughput, in Vivo Selection Platform for NADPH-Dependent Reactions Based on Redox Balance Principles. *ACS Synthetic Biology*, *7*(7), 1715–1721. https://doi.org/10.1021/acssynbio.8b00179

Zhang, X., Jantama, K., Moore, J. C., Shanmugam, K. T., & Ingram, L. O. (2007). Production of L-alanine by metabolically engineered Escherichia coli. *Applied Microbiology and Biotechnology*, *77*(2), 355–366. https://doi.org/10.1007/S00253-007-1170-Y

Zheng, L., Baumann, U., & Reymond, J. L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Research*, *32*(14), e115–e115. https://doi.org/10.1093/NAR/GNH110

Zhou, Y. J., Kerkhoven, E. J., & Nielsen, J. (2018). Barriers and opportunities in bio-based production of hydrocarbons. *Nature Energy 2018 3:11*, *3*(11), 925–935. https://doi.org/10.1038/s41560-018-0197-x

Zhou, Y., Wang, L., Yang, F., Lin, X., Zhang, S., & Zhao, Z. K. (2011). Determining the extremes of the cellular NAD(H) level by using an Escherichia coli NAD +-auxotrophic mutant. *Applied and Environmental Microbiology*, *77*(17), 6133–6140. https://doi.org/10.1128/AEM.00630-11

3. Construction and characterization of MoClo-compatible vectors for modular exploration of the protein expression space in *E. coli*

3.1 Abstract

Cloning methods are the backbone of modern synthetic biology research. The ability to generate custom DNA constructs with predictable protein expression capacity is fundamental to the engineering of biology. A popular technique is Golden Gate cloning, a form of modular cloning that enables one-pot rapid and reliable assembly of genetic parts. In this study, we expand on the existing MoClo toolkits by adding compatible low- (p15A) and medium-copy (pBR322) destination vectors. Together with existing high-copy vectors, these additional plasmids enable a maximum protein expression range covering a 500-fold difference in normalised fluorescence output. We further characterise the expression- and burden profiles of each vector and demonstrate their use for the optimization of growth-coupled enzyme expression. The vectors developed in this work are publicly available via Addgene.

3.2 Introduction

The field of synthetic biology is built upon the ability to manipulate DNA sequences rapidly and reliably. The availability of these genetic toolkits for predictably manipulating protein expression levels has allowed biology to become increasingly engineered. Engineering genetic circuitry relies on the correct assembly of functional genetic regions including promoters, ribosome binding sites (RBS), coding DNA sequences (CDS) and terminators, allowing the construction of custom DNA sequences with designed functionalities.

Effective DNA assembly methods are essential tools for any synthetic biologist. These assembly techniques include traditional restriction enzyme cloning, such as blunt-end or sticky-end cloning, and seamless cloning methods such as Gibson assembly (Gibson et al.,

2009). Although cloning using restriction enzymes has been common practice for decades, this technique was modernised with the conception of BioBricks (Knight, 2003). This method relied on the use of a set of standardised restriction sites to enable the modular assembly of characterised DNA fragments. The standardised assembly methods, modular parts repositories, and most importantly the widespread adoption by the synthetic biology community represented a leap forward in the field. However, BioBricks suffered from the same technical limitations as traditional restriction enzyme cloning, most notably low efficiencies with complex assemblies.

In 2008, a new DNA assembly method coined Golden Gate Assembly (GGA) was introduced, which was based on Type IIS restriction enzymes displaying restriction activity distal from their own recognition site (Engler et al., 2008). This activity removes the enzyme's own recognition site after restriction, a principle that can be exploited to drastically improve assembly efficiencies by cycling between restriction- and ligation steps to form increasing correctly assembled product within a reaction mix (Engler et al., 2008; Fromme & Klingenspor, 2007; Kotera & Nagai, 2008). Indeed, correct assemblies with up to 52 parts covering 40 kb have been reported using GGA (Pryor et al., 2022). In addition, GGA allows the use of both circular- and linear DNA as parts simultaneously, expanding the method's flexibility. The most commonly used Type IIS enzymes (BsaI, BbsI, BsmBI/BpiI) leave 4 bp overhangs, but a kit using SapI which leaves 3 bp overhangs has also been developed (Taylor et al., 2019).

Since the inception of GGA, many parts toolkits have been characterised and made publicly available for use in plants (Engler et al., 2014; Sarrion-Perdigones et al., 2013), yeast (Lee et al., 2015), fungi (Hernanz-Koers et al., 2018), cyanobacteria (Vasudevan et al., 2019) and those characterised more specifically for *Vibrio natriegens* (Stukenberg et al., 2021), *Pseudomonas putida* (Lammens et al., 2022) and *E. coli* (Andreou & Nakayama, 2018; Iverson et al., 2016; Moore et al., 2016; Taylor et al., 2019). Efforts to establish unified cross-genus or cross-kingdom GGA cloning standards are also underway (Blázquez et al., 2023; Chiasson et al., 2019; Valenzuela-Ortega & French, 2020). For *E. coli*, the CIDAR MoClo kit was the first

modular Golden Gate DNA parts library made publicly available via Addgene (Kit #1000000059) and it has been widely used since (Iverson et al., 2016). This kit contained a series of promoters and RBSs from the Anderson collection, as well as several inducible promoter parts, their repressors, and CDS parts coding for fluorescent proteins. All parts and destination vectors in the CIDAR MoClo kit were constructed using vectors with high-copy pUC19-derived origins of replication, but the kit lacked medium- or low copy destination vectors. For the purpose of expressing whole metabolic pathways or large genetic constructs, high copy vectors can become burdensome or lethal to *E. coli*, limiting the use of the kit. A variety of different plasmid backbones can be hosted by *E. coli*, exhibiting variable plasmid copy numbers based on their mode of replication or the presence of mutations (Jahn et al., 2016; Shao et al., 2021).

In this report, we describe the construction and characterization of a MoClo-compatible vector kit for use in *E. coli*. We generated vectors for level 1 assembly carrying either p15A- (low) or pBR322-derived (medium) origins of replication with either kanamycin or chloramphenicol resistance. Level 2 destination vectors were designed with ampicillin resistance and also carry either p15A- or pBR322-derived origins of replication. Since the p15A origin of replication is compatible with the pMB1-based pBR322- or pUC19 origins, this vector set enables the modular construction of dual plasmid systems in *E. coli*. Combined with the existing pUC19 vector, the plasmids in this kit enable an exploration of the protein expression space covering a 500-fold difference in normalised GFP expression. In addition, the use of these plasmids for optimizing enzyme expression in growth-coupled selection strategies was demonstrated. We expect this kit could be of use to synthetic biologists using Golden Gate cloning in *E. coli*, especially for the expression of metabolic pathways, burdensome constructs, or dual-plasmid systems. The plasmids constructed in this work are publicly available on Addgene (catalogue #217582-217609).

3.3 Materials and methods

3.3.1 Strains, plasmids, media, and growth conditions

All plasmids used and constructed in this work are available in Table S3.1. All primers used in this study can be found in Table S3.2 and synthesised DNA used for vector construction as detailed later can be found in Table S3.3. All routine cloning was performed using *E. coli* DH5α (NEB). Transformations were routinely performed via heatshock, and routine growth took place using LB medium, 250 rpm, 37°C. Cells were supplemented with appropriate antibiotics where required: carbenicillin (carb, 100 µg/mL); chloramphenicol (cm, 25 µg/mL); kanamycin (kan, 50 µg/mL). For blue-white screening plates were supplemented with 100 µg/mL 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal, Sigma) and 100 µM isopropyl β-D-1-thiogalactopyranoside (IPTG, Sigma). Modified minimal M9 medium was used as indicated in the text, comprised of 47.8 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 18.7 mM NH_4Cl, 2 mM $MgSO_4$, 100 µM $CaCl_2$, 134 µM ethylene diamine tetrachloroacetic acid (EDTA), 31 µM $FeCl_3 \cdot 6H_2O$, 6.2 µM $ZnCl_2$, 0.76 µM $CuCl_2 \cdot 2H_2O$, 0.42 µM $CoCl_2 \cdot 2H_2O$, 1.62 µM H_3BO_3, and 0.081 µM $MnCl_2 \cdot 4H_2O$, adjusted to pH 7.0. Filter sterilised D-glucose (1% w/v) was added as a carbon source (Sigma).

Table 3.1. Strains used in this work.

Strain	Genotype	Reference
E. coli DH5α	*fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	Purchased (New England Biolabs)
BW25113	*F- LAM- rrnB3 ΔlacZ4787 hsdR514 Δ(araBAD)567 Δ(rhaBAD)568 rph-1*	Purchased (DSMZ)
BW25113 Δ*adhE*	BW25113 Δ*adhE::FRT*	This study

3.3.2 Assembly of the medium copy MoClo destination vectors

Resistance markers (chloramphenicol-, kanamycin- and ampicillin), origins of replication (p15A and pBR322) and the expression construct for lacZα were synthesised as double strand DNA fragments (Twist, USA) with BsmBI-digestible flanks (Table S3.3). Kanamycin- and ampicillin markers were synthesised identically to those found in DVA and DVK vectors, while

the chloramphenicol marker was synthesised based on the sequence found in pKD3 but modified to carry a silent G435T mutation for domestication. Short DNA oligos introducing MoClo compatible overhangs A-H and containing the required BsaI and BbsI recognition sites were designed as two complementary oligos with BsmBI-digestible flanks (Table S3.2, JRN413-440). The oligo design differed depending on their use for level 1 or level 0/2 constructs, since BsaI and BbsI recognition sites had to be inverted accordingly. The synthesised fragments and oligos were designed to be assembled using GGA, with complementary overhangs between resistance marker- and ori fragments (GATT, 'X'), resistance marker fragment and oligo (TGGA, 'T'), ori fragment and oligo (TTCT, 'U'), oligo and one flank of the lacZα fragment (CCTG, 'V'), and oligo with the other flank of the lacZα fragment (GGGT, 'S'). These oligos were annealed by combining 100 µM of each primer in 50 µL deionised H_2O and heating to 98°C for 5 minutes, followed by cooling to 25°C at 0.1°C/s. Destination vectors were generated by combining 20 fmol of the antibiotic resistance marker, origin of replication and lacZα expression fragments, in addition to 1 µL of each 10x-diluted MoClo overhang fragment. Golden Gate assemblies were performed in a total of 20 µL with 2 µL of 10x T4 DNA ligase buffer (NEB), 10 units of BsmBI-HF (NEB) and 500 units of T4 DNA ligase (NEB). Reaction mixtures were cycled between incubations for 5 minutes at 42°C and 5 minutes at 16°C a total of 80 cycles in a thermocycler, with two final 10-minute steps at 42°C and 80°C before cooling to 4°C. The assembly mixture was transformed into *E. coli* DH5α and correct assemblies were identified by the formation of blue *E. coli* colonies on LB supplemented with the required antibiotic, X-gal and IPTG. Constructs were verified by Sanger sequencing (Eurofins, UK) over the MoClo overhang sites after plasmid extraction (NEB), using primers JRN273 and JRN274.

3.3.3 Routine Golden Gate assemblies

Routine Golden Gate cloning was performed mostly as described earlier (Iverson et al., 2016). Briefly, for level 1 assemblies including the GFP constructs used in this study, 20 fmol of destination vector was mixed with 40 fmol of each relevant part consisting of a promoter, RBS,

CDS, and terminator. All parts can be found in Table S3.1. Enzyme/buffer conditions and cycling conditions were identical to those used for the destination vector assemblies, except BsmBI-HF was replaced by BsaI-HFv2 (NEB) and a restriction temperature of 37ºC was used instead of 42ºC.

3.3.4 Gene deletion of *adhE* in *E. coli* BW25113

The deletion of *adhE* was performed by lambda red recombination mostly according to previous work (Jensen et al., 2015). Briefly, *E. coli* BW25113 was transformed with pSIJ8, encoding the lambda red recombination and flippase genes. Transformants were grown selectively overnight at 30°C and 1 mL of O/N culture was used to inoculate 50 mL of LB + carb in a shake flask. This culture was left to grow for 1 hour at 30ºC, followed by pSIJ8 induction for 45 minutes by addition of sterile 15 mM L-arabinose. Cells were made electrocompetent and finally resuspended in 300 µL ice-cold sterile 10% glycerol. DNA cassettes targeting *adhE* were generated by PCR amplification of pKD3 with JRN384 and JRN385 (Table S3.2). DNA was purified (PCR Cleanup Kit, NEB) and eluted in DI water. For electroporation of the marker, 50 µL of electrocompetent cells were mixed with 3.5 µL of marker (>250 ng/µL) and transferred to an ice-cold electroporation cuvette (0.1 cm gap). A 1.8 kV pulse was applied, and cells were recovered immediately for 2.5 hours at 30ºC before plating on LB + carb + cm (17 µg/mL) and incubated at 30ºC. Positive clones were identified by colony PCR of the *adhE* locus using primers JRN386 and JRN387. Integrated chloramphenicol markers were removed by inoculating 35 mL of LB + carb with 0.5 mL of overnight *E. coli* BW25113 *adhE::cmR* culture. After 1.5 hours shaking at 30°C, flippase expression was induced with 50 mM L-rhamnose for 4 hours before plating 10^{-5} and 10^{-6} dilutions on LB + carb plates. Strains with *adhE::FRT* genotypes (Δ*adhE*) were identified using the same *adhE* colony PCR primers and pSIJ8 was cured by a single overnight cultivation in LB at 42ºC.

3.3.5 Domestication of *adhE* and randomised *adhE** construct assembly

The *E. coli adhE* gene carries several endogenous BbsI recognition sites, which had to be removed prior to its use as a Golden Gate CDS part. Genomic DNA of *E. coli* MG1655 (carries an identical *adhE* nucleotide sequence to *E. coli* BW25113) was used as a template for amplification with primers JRN279+JRN280, JRN281+JRN282, JRN283+JRN284, JRN285+286 and JRN287+288 to generate five linear fragments of the *adhE* gene with silent mutations removing the internal BbsI recognition sites. All fragments were designed with overhangs including BbsI recognition sites to enable one-step assembly of the domesticated part into DVA_pUC19_CD, resulting in pL0_JRN012_CD. The A267T mutation was introduced by amplifying the domesticated level 0 part from pL0_JRN012_CD using primers JRN224+JRN225 and performing a Gibson assembly using an oligo 'bridge' (JRN226) between the two flanks, introducing the mutation and forming pL0_JRN017_CD (NEB HiFi assembly). The E568K mutation was introduced in the same way but using pL0_JRN017_CD as PCR template and using primers JRN189+JRN190 with oligo bridge JRN191, forming pL0_JRN018_CD. Correct assembly of A267T and/or E568K *adhE* variants was confirmed by Sanger sequencing (Eurofins, UK). For expression strength randomization, 6.67 fmol of DVK_pUC19_AE, DVK_pBR322_AE and DVK_p15A_AE were added to the Golden Gate reaction mix, along with 40 fmol of promoter part, RBS part, pL0_JRN018_CD, and terminator part (see text for promoter/RBS specification). Buffer/enzyme quantity and cycling conditions were identical to those described earlier in the Methods. The finalised assembly was transformed into *E. coli* DH5α, recovered for 1h and an aliquot plated on LB + IPTG + X-gal + kanamycin to assess library fidelity. The remaining recovered culture was washed twice in fresh LB to remove non-transformed DNA, resuspended in 5 mL of LB + kan and grown overnight for plasmid isolation the following day (Miniprep Kit, NEB).

3.3.6 Selection of *adhE** expression constructs

Three separate 50 µL aliquots of electrocompetent BW25113 Δ*adhE* were transformed by electroporation as described earlier with 1 µL of plasmid DNA library carrying randomised

*adhE** expression constructs. Each transformed culture was recovered immediately by incubating in LB at 37°C for 1 hour. After recovery, 50 µL of each replicate was plated on LB + kan to assess the diversity of the library pre-selection. The remaining recovered cells were washed three times in sterile 1x M9 salts and each replicate was used to inoculate a separate 250 mL flask containing 50 mL of M9 minimal medium supplemented with 1 mM glucose, 300 mM ethanol and 50 µg/mL kanamycin. The flasks were placed in a 37°C shaking incubator (225 rpm), and OD_{600} readings were taken daily. When OD_{600} values reached >1.0, 1 mL of grown culture was diluted 10^4- and 10^5-fold and 100 µL was plated on LB + kan plates. The grown culture was then passaged by diluting it 100-fold into fresh 50 mL of M9 medium supplemented as before, for two additional passages. Plasmid contents of clones were assessed by Sanger sequencing (Source Bioscience, UK) of amplified vector inserts (DreamTaq, ThermoFisher) using primers JRN273 and JRN274. Four colonies from each replicate population were sequenced at every step (pre-selection and passages 1, 2 and 3).

3.3.7 Microtiter plate assays

Microbial growth and fluorescence were assessed in a Tecan Spark plate reader (SparkControl v3.1 SP1 software), equipped with a humidity cassette. Assays were performed in 200 µL LB or M9 medium as indicated in the text and wells were sealed with BreatheEasy (Sigma-Aldrich) membranes to avoid contamination. Growth took place at 37°C and measured by absorbance at 600 nm while GFP fluorescence was assessed by monochromatic excitation at 480 nm and emission at 520 nm, using 30 flashes, a 20 nm bandwidth for both emission- and excitation, a 40 µs integration time, and a Z-position of 30 mm (measurement were taken from the top of the plate). A manual gain setting of 60 (A.U.) was used for fluorescence readings. Readings were taken every 15 minutes with agitation in between readings occurring in four separate stages: 1) 225 seconds, 3 mm amplitude, 90 rpm; 2) 225 seconds, 3 mm amplitude, 180 rpm; 3) 225 seconds, 2.5 mm amplitude, 108 rpm; 4) 225 seconds, 2.5 mm amplitude, 216 rpm. Growth rates were calculated through a log conversion of the blank-

medium corrected OD values, followed by determining the maximum linear slope of these values over time using a 6-12 hour sliding window per replicate.

3.4 Results

3.4.1 Expression level and burden assessment of the vector kit

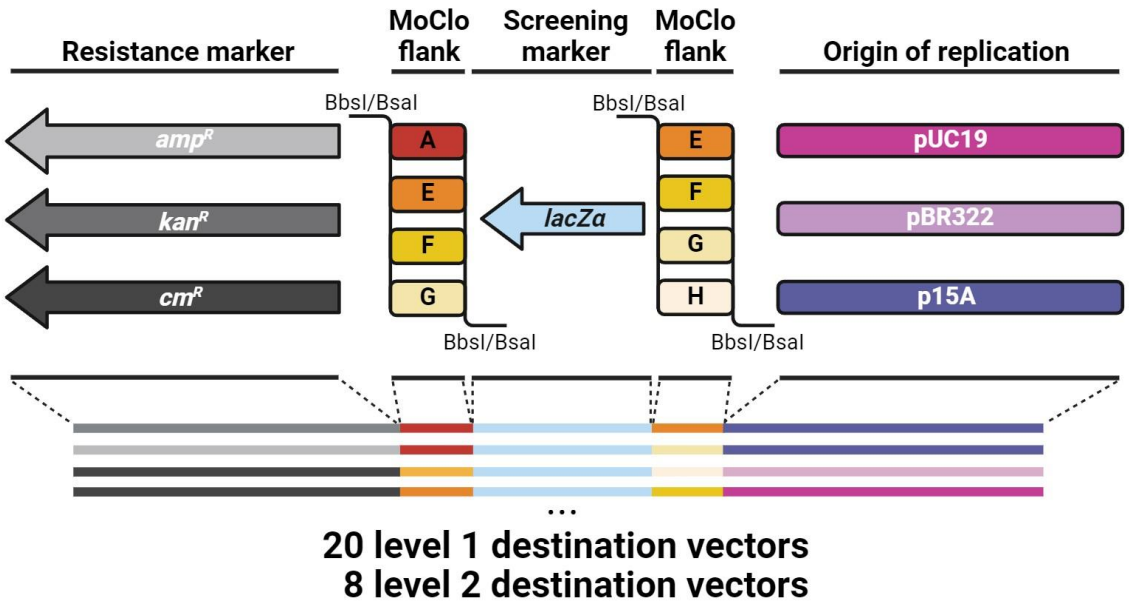


Figure 3.1. Constituent parts of the destination vectors generated in this study. Vectors were constructed from resistance markers (amp^R, kan^R, cm^R), MoClo-compatible flanks ('A', 'E', 'F', 'G' and 'H'), selectable marker (*lacZ*α) and origins of replication (pBR322 and p15A oris). These parts were assembled by GGA to form 20 new level 1 destination vectors and 8 new level 2 destination vectors to use for modular cloning.

The MoClo destination vectors were constructed using p15A- or pBR322 origins of replication to supplement already existing pUC19-based vectors. Ampicillin (level 0/2), kanamycin (level 1), or chloramphenicol (level 1) resistance markers were used for the new vectors. The overhang sequences were coded as follows: A (GGAG), E (GCTT), F (CGCT), G (TGCC) and H (ACTA). Level 1 destination vectors were generated with CIDAR MoClo compatible overhangs (A-E, A-F, E-F, F-G, G-H). Inclusion of an 'A-F' vector renders level 1 destination vectors compatible with the broader MoClo standard (Weber et al., 2011). Level 0/2 plasmids carried A-F, E-F, A-G and A-H overhangs. The nomenclature for these plasmids followed the DV[A/K/C]_p[15A/BR322/UC19]_[OVERHANG] standard, with A/K/C indicating antibiotic

resistance marker ampicillin/kanamycin/chloramphenicol (respectively), followed by the origin of replication, and finally the two overhang letter codes (e.g. AF) (see Table S3.1). In total, 28 new vectors were constructed with varying oris, MoClo overhangs, and markers (Figure 3.1).

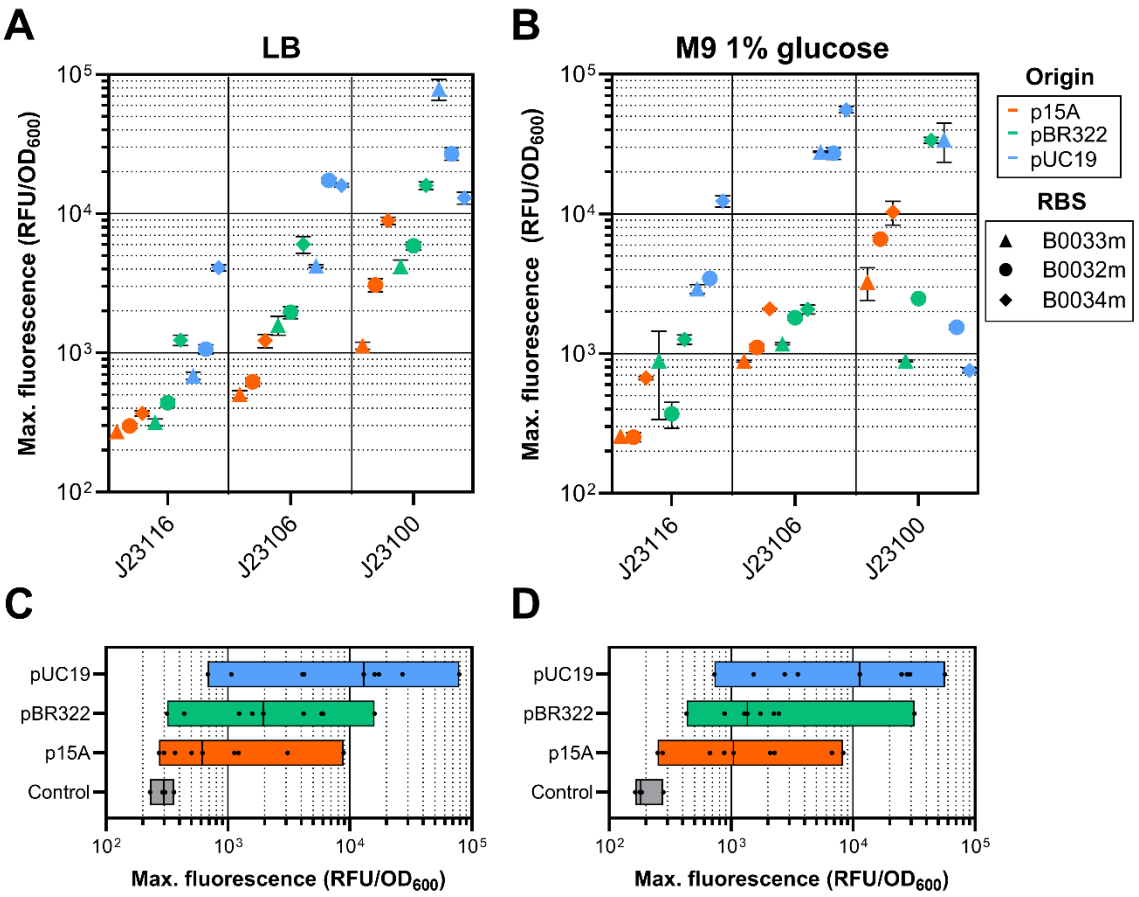


Figure 3.2. Fluorescence readings of MoClo-constructed GFP constructs. Normalised maximum fluorescence of *E. coli* BW25113 cultures grown in **(A)** LB or **(B)** M9 minimal medium with 1% glucose per promoter, RBS and origin of replication. Error bars represent the standard deviation of three biological replicates. **(C)** Range of normalised maximum fluorescence measurements of *E. coli* BW25113 cultures grown in LB or **(D)** M9 minimal medium 1% glucose per origin of replication. Black dots within bars represent individual constructs, each black line represents the median normalised fluorescence of all constructs per origin of replication.

To characterise the effect of each origin of replication on expression potential and burden, we constructed a set of 27 GFP constructs (pL1_JRN151_AE to pL1_JRN177_AE), with varying promoters, RBSs and either p15A-, pBR322- or pUC19-derived origins of replication and a kanamycin resistance marker (DVK_p15A_AE, DVK_pBR322_AE and DVK_pUC19_AE,

respectively). Characterised promoters and RBSs were selected from the CIDAR MoClo kit, consisting of constitutive promoters J23100 (rel. act. 1.0), J23107 (0.47), J23116 (0.16) and RBSs B0034m (rel. act. 1.0), B0032m (0.33) and B0033m (0.01) (Iverson et al., 2016). We used *E. coli* BW25113 as the host strain for this work, which is derived from *E. coli* K-12 and was the parental strain used for construction of the Keio collection of single gene knockouts (Baba et al., 2006). Since BW25113 is a frequently used host for metabolic engineering studies we aimed to characterise it in terms protein expression capacity and growth burden.

In general, the use of LB medium resulted in a more reliable relationship between the expected expression strength and measured GFP expression, with stronger promoters, RBSs, and higher copy vectors leading to higher GFP output (Figure 3.2A). Only extremely strong expression constructs with pUC19 backbones resulted in a deviation from the expected pattern in LB medium, which we attributed to growth being impacted due to expression burden, as quantified later. In terms of maximum normalised fluorescence, constructs assembled with a DVK_pUC19_AE backbone consistently achieved the highest values for every tested construct when grown on LB (except for the burdensome J23100/B0034m construct), indicating high expression of GFP even with a low strength promoter and RBS (Figure 3.2A). On the lower end of the GFP expression spectrum we mostly found constructs with p15A-derived origins of replication, and we were unable to distinguish between the DVK_p15A plasmid encoding a J23116/B0033m GFP construct and an empty vector control.

In glucose minimal medium, strains carrying DVK_pBR322_AE-derived plasmids displayed somewhat erratic expression profiles, such as J23100/B0033m displaying lower normalised fluorescence compared to J23106/B0033m (Figure 3.2B). This behaviour was mostly caused by the relatively poor growth of the strain when expressing certain constructs (Figure 3.3B, D). In general, vectors with a pBR322 backbone seemed to confer lower growth rates and lower final OD_{600} values in both LB and glucose minimal medium compared to low copy p15A and high copy pUC19 vectors (Figure 3.3). Perhaps the expression of *rop* and *bom* proteins from

the pBR322 origin of replication resulted in unwanted burden in the genetic context of certain GFP expression constructs.

In general, GFP expression from the respective backbones followed an expected trend where higher copy plasmids resulted in higher expression strength, although this profile was more distinctly present when using LB (Figure 3.2C) compared to M9 glucose medium (Figure 3.2D). The p15A- and pBR322 backbone plasmids allowed a more fine-grained exploration of the expression space, as can be seen from the improved coverage over normalised fluorescence values compared to just using DVK_pUC19_AE as a backbone (Figure 3.A, B). Besides improving coverage when using all three backbones, the use of p15A- or pBR322 backbones alone resulted in a smaller range of normalised fluorescence when weak- or medium strength constructs were used compared to the pUC19 backbone (Figure 3.2A) and B, J23116 and J23106). The use of lower-copy plasmids may therefore be beneficial if fine-tuning of constitutive protein expression is required within a specific expression range.

Across all three origins of replications used in this work, we observed the most consistent expression- and burden profiles with a p15A-derived origin of replication. Constructs assembled in pBR322- and pUC19 backbones displayed aberrant growth in minimal medium in some cases, especially when high strength promoters and RBSs were used (see Table S3.4 for μ_{max} and maximum OD_{600} values per construct and condition). This resulted in several outliers that deviated in terms of growth rates or final biomass yields (Figure 3.3B and D), while strains carrying DVK_p15A_AE plasmids exhibited a more-or-less stable relationship between expression strength and the theoretical strength of each constituent part, together with a mostly low burden profile (Figure 3.2A, B and Figure 3.3). Growth rates and final biomass yields of p15A-carrying *E. coli* were always equal to or higher than the other two backbones (Figure 3.3). The maximum normalised fluorescence achieved by GFP constructs in a p15A construct was around $1x10^4$ RFU/OD_{600} in both LB and M9 glucose. Compared to GFP expressed from a pUC19 backbone, the maximum RFU/OD_{600} was only ~8 fold (LB) or ~5 fold (M9 glucose) lower (Figure 3.2A, B). Considering the apparent low growth burden that

p15A imposes on its host, the DVK_p15A backbone may therefore be the most suitable choice if reliable profiling of various expression constructs is required.

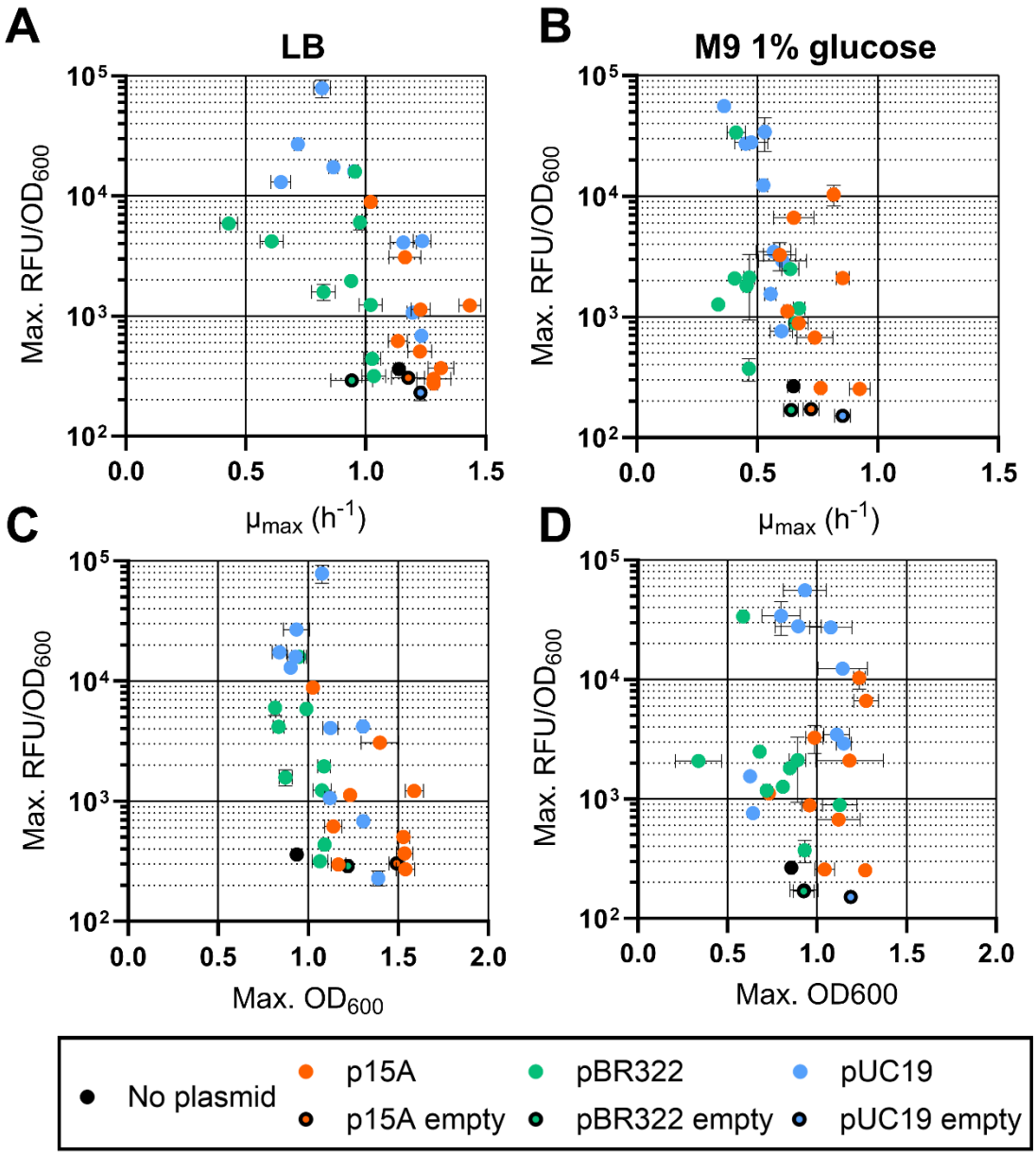


Figure 3.3. Burden assessment of MoClo-compatible destination vectors with different origins of replication. *E. coli* BW25113 carrying constitutive GFP expression constructs of varying strengths were grown in LB (**A, C**) or M9 minimal medium with 1% glucose (**B, D**) and assessed for their growth rates (**A, B**) and maximum achieved OD_{600} (**C, D**). Datapoints represent averages of biological triplicate measurements, and error bars represent their standard deviations.

3.4.2 Growth-coupled optimisation of *adhE* expression by randomised GGA

Besides being able to generate defined constructs at high efficiencies, another benefit of using modular cloning is that it allows efficient randomization of parts while still maintaining high fidelity in terms of the order of assembled DNA parts. By including a library of one or more parts in an assembly, the resulting constructs are expected to contain a random distribution of said parts. In this way, level 1 constructs can contain randomised promoters, RBSs, CDSs or terminators, and these randomised level 1 constructs can subsequently be used in level 2 assemblies to generate randomised expression levels of metabolic pathways or enzyme complexes.

If the output of these randomised constructs generates a detectable phenotype, this method can be used to screen for optimal metabolic pathway enzyme expression with relative ease. For example, the heterologous biosynthesis of carotenoids in *E. coli* has been used as a proof-of-principle for this approach, since the resulting colour formation of colonies can be linked to differentially expressed pathway enzymes (Taylor et al., 2019). An alternative to this screening approach is to use selection, where the activity of the enzyme(s) of interest contribute to cell fitness, making them growth coupled. Cell populations carrying a library of growth-coupled enzyme constructs can be used to determine optimal expression levels of the enzyme(s) with relatively ease, either by passaging in selective medium or growing populations in selective continuous cultures. To fully explore the expression space of the enzyme, using destination vectors with varying copy numbers should ensure selection for an optimal balance between expression level and plasmid burden.

We demonstrated this concept by optimizing expression of *adhE**, an oxygen-tolerant variant (A267T E568K) of the native *E. coli* bifunctional alcohol dehydrogenase *adhE* (Holland-Staley et al., 2000; Membrillo-Hernández et al., 2000). This enzyme confers *E. coli* the ability to grow on ethanol as a sole carbon source in aerobic environments (Figure 3.4A). To optimise *adhE** expression, we constructed a library of various promoters (J23103, J23116, J23107, J23106,

J23102, J23100), RBSs (B0033m, B0032m, B0034m) and destination vectors (DVK_pUC19, DVK_pBR322, DVK_p15A), along with the *adhE** CDS and a terminator (Figure 3.4B).

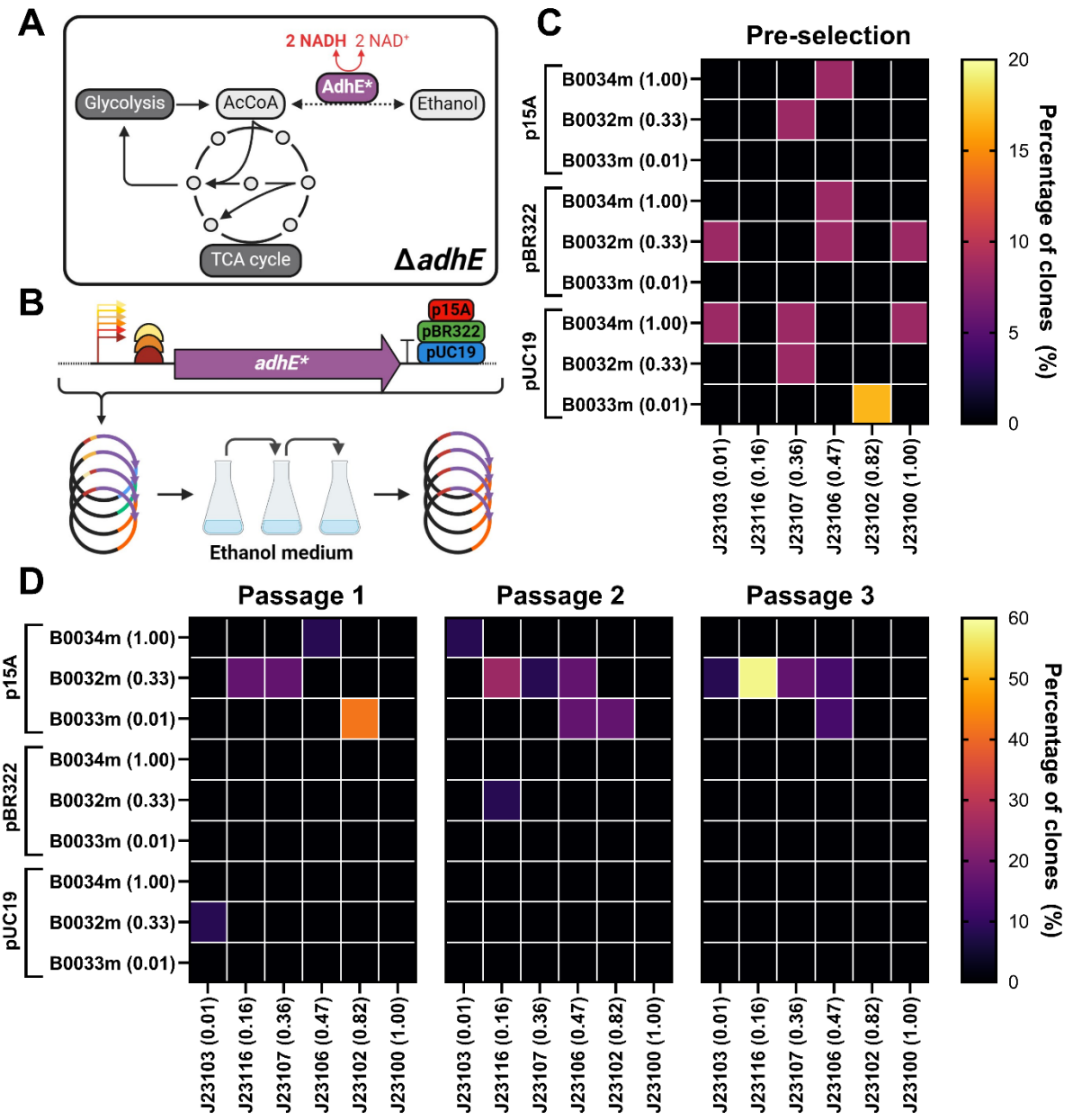


Figure 3.4. Growth-coupled selection for optimal *adhE expression levels using MoClo-mediated randomised construct assembly**. **A)** Metabolic pathway depicting the catabolism of ethanol into central metabolism by AdhE, the acetaldehyde intermediate is not depicted. **B)** Workflow of the selection experiment starting with randomised Golden Gate assembly of promoters, RBSs and destination vectors, yielding a diverse population of plasmids. Subsequent transformation and passaging on M9 minimal medium with 1 mM glucose + 300 mM ethanol results in selection for optimal expression of *adhE**. **C)** Percentage of clones carrying a certain *adhE** expression construct before selection on ethanol medium was imposed. **D)** Percentage of clones carrying a certain *adhE** expression construct after 1, 2 or 3 passages on M9 minimal medium with 1 mM glucose + 300 mM ethanol. Three separate populations were transformed and passaged, and the resulting construct compositions were combined in the heatmaps shown in panels C and D.

The transformed *adhE** expression library displayed a random distribution of promoters, RBSs and destination vectors before selection (Figure 3.4C). Once selection was imposed by supplying ethanol as a carbon source in M9 minimal medium, a single passage revealed that the majority of isolated clones carried *adhE** constructs in DVK_p15A backbones. This suggested that a plasmid with a low copy origin of replication conferred a growth benefit for the catabolism of ethanol by *adhE**. After the third passage, all examined clones carried DVK_p15A_AE backbones and the majority (~60%) of these expressed *adhE** with a J23116 promoter (rel. act. 0.16) and a B0032m RBS (rel. act. 0.33) (Figure 3.4D). This suggested that relatively low expression levels of *adhE** were sufficient to establish optimal growth with ethanol as a carbon source. After the first passage one particular variant was observed in half of the isolates in each replicate population (J23102/B0033m), but it's frequency in the population reduced after subsequent passages (Figure 3.4D). It is likely that the first passage into selective medium required metabolic adaptation (e.g. changes in regulation) for the catabolism of ethanol, whereas these adaptations were established in later passages. Optimal *adhE** expression levels for growth maximisation may have changed as a result of this.

It is unlikely that the catalytic properties of AdhE* itself were the bottleneck to growth since we observed selection for relatively low *adhE** expression levels. In fact, from our initial characterization of the GFP expression constructs in LB and M9 medium (Figure 3.2A, B) the J23116/B0032m combination in a DVK_p15A_AE backbone led to one of the lowest maximum normalised fluorescence readings of all tested constructs. This highlighted that providing vectors with a variety of copy numbers in growth-coupled selection experiments was important for identifying truly optimal expression levels. We concluded that the catabolism of ethanol into central metabolism may simply be unable to accommodate a higher flux from AdhE*, perhaps due to the high levels of NADH produced as a result of ethanol oxidation.

3.5 Discussion

In this report we constructed and characterised several MoClo-compatible vectors to add to the repertoire of modular parts available for GGA. These vectors were made to be compatible

specifically with the CIDAR MoClo kit (Iverson et al., 2016), although their overhangs are compatible with the MoClo standard in general (Weber et al., 2011). The effect of plasmid copy number was assessed in terms of GFP expression strength and growth burden, providing useful reference material for those using this set of plasmids, or those using similar plasmids.

The expression of GFP using various backbones and promoter/RBS combination revealed that on the whole, increasing copy number or promoter/RBS strength resulted in increases in normalised GFP expression, as expected. However, we found that deleterious effects on growth disrupted this pattern at the extremes of expression strength, most notably using pUC19- or pBR322 backbones with strong promoter/RBS parts. The pBR322 backbone also displayed some unexpectedly aberrant burden profiles when using seemingly nominal (i.e. not strong) expression constructs, both in LB and M9 glucose minimal medium. We currently do not have an explanation for these observations, but it highlighted that determining the expression- and burden profiles of plasmids before using them in further work is important to ensure the reproducibility of studies.

In addition, promoter and RBS sequences are subject to (genetic) context dependence, resulting in variable and possibly unexpected expression levels of target protein (Cardinale & Arkin, 2015; Carr et al., 2017). Although (genetic) context was somewhat controlled in this study, since only one *E. coli* strain and a limited number of backbones, promoters, RBSs and CDSs were used, computational tools and/or more extensive *in vivo* profiling of protein expression may be required to determine optimal construct composition when using this destination vector kit in different contexts (Salis et al., 2009). It may also be worthwhile to characterise the expression strength and burden of the plasmids carrying different resistance markers, as changing antibiotic resistance markers was reported to result in differences in expression profiles (Iverson et al., 2016).

Besides profiling expression strength and burden, the plasmid kit was used in a proof-of-principle experiment to optimise the expression of *adhE** for ethanol catabolism in *E. coli*. Since ethanol catabolism in this example is strictly growth-coupled, given it is essential for

carbon uptake, the expression of *adhE** was expected to be optimised for growth maximization. Indeed, we identified a preferred expression solution after only several rounds of passaging in ethanol-containing medium. Growth-coupled enzymes that are not optimally expressed can lead to deleterious effects on growth (Wortel et al., 2018). For example, overconsumption- or production of central metabolites may reduce biomass flux, and excess cofactor reduction or oxidation can lead to redox state imbalances. Growth-coupled bioproduction can also benefit from optimal enzyme expression (Dinh et al., 2018). By performing enzyme expression selection as described in this work, one or several optimal expression solutions can be identified from large populations of variants with minimal resource requirements. As mentioned earlier, the (genetic) context dependence of promoters and RBSs may result in unexpected expression levels, but growth-coupled approaches circumvent this issue by directly selecting against non-optimal protein expression solutions.

Although performing expression selection with one enzyme may be superfluous, since defined constructs could be manually constructed and evaluated, once multiple enzymes are involved (e.g. in a biosynthetic pathway), the combinatorial space quickly becomes too large to sample manually. This randomization approach therefore lends itself well to the optimisation of enzyme pathway expression, an approach explored in several studies (Gambacorta et al., 2022; Orsi et al., 2021; Sellés Vidal et al., 2021; Taylor et al., 2019). While finding expression optima is the main aim of such an approach, this method also identifies the expression solutions that are selected against, informing and streamlining future design of enzyme expression constructs. By using the vector kit designed in this work for example, optimal expression solutions may be limited to just one backbone type, as was the case with *adhE** expression, which may aid the future construction and optimisation efforts for single- or dual plasmid systems.

3.6 Conclusion

Golden Gate cloning has quickly become a staple method in synthetic biology research due to its high fidelity, efficiency and ease-of-use compared to other methods such as Gibson- or

restriction enzyme cloning. The standardization and characterisation of genetic parts, and their public availability, are important aspects for unlocking the full potential of this DNA assembly method.

In this study, we constructed MoClo-compatible medium-copy vectors and characterised their expression- and burden profiles. The addition of these lower-copy vectors to the original CIDAR MoClo repertoire allows the assembly of genetic constructs varying up to 500-fold in expression level. The addition of a p15A-derived destination vector also allows construction of multi-plasmid systems with compatible origins of replication. We demonstrated how using plasmids with different origins of replication can help identify optimum expression solutions in growth-coupled designs. The destination vectors described in this study are available on Addgene (catalog #217582-217609).

We expect this kit to be useful to synthetic biologists working with *E. coli* expression systems. This work adds to the existing set of publicly available MoClo destination vectors, including the original CIDAR MoClo kit and the more recent broad-host MoClo vector kit (Keating & Young, 2023), providing researchers with a wide range of options in terms of expression backbones.

Acknowledgements

This work was supported by the BBSRC (grant BB/M011224/1) and the Oxford Interdisciplinary Bioscience Doctoral Training Partnership. Figure schematics were created with Biorender.com.

3.7 References

Andreou, A. I., & Nakayama, N. (2018). Mobius Assembly: A versatile Golden-Gate framework towards universal DNA assembly. *PLOS ONE*, *13*(1), e0189892. https://doi.org/10.1371/JOURNAL.PONE.0189892

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, *2*, 2006.0008. https://doi.org/10.1038/MSB4100050

Blázquez, B., León, D. S., Torres-Bacete, J., Gómez-Luengo, Á., Kniewel, R., Martínez, I., Sordon, S., Wilczak, A., Salgado, S., Huszcza, E., Popłoński, J., Prieto, A., & Nogales, J. (2023). Golden Standard: a complete standard, portable, and interoperative MoClo tool for model and non-model proteobacteria. *Nucleic Acids Research*, *51*(19), e98–e98. https://doi.org/10.1093/NAR/GKAD758

Cardinale, S., & Arkin, A. P. (2012). Contextualizing context for synthetic biology – identifying causes of failure of synthetic biological systems. *Biotechnology Journal*, *7*(7), 856–866. https://doi.org/10.1002/BIOT.201200085

Carr, S. B., Beal, J., & Densmore, D. M. (2017). Reducing DNA context dependence in bacterial promoters. *PLoS ONE*, *12*(4). https://doi.org/10.1371/JOURNAL.PONE.0176013

Chiasson, D., Giménez-Oya, V., Bircheneder, M., Bachmaier, S., Studtrucker, T., Ryan, J., Sollweck, K., Leonhardt, H., Boshart, M., Dietrich, P., & Parniske, M. (2019). A unified multi-kingdom Golden Gate cloning platform. *Scientific Reports 2019 9:1*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-46171-2

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

Dinh, H. V., King, Z. A., Palsson, B. O., & Feist, A. M. (2018). Identification of growth-coupled production strains considering protein costs and kinetic variability. *Metabolic Engineering Communications*, *7*, e00080. https://doi.org/10.1016/J.MEC.2018.E00080

Engler, C., Kandzia, R., & Marillonnet, S. (2008). A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLOS ONE*, *3*(11), e3647. https://doi.org/10.1371/JOURNAL.PONE.0003647

Engler, C., Youles, M., Gruetzner, R., Ehnert, T. M., Werner, S., Jones, J. D. G., Patron, N. J., & Marillonnet, S. (2014). A Golden Gate modular cloning toolbox for plants. *ACS Synthetic Biology*, *3*(11), 839–843. https://doi.org/10.1021/SB4001504

Fromme, T., & Klingenspor, M. (2007). Rapid single step subcloning procedure by combined action of type II and type IIs endonucleases with ligase. *Journal of Biological Engineering*, *1*(1), 1–3. https://doi.org/10.1186/1754-1611-1-7

Gambacorta, F. V., Dietrich, J. J., Baerwald, J. J., Brown, S. J., Su, Y., & Pfleger, B. F. (2022). Combinatorial library design for improving isobutanol production in Saccharomyces cerevisiae. *Frontiers in Bioengineering and Biotechnology*, *10*. https://doi.org/10.3389/FBIOE.2022.1080024

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods 2009 6:5*, *6*(5), 343–345. https://doi.org/10.1038/nmeth.1318

Hernanz-Koers, M., Gandía, M., Garrigues, S., Manzanares, P., Yenush, L., Orzaez, D., & Marcos, J. F. (2018). FungalBraid: A GoldenBraid-based modular cloning platform for the assembly and exchange of DNA elements tailored to fungal synthetic biology. *Fungal Genetics and Biology*, *116*, 51–61. https://doi.org/10.1016/J.FGB.2018.04.010

Holland-Staley, C. A., Lee, K., Clark, D. P., & Cunningham, P. R. (2000). Aerobic activity of Escherichia coli alcohol dehydrogenase is determined by a single amino acid. *Journal of Bacteriology*, *182*(21), 6049–6054. https://doi.org/10.1128/JB.182.21.6049-6054.2000

Iverson, S. V., Haddock, T. L., Beal, J., & Densmore, D. M. (2016). CIDAR MoClo: Improved MoClo Assembly Standard and New *E. coli* Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synthetic Biology*, *5*(1), 99–103. https://doi.org/10.1021/ACSSYNBIO.5B00124

Jahn, M., Vorpahl, C., Hübschmann, T., Harms, H., & Müller, S. (2016). Copy number variability of expression plasmids determined by cell sorting and droplet digital PCR. *Microbial Cell Factories*, *15*(1), 1–12. https://doi.org/10.1186/S12934-016-0610-8

Jensen, S. I., Lennen, R. M., Herrgård, M. J., & 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), 17874. https://doi.org/10.1038/srep17874

Keating, K. W., & Young, E. M. (2023). Systematic Part Transfer by Extending a Modular Toolkit to Diverse Bacteria. *ACS Synthetic Biology*, *12*(7), 2061–2072. https://doi.org/10.1021/ACSSYNBIO.3C00104

Knight, T. (2003). *Idempotent Vector Design for Standard Assembly of Biobricks*. https://dspace.mit.edu/handle/1721.1/21168

Kotera, I., & Nagai, T. (2008). A high-throughput and single-tube recombination of crude PCR products using a DNA polymerase inhibitor and type IIS restriction enzyme. *Journal of Biotechnology*, *137*(1–4), 1–7. https://doi.org/10.1016/J.JBIOTEC.2008.07.1816

Lammens, E. M., Boon, M., Grimon, D., Briers, Y., & Lavigne, R. (2022). SEVAtile: a standardised DNA assembly method optimised for Pseudomonas. *Microbial Biotechnology*, *15*(1), 370–386. https://doi.org/10.1111/1751-7915.13922

Lee, M. E., DeLoache, W. C., Cervantes, B., & Dueber, J. E. (2015). A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synthetic Biology*, *4*(9), 975–986. https://doi.org/10.1021/SB500366V

Membrillo-Hernández, J., Echave, P., Cabiscol, E., Tamarit, J., Ros, J., & Lin, E. C. C. (2000). Evolution of the adhE gene product of Escherichia coli from a functional reductase to a dehydrogenase: Genetic and biochemical studies of the mutant proteins. *Journal of Biological Chemistry*, *275*(43), 33869–33875. https://doi.org/10.1074/jbc.M005464200

Moore, S. J., Lai, H. E., Kelwick, R. J. R., Chee, S. M., Bell, D. J., Polizzi, K. M., & Freemont, P. S. (2016). EcoFlex: A Multifunctional MoClo Kit for *E. coli* Synthetic Biology. *ACS Synthetic Biology*, *5*(10), 1059–1069. https://doi.org/10.1021/ACSSYNBIO.6B00031

Orsi, E., Claassens, N. J., Nikel, P. I., & Lindner, S. N. (2021). Growth-coupled selection of synthetic modules to accelerate cell factory development. *Nature Communications 2021 12:1*, *12*(1), 1–5. https://doi.org/10.1038/s41467-021-25665-6

Pryor, J. M., Potapov, V., Bilotti, K., Pokhrel, N., & Lohman, G. J. S. (2022). Rapid 40 kb Genome Construction from 52 Parts through Data-optimized Assembly Design. *ACS Synthetic Biology*, *11*(6), 2036–2042. https://doi.org/10.1021/ACSSYNBIO.1C00525

Salis, H. M., Mirsky, E. A., & Voigt, C. A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology 2009 27:10*, *27*(10), 946–950. https://doi.org/10.1038/nbt.1568

Sarrion-Perdigones, A., Vazquez-Vilar, M., Palací, J., Castelijns, B., Forment, J., Ziarsolo, P., Blanca, J., Granell, A., & Orzaez, D. (2013). GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiology*, *162*(3), 1618–1631. https://doi.org/10.1104/PP.113.217661

Sellés Vidal, L., Murray, J. W., & Heap, J. T. (2021). Versatile selective evolutionary pressure using synthetic defect in universal metabolism. *Nature Communications 2021 12:1*, *12*(1), 1–15. https://doi.org/10.1038/s41467-021-27266-9

Shao, B., Rammohan, J., Anderson, D. A., Alperovich, N., Ross, D., & Voigt, C. A. (2021). Single-cell measurement of plasmid copy number and promoter activity. *Nature Communications 2021 12:1*, *12*(1), 1–9. https://doi.org/10.1038/s41467-021-21734-y

Stukenberg, D., Hensel, T., Hoff, J., Daniel, B., Inckemann, R., Tedeschi, J. N., Nousch, F., & Fritz, G. (2021). The Marburg Collection: A Golden Gate DNA Assembly Framework for Synthetic Biology Applications in Vibrio natriegens. *ACS Synthetic Biology*, *10*(8), 1904–1919. https://doi.org/10.1021/ACSSYNBIO.1C00126

Taylor, G. M., Mordaka, P. M., & Heap, J. T. (2019). Start-Stop Assembly: a functionally scarless DNA assembly system optimized for metabolic engineering. *Nucleic Acids Research*, *47*(3), e17–e17. https://doi.org/10.1093/NAR/GKY1182

Valenzuela-Ortega, M., & French, C. E. (2020). Joint universal modular plasmids: A flexible platform for golden gate assembly in any microbial host. *Methods in Molecular Biology*, *2205*, 255–273. https://doi.org/10.1007/978-1-0716-0908-8_15

Vasudevan, R., Gale, G. A. R., Schiavon, A. A., Puzorjov, A., Malin, J., Gillespie, M. D., Vavitsas, K., Zulkower, V., Wang, B., Howe, C. J., Lea-Smith, D. J., & McCormick, A. J. (2019). CyanoGate: A Modular Cloning Suite for Engineering Cyanobacteria Based on the Plant MoClo Syntax. *Plant Physiology*, *180*(1), 39–55. https://doi.org/10.1104/PP.18.01401

Weber, E., Engler, C., Gruetzner, R., Werner, S., & Marillonnet, S. (2011). A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLOS ONE*, *6*(2), e16765. https://doi.org/10.1371/JOURNAL.PONE.0016765

Wortel, M. T., Noor, E., Ferris, M., Bruggeman, F. J., & Liebermeister, W. (2018). Metabolic enzyme cost explains variable trade-offs between microbial growth rate and yield. *PLOS Computational Biology*, *14*(2), e1006010. https://doi.org/10.1371/JOURNAL.PCBI.1006010

3.8 Supplementary data

Table S3.1. Plasmids used and constructed in this study.

Name	Contents	Resistance	Origin of replication	Reference
DVA	*lacZα*	amp^R	pMB1 (pUC19)	(Iverson et al., 2016)
DVK	*lacZα*	kan^R	pMB1 (pUC19)	(Iverson et al., 2016)
pKD3	FRT-cm^R-FRT, oriR6K, amp^R	amp^R,cm^R	R6K gamma	(Datsenko & Wanner, 2000)
pSIJ8	araC-P_{ara}-α-β-γ rhaRS-P_{rha}-FLP, amp^R	amp^R	pSC101	(Jensen et al., 2015)
pL0_JRN012_CD	Level 0 part containing domesticated *E. coli* MG1655 *adhE*	amp^R	pMB1 (pUC19)	This study
pL0_JRN017_CD	Level 0 part containing *E. coli* $adhE^{A267T}$	amp^R	pMB1 (pUC19)	This study
pL0_JRN018_CD	Level 0 part containing *E. coli* $adhE^{A267T/E568K}$	amp^R	pMB1 (pUC19)	This study
DVK_p15A_AE	*lacZα*	kan^R	p15A	This study
DVK_p15A_AF	*lacZα*	kan^R	p15A	This study
DVK_p15A_EF	*lacZα*	kan^R	p15A	This study
DVK_p15A_FG	*lacZα*	kan^R	p15A	This study
DVK_p15A_GH	*lacZα*	kan^R	p15A	This study
DVC_p15A_AE	*lacZα*	cm^R	p15A	This study
DVC_p15A_AF	*lacZα*	cm^R	p15A	This study
DVC_p15A_EF	*lacZα*	cm^R	p15A	This study
DVC_p15A_FG	*lacZα*	cm^R	p15A	This study
DVC_p15A_GH	*lacZα*	cm^R	p15A	This study
DVK_pBR322_AE	*lacZα*	kan^R	pMB1 (pBR322)	This study
DVK_pBR322_AF	*lacZα*	kan^R	pMB1 (pBR322)	This study
DVK_pBR322_EF	*lacZα*	kan^R	pMB1 (pBR322)	This study
DVK_pBR322_FG	*lacZα*	kan^R	pMB1 (pBR322)	This study
DVK_pBR322_GH	*lacZα*	kan^R	pMB1 (pBR322)	This study
DVC_pBR322_AE	*lacZα*	cm^R	pMB1 (pBR322)	This study
DVC_pBR322_AF	*lacZα*	cm^R	pMB1 (pBR322)	This study
DVC_pBR322_EF	*lacZα*	cm^R	pMB1 (pBR322)	This study
DVC_pBR322_FG	*lacZα*	cm^R	pMB1 (pBR322)	This study
DVC_pBR322_GH	*lacZα*	cm^R	pMB1 (pBR322)	This study
DVA_p15A_AF	*lacZα*	amp^R	p15A	This study
DVA_p15A_AG	*lacZα*	amp^R	p15A	This study
DVA_p15A_EH	*lacZα*	amp^R	p15A	This study
DVA_p15A_AH	*lacZα*	amp^R	p15A	This study

DVA_pBR322_AF	*lacZα*	amp^R	pMB1 (pBR322)	This study
DVA_pBR322_AG	*lacZα*	amp^R	pMB1 (pBR322)	This study
DVA_pBR322_EH	*lacZα*	amp^R	pMB1 (pBR322)	This study
DVA_pBR322_AH	*lacZα*	amp^R	pMB1 (pBR322)	This study
DVK_pUC19_AE	*lacZα*	kan^R	pMB1 (pUC19)	(Iverson et al., 2016)
DVA_pUC19_CD	*lacZα*	amp^R	pMB1 (pUC19)	(Iverson et al., 2016)
DVA_J23100_AB	J23100 promoter part	amp^R	pMB1 (pUC19)	(Iverson et al., 2016)
DVA_J23102_AB	J23102 promoter part	amp^R	pMB1 (pUC19)	(Iverson et al., 2016)
DVA_J23106_AB	J23106 promoter part	amp^R	pMB1 (pUC19)	(Iverson et al., 2016)
DVA_J23107_AB	J23107 promoter part	amp^R	pMB1 (pUC19)	(Iverson et al., 2016)
DVA_J23116_AB	J23116 promoter part	amp^R	pMB1 (pUC19)	(Iverson et al., 2016)
DVA_J23103_AB	J23103 promoter part	amp^R	pMB1 (pUC19)	(Iverson et al., 2016)
DVA_B0034m_BC	B0034m RBS part	amp^R	pMB1 (pUC19)	(Iverson et al., 2016)
DVA_B0032m_BC	B0032m RBS part	amp^R	pMB1 (pUC19)	(Iverson et al., 2016)
DVA_B0033m_BC	B0033m RBS part	amp^R	pMB1 (pUC19)	(Iverson et al., 2016)
DVA_B0015_DE	B0015 terminator	amp^R	pMB1 (pUC19)	(Iverson et al., 2016)
DVA_GFP_CD	Bba_E0040 part, GFP mut3b	amp^R	pMB1 (pUC19)	(Iverson et al., 2016)
pL1_JRN151_AE	J23116_B0033m_GFP_B0015	kan^R	p15A	This study
pL1_JRN152_AE	J23106_B0033m_GFP_B0015	kan^R	p15A	This study
pL1_JRN153_AE	J23100_B0033m_GFP_B0015	kan^R	p15A	This study
pL1_JRN154_AE	J23116_B0032m_GFP_B0015	kan^R	p15A	This study
pL1_JRN155_AE	J23106_B0032m_GFP_B0015	kan^R	p15A	This study
pL1_JRN156_AE	J23100_B0032m_GFP_B0015	kan^R	p15A	This study
pL1_JRN157_AE	J23116_B0034m_GFP_B0015	kan^R	p15A	This study
pL1_JRN158_AE	J23106_B0034m_GFP_B0015	kan^R	p15A	This study
pL1_JRN159_AE	J23100_B0034m_GFP_B0015	kan^R	p15A	This study
pL1_JRN160_AE	J23116_B0033m_GFP_B0015	kan^R	pMB1 (pBR322)	This study
pL1_JRN161_AE	J23106_B0033m_GFP_B0015	kan^R	pMB1 (pBR322)	This study
pL1_JRN162_AE	J23100_B0033m_GFP_B0015	kan^R	pMB1 (pBR322)	This study
pL1_JRN163_AE	J23116_B0032m_GFP_B0015	kan^R	pMB1 (pBR322)	This study
pL1_JRN164_AE	J23106_B0032m_GFP_B0015	kan^R	pMB1 (pBR322)	This study
pL1_JRN165_AE	J23100_B0032m_GFP_B0015	kan^R	pMB1 (pBR322)	This study
pL1_JRN166_AE	J23116_B0034m_GFP_B0015	kan^R	pMB1 (pBR322)	This study
pL1_JRN167_AE	J23106_B0034m_GFP_B0015	kan^R	pMB1 (pBR322)	This study

pL1_JRN168_AE	J23100_B0034m_GFP_B0015	kan^R	pMB1 (pBR322)	This study
pL1_JRN169_AE	J23116_B0033m_GFP_B0015	kan^R	pMB1 (pUC19)	This study
pL1_JRN170_AE	J23106_B0033m_GFP_B0015	kan^R	pMB1 (pUC19)	This study
pL1_JRN171_AE	J23100_B0033m_GFP_B0015	kan^R	pMB1 (pUC19)	This study
pL1_JRN172_AE	J23116_B0032m_GFP_B0015	kan^R	pMB1 (pUC19)	This study
pL1_JRN173_AE	J23106_B0032m_GFP_B0015	kan^R	pMB1 (pUC19)	This study
pL1_JRN174_AE	J23100_B0032m_GFP_B0015	kan^R	pMB1 (pUC19)	This study
pL1_JRN175_AE	J23116_B0034m_GFP_B0015	kan^R	pMB1 (pUC19)	This study
pL1_JRN176_AE	J23106_B0034m_GFP_B0015	kan^R	pMB1 (pUC19)	This study
pL1_JRN177_AE	J23100_B0034m_GFP_B0015	kan^R	pMB1 (pUC19)	This study

Table S3.2. DNA sequences of primers used in this study.

Primer name	Sequence	Purpose
JRN189	CTGGCGCTGCGCTTT	Introduce E568K into adhE
JRN190	TTCGAAGTGAGTTTCCGGATG	Introduce E568K into adhE
JRN191	ATCCGGAAACTCACTTCGAAAAGCTGGCGCTGCGCTTTATGGA	adhE E568K oligo bridge
JRN224	CGGCTATCTGTTGCAGG	Introduce A267T into adhE
JRN225	AACGTTCACGTACAGCGTC	Introduce A267T into adhE
JRN226	GTTGACTCTGTTTATGACGCTGTACGTGAACGTTTTACCACCCA CGGCGGCTATCTGTTGCAGGGTAAAGAGC	adhE A267T oligo bridge
JRN273	gctaaggatgatttctgg	Check MoClo vector insert
JRN274	agtgagcgaggaagc	Check MoClo vector insert
JRN279	AGTCGAAGACGAAATGGCTGTTACTAATGTCGCTG	Domesticate *adhE*
JRN280	AGTCGAAGACCAGTCATCTTCAGACAGAACACCAC	Domesticate *adhE*
JRN281	AGTCGAAGACGTTGACACTTTTGGTACCATCA	Domesticate *adhE*
JRN282	AGTCGAAGACGTGTTTTCAGACTGATCAGCGA	Domesticate *adhE*
JRN283	AGTCGAAGACCAAAACCCGTAACGCCATTATCT	Domesticate *adhE*
JRN284	AGTCGAAGACCATTCGAAATCTTTAGCGCG	Domesticate *adhE*
JRN285	AGTCGAAGACAGCGAAGATGCGGTAGAAAAAGCAGA	Domesticate *adhE*
JRN286	AGTCGAAGACAGAAGAACACTTCAGTTTCAACGC	Domesticate *adhE*
JRN287	AGTCGAAGACGATCTTCGAAGTAGAAGCGG	Domesticate *adhE*
JRN288	AGTCGAAGACGAACCTTTAAGCGGATTTTTTCGCT	Domesticate *adhE*
JRN386	Ctgttgtggaagccgttatag	Check *adhE* locus
JRN387	Ctaatgtttaaactcttttagtaaatcacag	Check *adhE* locus
JRN384	CCGTTTATGTTGCCAGACAGCGCTACTGATTAAGCGGATTTTTT CGCTTTtgtgtaggctggagctgcttc	Amplify cmR marker with adhE flanks from pKD3
JRN385	CGAGCAGATGATTTACTAAAAAAGTTTAACATTATCAGGAGAGC ATTATGcatatgaatatcctccttag	Amplify cmR marker with adhE flanks from pKD3
JRN413	GTCGTCTCCtggaagacatGGAGagagacctgGGAGACGAC	Oligo_A_L1_TV_fwd
JRN414	GTCGTCTCCgggtctctGCTTatgtcttctGGAGACGAC	Oligo_E_L1_SU_fwd
JRN415	GTCGTCTCCtggaagacatGCTTagagacctgGGAGACGAC	Oligo_E_L1_TV_fwd
JRN416	GTCGTCTCCgggtctctCGCTatgtcttctGGAGACGAC	Oligo_F_L1_SU_fwd
JRN417	GTCGTCTCCtggaagacatCGCTagagacctgGGAGACGAC	Oligo_F_L1_TV_fwd
JRN418	GTCGTCTCCgggtctctTGCCatgtcttctGGAGACGAC	Oligo_G_L1_SU_fwd
JRN419	GTCGTCTCCtggaagacatTGCCagagacctgGGAGACGAC	Oligo_G_L1_TV_fwd
JRN420	GTCGTCTCCgggtctctACTAatgtcttctGGAGACGAC	Oligo_H_L1_SU_fwd

JRN421	GTCGTCTCCtggaggtctcaGGAGatgtcttccctgGGAGACGAC	Oligo_A_L2_TV_fwd
JRN422	GTCGTCTCCtggaggtctcaGCTTatgtcttccctgGGAGACGAC	Oligo_E_L2_TV_fwd
JRN423	GTCGTCTCCgggttaagaagacgtCGCTagagaccttctGGAGACGAC	Oligo_F_L2_SU_fwd
JRN424	GTCGTCTCCtggaggtctcaCGCTatgtcttccctgGGAGACGAC	Oligo_F_L2_TV_fwd
JRN425	GTCGTCTCCgggttaagaagacgtTGCCagagaccttctGGAGACGAC	Oligo_G_L2_SU_fwd
JRN426	GTCGTCTCCgggttaagaagacgtACTAagagaccttctGGAGACGAC	Oligo_H_L2_SU_fwd
JRN427	GTCGTCTCCcaggtctctCTCCatgtcttccaGGAGACGAC	Oligo_A_L1_TV_rev
JRN428	GTCGTCTCCagaagacatAAGCagagacccGGAGACGAC	Oligo_E_L1_SU_rev
JRN429	GTCGTCTCCcaggtctctAAGCatgtcttccaGGAGACGAC	Oligo_E_L1_TV_rev
JRN430	GTCGTCTCCagaagacatAGCGagagacccGGAGACGAC	Oligo_F_L1_SU_rev
JRN431	GTCGTCTCCcaggtctctAGCGatgtcttccaGGAGACGAC	Oligo_F_L1_TV_rev
JRN432	GTCGTCTCCagaagacatGGCAagagacccGGAGACGAC	Oligo_G_L1_SU_rev
JRN433	GTCGTCTCCcaggtctctGGCAatgtcttccaGGAGACGAC	Oligo_G_L1_TV_rev
JRN434	GTCGTCTCCagaagacatTAGTagagacccGGAGACGAC	Oligo_H_L1_SU_rev
JRN435	GTCGTCTCCcagggaagacatCTCCtgagacctccaGGAGACGAC	Oligo_A_L2_TV_rev
JRN436	GTCGTCTCCcagggaagacatAAGCtgagacctccaGGAGACGAC	Oligo_E_L2_TV_rev
JRN437	GTCGTCTCCagaaggtctctAGCGacgtcttcttaacccGGAGACGAC	Oligo_F_L2_SU_rev
JRN438	GTCGTCTCCcagggaagacatAGCGtgagacctccaGGAGACGAC	Oligo_F_L2_TV_rev
JRN439	GTCGTCTCCagaaggtctctGGCAacgtcttcttaacccGGAGACGAC	Oligo_G_L2_SU_rev
JRN440	GTCGTCTCCagaaggtctctTAGTacgtcttcttaacccGGAGACGAC	Oligo_H_L2_SU_rev

Table S3.3. Synthesised DNA fragments used in this study.

Name	Sequence
p15A origin of replication fragment	TTCGTCTCCTTCTGGATCGGTTGTCGAGTAAGGATCTCCAGGCATCAAATAAAACGAAAGG CTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTA GAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATAGGATCCTAACTCGAGC CTAGGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGG CGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGATACTT AACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCCCCCTGAC AAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGA TACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTT ACCGGTGTCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCC GGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGAACCCCCCGTTCAGTCCGACCG CTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATGCAAAAGCACCA CTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAA GGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCA AAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTTTC AGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATCAGATA AAATATTTCTAGATTTCAGTGCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCAT ACGATATAAGTTGTTACTAGTGCTTGGATTCTCACCAATAAAAAACGCCCGGCGGCAACCG AGCGTTCTGAACAAATCCAGATGGAGTTCTGAGGTCATTACTGGATCTATCAACAGGAGTC CAAGCGAGCTCTCGAACCCCAGAGTCCCGCGATTGGAGACGAC
pBR322 origin of replication fragment	TTCGTCTCCTTCTAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATT TTTCTCTGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACC GGGCATGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCAT TACCCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACC GCCCTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACG AGCTGGACGCGGATGAACAGGCAGACATCTGTGAATCGCTTCACGACCACGCTGATGAGC TTTACCGCAGCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGC TCCCGGCGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAG GGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGCAGCCATGACCCAGTCACGTAGCG ATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCA CCATTGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCT TCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCA GCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAAC ATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTT TTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGG CGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGC TCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGC GTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCA AGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACT ATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTA ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTA ACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTT CGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTT

	TTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATC TTTTCTACCGGTCCGTCCAAAAAAAAAGGCTCCAAAAGGAGCCTTTAATTGTATCGGTGATT GGAGACGAC
amp^R fragment	TTCGTCTCCGATTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCG TTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCA TCTGGCCCCAGTGCTGCAATGATACCGCGCGACCCACGCTCACCGGCTCCAGATTTATCA GCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGC CTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGT TTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATG GCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCA AAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTT ATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCT TTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAG TTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTG CTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGAT CCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAG CGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGAC ACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTA TTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGC GCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC TATAAAAATAGGCGTATCACGAGGCAGAATTTCAGATAAAAAAAAATCCTTAGCTTTCGCTAA GGATGATTTCTGGAATTCGCGGCCGCTTCTAGAGACTAGTGGAGGAGACGAC
kan^R fragment	TTCGTCTCCGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGG ATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGC AGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAAT ACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGA CGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTGTTCAACAGG CCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATT GCGCCTGAGCAAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCG AATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATA TTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCA TCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTA GTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAAC TCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTAT CGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTGGA GCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAG ACAGTTTTATTGTTCATGATGATATATTTTTATCTTGTGCAATGTAACATCAGAGATTTTGAG ACACAACGTGGCTTTGTTGAATAAATCGAACTTTTGCTGAGTTGAAGGATCAGCTCGAGTG CCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCAC GAGGCAGAATTTCAGATAAAAAAAATCCTTAGCTTTCGCTAAGGATGATTTCTGGAATTCGC GGCCGCTTCTAGAGACTAGTGGAGGAGACGAC
cm^R fragment	TTCGTCTCCGATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCA TTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCA GCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGT CCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGATACGAA AAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACAT CTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGA AAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACC AGCTCACCGTCTTTCATTGCCATACGTAATTCCGGATGAGCATTCATCAGGCGGGCAAGAA TGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTA ATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAAT GTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATT TTAGCTTCCTTAGCTCCTGAAAATCTCGACAACTCAAAAAATACGCCCGGTAGTGATCTTAT TTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCATGCCACCTGACGTCTAAGAAA CCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCAGAATTTCAGATAA AAAAAATCCTTAGCTTTCGCTAAGGATGATTTCTGGAATTCGCGGCCGCTTCTAGAGACTA GTGTGGAGGAGACGAC
lacZα fragment	GTCGTCTCCCCTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATA CCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGC GGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTT GGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGC TCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGG TCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCG GAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTT GCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGG CCAACGCGCGGGGGTGGAGACGAC

Table S3.4. Growth parameters of *E. coli* BW25113 transformed with various GFP expression constructs. Strains were cultivated either in M9 minimal medium with 1% glucose, or rich LB medium. Values for maximum growth rates (μ_{max}) and maximum OD_{600} represent averages of biological triplicate measurements, and their respective standard deviations are provided.

			M9 1% glucose medium				LB medium			
Backbone	**RBS**	**Promoter**	μ_{max} (h^{-1})	Stdev (h^{-1})	Max. OD_{600}	Stdev	μ_{max} (h^{-1})	Stdev (h^{-1})	Max. OD_{600}	Stdev
No plasmid			0.649	0.0246	0.857	0.015288	1.14	0.0138	0.938	0.00821
p15A empty			0.723	0.0322	0.927	0.078063	1.18	0.0687	1.49	0.0421
pBR322 empty			0.639	0.0300	0.925	0.058251	0.942	0.0876	1.22	0.0159
pUC19 empty			0.853	0.0325	1.19	0.017071	1.23	0.0155	1.39	0.0253
p15A	**B0033m**	**J23116**	0.763	0.0101	1.04	0.056157	1.28	0.0151	1.54	0.0514
		J23106	0.671	0.0389	0.958	0.043467	1.23	0.0494	1.53	0.0356
		J23100	0.592	0.0676	0.985	0.049887	1.23	0.0392	1.23	0.00458
	B0032m	**J23116**	0.924	0.0415	1.27	0.010488	1.28	0.0709	1.17	0.0412
		J23106	0.623	0.0139	0.731	0.020763	1.13	0.0397	1.14	0.0478
		J23100	0.651	0.0838	1.27	0.068533	1.16	0.0661	1.40	0.1036
	B0034m	**J23116**	0.738	0.0741	1.12	0.118939	1.31	0.0535	1.54	0.0457
		J23106	0.854	0.0259	1.18	0.187774	1.43	0.0458	1.59	0.0499
		J23100	0.815	0.0126	1.24	0.031429	1.02	0.0264	1.03	0.0281
pBR322	**B0033m**	**J23116**	0.466	0.0232	0.891	0.046621	1.03	0.0497	1.07	0.0432
		J23106	0.672	0.0244	0.718	0.020014	0.824	0.0493	0.874	0.0369
		J23100	0.655	0.0214	1.13	0.09564	0.607	0.0484	0.836	0.0298
	B0032m	**J23116**	0.464	0.0119	0.930	0.026595	1.03	0.0357	1.09	0.0204
		J23106	0.455	0.0151	0.848	0.014656	0.940	0.00255	1.09	0.0357
		J23100	0.636	0.0355	0.680	0.017229	0.429	0.0364	0.991	0.0172
	B0034m	**J23116**	0.337	0.0079	0.808	0.011385	1.02	0.0482	1.08	0.0512
		J23106	0.405	0.00773	0.338	0.129651	0.976	0.0166	0.818	0.0222
		J23100	0.411	0.0382	0.587	0.018368	0.956	0.00913	0.951	0.0405
pUC19	**B0033m**	**J23116**	0.603	0.100	1.15	0.041369	1.23	0.00895	1.31	0.0189
		J23106	0.474	0.0691	0.894	0.128218	1.24	0.0372	1.30	0.0153
		J23100	0.531	0.0131	0.799	0.107182	0.818	0.0349	1.08	0.0104
	B0032m	**J23116**	0.567	0.0723	1.11	0.071954	1.20	0.0237	1.12	0.0277
		J23106	0.451	0.0169	1.08	0.118668	0.865	0.0190	0.841	0.0401
		J23100	0.554	0.0236	0.626	0.004239	0.717	0.0151	0.936	0.0715
	B0034m	**J23116**	0.524	0.0167	1.14	0.137828	1.16	0.0547	1.12	0.0423
		J23106	0.361	0.00275	0.931	0.120334	0.951	0.0190	0.931	0.0446
		J23100	0.599	0.0462	0.642	0.001899	0.647	0.0407	0.904	0.00473

2

4. Hydrocarbon biosynthesis in *E. coli* using non-sugar carbon sources

4.1 Abstract

The commercial production of drop-in biofuels has been a long-standing goal within the field of white biotechnology. Progress towards this goal has not only been hindered by low production metrics, such as limited yields and titres, but also by the scope of substrates used to establish these processes. In this work, we assessed the use of C_2 carbon sources acetate and ethanol for the biosynthesis of fatty alcohols, aldehydes, and alkanes in *E. coli*, and compared their efficacy to glucose. The cyanobacterial acyl-ACP reductase (AAR) and aldehyde deformylating oxygenase (ADO) enzymes were used for this purpose, and several optimization steps were taken to improve hydrocarbon production via this pathway. These included optimizing AAR/ADO expression levels, deletion of competing pathways, KCl supplementation, and the expression of fatty acid degradation regulator *fadR*. In addition, this work suggested that the *in vivo* oxygen tolerance of the *E. coli* bifunctional alcohol dehydrogenase (*adhE*) is considerably higher than previously assumed. In summary, we find that both acetate and ethanol performed as well or better than glucose in bioproduction metrics such as product yield, while glucose was more suitable for reaching a high productivity.

4.2 Introduction

July 2023 was the hottest month ever recorded since temperature records were first kept, with an average of 1.18^{o}C above the average global July temperature between 1951-1980 (O'Shea, 2023). With the effects of climate change becoming increasingly severe and more noticeable in day-to-day life, it is necessary to rapidly make the transition to a low-emission future.

Liquid transportation fuels including diesel, gasoline and kerosene are still expected to play a significant role in commercial and personal transportation even after the 2050 goal of net zero,

even if all government climate targets are met (International Energy Agency, 2023). Sectors including aviation and shipping are challenging to decarbonise in particular, due to the exceptional energy density of hydrocarbon fuels and the concurrent difficulty of replacing these with lower density alternatives, such as batteries. Therefore, hydrocarbon fuels that can be produced renewably represent attractive and necessary alternatives to the current fossil fuels.

The biosynthesis of fuel-length hydrocarbons such as alkenes, alkanes, aldehydes, and alcohols has been identified in various branches of life (Bourdenx et al., 2011; Holze et al., 2020; Ladygina et al., 2006). Some of these products are identical to those found in fossil fuels, making it theoretically possible to biosynthesise chemically identical 'drop-in' biofuels using the relevant enzymatic pathways. The most widely studied of these is the cyanobacterial alkane biosynthesis pathway discovered in 2010, which utilises the acyl-ACP reductase (AAR) and aldehyde deformylating oxygenase (ADO) enzymes (Schirmer et al., 2010). Many studies have expressed this pathway in *E. coli* to biosynthesise alkanes in various conditions, using metabolic modifications to improve production. However, the use of substrates beyond glucose has rarely been explored, even though the substrate choice impacts process economics, performance limits, and overall process sustainability. Other work using *Acinetobacter baylyi* and *A. woodii* has demonstrated that CO_2-derived acetate can be used as the sole carbon source for alkane biosynthesis (Lehtinen et al., 2018).

To produce low-value, high-volume products such as transportation fuels through a bioprocess, these must operate near theoretical maxima in terms of carbon yields (Caspeta & Nielsen, 2013; Kang & Nielsen, 2016). In addition, substrates must be affordable and ideally originate from waste streams or be closely derived from CO_2. Since the initial characterization of the AAR-ADO pathway, all studies using *E. coli* as a chassis for alkane production have used glucose as a substrate, often supplemented with rich medium such as yeast extract. Besides the high cost of glucose, unmodified *E. coli* will decarboxylate glucose (C_6) twice to produce two molecules of acetyl-CoA (C_2) that can enter fatty acid biosynthesis, a significant carbon loss of 33%. Unless synthetic glycolytic pathways are used that reduce carbon loss,

such as non-oxidative glycolysis (Bogorod et al., 2013), more suitable carbon- and energy sources must be explored to drive biological hydrocarbon synthesis towards economic viability.

To minimise carbon loss for bioproduction of long-chain compounds, substrates entering central metabolism at the acetyl-CoA node are an obvious choice since it is the starting point for fatty acid biosynthesis. Compounds such as acetate and ethanol can be activated or oxidised, respectively, to acetyl-CoA without carbon loss representing an attractive alternative to carbon-inefficient glucose. Acetate and ethanol have been compared to glucose for the production of PHB and other acetyl-CoA derived compounds, where the authors found that both performed as well if not better than glucose in terms of g/g product yield per substrate (Sun et al., 2020). A crucial factor for the efficacy of alkane biosynthesis is the availability of sufficient reducing power to operate fatty acid biosynthesis and the AAR/ADO pathway. Since ethanol catabolism releases two mol NADH per mol ethanol, it can provide more reducing power than glucose per carbon unit. In contrast, acetate requires ATP input to activate it to acetyl-CoA, meaning that reducing power must subsequently be derived from decarboxylation reactions in central catabolism, reducing yields. For this reason, we also explored using formate as an electron donor in acetate-fed cultures. Formate can be directly derived from CO_2 by electrocatalysis, and its oxidation to yield NAD(P)H can be used to provide reducing power *in vivo* (Partipilo et al., 2023).

In this work, we performed several optimisation steps to improve the activity of AAR alone, or in combination with ADO, using glycolytic substrates. We then explored ethanol and the combination of acetate and formate as carbon- and energy sources for hydrocarbon biosynthesis. The addition of KCl significantly improved hydrocarbon biosynthesis via AAR activity, as did the deletion of thioesterase C (*tesC/fadM/ybaW*). Other findings included that the wildtype *E. coli* bifunctional aldehyde/alcohol dehydrogenase *adhE* was significantly more oxygen tolerant *in vivo* than reported in seminal works on the topic (Holland-Staley et al., 2000; Membrillo-Hernández et al., 2000). Finally, the expression of transcriptional regulators *fadR*

and *iclR* was explored as a method to block biomass formation on acetate and ethanol, as a proof-of-principle for potential growth-decoupled production of acetyl-CoA derivatives including alkanes.

In summary, this work demonstrated that both acetate and ethanol are viable alternative carbon sources for hydrocarbon biosynthesis in *E. coli*, reaching titres and product/substrate yields similar to or better than glucose-fed cultures. Further exploration and optimization of using glucose alternatives may contribute to attaining economically viable hydrocarbon biosynthesis.

4.3 Materials and method

4.3.1 Strains, plasmids, and routine growth conditions

For general cloning and propagation purposes *E. coli* DH5α and BW25113 were grown using solid or liquid LB medium at 37°C (shaking at 250 rpm where relevant) with supplementation of antibiotics where appropriate: carbenicillin (carb, 100 µg/mL), kanamycin (kan, 50 µg/mL) or chloramphenicol (cm, 25 µg/mL). Where necessary, modified M9 minimal with the following composition was used: 47.8 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 18.7 mM NH_4Cl, 2 mM $MgSO_4$, 100 µM $CaCl_2$, 134 µM ethylene diamine tetrachloroacetic acid (EDTA), 100 µM $FeCl_3 \cdot 6H_2O$, 6.2 µM $ZnCl_2$, 0.76 µM $CuCl_2 \cdot 2H_2O$, 0.42 µM $CoCl_2 \cdot 2H_2O$, 1.62 µM H_3BO_3, and 0.081 µM $MnCl_2 \cdot 4H_2O$, adjusted to pH 7.0. Modified M9 medium was supplemented with antibiotics and carbon sources (e.g. D-glucose, sodium acetate, sodium formate, or ethanol) as required for the specific strain and assay conditions (all >99.5% purity, Sigma-Aldrich).

Table 4.1. Strains used and engineered in this study

Strain name	Genotype	Reference
E. coli DH5α	*fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	New England Biolabs, UK
E. coli BW25113	*F- LAM- rrnB3 ΔlacZ4787 hsdR514 Δ(araBAD)567 Δ(rhaBAD)568 rph-1*	DSMZ, Germany
BW25113 Δ*tesC*	*E. coli* BW25113 Δ*tesC::FRT*	This study
BW25113 Δ*tesC ΔyqhC-yqhD-dkgA*	BW25113 Δ*tesC* Δ*yqhC-yqhD-dkgA::FRT*	This study
BW25113 Δ*adhE*	*E. coli* BW25113 Δ*adhE::FRT*	Chapter 3

Table 4.2. Plasmids used and constructed in this work.

Name	Contents	Reference
pKD3	FRT-cm^R-FRT, oriR6K, amp^R	(Datsenko & Wanner, 2000)
pSIJ8	araC-P$_{ara}$-α-β-γ rhaRS-P$_{rha}$-FLP, amp^R	(Jensen et al., 2015)
DVC_p15A_AE	Cm^R level 1 destination vector, low copy p15A ori, AE cloning sites	Chapter 3
DVC_p15A_FG	Cm^R resistant level 1 destination vector, p15A ori, FG cloning sites	Chapter 3
DVK_p15A_FG	Kan^R resistant level 1 destination vector, p15A ori, FG cloning sites	Chapter 3
DVC_pBR322_AE	Cm^R level 1 destination vector, pBR322 ori, AE cloning sites, AE cloning sites	Chapter 3
DVA_p15A_AF	Kan^R level 1 destination vector, p15A ori, AF cloning sites	Chapter 3
DVA_p15A_EF	Kan^R level 1 destination vector, p15A ori, EF cloning sites	Chapter 3
DVA_pBR322_AF	Kan^R level 1 destination vector, pBR322 ori, AF cloning sites	Chapter 3
DVA_pUC19_CD	Amp^R level 0 destination vector, CD cloning sites (renamed to include ori type)	(Iverson et al., 2016)
DVA_J23100_AB	Amp^R vector containing J23100 constitutive Anderson promoter part, level 0 part	(Iverson et al., 2016)
DVA_J23107_AB	Amp^R vector containing J23107 constitutive Anderson promoter part, level 0 part	(Iverson et al., 2016)
DVA_J23116_AB	Amp^R vector containing J23116 constitutive Anderson promoter part, level 0 part	(Iverson et al., 2016)
DVA_B0034m_BC	Amp^R vector containing B0034m RBS part, level 0 part	(Iverson et al., 2016)
DVA_B0015m_BC	Amp^R vector containing B0015m terminator part, level 0 part	(Iverson et al., 2016)
pDS_120_1R1_kan	Kan^R destination vector, low copy RSF1010 ori, AF cloning sites	(Stukenberg et al., 2021)
pMC0_2_42_Ptet*	Cm^R vector containing *tetR*/P$_{tet}$ part	(Stukenberg et al., 2021)
pST_116_lvl2_cam	Cm^R vector containing *lacI*/P$_{tac}$ part	(Stukenberg et al., 2022)
pET3a_luxCDE_PCC7942_ADO	Contains the *luxCDE* operon from *P. luminescens* and aldehyde deformylating oxygenase (ADO) from *S. elongatus* PCC7942	Habgood, R., 2020 (unpublished data)
pL0_JRN001_CD	DVA_pUC19_CD containing *S. elongatus* PCC7942 aldehyde deformylating oxygenase (ADO)	This study
pL0_JRN002_CD	DVA_pUC19_CD containing *S. elongatus* PCC7942 acyl-ACP reductase (AAR)	This study
pL0_JRN005_CD	DVA_pUC19_CD containing domesticated *Pseudomonas sp. 101* NAD$^+$-dependent formate dehydrogenase	This study
pL0_JRN012_CD	DVA_pUC19_CD containing level 0 part containing domesticated *E. coli* MG1655 *adhE*	Chapter 3

pL0_JRN017_CD	DVA_pUC19_CD containing level 0 part containing domesticated *E. coli* MG1655 *adhE*^{A267T}	Chapter 3
pL0_JRN013_CD	DVA_pUC19_CD containing level 0 part containing domesticated *E. coli* MG1655 *adhE*^{E568K}	This study
pL0_JRN018_CD	DVA_pUC19_CD containing level 0 part containing domesticated *E. coli* MG1655 *adhE*$^{A267T/E568K}$	Chapter 3
pL1_JRN067_FG	DVK_p15A_FG containing J23100-B0034m-*Pseudomonas sp.* NAD^+-dep. FDH	This study
pL1_JRN080_AE	DVK_p15A_AE *lacI*/P_{tac}-B0034m-*adhE*	This study
pL1_JRN081_AE	DVK_p15A_AE *lacI*/P_{tac}-B0034m-*adhE*^{A267T}	This study
pL1_JRN082_AE	DVK_p15A_AE *lacI*/P_{tac}-B0034m-*adhE*^{E568K}	This study
pL1_JRN083_AE	DVK_p15A_AE *lacI*/P_{tac}-B0034m-*adhE*$^{A267T/E568K}$	This study
pL1_JRN094_FG	DVK_p15A_FG J23100-B0034m-*adhE*$^{A267T/E568K}$	This study
pL1_JRN095_FG	DVK_p15A_FG J23107-B0034m-*adhE*$^{A267T/E568K}$	This study
pL1_JRN096_FG	DVK_p15A_FG J23116-B0034m-*adhE*$^{A267T/E568K}$	This study
pL1_JRN097_AF	pRSF1010 containing *lacI*/P_{tac}-B0034m-*fadR*	This study
pL1_JRN098_AF	pRSF1010 containing *tetR*/P_{tet}-B0034m-*fadR*	This study
pL1_JRN115_AF	pRSF1010 containing *lacI*/P_{tac}-B0034m-*iclR*	This study
pL1_JRN116_AF	pRSF1010 containing *tetR*/P_{tet}-B0034m-*iclR*	This study
pL1_JRN132_AE	DVC_p15A_AE containing *lacI*/P_{tac}-B0034m-AAR	This study
pL1_JRN133_AE	DVC_pBR322_AE containing *lacI*/P_{tac}-B0034m-AAR	This study
pL1_JRN139_FG	DVC_p15A_FG J23100-B0034m-*adhE*$^{A267T/E568K}$	This study
pL1_JRN145_EF	DVC_p15A_EF containing P_{tac}-B0034m-ADO	This study
pL2_JRN072_AF	DVA_pBR322_AF containing *lacI*/P_{tac}-B0034m-AAR and P_{tac}-B0034m-ADO	This study

4.3.2 Cloning procedures

All primer sequences can be found in supplementary table S4.1. Coding DNA sequences of *Synechococcus elongatus* PCC7942 acyl-ACP reductase (AAR) as well as *Pseudomonas sp.* 101 formate dehydrogenase (PsFDH), were ordered as codon optimised gene fragments (Twist Bioscience, USA) (Table S4.2). The plasmids used in this study were routinely constructed by Golden Gate Assembly (GGA), using a selection of parts from the CIDAR MoClo kit (Iverson et al., 2016), Marburg collection MoClo kit (Stukenberg et al., 2021) and self-constructed MoClo-compatible medium copy vectors (see Chapter 3). Several custom level 0 parts were cloned into DVA_pUC19_CD by amplifying them using PCR and introducing compatible overhangs, including *S. elongatus* PCC7942 AAR (JRN275+276), *S. elongatus* PCC7942 ADO (JRN239+240), *E. coli fadR* (JRN498+499) and *E. coli iclR* (JRN504+505). PsFDH was domesticated and cloned into DVA_pUC19_CD by amplifying fragments up- and

downstream of the target site introducing silent mutations through primer overhangs (JRN253, JRN254, JRN255, JRN256). The E568K mutation was introduced in domesticated *E. coli adhE* by linearizing pL0_JRN012_CD using JRN189+190 and performing Gibson assembly using this linearised part and an oligo (JRN191) introducing the mutation, as per manufacturer protocol specifications (HiFi Assembly, NEB).

To construct low copy *fadR* and *iclR* expression plasmids (pL1_JRN097_AF, pL1_JRN098_AF, pL1_JRN115_AF, pL1_JRN116_AF) we used the pDS_120_1R1_kan backbone from the Marburg collection MoClo kit. This vector is compatible with the CIDAR MoClo standard for level 1 plasmid construction, as it requires BsaI-mediated restriction. The *tetR*/P_{tet} part from the Marburg collection (pMC0_2_42_Ptet*) was used to generate anhydrotetracycline (aTC) inducible level 1 constructs (Meyer et al., 2018; Stukenberg et al., 2021). To generate IPTG-inducible constructs expressing *lacI*, the *lacI*/P_{tac} part from pST_116_LVL2_cam was amplified using primers oRZ36 and oRZ37, introducing BsaI-digestible flanks for MoClo assembly (Table S4.1) (Meyer et al., 2018; Stukenberg et al., 2022). The IPTG-inducible promoter regions (without *lacI*) found in the CIDAR MoClo kit contain CAP binding sites, meaning they are repressed during glucose catabolism. To construct level 1 IPTG-inducible parts that do not contain CAP binding sites nor carry *lacI,* complementary oligos were designed (JRN492+JRN493) coding for the hybrid P_{tac} promoter (Meyer et al., 2018) carrying BsaI-digestible overhangs to use as MoClo-compatible level 0 parts. These oligos were annealed by combining 100 µM of each in 50 µL DI water, heating at 98^{o}C for 5 minutes, and cooling to 25^{o}C at 0.1^{o}C/s. A 1 µL aliquot of 10x diluted annealed oligo was used in level 1 construct assembly.

Golden Gate cloning protocols were performed mostly as described previously (Iverson et al., 2016). Briefly, we used 20 fmol of destination vector and 40 fmol of each constituent part in a total of 20 µL reaction volume, containing 2 µL 10x T4 DNA ligase buffer (NEB), 500 units of T4 DNA ligase (NEB) and 10 units of either BsaI-HFv2 (NEB) or BbsI-HF (NEB) for level 1 or 0/2 assemblies, respectively.

4.3.3 Gene deletions

The deletion of *tesC* and the *yqhC-yqhD-dkgA* operon were performed by lambda red recombination mostly according to previous work (Jensen et al., 2015). Briefly, *E. coli* BW25113 was transformed with pSIJ8, encoding the lambda red recombination and flippase genes. Transformed cells were grown selectively overnight at 30°C and 1 mL of O/N culture was used to inoculate 50 mL of LB + carb in a shake flask. This culture was left to grow for 1 hour at 30ºC, followed by pSIJ8 induction for 45 minutes by addition of sterile 15 mM L-arabinose. Cells were made electrocompetent and finally resuspended in 300 µL ice-cold sterile 10% glycerol. DNA cassettes targeting the relevant locus were generated by PCR amplification of pKD3 (see primer table S4.1). DNA was purified (NEB) and eluted in DI water. For electroporation of the marker, 50 µL of electrocompetent cells were mixed with 3.5 µL of marker (>250 ng/µL) and transferred to an ice-cold electroporation cuvette (0.1 cm gap). A 1.8 kV pulse was applied, and cells were recovered immediately for 2.5 hours at 30ºC before plating on LB + carb + cm (17 µg/mL) and incubated at 30ºC. Positive clones were identified by colony PCR using DreamTaq Green 2x Master Mix (see Table S4.1 for relevant primers) (ThermoFisher). Integrated chloramphenicol markers were removed by inoculating 35 mL of LB + carb with 0.5 mL of overnight *E. coli* BW25113 with marker inserted. After 1.5 hours shaking at 30°C, flippase expression was induced with 50 mM L-rhamnose for 4 hours before plating 10^{-5} and 10^{-6} dilutions on LB + carb plates. Clones with removed markers were identified by colony PCR and pSIJ8 was cured by a single overnight cultivation in LB at 42ºC.

4.3.4 Microtiter plate growth assays

For plate reader growth assays strains were first streaked on LB agar supplemented with antibiotics where required, and with filter-sterilised carbon sources intended to be used in the growth assay. Growth took place at 30°C or 37°C, depending on the exact assay. Single colonies were isolated and pre-cultured in triplicate in liquid LB medium overnight (250 rpm), again including supplements where required, before inoculating them at a 1:100 v/v ratio into M9 minimal medium. These precultures were used to inoculate 200 µL M9 minimal medium

to an OD_{600} of 0.01 with pre-cultures washed three times in sterile 1x M9 salts, in a 96-well clear flat bottom microtiter plate (Greiner). A Tecan Spark plate reader was used for microtiter plate growth assays using SparkControl software (v3.1 SP1). To avoid contamination and evaporation, plates were covered with a Breathe-Easy Sealing Membrane (Sigma-Aldrich) and cultivation was performed using a small humidity cassette (Tecan) re-filled every 24h. Growth assay were performed at 30- or 37°C and OD_{600} measurements were taken every 15 minutes, using the following four-step shaking profile: 1) double orbital, 3 mm amplitude, 90 rpm frequency, 225 seconds 2) orbital shaking, 3 mm amplitude, 180 rpm, 225 seconds 3) double orbital, 2.5 mm amplitude, 108 rpm, 225 seconds, 4) orbital, 2.5 mm amplitude, 216 rpm, 224 seconds.

4.3.5 Hydrocarbon production via AAR and AAR-ADO

To prepare inoculum for production experiments, strains carrying AAR or AAR/ADO plasmids (and other plasmids required for AdhE, PsFDH, or FadR expression) were streaked out on selective LB agar at 30°C. Single colonies were then inoculated as biological duplicates or triplicates in LB liquid medium (shaking, 250 rpm) supplemented with appropriate antibiotics and filter-sterilised carbon sources (e.g. acetate/ethanol/glucose). After O/N cultivation at 30°C, liquid LB cultures were diluted 100-fold into M9 minimal medium pre-cultures and grown O/N, shaking (250 rpm) at 30°C, supplemented with required antibiotics and the carbon source(s) to be used in the subsequent production experiment as indicated in the text and figure legends. Grown pre-cultures were then diluted 100-fold into 20 mL M9 minimal medium in 100 mL shake flasks, supplemented with antibiotics and filter-sterilised carbon sources as indicated in the text and figure legends. Production experiments were performed at 30°C with 150 rpm shaking using either biological duplicates or triplicates as indicated in text and figure legends. Where necessary, a stock solution of 3 M filter-sterilised potassium chloride (KCl) in DI water was diluted in the M9 medium to a concentration of 100- or 250 mM. If induction of genetic constructs was required, cultures were grown to an OD_{600} of ~0.2 and induced with either IPTG (isopropyl β-D-1-thiogalactopyranoside) or aTC (anhydrotetracycline

hydrochloride) at the appropriate concentrations as indicated in the text and figure legends. Growth was measured daily by OD_{600} measurement, and a 1 mL sample was taken simultaneously for GC-FID and HPLC analysis. If not used immediately, these samples were stored at -80°C.

4.3.6 Gas chromatography (GC-FID)

Samples for GC-FID were routinely prepared by addition of 650 µL ethyl acetate (Sigma-Aldrich) to 500 µL of culture broth sample. The suspension was vortexed vigorously for 10 minutes before spinning down at 14000 *g* for 3 minutes. The resulting upper fraction was used for GC-FID analysis. If separate analysis of pellet and supernatant fractions was required, whole culture samples were first spun down for 10 minutes at 1500 *g*. Supernatant fractions were removed and extracted as before, whereas the remaining pellet fraction was first resuspended in 500 µL of sterile 1x M9 salts before extraction took place. A Shimadzu GC-2010 equipped with a flame ionization detector (FID) and an SH-5 column (30 m length, 0.25 mm ID, 0.25 µm film thickness) was used for analysis with a splitless injection of 1 µL sample using the following oven program: hold at 40°C for 5 minutes, ramp to 320°C at 25°C/minute, hold at 320°C for 2 minutes. Standards of aldehydes (C_{14}, $C_{14:1}$, C_{16}, $C_{16:1}$, C_{18}, $C_{18:1}$), alcohols (C_{14}, $C_{14:1}$, C_{16}, $C_{16:1}$, C_{18}) and alk(a/e)nes (C_{13}, C_{15}, $C_{15:1}$, C_{17}, $C_{17:1}$) (purchased from Sigma-Aldrich, Ambeed, or Biosynth, all >95% grade) were prepared in 100% ethyl acetate and quantified using the same analytical method.

4.3.7 High performance liquid chromatography (HPLC)

Whole culture samples taken for HPLC analysis were spun down at 14000 *g* for 10 minutes, after which 400 µL of the supernatant fraction was removed and diluted with 400 µL of 250 mM propionic acid. A suitable concentration range of organic acid standards (BioRad, USA) and ethanol were prepared in DI water and diluted 2-fold in 250 mM propionic acid before HPLC analysis. A Shimadzu LC2030 Prominence-i device was used for HPLC analysis, equipped with a photodiode array (PDA, 210 nm), refractive index detector (RID-20A), and an

Aminex HPX-87H column (Bio-Rad, USA). A mobile phase consisting of 5 mM H_2SO_4 was used at a flow rate of 0.8 mL/min and a column temperature of 60°C, with a sample injection volume of 10 µL.

4.4 Results

4.4.1 Optimization of acyl-ACP reductase (AAR) activity with glycolytic carbon sources

Before attempting to produce alkanes or use acetate and ethanol as carbon sources, we aimed to establish and optimise the activity of the first enzyme in the alkane biosynthesis pathway, acyl-ACP reductase (AAR), which generates aldehydes from acyl-ACP substrate (Figure 4.1A). The activity of AAR was first assessed using constructs with an IPTG-inducible P_{tac} promoter, a strong RBS (B0034m) and on plasmid backbones with either a p15A (pL1_JRN132_AE) or pBR322 (pL1_JRN133_AE) origin of replication (Figure 4.1B). Induction of pL1_JRN132_AE or pL1_JRN133_AE with 1000 µM IPTG resulted in significantly higher titres of total hydrocarbons (aldehydes and alcohols) compared to non-induced controls (Figure 4.1C). Differences between hydrocarbon titres of BW25113 pL1_JRN133_AE induced with 100- or 1000 µM IPTG were insignificant, as were the differences between pL1_JRN132_AE and pL1_JRN133_AE at all induction levels. The main products resulting from AAR induction were C_{16} alcohols and C_{16} aldehydes.

Next, we assessed the effect of potassium chloride (KCl) as an additive in the production medium. The AAR enzyme has previously been found to be highly unstable, and KCl was reported to significantly stabilise AAR *in vitro* and *in silico* (Lin et al., 2013; Sharma et al., 2021). Even so, no *in vivo* production studies that use KCl supplementation have been reported to date. To determine the toxicity of *E. coli* to KCl we performed growth characterization of *E. coli* BW25113 in M9 glucose medium supplemented with various concentrations of KCl (Figure S4.1). Concentrations up to 250 mM did not significantly affect growth, although higher concentrations led to increased lag phases.

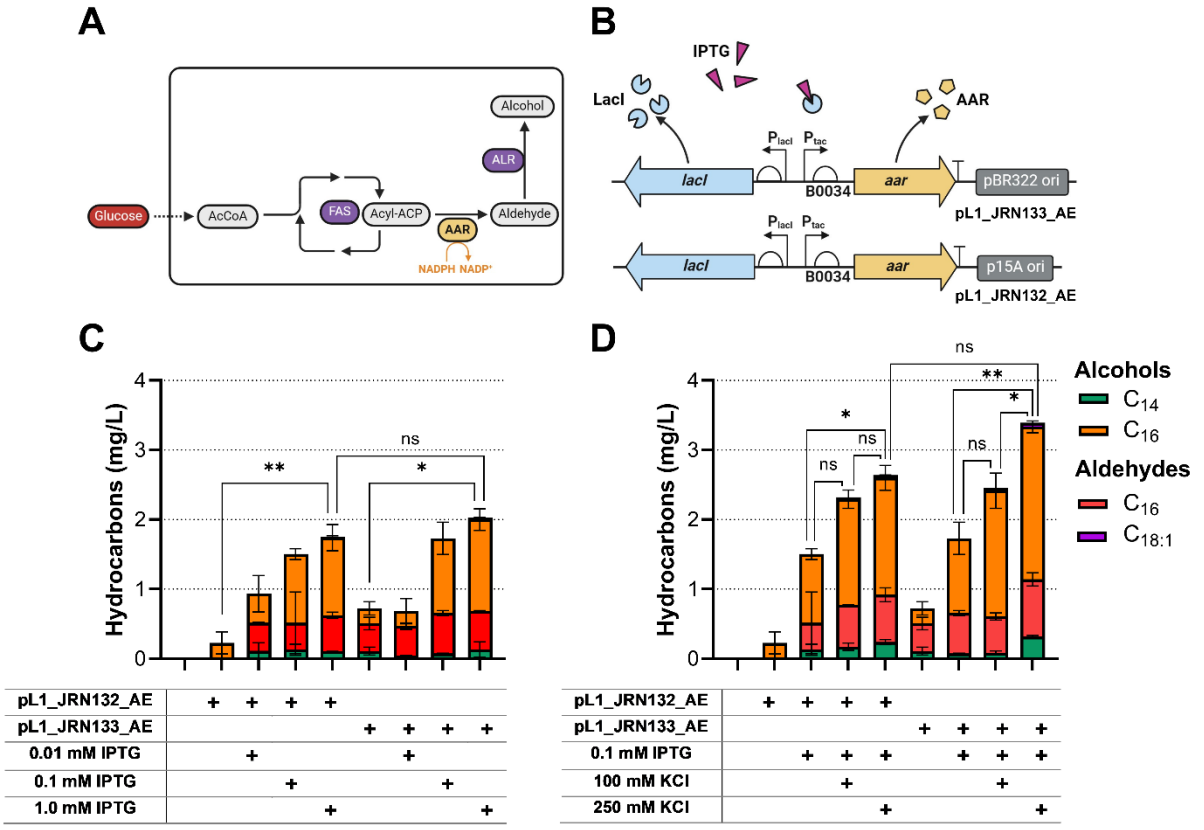


Figure 4.1. Hydrocarbon biosynthesis in *E. coli* BW25113 expressing acyl-ACP reductase (AAR). (**A**) Schematic of the metabolic pathway leading to aldehyde and ultimately alcohol biosynthesis through AAR and alcohol reductase (ALR) activity. (**B**) Schematic of the IPTG-inducible AAR expression constructs pL1_JRN132_AE (top) and pL1_JRN133_AE (bottom), the only difference being their origin of replication (p15A and pBR322, resp.). (**C**) Induction of AAR from pL1_JRN132_AE or pL1_JRN133_AE leads to hydrocarbon biosynthesis, mostly in the form of C_{16} alcohol. (**D**) Expression of AAR in the presence of 100- or 250 mM KCl leads to significantly higher hydrocarbon titres compared to controls without KCl addition. Data represents the averages of biological duplicates, and errors bars represent their standard deviations. Statistical analysis (not explicitly visualised for each result) was performed by comparing means using a one-way ANOVA and post-hoc Tukey's multiple comparison test. P-value significance indicated are as follows: ns (p-value >0.05); * (p-value <0.05), ** (p-value <0.01), *** (p-value <0.001) or **** (p-value <0.0001).

After supplementing production medium with 250 mM KCl we observed significant increases in total hydrocarbon titres in induced cultures of BW25113 pL1_JRN132_AE (~75% increase) and BW25113 pL1_JRN133_AE (~96% increase) (Figure 4.1D). In addition, in the case of BW25113 pL1_JRN133_AE (but not pL1_JRN132_AE) the addition of 250 mM KCl resulted in a significantly higher titre of hydrocarbons compared to supplementation with 100 mM KCl (Figure 4.1D). In these conditions, no effects of plasmid origin of replication (p15A or pBR322) on total hydrocarbon titres were detected.

Thioesterases are responsible for hydrolysing the thioester bond between a fatty acid and CoA or ACP, releasing free fatty acids and CoA or ACP, respectively. Since AAR utilises acyl-ACP or acyl-CoA as its substrate, thioesterase activity directly competes with AAR's substrate pool (Figure 4.2A). Previous work has shown that deletion of known thioesterases in *E. coli*, especially *tesC* (*ybaW/fadM*), can lead to significantly increased titres of fatty alcohols when expressing the fatty acyl-CoA reductase (Liu et al., 2016). We therefore decided to delete *tesC,* resulting in BW25113 Δ*tesC*, and transformed this strain with AAR-expression plasmids pL1_JRN132_AE and pL1_JRN133_AE to assess their production potential. High induction of pL1_JRN133_AE (1000 µM IPTG) and 250 mM KCl addition led to the formation of ~9 mg/L of total hydrocarbons, of which the majority (7 mg/L) was C_{16} alcohol (Figure 4.2B). Compared to *E. coli* BW25113, the additional *ΔtesC* deletion led to a ~30% higher hydrocarbon titre when using pL1_JRN133_AE induced with 0.1 mM IPTG, indicating that deletion of this competing pathway was indeed leading to a higher AAR flux. Hydrocarbon biosynthesis titres of BW25113 Δ*tesC* were also significantly higher when using pL1_JRN133_AE compared to pL1_JRN132_AE induced with 0.1- or 1 mM IPTG, demonstrating that in these conditions a higher copy number plasmid led to increased product synthesis (Figure 4.2B).

To potentially further increase carbon flux through fatty acid synthesis towards AAR, we transformed BW25113 Δ*tesC* carrying either pL1_JRN132_AE or pL1_JRN133_AE constructs with the additional pL1_JRN097_AF or pL1_JRN098_AF low-copy plasmids, which carry the *E. coli* fatty acid degradation regulator (*fadR*) under *lacI* or *tetR* control, respectively (Figure 4.3B). This regulator acts as a transcriptional activator for genes involved in fatty acid biosynthesis, including *accABCD, fabD, fabH, fabG, fabA, fabZ and fabB*, and as a transcriptional repressor of beta oxidation genes (Figure 4.3A) (Cronan, 2021; Janßen & Steinbüchel, 2014). Expression of *fadR* has been found to stimulate fatty acid biosynthesis in *E. coli*, and this has been exploited to significantly improve titres of free fatty acids and alkanes (Song et al., 2016; Zhang et al., 2012). First, *fadR* under *lacI* control was co-expressed with AAR at various IPTG induction levels. The growth of BW25113 Δ*tesC* pL1_JRN133_AE

pL1_JRN097_AF induced with 0.1- or 1 mM IPTG was somewhat inhibited in the first 18 hours, probably owing to the high expression burden of AAR and several Fab enzymes (Figure 4.3C)

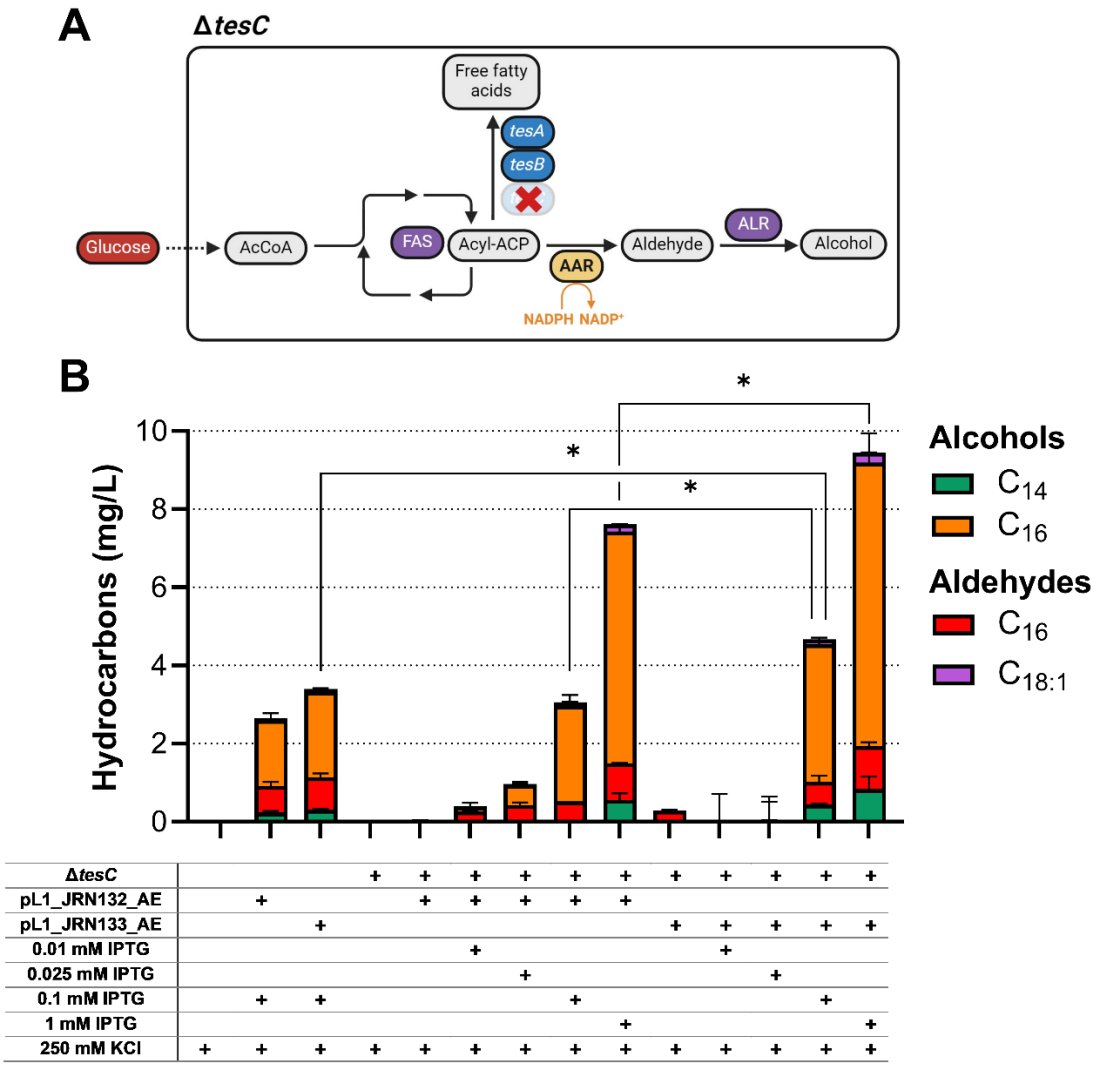


Figure 4.2. Effects of deleting *tesC* on hydrocarbon biosynthesis via acyl-ACP reductase (AAR). (A) Simplified schematic of *E. coli* metabolism; the *ΔtesC* mutation partially blocks thioesterase activity leaving a larger pool of acyl-ACP available for reduction to aldehydes by AAR. Endogenous alcohol reductases (ALRs) can further reduce aldehydes to the corresponding alcohols. **(B)** Hydrocarbon biosynthesis of AAR-expressing *E. coli* cultures, with combinatorial screening of conditions including use of *E. coli* Δ*tesC*, copy number of the AAR-carrying plasmid (p15A, pL1_JRN132_AE or pBR322, pL1_JRN133_AE) and induction level of P_{tac}-AAR by IPTG concentration. All measurements were deduced from biological duplicates, errors bars represent the standard deviation thereof. Statistical analysis (not explicitly visualised for each result) was performed by comparing means using a one-way ANOVA and post-hoc Tukey's multiple comparison test. P-value significance indicated are as follows: ns (p-value >0.05); * (p-value <0.05), ** (p-value <0.01), *** (p-value <0.001) or **** (p-value <0.0001).

Interestingly, we found that the addition of 0.01 mM IPTG resulted in significantly higher OD_{600} values after 18 hours of growth compared to non-induced cultures or those not carrying the *fadR* plasmid. We found that strong induction with 1 mM IPTG did not significantly improve hydrocarbon titres compared to expressing AAR at the same strength without FadR (Figure 4.2B and 4.3D). In the 1 mM IPTG condition, the use of either a p15A- or pBR322 backbone to express AAR did not result in significant titre differences when FadR was also expressed (Figure 4.3D). However, FadR and AAR expression using 0.1 mM IPTG resulted in significant increases in hydrocarbon titre compared to induction of just AAR with 0.1 mM IPTG, both when using pL1_JRN132_AE (2.2-fold increase) or pL1_JRN133_AE (1.7-fold increase) (Figure 4.3D). This indicated that the presence of FadR could improve flux through AAR, but that high induction conditions were sufficient to maximise AAR-mediated hydrocarbon production without additional FadR expression, possibly indicating a saturation of AAR activity in the tested conditions. Since the *lacI*/P_{tac}-*fadR* construct was controlled by the same inducer as AAR, it was not possible to assess high AAR- and low *fadR* induction simultaneously. We therefore used an aTC-inducible *tetR*/P_{tet}-*fadR* construct, pL1_JRN098_AF, to control FadR expression independently of AAR. The growth of strains carrying pL1_JRN098_AF displayed clear responses to the aTC induction level (Figure 4.3E). Firstly, high aTC induction (100 ng/mL) resulted in a significant growth inhibition which was more pronounced than inducing pL1_JRN097_AF with 1 mM IPTG. Secondly, similar to the IPTG-induced *fadR* construct, the low induction of *fadR* (1 ng/mL aTC) resulted in significantly higher OD_{600} at t_{18} compared to a non-induced culture, indicating a growth improvement that could be attributed to the expression of *fadR*. As for hydrocarbon biosynthesis, the expression of AAR using 1 mM IPTG and *fadR* without induction or 2.5 ng/mL aTC resulted in a significantly increased hydrocarbon titre compared to cultures with no *fadR* expression construct, or those induced with 1- or 100 ng/mL aTC (Figure 4.3F).

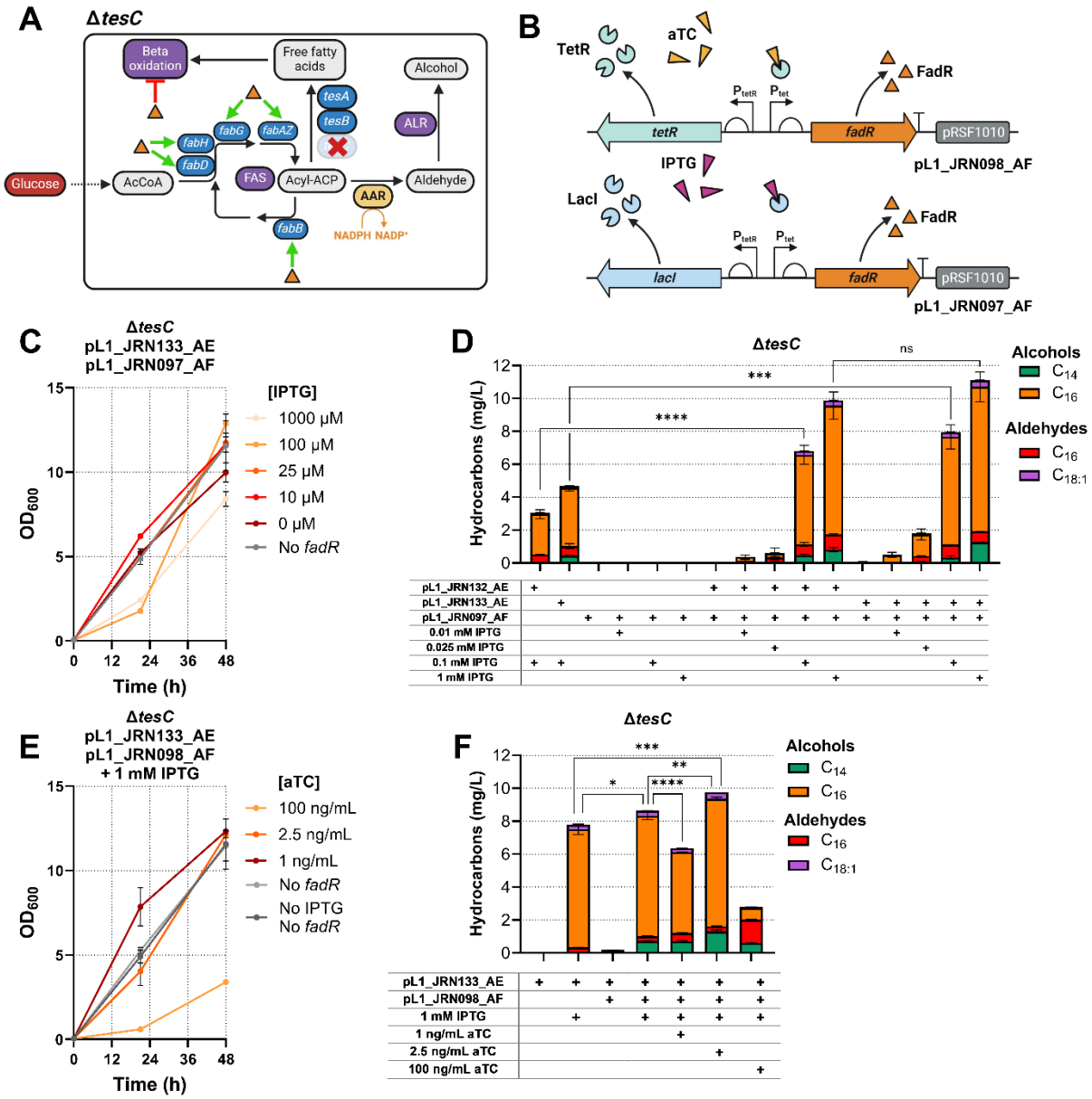


Figure 4.3. Effects of fatty acid degradation regulator (*fadR*) expression on growth and hydrocarbon biosynthesis using acyl-ACP reductase (AAR) in *E. coli* Δ*tesC*. (A) Schematic of *E. coli* metabolism including fatty acid biosynthesis genes transcriptionally regulated by *fadR*. Green arrows indicate upregulation, red T indicates repression. **(B)** Genetic constructs used for expression of *fadR*: pL1_JRN098_AF (top) or pL1_JRN097_AF (bottom), both in a low copy pRSF1010 backbone. **(C)** Growth of *E. coli* Δ*tesC* pL1_JRN133_AE pL1_JRN097_AF under different IPTG induction levels. **(D)** Hydrocarbon biosynthesis of *E. coli* Δ*tesC* cultures, with combinatorial screening of conditions including copy number of the AAR-carrying plasmid (p15A in pL1_JRN132_AE or pBR322 in pL1_JRN133_AE), presence of pL1_JRN097_AF, and IPTG induction levels. **(E)** Growth of *E. coli* Δ*tesC* carrying both pL1_JRN133_AE and pL1_JRN098_AF under different aTC induction levels. The pL1_JRN133_AE plasmid was induced with 1 mM IPTG unless otherwise indicated. **(F)** Hydrocarbon biosynthesis in *E. coli* Δ*tesC* pL1_JRN133_AE cultures with or without 1 mM IPTG induction and in the presence of pL1_JRN098_AF under various aTC induction levels. All measurements were deduced from biological duplicates, errors bars represent the standard deviation thereof. Statistical analysis (not explicitly visualised for each result) was performed by comparing means using a one-way ANOVA and post-hoc Tukey's multiple comparison test. P-value significance indicated are as follows: ns (p-value >0.05); * (p-value <0.05), ** (p-value <0.01), *** (p-value <0.001) or **** (p-value <0.0001).

Supplementation with 100 ng/mL aTC led to low hydrocarbon biosynthesis, however, it was notable that C_{16} aldehyde levels were around 5-fold higher compared to cultures without *fadR* induction even though C_{16} alcohol production was lower. This indicated that AAR was active but that the conversion of aldehydes to alcohols by endogenous enzymes was still ongoing. We found no significant difference in hydrocarbon production between strains expressing *fadR* from either a *lacI*/P_{tac} or a *tetR*/P_{tet} construct.

The results presented in Figure 4.3 seemed to point to a saturation of AAR flux, since increased plasmid copy number and *fadR* overexpression did not result in meaningful improvements in hydrocarbon titre compared to the previous results, at least not when using high induction levels (Figure 4.2). A recent study showed that introduction of a His-12 tag on the N-terminus of AAR improved alkane biosynthesis approximately 7-fold (Han et al., 2023). Our expectation was that this modification could contribute to an improvement of AAR flux by improving its stability. To assess what effect this modification had on AAR activity without ADO, we used the same *lacI*/P_{tac} construct as before but with AAR carrying an N-terminal His-12 tag and grew it in the same KCl-supplemented glucose minimal medium. Contrary to expectations, we found that expression of the His-12 tagged AAR with either 0.1- or 1 mM IPTG resulted in a 48h lag phase compared to non-induced controls. Once the strains had fully grown, we did not detect any hydrocarbons in the culture broth indicating that the addition of the 12-His tag had resulted in a complete loss of AAR activity (data not shown). Finally, we attempted to use gluconate, a more oxidised sugar than glucose, for hydrocarbon biosynthesis. We reasoned that since gluconate enters central metabolism in the pentose phosphate pathway, its partial catabolism via 6-phosphogluconate dehydrogenase (*gnd*) would lead to increased levels of NADPH – a cofactor in high demand for fatty acid synthesis and AAR activity. While *E. coli* Δ*tesC* grew similarly well on 100 mM gluconate as on 100 mM glucose, we observed almost no hydrocarbon production using gluconate (data not shown).

4.4.2 Alkane biosynthesis with glucose as a carbon source

Having identified several optimised conditions for AAR activity, we next moved towards alkane biosynthesis by constructing a level 2 plasmid containing IPTG-inducible *lacI*/P_{tac}-AAR and P_{tac}-ADO constructs in a DVA_pBR322_AF backbone forming pL2_JRN072_AF (Figure 4.4A). The accumulation of alkanes in *E. coli* relies on the availability of fatty aldehyde substrate produced by AAR, which can be hindered by the activity of endogenous aldehyde reductases (ALRs). In fact, endogenous ALRs were responsible for the fatty alcohol production reported in the previous results, and indeed multiple studies have characterised and/or deleted groups of endogenous *E. coli* ALRs often with the goal of improving aldehyde accumulation (Kunjapur et al., 2014; Rodriguez & Atsumi, 2014). We therefore deleted the *yqhC-yqhD-dkgA* operon in the existing Δ*tesC* background, encoding two broad specificity NADPH-dependent ALRs (*yqhD* and *dkgA*) and an ALR transcriptional activator *yqhC*, with the aim to reduce alcohol formation (Figure 4.4B).

Besides testing the effect of deleting this ALR operon, we used 250 mM KCl supplementation as before, but did not include *fadR* overexpression. Several studies exploring AAR-ADO expression levels have reported that low induction of AAR-ADO led to the highest production (Fatma et al., 2018; Song et al., 2015). Therefore, we assessed a series of IPTG concentrations (0.01-, 0.025-, 0.1- and 1 mM) for their effect on alkane production, with the aim of identifying an optimal expression level.

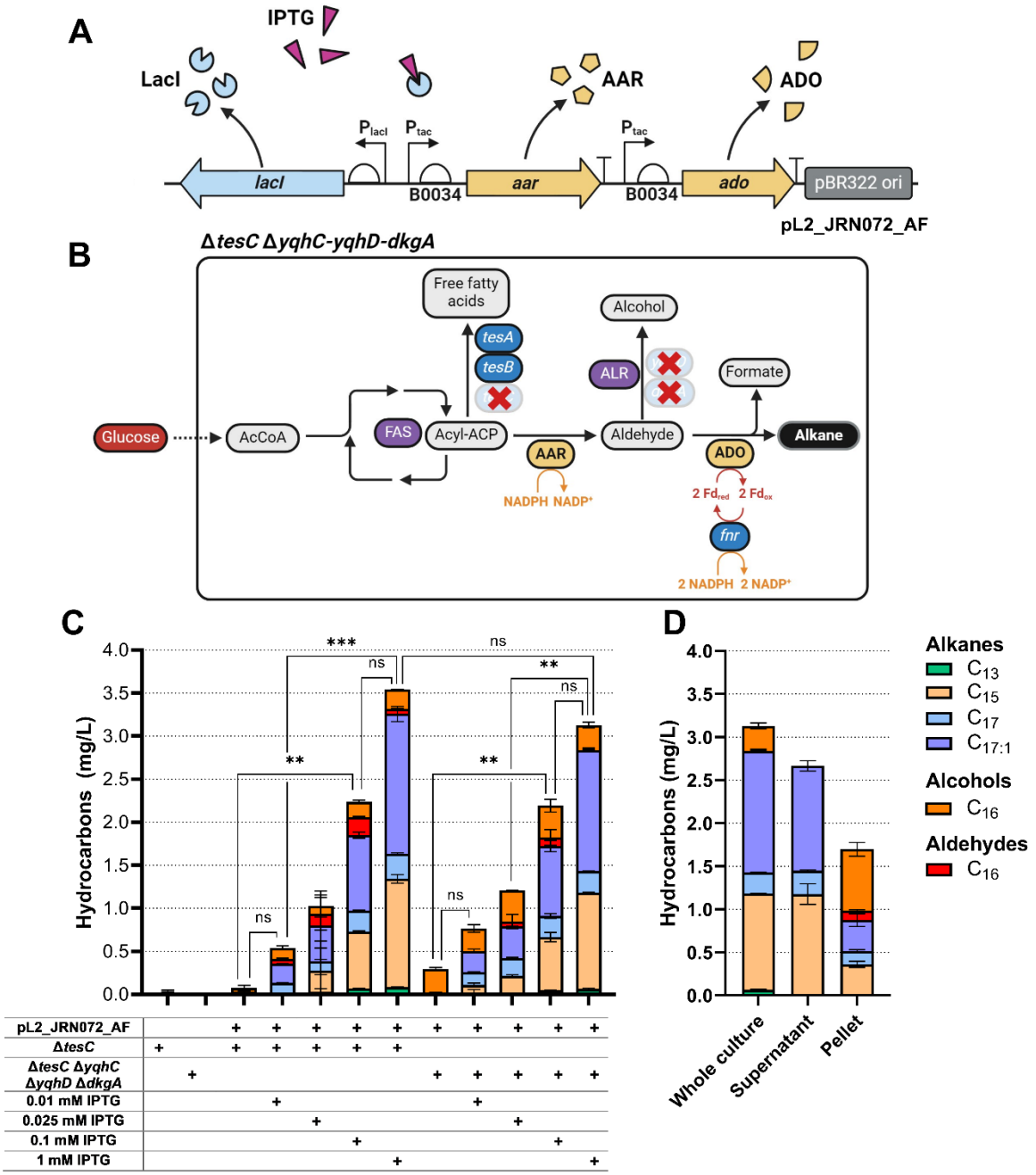


Figure 4.4. Alkane biosynthesis in *E. coli* using glucose. (A) Schematic of pL2_JRN072_AF used for alkane biosynthesis, featuring two separate IPTG inducible P_{tac} promoters and strong B0034m RBSs driving AAR and ADO expression. **(B)** Schematic of *E. coli* metabolism including the acyl-ACP reductase (AAR) and aldehyde deformylating oxygenase (ADO) resulting in alkane and formate production. The deletion of the Δ*yqhC-yqhD-dkgA* operon, encoding two endogenous aldehyde reductases (ALRs) was expected to reduce aldehyde reductase activity. **(C)** Alkane and alcohol/aldehyde biosynthesis in *E. coli* Δ*tesC*, or with the additional Δ*yqhC-yqhD-dkgA* operon deleted. Hydrocarbon biosynthesis generally increased at higher IPTG concentrations. **(D)** Analysis of the supernatant and pellet fractions of *E. coli* Δ*tesC* Δ*yqhC-yqhD-dkgA* pL2_JRN072_AF induced with 1 mM IPTG, compared to the whole culture extractions. All measurements were biological duplicates, errors bars represent the standard deviation thereof. Statistical analysis (not explicitly visualised for each result) was performed by comparing means using a one-way ANOVA and post-hoc Tukey's multiple comparison test. P-value significance indicated are as follows: ns (p-value >0.05); * (p-value <0.05), ** (p-value <0.01), *** (p-value <0.001) or **** (p-value <0.0001).

Using AAR-ADO induction levels 0.1 or 1 mM IPTG resulted in increased alkane titre compared to non or weakly induced (0.01 or 0.025 mM IPTG) conditions, reaching a maximum of ~3 mg/L when induced with 1 mM IPTG (Figure 4.4C). The fatty aldehyde and alcohol titres were strongly reduced compared to constructs solely expressing AAR, presumably as aldehyde substrates were deformylated to alkanes instead of released extracellularly and/or reduced to alcohols by ALRs. Interestingly, the previously measured maximum hydrocarbon titre of AAR was around 10 mg/L, so the ~3 mg/L of alk(a/e)nes achieved through expression of both AAR and ADO constitutes an approximate 30% efficiency in channelling native fatty acyl substrate to alkanes via AAR. Low titres of aldehyde or alcohol product were detected besides alkanes (total ~0.3 mg/L at 1 mM IPTG), which contrasts with other studies expressing AAR-ADO reporting that the majority of hydrocarbon products were alcohols (Fatma et al., 2018; Schirmer et al., 2010). Contrary to expectations, the deletion of the aldehyde reductase operon *yqhC-yqhD-dkgA* did not result in significant changes in the hydrocarbon titres. After separating the pellet and supernatant fractions of the BW25113 Δ*tesC* Δ*yqhC-yqhD-dkgA* pL2_JRN072_AF strain induced with 1 mM IPTG, we found that the majority (67%) of alkane product was found extracellularly, corroborating initial work on the characterization of alkane biosynthesis in *E. coli* (Schirmer et al., 2010) (Figure 4.4D). We note that the sum of hydrocarbon products detected in supernatant and pellet fractions combined was around 35% higher than the titres detected when extracting from whole culture samples. Interestingly, fatty aldehyde and alcohol products were also not detected in the supernatant fraction, suggesting that they were all cell-associated.

After confirming and optimizing the functionality of both AAR and the AAR-ADO pathway, we focused our attention on using non-sugar carbon sources for hydrocarbon synthesis.

4.4.3 Characterization of *E. coli* growth using acetate and formate

Acetate enters *E. coli* central metabolism at the acetyl-CoA node through the activity of either phosphoacetyl transferase (*pta*) and acetate kinase (*ackA*) or acetyl-CoA synthase (*acs*). Although both these pathways retain all carbon in acetate to acetyl-CoA, a drawback of using

acetate as a carbon source is that no reducing power is generated in its assimilation, requiring either further decarboxylation in electron-yielding reactions in central metabolism or the addition of an external electron donor. We envisaged using acetate as the carbon-supplying substrate for alkane biosynthesis, while formate could be used as an electron donor to provide sufficient reducing power.

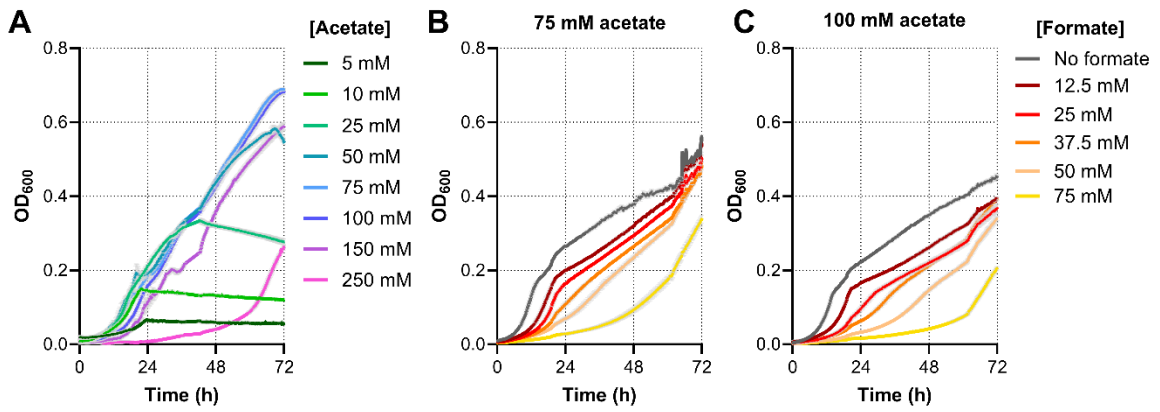


Figure 4.5. Characterising *E. coli* tolerance and growth using acetate and formate. (A) *E. coli* BW25113 was grown in microtiter plates using M9 medium supplemented with various concentrations of sodium acetate. Growth of BW25113 was also assessed using either **(B)** 75 mM- or **(C)** 100 mM sodium acetate and various concentrations of sodium formate. All data represents averages of biological triplicate, and errors bars (grey) represent their standard deviations.

To determine optimal concentrations of acetate and/or formate to use for production experiments, *E. coli* BW25113 was grown in M9 medium supplemented with 5 - 250 mM acetate (0.4 – 20 g/L) (Figure 4.5A). The use of 75- and 100 mM acetate resulted in a maximization of biomass formation (OD_{600}), whereas concentrations of 5-50 mM of acetate resulted in lower OD_{600} values presumably as acetate levels were not sufficient to maximise growth. In contrast, 150- or 250 mM acetate supplementation also resulted in lower OD_{600} values compared to 75- and 100 mM acetate, although this can be attributed to the toxicity of acetate at these concentrations. The 75- and 100 mM concentrations of acetate were deemed suitable for maximizing biomass formation and growth rate. Using these acetate concentrations, the additional supplementation of formate at 12.5 - 75 mM (0.85 – 5.1 g/L) was deleterious to growth regardless of concentration (Figure 4.5B and C). We note that the

no formate control in these two panels depicted poorer growth than the same conditions in Figure 4.5A, which we attributed to differences in the duration of minimal medium pre-culturing. The deleterious effects of formate on growth were dependent on formate concentration, with higher concentrations leading to longer lag phases, especially when 75 mM formate was supplemented. Using 75 mM acetate resulted in better growth at all formate concentration compared to using 100 mM acetate, presumably as total acid concentrations were lower. Based on this data, we decided to use an acetate concentration of 75 mM for all production experiments, with supplemental 50 mM formate where required.

4.4.4 Oxygen tolerance of *adhE* and characterization of *E. coli* growth using ethanol

Ethanol is fermentatively produced by *E. coli* during the anaerobic catabolism of glycolytic carbon sources, where its formation acts as an electron sink for NADH to recycle NAD^+ required for glycolytic flux. Ethanol production is primarily carried out by the bifunctional alcohol/aldehyde dehydrogenase *adhE*, an enzyme with two distinct domains sequentially carrying out the necessary reductions to produce ethanol from acetyl-CoA via an acetaldehyde intermediate, consuming 2 mol NADH per mol ethanol produced in the process. In recent years, the higher order structural organisation of *adhE* has also been studied in great detail, revealing that the protein organises into spirosomes (Kim et al., 2019, 2020). This configuration is essential for activity in the forward ethanol-producing reaction, whereas spirosome formation is not required for ethanol consumption.

The activity of *adhE* was found to be dependent on oxygen concentrations at both the transcriptional- and protein level. Anaerobic conditions led to around a ten-fold increase in expression, congruent with the role *adhE* plays in fermentative metabolism. The structural protein was also reported to be oxygen sensitive, and two seminal works identified the residue E568 as conferring oxygen sensitivity; gene products with a non-acidic amino acid (most notably E568K) at that position were shown to be tolerant to oxygen (Holland-Staley et al., 2000; Membrillo-Hernández et al., 2000).

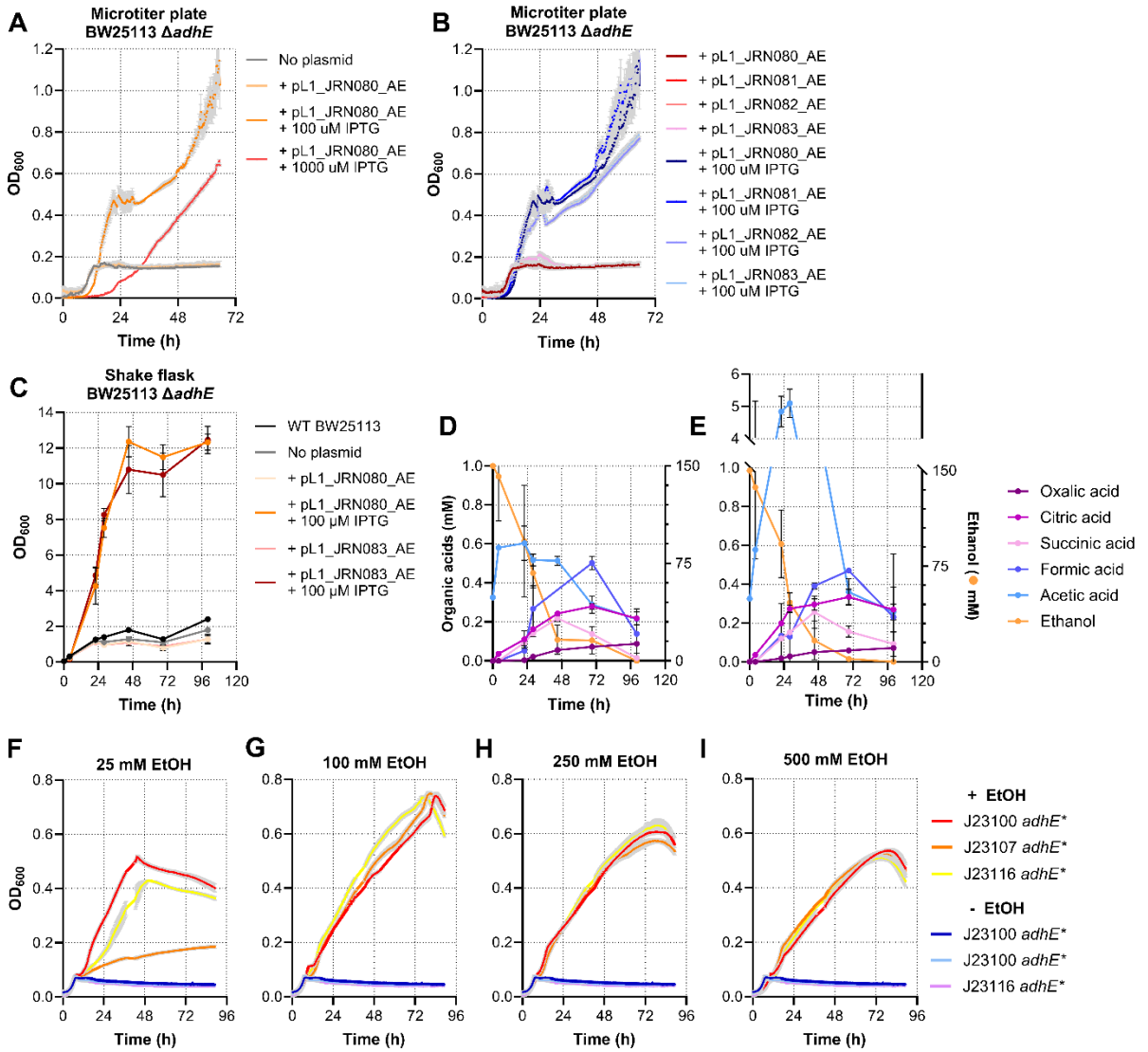


Figure 4.6. Oxygen tolerance, metabolite biosynthesis and expression optimization for AdhE-mediated growth on ethanol. Growth using 1 mM glucose + 150 mM ethanol was assessed in microtiter plates, first using IPTG-inducible constructs of **(A)** wildtype *adhE* (pL1_JRN080_AE) and **(B)** several known variants of *adhE* reported to be involved in oxygen tolerance: *adhE*^{A267T} (pL1_JRN081_AE), *adhE*^{E568K} (pL1_JRN082_AE) *adhE*$^{A267T/E568K}$ also referred to as *adhE** (pL1_JRN083_AE). **(C)** To further confirm the oxygen tolerance of wildtype *adhE* for ethanol oxidation, *E. coli* was grown with or without induction of *adhE* or *adhE** in well-aerated shake flasks containing 1 mM glucose + 150 mM ethanol. The biosynthesis of organic acids and consumption of ethanol was measured by HPLC in cultures expressing **(D)** wildtype *adhE* or **(E)** *adhE**. Finally, the effects of promoter strength driving constitutive *adhE** expression on growth of *E. coli* BW25113 Δ*adhE was* assessed in microtiter growth assays using 1 mM glucose and **(F)** 25-, **(G)** 100-, **(H)** 250-, or **(I)** 500 mM ethanol. The following promoters were tested in combination with a B0034m RBS: J23100 (pL1_JRN094_FG), J23107 (pL1_JRN095_FG) and J23116 (pL1_JRN096_FG). All growth assays and metabolite concentrations were derived from biological triplicate measurements, error bars represent their respective standard deviations.

Besides E568K, a second mutation A267T was shown to stabilise *adhE* and improve acetaldehyde dehydrogenase activity in the reverse, ethanol-consuming reaction. The resulting double mutant *adhE*$^{E568K/A267T}$ (*adhE**) has been used in studies requiring aerobic ethanol consumption (Sun et al., 2020).

To assess *E. coli* growth on ethanol and confirm WT *adhE* oxygen sensitivity, we transformed *E. coli* Δ*adhE* with pL1_JRN080_AE, carrying an IPTG-inducible *lacI*/P_{tac}-*adhE* construct (Figure 4.6A). To our surprise, inducing WT *adhE* expression using 0.1 or 1 mM IPTG was sufficient to support growth on in M9 medium with 1 mM glucose and 150 mM ethanol, while non-induced cultures or those not carrying *adhE* were unable to grow after the glucose was consumed. Strong induction with 1 mM IPTG resulted in poorer initial growth compared to 0.1 mM IPTG supplementation, possibly due to the high expression burden. We note that supplementation of the ethanol medium with a small amount of glucose was necessary to 'kickstart' growth on ethanol, as inoculation of single colonies into minimal medium pre-cultures with ethanol as sole carbon source did not result in growth over 96h (data not shown). To assess whether the mutations identified in earlier works (A267T, E568K) had any effect on growth behaviour, we constructed lacI/P_{tac} constructs with *adhE*^{A267T}, *adhE*^{E568K}, and double mutant *adhE**. The constructs were induced using 0.1 mM IPTG, and expression of WT *adhE* resulted in similar growth to cultures expressing *adhE* carrying A267T, E568K or both mutations when cultivated aerobically in microtiter plates (Figure 4.6B). Expression of the *adhE* and *adhE*^{A267T} constructs seemed to result in higher growth rates from 48 hours onwards, although we could observe clumping of cell material in the microtiter plate wells at the end of the assay which could be responsible for these higher OD_{600} values (not shown). The use of microtiter plates for cultivation may limit oxygen gas-transfer due to the limited volume to surface area and increased surface tension effects (Duetz, 2007). To rule out that micro- or local anaerobic conditions were causing the observed activity of WT *adhE* enzymes, we repeated the assay in aerobically agitated shake flasks (Figure 4.6C). We still observed growth of cultures expressing unmodified *adhE* at rates similar to *adhE**. Sequencing of three isolates

at the end of cultivation confirmed that no spontaneous mutation had occurred in the WT *adhE* coding sequence. These results indicated that the *adhE* enzyme is perhaps not as oxygen sensitive *in vivo* as previously thought, at least not in the ethanol-oxidizing direction, which contradicted two seminal works in the field (Holland-Staley et al., 2000; Membrillo-Hernández et al., 2000). To identify any differences in the metabolism of *adhE*- or *adhE**-expressing cultures, we performed HPLC analysis on culture supernatant over the course of the shake flask assay. The production profiles of oxalic-, citric-, succinic- and formic acid were similar in either culture, as was the consumption profile of ethanol (Figure 4.6D and E). However, the production profile of acetic acid (acetate) revealed a difference between the two cultures. The strain expressing WT *adhE* generated a maximum of ~0.6 mM acetate (Figure 4.6D), while the expression of mutated *adhE** resulted in a maximum acetate concentration of around ~5 mM (Figure 4.6E). The formic acid measured in culture broth of both ethanol-grown strains implies the activity of a formate-producing enzyme, which is most commonly pyruvate-formate lyase (*pflB*) in *E. coli*. This enzyme is often assumed to be oxygen sensitive, so its activity is unexpected in aerobically grown *E. coli* cultures, implying that either the conditions are not fully aerobic or that these specific conditions enable aerobic formate production.

Despite these interesting observations, we still chose to continue our work with the double mutant *adhE** due to its use in other ethanol utilisation studies. We aimed to optimise the constitutive expression of *adhE** for the purpose of ethanol utilisation in hydrocarbon production experiments. Three constructs with constitutive promoters of various strengths (J23100, J23107 and J23116) driving expression of *adhE** were constructed in a p15A backbone. The ethanol tolerance of *E. coli* was simultaneously assessed by supplementing *adhE**-expressing cells with 1 mM glucose and 25-500 mM ethanol (~0.05 % to 1.1 % v/v), where we found that 250- and 500 mM ethanol supplementation resulted in lower final OD_{600} values than when using 100 mM EtOH (Figure 4.6G-I). In terms of *adhE** expression strength, lower strength promoters (J23107 or J23116) worked similarly well to high-strength J23100 at higher ethanol concentrations, although with 25 mM ethanol the J23100 *adhE** exhibited a

higher final OD_{600} compared to the other constructs (Figure 4.6F). In general, we did not detect any deleterious effects of strong *adhE** expression, such as longer lag phases, suggesting that the burden from these constructs was not limiting to growth. In our previous study characterizing the MoClo-compatible plasmid backbones (see Chapter 3), we indeed observed that the DVK_p15A backbone displayed a relatively low burden even when carrying strongly expressed constructs. We therefore decided to use the J23100_B0034m_*adhE** construct (pL1_JRN094_FG) for experiments involving ethanol utilisation.

4.4.5 Hydrocarbon biosynthesis using acetate and formate or ethanol

Having characterised the growth of *E. coli* on acetate/formate and ethanol, we sought to use these as substrate for hydrocarbon synthesis. We incorporated earlier findings from the optimization of hydrocarbon biosynthesis using glucose, such as the use of a Δ*tesC* mutant, the addition of 250 mM KCl, and the use of 1 mM IPTG to induce expression of AAR from a plasmid with a pBR322 origin of replication.

The use of acetate and the combination of acetate, formate, and formate dehydrogenase (FDH) expression was first assessed in BW25113 Δ*tesC* (Figure 4.7B). The construct for AAR expression was the same as used before (pL1_JRN133_AE), while several strains now also carried a constitutively expressed FDH (pL1_JRN067_FG) (Figure 4.7A). The use of 75 mM acetate with expression of AAR using 1 mM IPTG yielded around 5.5 mg/L hydrocarbons (Figure 4.7C). The additional presence of an FDH resulted in a significant improvement in titre (~10 mg/L), even though no formate was added in this condition. In fact, the supplementation of 50 mM formate in addition to expression of an FDH led to a significant decrease in hydrocarbon titre compared to the strain expressing FDH but lacking formate. In all cases, C_{16} alcohol was produced the most, followed by lower levels of C_{18} alcohol, $C_{18:1}$ aldehyde and C_{18} aldehyde.

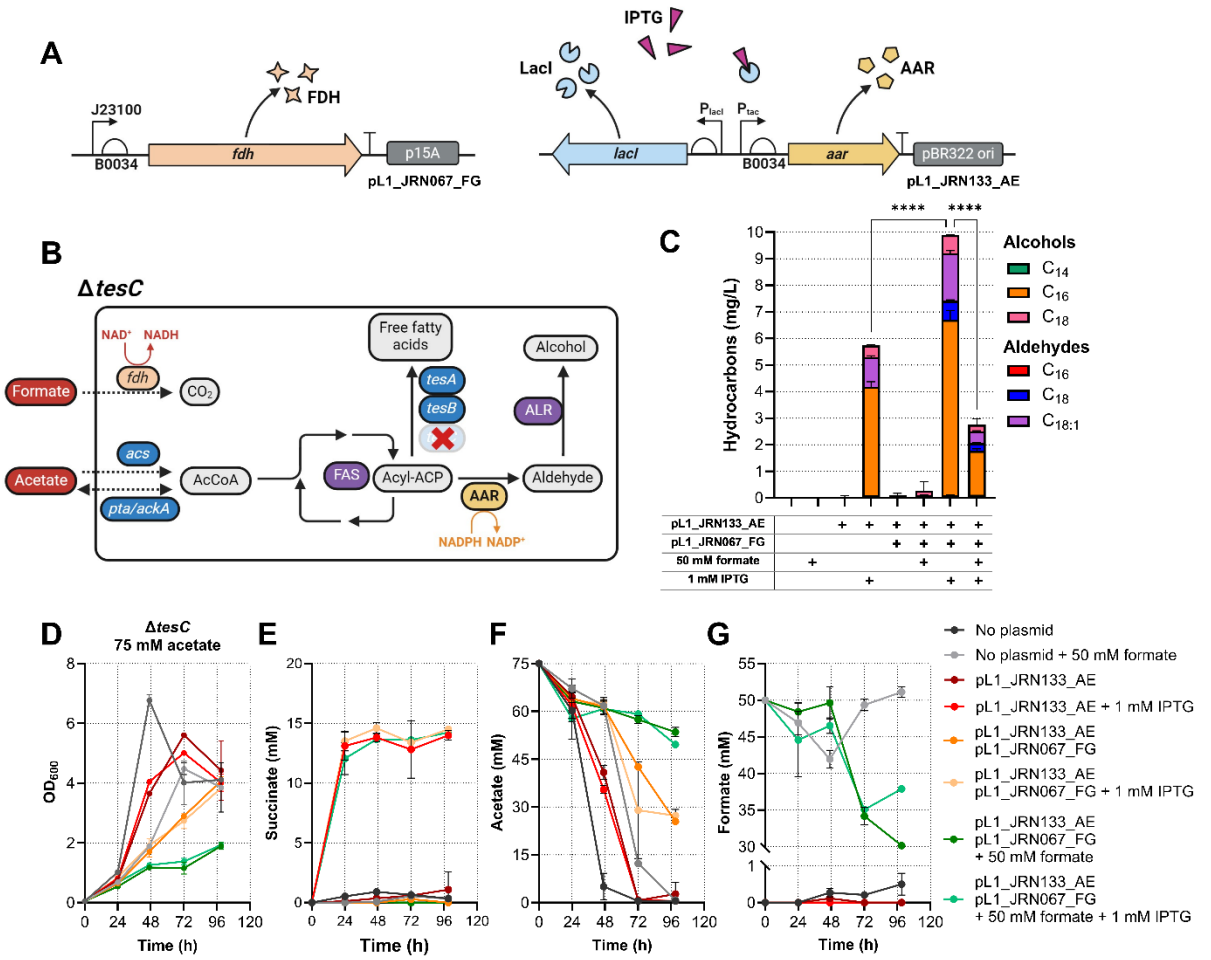


Figure 4.7. Hydrocarbon biosynthesis using acetate and formate. (A) Schematic of FDH expression construct pL1_JRN067_FG (left) and AAR-expression construct pL1_JRN133_AE (right) used in this work. **(B)** Schematic of *E. coli* Δ*tesC* metabolism indicating the routes for acetate utilisation (via *acs* or *ackA/pta*), the decarboxylation of formate to yield NADH via FDH activity, and the biosynthesis of aldehydes via AAR. **(C)** Alcohol and aldehyde biosynthesis in *E. coli* Δ*tesC* as a result of AAR induction, FDH expression and the addition of 50 mM formate. **(D)** Growth curves of *E. coli* Δ*tesC* grown on 75 mM acetate, carrying plasmids for AAR expression, FDH expression, or both, with or without 1 mM IPTG addition and with or without 50 mM formate addition (see legend on the right, relevant for graphs **D-G**). Using HPLC the concentrations of **(E)** succinate, **(F)** acetate, and **(G)** formate were determined over the course of the production experiment. The data represent averages of biological triplicate measurements, and errors bars represent their standard deviations. Statistical analysis (not explicitly visualised for each result) was performed by comparing means using a one-way ANOVA and post-hoc Tukey's multiple comparison test. P-value significance indicated are as follows: ns (p-value >0.05); * (p-value <0.05), ** (p-value <0.01), *** (p-value <0.001) or **** (p-value <0.0001).

The growth behaviour of the tested strains differed significantly as a result of their medium conditions and plasmid cargo (Figure 4.7D). Supplementing *E. coli* BW25113 Δ*tesC* with 50 mM formate alongside 75 mM acetate resulted in significantly poorer growth within the first

48h compared to the condition lacking formate, an effect we observed in our earlier characterization of growth on acetate/formate mixtures (Figure 4.5). The growth of strains carrying AAR with a pBR322 origin of replication (pL1_JRN133_AE) resulted in a slight decrease in growth performance compared to the non-plasmid control, although the induction of AAR with 1 mM IPTG did not seem to affect growth significantly. The additional presence of FDH (pL1_JRN067_FG) did result in considerable decreases in growth rate however, which is reflected in the acetate consumption profile (Figure 4.7F). Finally, the combined presence of AAR, FDH and 50 mM formate resulted in a severe growth deficiency. These cultures reached OD_{600} values of ~2 and consumed ~20-25 mM acetate over the course of 100h of growth, compared to the no-plasmid no-formate cultures which consumed all 75 mM acetate within 72h (Figure 4.7D and F). The expression of FDH in formate-supplemented strains did result in the consumption of 10-20 mM formate over the course of the growth assay, while the non-FDH expressing strain still retained 50 mM formate at the end of the production experiment (Figure 4.7G). We note that BW25113 WT (no plasmid) was found to produce ~0.5 mM formate after 100h (Figure 4.7G). Although this was not observed for BW25113 pL1_JRN133_AE, we cannot exclude that strains may have produced trace levels of formate, potentially acting as a source of reducing power for strains expressing FDH. Another interesting observation resulting from HPLC analysis of the cultures revolved around succinate biosynthesis (Figure 4.7E). These measurements revealed that 10-15 mM of succinate was present in the medium over the course of the production experiment, but only in cultures induced with 1 mM IPTG. In our previous work on ethanol utilisation we also performed HPLC analysis on IPTG-induced cultures but did not detect high levels of succinate in the medium (Figure 4.6D, E), suggesting that succinate formation was occurring due to the cultivation conditions used in this particular experiment.

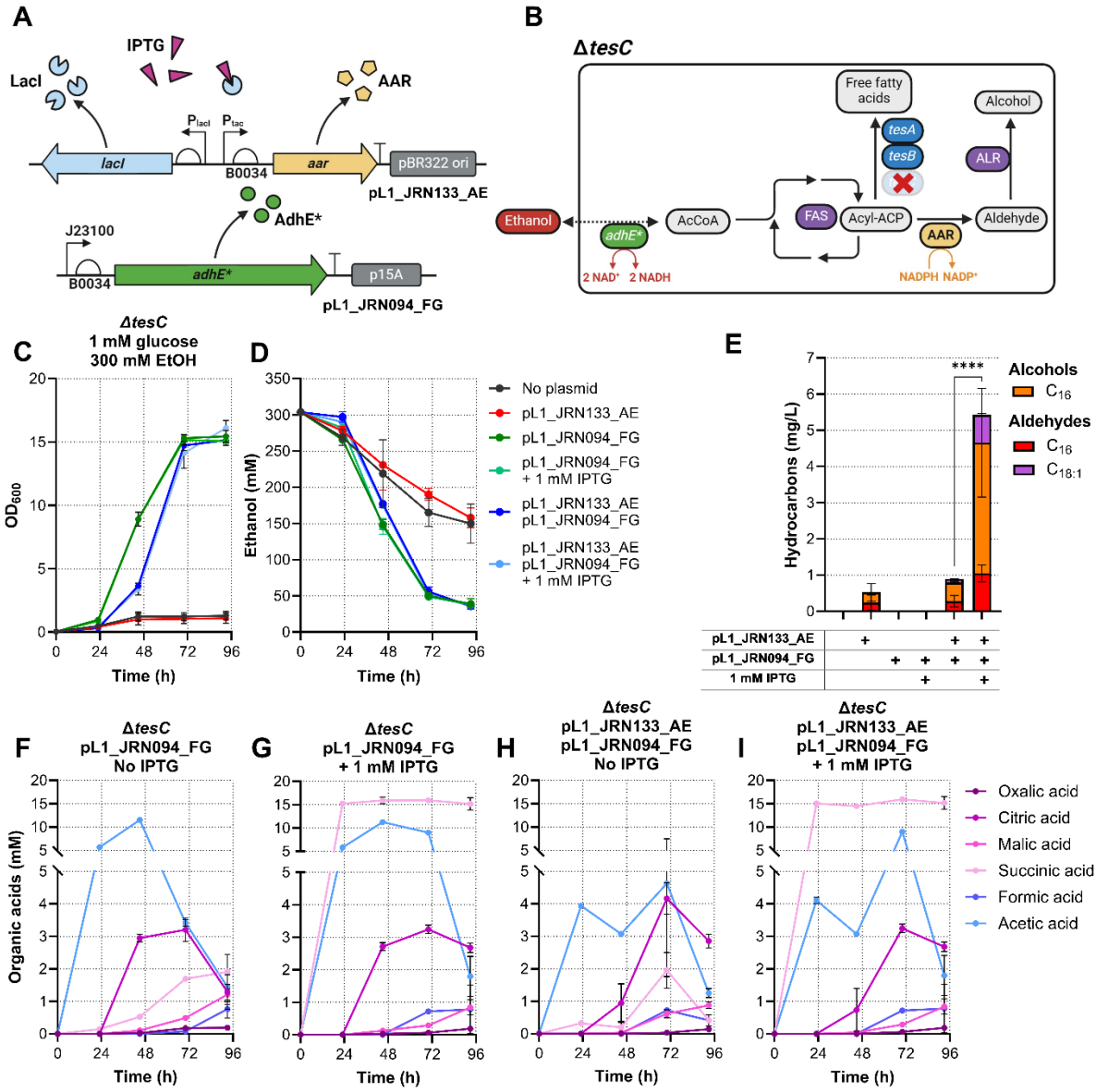


Figure 4.8. Hydrocarbon biosynthesis using ethanol as a carbon source. (A) Schematic of AAR expression construct pL1_JRN133_AE (top) and *adhE** expression construct pL1_JRN094_FG (bottom) used in this work. **(B)** Schematic of *E. coli* Δ*tesC* metabolism indicating ethanol assimilation via *adhE** and aldehyde biosynthesis via AAR. **(C)** Various *E. coli* Δ*tesC* strains carrying pL1_JRN133_AE (induced or not), pL1_JRN094_FG, or a combination. Growth took place using 1 mM glucose and 300 mM EtOH, and **(D)** their ethanol consumption profiles were measured. **(E)** Alcohol and aldehyde biosynthesis upon AAR induction in *E. coli* Δ*tesC* carrying both AdhE*-* and AAR- plasmids grown using 1 mM glucose and 300 mM ethanol. The profiles of organic acids were measured over the course of the production experiment by HPLC. Data presented originated from strains expressing AdhE* and **(F)** without AAR and without IPTG, **(G)** without AAR but supplemented with 1 mM IPTG, **(H)** with AAR but without IPTG and **(I)** with AAR and with 1 mM IPTG supplementation. The data represents averages of biological triplicate measurements, and errors bars represent their standard deviations. Statistical analysis (not explicitly visualised for each result) was performed by comparing means using a one-way ANOVA and post-hoc Tukey's multiple comparison test. P-value significance indicated are as follows: ns (p-value >0.05); * (p-value <0.05), ** (p-value <0.01), *** (p-value <0.001) or **** (p-value <0.0001).

Next the utilisation of ethanol for hydrocarbon synthesis was assessed using the same BW25113 Δ*tesC* strain but carrying pL1_JRN094_FG, responsible for constitutive expression of *adhE**, along with pL1_JRN133_AE for inducible AAR expression (Figure 4.8A and B). An ethanol concentration of 300 mM was chosen as this represented the same number of carbon units in the substrate as the 100 mM glucose condition used in previous experiments (600 carbon units). The growth of BW25113 Δ*tesC* was dependent on the presence of *adhE**, as expected, and the additional presence of the AAR plasmid seemed to slow growth in the first 48h, presumably due to the effects of plasmid burden and antibiotic addition (Figure 4.8C).

At the end of the growth assay, almost all the ethanol (>85%) was consumed in *adhE**-expressing cultures, but we also measured significant evaporation of ethanol in the no plasmid controls resulting in a loss of around half of the starting EtOH over 93 hours (Figure 4.8D). Nevertheless, hydrocarbon biosynthesis using ethanol was achieved by expression of AAR, reaching a titre of around 5.5 mg/L again consisting mostly of C_{16} alcohol and smaller amounts of C_{18}- and $C_{18:1}$ aldehydes (Figure 4.8E). To profile any differences in metabolite biosynthesis between the strains, we performed HPLC analysis on all culture broths over the course of the production experiment. Similar to the acetate-grown cultures, we observed significant succinate biosynthesis (~15 mM) in IPTG-supplemented cultures, which we also observed in strains lacking the AAR plasmid (Figure 4.8G and I). This suggested that the addition of IPTG itself was leading to succinate biosynthesis, an observation we currently do not have an explanation for. Apart from succinate, acetate production occurred in all assessed conditions, reaching maximum concentrations of around 10 mM except in strains expressing *adhE** and carrying an uninduced AAR plasmid where concentrations only reached ~4 mM (Figure 4.8F-I). Another surprising finding was that formate was being produced in all conditions, albeit at low levels (~0.5 mM after 93h). Formate is thought to be a strictly anaerobic fermentation product since its formation requires both transcription of the oxygen-regulated *pflAB* operon and activity of the oxygen sensitive pyruvate-formate lyase structural gene *pflB*. The AdhE protein, which is being constitutively expressed in this study, is also thought to act as a

pyruvate-formate lyase-deactivase further raising the question of how formate biosynthesis is taking place (Kessler et al., 1991). Regardless, the results presented above demonstrate that both acetate and ethanol can be used as carbon sources for AAR-mediated hydrocarbon biosynthesis.

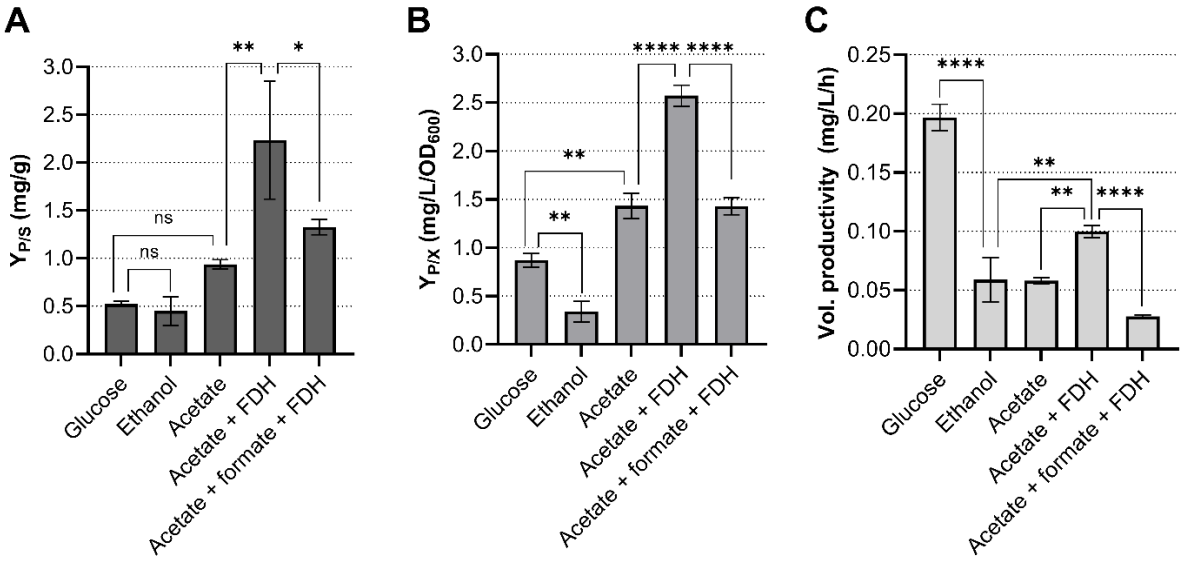


Figure 4.9. **Comparison of bioproduction metrics for AAR-mediated hydrocarbon synthesis. (A)** Yield of all hydrocarbon products per unit substrate (mg/g). **(B)** Concentration of all hydrocarbon products normalised for final biomass formation (mg/L/OD_{600}). **(C)** Volumetric productivity (mg/L/h) of hydrocarbon product biosynthesis per carbon source used. Carbon sources and concentrations: glucose (100 mM), ethanol (1 mM glucose and 300 mM ethanol), acetate (75 mM), acetate (75 mM) and expression of FDH, acetate (75 mM) with formate (50 mM) and expression of FDH. The data represent averages of biological duplicate (glucose) or triplicate (other conditions) measurements, and errors bars represent their standard deviations. *S*tatistical analysis (not explicitly visualised for each result) was performed by comparing means using a one-way ANOVA and post-hoc Tukey's multiple comparison test. P-value significance indicated are as follows: ns (p-value >0.05); * (p-value <0.05), ** (p-value <0.01), *** (p-value <0.001) or **** (p-value <0.0001).

To summarise, the hydrocarbon biosynthesis performance of the various approaches was compared using several standardised bioproduction metrics. These included the mass yield of product per substrate, the concentration of product normalised for biomass formation, and the volumetric productivity. One of the most important metrics in terms of bioprocess economics is product yield per substrate, and the results from this work indicated that both ethanol and acetate performed equally to glucose in this regard (Figure 4.9A). In fact, acetate-

supplied cultures expressing FDH resulted in a ~4-fold higher yield and a ~2.5-fold higher hydrocarbon concentration per biomass than the glucose condition (Figure 4.9A and B). Glucose use did clearly lead to a higher volumetric productivity however, which can be attributed to the high growth rate of *E. coli* on glucose and the concomitant short production times (Figure 4.9C).

4.4.6 Alkane biosynthesis using ethanol as a carbon source

In terms of generating fatty-acid derived products, ethanol is an attractive candidate as it provides reducing power and enters metabolism as acetyl-CoA, the starting point of fatty acid biosynthesis. The results presented earlier indicated that ethanol could be used as a C-source for hydrocarbon synthesis. To take this one step further towards alkane biosynthesis, we transformed BW25113 Δ*tesC* Δ*yqhC-yqhD-dkgA* with pL2_JRN072_AF encoding the IPTG-inducible AAR-ADO pathway, and pL1_JRN139_FG encoding a constitutively expressed *adhE** (Figure 4.10A and B). This latter plasmid was identical to previously used pL1_JRN094_FG, except it carried cm^R instead of kan^R. Since the consumption of ethanol via AdhE* was essential for growth, we did not add antibiotics to select for the pL1_JRN139_FG.

First, the growth of strains carrying both AdhE* and AAR-ADO expression constructs seemed to benefit from inducing the alkane biosynthesis pathway (Figure 4.10C). All strains induced with 1 mM IPTG had amassed significantly higher levels of biomass at 72 hours compared to non-induced AAR-ADO carrying strains, verified by the ethanol consumption profiles (Figure 4.10D). Ethanol-grown strains expressing AAR-ADO were able to accumulate ~2.3 mg/L alkanes, consisting mainly of $C_{17:1}$ and C_{15} hydrocarbons (Figure 4.10E). As was the case with glucose-grown cells expressing AAR-ADO, we observed almost no accumulation of alcohols and aldehydes. The separation of pellet and supernatant fractions again indicated that alkanes accumulated mainly extracellularly (Figure 4.10F). Besides this, the sum of hydrocarbon titres was higher when pellet and supernatant when separated, presumably due to a more efficient extraction.

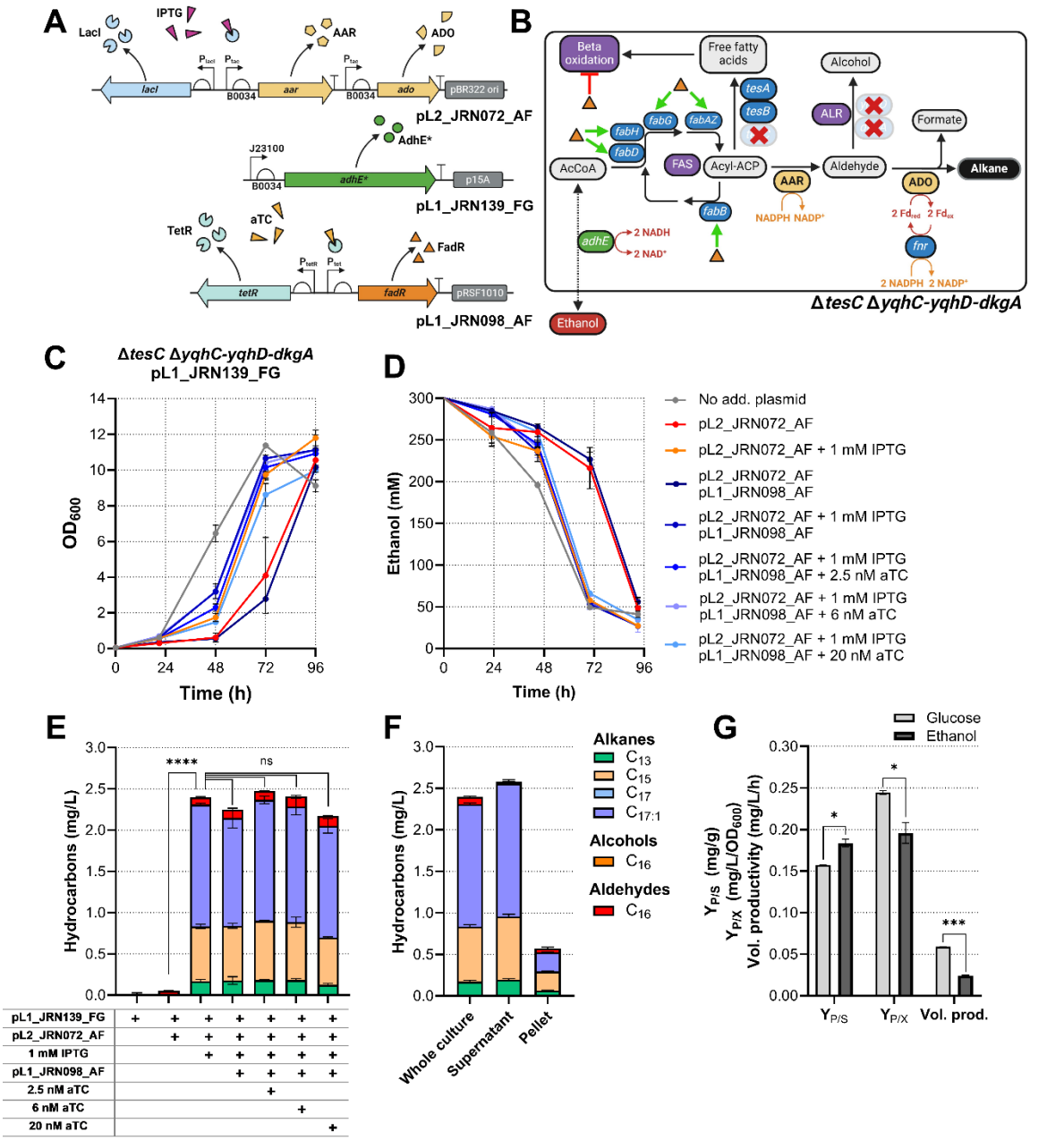


Figure 4.10. Alkane biosynthesis using ethanol as a carbon source. (A) Schematic representation of pL1_JRN072_AF (top), pL1_JRN139_FG (middle) and pL1_JRN098_FG (bottom) plasmids. **(B)** Schematic of *E. coli* Δ*tesC* Δ*yqhC-yqhD-dkgA* metabolism indicating ethanol assimilation via AdhE*, alkane biosynthesis via AAR-ADO and induction of fatty acid synthesis genes by FadR. **(C)** Growth data of *E. coli* Δ*tesC* Δ*yqhC-yqhD-dkgA* grown using 1 mM glucose and 300 mM EtOH. Strains all carried pL1_JRN139_FG, and some carried pL1_JRN072_AF and pL1_JRN098_FG induced at various concentrations as indicated. **(D)** Ethanol consumption profiles of the same strains. **(E)** Hydrocarbon production profiles of ethanol-grown *E. coli* Δ*tesC* Δ*yqhC-yqhD-dkgA* dependent on the induction of the AAR-ADO pathway, and on induction of *fadR*. **(F)** Hydrocarbon presence in the whole culture, pellet, and supernatant fractions of *E. coli* Δ*tesC* Δ*yqhC-yqhD-dkgA* pL1_JRN139_FG pL1_JRN133_AE grown using 300 mM ethanol and induced with 1 mM IPTG. **(G)** Bioproduction metrics comparing glucose and ethanol for alkane biosynthesis. Statistical analysis was performed by comparing means using a one-way ANOVA and post-hoc Tukey's multiple comparison test (panel E). A two-tailed equal variance t-test was used for data in panel G. P-value significance indicated are as follows: ns (p-value >0.05); * (p-value <0.05), ** (p-value <0.01), *** (p-value <0.001) or **** (p-value <0.0001).

Although we did not observe a significant effect of *fadR* overexpression on hydrocarbon titres when grown on glucose, we attempted this again in the context of alkane biosynthesis using ethanol. To do so, we transformed the strains already carrying pL1_JRN133_AE and pL1_JRN139_FG with pL1_JRN098_AF, carrying an aTC-inducible *tetR*/P_{tet}-*fadR* construct (Figure 4.10A). The final hydrocarbon titres were not significantly improved by overexpression of *fadR* (Figure 4.10E).

Finally, we compared glucose and ethanol as carbon sources for alkane biosynthesis using several bioproduction metrics (Figure 4.10G). The alkane yield per substrate (mg/g) were around 15% higher when using ethanol, indicating the viability of this in terms of substrate conversion to product. The alkane yield per biomass (mg/L/OD_{600}) and volumetric productivity (mg/L/h) were significantly lower compared to glucose however, owing to the high biomass accumulation and high growth rates achieved with glucose as a carbon source.

4.4.7 Growth-decoupling using C_2 carbon sources by regulatory control

A challenge for achieving high-yield bioproduction is the tendency for cells to divert carbon and energy towards growth instead of product synthesis. Decoupling growth from product synthesis is therefore required to maximise product yield. Since acetyl-CoA is both an essential building block for biomass formation and the starting metabolite for alkane biosynthesis, exerting control over its fate in cell metabolism can offer a way to decouple growth from alkane biosynthesis.

Growth of *E. coli* on acetate or ethanol as sole carbon sources requires expression of the glyoxylate shunt (or glyoxylate bypass), catalysed by *aceBAK.* This operon consists of isocitrate lyase (*aceA*), malate synthase (*aceB*) and isocitrate dehydrogenase kinase (*aceK*). Flux through the TCA cycle without glyoxylate shunt activity would result in complete oxidation of the C_2 substrates by the two sequential decarboxylation reactions catalysed by isocitrate dehydrogenase (*icd*) and the α-ketoglutarate dehydrogenase complex (*lpd, sucA, sucB*). Therefore, flux from acetyl-CoA towards biomass formation is fully reliant on functionality of

the glyoxylate shunt when *E. coli* is grown on carbon sources entering metabolism as acetyl-CoA, such as fatty acids, acetate, or ethanol (Figure 4.11A).

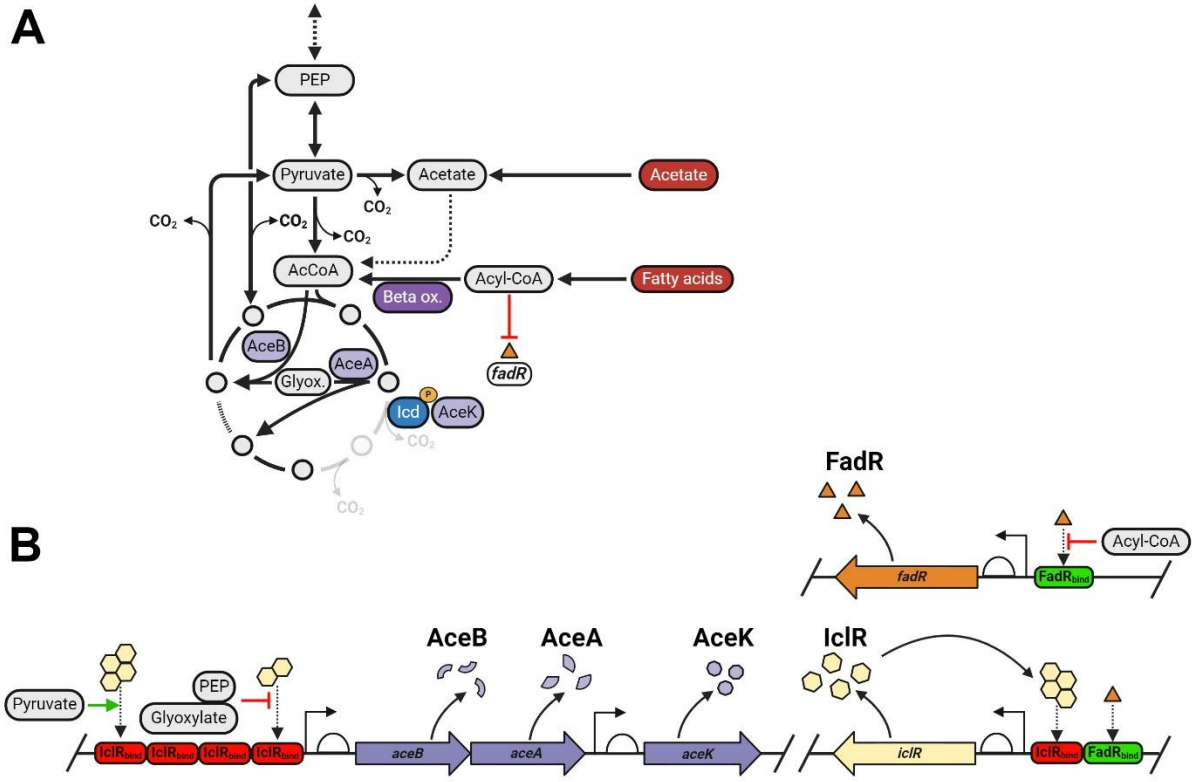


Figure 4.11. **The glyoxylate shunt and the genomic organisation of the *aceBAK* operon**. **(A)** Schematic representing *E. coli* central metabolism, with indicated glyoxylate shunt enzymes AceA and AceB, and the Icd-phosphorylating enzyme AceK. The assimilation of acetate and fatty acids requires glyoxylate shunt activity to avoid the two sequential decarboxylation steps in the lower TCA cycle. **(B)** The genomic *aceBAK* operon and several of its regulatory features. The expression of *aceBAK* is negatively transcriptionally regulated by IclR, indicated by the red $IclR_{bind}$ binding regions. The active DNA-binding tetrameric form of IclR is stabilised by pyruvate, and formation of the dimeric inactive form is promoted by glyoxylate or PEP. IclR transcription is also self-regulated, as indicated by the $IclR_{bind}$ region upstream of its own promoter. The expression of FadR, a positive transcriptional regulator of IclR, is also self-regulated. FadR interaction with fatty acyl-CoA molecules disrupts its DNA binding activity, resulting in transcriptional inactivation.

The *aceBAK* operon is subject to complex multi-level regulation (Figure 4.11B). The direct regulator for the operon is the isocitrate lyase regulator (IclR), which itself is subject to allosteric effects by phosphoenolpyruvate (PEP), pyruvate, and glyoxylate (Cortay et al., 1991;

Lorca et al., 2007). Pyruvate stabilises the active tetrameric state of the IclR protein, which enhances binding affinity of IclR to the upstream promoter region of *aceBAK*, resulting in repression of the operon. In contrast, PEP and glyoxylate stabilise the inactive dimeric state of IclR, reducing DNA binding of IclR to the *aceBAK* promoter region thereby de-repressing it. These interactions reflect the physiological role of the glyoxylate shunt in mitigating carbon loss of substrates entering metabolism directly as acetyl-CoA. Besides allosteric interactions on the protein level, *iclR* is also subject to transcriptional self-regulation. The *iclR* promoter region includes an IclR binding site that represses transcription when bound (Gui, Sunnarborg, Pan, et al., 1996).

Another regulatory protein indirectly involved in *aceBAK* operon expression is the fatty acid degradation regulator *fadR*, which acts as a positive transcriptional regulator of *iclR* (Gui, Sunnarborg, & Laporte, 1996). The activation of exogenous fatty acids to fatty acyl-CoA molecules results in the repression of *fadR*, which in turn de-represses fatty acid catabolism genes enabling beta oxidation to occur. Since acetyl-CoA is the terminal product of fatty acid degradation, the glyoxylate shunt is essential for carbon retention in fatty acid-grown *E. coli*. Since *fadR* is a transcriptional activator of *iclR*, the fatty acyl-CoA induced repression of *fadR* also reduces expression of *iclR*, thereby enhancing glyoxylate shunt activity.

As we have demonstrated in earlier results, the overexpression of *fadR* can lead to significantly improved titres of hydrocarbons (only when flux through the terminal enzyme has not been maximised). Besides its direct effect on product synthesis, we were interested in exploring its involvement with *iclR* and the *aceBAK* operon. In this work we have explored the use of acetate and ethanol which are both C_2 carbon sources entering metabolism as acetyl-CoA, hence their catabolism is entirely dependent on glyoxylate shunt activity. Controlling the expression of both *iclR* and *fadR* was therefore explored as a potential method to control flux between acetyl-CoA, hydrocarbon product synthesis and biomass formation.

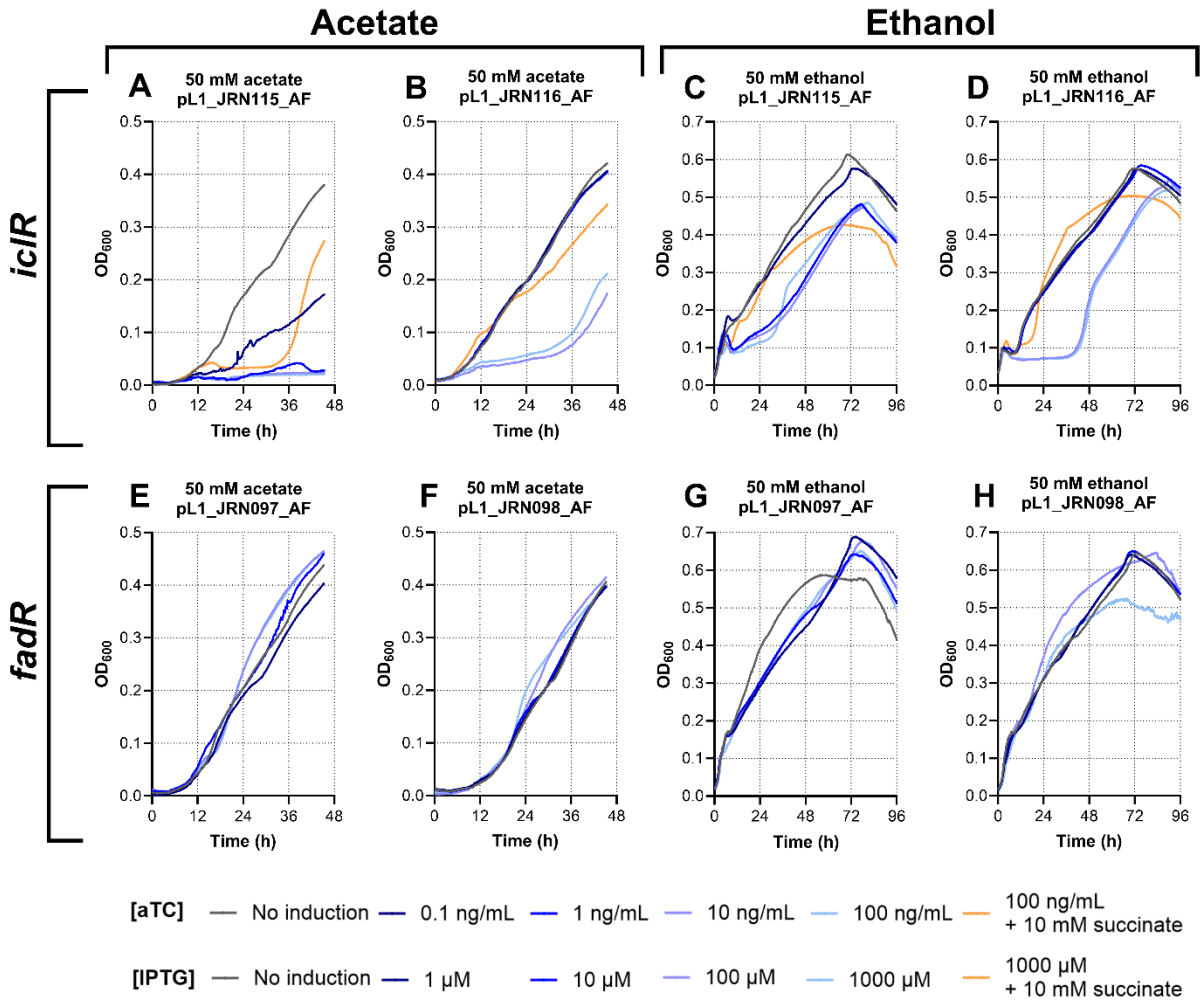


Figure 4.12. Growth inhibition on C_2 carbon sources by regulatory control of the glyoxylate shunt. The *iclR* (**A-D**) or *fadR* (**E-H**) transcriptional regulators were expressed at different induction levels to assess their effects on growth using 50 mM acetate (**A, B, E, F**) or 1 mM glucose and 50 mM ethanol (**C, D, G, H**). The coding sequences for the regulators were cloned under control of inducible *lacI*/P_{tac} (**A, C, E, G**) or *tetR*/P_{tet} (**B, D, F, H**) promoters. The *lacI*/P_{tac} constructs were induced using IPTG, and the *tetR*/P_{tet} constructs with aTC, at the concentrations indicated. The following plasmids were used: *lacI*/P_{tac}-*fadR* (pL1_JRN097_AF), *tetR*/P_{tet}-*fadR* (pL1_JRN098_AF), *lacI*/P_{tac}-*iclR* (pL1_JRN115_AF) and *tetR*/P_{tet}-*iclR* (pL1_JRN116_AF). All datapoints represent averages of biological triplicate measurements, and errors bars represent the standard deviation thereof.

In addition to the existing inducible *fadR* plasmids used previously, *iclR* was cloned under control of the *lacI*/P_{tac} and *tetR*/P_{tet} inducible promoter systems in low copy pRSF1010 plasmids (forming pL1_JRN115_AF and pL1_JRN116_AF, respectively). All *fadR* and *iclR* constructs were assessed for their ability to repress *E. coli* growth on acetate or ethanol as sole carbon sources at various induction levels (Figure 4.12). Even low induction levels of *iclR* (1 µM IPTG) resulted in an approximate halving of growth rate on acetate as a sole carbon source compared to the uninduced state (Figure 4.12A). Higher IPTG levels resulted in

complete cessation of growth, although growth could be rescued by addition of 10 mM succinate, which enters the TCA cycle at the C_4 branch and therefore circumvents the glyoxylate shunt. We observed a similar pattern using the *tetR*/P_{tet} system at 10- and 100 ng/mL induction, although surprisingly the culture was able to re-initiate growth after 36 hours (Figure 4.12B). Similarly to the *lacI*/P_{tac} system, succinate supplementation rescued growth in the induced condition. Strains supplemented with ethanol were grown for 96h instead of 48h, and we observed relatively leaky growth of strains carrying *lacI*/P_{tac}-*iclR* constructs grown using ethanol compared to acetate (Figure 4.12C). In contrast, the strong expression of *iclR* from an aTC-inducible construct (10- or 100 ng/mL) led to a complete growth cessation for around 36 hours, indicating tight glyoxylate shunt repression (Figure 4.12D). Interestingly, these strains also recovered and were able to achieve maximum growth rates equal to non-induced strains. Another notable observation was that the time until growth resumed normally was not dependent on the supplemented aTC concentration (Figure 4.12B and D). This indicated that declining aTC concentrations over time was not the cause of the *aceBAK* de-repression, suggesting that another mechanism was responsible for restoring growth.

When the same systems were used to express *fadR*, we observed no deleterious effect on growth (Figure 4.12E-H). It was clear that *iclR* activation by *fadR* overexpression was not pronounced enough to cause noticeable glyoxylate shunt repression. It is known that *iclR* is self-regulating: presence of the protein negatively regulates its own expression (Gui, Sunnarborg, Pan, et al., 1996). However, this is dependent on a putative binding site for *iclR* located in its own promoter region. Since the synthetic constructs used to express *iclR* in this study lacked the native self-regulatory binding sites for *iclR, its* transcriptional self-regulation was avoided. It is likely that with the single-copy native *iclR* gene on the chromosome even strong *fadR* overexpression cannot overcome the self-regulating properties of *iclR.*

4.5 Discussion and conclusion

The biosynthesis of hydrocarbons represents an opportunity for producing sustainable fuels and chemicals. As a result, this field has garnered much attention over the years as these technologies move towards commercial viability. Even so, several significant barriers stand in the way of economically producing microbial hydrocarbons, including the low efficiencies of the involved pathways, the instability of microbial production processes, and the use of costly and unsustainable carbon sources.

In this study, we sought to identify optimal hydrocarbon production conditions via the well-studied AAR-ADO biosynthetic pathway. By first studying only the activity of AAR, we could separate the factors affecting this prior enzyme in the pathway from the factors affecting total pathway activity. In addition, we aimed to explore the use of acetate, formate, and ethanol as carbon- and energy sources for hydrocarbon biosynthesis in *E. coli*, which has not been demonstrated before.

Firstly, the addition of potassium chloride was found to improve production via AAR, as supplementation with 250 mM KCl resulted in a 2-fold increase in total hydrocarbon titre in some conditions. Previous studies have reported 10- to 20- fold improvements in AAR activity as a result of KCl supplementation *in vitro* (Lin et al., 2013; Warui et al., 2015). Other *in silico* studies have also found that KCl was expected to stabilise AAR (Sharma et al., 2021). We therefore conclude that the positive effect of KCl supplementation on hydrocarbon titres was due to stabilization of AAR, although the *in vivo* effect was apparently less pronounced than when using purified AAR protein. Even so, it was apparent that AAR was still inefficient despite KCl supplementation; hydrocarbon titres beyond 10 mg/L could not be reached despite cloning AAR in a multi-copy plasmid containing a strong inducible promoter, strong RBS, inducing its expression at high levels, and expressing transcriptional regulator *fadR* to increase fatty acid biosynthesis flux. We note that several studies have focussed on engineering AAR and did indeed find that enzyme performance is suboptimal, marking it as an important target for

engineering alkane biosynthesis (Han et al., 2023; Kudo et al., 2019; Lin et al., 2013; Sharma et al., 2021).

The deletion of thioesterase C (*tesC*/*fadM*/*ybaW*) was also found to be a successful strategy for improving hydrocarbon synthesis. This approach was based on previous work demonstrating improved fatty alcohol biosynthesis by thioesterase deletions (Liu et al., 2016). The deletion of *tesC* resulted in a ~30% increase in hydrocarbon titre, presumable as higher levels of acyl-ACP remained available for their reduction to aldehydes. This increase was similar to what Liu and coworkers observed.

The co-expression of fatty acid regulator *fadR* was only beneficial when lower AAR induction levels were used, suggesting that AAR activity was not saturated at lower induction levels but that a maximum flux was reached when it was induced strongly. Possibly, the Δ*tesC* mutation and low AAR induction already induced FAS to some extent due to the lower availability of free fatty acids. With maximised AAR expression, FAS may already be highly expressed and additional *fadR* expression would have no effect. A previous study had attempted the same *fadR* overexpression approach and reported an approximate doubling of alkane/alkene titre (Song et al., 2015). The authors used extremely low induction levels of AAR-ADO (20 µM IPTG) to achieve this 2-fold improvement, which was similar to the fold improvements we measured when using lower (0.1 mM IPTG) induction conditions for the expression of only AAR.

When AAR and ADO were expressed simultaneously, alkane titres reached around 3 mg/L in contrast to the ~10 mg/L of aldehydes and alcohols that were produced by expression of AAR alone. While this confirmed that ADO is indeed limiting to alkane biosynthesis via the AAR-ADO pathway, the ~30% efficiency of ADO in converting AAR product to alkanes suggested that both proteins were at least somewhat matched in terms of catalytic efficiency in these conditions. This might not be surprising considering the enzymes catalyse succeeding steps in the same enzymatic pathway and originate from the same organism, although their expression is probably independently regulated natively (Klähn et al., 2014). Even so, studies

on the AAR-ADO pathway do not often measure activities of solely AAR, which can offer insights regarding metabolic bottlenecks. Interestingly, in attempts to co-localise AAR and ADO using zinc finger scaffolds, Rahman and coworkers found that a ratio of 3:1 ADO:AAR was optimal for maximising alkane biosynthesis (Rahman et al., 2014). This suggested that the 30% efficiency we measured was in line with expected catalytic efficiencies per enzyme. We also note that the supplementation of KCl may have affected AAR-ADO interaction, since previous *in vitro* worked showed that AAR-ADO interactions were completely abolished by addition of 250 mM NaCl (Chang et al., 2020). Then again, KCl was found to significantly increase AAR activity *in vitro* while NaCl did not, which points to potassium playing a different role to sodium in terms of AAR's catalytic activity (Lin et al., 2013). Comparative production experiments using AAR-ADO with and without KCl supplementation could provide more insight about its effects.

Using ethanol as a carbon source required the expression of *adhE*, the *E. coli* bifunctional aldehyde/alcohol dehydrogenase. This enzyme -and other *adhE* homologs- are often described as oxygen sensitive, and work from over two decades ago had led to the discovery of oxygen tolerant *adhE** variants carrying E568K and A267T mutations (Holland-Staley et al., 2000; Membrillo-Hernández et al., 2000). In contrast, we found that overexpression of wildtype *adhE* was sufficient to enable growth on ethanol at rates equal to *adhE**. We did observe more acetate formation in *adhE**-expressing cultures compared to WT *adhE*. This could be explained by a higher acetyl-CoA flux being counteracted with a higher acetate secretion flux, although growth was indistinguishable between strains expressing either variant. Another interesting finding was the detection of formate in ethanol-grown aerobic shake flask cultures, regardless of whether *adhE* or *adhE** was used. Formate is mainly produced through *pflB* activity in *E. coli*, an enzyme generally understood to be fermentative and oxygen sensitive, meaning that either our specific growth conditions imparted a higher O_2 tolerance, or that despite agitation in aerobic shake flasks the conditions were still sufficiently micro- or anaerobic. A previous study on the transcriptome of *E. coli* reported an upregulation of *pflB*

upon sudden ethanol exposure, possibly explaining why an enzyme such as *pflB* might be expressed in ethanol-grown *E. coli* (Cao et al., 2017).

We also considered whether the expression levels of *adhE* in this work could be responsible, but in the work by Holland-Stanley *et al.* a very similar plasmid (p15A origin, lacUV5 promoter) was used (Holland-Staley et al., 2000). In contrast, Membrillo-Hernández *et al*. used a pBR322-derived plasmid with an IPTG-inducible promoter to express *adhE** and noted that strong induction resulted in no growth in aerobic conditions (Membrillo-Hernández et al., 2000). Although we did observe that strong IPTG induction resulted in poorer growth, we did not observe growth cessation. Indeed, in this work we settled on using a strong constitutively expressed construct for bioproduction. One important distinction between these seminal works and our assays was the supplementation with low amounts of glucose (1 mM). We noticed that direct transfer of rich medium grown cells to ethanol minimal medium did not result in growth, and low glucose supplementation allowed O/N growth on ethanol, although we observed this effect for populations expressing either *adhE* as well as *adhE**. Direct measurement of dissolved oxygen or growth in oxygen-controlled chemostats could clarify whether *adhE* (and possibly *pflB*) are active despite *E. coli* residing in aerobic conditions.

The use of acetate and ethanol for hydrocarbon biosynthesis revealed that both were suitable carbon sources for AAR-mediated production compared to glucose. The use of ethanol resulted in equal $Y_{P/S}$, but lower $Y_{P/X}$ and lower productivity. In contrast, acetate outperformed glucose in both $Y_{P/S}$ and $Y_{P/X}$ although productivity was still lower. Other notable results were the improvements as a result of expressing FDH in addition to using acetate (4-fold increased $Y_{P/S}$ and 2.5-fold increased $Y_{P/X}$ compared to glucose). Although we cannot foresee a role of FDH in this condition, since no formate was supplied and barely any was produced in non-FDH carrying strains, we did observe significant reductions in growth rate as a result of FDH expression. Our expectation is therefore that this reduced growth rate resulted in higher product yields, possibly due to reduced rates of acetate oxidation to form biomass, or due to lower FAS flux allowing a higher proportion of acyl-ACP to be catalysed by AAR. Modulating

the growth rate of *E. coli* may therefore be a worthwhile strategy to improve product synthesis. In fact, the growth decoupling work using C_2 carbon sources and *iclR* expression may be a method to achieve a reduced growth rate. Alkane biosynthesis with ethanol as sole carbon source resulted in 2 - 2.5 mg/L titres, with higher $Y_{P/S}$ values but lower $Y_{P/X}$ and productivity values. The use of acetate for alkane biosynthesis would be a logical next step in this work. One study has used acetate to biosynthesis alkanes using *Acinetobacter baylyi,* resulting in a final titre of 74 µg/mL (Lehtinen et al., 2018).

Compared to other studies using AAR or AAR/ADO our work led to considerably lower titres of aldehydes, alcohols, and alkanes, although there may be several reasons for this. For example, the first study describing AAR/ADO activity reached alkane concentrations of around 300 mg/L, a factor 100 above our achieved titres (Schirmer et al., 2010). A first reason for this may be the carbon source quantity, Schirmer *et al.* supplemented their strains with ~166 mM glucose, while in this study 100 mM was used. This cannot explain the large gap in titres however, although differences in cultivation conditions might. Several studies used glass tubes or vials for production experiments instead of shake flasks, including initial work by Schirmer *et al.,* which may significantly change the dynamics of agitation and oxygen availability (Cao et al., 2016; Harger et al., 2013; Schirmer et al., 2010; Sheppard et al., 2016). Other studies performed production experiments in LB or supplemented their cultures with undefined medium components, such as yeast extract, which likely has a considerable impact on cell density and carbon availability (Cao et al., 2016; Choi & Lee, 2013; Sheppard et al., 2016; Song et al., 2016). Another important aspect was the timing of plasmid induction. In our work, plasmids were induced relatively soon after inoculation (at OD_{600} ~0.2) to allow full expression over the course of growth. However, other studied induced strains at later points, for example at OD_{600} 1-1.5 (Cao et al., 2016), 0.7-0.9 (Sheppard et al., 2016) and 0.6 (Song et al., 2016), to name several. Possibly, a combination of these factors could be optimised further in our work to achieve higher hydrocarbon titres.

In summary, this work identified several novel optimization strategies that led to improved hydrocarbon biosynthesis in *E. coli* via AAR, including KCl supplementation and the deletion of Δ*tesC*. Furthermore, we demonstrated for the first time that acetate and ethanol are viable carbon sources for hydrocarbon biosynthesis in *E. coli*, even outperforming glucose in some metrics. As such, this work represents a preliminary study to help enable a shift towards using more sustainable and economically viable carbon sources for biofuel production. Hopefully, these insights can aid in pushing hydrocarbon biosynthesis towards a scalable future.

Acknowledgements

We thank Reza Rohani and Yutong Yin for their insightful comments and advice. This work was supported by the BBSRC (grant BB/M011224/1) and the Oxford Interdisciplinary Bioscience Doctoral Training Partnership. Figure schematics were created using Biorender.com.

4.6 References

Bogorad, I. W., Lin, T. S., & Liao, J. C. (2013). Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature 2013 502:7473*, *502*(7473), 693–697. https://doi.org/10.1038/nature12575

Bourdenx, B., Bernard, A., Domergue, F., Pascal, S., Léger, A., Roby, D., Pervent, M., Vile, D., Haslam, R. P., Napier, J. A., Lessire, R., & Joubès, J. (2011). Overexpression of Arabidopsis ECERIFERUM1 Promotes Wax Very-Long-Chain Alkane Biosynthesis and Influences Plant Response to Biotic and Abiotic Stresses. *Plant Physiology*, *156*(1), 29–45. https://doi.org/10.1104/PP.111.172320

Cao, H., Wei, D., Yang, Y., Shang, Y., Li, G., Zhou, Y., Ma, Q., & Xu, Y. (2017). Systems-level understanding of ethanol-induced stresses and adaptation in E. coli. *Scientific Reports*, *7*. https://doi.org/10.1038/SREP44150

Cao, Y. X., Xiao, W. H., Zhang, J. L., Xie, Z. X., Ding, M. Z., & Yuan, Y. J. (2016). Heterologous biosynthesis and manipulation of alkanes in Escherichia coli. *Metabolic Engineering*, *38*, 19–28. https://doi.org/10.1016/J.YMBEN.2016.06.002

Caspeta, L., & Nielsen, J. (2013). Economic and environmental impacts of microbial biodiesel. *Nature Biotechnology 2013 31:9*, *31*(9), 789–793. https://doi.org/10.1038/nbt.2683

Chang, M., Shimba, K., Hayashi, Y., & Arai, M. (2020). Electrostatic interactions at the interface of two enzymes are essential for two-step alkane biosynthesis in cyanobacteria. *Bioscience, Biotechnology, and Biochemistry*, *84*(2), 228–237. https://doi.org/10.1080/09168451.2019.1677142

Choi, Y. J., & Lee, S. Y. (2013). Microbial production of short-chain alkanes. *Nature*, *502*(7472), 571–574. https://doi.org/10.1038/nature12536

Cortay, J.-C., Negre, D., Galinier, A., Duclos, B., Perriere, G., & Cozzone, A. J. (1991). Regulation of the acetate operon in Escherichia coli: purification and functional characterization of the IclR repressor. *The EMBO Journal*, *10*(3), 675–679. https://doi.org/10.1002/J.1460-2075.1991.TB07996.X

Cronan, J. E. (2021). The Escherichia coli FadR transcription factor: Too much of a good thing? *Molecular Microbiology*, *115*(6), 1080–1085. https://doi.org/10.1111/MMI.14663

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

Duetz, W. A. (2007). Microtiter plates as mini-bioreactors: miniaturization of fermentation methods. *Trends in Microbiology*, *15*(10), 469–475. https://doi.org/10.1016/J.TIM.2007.09.004

Fatma, Z., Hartman, H., Poolman, M. G., Fell, D. A., Srivastava, S., Shakeel, T., & Yazdani, S. S. (2018). Model-assisted metabolic engineering of Escherichia coli for long chain alkane and alcohol production. *Metabolic Engineering*, *46*, 1–12. https://doi.org/10.1016/j.ymben.2018.01.002

Gui, L., Sunnarborg, A., & Laporte, D. C. (1996). Regulated expression of a repressor protein: FadR activates iclR. *Journal of Bacteriology*, *178*(15), 4704–4709. https://doi.org/10.1128/JB.178.15.4704-4709.1996

Gui, L., Sunnarborg, A., Pan, B., & Laporte, D. C. (1996). Autoregulation of iclR, the gene encoding the repressor of the glyoxylate bypass operon. *Journal of Bacteriology*, *178*(1), 321–324. https://doi.org/10.1128/JB.178.1.321-324.1996

Han, J., Asano, K., Matsumoto, T., Yamada, R., & Ogino, H. (2023). Engineering acyl-ACP reductase with fusion tags enhances alka(e)ne synthesis in Escherichia coli. *Enzyme and Microbial Technology*, *168*, 110262. https://doi.org/10.1016/J.ENZMICTEC.2023.110262

Harger, M., Zheng, L., Moon, A., Ager, C., An, J. H., Choe, C., Lai, Y. L., Mo, B., Zong, D., Smith, M. D., Egbert, R. G., Mills, J. H., Baker, D., Pultz, I. S., & Siegel, J. B. (2013). Expanding the product profile of a microbial alkane biosynthetic pathway. *ACS Synthetic Biology*, *2*(1), 59–62. https://doi.org/10.1021/sb300061x

Holland-Staley, C. A., Lee, K., Clark, D. P., & Cunningham, P. R. (2000). Aerobic activity of Escherichia coli alcohol dehydrogenase is determined by a single amino acid. *Journal of Bacteriology*, *182*(21), 6049–6054. https://doi.org/10.1128/JB.182.21.6049-6054.2000

Holze, H., Schrader, L., & Buellesbach, J. (2020). Advances in deciphering the genetic basis of insect cuticular hydrocarbon biosynthesis and variation. *Heredity 2020 126:2*, *126*(2), 219–234. https://doi.org/10.1038/s41437-020-00380-y

International Energy Agency. (2023). *The Oil and Gas Industry in Net Zero Transitions*.

Iverson, S. V., Haddock, T. L., Beal, J., & Densmore, D. M. (2016). CIDAR MoClo: Improved MoClo Assembly Standard and New *E. coli* Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synthetic Biology*, *5*(1), 99–103. https://doi.org/10.1021/ACSSYNBIO.5B00124

Janßen, H. J., & Steinbüchel, A. (2014). Fatty acid synthesis in Escherichia coli and its applications towards the production of fatty acid based biofuels. *Biotechnology for Biofuels 2014 7:1*, *7*(1), 1–26. https://doi.org/10.1186/1754-6834-7-7

Jensen, S. I., Lennen, R. M., Herrgård, M. J., & 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), 17874. https://doi.org/10.1038/srep17874

Kang, M.-K., & Nielsen, J. (2016). Biobased production of alkanes and alkenes through metabolic engineering of microorganisms. *Journal of Industrial Microbiology & Biotechnology 2016 44:4*, *44*(4), 613–622. https://doi.org/10.1007/S10295-016-1814-Y

Kessler, D., Leibrecht, I., & Knappe, J. (1991). Pyruvate-formate-lyase-deactivase and acetyl-CoA reductase activities of Escherichia coli reside on a polymeric protein particle encoded by adhE. *FEBS Letters*, *281*(1–2), 59–63. https://doi.org/10.1016/0014-5793(91)80358-A

Kim, G., Azmi, L., Jang, S., Jung, T., Hebert, H., Roe, A. J., Byron, O., & Song, J. J. (2019). Aldehyde-alcohol dehydrogenase forms a high-order spirosome architecture critical for its activity. *Nature Communications 2019 10:1*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-12427-8

Kim, G., Yang, J., Jang, J., Choi, J. S., Roe, A. J., Byron, O., Seok, C., & Song, J. J. (2020). Aldehyde-alcohol dehydrogenase undergoes structural transition to form extended spirosomes for substrate channeling. *Communications Biology 2020 3:1*, *3*(1), 1–9. https://doi.org/10.1038/s42003-020-1030-1

Klähn, S., Baumgartner, D., Pfreundt, U., Voigt, K., Schön, V., Steglich, C., & Hess, W. R. (2014). Alkane biosynthesis genes in cyanobacteria and their transcriptional organization. *Frontiers in Bioengineering and Biotechnology*, *2*(JUL). https://doi.org/10.3389/FBIOE.2014.00024

Kudo, H., Hayashi, Y., & Arai, M. (2019). Improving hydrocarbon production by engineering cyanobacterial acyl-(acyl carrier protein) reductase. *Biotechnol Biofuels*, *12*, 291. https://doi.org/10.1186/s13068-019-1623-4

Kunjapur, A. M., Tarasova, Y., & Prather, K. L. J. (2014). *Synthesis and Accumulation of Aromatic Aldehydes in an Engineered Strain of Escherichia coli*. https://doi.org/10.1021/ja506664a

Ladygina, N., Dedyukhina, E. G., & Vainshtein, M. B. (2006). A review on microbial synthesis of hydrocarbons. *Process Biochemistry*, *41*(5), 1001–1014. https://doi.org/10.1016/J.PROCBIO.2005.12.007

Lehtinen, T., Virtanen, H., Santala, S., & Santala, V. (2018). Production of alkanes from CO2 by engineered bacteria. *Biotechnology for Biofuels*, *11*(1), 228. https://doi.org/10.1186/s13068-018-1229-2

Lin, F., Das, D., Lin, X. N., & Marsh, E. N. G. (2013). Aldehyde-forming fatty acyl-CoA reductase from cyanobacteria: expression, purification and characterization of the recombinant enzyme. *The FEBS Journal*, *280*(19), 4773–4781. https://doi.org/10.1111/FEBS.12443

Liu, Y., Chen, S., Chen, J., Zhou, J., Wang, Y., Yang, M., Qi, X., Xing, J., Wang, Q., & Ma, Y. (2016). High production of fatty alcohols in Escherichia coli with fatty acid starvation. *Microbial Cell Factories*, *15*(1), 129. https://doi.org/10.1186/S12934-016-0524-5

Lorca, G. L., Ezersky, A., Lunin, V. V., Walker, J. R., Altamentova, S., Evdokimova, E., Vedadi, M., Bochkarev, A., & Savchenko, A. (2007). Glyoxylate and Pyruvate Are Antagonistic Effectors of the Escherichia coli IclR Transcriptional Regulator. *Journal of Biological Chemistry*, *282*(22), 16476–16491. https://doi.org/10.1074/JBC.M610838200

Membrillo-Hernández, J., Echave, P., Cabiscol, E., Tamarit, J., Ros, J., & Lin, E. C. C. (2000). Evolution of the adhE gene product of Escherichia coli from a functional reductase to a dehydrogenase: Genetic and biochemical studies of the mutant proteins. *Journal of Biological Chemistry*, *275*(43), 33869–33875. https://doi.org/10.1074/jbc.M005464200

Meyer, A. J., Segall-Shapiro, T. H., Glassey, E., Zhang, J., & Voigt, C. A. (2018). Escherichia coli “Marionette” strains with 12 highly optimized small-molecule sensors. *Nature Chemical Biology 2018 15:2*, *15*(2), 196–204. https://doi.org/10.1038/s41589-018-0168-3

O’Shea, C. A. (2023). NASA Clocks July 2023 as Hottest Month on Record Ever Since 1880 - NASA. *NASA.Gov*. https://www.nasa.gov/news-release/nasa-clocks-july-2023-as-hottest-month-on-record-ever-since-1880/

Partipilo, M., Claassens, N. J., & Slotboom, D. J. (2023). A Hitchhiker’s Guide to Supplying Enzymatic Reducing Power into Synthetic Cells. *ACS Synthetic Biology*, *12*(4), 947–962. https://doi.org/10.1021/ACSSYNBIO.3C00070

Rahman, Z., Sung, B. H., Yi, J. Y., Bui, L. M., Lee, J. H., & Kim, S. C. (2014). Enhanced production of n-alkanes in escherichia coli by spatial organization of biosynthetic pathway enzymes. *Journal of Biotechnology*, *192*(Part A), 187–191. https://doi.org/10.1016/j.jbiotec.2014.10.014

Rodriguez, G. M., & Atsumi, S. (2014). Toward aldehyde and alkane production by removing aldehyde reductase activity in Escherichia coli. *Metabolic Engineering*, *25*, 227–237. https://doi.org/10.1016/j.ymben.2014.07.012

Schirmer, A., Rude, M. A., Li, X., Popova, E., & Del Cardayre, S. B. (2010). Microbial biosynthesis of alkanes. *Science*, *329*(5991), 559–562. https://doi.org/10.1126/science.1187936

Sharma, A., Shakeel, T., Gupta, M., Rajacharya, G. H., & Yazdani, S. S. (2021). Biophysical and structural studies reveal marginal stability of a crucial hydrocarbon biosynthetic enzyme acyl ACP reductase. *Scientific Reports 2021 11:1*, *11*(1), 1–21. https://doi.org/10.1038/s41598-021-91232-0

Sheppard, M. J., Kunjapur, A. M., & Prather, K. L. J. (2016). Modular and selective biosynthesis of gasoline-range alkanes. *Metabolic Engineering*, *33*, 28–40. https://doi.org/10.1016/j.ymben.2015.10.010

Song, X., Yu, H., & Zhu, K. (2015). Improving alkane synthesis in Escherichia coli via metabolic engineering. *Applied Microbiology and Biotechnology 2015 100:2*, *100*(2), 757–767. https://doi.org/10.1007/S00253-015-7026-Y

Song, X., Yu, H., & Zhu, K. (2016). Improving alkane synthesis in Escherichia coli via metabolic engineering. *Applied Microbiology and Biotechnology*, *100*(2), 757–767. https://doi.org/10.1007/S00253-015-7026-Y/FIGURES/4

Stukenberg, D., Hensel, T., Hoff, J., Daniel, B., Inckemann, R., Tedeschi, J. N., Nousch, F., & Fritz, G. (2021). The Marburg Collection: A Golden Gate DNA Assembly Framework for Synthetic Biology Applications in Vibrio natriegens. *ACS Synthetic Biology*, *10*(8), 1904–1919. https://doi.org/10.1021/ACSSYNBIO.1C00126

Stukenberg, D., Hoff, J., Faber, A., & Becker, A. (2022). NT-CRISPR, combining natural transformation and CRISPR-Cas9 counterselection for markerless and scarless genome editing in Vibrio natriegens. *Communications Biology*, *5*(1). https://doi.org/10.1038/S42003-022-03150-0

Sun, S., Ding, Y., Liu, M., Xian, M., & Zhao, G. (2020). Comparison of Glucose, Acetate and Ethanol as Carbon Resource for Production of Poly(3-Hydroxybutyrate) and Other Acetyl-CoA Derivatives. *Frontiers in Bioengineering and Biotechnology*, *8*. https://doi.org/10.3389/FBIOE.2020.00833

Warui, D. M., Pandelia, M.-E., Rajakovich, L. J., Krebs, C., Bollinger, J. M., & Booker, S. J. (2015). Efficient Delivery of Long-Chain Fatty Aldehydes from the *Nostoc punctiforme* Acyl–Acyl Carrier Protein Reductase to Its Cognate Aldehyde-Deformylating Oxygenase. *Biochemistry*, *54*(4), 1006–1015. https://doi.org/10.1021/bi500847u

Zhang, F., Ouellet, M., Batth, T. S., Adams, P. D., Petzold, C. J., Mukhopadhyay, A., & Keasling, J. D. (2012). Enhancing fatty acid production by the expression of the regulatory transcription factor FadR. *Metabolic Engineering*, *14*(6), 653–660. https://doi.org/10.1016/J.YMBEN.2012.08.009

4.7 Supplementary data

Table S4.1. Primers used in this study

Name	Sequence	Purpose
oRZ36	AAGGTCTCGggagctcagataaaatatttgc	Amplify *lacI*/P_{tac} promoter from pST_116_LVL2_cam. Primers designed by Rohani, R. (2023, unpublished data)
oRZ37	TTGGTCTCGagtaaattgtgagcgctc	Amplify *lacI*/P_{tac} promoter from pST_116_LVL2_cam. Primers designed by Rohani, R. (2023, unpublished data)
JRN492	gaggtctcaGCTTTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGCTCACAATTTACTagagaccag	P_{tac} promoter with EB overhangs oligo, fw
JRN493	ctggtctctAGTAAATTGTGAGCGCTCACAATTCCACACATTATACGAGCCGATGATTAATTGTCAACAAAGCtgagacctc	P_{tac} promoter with EB overhangs oligo, rev
JRN388	CAGTTATGACCTCTGTACTTATAACAACAACGTAAGGTTATTGCGCTATGtgtgtaggctggagctgcttc	Amplify cmR marker from pKD3, *tesC* flanks, fw
JRN389	CGGCTTGAGCATCCGGCACCACAAAACGTTTACTTAACCATCTGCTCCAGcatatgaatatcctccttag	Amplify cmR marker from pKD3, *tesC* flanks, rev
JRN312	CGTTCCCGGTTGCTGTACCGGGAACGTATTTAATTCCCCTGCATCGCCCGtgtgtaggctggagctgcttc	Amplify cmR marker from pKD3, *yqhC-yqhD-dkgA* flanks, fw
JRN313	acattaccgccaccgggagaatttgcatgttagccgccgaactggtcaggcatatgaatatcctccttag	Amplify cmR marker from pKD3, *yqhC-yqhD-dkgA* flanks, rev
JRN390	tgccagactgaagaggtc	Check *tesC* locus
JRN391	agcgatgaagcagaaaacc	Check *tesC* locus
JRN318	cacgctgtttgtggtgatta	Check *yqhC-yqhD-dkgA* locus
JRN319	tgatgatgtccttatatttggc	Check *yqhC-yqhD-dkgA* locus
JRN239	TGAGGAAGACCAAATGCCGCAGCTGGAAGCAAGC	Primer used to amplify PCC7942_ADO as level 0 part
JRN240	TGAGGAAGACCAACCTTTAAACTGCTGCCAGACCATAG	Primer used to amplify PCC7942_ADO as level 0 part
JRN253	AACTGAAGACAGAATGGCTAAAGTTCTGTGCGTTCT	Amplify PsFDH with 'CD' overhangs for GGA, fw
JRN254	AACTGAAGACCAATCCTCACGGGTAGCGTGCC	Domesticate PsFDH, rv
JRN255	AACTGAAGACACGGATATGTACCCGGTTTGCG	Domesticate PsFDH, fw
JRN256	GCAACTGAAGACACACCTTTAAACAGCTTTCTTGAATTTAGCA	Amplify PsFDH with 'CD' overhangs for GGA, rev
JRN498	TGGAAGACTGAatggtcattaaggcgc	Amplify *E. coli fadR* with 'CD' overhangs for GGA
JRN499	ACGAAGACCTACCTttatcgcccctgaat	Amplify *E. coli fadR* with 'CD' overhangs for GGA
JRN504	TGGAAGACTGAatggtcgcacccattc	Amplify *E. coli iclR* with 'CD' overhangs for GGA
JRN505	ACGAAGACCTACCTtcagcgcattccacc	Amplify *E. coli iclR* with 'CD' overhangs for GGA

JRN189	CTGGCGCTGCGCTTT	Linearise *adhE* at the E568 locus, fw
JRN190	TTCGAAGTGAGTTTCCGGATG	Linearise *adhE* at the E568 locus, rev
JRN191	ATCCGGAAACTCACTTCGAAAAGCTGGCGCTG CGCTTTATGGA	Oligo to introduce E568K mutation

Table S4.2. Synthesised DNA fragments used in this study

Pseudomonas sp. 101 formate dehydrogenase (PsFDH)	TAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTA ACTTTAAGAAGGAGATATACATATGCATCATCACCATCACCACGCTAAAGTTCTGTG CGTTCTGTACGACGACCCGGTTGACGGTTACCCGAAAACCTACGCTCGTGACGACC TGCCGAAAATCGACCACTACCCGGGTGGTCAGACCCTGCCGACCCCGAAAGCTATC GACTTCACCCCGGGTCAGCTGCTGGGTTCTGTTTCTGGTGAACTGGGTCTGCGTAA ATACCTGGAATCTAACGGTCACACCCTGGTTGTTACCTCTGACAAAGACGGTCCGG ACTCTGTTTTCGAACGTGAACTGGTTGACGCTGACGTTGTTATCTCTCAGCCGTTCT GGCCGGCTTACCTGACCCCGGAACGTATCGCTAAAGCTAAAAACCTGAAACTGGCT CTGACCGCTGGTATCGGTTCTGACCACGTTGACCTGCAATCTGCTATCGACCGTAA CGTTACCGTTGCTGAAGTTACCTACTGCAACTCTATCTCTGTTGCTGAACACGTTGT TATGATGATCCTGTCTCTGGTTCGTAACTACCTGCCGTCTCACGAATGGGCTCGTAA AGGTGGTTGGAACATAGCTGACTGCGTAAGCCACGCTTACGACCTGGAAGCTATGC ACGTTGGTACCGTTGCTGCTGGTCGTATCGGTCTGGCTGTTCTGCGTCGTCTGGCT CCGTTCGACGTTCACCTGCACTACACCGACCGTCACCGTCTGCCGGAATCTGTTGA AAAAGAACTGAACCTGACCTGGCACGCTACCCGTGAAGACATGTACCCGGTTTGCG ACGTTGTTACCCTGAACTGCCCGCTGCACCCGGAAACCGAACACATGATCAACGAC GAAACCCTGAAACTGTTCAAACGTGGTGCTTACATCGTTAACACCGCTCGTGGTAAA CTGTGCGACCGTGACGCTGTTGCTCGTGCTCTGGAATCTGGTCGTCTGGCTGGTTA TGCGGGTGACGTGTGGTTCCCCCAGCCGGCTCCGAAAGACCACCCGTGGCGTACC ATGCCGTACAACGGTATGACCCCGCACATCTCTGGTACCACCCTGACCGCTCAGGC TCGTTACGCTGCTGGTACCCGTGAAATCCTGGAATGCTTCTTCGAAGGTCGTCCGA TCCGTGACGAATACCTGATCGTTCAGGGTGGTGCTCTGGCTGGTACCGGTGCTCAC TCTTACTCTAAAGGTAACGCTACCGGTGGTTCTGAAGAAGCTGCTAAATTCAAAAAA GCTGTTTAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG
Synochococcus elognatus PCC7942 acyl-ACP reductase (AAR)	TAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTA ACTTTAAGAAGGAGATATACATATGTTCGGCCTGATCGGGCACCTGACGAGCCTTG AACAGGCGCGTGACGTCTCCCGTCGTATGGGCTATGATGAGTACGCCGACCAAGG TTTGGAATTCTGGAGCAGCGCTCCACCGCAGATAGTTGATGAAATCACCGTAACATC AGCCACCGGCAAAGTGATTCATGGGCGATATATTGAGTCATGTTTCCTGCCGGAGA TGCTGGCAGCGCGCCGTTTTAAGACGGCTACGCGTAAAGTCCTCAATGCCATGAGT CATGCTCAGAAACATGGCATTGATATATCGGCTCTGGGCGGCTTTACCTCTATTATC TTTGAAAACTTTGATTTGGCCAGCCTAAGACAAGTACGCGACACCACCCTGGAGTTC GAGCGGTTCACAACCGGTAATACTCACACCGCTTATGTGATCTGCCGCCAGGTGGA AGCTGCGGCAAAAACGCTTGGCATTGATATCACTCAAGCGACTGTTGCGGTCGTAG GAGCAACGGGTGATATTGGTAGTGCCGTTTGCCGCTGGCTGGACCTGAAACTGGGT GTTGGGGATCTCATTCTGACCGCGCGTAACCAGGAGCGTTTAGACAACCTGCAGGC GGAACTGGGTCGGGGCAAAATTCTGCCGTTGGAAGCAGCGCTGCCGGAAGCGGAT TTTATTGTGTGGGTGGCCTCGATGCCTCAGGGTGTTGTGATCGACCCAGCCACCTT AAAACAACCGTGCGTCCTCATCGATGGTGGTTACCCCAAGAACCTTGGCTCTAAAGT TCAAGGGGAAGGCATTTACGTACTTAACGGAGGCGTGGTGGAACACTGTTTTGATA TTGATTGGCAGATCATGTCCGCAGCGGAAATGGCCCGCCCTGAAAGGCAGATGTTC GCATGTTTTGCAGAGGCCATGCTGTTAGAATTTGAAGGATGGCACACAAATTTTTCG TGGGGGCGCAATCAGATCACGATCGAGAAGATGGAAGCCATTGGAGAAGCATCCG TCCGCCACGGCTTCCAGCCGTTAGCGTTGGCGATTGGTAGCAGTGGTCATCATCAT CATCATCACTGACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTT TTG

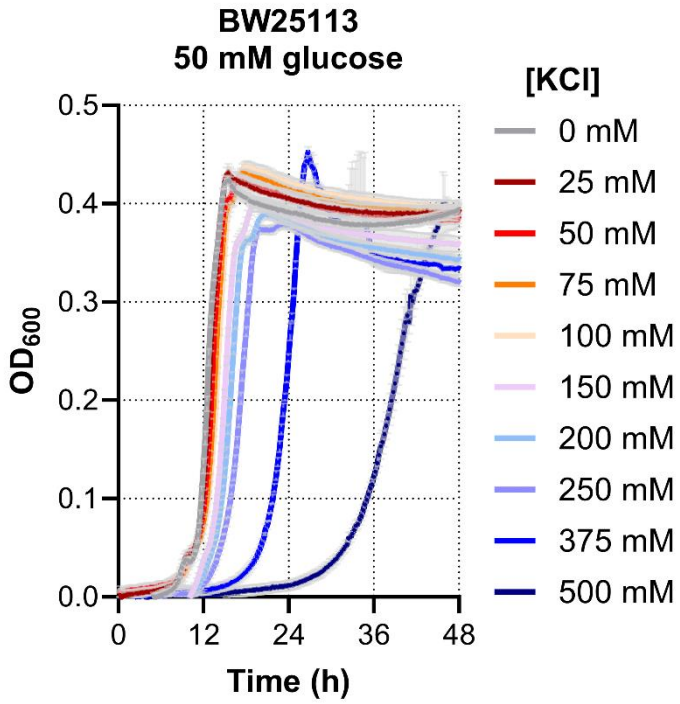


Figure S4.1. Tolerance of *E. coli* BW25113 to potassium chloride. Growth of *E. coli* BW25113 was assessed in modified M9 minimal medium supplemented with 50 mM glucose and varying concentrations of KCl. Datapoints represent averages of biological triplicate measurements, and errors bars represent their standard deviations.

5. Novel metabolic strategies for growth-coupling alkane biosynthesis in *E. coli*

5.1 Statement of collaborative contributions

This work was conducted in partial collaboration with Ari Satanowski (Max Plank Institute of Terrestrial Biology). He was responsible for construction of the $C_1(G)_{aux}$ and $C_1(G)S_{aux}$ strains used in this work, and he retains the right to first publication of this material.

5.2 Abstract

The field of microbial alkane biosynthesis has long been held back by the inefficient enzymes involved. This is partially due to a lack of specific, robust, and high-throughput methods to detect or select for alkane biosynthesis. Enzyme engineering approaches are therefore limited to rational mutagenesis and low-throughput analytical quantification of enzyme activity. In this work, we sought to explore several novel strategies for growth-coupling ADO activity to ultimately enable a high-throughput growth-based enzyme engineering pipeline. A first approach focussed on using imbalances in cofactor redox states to drive the biosynthesis of alkanes. Flux balance analysis indicated that a deficiency in NADPH oxidation could be used to establish weakly growth-coupled alkane biosynthesis. We also found that the formate by-product resulting from ADO activity could be stringently growth-coupled to NADH regeneration and ATP synthesis through formate dehydrogenase activity, given that the host cell is deficient in NADH regeneration. In another approach, strains dependent on formate for the production of 5,10-mTHF displayed significant improvements in growth as a result of AAR-ADO pathway expression. This latter result represents the first example of microbial growth being directly coupled to ADO product synthesis. Although this work does not include any enzyme engineering efforts, it presents several solutions to a longstanding problem within the field of alkane biosynthesis.

5.3 Introduction

Engineering hydrocarbon-producing enzymes is challenging due to a lack of robust, sensitive, and high-throughput screening and/or selection methods. As detailed more extensively in Chapter 2, some physical properties of hydrocarbon molecules, such as their relative insolubility in aqueous solutions, significantly limit our ability to measure their production *in vivo* with accuracy and in high throughput. In addition, compartmentalization of the target product is an essential requirement of using *in vivo* screening or selection methods, but hydrocarbon products can often passively diffuse out of the cell or are volatile at shorter chain lengths. Product leakage from the cell and/or product uptake by other cells in the population limits the use of *in vivo* enzyme engineering techniques, due to the connection between a variant's sequence and its activity being (partially) lost.

One of the most extensively studied hydrocarbon-producing enzymes is the aldehyde deformylating oxygenase (ADO), a ferritin-like diiron oxygenase (Schirmer et al., 2010). This enzyme catalyses the cleavage of a formate group from C_4-C_{18} aldehydes, forming a C_{n-1} alkane product which can be used as a direct fossil fuel replacement. However, its poor activity is noted as a key limitation to biological alkane production (Marsh & Waugh, 2013). Indeed, from the results gathered in Chapter 4 it was clear that ADO was the bottleneck to alkane biosynthesis in the AAR-ADO pathway.

Attempts to engineer ADO have been limited mostly to (semi)-rational approaches, including targeting non-conversed residues (Kudo et al., 2019), conserved residues (Wang et al., 2017), engineering the enzyme access tunnel (Bao et al., 2016; Khara et al., 2013; L. Zhang et al., 2016) and using consensus sequences (Shakeel et al., 2018). While these approaches have yielded useful findings, especially those related to substrate specificity, the necessary order of magnitude increases in ADO catalytic activity have not been achieved. Since rational approaches have struggled to establish highly improved variants, it is likely that the relationships between sequence, structure, and function of ADO, and the interactions of ADO with a non-native cellular environment, are poorly understood. Exploiting the principles of

directed evolution could overcome these issues by taking a ‘blind’ approach to engineering ADO (Arnold, 1998), potentially yielding mutants with higher catalytic activity and those that are better suited to heterologous overexpression. A handful of directed evolution approaches targeting ADO have been attempted, both identifying beneficial ADO mutations (D. Chen et al., 2023; Green, 2021; Liu et al., 2020). These relied either on a) screening for alkane biosynthesis via a GFP-linked biosensor or b) mitigating aldehyde toxicity through ADO activity. As discussed in Chapter 2, biosensor approaches can be limited by their substrate specificity and may enable ‘cheating’ populations over time, while resolving aldehyde toxicity does not directly target the products generated by ADO. The further development of robust, sensitive, and high-throughput screening or selection methods for ADO is therefore still a crucial aspect of the field.

Growth coupling represents a powerful method for ultra-high throughput enzyme selection. These approaches require that the enzyme of interest contributes to cell fitness, which is usually achieved by engineering a synthetic form of auxotrophy and allowing the target enzyme to rescue the deficiency (J. Chen et al., 2022; Orsi et al., 2021). To establish such an approach for ADO its products must be coupled to cell metabolism, which is challenging with hydrocarbon products as detailed in Chapter 2. However, the catalytic activity of ADO involves several other molecules besides the production of alkanes, including the oxidation of NADPH (via ferredoxin-mediated electron transfer), consumption of O_2 and the production of formate (Li et al., 2012; Warui et al., 2011; J. Zhang et al., 2013).

In recent years the development of cofactor auxotrophic microbes has enabled growth-coupling and enzyme engineering of oxidoreductases requiring the canonical NAD^+/NADH or $NADP^+$/NADPH cofactors, as well as non-canonical NMN^+/NMNH or NCD^+/NCDH cofactor pairs (Nielsen et al., 2023). The basic principle of cofactor auxotrophy involves blocking the cell’s native ability to cycle essential redox cofactors between their reduced or oxidised states. This causes a cessation of growth that can be rescued by the introduction of enzymes that perform the necessary redox reaction. A crucial benefit of growth-coupling via redox cofactor

auxotrophy is that the desired (non-cofactor) product of the enzyme does not have to be interacted with to establish a growth-coupled system. Enzymes that generate hydrophobic, insoluble, or gaseous products such as alkanes are therefore ideal targets for this approach.

The catalytic activity of ADO and the prior enzyme in the pathway, AAR, both require electron input ultimately in the form of NADPH. Therefore, a strain unable to oxidise NADPH natively could theoretically be used to select for either enzyme activity or that of the whole pathway through the oxidation rate of NADPH resulting from AAR/ADO activity. Strains deficient in NADPH oxidation have been used previously for the engineering of various NADPH-dependent enzymes, including a carboxylic acid reductase (Kramer et al., 2020), lactate dehydrogenase (L. Zhang et al., 2018), 4-hydroxybenzoate hydroxylase (Maxel, Aspacio, et al., 2020), and a cyclohexanone monooxygenase (Maxel et al., 2021; Maxel, Zhang, et al., 2020).

The ADO enzyme also produces formate in an equimolar ratio to alkanes, which means that formate production acts as a proxy for alkane biosynthesis (Warui et al., 2011). Some studies have used formate production to measure ADO activity *in vitro*, but none have attempted to use formate as an intracellular growth-coupling target. An *in vivo* formate-based approach would avoid most of the issues involved with the directed evolution of hydrocarbon production, such as the low solubility, passive diffusion, or evaporation of the target products, and instead allows the selection strategy to utilise the highly soluble and non-volatile formate byproduct as the selective molecule.

This chapter explores two ways to utilise ADO-produced formate to achieve growth-coupling. The first of these involves using a redox cofactor auxotroph as described earlier, but instead of rescuing an NADPH oxidation deficiency, we chose to establish selection for NAD^+ reduction. Since formate oxidation can be coupled to NAD^+ reduction, a strain deficient in the native enzymatic reactions responsible for this could be rescued by coupling endogenous formate biosynthesis to NAD^+ reduction. An *E. coli* strain displaying NADH auxotrophy had been established in previous work, however this strain was auxotrophic to all reducing power

as opposed to just NADH (Wenk et al., 2020). Since ADO catalytic activity requires reducing power, this previously described strain could not be used in its current form to select for formate production by ADO. In Chapter 6, we designed and constructed such an NADH auxotrophic *E. coli* through a combination of flux balance analysis, genetic engineering, and growth characterization. In the current chapter, we performed flux balance analysis using adapted models to assess the theoretical viability of this ADO growth-coupling method.

A second method to establish growth-coupling to formate relies on using it as a C_1 carbon source. Novel pathways to fix C_1 molecules such as formate and methanol have generated much interest in recent years, since they represent ways to fix derivatives of CO_2. The reductive glycine pathway is one such pathway, which utilises formate as a C_1 donor for the tetrahydrofolate (THF) pool (Claassens et al., 2022; Kim et al., 2020). This pathway requires heterologous expression of three enzymes, starting with a formate-THF ligase (*ftfL*) to form the intermediate 10-formyl-THF, followed by a methenyl-THF cyclohydrase (*fchA*) to form 5,10-methenyl-THF and a methylene-THF dehydrogenase (*mtdA*) to ultimately produce 5,10-methylene-THF. This C_1 molecule is required for the biosynthesis of, among others, essential metabolites methionine and thymine, which ensures formate growth-dependence. Glycine, a C_2 amino acid, can also be converted to serine (C_3) by an additional reaction with 5,10-methylene-THF via the activity of serine hydroxymethyltransferase (*glyA*), releasing THF. Serine can then enter central metabolism, for example when converted further to pyruvate. Using strains deficient in the biosynthesis of 5,10-methylene-THF and/or serine, growth could be made fully dependent on formate supplementation, offering a highly stringent selection method for formate (Yishai et al., 2017).

In this chapter, we designed novel approaches to establish growth-coupling of ADO based on cofactor auxotrophy ($NADP^+$ or NADH deficiency) and the use of formate as a precursor for C_1 molecule synthesis via the reductive glycine pathway. We verified the (theoretical) functionality of these strategies by flux balance analysis and growth assays. This work also demonstrated that formate produced by the AAR-ADO pathway can rescue growth of a

formate-dependent strain via activity of the reductive glycine pathway, resulting in the first system where AAR-ADO activity has been growth-coupled. In summary, this work demonstrated how a wider array of growth-coupled designs are possible when considering all products and substrates of a pathway or enzyme, as opposed to just the focal product of interest.

5.4 Materials and methods

5.4.1 Strains, plasmids, media and routine growth

Routine growth of *E. coli* was performed using either LB agar or LB liquid medium, at 37°C and shaking at 250 rpm, where necessary. Antibiotics were supplemented as required: carbenicillin (carb, 100 µg/mL), kanamycin (kan, 50 µg/mL), chloramphenicol (cm, 25 µg/mL). M9 minimal medium was routinely used for growth assays, composed of: 47.8 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 18.7 mM NH_4Cl, 2 mM $MgSO_4$, 100 µM $CaCl_2$, 134 µM ethylene diamine tetrachloroacetic acid (EDTA), 31 µM $FeCl_3{\cdot}6H_2O$, 6.2 µM $ZnCl_2$, 0.76 µM $CuCl_2{\cdot}2H_2O$, 0.42 µM $CoCl_2{\cdot}2H_2O$, 1.62 µM H_3BO_3, and 0.081 µM $MnCl_2{\cdot}4H_2O$, adjusted to pH 7.0. For expression of alkane biosynthesis enzymes, 100 µM $FeCl_3{\cdot}6H_2O$ was used. Carbon sources including D-glucose, glycerol, D-fructose, glycine, and sodium formate (all >99% grade, Sigma-Aldrich) were prepared in DI water and filter sterilised. All primers mentioned in the methods can be found in supplementary table S5.1.

Table 5.1. Strains used and constructed in this work

Name	Genotype	Reference
BW25113	*F- LAM- rrnB3 ΔlacZ4787 hsdR514 Δ(araBAD)567 Δ(rhaBAD)568 rph-1*	Purchased (DSMZ)
ΔP	BW25113 *Δpgi*::FRT	(Baba et al., 2006)
ΔPS	ΔP *ΔsthA*::FRT	This study
$C_1(G)_{aux}$	*ΔfrmRAB ΔgcvTHP ΔglyA ΔpflAB SS9::Prom.S-rbsC:MeFtfL-MeFch-MeMtdA*	Satanowski, A. (unpublished data)
$C_1(G)S_{aux}$	*ΔfrmRAB ΔgcvTHP ΔserA SS9::Prom.S-rbsC:MeFtfL-MeFch-MeMtdA-Kan*	Satanowski, A. (unpublished data)

Table 5.2. Plasmids used in this work

Name	Contents	Reference
pKD3	FRT-*cm*^{R}-FRT, oriR6K, *amp*^{R}	(Datsenko & Wanner, 2000)
pSIJ8	*araC*-P_{ara}-α-β-γ *rhaRS*-P_{rha}-FLP, *amp*^{R}	(Jensen et al., 2015)
DVA_pBR322_AF	*lacZα*, *amp*^{R}	See chapter 3
pL2_JRN072_AF	*lacI*/P_{tac}-B0034m-AAR P_{tac}-B0034m-ADO, *amp*^{R}, pBR322	See chapter 4
pL1_JRN097_AF	*lacI*/P_{tac}-B0034m-*fadR,* pRSF1010, *kan*^{R}	See chapter 4

5.4.2 Constructing Δpgi ΔsthA

The BW25113 Δ*pgi::FRT* strain (ΔP) was purchased with the knockout already introduced (Horizon Discovery, UK), and the deletion was confirmed using primers JR211+JRN212. The additional deletion of *sthA* was performed mostly as described elsewhere (Jensen et al., 2015). Briefly, ΔP was transformed with pSIJ8 by heatshock and cultivated at 30°C for all following steps. An O/N clonal culture of ΔP grown in LB was used to inoculate 50 mL of LB medium using 1 mL inoculum, which was grown for 1.5 hours at 37°C, 250 rpm, before inducing with 15 mM sterile L-arabinose for an additional 45 minutes to induce expression of lambda red recombination machinery. The culture was subsequent rapidly cooled in an ice slurry for 30 minutes, spun down at 1500 *g*, washed twice in ice-cold DI water spinning down between steps, and finally resuspended in 300 µL 10% ice-cold glycerol. A 50 µL aliquot of electrocompetent cells was mixed with 3.5 µL of >250 ng/µL linear *cmR* cassette, previously amplified using pKD3 template and primers with *sthA*-targeting flanks JRN126+JRN127 (DreamTaq, ThermoFisher). The mixture was transferred to a pre-chilled electroporation cuvette (0.1 cm gap) and a 1.8 kV pulse applied. Cells were immediately recovered in LB + 10 mM succinate + 20 mM glycerol + 20 mM acetate + 1 mM L-arabinose + carb at 30°C for 2.5 hours. After recovery, cells were plated on identically supplemented LB agar plates, except arabinose was omitted and 25 µg/mL chloramphenicol was added. Correct clones were identified by colony PCR (JRN079+JRN080). The *cmR* marker was removed by inoculating 35 mL LB + carb with 0.5 mL of an overnight culture of ΔP *sthA::cmR* pSIJ8, cultivating for 1.5 hours at 30°C and subsequently inducing flippase expression with 50 mM L-rhamnose for 4

hours. A 10^{-5} and 10^{-6} dilution was plated on LB + carb plates, incubated at 30°C, and correct clones were identified by colony PCR checking the *sthA* locus. Removal of pSIJ8 was achieved by a single O/N cultivation at 42°C.

5.4.3 Growth-rescue of ΔPS

All growth assays using ΔPS were performed in M9 minimal medium, supplemented with sterile carbon sources as indicated in the text. For aldehyde supplementation assays of ΔPS, M9 minimal medium was supplemented with 0.1% (v/v) Triton X-100 to enable solubilization of the aldehydes. Hexanal and octanal (>99.5%, Sigma-Aldrich) could be directly added to M9 triton medium. Hexadecanal (>95%, Sigma-Aldrich) was first dissolved in 100% v/v ethanol to a concentration of 100 mM before dilution into the M9 triton medium. First, ΔPS (transformed with pL1_JRN097_AF for the *fadR* growth-rescue experiments) was streaked on LB agar plates supplemented with relevant antibiotics to isolate single colonies, and the presence of Δ*pgi* and Δ*sthA* mutations were confirmed by colony PCR (Table S5.1, primers JRN211+JRN212 and JRN079+JRN080). Three colonies were picked as biological replicates and grown O/N in liquid LB medium at 37°C supplemented with either 20 mM glycerol or 10 mM fructose (see text), after which they were diluted 1:100 fold in liquid M9 minimal medium supplemented with 20 mM glycerol, and grown again O/N at 37°C. The resulting pre-cultures were washed 3 times using sterile 1x M9 salts and washed cells were used as inoculum for plate reader growth assays. Filter sterilised isopropyl β-D-1-thiogalactopyranoside (IPTG, Sigma) was used for induction of *fadR* at concentrations indicated in the text.

5.4.4 Growth characterization and rescue of C_1 auxotrophs

The C_1 auxotrophic strains ($C_1(G)_{aux}$ and $C_1(G)S_{aux}$) were cultivated with 2 mM glycine supplementation regardless of whether LB or M9 minimal medium was used, along with either 50 mM- or 100 mM sterile D-glucose as indicated in text and figure legends. For non-selective growth 1 mM sodium formate was also supplemented. For the AAR-ADO growth-rescue experiments, the $C_1(G)S_{aux}$ strain was transformed with pL2_JRN072_AF via a standard

heatshock transformation. Pre-cultures of C_1 auxotrophs were grown similarly to ΔPS, briefly: they were streaked on LB agar, gene deletions of single colonies were confirmed (Table S5.1, primers Ec127+Ec128 and Ec051+Ec052), three biological replicates were grown in liquid LB and transferred at a 1:100 dilution into M9 minimal medium as a final preculture. Growth of C_1 auxotrophs was performed at 37°C and in the presence of antibiotics, where appropriate. Before inoculation of either strain for growth assays, precultures were washed 3 times using sterile 1x M9 salts to avoid carryover of formate or other metabolites in the medium.

5.4.5 Flux balance analysis

The *e_coli_core* metabolic model, based on *Escherichia coli* MG1655, was retrieved from the BiGG Models database and used for all metabolic modelling simulations (King et al., 2016; Orth et al., 2010). Flux balance analysis was carried out in Python (v3.8) using COBRApy (v0.20.0) (Ebrahim et al., 2013). The pyruvate formate lyase (PFL) reaction was deactivated since this reaction only operates under strictly anaerobic conditions. Several reactions and their transport reactions, if necessary, were added to the model manually: pyruvate oxidase in the decarboxylating direction (POX), C_{18} acyl-ACP synthesis (C180SN), C_{18} acyl-ACP reductase (AAR18), C_{17} alkane biosynthesis by ADO (ALDDC17), NAD^+-dependent formate dehydrogenase (FDH_NAD), an $NADP^+$-dependent phosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPC) and a non-phosphorylating variant (GAPN). Fructose, acetate, succinate, and formate were set as carbon sources depending on the simulation (lower bounds between 0 to -50 A.U., see text for details). Where relevant, C_{18} fatty aldehydes were either made freely available for uptake (lower bound -1000) or restricted (lower bound 0 to -20, see text for details). All FBA simulations were run with biomass maximization as the objective.

5.4.6 Microtiter plate growth assays

Microbial growth in microtiter plates was assessed in a Tecan Spark plate reader (SparkControl v3.1 SP1 software), equipped with a small humidity cassette. Assays were

performed in 200 µL M9 minimal medium with a starting OD_{600} of 0.01 and wells were sealed with BreatheEasy (Sigma-Aldrich) membranes to avoid contamination. Growth took place at 37°C and measured by absorbance at 600 nm every 15 minutes with agitation in between readings occurring in four separate stages: 1) 225 seconds, 3 mm amplitude, 90 rpm; 2) 225 seconds, 3 mm amplitude, 180 rpm; 3) 225 seconds, 2.5 mm amplitude, 108 rpm; 4) 225 seconds, 2.5 mm amplitude, 216 rpm. Growth rates were calculated through a log conversion of the blank-medium corrected OD values, followed by determining the maximum linear slope of these values over time using a 6-12 hour sliding window per replicate.

5.5 Results

5.5.1 $NADP^+$ auxotrophy for growth-coupled alkane biosynthesis

The biosynthesis of alkanes in *E. coli* is achieved through the fatty acid synthesis (FAS) pathway, which is initiated at the acetyl-CoA node and requires multiple cycles of carbon- and reducing power input, the latter of which is delivered mainly by NADPH (see Figure 5.1 for an overview of NADP(H)-dependent pathways in central metabolism and fatty acid biosynthesis). The FAS pathway in *E. coli* involves many enzymes, starting with acetyl-CoA carboxylating enzymes (*accABCD*) and followed by the *fab* enzymes (*fabD, fabH, fabG, fabA, fabZ, fabI, fabB and fabF*) (Janßen & Steinbüchel, 2014). Of these, *fabG* and *fabI* utilise the NADPH cofactor, although *fabI* has a more promiscuous cofactor specificity and can also utilise NADH (Bergler et al., 1996). Chain elongation via fatty acid biosynthesis is therefore dependent on NADPH availability, a property that can be exploited for growth-coupling using cofactor auxotrophic strains. If other pathways for NADPH consumption are removed, and the cell is engineered to overproduce NADPH, flux through NADPH-consuming pathways fatty acid- and alkane biosynthesis could theoretically be coupled to growth. Of course, fatty acid biosynthesis is already strictly growth-coupled since it is essential for cell metabolism, but flux through this pathway and the terminal alkane biosynthesis enzymes could potentially be increased by creating imbalances in the $NADP^+$/NADPH ratio.

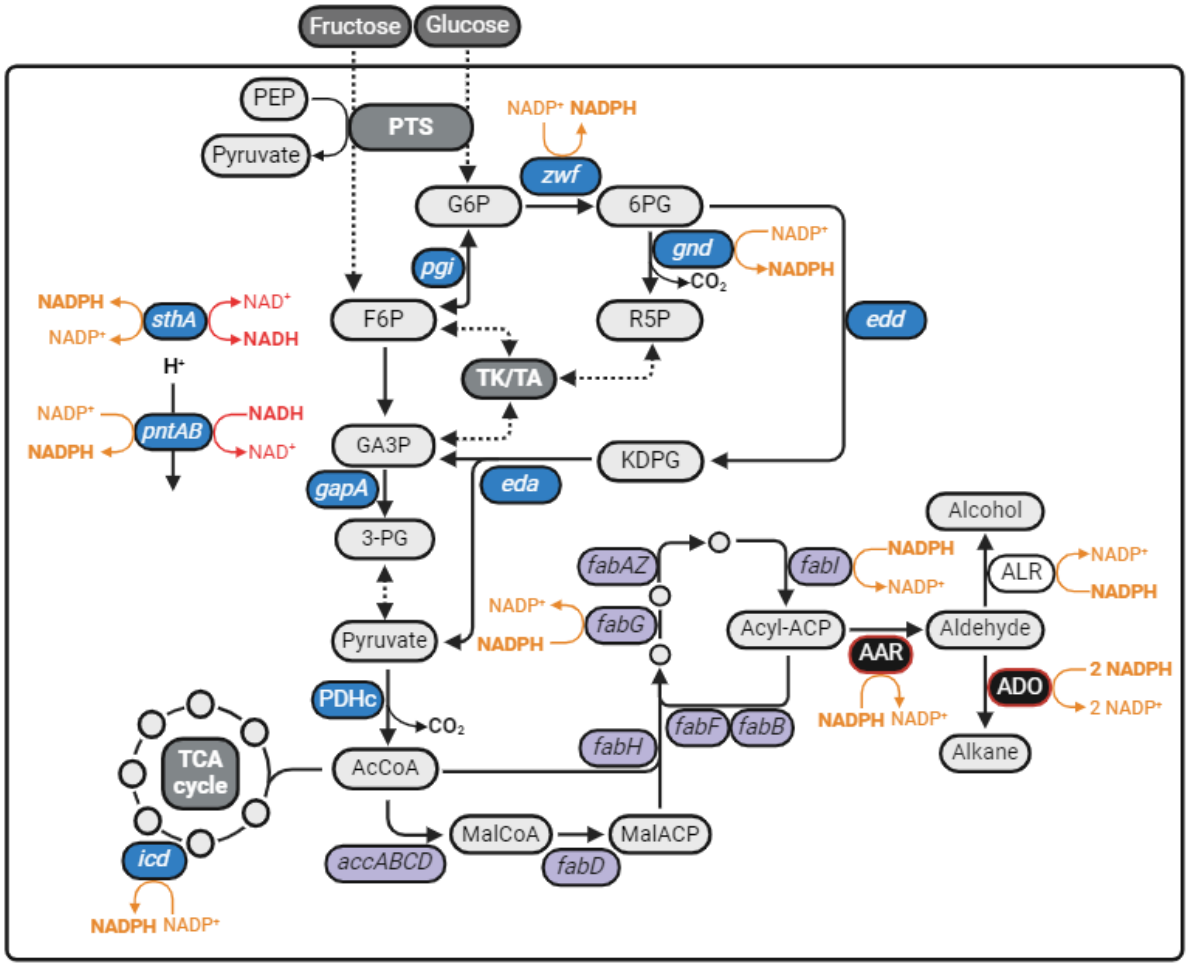


Figure 5.1. Schematic overview of *E. coli* central metabolism and fatty acid biosynthesis, with a focus on NADP(H)-utilising enzymes. Gene names of central metabolic enzymes are represented in blue boxes, while gene names of fatty acid synthesis genes are represented in purple boxes. Alkane biosynthesis enzymes are shown in black boxes with red rim. Abbreviations: TK/TA (transketalase/transaldolase reactions); PTS (phosphotransferase system); PDHc (pyruvate dehydrogenase complex); *icd* (isocitrate dehydrogenase); *pntAB* (pyridine nucleotide transhydrogenase); *sthA* (soluble transhydrogenase); *pgi* (glucose-6-phosphate isomerase); *zwf* (glucose-6-phosphate dehydrogenase); *gnd* (6-phosphogluconate dehydrogenase); *edd* (phosphogluconate dehydrogenase); *eda* (2-keto-3-deoxygluconate 6-phopshate aldolase); *accABCD* (acetyl-CoA carboxylase subunit A-D); *fabD* (malonyl CoA-acyl carrier protein transacylase); *fabB, fabF, fabH* (3-ketoacyl-ACP synthase I, II and II, respectively); *fabG* (3-ketoacyl-ACP reductase); *fabAZ* (3-hydroxyacyl-ACP dehydrase); *fabI* (enoyl-ACP reductase); AAR (acyl-ACP reductase), ADO (aldehyde deformylating oxygenase), ALR (aldehyde reductase).

Constructing a suitable selection strain would require the overproduction of NADPH and a deficiency in recycling back $NADP^+$. This strategy has previously been used to engineer NADPH-dependent enzymes, which we have discussed at length in section 2.7.2. Briefly, to overproduce NADPH, flux from glucose catabolism can be redirected through the NADPH-

generating pentose phosphate pathway by deleting *pgi* in addition to the soluble transhydrogenase *sthA*, the latter of which avoids the consumption of excess NADPH to form NADH (Maxel, Aspacio, et al., 2020; Sauer et al., 2004). Alternatively, glycolysis in general can be made fully NADPH-producing by replacing the native NAD^+-dependent *gapA* gene with an $NADP^+$-dependent version such as *gapN* (Kramer et al., 2020).

We constructed the Δ*pgi* Δ*sthA* (ΔPS) strain and assessed its growth on glucose and glycerol (Figure 5.2A and B). To determine whether NADPH-consuming pathways could incite growth-coupled behaviour in this mutant, we also compared the growth of BW25113 WT and ΔPS with aldehyde supplementation (Figure 5.2C and D). Endogenous *E. coli* aldehyde reductases (ALRs) are mostly NADPH-dependent (Rodriguez & Atsumi, 2014), so supplemented aldehydes could act as an electron acceptor for NADPH through ALR-mediated reduction of aldehydes to fatty alcohols.

The use of glucose in ΔPS as opposed to the WT resulted in the strain not entering an exponential growth phase at least for the ~140 hours of the growth assay. When using glycerol, ΔPS growth fully recovered reaching maximum OD within 48 hours (Figure 5.2B). Since glycerol enters central metabolism ‘below’ the *pgi* reaction, this ensures carbon can enter central metabolism without being forced through the PPP, avoiding the concomitant NADPH release. The addition of aldehydes (5 mM hexanal/octanal or 1 mM hexadecanal) to ΔPS grown on 10 mM glucose resulted in a clear exponential growth phase starting around 72h of cultivation. Hexadecanal supplementation resulted in slightly lower growth rates compared to using hexanal or octanal, although this can be attributed to its lower concentration (1 mM as opposed to 5 mM) and its presumed lower uptake rate into *E. coli*. These results indicated that growth of ΔPS using glucose was considerably improved by aldehyde supplementation due to ALR activity oxidizing excess NADPH. When the WT was supplemented with the same aldehydes, we observed no difference in growth, indicating that the aldehydes were not toxic at the indicated concentrations (Figure 5.2C).

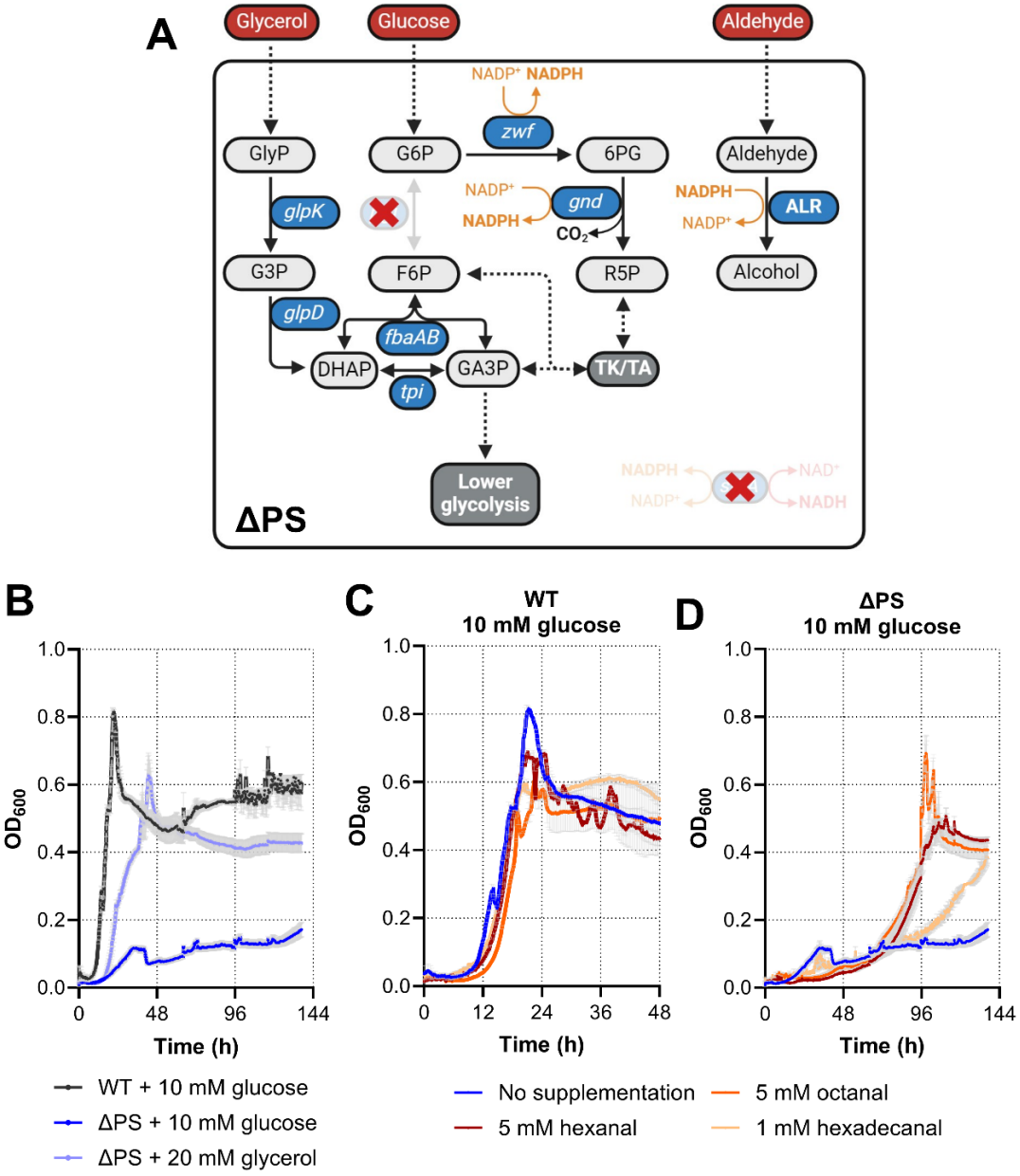


Figure 5.2. Growth rescue of *E. coli* BW25113 Δ*pgi* Δ*sthA* through aldehyde reductase activity. (A) Schematic of relevant cell metabolism, including the *Δpgi* and *ΔsthA* deletions (ΔPS), forcing flux into the PPP and blocking transhydrogenase activity. The NADPH-consuming aldehyde reductase (ALR) activity is also indicated. **(B)** Growth of *E. coli* BW25113 WT on glucose and ΔPS on glucose, resulting in a growth defect, or glycerol, which enters metabolism beyond the *pgi* reaction thereby restoring growth. **(C)** No inhibitory effects on growth were detected when *E. coli* BW25113 was supplemented with 1 mM hexadecanal or 5 mM hexanal or octanal. **(D)** Addition of aldehydes to ΔPS grown on glucose resulted in a partial restoration of growth compared to a non-supplemented condition. All growth data represent averages of biological triplicate measurements, and error bars represent their standard deviations. Abbreviations: TK/TA (transketalase/transaldolase reactions); *sthA* (soluble transhydrogenase); *pgi* (glucose-6-phosphate isomerase); *zwf* (glucose-6-phosphate dehydrogenase); *gnd* (6-phosphogluconate dehydrogenase); *glpK* (glycerol kinase); *glpD* (glycerol 3-phosphate dehydrogenase); *fbaAB* (fructose-bisphosphate aldolases); *tpi* (triosephosphate isomerase); ALR (aldehyde reductase).

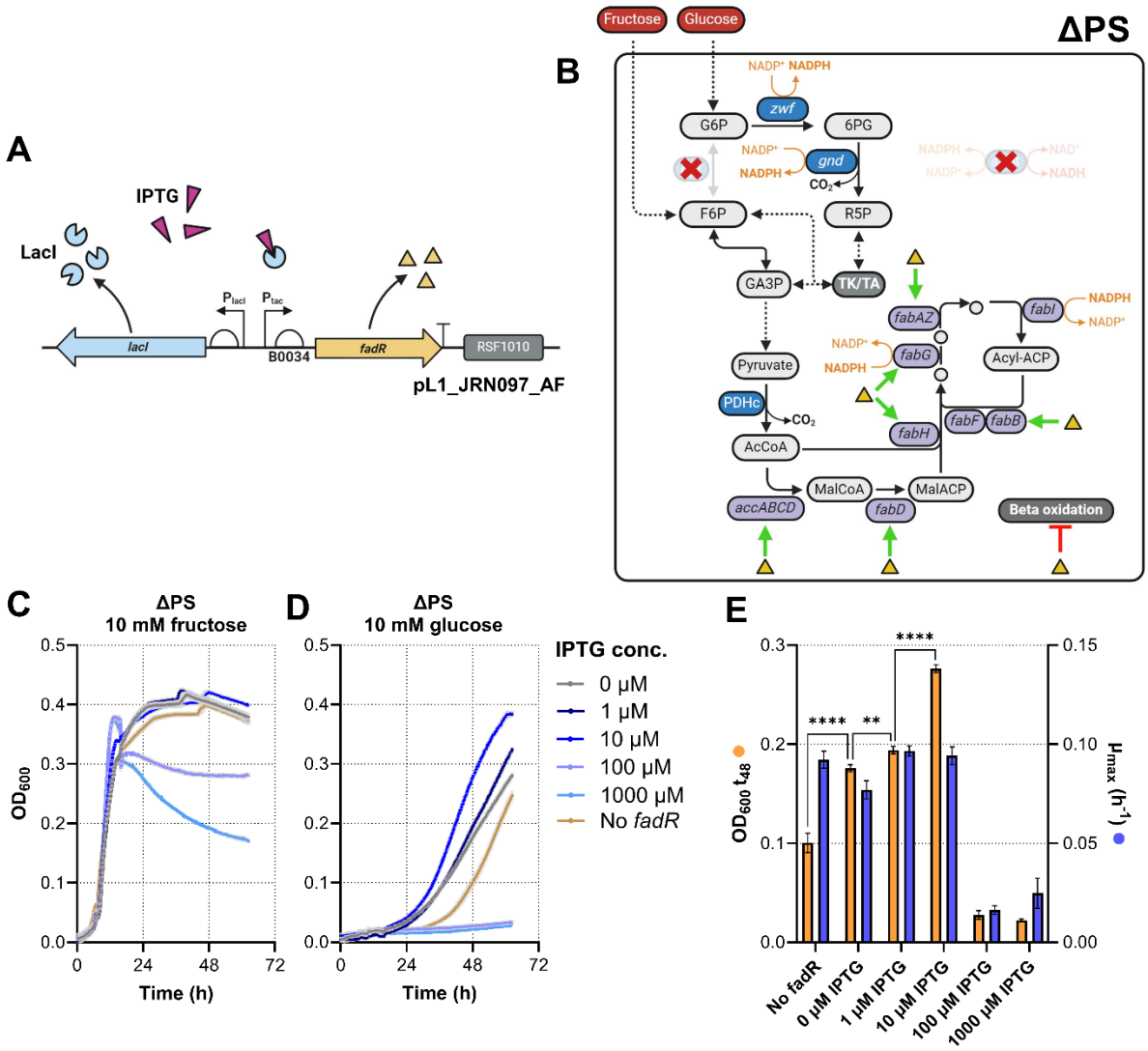


Figure 5.3. Growth rescue of *E. coli* BW25113 Δ*pgi* Δ*sthA* through induction of fatty acid biosynthesis. **(A)** Schematic of pL1_JRN097_AF carrying *fadR* under *lacI* control, inducible by IPTG. **(B)** Metabolic pathways involved in fructose/glucose catabolism and fatty acid biosynthesis. Deletion of *pgi* and *sthA* is indicated by red crosses. Blue and purple boxes contain names of *E. coli* genes coding for relevant enzymes, light grey boxes indicate metabolites, dark grey boxes indicate a collection of metabolic reactions, and red boxes indicate substrates. Certain enzymatic reactions involve the NADP(H) cofactor as indicated. Yellow triangles indicate *fadR* protein, with green arrows representing upregulation and red T symbols representing repression. The ΔPS strain carried pL1_JRN097_AF except in the 'No *fadR*' condition and was grown in M9 minimal medium with **(C)** 10 mM fructose or **(D)** 10 mM glucose. **(E)** OD_{600} values at t_{48} and the maximum growth rates of ΔPS grown using 10 mM glucose and varying amounts of IPTG. Growth curves represent the average of biological triplicate measurements, and error bars represent the standard deviation thereof. Where indicated, statistical analysis was performed by comparing means using a one-way ANOVA and post-hoc Tukey's multiple comparison test. P-value significance indicated are as follows: ns (p-value >0.05); * (p-value <0.05), ** (p-value <0.01), *** (p-value <0.001) or **** (p-value <0.0001).

This NADP(H)-linked growth coupling experiment specifically targeted one type of enzymatic reaction. To attempt a broader whole-cell approach, we sought to use ΔPS as a growth-coupling strain for an entire pathway. As previously mentioned, fatty acid biosynthesis requires NADPH for its activity. Fatty acid synthesis is regulated by *fadR* and *fabR*, the former of which has been used previously to significantly upregulate fatty acid biosynthesis genes resulting in higher fatty acid-derived product titres (Figure 5.3B). We used pL1_JRN097_AF, a low-copy plasmid encoding *fadR* under IPTG inducible control, to attempt to increase the rate of fatty acid synthesis in ΔPS and therefore also the NADPH consumption rate (Figure 5.3A).

The expression of *fadR* in fructose-grown strains resulted in identical growth rates irrespective of the induction level used (Figure 5.3C). However, sharp decreases in stationary phase OD_{600} were seen for strains carrying highly induced *fadR*, indicating a defect in cell maintenance after fructose was consumed. When grown on glucose, ΔPS carrying non-induced pL1_JRN097_AF already displayed a shorter lag phase than ΔPS without a plasmid, which we presumed was due to leaky expression of the IPTG-inducible *fadR* construct (Figure 5.3D). When induced with low levels of IPTG (1 and 10 μM), we observed additional shortening of the lag phase compared to non-induced- and no-plasmid controls, although growth rates were not significantly different (Figure 5.3D and E). At high IPTG induction levels (100 and 1000 μM) growth using glucose was negatively impacted, resulting in more-or-less no growth over the ~60 hours of the assay. The reduced lag phases of *fadR*-expressing ΔPS indicated that fatty acid biosynthesis was (weakly) growth coupled, but with significant non-selective growth also taking place. We note that ΔPS was able to grow in what we assumed to be selective conditions, which contradicts previous results (Figure 5.2). However, the preculture conditions of these two experiments differed, since for the *fadR* growth-rescue work 10 mM fructose was used as the preculture carbon source as opposed to 20 mM glycerol. The switch from glycerol to glucose in the previous work represents a more significant metabolic rearrangement than a switch between fructose- and glucose utilisation, which enter at nearly identical points in

central metabolism. This may have resulted in the longer-lasting growth defect when using glycerol-grown precultures.

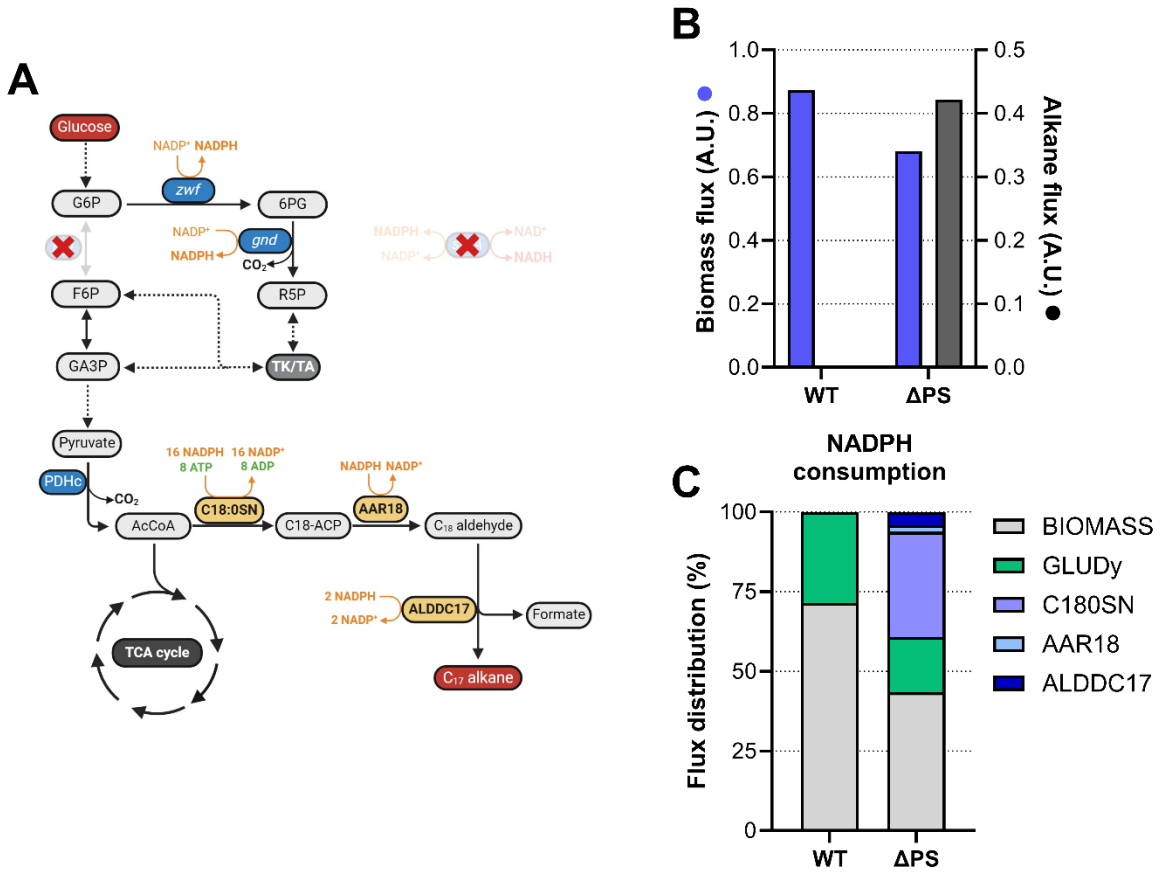


Figure 5.4. Flux balance analysis of ΔPS for growth-coupled alkane biosynthesis. (A) Schematic of cell metabolism involving *pgi* and *sthA* deletions, and the introduced alkane biosynthesis pathway consisting of C18:0SN (C_{18} acyl-ACP synthesis), AAR18 (acyl-ACP reductase, AAR) and ALDDC17 (aldehyde deformylating oxygenase, ADO). **(B)** Biomass- and alkane fluxes of an unmodified model (WT) and ΔPS both equipped with C18:0SN, AAR18 and ALDDC17 reactions, indicating reduced biomass flux in the latter but accompanied by alkane-producing flux. **(C)** Distribution of NADPH-consuming flux in WT and ΔPS models, demonstrating that flux is passing through the alkane biosynthesis pathway, but that AAR and ADO contribute little to the total NADPH consumption. GLUDy represents the glutamate dehydrogenase reaction.

To assess if growth coupling alkane biosynthesis using such an NADPH overproducer was theoretically feasible, we generated models of WT and ΔPS for flux balance analysis. Reactions for fatty acid synthesis (C180SN), acyl-ACP reductase (AAR18), and aldehyde deformylating oxygenase (ALDDC17) activity were added to the models manually (Figure 5.4A). When the WT model was run with 10 A.U. glucose as sole carbon input, no alkane flux

was reported, and all NADPH was consumed either for biomass formation or via the L-glutamate-producing GLUDy reaction (Figure 5.4B and C).

In contrast, running the ΔPS model in the same conditions resulted in lower biomass flux, but also alkane-producing flux, and around 40% of all NADPH was now consumed via the C180SN, AAR18 and ALDDC17 reactions. In fact, removing the C180SN reaction resulted in an infeasible solution, indicating that an NADPH sink was essential for ΔPS biomass flux (data not shown).

Despite some encouraging results using the ΔPS strain, both *in vivo* and *in silico*, we foresaw several issues moving forward with this type of selection system. The first was that NADPH is required for a wide variety of anabolic reactions in *E. coli*, meaning that coupling growth to NADPH consumption solely via chain elongation and alkane biosynthesis would require an extensive -and possibly infeasible- degree of metabolic engineering. Secondly, as demonstrated in Figure 5.4, metabolic modelling of ΔPS indicated that alkanes could be biosynthesised in a growth-coupled manner via the native fatty acid synthesis pathway (Figure 5.4B). However, the contribution of AAR (AAR18) and ADO (ALDDC17) to overall NADPH consumption was minimal (~6% of total NADPH-consuming flux), while fatty acid synthesis (C18:0SN) was responsible for 33% of the NADPH-consuming flux (Figure 5.4C). This indicated that the terminal AAR-ADO reactions could only ever be weakly growth-coupled, since the prior fatty acid synthesis flux would be required for their formation. Finally, the intermediate products of the AAR-ADO pathway are aldehydes, which we have shown can quite easily relieve $NADP^+$ auxotrophy *in vivo* through the action of endogenous ALRs. Whether aldehydes are produced via AAR or supplemented directly, selection will act on the numerous ALRs in *E. coli* capable of oxidizing NADPH. To circumvent this, again more metabolic engineering would be required to establish strains which not only overproduce NADPH and are unable to regenerate $NADP^+$, but also lack ALR activity. We deemed this level of metabolic engineering infeasible for this study, and therefore attempted to find more sensitive and ADO-specific selection methods to establish alkane growth-coupling.

5.5.2 Growth-coupled alkane biosynthesis via NADH auxotrophy

As we described in the previous section, using $NADP^+$/NADPH cofactor cycling as a selection method comes with some challenges. Not least of these is its low selectivity: since FAS and any other NADPH-consuming pathway could potentially rescue growth, the terminal AAR/ADO enzymes were expected to contribute little to $NADP^+$ regeneration rate when biosynthesizing alkanes from a carbon source such as glucose. Since we sought to specifically target alkane biosynthesis by ADO in this work, we required a way to couple ADO activity to growth while still retaining the capacity for high-throughput enzyme engineering.

The ADO enzyme produces formate in a 1:1 equimolar ratio to alkanes (Warui et al., 2011). Unlike other hydrocarbon-forming enzymes that release CO_2 or CO as byproducts, the release of formate represents an opportunity in terms of measuring ADO's catalytic properties *in vivo*, since formate is highly soluble and can be linked to several metabolic processes.

Formate dehydrogenases can catalyse the oxidation of formate, releasing CO_2 and NAD(P)H in the process. Since this process reduces a redox cofactor, we envisioned indirectly coupling ADO activity to the cycling of redox cofactors using formate as the intermediate electron carrier. We chose to focus on coupling ADO activity to NAD^+/NADH cycling. This is because ADO activity requires input of 4 electrons (2 NADPH via ferredoxin), and the oxidation of formate yields 2 electrons, so coupling ADO activity to the release of NADPH by formate oxidation would not even provide sufficient reducing power to drive the ADO reaction itself. We therefore aimed to construct a strain capable of generating high levels of NAPDH to drive ADO or AAR-ADO activity together with fatty acid biosynthesis, while still being sensitive to NADH. The design, construction, and validation of this 'NADH auxotroph' is extensively detailed in Chapter 6. In that chapter, the strain is presented more generally as being capable of selecting for any NADH-producing reaction. In this current chapter, we focused on demonstrating the theoretical validity of using this strain as a host for growth-coupling ADO activity.

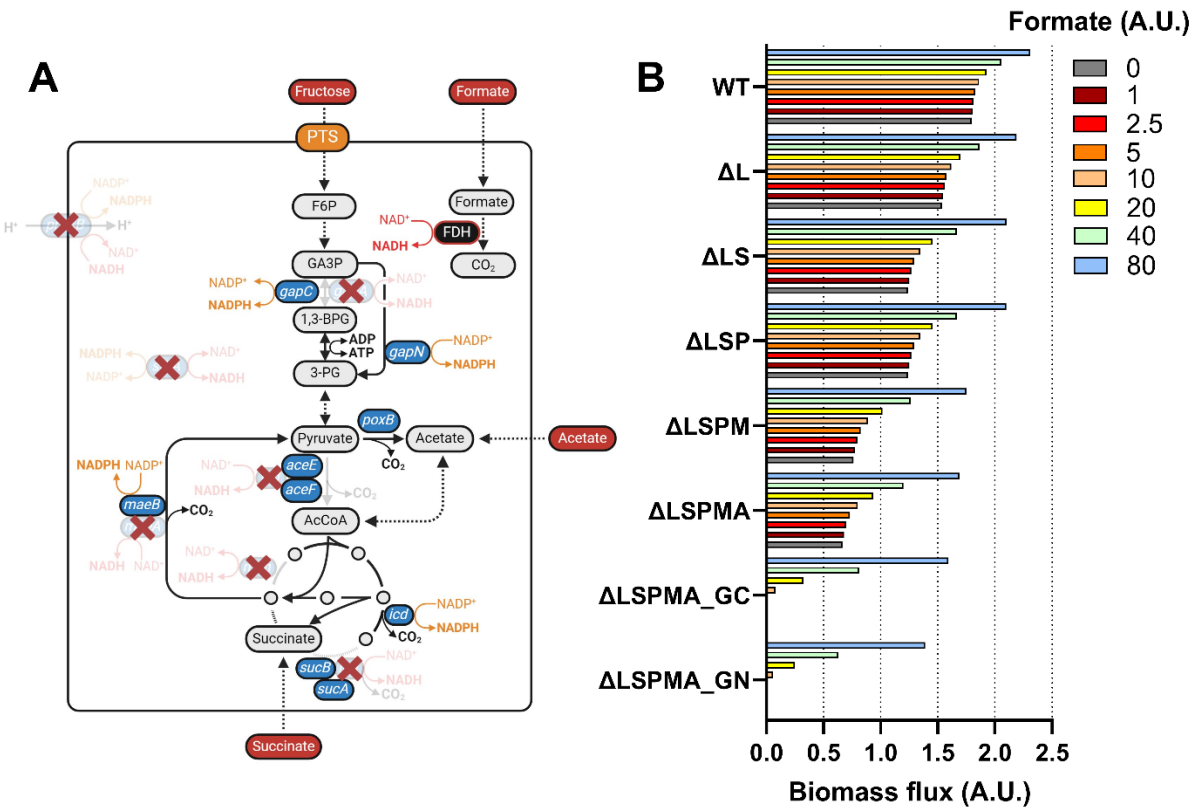


Figure 5.5. Dependence of biomass flux on formate availability in NADH-deficient *E. coli* models. (A) Partial schematic of engineered cell metabolism represented in the E_coli_core model, with modifications representing the gene deletions of *mdh* (M), *maeA* (A), *lpd* (L), *sthA* (S), *pntAB* (P) and the replacement of *gapA* with either *gapC* (GC) or *gapN* (GN). The formate dehydrogenase (FDH) reaction was added to the model manually. Dotted lines represent substrate import reactions or reactions spanning multiple enzymatic reactions. **(B)** The effect of formate availability (A.U.) on biomass flux (A.U.) per *in silico* mutant. Carbon source availability was set to 10 A.U. fructose, 20 A.U. acetate and 10 A.U. succinate. Abbreviations: *aceE* (pyruvate dehydrogenase); *aceF* (dihydrolipoamide acetyltransferase); *gapA* (NAD^+-dependent glyceraldehyde-3-phosphate dehydrogenase); *gapC* ($NADP^+$-dependent glyceraldehyde-3-phosphate dehydrogenase); *gapN* (non-phosphorylating $NADP^+$-dependent glyceraldehyde-3-phosphate dehydrogenase); *icd* (isocitrate dehydrogenase); *lpd* (dihydrolipoyl dehydrogenase); *maeA* (NAD^+-dependent malic enzyme); *maeB* ($NADP^+$-dependent malic enzyme); *mdh* (malate dehydrogenase); *pntAB* (pyridine nucleotide transhydrogenase); *poxB* (pyruvate oxidase) *sthA* (soluble transhydrogenase); *sucA* (2-oxoglutarate dehydrogenase); *sucB* (dihydrolipoyllysine succinyltransferase); FDH (formate dehydrogenase); PTS (phosphotransferase system).

We used the FBA model developed in Chapter 6 for this work, featuring the simplified e_coli_core model with knockouts in malate dehydrogenase (MDH, M), malic enzyme (MAEA, A), soluble transhydrogenase (STHA, S), membrane-bound transhydrogenase (PNTAB, P), lipoamide dehydrogenase (LPD, L) and a replacement of the native NAD^+-dependent glyceraldehyde-3-phosphate dehydrogenase (GAPDH) with $NADP^+$-dependent phosphorylating (*gapC*, GC) or non-phosphorylation variants (*gapN*, GN) (Figure 5.5A). To

assess the potential of (AAR)-ADO growth-coupling, we altered the model to include the necessary reactions coding for a formate dehydrogenase (FDH), fatty acid synthesis (C18:0SN), AAR (AAR18) and ADO (ALDDC17) reactions.

Before assessing *in silico* growth-coupling of alkane biosynthesis, the dependence of biomass flux on formate availability was determined (Figure 5.5B). Several models constructed in Chapter 6 were used for this purpose including some with reduced, but non-zero, NADH flux taking place in central metabolism when provided fructose as a carbon source (ΔL, ΔLS, ΔLSP, ΔLSPM, ΔLSPMA). Two *in silico* NADH auxotrophs, ΔLSPMA_GC and ΔLSPMA_GN, were also included. We observed that all models harbouring the NADH-yielding GAPDH reaction were able to establish biomass flux without any formate availability, indicating their NADH autotrophy. In contrast, the ΔLSPMA_GC and GN variants were unable to establish biomass flux without at least 10 A.U. formate present. Biomass flux of these two models increased with formate availability, indicating a biomass flux dependence on formate oxidation. The other GAPDH-carrying strains also displayed increasing biomass fluxes with increasing formate presence, although the fold change in biomass flux as a function of formate availability differed between the models. Those with less NADH-generating reactions remaining benefitted more in terms of biomass flux when formate was made available compared to those with more NADH-generating left intact in the model. This indicated that the biomass flux response to formate availability was dependent on the presence (or absence) of NADH-generating steps in metabolism.

Having confirmed that the model's biomass fluxes were formate-dependent to varying degrees depending on the availability of NADH-generating reactions, we set out to attempt *in silico* growth-coupling to the alkane- and formate producing ADO enzyme (ALDDC17) (Figure 5.6A). The principle of this approach relies on feeding aldehyde substrate to ADO, which performs a deformylation to yield formate and a C_{n-1} alkane. The released formate can then be used to drive NAD^+ reduction by formate dehydrogenase, generating a pool of NADH destined mainly

to operate oxidative phosphorylation. The ATP generated from this process can subsequently be used for growth and maintenance.

To simulate the deformylation of aldehydes supplied in the medium by ADO, we added a freely available source of C_{18} aldehyde to the extracellular environment in the model, representing a real-life substrate of the ADO enzyme. Indeed, we have measured C_{17} and $C_{17:1}$ alkane production in Chapter 4. Using the same set of models used to assess biomass flux dependence on formate oxidation, we now included the ALDDC17 reaction along with the FDH reaction and set the C_{18}-al substrate lower bound to -1000 A.U., simulating a freely available substrate. The results indicated that running ΔLSPM, ΔLSPMA and ΔLSPMA_GC or GN models resulted in C_{17} alkane flux, and alkane flux was highest using the ΔLSPMA_GC or GN models (Figure 5.6B). Along with the formate dependence data this indicated that NADH deficiencies were indeed driving the uptake and deformylation of the C_{18}-al substrate in these models.

Although ΔLSPMA_GC and ΔLSPMA_GN displayed almost identical biomass fluxes, the alkane production flux of the GN variant was around 1.5-fold higher than the GC variant. We attribute this to the GN reaction being non-phosphorylating, i.e. the reaction generates 3-phosphoglycerate directly, and not 1,3-bisphosphoglycerate, which lowers the potential ATP yield of glycolysis. This means that a higher NADH flux would be required to achieve the same level of ATP flux as ΔLSPMA_GC does with the same carbon input. As a result, ΔLSPMA_GN deformylates more C_{18}-al to compensate. We also assessed ΔLSPMA_GN alone with defined availabilities of C_{18}-al and found that biomass flux could be established with 5 A.U. of C_{18}-al (Figure 5.6C). Interestingly, this equates to 5 A.U. of formate generation, whereas the formate dependence results indicated that 10 A.U. of formate was required to establish any biomass flux (Figure 5.5B).

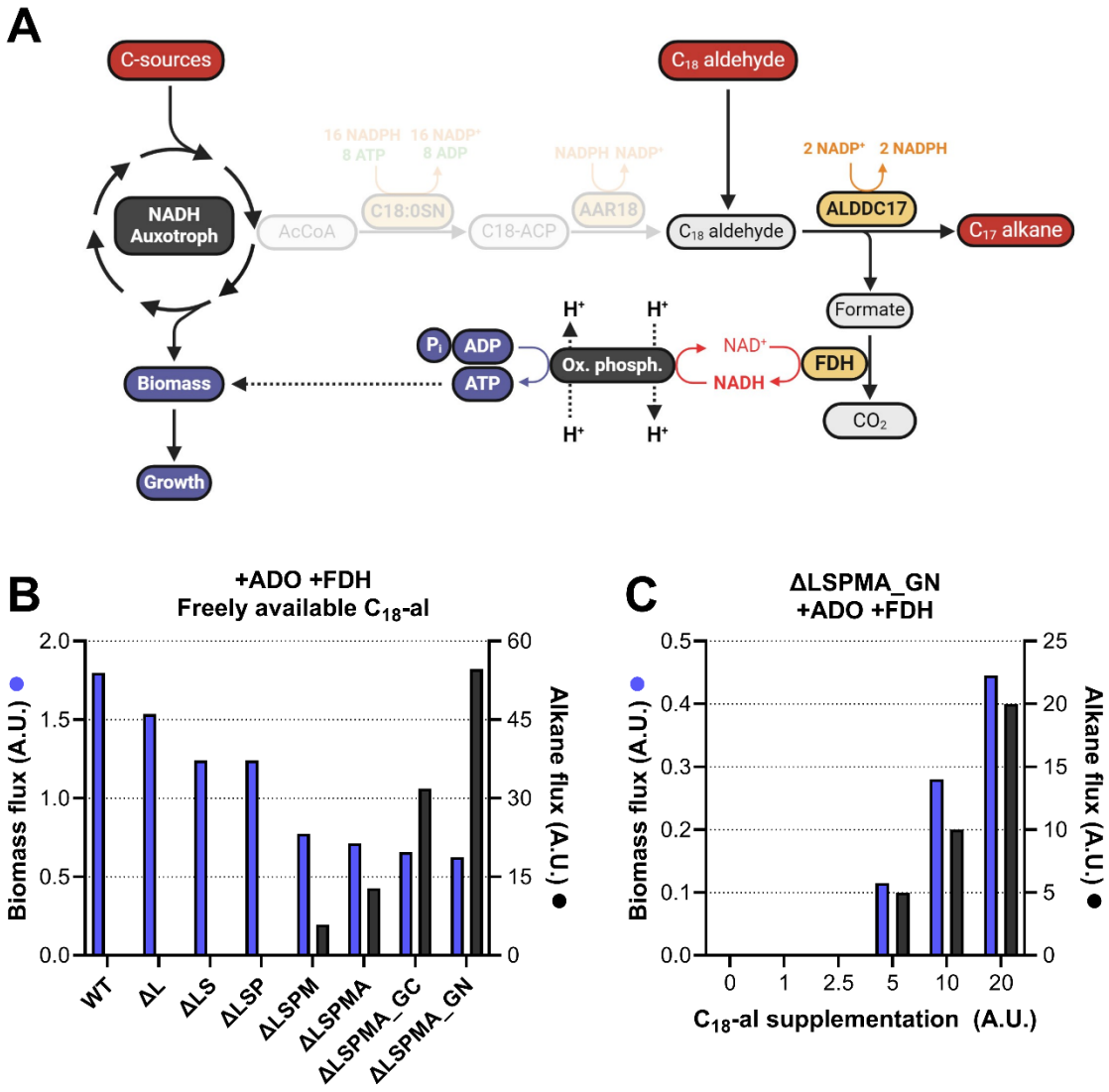


Figure 5.6. Biomass- and alkane fluxes of in silico mutants with FDH and ADO. **(A)** Schematic of the principle of ADO-mediated growth rescue via formate oxidation. The ADO enzyme (*in silico*: ALDDC17) generates formate through the 'feeding' of aldehyde substrate. Through the activity of formate dehydrogenase (FDH), NADH can be generated to drive oxidative phosphorylation leading to biomass formation (note that the diagram does not indicate the correct stoichiometry of ATP synthesis by NADH oxidation). **(B)** Biomass- and alkane fluxes of several *in silico* models run with 10 A.U. fructose, 20 A.U. acetate and 10 A.U. succinate, in addition to a freely available pool of C_{18} aldehyde. All models were equipped with ADO and FDH reactions, and the required aldehyde uptake reactions. **(C)** Biomass and alkane flux values using the ΔLSPMA_GN model run with the same carbon source settings as figure B, but with a defined supply of C_{18} aldehyde as indicated.

An important distinction between the two is the consumption of NADPH by ADO (ALDDC17) in the alkane-forming approach. It is possible that the removal of NADH-generation reactions and the introduction of GN, along with native NADPH-generating reactions, results in an excess of NADPH that requires balancing with sufficient levels of NADH to enable biomass

flux. Alternatively, NADPH can be consumed to reduce the NADPH/NADH level, which occurs through ALDDC17 activity. In general however, the aldehyde feeding approach seemed a viable strategy for growth-coupling ADO activity, and the stringency of this growth coupling is somewhat customizable through the choice of strain.

The second *in silico* approach we probed featured the endogenous biosynthesis of aldehydes via the fatty acid biosynthesis pathway and AAR activity, as opposed to their supplementation in the medium. Instead of using all models we focused solely on the ΔLSPMA_GC and GN variants due to their total NADH auxotrophy. Generating the substrate for ADO comes at a high cost in terms of carbon and reducing power; 9 acetyl-CoA, 16 NADPH and 8 ATP are required for the biosynthesis of C_{18}-acyl ACP alone (C18:0SN) (Figure 5.7A). Another NADPH is subsequently required for the generation of aldehydes via AAR (AAR18) and finally 2 NADPH are needed to generate alkanes and formate via ADO. This means that for each potential mol of NADH released from formate oxidation, 19 mol NADPH would be required.

After introducing the required reactions, both models were run using 10 A.U. fructose, 20 A.U. acetate and 10 A.U. succinate. However, the ΔLSPMA_GN model returned an infeasible solution with these conditions, and we found that supplying 50 A.U. succinate restored biomass flux. Using these settings, ΔLSPMA_GC attained a biomass flux around 0.2 A.U. while ΔLSPMA_GN was only able to achieve a fraction of this (0.002) (Figure 5.7B). Despite this large difference in biomass flux, the alkane fluxes of both models were strikingly similar. We found that the NADPH production profile of either strain displayed a core difference, namely that ΔLSPMA_GN was barely utilising GN to metabolise fructose (Figure 5.7C). Instead, all NADPH flux was generated from ME2 ($NADP^+$-dependent malic enzyme), explaining the need for succinate supplementation since this compound enters the TCA cycle making it available for the ME2 reaction. The clear cause of this is that GC permits more glycolytic ATP generation, whereas GN does not. The high levels of NADPH required to generate formate, NADH and finally ATP in this model causes an extreme constraint on ΔLSPMA_GN biomass flux, since all ATP must be generated from this highly NADPH-

consuming pathway first. This is reflected in the NADPH consumption profiles of both strains, which is overwhelmingly represented by the C180SN reaction, followed by ADO and AAR (Figure 5.7D). Only in ΔLSPMA_GC a small amount of NADPH is consumed as biomass flux.

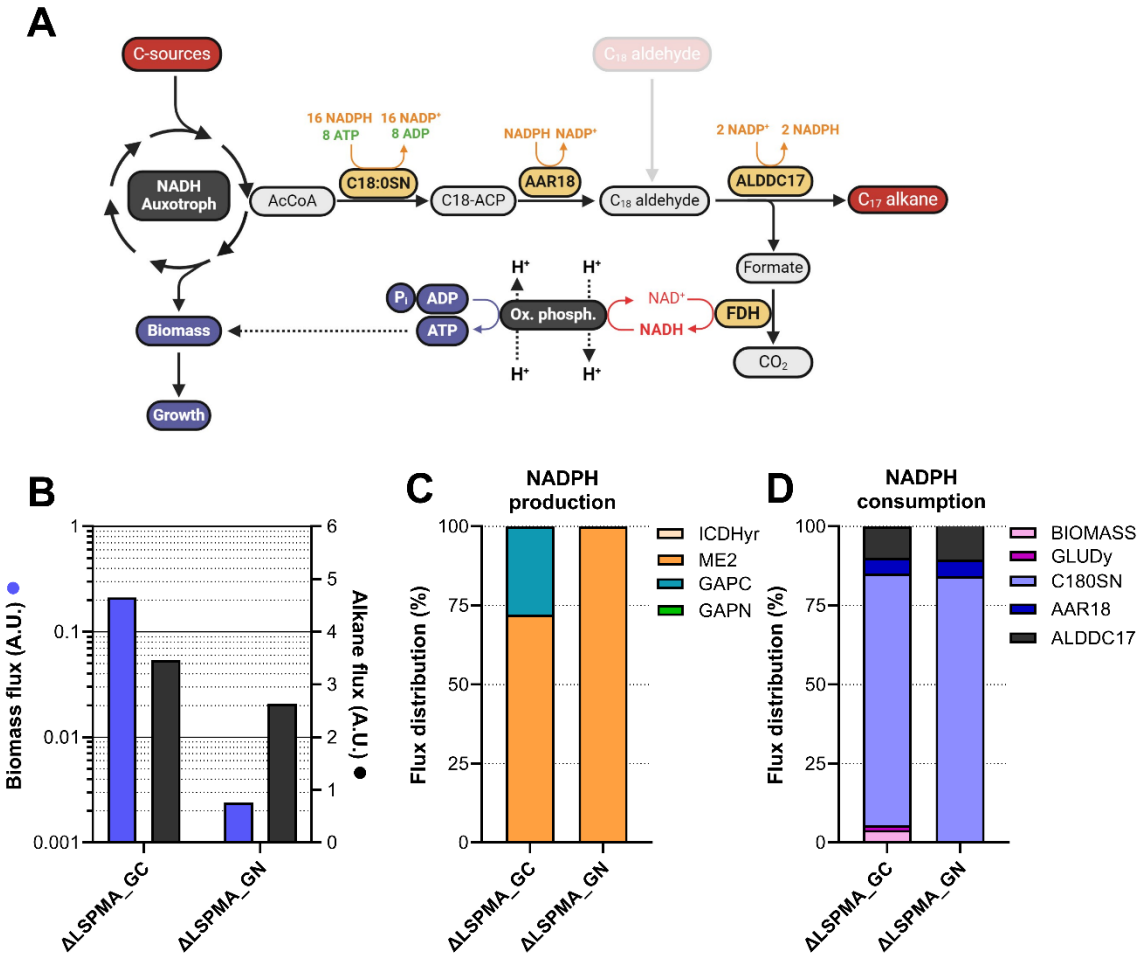


Figure 5.7. Self-sufficient *in silico* alkane and formate biosynthesis via fatty acid synthesis and AAR/ADO activity. (A) ADO-mediated growth rescue using formate oxidation, but instead of supplying aldehydes for uptake directly the model has the additional C180SN (C_{18} acyl-ACP synthesis) and AAR18 (C_{18} acyl-ACP reductase) reactions added, allowing alkane- and formate biosynthesis (note that the diagram does not indicate the correct stoichiometry of ATP synthesis by NADH oxidation). **(B)** Biomass- and alkane fluxes of ΔLSPMA_GC- and GN, using 10 A.U. fructose, 20 A.U. acetate and 50 A.U. succinate as available carbon sources. Besides the C180SN and AAR18 reactions, both models also had the ALDDC17 and FDH reactions added. Note the logarithmic scale for biomass flux. We assessed the flux distribution of reactions **(C)** producing and **(D)** consuming NADPH in both ΔLSPMA_GC- and GN metabolic models.

In summary, the FBA results indicated the theoretical viability of growth-coupling ADO using an NADH-auxotrophic strain. An important finding was that substrate level ATP generation largely determined the degree to which biomass formation was dependent on ADO flux. Indeed, it appeared that using a phosphorylating $NADP^+$-dependent GAPDH is an important requirement if the substrate for ADO were to be biosynthesised, although an aldehyde feeding approach could be exploited using either strain.

5.5.3 Growth-coupled alkane biosynthesis via the reductive glycine pathway

Formate has been intensively studied for its potential as a C_1 building block acting as an intermediate between CO_2, which is challenging to fix directly, and central microbial metabolism. In research pioneered by Bar-Even and coworkers, the reductive glycine pathway has been established to achieve formate incorporation into *E. coli* by linking its uptake to 5,10-methylene-THF (5,10-mTHF) biosynthesis, an essential cofactor acting as a methyl donor required for the biosynthesis of thymidine, methionine, purines, and CoA (Bar-Even, 2016; Bar-Even et al., 2013; Yishai et al., 2017). By knocking out pathways natively producing 5,10-mTHF, and introducing the required heterologous formate-fixing pathways, its assimilation could be made essential for growth. In previous work, a strain was developed displaying auxotrophy for these C_1 units and glycine ($C_1(G)_{aux}$) by the deletion of *glyA* and *gcvTHP,* knocking out the glycine cleavage system and through this the native *E. coli* 5,10-mTHF synthesis route (Yishai et al., 2017) (Figure 5.8A). In addition, the authors also described construction of a strain auxotrophic for 5,10-mTHF as well as glycine and serine ($C_1(G)S_{aux}$). This was achieved by deleting *gcvTHP* and *serA* to abolish serine biosynthesis via glycolysis while retaining *glyA* to allow formate-dependent serine biosynthesis via hydroxymethyltransferase activity (Figure 5.8B). Both $C_1(G)_{aux}$ and $C_1(G)S_{aux}$ are expected to require glycine supplementation for growth, although enzymes including *kbl* and *ltaE* may nevertheless contribute to (low flux) glycine biosynthesis (Yishai et al., 2017).

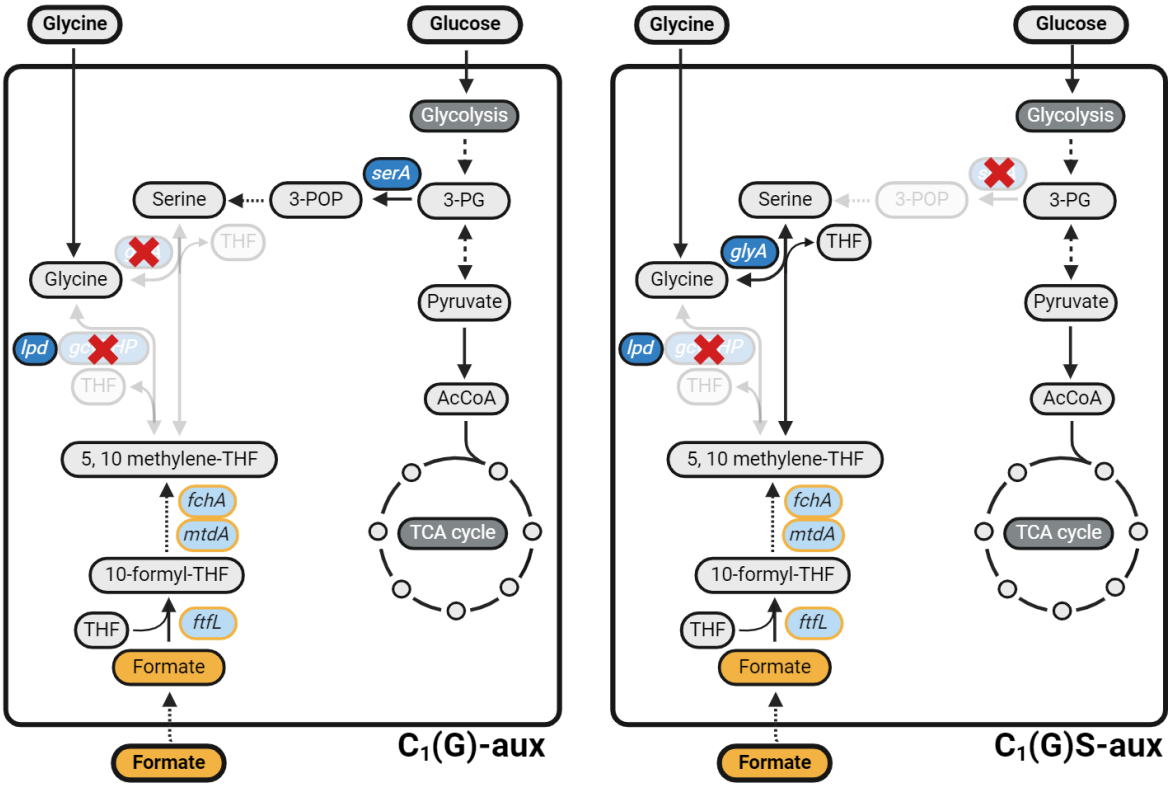


Figure 5.8. Schematic of the metabolic architecture of formate dependent $C_1(G)_{aux}$ and $C_1(G)S_{aux}$ strains. Dark blue boxes represent native *E. coli* genes, light blue boxes represent chromosomally integrated heterologous genes. In both strains, formate is assimilated by formate tetrahydrofolate ligase (*ftfL*), methylene tetrahydrofolate dehydrogenase (*mtdA*) and methenyl tetrahydrofolate cyclohydrolase (*fchA*), resulting in 5,10-mTHF formation. This cofactor is required for biosynthesis of e.g. methionine and thymidine. In $C_1(G)_{aux}$ (**A**), the glycine cleavage system (*gcvTHP* and *lpd*) and serine hydroxymethyltransferase (*glyA*) were inactivated to block 5,10-mTHF synthesis from glycine and/or serine. In $C_1(G)S_{aux}$ **(B)**, 3-phosphoglycerate dehydrogenase (*serA*) was deleted but *glyA* left intact, to enable 5,10-mTHF-dependent serine biosynthesis.

In unpublished work by Satanowski and coworkers, the gene insertions and knockouts required to establish the $C_1(G)_{aux}$ and $C_1(G)S_{aux}$ metabolic schemes were constructed in *E. coli* SIJ488 (Jensen et al., 2015) to enable more efficient gene editing. Specifically, this means that $C_1(G)_{aux}$ and $C_1(G)S_{aux}$ carried the Δ*gcvTHP* Δ*glyA* Δ*frmRAB* and Δ*gcvTHP* Δ*serA* Δ*frmRAB* deletions, respectively, as well as an integrated *ftfL-fchA-mtdA* cassette required for formate assimilation. These changes were identical to those previously engineered in *E. coli* MG1655 (Yishai et al., 2017), except they carried additional Δ*frmRAB* to avoid formaldehyde oxidation. These new *E. coli* SIJ488-based strains were kindly shared with us, and we first set

out to establish the formate growth-dependence of both $C_1(G)_{aux}$ and $C_1(G)S_{aux}$ to identify their relevant dynamic ranges.

Surprisingly, we found that $C_1(G)_{aux}$ displayed growth in conditions lacking formate (Figure 5.9A). After a 3-day lag phase we observed growth of $C_1(G)_{aux}$ in all conditions with glycine supplementation regardless of formate concentration, although strains supplemented with >100 µM formate displayed immediate growth and the supplementation of 1000 µM formate yielded the best growth performance. This conflicted with data from the original work describing these strains (Yishai et al., 2017), although we cannot rule out trace contamination of our medium with formate or cell debris. In contrast to $C_1(G)_{aux}$, strain $C_1(G)S_{aux}$ displayed somewhat more robust growth, as strains not supplemented with formate were unable to enter exponential growth (Figure 5.9B). A clear 'switch' in growth behaviour was visible between 100-150 µM formate supplementation. With 100 µM formate supplementation the maximum OD_{600} reached around 0.1, while this value was 0.8 with 150 µM formate supplementation (Figure 5.9B and D). The same 'switch' was measured in the strain's maximum growth rates. Supplementation of 100 µM formate led to μ_{max} values of 0.017 h^{-1}, which was in the same range as strains supplemented with 0-75 µM formate. Between 100-1000 µM formate the maximum growth rates increased ten-fold to around 0.19 h^{-1}.

Regardless of the strain's response to high formate concentrations, more importantly the low- and no-formate controls behaved as expected, displaying virtually no growth compared to strains sufficiently supplemented with formate over the course of 96 hours. The clear response of μ_{max} values to formate concentrations of 150-1000 µM was encouraging in terms of the potential of $C_1(G)S_{aux}$ to be used in growth-based selection campaigns. Glycine concentrations were also not limiting to growth, as supplementation with 4-, 8- or 16 mM did not result in improvements in growth (Figure S5.1A). We therefore decided to use $C_1(G)S_{aux}$ as the selection strain for proof-of-principle assays with an alkane biosynthesis pathway.

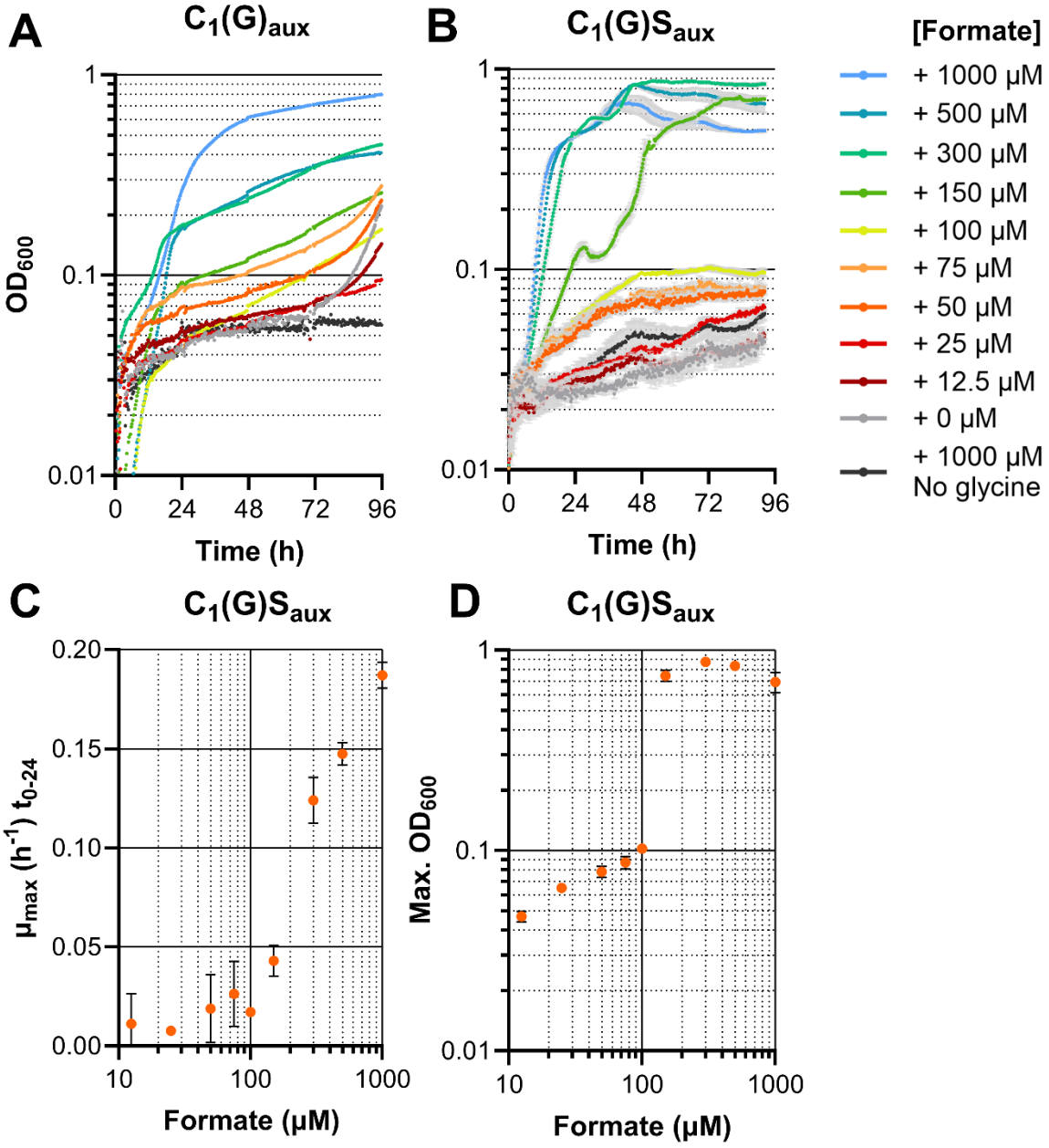


Figure 5.9. Growth dependence of $C_1(G)_{aux}$ and $C_1(G)S_{aux}$ on formate. Both strains were grown in M9 minimal medium with 50 mM glucose and 2 mM glycine unless otherwise indicated. **(A)** $C_1(G)_{aux}$ and **(B)** $C_1(G)S_{aux}$ growth as a function of formate concentration. **(C)** Maximum growth rates achieved within the first 24 hours of $C_1(G)S_{aux}$ cultivation as a function of formate concentration. **(D)** Maximum observed OD_{600} of $C_1(G)S_{aux}$ as a function of formate concentration. All datapoints represent the averages of biological triplicate measurements, and error bars represented the standard deviation of these averages. Error bars were omitted in figure A for visual clarity.

In Figure 5.10, two different approaches to growth-coupling ADO activity are presented. The first consists of direct feeding of aldehydes into the culture broth of ADO-expressing cells, in theory resulting in aldehyde deformylation and concomitant growth through formate

assimilation. The second approach relies on the use of endogenous fatty acid synthesis (FAS) to generate acyl-ACP as a substrate for AAR which, in turn, generates aldehydes. These can subsequently be deformylated by ADO, generating alkanes and providing a source of 5,10-mTHF via the formate byproduct.

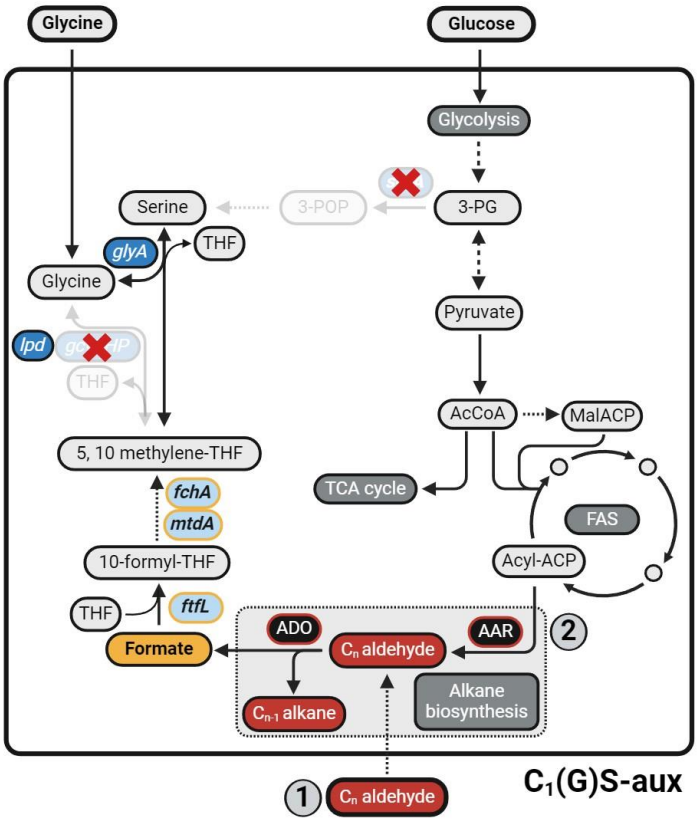


Figure 5.10. Proposed metabolic architecture for growth-coupling (AAR-)ADO activity via the reductive glycine pathway. Dark blue boxes represent native *E. coli* genes, light blue boxes represent chromosomally integrated heterologous genes, black boxes represent alkane biosynthesis enzymes expressed from a plasmid. Strain C_1(G)S-aux requires formate for 5,10-mTHF synthesis, providing C_1 units for methionine, thymidine, serine, purines, and CoA biosynthesis. Besides direct supplementation, formate availability can be achieved through ADO activity acting on (**1**) exogenously supplied aldehydes or (**2**) aldehydes produced endogenously by fatty acid synthesis (FAS) and AAR activity.

We first attempted the aldehyde feeding approach as this would grant control over the substrate carbon chain length and the substrate concentration. However, supplementation of 1 mM C_{16} or C_8 aldehyde substrates did not result in growth of C_1(G)S_{aux} irrespective of ADO expression construct (low/medium/high copy vector) or expression strength (data not shown).

However, this approach was attempted without additional supplementation of formate, and it was possible that any formate generated by ADO was insufficient to enter the required dynamic range as established in Figure 5.9. The uptake of aldehydes may also have been limiting, although the results in Figure 5.2 suggested that *E. coli* was able to take up aldehydes supplied exogenously.

We therefore turned to using AAR-mediated aldehyde biosynthesis. This enzyme produces aldehydes intracellularly from native fatty acyl-ACPs and there is evidence to suggest it interacts directly with ADO to shuttle aldehydes into ADO's enzyme access tunnel, which may improve aldehyde availability for ADO compared to feeding aldehydes exogenously (Chang et al., 2020; Gao et al., 2020). In Chapter 4 we investigated alkane biosynthesis in *E. coli* using AAR and ADO, leading to the establishment of a functional alkane biosynthesis pathway. The $C_1(G)S_{aux}$ strain was therefore transformed with either an empty plasmid (DVA_pBR322_AF) or the IPTG-inducible alkane biosynthesis plasmid (pL2_JRN072_AF), and we used the same optimised conditions identified from our alkane biosynthesis work (see Chapter 4). This included using minimal medium with 100 mM glucose supplemented with 250 mM KCl and 2 mM glycine to ensure growth of $C_1(G)S_{aux}$. Supplementation of 250 mM KCl was not detrimental to growth of $C_1(G)S_{aux}$ (figure S5.1B).

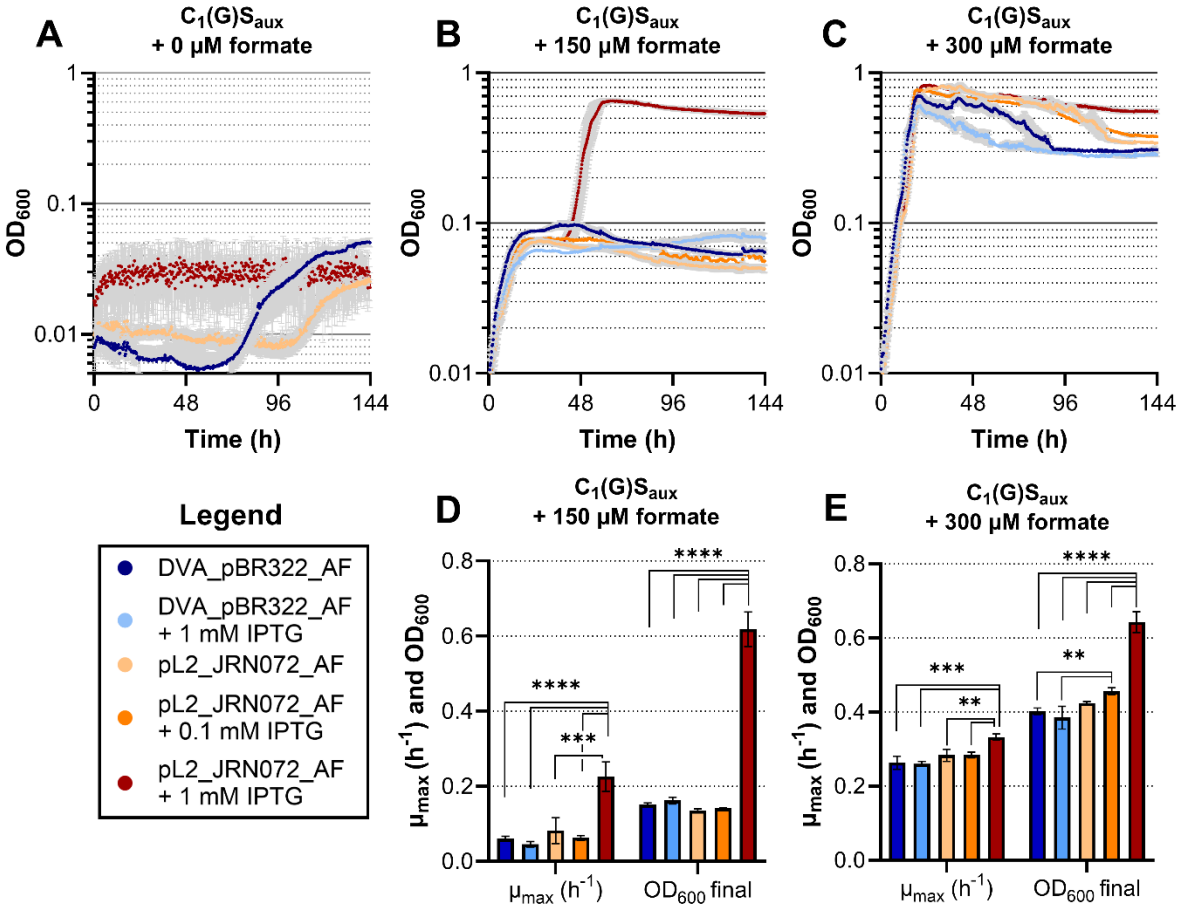


Figure 5.11. Formate-based growth coupling of the AAR-ADO alkane biosynthesis pathway. The formate-dependent $C_1(G)S_{aux}$ strain was transformed with either an empty vector or pL2_JRN072_AF encoding an IPTG-inducible AAR-ADO pathway. Strains were cultivated in M9 minimal medium supplemented with 100 mM glucose, 2 mM glycine, 250 mM KCl and either (**A**) no formate, (**B**) 150 µM formate or (**C**) 300 µM formate. (**D**) Growth of $C_1(G)S_{aux}$ with 150 µM supplemented formate and a highly induced alkane biosynthesis plasmid displayed significantly higher maximum growth rates and OD_{600} values compared to the other conditions. (**E**) The supplementation of 300 µM formate and strong induction of AAR-ADO resulted in a higher final OD_{600} value compared to other conditions, and significantly higher growth rates compared to empty vector controls, but not compared to other AAR-ADO plasmid conditions. All data were averages derived from biological triplicate measurements. Error bars represented the standard deviation of this average. Statistical analysis (not explicitly visualised for each result) was performed by comparing means using a one-way ANOVA and post-hoc Tukey's multiple comparison test. P-value significance indicated are as follows: ns (p-value >0.05); * (p-value <0.05), ** (p-value <0.01), *** (p-value <0.001) or **** (p-value <0.0001).

Growth of $C_1(G)S_{aux}$ within the first 48h was only observed with supplemental formate (Figure 5.11A-C), although we did observe slow growth of $C_1(G)S_{aux}$ in conditions lacking formate between 72-144 hours (Figure 5.11A). This could be caused by trace formate in the medium, for example by insufficient washing, although we note that the achieved OD_{600} values were very low (0.02-0.05). Inducing the expression of AAR-ADO pathway genes without

supplemental formate did not result in growth (Figure 5.11A). In conditions with 150 µM supplemented formate the cultures reached OD_{600} values of around 0.08-0.10 before stagnating (Figure 5.11B). However, $C_1(G)S_{aux}$ pL2_JRN072_AF was able to enter a secondary phase of exponential growth only when induced with 1 mM IPTG, reaching a maximum OD_{600} of around 0.6 within ~48h. This improved growth was also reflected in the achieved maximum growth rate over the course of the assay, which exceeded all other tested strains in this condition (Figure 5.11D). This was strong evidence that the activity of the alkane biosynthesis pathway was responsible for growth beyond the initial growth phase.

Supplementing the $C_1(G)S_{aux}$ strain with 300 µM formate resulted in a higher basal level of growth in general, with each condition leading to a maximum OD_{600} of ~0.6-0.8. Over the course of the assay we observed a decline in OD_{600} amongst all strains, which is often associated with cell lysis and therefore cell viability. After 144 hours of cultivation, we measured a significantly higher OD_{600} value for $C_1(G)S_{aux}$ pL2_JRN072_AF under high induction as opposed to all other tested strains, and significant differences between weakly induced $C_1(G)S_{aux}$ pL2_JRN072_AF and both empty vector controls. Strongly induced $C_1(G)S_{aux}$ pL2_JRN072_AF also achieved maximum growth rates significantly higher than the empty vector controls, as well as the non- and weakly induced $C_1(G)S_{aux}$ pL2_JRN072_AF (Figure 5.11E).

Taken together, the data implied that (strong) expression of the AAR-ADO enzymes in $C_1(G)S_{aux}$ significantly improved its fitness in selective conditions compared to those carrying an empty- or non-induced plasmid. These results therefore suggested that growth had indeed been successfully coupled to alkane biosynthesis pathway activity, most notably when supplemental formate was low (150 µM).

5.6 Discussion

Growth-coupling techniques are used to link the biosynthesis of a target compound to cell fitness. From a bioproduction perspective this offers a way to ensure product synthesis at defined minimum yields and/or rates, and from an enzyme engineering angle growth-coupling offers an attractive low-cost method to select for a certain enzymatic activity. Products such as hydrocarbons are challenging to growth-couple due some of their inherent properties, therefore in this chapter we described several potential approaches to couple microbial growth to this important class of molecules.

Firstly, manipulation of the $NADP^+$/NADPH redox cofactor ratio was explored as a method to growth-couple hydrocarbon synthesis. To do so, we generated ΔPS that displayed a deficiency in its ability to regenerate $NADP^+$ when grown on glucose. Using this strain, we were able to partially couple growth to aldehyde reductase activity and fatty acid biosynthesis activity. Although bioconversions of carboxylic acids to alcohols had been performed before using this strain (Kramer et al., 2020), this is the first report demonstrating that stimulating fatty acid biosynthesis in an NADPH overproducing strain can result in weak growth coupling. We note that in other work, the additional deletions of *edd, eda,* and *qor* (an NADPH sink) were implemented to ensure all carbon flux was funneled through the pentose phosphate pathway and $NADP^+$ could not be regenerated (Maxel, Aspacio, et al., 2020; Maxel, King, et al., 2020; Maxel, Zhang, et al., 2020). Besides the removal of *edd* and *eda*, the replacement of *gapA* with an NADPH-dependent version such as *gapN* could further enforce the cycling of NADPH through endogenous or exogenous reactions, since a proportion of carbon exiting the pentose phosphate pathway still releases reducing power via *gapA* (L. Zhang et al., 2018).

One limitation of this approach is that it is not specific for the production of a certain hydrocarbon from a sole carbon source such as glucose. The *in silico* results indicated that the NADPH-consuming AAR/ADO enzymes could be partially coupled to growth by using an $NADP^+$-deficient strain such as ΔPS, as we observed alkane producing flux in models run with glucose as input carbon source. However, the high NADPH consumption due to fatty acid

synthesis meant that the terminal AAR/ADO enzymes contributed relatively little to the overall consumption of NADPH. The FBA data provided evidence for this, since only a small proportion of consumed NADPH could be attributed to AAR/ADO activity. The selective pressure for any of these terminal enzymes to catalyse its reaction would therefore be relatively weak, which is undesirable if enzyme engineering is the aim. Alternatively, the substrate for the terminal enzyme of interest can also be added directly to the medium, which has been used before to engineer a carboxylic acid reductase enzyme using an $NADP^+$ deficient strain (Kramer et al., 2020). Although this increases the selection pressure acting on the target enzyme, this raises other challenges such as ensuring sufficient substrate uptake. In addition, the NADPH-dependent degradation of the supplemented substrate by competing endogenous enzymes would have to be avoided. For compounds such as aldehydes -the substrate for ADO- this may lead to growth rescue independent of ADO activity, as we have demonstrated in this work.

To circumvent this issue of low product specificity using $NADP^+$ auxotrophy, we next focused on the formate-producing property of ADO. We hypothesised that any selection for formate could be used to specifically select for alkane biosynthesis by ADO. In the context of *in vivo* directed evolution, using formate as the selective metabolite yields several benefits over direct alkane quantitation or detection. Firstly, formate is relatively non-toxic and highly soluble. Secondly, since formate is a terminal co-product of the AAR-ADO pathway, it is possible to establish selection for alkane production from any substrate and for the entire metabolic pathway. Thirdly, in this approach formate is consumed while alkanes can accumulate freely, reducing the chance of selection escape due to extracellular leakage. Formate is also not normally produced in aerobically growing *E. coli*, further reducing the chance of false positives.

Formate is often used as an electron donor for the reduction of NAD^+ or $NADP^+$ through expression of formate dehydrogenases. Since ADO, AAR, and fatty acid biosynthesis all require NADPH input, we could not growth-couple ADO-activity by linking its formate production to $NADP^+$ reduction. Instead, we designed a strain auxotrophic for NADH but

autotrophic for NADPH to ensure selection would act only on NADH-producing pathways. Coupling enzyme activity to ATP availability, as is the case when limiting NADH generation and therefore respiration, is known to be a stringent growth coupling strategy (Alter & Ebert, 2019). The *in silico* results demonstrated that this approach is theoretically viable, as demonstrated by a) the uptake and deformylation of freely available aldehydes to maximise biomass flux and b) the self-sufficient biosynthesis of aldehydes to maximise biomass flux. Both approaches also demonstrated that substrate-level ATP generation was a core modulator of the coupling strength, as ΔLSPMA_GC models achieved higher biomass fluxes with a lower ADO flux requirement compared to when ΔLSPMA was equipped with non-phosphorylating GN. In fact, using respiration-deficient strains to select for alkane biosynthesis with ATP availability as the selective principle has been suggested before (Sheppard et al., 2016). The construction and characterization of both ΔLSPMA_GC and ΔLSPMA_GN strains is elaborated on at length in Chapter 6. To some extent, the experimental characterization of these strains in Chapter 6 echoes some of the FBA results we gathered. For example, growth of ΔLSPMA_GN was hindered by a high NADPH/$NADP^+$ ratio, requiring either an NADPH sink or a high flux NADH-generating pathway to restore growth (see Chapter 6). Due to time restraints, we were unable to test NADH-linked ADO growth-coupling, but the results from this work suggested that it may be worthwhile.

Finally, we designed an approach similarly based on formate, but rather than using formate as an energy carrier we sought to use it as an essential C_1 donor through the reductive glycine pathway. This approach relies on formate being made essential for the formation of the 5,10-mTHF cofactor, which requires the introduction of formate-fixing reactions and the removal of endogenous 5,10-mTHF-forming pathways. The involved metabolic principles and the characterization of these strains has been described in earlier work (Yishai et al., 2017). Our approach involved repurposing these strains as selection hosts for ADO activity. We found that the $C_1(G)S_{aux}$ strain, carrying deletions in *serA* and *gcvTHP*, displayed a suitable dynamic range for testing ADO-mediated formate biosynthesis. The supplementation of aldehydes

directly in ADO-expressing cultures did not result in growth. Besides the activity of ADO simply being too low, another reason could be the poor intracellular availability of long-chain aldehydes when supplied exogenously. Other work has shown that exogenously added aldehydes can be converted to alkanes using ADO-expressing *E. coli* (Shakeel et al., 2020), suggesting that either the strain's genetic background, growth conditions, enzyme expression conditions or other miscellaneous differences could be limiting in our work. One considerable issue for direct aldehyde supplementation is the unwanted activity of aldehyde reductases, which have been shown to effectively convert aldehydes to alcohols in several works (Fatma et al., 2016; Rodriguez & Atsumi, 2014).

The results obtained using an AAR-ADO alkane biosynthesis pathway as optimised in Chapter 4 were more encouraging. In this approach, aldehydes were biosynthesised *in vivo* through the native fatty acid biosynthesis pathway and AAR activity. The $C_1(G)S_{aux}$ strain supplemented with 150 µM formate displayed significantly improved growth when the AAR-ADO pathway was expressed, leading to higher final OD_{600} and μ_{max} values compared to an uninduced control. When supplemented with 300 µM formate, $C_1(G)S_{aux}$ grew fully without induction of AAR-ADO, but the subsequent reduction in OD_{600} due to cell death was lower in cells with a strongly induced AAR-ADO pathway. In addition, maximum growth rates of $C_1(G)S_{aux}$ expressing AAR-ADO were also higher than the empty vector control when supplied 300 µM formate. These results indicated that a fitness benefit was conferred by the expression of the AAR-ADO pathway in the formate-dependent $C_1(G)S_{aux}$ strain. A necessary next step would be to measure alkane production using analytical techniques to confirm the pathway's activity, and to assess growth- and production behaviour in conditions with only AAR or ADO present (i.e. an incomplete pathway). These assays could definitively demonstrate that the activity of AAR-ADO together results in growth-rescue of $C_1(G)S_{aux}$. Other verification methods could include C^{13} labelling of the supplemented formate, to demonstrate that the metabolites formed from formate assimilation (all 5,10-mTHF derivatives) are mixed in terms of carbon

origin – i.e. the metabolite pool would consist of a mix of C^{13} labelled supplemented formate and ADO-produced C^{12} formate, but only when AAR-ADO expression is induced.

The growth-coupling of $C_1(G)S_{aux}$ to AAR-ADO activity represents, to our knowledge, the first system where microbial growth has been directly metabolically coupled to product synthesis by ADO, and therefore to alkane biosynthesis. Considering the level of (commercial) attention that alkane production via the AAR-ADO pathway has attracted, and the high global demand for biofuel production in general, this work warrants further study especially in the context of enzyme engineering. Some of the engineering work that could be undertaken using a growth-coupled system includes optimizing expression of AAR/ADO and electron transfer proteins, screening for optimal production culture conditions, and of course directed evolution of AAR and/or ADO to improve stability and catalytic efficiency. The fact that $C_1(G)S_{aux}$ expressing AAR-ADO still required supplemental formate to grow indicated that the dynamic range was appropriate for selection. Formate supplementation can also be dynamically reduced as required to ensure selective pressure for AAR-ADO flux. In addition, enzyme expression levels of AAR-ADO can be reduced to avoid their activity from saturating growth. Finally, the expression of formate dehydrogenases could be used to continuously consume formate produced by AAR-ADO to maintain a high selective pressure for formate production. Further work using this AAR-ADO selection system could benefit considerably from the use of turbidostats combined with *in vivo* continuous evolution systems, such as eMutaT7 or EvolvR, for high throughput evolution of this important alkane biosynthesis pathway.

5.7 Conclusion

The AAR-ADO pathway represents one of the most comprehensively studied routes towards drop-in biofuel production. However, despite continued efforts, the activity of ADO has not been improved by the levels required, pointing to a lack of our understanding of the enzyme's sequence-structure-function relationship. The challenge of growth-coupling alkane biosynthesis was approached from several angles in this work, using a combination of theoretical and experimental data. Two methods exploiting the formate-producing property of

ADO were found to be viable approaches for stringently coupling alkane biosynthesis to *E. coli* growth. Further implementation of this work could enable growth-coupled engineering of microbial alkane biosynthesis, providing a tool to help solve several of the long-standing issues in this field.

Acknowledgements

I would like to thank Ari Satanowski for providing the formate-sensitive strains and for insightful discussions over the course of our collaboration. This work was supported by the BBSRC (grant BB/M011224/1) and the Oxford Interdisciplinary Bioscience Doctoral Training Partnership.

5.8 References

Alter, T. B., & Ebert, B. E. (2019). Determination of growth-coupling strategies and their underlying principles. *BMC Bioinformatics 2019 20:1*, *20*(1), 1–17. https://doi.org/10.1186/S12859-019-2946-7

Arnold, F. H. (1998). When blind is better: Protein design by evolution. *Nature Biotechnology* 1998 16:7, 16(7), 617–618. https://doi.org/10.1038/nbt0798-617

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, *2*, 2006.0008. https://doi.org/10.1038/MSB4100050

Bao, L., Li, J. J., Jia, C., Li, M., & Lu, X. (2016). Structure-oriented substrate specificity engineering of aldehyde-deformylating oxygenase towards aldehydes carbon chain length. *Biotechnology for Biofuels*, *9*(1), 185. https://doi.org/10.1186/s13068-016-0596-9

Bar-Even, A. (2016). Formate Assimilation: The Metabolic Architecture of Natural and Synthetic Pathways. *Biochemistry*, *55*(28), 3851–3863. https://doi.org/10.1021/ACS.BIOCHEM.6B00495

Bar-Even, A., Noor, E., Flamholz, A., & Milo, R. (2013). Design and analysis of metabolic pathways supporting formatotrophic growth for electricity-dependent cultivation of microbes. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1827*(8–9), 1039–1047. https://doi.org/10.1016/J.BBABIO.2012.10.013

Bergler, H., Fuchsbichler, S., Högenauer, G., & Turnowsky, F. (1996). The Enoyl-[Acyl-Carrier-Protein] Reductase (FabI) of Escherichia coli, which Catalyzes a Key Regulatory Step in Fatty Acid Biosynthesis, Accepts NADH and NADPH as Cofactors and is Inhibited by Palmitoyl-CoA. *European Journal of Biochemistry*, *242*(3), 689–694. https://doi.org/10.1111/J.1432-1033.1996.0689R.X

Chang, M., Shimba, K., Hayashi, Y., & Arai, M. (2020). Electrostatic interactions at the interface of two enzymes are essential for two-step alkane biosynthesis in cyanobacteria. *Bioscience, Biotechnology, and Biochemistry*, *84*(2), 228–237. https://doi.org/10.1080/09168451.2019.1677142

Chen, D., Xu, S., Li, S., Tao, S., Li, L., Chen, S., & Wu, L. (2023). Directly Evolved AlkS-Based Biosensor Platform for Monitoring and High-Throughput Screening of Alkane Production. *ACS Synthetic Biology*, *12*(3), 832–841. https://doi.org/10.1021/ACSSYNBIO.2C00620

Chen, J., Wang, Y., Zheng, P., & Sun, J. (2022). Engineering synthetic auxotrophs for growth-coupled directed protein evolution. *Trends in Biotechnology*. https://doi.org/10.1016/J.TIBTECH.2022.01.010

Claassens, N. J., Satanowski, A., Bysani, V. R., Dronsella, B., Orsi, E., Rainaldi, V., Yilmaz, S., Wenk, S., & Lindner, S. N. (2022). Engineering the Reductive Glycine Pathway: A Promising Synthetic Metabolism Approach for C1-Assimilation. *Advances in Biochemical Engineering/Biotechnology*, *180*, 299–350. https://doi.org/10.1007/10_2021_181

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

Ebrahim, A., Lerman, J. A., Palsson, B. O., & Hyduke, D. R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology 2013 7:1*, *7*(1), 1–6. https://doi.org/10.1186/1752-0509-7-74

Fatma, Z., Jawed, K., Mattam, A. J., & Yazdani, S. S. (2016). Identification of long chain specific aldehyde reductase and its use in enhanced fatty alcohol production in *E. coli*. *Metabolic Engineering*, *37*, 35–45. https://doi.org/10.1016/J.YMBEN.2016.04.003

Gao, Y., Zhang, H., Fan, M., Jia, C., Shi, L., Pan, X., Cao, P., Zhao, X., Chang, W., & Li, M. (2020). Structural insights into catalytic mechanism and product delivery of cyanobacterial acyl-acyl carrier protein reductase. *Nature Communications 2020 11:1*, *11*(1), 1–11. https://doi.org/10.1038/s41467-020-15268-y

Green, L. (2021). Directed Evolution and Synthetic Biology for the Production of Next Generation Biofuels — Research Explorer The University of Manchester [The University of Manchester]. In *PhD thesis*. https://research.manchester.ac.uk/en/studentTheses/directed-evolution-and-synthetic-biology-for-the-production-of-ne

Janßen, H. J., & Steinbüchel, A. (2014). Fatty acid synthesis in Escherichia coli and its applications towards the production of fatty acid based biofuels. *Biotechnology for Biofuels 2014 7:1*, *7*(1), 1–26. https://doi.org/10.1186/1754-6834-7-7

Jensen, S. I., Lennen, R. M., Herrgård, M. J., & 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), 17874. https://doi.org/10.1038/srep17874

Khara, B., Menon, N., Levy, C., Mansell, D., Das, D., Marsh, E. N. G., Leys, D., & Scrutton, N. S. (2013). Production of Propane and Other Short-Chain Alkanes by Structure-Based Engineering of Ligand Specificity in Aldehyde-Deformylating Oxygenase. *ChemBioChem*, *14*(10), 1204–1208. https://doi.org/10.1002/cbic.201300307

Kim, S., Lindner, S. N., Aslan, S., Yishai, O., Wenk, S., Schann, K., & Bar-Even, A. (2020). Growth of *E. coli* on formate and methanol via the reductive glycine pathway. *Nature Chemical Biology 2020 16:5*, *16*(5), 538–545. https://doi.org/10.1038/s41589-020-0473-5

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O., & Lewis, N. E. (2016). BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, *44*(D1), D515–D522. https://doi.org/10.1093/NAR/GKV1049

Kramer, L., Le, X., Rodriguez, M., Wilson, M. A., Guo, J., & Niu, W. (2020). Engineering Carboxylic Acid Reductase (CAR) through a Whole-Cell Growth-Coupled NADPH Recycling Strategy. *ACS Synthetic Biology*, *9*(7), 1632–1637. https://doi.org/10.1021/ACSSYNBIO.0C00290

Kudo, H., Hayashi, Y., & Arai, M. (2019). Identification of non-conserved residues essential for improving the hydrocarbon-producing activity of cyanobacterial aldehyde-deformylating oxygenase. *Biotechnology for Biofuels*, *12*(1), 89. https://doi.org/10.1186/s13068-019-1409-8

Li, N., Chang, W. C., Warui, D. M., Booker, S. J., Krebs, C., & Bollinger, J. M. (2012). Evidence for only oxygenative cleavage of aldehydes to alk(a/e)nes and formate by cyanobacterial aldehyde decarbonylases. *Biochemistry*, *51*(40), 7908–7916. https://doi.org/10.1021/bi300912n

Liu, Y., Chen, J., Khusnutdinova, A. N., Correia, K., Diep, P., Batyrova, K. A., Nemr, K., Flick, R., Stogios, P., Yakunin, A. F., & Mahadevan, R. (2020). A novel C-terminal degron identified in bacterial aldehyde decarbonylases using directed evolution. *Biotechnology for Biofuels*, *13*(1), 114. https://doi.org/10.1186/s13068-020-01753-5

Marsh, E. N. G., & Waugh, M. W. (2013). Aldehyde decarbonylases: Enigmatic enzymes of hydrocarbon biosynthesis. In *ACS Catalysis* (Vol. 3, Issue 11, pp. 2515–2521). American Chemical Society. https://doi.org/10.1021/cs400637t

Maxel, S., Aspacio, D., King, E., Zhang, L., Acosta, A. P., & Li, H. (2020). A Growth-Based, High-Throughput Selection Platform Enables Remodeling of 4-Hydroxybenzoate Hydroxylase Active Site. *ACS Catalysis*, *10*(12), 6969–6974. https://doi.org/10.1021/ACSCATAL.0C01892

Maxel, S., King, E., Zhang, Y., Luo, R., & Li, H. (2020). Leveraging Oxidative Stress to Regulate Redox Balance-Based, In Vivo Growth Selections for Oxygenase Engineering. *Cite This: ACS Synth. Biol*, *2020*, 3133. https://doi.org/10.1021/acssynbio.0c00380

Maxel, S., Saleh, S., King, E., Aspacio, D., Zhang, L., Luo, R., & Li, H. (2021). Growth-Based, High-Throughput Selection for NADH Preference in an Oxygen-Dependent Biocatalyst. *ACS Synthetic Biology*, *10*. https://doi.org/10.1021/ACSSYNBIO.1C00258

Maxel, S., Zhang, L., King, E., Acosta, A. P., Luo, R., & Li, H. (2020). In Vivo, High-Throughput Selection of Thermostable Cyclohexanone Monooxygenase (CHMO). *Catalysts*, *10*(8), 935. https://doi.org/10.3390/catal10080935

Nielsen, J. R., Weusthuis, R. A., & Huang, W. E. (2023). Growth-coupled enzyme engineering through manipulation of redox cofactor regeneration. *Biotechnology Advances*, 108102. https://doi.org/10.1016/J.BIOTECHADV.2023.108102

Orsi, E., Claassens, N. J., Nikel, P. I., & Lindner, S. N. (2021). Growth-coupled selection of synthetic modules to accelerate cell factory development. *Nature Communications 2021 12:1*, *12*(1), 1–5. https://doi.org/10.1038/s41467-021-25665-6

Orth, J. D., Fleming, R. M. T., & Palsson, B. Ø. (2010). Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide . *EcoSal Plus*, *4*(1). https://doi.org/10.1128/ECOSALPLUS.10.2.1

Rodriguez, G. M., & Atsumi, S. (2014). Toward aldehyde and alkane production by removing aldehyde reductase activity in Escherichia coli. *Metabolic Engineering*, *25*, 227–237. https://doi.org/10.1016/j.ymben.2014.07.012

Sauer, U., Canonaco, F., Heri, S., Perrenoud, A., & Fischer, E. (2004). The Soluble and Membrane-bound Transhydrogenases UdhA and PntAB Have Divergent Functions in NADPH Metabolism of Escherichia coli*. *Journal of Biological Chemistry*, *279*(8), 6613–6619. https://doi.org/10.1074/JBC.M311657200

Schirmer, A., Rude, M. A., Li, X., Popova, E., & Del Cardayre, S. B. (2010). Microbial biosynthesis of alkanes. *Science*, *329*(5991), 559–562. https://doi.org/10.1126/science.1187936

Shakeel, T., Fatma, Z., & Yazdani, S. S. (2020). In vivo Quantification of Alkanes in Escherichia coli. *BIO-PROTOCOL*, *10*(8). https://doi.org/10.21769/bioprotoc.3593

Shakeel, T., Gupta, M., Fatma, Z., Kumar, R., Kumar, R., Singh, R., Sharma, M., Jade, D., Gupta, D., Fatma, T., & Yazdani, S. S. (2018). A consensus-guided approach yields a heat-stable alkaneproducing enzyme and identifies residues promoting thermostability. *Journal of Biological Chemistry*, *293*(24), 9148–9161. https://doi.org/10.1074/jbc.RA117.000639

Sheppard, M. J., Kunjapur, A. M., & Prather, K. L. J. (2016). Modular and selective biosynthesis of gasoline-range alkanes. *Metabolic Engineering*, *33*, 28–40. https://doi.org/10.1016/j.ymben.2015.10.010

Wang, Q., Bao, L., Jia, C., Li, M., Li, J. J., & Lu, X. (2017). Identification of residues important for the activity of aldehyde-deformylating oxygenase through investigation into the structure-activity relationship. *BMC Biotechnology*, *17*(1), 1–9. https://doi.org/10.1186/s12896-017-0351-8

Warui, D. M., Li, N., Nørgaard, H., Krebs, C., Bollinger, J. M., & Booker, S. J. (2011). Detection of formate, rather than carbon monoxide, as the stoichiometric coproduct in conversion of fatty aldehydes to alkanes by a cyanobacterial aldehyde decarbonylase. *Journal of the American Chemical Society*, *133*(10), 3316–3319. https://doi.org/10.1021/ja111607x

Wenk, S., Schann, K., He, H., Rainaldi, V., Kim, S., Lindner, S. N., & Bar-Even, A. (2020). An “energy-auxotroph” *Escherichia coli* provides an in vivo platform for assessing NADH regeneration systems. *Biotechnology and Bioengineering*, *117*(11), 3422–3434. https://doi.org/10.1002/bit.27490

Yishai, O., Goldbach, L., Tenenboim, H., Lindner, S. N., & Bar-Even, A. (2017). Engineered Assimilation of Exogenous and Endogenous Formate in Escherichia coli. *ACS Synthetic Biology*, *6*(9), 1722–1731. https://doi.org/10.1021/ACSSYNBIO.7B00086

Zhang, J., Lu, X., & Li, J. J. (2013). Conversion of fatty aldehydes into alk (a/e)nes by in vitro reconstituted cyanobacterial aldehyde-deformylating oxygenase with the cognate electron transfer system. *Biotechnology for Biofuels*, *6*(1), 86. https://doi.org/10.1186/1754-6834-6-86

Zhang, L., King, E., Luo, R., & Li, H. (2018). Development of a High-Throughput, in Vivo Selection Platform for NADPH-Dependent Reactions Based on Redox Balance Principles. *ACS Synthetic Biology*, *7*(7), 1715–1721. https://doi.org/10.1021/acssynbio.8b00179

Zhang, L., Liang, Y., Wu, W., Tan, X., & Lu, X. (2016). Microbial synthesis of propane by engineering valine pathway and aldehyde-deformylating oxygenase. *Biotechnology for Biofuels*, *9*(1), 1–9. https://doi.org/10.1186/S13068-016-0496-Z

5.9 Supplementary Data

Table S5.1. Primers used in this study

Name	Sequence	Purpose
Ec127_serA_Ver_F	gcttgtttgctaaccgtgg	Confirm *serA* deletion
Ec128_serA_Ver_R	gggctttaccgttacagc	Confirm *serA* deletion
Ec051_glyA_Ver_F	cgtagcctgaaggtaatcg	Confirm *glyA* deletion
Ec052_glyA_Ver_R	ttgcaaagttttgtaggccg	Confirm *glyA* deletion
JRN079	TCGTCCTGGCTCACCG	Confirm *sthA* deletion
JRN080	GGGGATCAATTGGCTTAC	Confirm *sthA* deletion
JRN126	caagaatggatggccatttcgataaagttttaaaacaggcggtttaaa cctgtgtaggctggagctgcttc	Amplify *cmR* from pKD3 with *sthA* flanks
JRN127	tgttaccattctgttgcttttatgtataagaacaggtaagccctaccatgc atatgaatatcctccttag	Amplify *cmR* from pKD3 with *sthA* flanks
JRN211	GCGGTCAGTAGTTCAG	Check *pgi* flanks
JRN212	GTCGGCATTGTTATTAAGG	Check *pgi* flanks

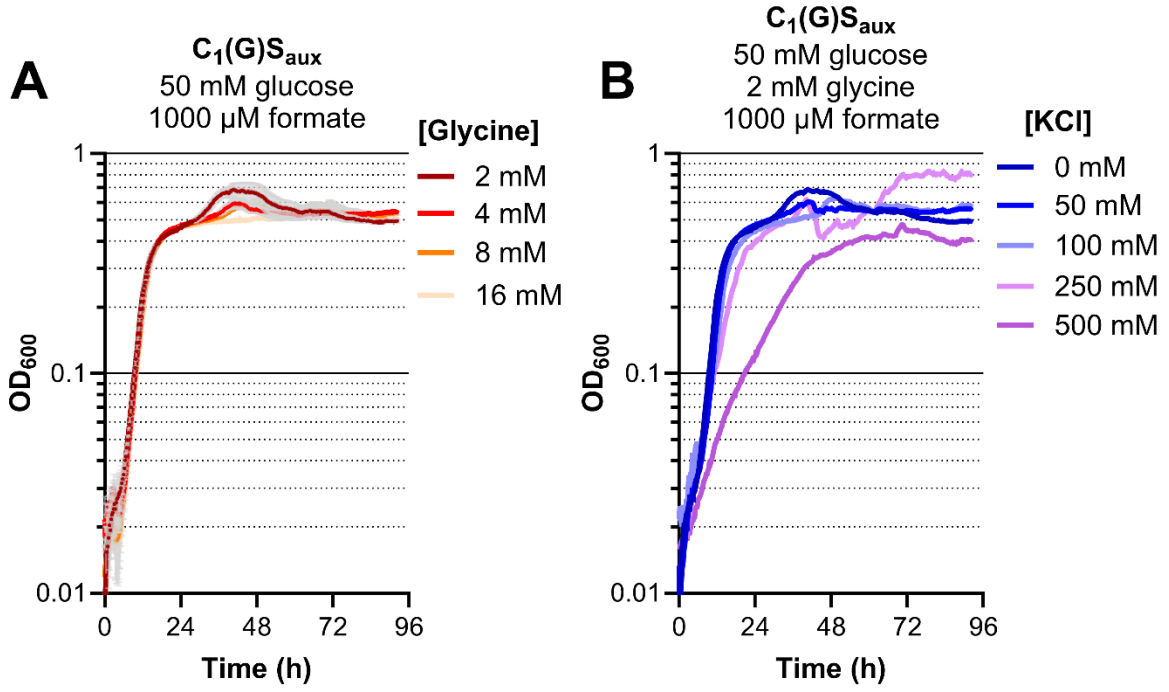


Figure S5.1. Growth response of $C_1(G)S_{aux}$ to supplements. **(A)** Additional glycine was supplemented to assess whether 2 mM was sufficient to saturate growth using 50 mM glucose and 1 mM formate. **(B)** The addition of 250 mM KCl was reported to be beneficial for AAR/ADO-mediated alkane biosynthesis (see Chapter 4). The supplementation of KCl was therefore assessed for its potential effects on growth of $C_1(G)S_{aux}$ supplemented with glucose, glycine, and formate. Only the addition of 500 mM KCl resulted in a reduction of growth rate and final biomass formation. All data were gathered as biological triplicates and errors bars represent their standard deviations.

6. A sensitive NADH-auxotrophic *E. coli* for use as an enzyme engineering platform

6.1 Abstract

Broadly employable enzyme engineering platforms based on growth-coupling principles combine an extremely high throughput with the flexibility to engineer multiple enzymes. Strains auxotrophic for the reduced- or oxidised state of a certain redox cofactor pair, such as NAD^+/NADH or $NADP^+$/NADPH, represent such platforms, since cofactor pairs are involved in a multitude of enzymatic reactions across cell metabolism. In this study, *E. coli* was made auxotrophic for the reduced state of the NAD^+/NADH cofactor pair by removing NADH-generating pathways in central metabolism. The resulting strain, dubbed NADHaux, could be rescued using both endogenous and exogenous NADH-releasing metabolic pathways. A 'cheating' NADHaux strain was also uncovered with a mutation in *nuoF* (E183A), resulting in a cofactor specificity switch affecting oxidative phosphorylation. In summary, NADHaux represents a broadly employable NADH-specific cofactor auxotroph, which can be used for engineering NADH-generating enzymes or pathways. Besides its use as an engineering platform, its characterization pointed to several interesting features of redox metabolism in *E. coli*.

6.2 Introduction

Enzyme engineering efforts are necessary for driving bioprocesses towards commercial viability. In Nature, the biochemical properties of enzymes often gravitate towards moderate efficiencies that are sufficient to run their destined enzymatic reaction, but not exceedingly so (Bar-Even et al., 2011). In contrast, bioproduction processes require highly specialised enzymes often exhibiting certain extreme properties, such as high catalytic efficiencies, strict substrate specificities, or stability at high temperatures. Ideally, enzyme engineering methods

can be used improve all three of these properties towards the desired performance properties. The first step in an enzyme engineering pipeline involves introducing diversity, either rationally by using structural information, or randomly in case this data is lacking or insufficiently informative. Over the past decades, generating this genetic diversity has become increasingly accessible as DNA synthesis costs have plummeted and considerable advances have been made in the areas of *in vitro* and *in vivo* DNA mutagenesis. Acquiring a library of variants is therefore no longer the bottleneck limiting enzyme engineering campaigns, but reliably identifying improved clones from that library remains a significant challenge.

Strategies for screening or selecting enzymes vary vastly in terms of costs, throughput and sensitivity, depending on the variant library size and the required activity measurement (Longwell et al., 2017; Xiao et al., 2015). The most challenging screening or selection protocols are those that attempt to characterise vast variant libraries, such as those generated by random mutagenesis methods, while also attempting to measure enzyme activity that is difficult to quantify. This could be because the enzyme activity is low, requiring highly sensitive detection methods, or because the product or substrate of the enzyme do not lend themselves to rapid, affordable, and accurate quantification.

A popular way of characterizing large variant libraries is to use a growth-coupled approach, where enzyme activity is linked to cell fitness via essential carbon metabolism or the cell's energy state (Alter & Ebert, 2019; Von Kamp & Klamt, 2017). Cells expressing enzymes with a higher catalytic efficiency will be selected for in the population and enriched over multiple generations. In contrast to screening methods the observation of distinct isolates is not required for quantifying their activity, making growth-based approaches relatively simple and accessible for researchers. The throughput of growth-coupled approaches is also amongst the highest among directed evolution processes, limited only by the transformation efficiency of the host cell, or by the mutation rate when using *in vivo* mutagenesis tools. Although these properties seem alluring, setting up a growth-coupled selection method may take a considerable investment in terms of time and resources, often requiring (extensive) genetic

engineering and proof-of-principle work. Therefore, enzyme engineering platforms that can be used for multiple engineering campaigns are of outstanding interest for cost- and time-effective enzyme development.

Growth coupled approaches often rely on relieving some form of synthetically introduced auxotrophy (J. Chen et al., 2022; Orsi et al., 2021). The imposed auxotrophy can either be specific for a certain metabolite, for example an essential amino acid, or more universal such as ATP availability or the cell's redox state (Alter & Ebert, 2019). The latter can be exploited by introducing auxotrophy for either the reduced or oxidised form of a certain cofactor pair, such as NAD^+/NADH or $NADP^+$/NADPH. These cofactors are ubiquitous to microbial life and participate in more than 25% of all cellular reactions, making them attractive targets for the generation of broadly applicable enzyme engineering platforms. We recently reviewed the generation of these cofactor auxotrophic strains and their use as enzyme engineering platforms, where we found that a strain auxotrophic specifically for NADH had not yet been described (Nielsen et al., 2023).

The NAD^+/NADH cofactor pair is an essential redox mediator in *E. coli* metabolism. It plays a central role in glycolysis and concomitant oxidative or fermentative metabolism. In oxygen-rich conditions, NADH generated during substrate catabolism is oxidised to NAD^+ by NADH dehydrogenase, principally via NADH:ubiquinone oxidoreductase 1 (NDH-1), a large multi-subunit membrane complex encoded by the *nuo* operon (Friedrich, 1998). The released electrons are then used to generate a proton motive force via the electron transport chain with oxygen as the terminal electron acceptor, ultimately driving ATP synthase. In anaerobic conditions, NADH-consuming fermentation pathways generate products such as lactate and ethanol in order to regenerate NAD^+ to drive glycolytic flux (De Graef et al., 1999). Besides its role in central metabolism, the NAD^+/NADH redox couple is also required to biosynthesise of a range of essential cofactors and metabolites including, among others, pyroxidal-5-phosphate, arginine, tyrosine, histidine, proline, serine, and threonine, acting either as a reducing- or oxidising agent in the respective metabolic pathways (Kanehisa & Goto, 2000).

Previous work describing NADH auxotrophy included the development of an 'energy auxotrophic' *E. coli* strain carrying the Δ*lpd* mutation, which displayed auxotrophy for reducing power when grown on acetate (Wenk et al., 2020).This required the cell to derive all of its reducing power from an electron donor supplied externally. The exact choice of electron donor was demonstrated to be flexible, as growth could be rescued using formate-, ethanol- or methanol, provided the appropriate dehydrogenase was also present. Due to the Δ*lpd* mutant being auxotrophic to both NADH and NADPH when grown on acetate, this particular strain was effectively unable to discern between the reduction of either cofactor pair. To obtain a strain that is auxotrophic specifically to NADH, the cell must be able to derive a supply of NADPH from a chosen carbon source, whilst still being unable to reduce NAD^+ as a result of this carbon catabolism.

In this work, we developed an NADH-specific cofactor auxotroph, NADHaux, for use in aerobic conditions. We demonstrated how various NADH-generating enzymes and pathways, both endogenous and heterologous, could be used to rescue its growth. In addition, we uncovered an NADHaux mutant that had adapted to grow without being supplied NADH through a rescue pathway, shedding light on the adaptation of *E. coli* to extreme cofactor redox states.

6.3 Methods and materials

6.3.1 Strains, plasmids, media and routine growth

Routine cloning and propagation was performed using LB medium or LB agar with relevant antibiotics: carbenicillin (carb, 100 µg/mL), kanamycin (kan, 50 µg/mL) and chloramphenicol (cm, 25 µg/mL with a multi-copy marker, 17 µg/mL when single copy). Minimal M9 medium was routinely used for growth assays, composed of: 47.8 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 18.7 mM NH_4Cl, 2 mM $MgSO_4$, 100 µM $CaCl_2$, 134 µM ethylene diamine tetrachloroacetic acid (EDTA), 31 µM $FeCl_3 \cdot 6H_2O$, 6.2 µM $ZnCl_2$, 0.76 µM $CuCl_2 \cdot 2H_2O$, 0.42 µM $CoCl_2 \cdot 2H_2O$, 1.62 µM H_3BO_3, and 0.081 µM $MnCl_2 \cdot 4H_2O$, adjusted to pH 7.0. All strains were first streaked on LB agar to obtain single colonies, and gene deletion integrity

was confirmed by colony PCR where relevant. Isolates were then pre-cultured in LB liquid medium supplemented with filter sterilised 10 mM D-fructose, 10 mM D-sorbitol, or 20 mM glycerol, in addition to 20 mM sodium acetate, with 10 mM sodium succinate being added for some experiments as indicated in the text. When pre-culturing mutants for formate- or ethanol dependent growth, cultures were additionally supplemented with 20 mM sodium formate or 100 mM ethanol, respectively. The liquid LB pre-culture was then diluted 100-fold into M9 minimal medium pre-cultures, supplemented identically to the LB cultures. Once grown, these minimal medium pre-cultures were washed three times in sterile 1x M9 salts, and used as inoculum for microtiter plate growth assays. All pre-cultures for growth assays were performed at 37°C, whereas routine growth for gene deletions was performed at 30°C to retain pSIJ8.

Table 6.1. Strains used and constructed in this work.

Strain	Genotype	Reference
E. coli DH5α	*fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	Purchased (New England Biolabs)
E. coli BW25113	*F- LAM- rrnB3 ΔlacZ4787 hsdR514 Δ(araBAD)567 Δ(rhaBAD)568 rph-1*	DSMZ, Germany
BW25113 *lpd::kanR*	*E. coli* BW25113 *lpd::kanR*	Purchased (Horizon Discovery, UK) (Baba et al., 2006)
ΔL	*E. coli* BW25113 *Δlpd::FRT*	This study
ΔLS	ΔL *ΔsthA::FRT*	This study
ΔLP	ΔL *ΔpntAB::FRT*	This study
ΔLSP	ΔLP *ΔsthA::FRT*	This study
ΔLSPM	ΔLSP *Δmdh::FRT*	This study
ΔLSPMA	ΔLSPM *ΔmaeA::FRT*	This study
ΔLS_GN	ΔLS *gapA::gapN-FRT*	This study
ΔLSPMA_GC	ΔLSPMA *gapA::gapC-FRT*	This study
NADHaux	ΔLSPMA *gapA::gapN-FRT*	This study
M1	NADHaux *nuoF*^{E183A} *gltA*^{D229G}	This study

Table 6.2. Plasmids used and constructed in this work.

Plasmid	Contents	Reference
pKD3	FRT-cm^R-FRT, oriR6K, amp^R	(Datsenko & Wanner, 2000)
pSIJ8	araC-P$_{ara}$-α-β-γ rhaRS-P$_{rha}$-FLP, amp^R	(Jensen et al., 2015)
pKD3-*gapC*	*gapC*-FRT-*cm^R*-FRT	This study
pKD3-*gapN*	*gapN*-FRT-*cm^R*-FRT	This study
DVA_pUC19_CD	Amp^R level 0 destination vector, CD cloning sites	(Iverson et al., 2016)
DVK_p15A_FG	Dest. vector, p15A ori, kan^R, *lacZα,* FG flanks	See chapter 3

DVK_pBR322_AE	Dest. vector, pBR322 ori, kan^{R}, *lacZα*, AE flanks	See chapter 3
DVK_p15A_AE	Dest. vector, p15A ori, kan^{R}, *lacZα,* AE flanks	See chapter 3
DVA_J23107_FB	J23107 promoter part, FB flanks	(Iverson et al., 2016)
DVA_J23116_FB	J23116 promoter part, FB flanks	(Iverson et al., 2016)
DVA_J23100_AB	J23100 promoter part, AB flanks	(Iverson et al., 2016)
DVA_J23102_AB	J23102 promoter part, AB flanks	(Iverson et al., 2016)
DVA_J23106_AB	J23106 promoter part, AB flanks	(Iverson et al., 2016)
DVA_J23107_AB	J23107 promoter part, AB flanks	(Iverson et al., 2016)
DVA_J23116_AB	J23116 promoter part, AB flanks	(Iverson et al., 2016)
DVA_J23103_AB	J23103 promoter part, AB flanks	(Iverson et al., 2016)
DVA_B0034m_BC	B0034m RBS part, BC flanks	(Iverson et al., 2016)
DVA_B0032m_BC	B0032m RBS part, BC flanks	(Iverson et al., 2016)
DVA_B0033m_BC	B0033m RBS part, BC flanks	(Iverson et al., 2016)
DVA_B0015_DE	B0015 terminator part, DE flanks	(Iverson et al., 2016)
pL0_JRN005_CD	DVA_pUC19_CD containing domesticated *Pseudomonas sp. 101* NAD^{+}-dependent formate dehydrogenase	See chapter 4
pL0_JRN018_CD	DVA_pUC19_CD containing domesticated *E. coli* MG1655 *adhE*$^{A267T/E568K}$ (*adhE**)	See chapter 3
pL1_JRN067_FG	J23100-B0034m-NAD^{+}-dep. FDH	See chapter 4
pL1_JRN068_FG	J23107-B0034m-NAD^{+}-dep. FDH	This study
pL1_JRN069_FG	J23116-B0034m-NAD^{+}-dep. FDH	This study
pL1_JRN094_FG	J23100-B0034m-*adhE**	See chapter 4
pL1_JRN095_FG	J23107-B0034m-*adhE**	This study
pL1_JRN096_FG	J23116-B0034m-*adhE**	This study

6.3.2 Flux balance analysis

The *e_coli_core* metabolic model, based on *Escherichia coli* MG1655, was retrieved from the BiGG Models database and used for all metabolic modelling simulations (King et al., 2016; Orth et al., 2010). Flux balance analysis was carried out in Python (v3.8) using COBRApy (v0.20.0) (Ebrahim et al., 2013). Several modifications were introduced: the pyruvate formate lyase (PFL) reaction was deactivated since it operates in anaerobic conditions, and proton number translocated by the proton-driven transhydrogenase (THD2) was set to 1 instead of 2 (Wenk et al., 2020). Several reactions and their transport reactions, if necessary, were added to the model manually: pyruvate oxidase in the decarboxylating direction (POX), NAD^{+}-dependent formate dehydrogenase (FDH_NAD), an NADP^{+}-dependent phosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPC) and a non-phosphorylating variant

(GAPN). Fructose, acetate, succinate were set as carbon sources depending on the simulation (lower bounds set between -10 to -20 A.U., see text for details). All FBA simulations were run with biomass maximization as the objective.

6.3.3 Plasmid construction

Plasmids for expression of FDH and AdhE were generated by Golden Gate Assembly, using the CIDAR MoClo kit and the corresponding standardised overhangs and cloning methods (Iverson et al., 2016). We also used MoClo compatible low- and medium copy plasmids developed in other work (see Chapter 3). Standard GGA conditions were as follows: 20 fmol of vector was combined with 40 fmol of insert(s) in a total volume of 20 µL, containing 2 µL of 10x T4 DNA ligase buffer (NEB), 10 units of BsaI-HFv2 or BbsI-HF for level 1 or 0 assemblies respectively (NEB), and 500 units of T4 DNA ligase (NEB). Reaction mixtures were cycled between 5-minute incubations at 37^{o}C and 16^{o}C for total of 80 cycles in a thermocycler, with two final 10-minute steps at 42^{o}C and 80^{o}C before cooling to 4^{o}C. The *gapC* and *gapN* integration cassettes were constructed by first cloning either *gapC* or *gapN* into pKD3. This was achieved by amplification of pKD3 with JRN152+JRN153 (*gapC*) or JRN154+JRN155 (*gapN*) to linearise the backbone with the required complementary overhangs (for primer sequences, see Table S6.1). The *gapC* and *gapN* synthesised fragments (Table S6.2) were then assembled into linearised pKD3 upstream of the FRT-cm^{R}-FRT cassette by HiFi assembly, using conditions outlined by the manufacturer, forming pKD3-*gapC* and pKD3-*gapN* (NEB).

6.3.4 Microtiter plate growth assays

Microtiter plate experiments were performed using a Tecan Spark plate reader with SparkControl software (v3.1 SP1) (Tecan, USA). Clear flat-bottom 96-well plates were used filled with 200 µL of M9 minimal medium supplemented with carbon sources as indicated in the text (Greiner, USA). M9 minimal medium was inoculated to an OD_{600} of 0.01 with washed pre-cultures. To avoid contamination and evaporation, plates were covered with a Breathe-

Easy Sealing Membrane (Sigma-Aldrich) and cultivation was performed using a small humidity cassette (Tecan, USA) re-filled every 24h. Growth assay were performed at 37ºC and OD_{600} measurements were taken every 15 minutes, using the following four-step shaking profile: 1) double orbital, 3 mm amplitude, 90 rpm frequency, 225 seconds 2) orbital shaking, 3 mm amplitude, 180 rpm, 225 seconds 3) double orbital, 2.5 mm amplitude, 108 rpm, 225 seconds, 4) orbital, 2.5 mm amplitude, 216 rpm, 224 seconds. Growth rates were calculated through a log conversion of the blank-medium corrected OD values, followed by determining the maximum linear slope of these values over time using a 6-12 hour sliding window per replicate.

6.3.5 Gene deletions

Lambda red recombination was used for gene deletions, as outlined in previous work (Datsenko & Wanner, 2000; Jensen et al., 2015). Briefly, dsDNA cassettes were prepared by PCR amplification (DreamTaq, ThermoFisher) of the FRT-flanked cm^R marker from plasmid pKD3 using oligos with 50 bp overhangs targeting either side of the gene locus (see Table S6.1 for all targeting primers) (Baba et al., 2006; Datsenko & Wanner, 2000). For integrating *gapC* or *gapN* at the *gapA* locus, pKD3-*gapC* or pKD3-*gapN* were used as PCR templates for cassette generation, respectively. *E. coli* BW25113 carrying pSIJ8 was first grown O/N in liquid LB + carb at 30°C, and 1 mL of O/N culture was used to inoculate 50 mL of liquid LB + carb in a shake flask (agitated at 250 rpm). This culture was grown to an OD_{600} of ~0.3 at 30°C, after which the culture was induced with 15 mM L-arabinose (final concentration) for 45 minutes. Cells were then cooled in an ice slurry for 30 minutes, spun down (1500*g*, 10 mins) and washed twice in ice-cold sterile deionised water. Finally, cells were resuspended in 300 µL ice-cold sterile 10% glycerol. Electrocompetent cells (50 µL) were mixed with 3.5 µL dsDNA cassette (>250 ng/µL in DI water), the mixture transferred to pre-chilled 0.1 cm electroporation cuvettes, and a 1.8 kV pulse was applied. Cells were immediately recovered for 2.5 hours in LB, supplemented with 20 mM glycerol, 20 mM acetate, 10 mM sorbitol, 10 mM succinate, 1 mM L-arabinose and 100 µg/mL carbenicillin. Cells were subsequently plated on LB agar supplemented identically to the recovery medium, with an additional 17 µg/mL

chloramphenicol to select for marker presence and incubated at 30°C. Time for correct clone appearance varied between 1 day to 1 week, depending on the particular deletion and existing genetic background. Correct clones were confirmed by colony PCR (see Table S6.1 for primer sequences), and these were inoculated and re-streaked at least once to ensure clonality. To remove the integrated cm^R marker, 0.5 mL of an O/N LB culture supplemented identically to the electroporation recovery medium (but lacking L-arabinose) was inoculated into 35 mL of LB (also with carbon source supplements), and once its OD_{600} reached ~0.3 flippase expression was induced by addition of 50 mM sterile L-rhamnose. Induction was maintained for 4 hours, or overnight in strains carrying *gapA::gapC*-FRT-cm^R-FRT or *gapA::gapN*-FRT-cm^R-FRT and induced cultures were subsequently diluted 10^{-5}- and 10^{-6}-fold and plated on LB agar with carbenicillin. Correct clones were identified by colony PCR and re-streaked as described earlier. When needed, plasmid pSIJ8 was cured by an overnight cultivation at 42°C.

6.3.6 Formate dehydrogenase expression strength selection

Expression libraries with randomised destination vectors, promoters and RBSs were generated by Golden Gate. Similar to the routine GGA method described above, we used 20 fmol of destination vector (DVK_p15A_AE, DVK_pBR322_AE), 40 fmol promoter part (DVA_J23100_AB, DVA_J23102_AB, DVA_J23106_AB, DVA_J23107_AB, DVA_J23116_AB, DVA_J23103_AB), 40 fmol RBS part (DVA_B0034m_BC, DVA_B0032m_BC, DVA_B0033m_BC), 40 fmol CDS (pL0_JRN005_CD) and 40 fmol terminator part (DVA_B0015_DE). When randomizing destination vectors, we ensured that the total amount of destination vector in the assembly did not exceed 20 fmol (e.g. 10 fmol each when using two destination vectors). Cycling conditions were as previously mentioned. We transformed three aliquots of chemically competent BW25113 Δ*lpd* with 4 µL of Golden Gate reaction mix, recovering for 1h at 37°C. A 50 µL aliquot of each recovered replicate culture was plated on LB agar supplemented with kanamycin, 100 µg/mL 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal, Sigma-Aldrich) and 100 µM isopropyl β-D-1-thiogalactopyranoside (IPTG, Sigma-Aldrich) to assess library fidelity and transformation

efficiency. The remaining cells were spun down, washed three times in sterile 1x M9 salts and finally transferred to a 250 mL shake flask containing 50 mL of sterile M9 medium with 20 mM sodium acetate and 40 mM sodium formate. Cultures were grown at 37°C and shaken at 250 rpm, and growth was monitored daily. Once OD_{600} values plateaued, cultures were diluted 10^4- and 10^5-fold and plated on LB agar plates supplemented with kanamycin. Grown cultures were passaged by 100-fold dilution into flasks with fresh medium. Cultures were passaged another two times, and eight colonies from the plates of every replicate were sequenced at each step to determine the promoter, RBS, and origin of replication of the isolates (Source Bioscience, UK).

6.3.7 Whole genome sequencing

Clonal cultures of BW25113 WT and mutant M1 strains were isolated through first streaking on LB agar supplemented with filter sterilised 20 mM D-sorbitol, 20 mM sodium acetate and 10 mM sodium succinate. A single colony was picked and the relevant genomic loci checked to assess gene deletion fidelity, after which it was cultivated in liquid LB at 37°C supplemented as the agar. Following this, gDNA was isolated (NEB, gDNA extraction kit) and samples sequenced (Illumina sequencing, 2x 150 bp reads) at a coverage depth of ~1500 reads on average (Source Bioscience, UK). Reads were trimmed using fastp v0.23.4 (S. Chen et al., 2018) and aligned to reference genome CP009273 (Grenier et al., 2014) using breseq v0.38.3 in consensus mode (Deatherage & Barrick, 2014). Using other tools in the breseq package we performed a genome subtraction of WT BW25113 from the M1 mutant to identify unique mutations.

6.4 Results

6.4.1 Design and metabolic modelling to achieve NADH auxotrophy

Sugar metabolism in *E. coli* provides an abundant source of reducing power through the Embden-Meyerhof-Parnis- (EMP), oxidative pentose phosphate- (oxPPP) and Entner-

Doudoroff (ED) pathways. In terms of pyruvate formation, the EMP pathway generates NADH through glyceraldehyde-3-phosphate dehydrogenase (*gapA,* GAPDH), generating 2 mol NADH and net 2 ATP per mol hexose. In contrast, oxPPP generates 2 mol NADPH in addition to 1.66 mol NADH and 1.66 mol ATP per mol hexose (if all carbon re-enters glycolysis), at the expense of a decarboxylation catalysed by 6-phosphogluconate dehydrogenase (*gnd*). The ED pathway releases 1 NADPH, 1 NADH and 1 ATP per mol hexose, but retains all carbon. Pyruvate-forming flux through unmodified EMP, oxPPP or ED pathways therefore always generates NADH through *gapA*. This enzyme is therefore essential for growth on sugar substrates, as has been experimentally determined in several works (Baba et al., 2006; Della Seta et al., 1997; Goodall et al., 2018), meaning that we could not simply remove *gapA* to avoid its NADH-generating activity. Replacement of the *gapA* gene with an $NADP^+$ dependent version has been explored in several studies aiming to improve NADPH-yields from glycolysis, with phosphorylating *gapC* from *Clostridium acetobutylicum* and non-phosphorylating *gapN* from *Streptococcus mutans* successfully being used (Martínez et al., 2008; Spaans et al., 2015; Takeno et al., 2016; Zhang et al., 2018). Another approach would be to circumvent *gapA* activity altogether through glycolytic bypasses such as the phosphoketolase- or methylglyoxal bypass, but since we aimed to generate a pool of NADPH for anabolic reactions, we chose to replace *gapA* with an $NADP^+$-dependent version (Iacometti et al., 2022; Lin et al., 2018) .

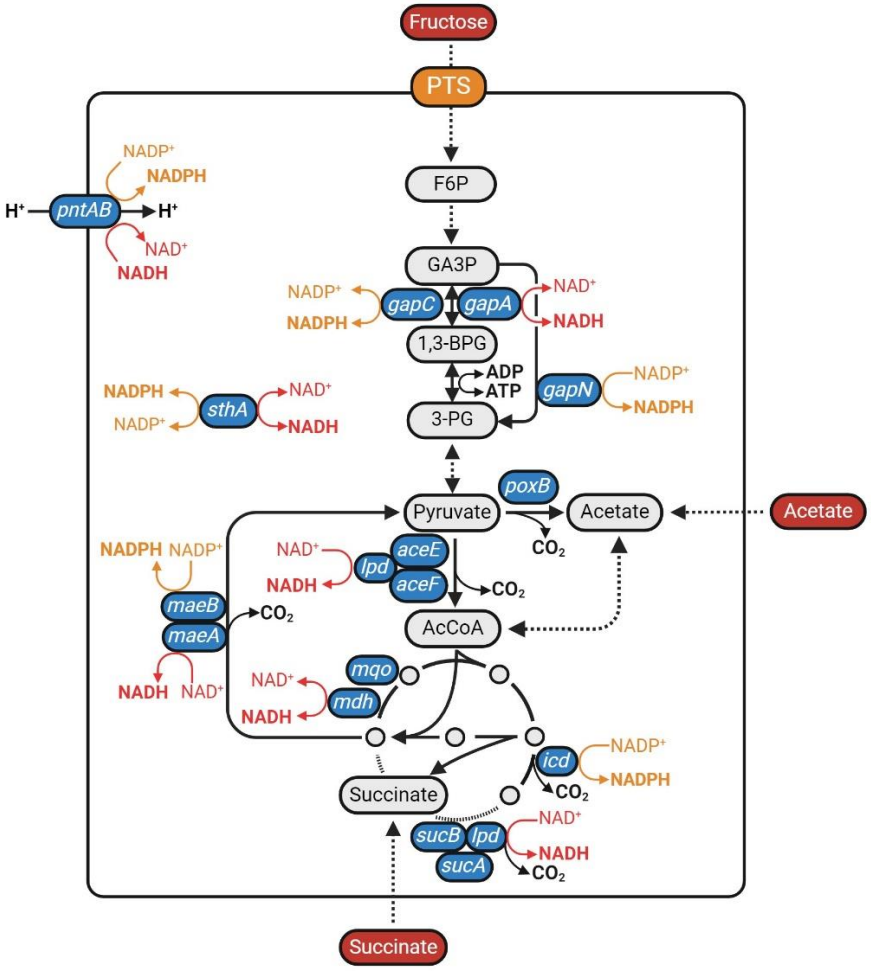


Figure 6.1. Simplified schematic of *E. coli* central metabolism focusing on reactions involving redox cofactors. This diagram displays several central metabolic NAD^+/NADH and $NADP^+$/NADPH requiring enzymes or enzyme complexes, excluding fermentative reactions and pentose phosphate pathway reactions. The reactions catalysed by *gapC* and *gapN* GAPDH enzymes were also indicated. Abbreviations: *aceE* (pyruvate dehydrogenase); *aceF* (dihydrolipoamide acetyltransferase); *gapA* (NAD^+-dependent glyceraldehyde-3-phosphate dehydrogenase); *gapC* ($NADP^+$-dependent glyceraldehyde-3-phosphate dehydrogenase); *gapN* (non-phosphorylating $NADP^+$-dependent glyceraldehyde-3-phosphate dehydrogenase); *icd* (isocitrate dehydrogenase); *lpd* (dihydrolipoyl dehydrogenase); *maeA* (NAD^+-dependent malic enzyme); *maeB* ($NADP^+$-dependent malic enzyme); *mdh* (malate dehydrogenase); *mqo* (malate:quinone-oxidoreductase) *pntAB* (pyridine nucleotide transhydrogenase); *poxB* (pyruvate oxidase) *sthA* (soluble transhydrogenase); *sucA* (2-oxoglutarate dehydrogenase); *sucB* (dihydrolipoyllysine succinyltransferase).

Several other enzymes in lower glycolysis and the TCA cycle also require the NAD^+/NADH cofactor pair, such as the pyruvate dehydrogenase complex (PDHc, *lpd, aceE, aceF*), alpha ketoglutarate dehydrogenase complex (αKGDHc, *lpd, sucA, sucB*), malate dehydrogenase (*mdh*) and the decarboxylating NAD^+-dependent malic enzyme (*maeA/sfcA*) (Figure 6.1). *E. coli* also harbours two transhydrogenase enzymes, a soluble (*sthA*) and a membrane-bound variant (*pntAB*). The former contributes towards the oxidation of NADPH when NADH

availability is limited, such as during growth on acetate, while the latter contributes significantly to NADPH-generating flux in aerobic glucose-grown *E. coli* (Sauer et al., 2004).

To investigate the effects of removing NADH-generating reactions on biomass flux we used a simplified metabolic model of *E. coli* central metabolism, E_coli_core. Using several *in silico* generated mutants, we investigated overall biomass flux and the distribution of fluxes through NADH-generating reactions (Figure 6.2). When an unmodified model was supplied with fructose, a carbon source entering upper glycolysis as fructose-6-phosphate (F6P), the majority of NADH was generated by GAPDH, PDHc and αKGDHc, together contributing to around 80% of total NADH flux. Introducing the Δ*lpd* deletion (ΔL) resulted in a slight decrease in biomass formation, but mostly caused a dramatic shift in the distribution of NADH-generating fluxes, with around 75% now being generated by STHA (Figure 6.2B). Additional deletion of STHA again led to a flux reorganisation, now utilising MDH and MAEA to generate ~50% of the model's NADH supply and resulting in a further biomass flux decrease.

The *in silico* deletion of PNTAB resulted in no significant differences in NADH flux distribution nor could we detect any flux through PNTAB. We note that this does not accurately reflect transhydrogenase activity *in vivo*, as SthA and PntAB act independently of one another *in vivo* and will perform their respective redox reactions even when one of the two is inactivated. Indeed, several studies describing the construction of cofactor auxotrophic strain have found that deleting only one transhydrogenase led to growth defects or non-selective growth (Lindner et al., 2018; Maxel et al., 2020; Sellés Vidal et al., 2021).

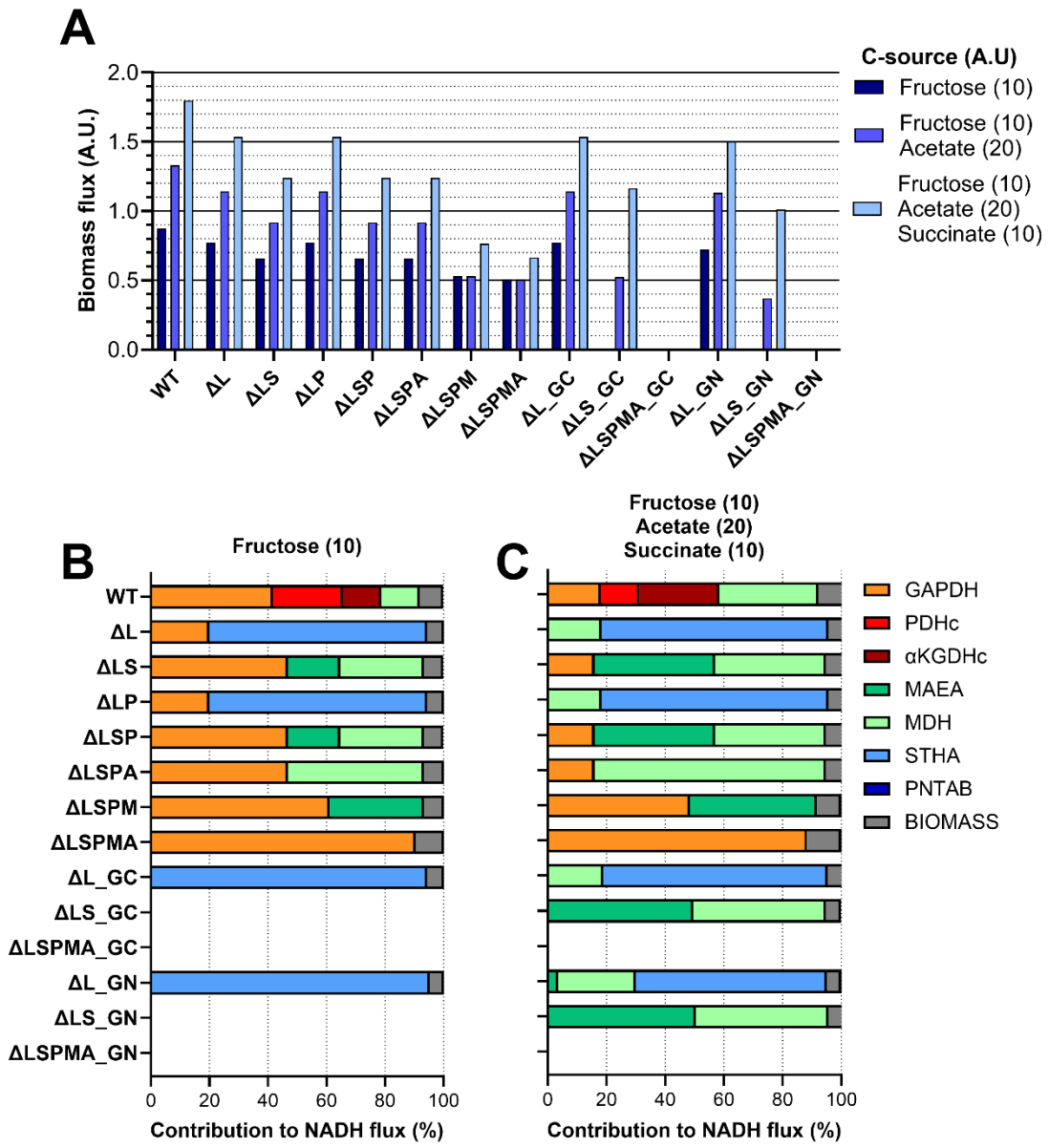


Figure 6.2. Flux balance analysis results of NADH-deficient *E. coli* models. (A) Flux towards biomass formation (A.U.) of a series of *in silico* mutants lacking NADH-regenerating enzymes when provided with 10 A.U. fructose, 10 A.U. fructose and 20 A.U. acetate, or 10 A.U. fructose, 20 A.U. acetate and 10 A.U. succinate. The distribution of NADH-generating reaction fluxes of *in silico* mutants displayed as percentages of the total NADH flux, modelled with either **(B)** 10 A.U. fructose or **(C)** 10 A.U. fructose, 20 A.U. acetate and 10 A.U. succinate as available carbon sources.

The removal of all NADH-generating reactions except GAPDH (ΔLSPMA) resulted in all NADH being generated by GAPDH and by the BIOMASS reaction which results in low NADH flux. The replacement of GAPDH by either phosphorylating GapC (GC) or non-phosphorylating GapN (GN) both resulted in a complete lack of biomass- and NADH flux in a ΔLSPMA

background (Figure 6.2A and B). Interestingly, supplying only fructose resulted in no biomass flux in ΔLS_GN and ΔLS_GC models while the addition of acetate or acetate and succinate restored biomass formation via MDH and MAEA flux, indicating that a baseline level of TCA cycle flux was required to establish biomass flux in ΔLS_GC/GN (Figure 6.2A and C). Introducing GC or GN in a ΔLSPMA background resulted in no NADH- nor biomass flux in either model, regardless of the carbon sources being made available. The FBA results therefore suggested that these two strains represented NADH auxotrophs, providing a blueprint for the ensuing metabolic engineering.

6.4.2 Confirming Δ*lpd* phenotype and rescuing its growth

In this study, we used *E. coli* BW25113 *Δlpd* (ΔL) as the 'base' strain to start construction of an NADH-specific cofactor auxotroph. Previous work also demonstrated growth-rescue of ΔL using heterologously expressed pathways, one of these being formate oxidation by formate dehydrogenase (FDH), to provide reducing power (Figure 6.3A) (Wenk et al., 2020). We elaborated on this work by performing a selection experiment using a library of FDH constructs with randomised promoters and RBSs to identify its optimal expression level (Figure 6.3B). We found that low-strength promoters and a high-strength RBS were selected for after 3 serial passages, with >50% of isolates clones carrying a J23116-B0034m construct in a p15A backbone (Figure 6.3C). Both p15A and pBR322 origins of replication were represented in the clones after 3 passages.

The ΔL strain was also grown with acetate and/or glycerol to demonstrate that it was unable to grow on acetate, but that glycolytic flux from glycerol catabolism would provide it with sufficient reducing power to restore growth (Figure 6.3D). Besides confirming this, the data also suggested that *E. coli* BW25113 ΔL displayed PoxB activity, since growth on glycerol as a sole carbon source was possible (Li et al., 2006; Wenk et al., 2020). Note that the final OD_{600} values of glycerol- and glycerol/acetate-grown strains were identical, since ΔL growth was limited not by total carbon availability, but by reducing power provided by glycerol only.

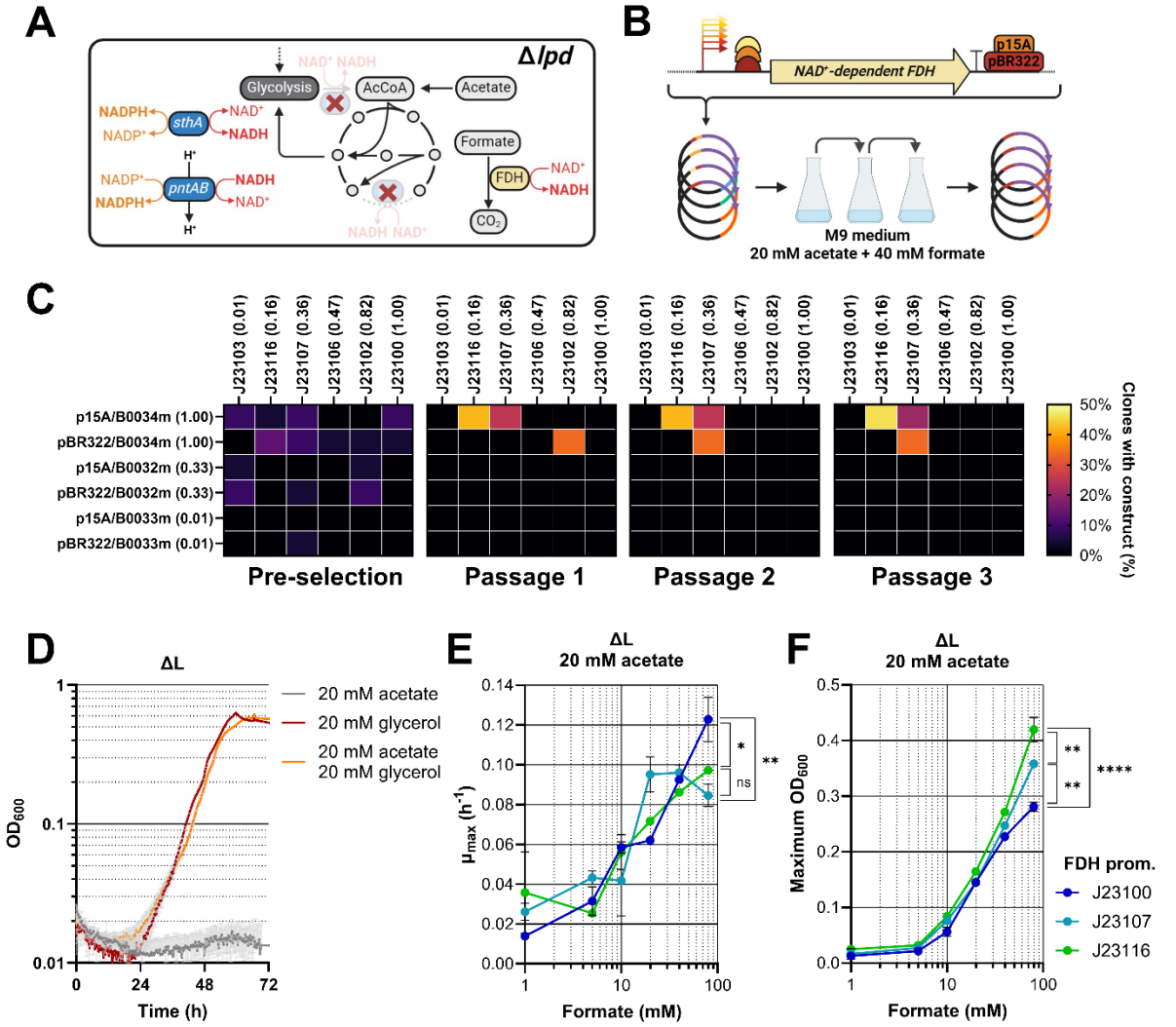


Figure 6.3. Characterization of *E. coli* Δ*lpd* (ΔL) and rescuing growth by formate oxidation. **(A)** Schematic of central metabolism involving the *lpd* gene product. Deletion of *lpd* renders the pyruvate- and α-ketoglutarate dehydrogenase complexes inactive. **(B)** Schematic of Golden Gate-mediated expression strength randomization. Promoters, RBSs and backbones were combined randomly to generate FDH constructs of varying expressing strengths. Libraries were transformed into ΔL and passaged in selective conditions. **(C)** Distribution of FDH constructs during selective passaging of ΔL using acetate and formate medium with randomised FDH expression constructs. **(D)** Growth characterization of ΔL using acetate, glycerol, or acetate and glycerol. **(E)** Maximum growth rates and **(F)** maximum OD_{600} values of ΔL grown on 20 mM acetate as a function of the formate concentration and FDH promoter (all constructs carried a B0034m RBS and DVK_p15A_FG backbone). All data in figures **C-F** were averages derived from biological triplicate measurements. Error bars, where relevant, represented the standard deviation of this average. Where indicated, statistical analysis was performed by comparing means using a one-way ANOVA and post-hoc Tukey's multiple comparison test. P-value significance indicated are as follows: ns (p-value >0.05); * (p-value <0.05), ** (p-value <0.01), *** (p-value <0.001) or **** (p-value <0.0001).

Separate growth characterization of FDH expression constructs with stronger promoters (J23100 or J23107) revealed these to exhibit lower final OD_{600} values with 80 mM formate supplementation, in a promoter strength-dependent manner (Figure 6.3F). In general, we did

not observe significant differences in μ_{max} trends per different promoter, although when using 80 mM formate the J23100 promoter led to significantly higher growth rates than J23116 and J23107 (Figure 6.3E).

To confirm ΔL behaviour using a difference electron source we used the *E. coli* bifunctional alcohol/aldehyde reductase *adhE* carrying A267T and E568K mutations (designated *adhE**) known to increase aerobic activity. The *adhE** gene was cloned with high (J23100), medium (J23107), and low (J23116) strength constitutive promoters and transformed into ΔL. In contrast to acetate, two mol NADH is generated per mol of oxidised ethanol, meaning that ΔL should be able to grow on ethanol as a sole carbon source. We found that a small amount of glycolytic carbon source was required to 'kickstart' growth on ethanol (in this case 1 mM glycerol), and growth was achieved using ethanol with an *adhE** under high constitutive expression (promoter J23100 + RBS B0034m) (Figure S6.1). Ethanol concentrations as low as 20 mM led to growth using this construct, whereas *adhE** expressed using J23107 or J23116 promoters could only support growth in conditions with 100 mM ethanol.

Taken together, several important aspects of the Δ*lpd* mutant were assessed to confirm the strain operated as described in previous work (Li et al., 2006; Wenk et al., 2020). This included its inability to grow on acetate as a sole carbon source, and its ability to be rescued by glycolytic carbon sources or the oxidation of exogenous compounds such as formate and ethanol.

6.4.3 Deletion of central metabolic NADH-regenerating enzymes

After confirming that ΔL displayed the expected phenotypes, we proceeded with deleting the required central metabolic NADH-generating enzymes and transhydrogenases (*sthA, pntAB, maeA* and *mdh*) as predicted by our metabolic modelling results, except *gapA*. We then assessed the growth of these mutants on 20 mM glycerol, 10 mM D-fructose or 10 mM D-sorbitol as sole glycolytic carbon sources in addition to 20 mM acetate (Figure 6.4). We opted to use D-sorbitol as a carbon source since its catabolism results in the release of NADH

through the activity of sorbitol-6-phosphate dehydrogenase (*srlD*) (Novotny et al., 1984). Considering that D-fructose uptake does not generate NADH and *srlD* activity generates D-fructose-6P, fructose and sorbitol catabolism differ by the NADH released in the latter. As such, these carbon sources were chosen to assess growth of various NADH deficient strains throughout this study.

Firstly, we noticed that the mutant strains were surprisingly robust in terms of their ability to retain normal growth behaviour compared to ΔL. Indeed, we found no significant differences between the maximum growth rates of ΔLSMPA, theoretically the most NADH-deficient strain of this group, and ΔL using acetate and glycerol (0.199 h^{-1} vs. 0.268 h^{-1}, resp., p=0.136) acetate and fructose (0.157 h^{-1} vs. 0.241 h^{-1}, resp., p=0.301) or acetate and sorbitol (0.137 h^{-1} vs 0.153 h^{-1}, resp., p=0.997). Wildtype BW25113 had a significantly higher growth rate compared to ΔL and ΔLSPMA when grown with acetate and fructose (0.430 h^{-1}) and acetate and sorbitol (0.367 h^{-1}), although growth rates of BW25113 using acetate and glycerol (0.285 h^{-1}) were not significantly higher than that of ΔL. We note that these strains were precultured in medium containing glycerol and acetate, and as such we expected that the switch to sorbitol -requiring expression of *srlD*- would result in a lag phase. We did not observe long lag phases for sorbitol growth in ΔL, ΔLS, ΔLP, ΔLSP or ΔLSPA mutants, but ΔLSMPA required a 96h lag phase before growth initiated. Interestingly, ΔLSPM displayed a complete inability to grow on sorbitol and acetate during the 144h experiment. We currently do not have an explanation for this behaviour especially since growth on fructose/glycerol and acetate was normal, although we expect that given enough time ΔLSPM would also have initiated growth on sorbitol.

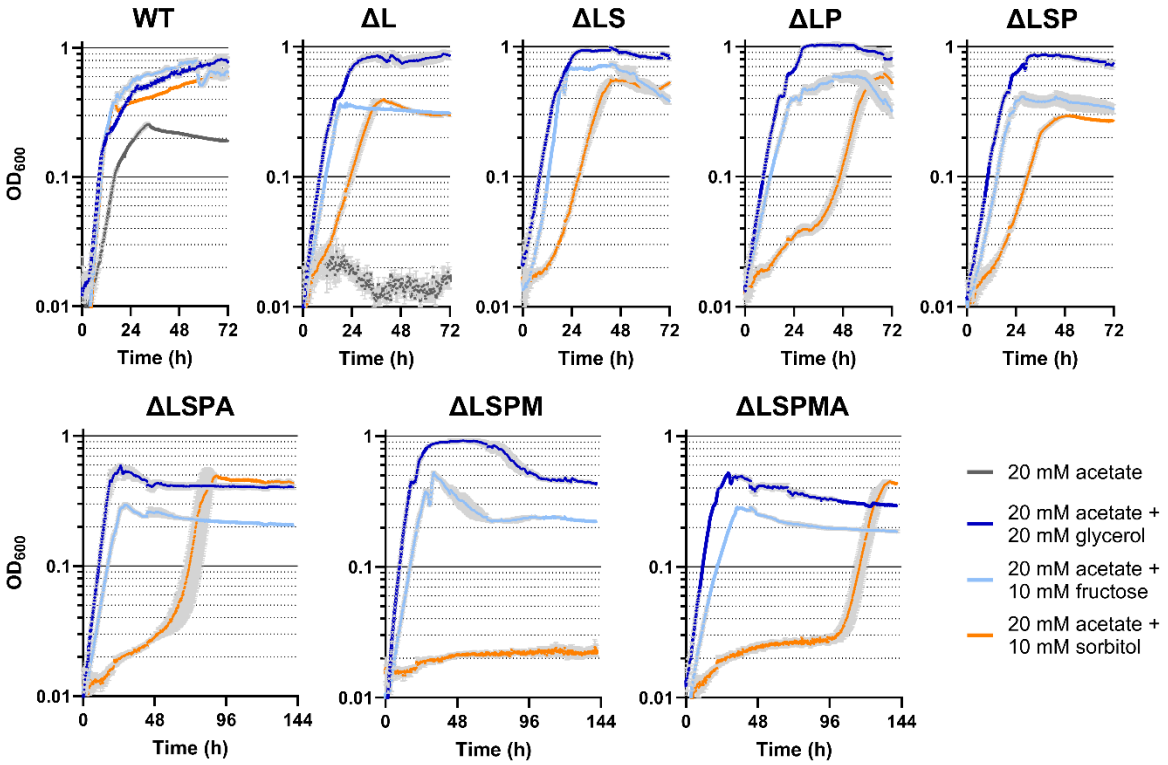


Figure 6.4. Growth characterization of *E. coli* strains lacking NADH-generating central metabolic enzymes. Sequential deletion of genes involving NADH-regeneration resulted in strains Δ*lpd* (ΔL), Δ*lpd* Δ*sthA* (ΔLS), Δ*lpd* Δ*pntAB* (ΔLP), Δ*lpd* Δ*sthA* Δ*pntAB* (ΔLSP), Δ*lpd* Δ*sthA* Δ*pntAB* Δ*maeA* (ΔLSPA), Δ*lpd* Δ*sthA* Δ*pntAB* Δ*mdh* (ΔLSPM) and Δ*lpd* Δ*sthA* Δ*pntAB* Δ*mdh* Δ*maeA* (ΔLSPMA). These mutants, and WT BW25113, were grown using combinations of acetate and various glycolytic carbon sources, as indicated. All datapoints represent averages derived from biological triplicate measurements. Maximum growth rate measurements as mentioned in the text were compared using a one-way ANOVA and post-hoc Tukey's multiple comparison test.

An important observation was that the achieved maximum OD_{600} values of ΔLSPA and ΔLSPMA grown on sorbitol and acetate were significantly higher than on fructose and acetate. This could be attributed to the fact that the catabolism of sorbitol to F6P generates NADH, whereas that of fructose does not. Surprisingly, this behaviour does not seem consistent when compared to the double transhydrogenase knockout ΔLSP, which displayed a maximum OD_{600} of ~0.2 when grown on sorbitol and acetate, while the theoretically more NADH-deficient ΔLSPA and ΔLSMPA mutants reached around 50% higher values. This pointed to the complex interplay of cofactor balancing in *E. coli* for efficient substrate assimilation, and the disruption caused by removing these balancing mechanisms.

In general, growth on glycerol in ΔL and all derived strains resulted in higher maximum OD_{600} values compared to using fructose or sorbitol. Since the same number of carbon units were entering metabolism in all cases, theoretically one would expect identical final OD_{600} values, as we observed in the BW25113 WT strain. Since glycerol enters glycolysis in the form of dihydroxyacetone phosphate (DHAP), it only requires one further isomerization reaction catalysed by *tpi* to generate glyceraldehyde-3-phosphate (GA3P), the substrate for NADH-producing GapA. The growth of a strain exhibiting insufficient NADH production, such as ΔL or its derivates, could benefit from substrates entering glycolysis in close metabolic vicinity to GapA, as this could result in a higher NADH-generating flux. In contrast, some of the carbon from catabolised fructose or sorbitol may instead be redirected to the pentose phosphate pathway since fructose-6-phosphate can be converted to glucose-6-phosphate by *pgi*. Glycerol catabolism can also directly involve the NADH-generating *gldA* enzyme, however this enzyme should only be active anaerobically hence we assumed it was not functional in the tested conditions.

6.4.4 Characterization of mutants carrying an $NADP^+$-dependent GAPDH

The growth behaviour of ΔL and derived strains revealed that *E. coli* is surprisingly versatile in terms of its ability to balance NADH flux even when all bar one of its central NADH-generating enzymes were deleted. This indicated that the only remaining central NADH-generating enzyme, GapA, must be capable of satisfying almost all the demand for NADH in a ΔLSMPA background. This was corroborated by our metabolic modelling results, which indicated that only strains with an $NADP^+$-dependent GAPDH would display significant growth defects without an external NADH-generating pathway (Figure 6.2).

To verify these *in silico* results, we first assessed an intermediate strain, ΔLS_GN, for its ability to grow with- and without a source of NADH. We chose this variant since the *in silico* results suggested that these adaptations (Δ*lpd* Δ*sthA gapA::gapN*) represented the minimal set of required modifications to cause a considerable NADH deficiency resulting in growth defects.

This strain carried a chromosomal replacement of the native *gapA* with *gapN,* a strictly $NADP^+$-dependent non-phosphorylating GAPDH.

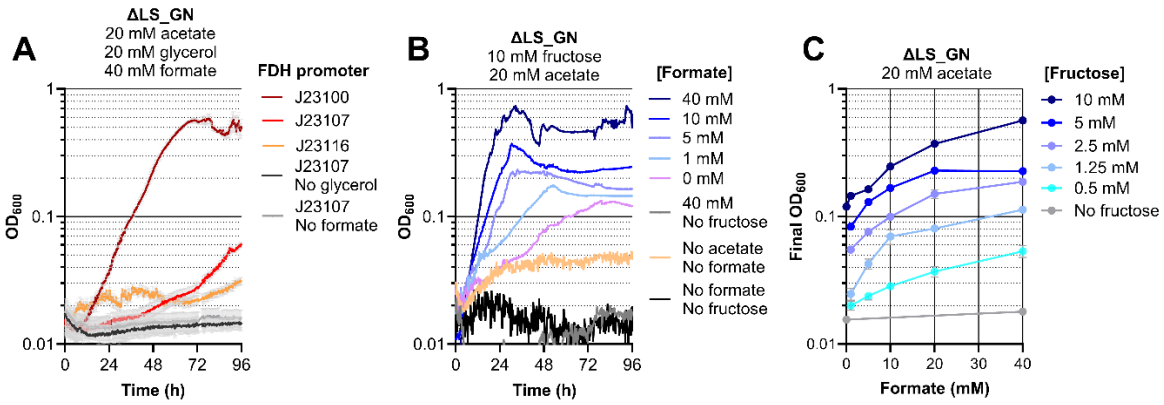


Figure 6.5. Characterization of ΔLS_GN growth. (A) The ΔLS_GN strain was grown on acetate, glycerol and formate medium carrying FDH constructs with either J23100 (pL1_JRN067_FG), J23107 (pL1_JRN068_FG) or J23116 (pL1_JRN069_FG) promoters. **(B)** Growth of ΔLS_GN carrying pL1_JRN067_FG was also assessed in fructose + acetate medium, where it as clear that formate was not strictly required for biomass formation although its supplementation contributed to more rapid growth. **(C)** The effects of fructose- and formate concentration on final biomass formation. The ΔLS_GN strain grew to higher final OD_{600} values as a function of both the formate and fructose concentration, as opposed to just formate supplementation. This confirmed that background NADH reactions were still operational. All data in figures were averages derived from biological triplicate measurements. Error bars represented the standard deviation of this average.

We transformed ΔLS_GN with constitutive NAD^+-dependent FDH constructs with varying expression strengths and grew these in minimal medium supplemented with acetate, glycerol, and formate (Figure 6.5A). Use of the strong J23100 promoter resulted in considerably reduced lag phase compared to medium- and low-strength promoters. This contrasted with the ΔL strain, carrying an intact *gapA* gene, which benefitted from relatively low expression of FDH. This could be explained by the demand for NADH presumably being much greater in ΔLS_GN, due to the replacement of *gapA* with an $NADP^+$-dependent version and the deletion of the soluble transhydrogenase. In addition, strains carrying a *gapA::gapN* replacement were expected to generate considerable quantities of NADPH through glycolysis, increasing the NADPH/NADH ratio. Possibly, higher expression of NADH-generating FDH prevented this redox cofactor ratio from becoming inhibitory.

Due to both NAD^+-dependent malic enzyme and malate dehydrogenase still being present in ΔLS_GN, the strain still harboured pathways capable of NADH-regeneration. Indeed, we observed growth of ΔLS_GN on fructose and acetate without formate supplementation (Figure 6.5B). A notable observation in this same graph was that no growth was observed in conditions with only 10 mM fructose. The requirement for acetate could have to do with *poxB* expression, which is essential for pyruvate to acetyl-CoA flux in ΔLS_GN and known to be induced by acetate (Millard et al., 2021). However, our FBA modelling also indicated that no growth would be possible by addition of fructose alone. Upon addition of acetate to the ΔLS_GN model, flux through MDH and MAEA reactions was enabled, generating NADH (Figure 6.2B and C).

In addition, the relationship between formate addition and OD_{600} was affected by the concentration of supplemented fructose; a higher fructose concentration led to higher final OD_{600} values (Figure 6.5C). While this might seem intuitive at first considering that fructose is limiting to growth, this behaviour also occurred when no formate was supplemented. This indicated that remaining NADH-generating enzymes in central metabolism were capable of generating NADH as a result of higher carbon availability in central metabolism, presumably through the Mdh and MaeA reactions.

Since ΔLS_GN still displayed background growth, we turned to the ΔLSPMA strain. This strain harboured all required deletions to establish NADH auxotroph, except it still carried the native *gapA* gene. The relatively fast growth of ΔLSPMA on fructose demonstrated the important role *gapA* fulfils in maintaining intracellular redox- and energy homeostasis. To abolish NADH-generation through glycolysis we replaced *gapA* with *gapC*, an $NADP^+$-dependent phosphorylating GAPDH, yielding ΔLSPMA_GC (Figure 6.6A). This mutant grew rapidly on sorbitol with final OD_{600} values dependent on sorbitol concentration but struggled to grow on fructose (Figure 6.6B and C). Even so, after 48h we observed that fructose-supplemented strains were able to grow, which we explained by the slightly promiscuous cofactor specificity of *gapC* as reported in earlier work (Zhang et al., 2018). This assumption was corroborated by increases in fructose concentration leading to increased growth rates. We hypothesised

that increased glycolytic flux caused by higher fructose concentrations could lead to higher NADH synthesis by the promiscuous *gapC* enzyme, stimulating growth. In contrast, growth rates using sorbitol did not increase as a function of its concentration; growth rate was already saturated at 1 mM sorbitol presumably due to the obligatory NADH synthesis through sorbitol catabolism (Figure 6.6B).

Omitting acetate in succinate- and sorbitol-supplemented medium resulted in an extended lag phase compared to those supplemented with even low (1 mM) amounts of acetate (Figure 6.6E). After 72h the growth rate of the acetate-lacking strain achieved similar values to that of acetate-supplemented strains. This suggested that acetate may not be a main source of biomass-forming carbon, although its presence may have regulatory effects such inducing *poxB* expression as mentioned earlier. Omitting succinate resulted in a more severe growth defect; when 1 mM succinate was added to 10 mM sorbitol and 20 mM acetate, we observed biphasic growth with an initial fast growth phase followed by a significant slowing of growth, presumably switching at the point of succinate depletion (Figure 6.6D). Growth without initial succinate supplementation also resulted in a severe growth deficiency, reaching a final OD_{600} of only ~0.04 after 144 hours. This suggested that succinate plays a critical role in establishing a viable redox state in these NADH-deficient conditions.

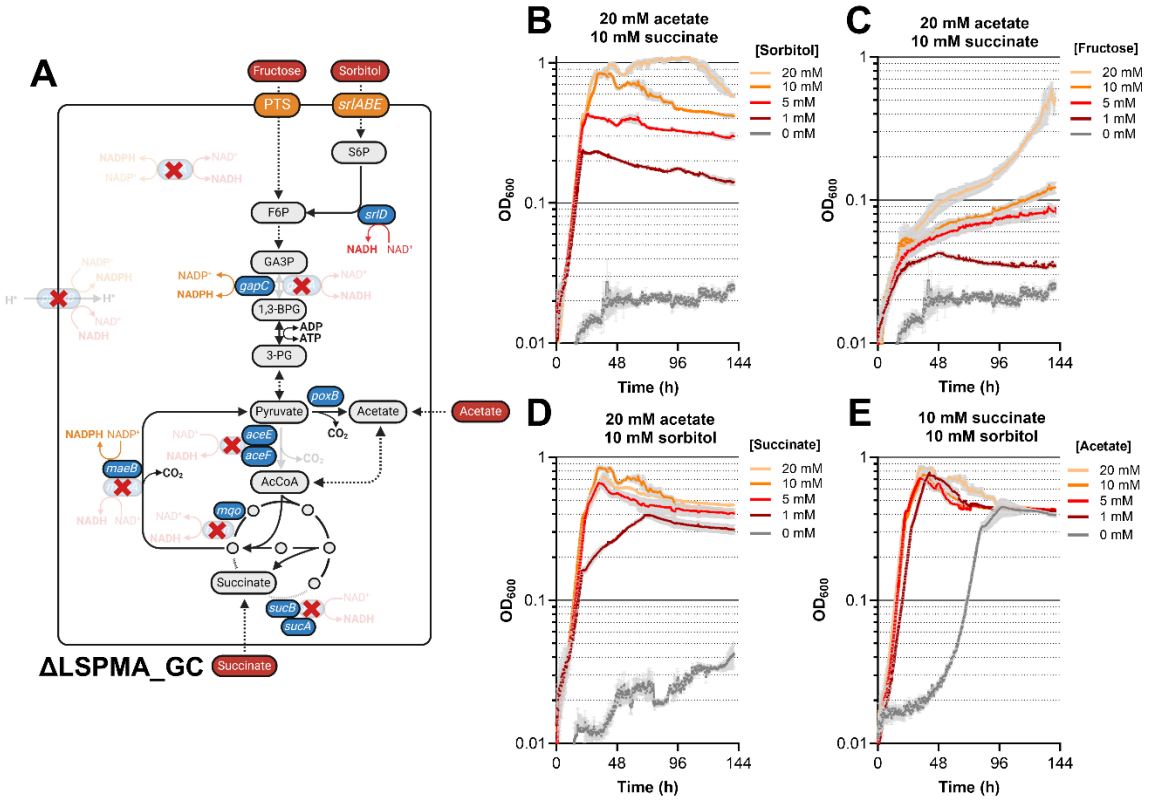


Figure 6.6. Growth characterization of ΔLSPMA_GC. **(A)** Schematic of cell metabolism including the knocked-out genes indicated with red crosses. Metabolites are indicated by grey boxes, while relevant *E. coli* genes are represented by blue boxes often with their cofactor usage indicated. Red boxes indicate substrates, orange boxes represent membrane-bound substrate transport protein (complexes). Dotted lines represent either transport reactions or metabolic pathways consisting of more than one reaction. The effect of using fructose or NADH-generating sorbitol on the growth of ΔLSPMA_GC was assessed by growth assays using acetate, succinate, and various concentrations of **(B)** D-sorbitol or **(C)** D-fructose. The succinate and acetate requirements of ΔLSPMA_GC were assessed by growing the strain with **(D)** acetate, sorbitol, and varying concentrations of succinate or **(E)** succinate, sorbitol, and varying concentrations of acetate. All measurements represented averages of biological triplicates and error bars represent the standard deviation of these values. Abbreviations: *aceE* (pyruvate dehydrogenase); *aceF* (dihydrolipoamide acetyltransferase); *gapA* (NAD^+-dependent glyceraldehyde-3-phosphate dehydrogenase); *gapC* ($NADP^+$-dependent glyceraldehyde-3-phosphate dehydrogenase); *lpd* (dihydrolipoyl dehydrogenase); *maeA* (NAD^+-dependent malic enzyme); *maeB* ($NADP^+$-dependent malic enzyme); *mdh* (malate dehydrogenase); *mqo* (malate:quinone-oxidoreductase) *pntAB* (pyridine nucleotide transhydrogenase); *poxB* (pyruvate oxidase); *srlD* (D-sorbitol-6-phosphate dehydrogenase); *sthA* (soluble transhydrogenase); *sucA* (2-oxoglutarate dehydrogenase); *sucB* (dihydrolipoyllysine succinyltransferase).

6.4.5 Growth of NADHaux using a glycolytic NADH source reveals $NADP^+$ auxotrophy

Characterization of ΔLS_GN and ΔLSPMA_GC granted insights into the cultivation conditions required to rescue NADH deficient strains. However, neither of these strains were truly auxotrophic to NADH since both were able to grow using acetate (and succinate, for ΔLSPMA_GC) and fructose as carbon sources. We assumed that the promiscuous cofactor specificity of *gapC* was causing low, but sufficient, NADH flux to rescue growth. To avoid this issue, another strain was generated dubbed ΔLSPMA_GN, or NADHaux, featuring a *gapA::gapN* gene replacement in a ΔLSPMA background known for its stringent $NADP^+$-specificity (Zhang et al., 2018).

In our attempts to culture NADHaux using sorbitol as a source of NADH we noticed extreme growth deficiencies both in rich- and minimal media. Using native glycolytic metabolism to generate NADH differs from the use of an external electron donor, such as formate, in terms of how NAD^+/NADH and $NADP^+$/NADPH cofactor pools are balanced. In the former, the NADH/NADPH ratio during substrate catabolism is fixed by the stoichiometry of the engineered glycolytic pathway, whereas the latter allows a decoupling of this ratio, since NADH can be generated using a freely available electron donor. The deleterious effects of a *gapA*::*gapN* modification on $NADP^+$/NADPH ratios and growth had previously been reported (Zhang et al., 2018), which prompted us to assess whether this could also be responsible for the lack of growth of NADHaux.

We attempted to use several NADPH-consuming pathways to relieve the presumed $NADP^+$ auxotrophy and enable growth on sorbitol as sole NADH-generating carbon source (Figure 6.7A). Firstly, we found that the supplementation of 5 mM octanal, a C_8 aliphatic aldehyde, enabled growth of NADHaux on LB agar supplemented with sorbitol, acetate, and succinate after 48h (Figure 6.7C), while a non-supplemented strain did not form visible colonies within this time (Figure 6.7B). *E. coli* carries many NADPH-dependent endogenous aldehyde reductases capable of reducing aldehydes to their corresponding alcohols. Next, we took advantage of the fact that the *lpd* deletion causes an accumulation of α-ketoglutarate. This is

a substrate for the transaminating glutamate synthase (GOGAT, *gltD* and *gltB*), which requires L-glutamine and NADPH in addition to α-KG to generate two L-glutamate. Upon addition of 1-10 mM L-glutamine to LB agar with sorbitol, acetate, and succinate, we observed growth of NADHaux within 48h (Figure 6.7D). Finally, another native NADPH sink used by *E. coli* to protect it against oxidative stress is the glutathione pool, which can be cycled between its reduced- (GSG) and oxidised states (GSSG). We sought to stimulate NADPH consumption by oxidising the glutathione pool through supplementation of either hydrogen peroxide (H_2O_2), a potent oxidiser, or GSSG directly. Although supplementation with H_2O_2 (1-5 mM) did not result in visible growth on solid medium after 96h (data not shown), we were able to achieve growth on LB agar supplemented with sorbitol, acetate, succinate and 1- or 2.5 mM GSSG after 48h (Figure 6.7E). Higher levels of GSSG (5-10 mM) resulted in no visible growth on solid medium.

Pre-cultures of NADHaux in liquid M9 minimal medium also displayed expected growth behaviour; NADHaux supplemented with acetate, succinate, and 5 mM octanal (Figure 6.7F) or 10 mM L-glutamine (Figure 6.7G) were able to form biomass when supplied with 10 mM sorbitol, but not when supplied 10 mM fructose. Taken together, this was evidence that careful balancing of both $NADP^+$/NADPH and NAD^+/NADH levels was required to successfully establish growth of NADHaux using an NADH source coupled to carbon metabolism. Unfortunately, we were unable to establish reproducible growth of NADHaux using sorbitol as sole source of NADH in plate reader assays.

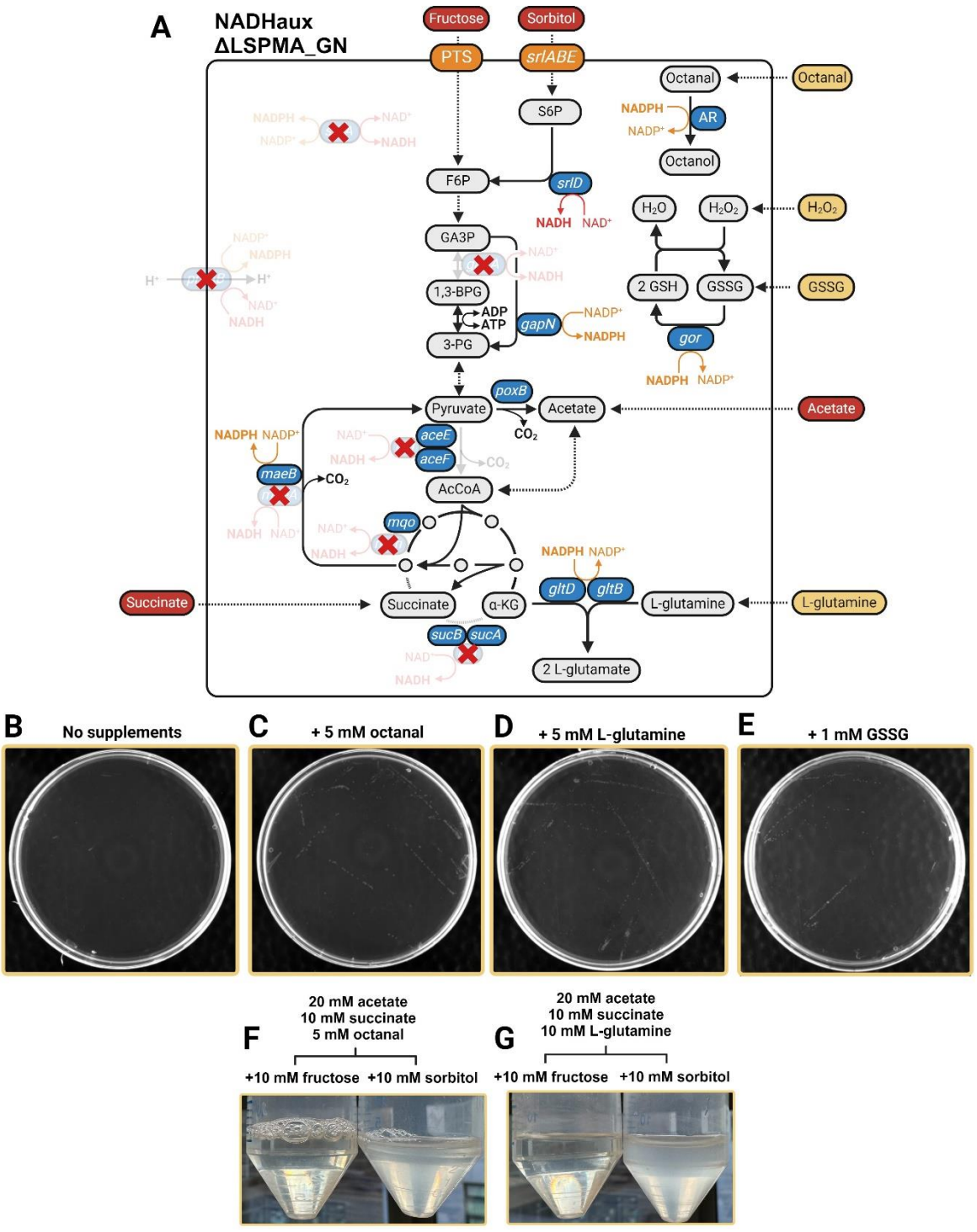


Figure 6.7. Growth characterization of NADHaux reveals auxotrophy for NADP⁺. (A) Schematic of cell metabolism indicating the tested NADPH-consuming pathways. Growth of NADHaux was performed on LB agar with 10 mM sorbitol, 20 mM acetate, 10 mM succinate and **(B)** no additional supplements, **(C)** 5 mM octanal, **(D)** 5 mM L-glutamine or **(E)** 1 mM GSSG, and imaged after 48h. Precultures of NADHaux grown in M9 minimal medium with acetate, succinate and either **(F)** octanal or **(G)** L-glutamine after 144 hours, demonstrating that sorbitol supplementation rescued growth, whereas the addition of fructose did not.

6.4.6 Growth rescue of NADHaux via beta oxidation

The oxidation of sugar alcohols such as sorbitol represents one endogenous pathway that can be used to rescue the NADH-auxotroph. However, to demonstrate the flexibility of the platform, we attempted to use the fatty acid beta oxidation pathway as an alternative source of reducing power. Fatty acid catabolism involves *fadL* and *fadD*-mediated import of exogenous fatty acids, resulting in the formation of a fatty acyl-CoA product. Through the activity of *fadE, fadB* and *fadA* these acyl-CoA's are subsequently degraded in cycles, resulting in the release of one acetyl-CoA, NADH and $FADH_2$ per cycle (Figure 6.8A).

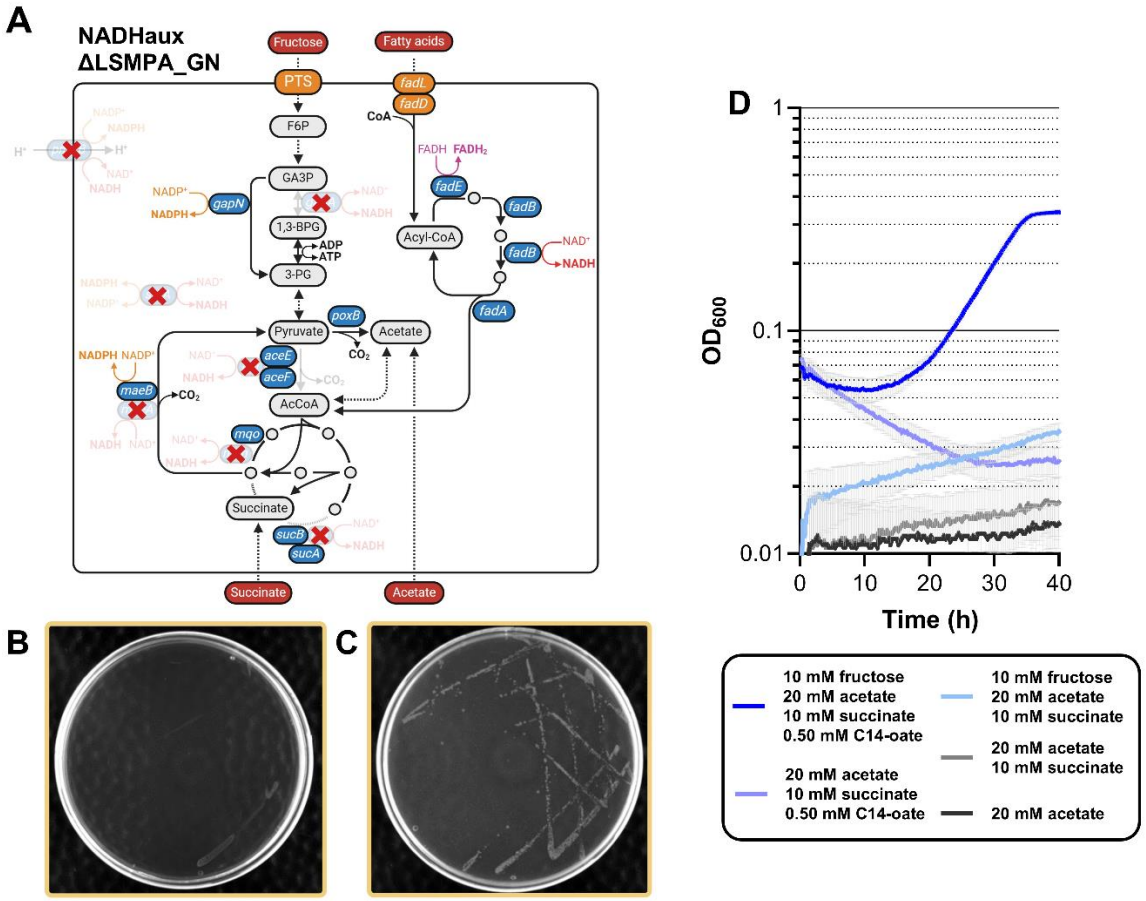


Figure 6.8. Growth rescue of NADHaux via beta oxidation. (**A**) The beta oxidation pathway catabolises fatty acids to acetyl-CoA, releasing reducing power in the form of $FADH_2$ and NADH. (**B)** We observed no growth of NADHaux after 48h of growth on LB supplemented with 10 mM fructose, 20 mM acetate and 10 mM succinate. (**C**) With additional supplementation of 1 mM tetradecanoate colonies formed within 48h. (**D**) NADHaux was only able to grow in liquid minimal medium when supplemented with fatty acids and fructose in addition to acetate and succinate. Growth curve datapoints represent the averages of biological triplicate measurements and the errors bars (light grey) the respective standard deviations.

We pre-cultured the NADH-auxotroph in liquid LB supplemented with 1 mM tetradecanoate (C_{14} fatty acid), 20 mM acetate, 10 mM succinate and 10 mM fructose, subsequently streaking this culture on LB agar plates supplemented with fructose, acetate, succinate and either with or without 1 mM tetradecanoate. After 48h, we observed significant growth on the fatty acid-supplemented plate (Figure 6.8B, C). To demonstrate that fatty acids and not LB medium components were solely responsible for the observed growth, we also pre-cultured NADHaux in liquid M9 minimal medium supplemented with fructose, acetate, succinate, and 1 mM tetradecanoate, and subsequently assessed its growth in various conditions. Medium lacking fructose did not support growth, since substrates entering upper glycolysis are essential due to the irreversible *gapN* reaction. Only the NADHaux strain supplemented with fructose, acetate, succinate and tetradecanoate displayed clear growth within 40 hours (Figure 6.8D). Note that in the condition containing fructose, acetate, and succinate but lacking fatty acids a modest increase in OD_{600} can be seen over the course of the growth assay. Normally we would not expect this behaviour, but it is likely that the fatty acid-containing preculture medium resulted in the expression of beta oxidation pathway enzymes. It is therefore possible that the observed growth was caused by degradation of endogenous fatty acids by beta oxidation, essentially using fatty acid biosynthesis and subsequent fatty acid degradation as a multi-pathway transhydrogenase to transfer reducing power from NADPH (fatty acid biosynthesis) to NAD^+ (beta oxidation). This activity could in theory be blocked by knockout of either *fadD* or *fadE*, however due to the relatively low growth of NADHaux in this condition we did not deem it essential to perform this knockout, especially since preculture conditions normally lacked fatty acid supplementation.

6.4.7 Growth rescue of NADHaux by formate oxidation

To demonstrate that NADHaux could also be rescued by heterologous enzyme activity, we turned to using formate oxidation as a proof-of-principle rescue pathway. Guided by the ΔLS_GN growth-rescue results, we chose to use the strong constitutively expressed formate dehydrogenase construct pL1_JRN067_FG. We transformed NADHaux with this plasmid and

characterised its growth in minimal medium, with supplementation of fructose, acetate, formate, and/or succinate as indicated in the text and figures.

In the presence of 10 mM fructose and 20 mM acetate, NADHaux growth could be rescued through the supplementation of 20-, 40-, and 80 mM formate (Figure 6.9A). The lag phase for this growth was considerable, only displaying exponential growth after around 96h of cultivation, and reaching maximum growth rates of 0.071-, 0.11- and 0.13 h^{-1} (20-, 40-, and 80 mM formate, respectively). We observed no significant growth of strains supplemented with 0-10 mM formate, nor strains lacking fructose or acetate. Since the supplementation of succinate was previously shown to improve growth of ΔLSMPA_GC, we supplemented NADHaux pL1_JRN067_FG with an additional 10 mM succinate besides fructose and acetate. With the supplementation of succinate, the growth of NADHaux was significantly more sensitive to the concentration of formate in the medium (Figure 6.9B). Clearly discernible growth took place within 24h after assay initiation and with as little as 1 mM formate supplementation, while growth saturation occurred with addition of 10-80 mM formate. Growth rates were also significantly higher than that NADHaux lacking succinate, with a 1.9-fold higher μ_{max} at 20 mM formate supplementation, and 1.3-fold higher μ_{max} when supplied 40 mM formate. When using 80 mM formate, μ_{max} values between succinate-lacking and succinate-supplemented NADHaux were not significantly different. As before, removing acetate and succinate, removing fructose, or not providing any formate, led to no growth over the course of the assay. When comparing the growth sensitivity of the original ΔL strain and NADHaux to the concentration of formate we observed that in conditions without succinate, ΔL and NADHaux performed similarly (Figure 6.9D and E). With succinate supplementation NADHaux displayed increased biomass formation per unit formate and higher μ_{max} values until growth saturation occurred. It would be interesting to investigate if succinate supplementation in ΔL cultures would result in a similar increase in growth rates and formate sensitivity.

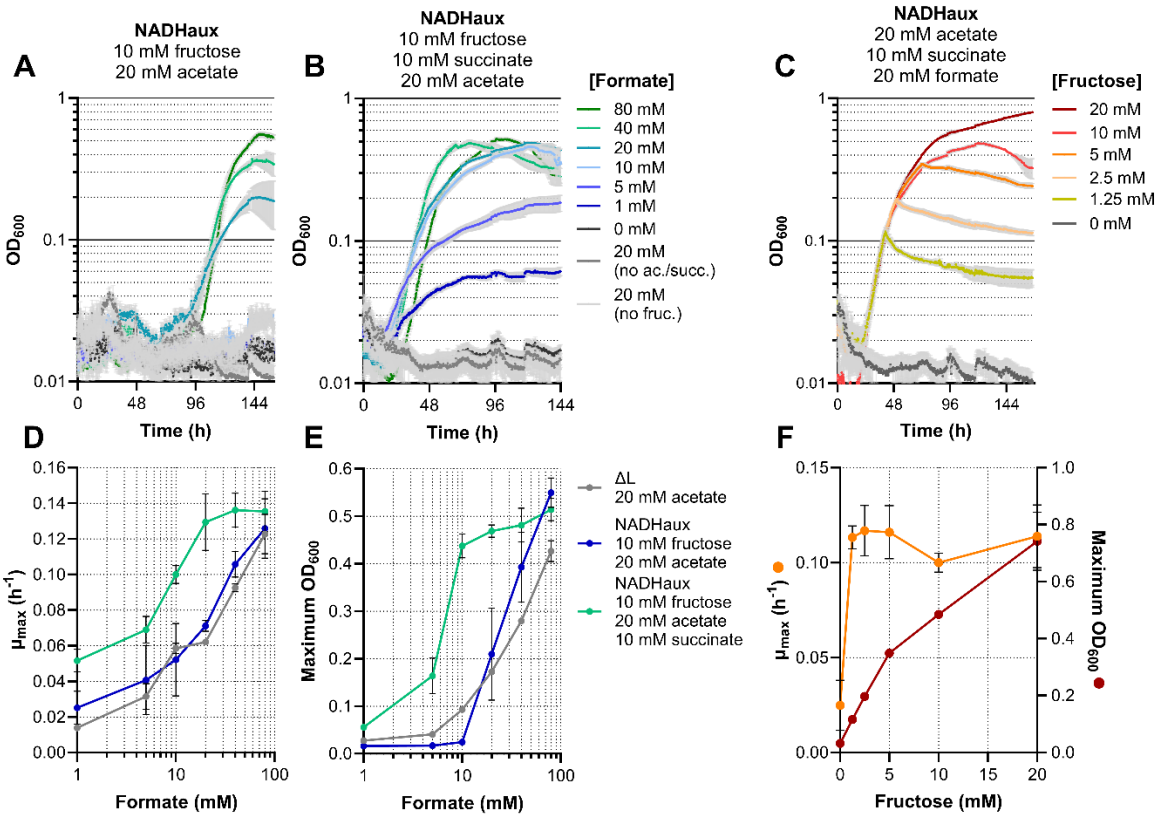


Figure 6.9. Formate-dependent growth of NADHaux. All NADHaux strains carried pL1_JRN067_FG for constitutive expression of NAD$^+$-dependent FDH. NADHaux was supplemented with various concentrations of formate in addition to **(A)** fructose and acetate or **(B)** fructose, acetate, and succinate. **(C)** The growth dependence of NADHaux on fructose when supplied with acetate, succinate and formate was also established. Comparisons of **(D)** maximum growth rates and **(E)** maximum OD_{600} values between ΔL grown on acetate, NADHaux grown on fructose and acetate, and NADHaux grown on fructose, acetate, and succinate, as a function of formate concentration. **(F)** The growth rate of NADHaux was not affected by fructose concentration, but biomass formation displayed a dependence on fructose concentration. All datapoints represent averages of triplicate measurements, and error bars represent their respective standard deviations.

The growth saturation of succinate-supplemented NADHaux with 10-80 mM formate addition was assumed to be related to insufficient fructose availability. The presence of the irreversible *gapN* reaction demanded a constant flux of fructose entering metabolism to ensure essential upper glycolytic compounds could be biosynthesised. To assess whether fructose was limiting, we used a growth-saturating concentration of formate (20 mM), acetate (20 mM) and succinate (10 mM), while varying the fructose concentration from 0-20 mM (Figure 6.9C). As expected, growth with 10 mM fructose resulted in a similar maximum OD_{600} as observed in the

previous assay (~0.5) and decreasing fructose concentration to 5 mM or lower also decreased the maximum achievable OD_{600}. In contrast, a 20 mM fructose concentration increased the maximum OD_{600} to around 0.8, although growth was still ongoing when the assay was stopped, indicating that fructose availability was indeed limiting growth. Maximum growth rates were identical using any concentration of fructose, indicating that growth rates were not limited by fructose availability, while final OD_{600} values were dependent on fructose concentration (Figure 6.9F).

6.4.8 An NADHaux 'cheater' carries mutations in NADH:ubiquinone oxidoreductase

In our attempts to grow the NADHaux strain using glycolytic carbon sources in addition to acetate and succinate, we noticed the occurrence of a 'cheating' *E. coli* strain that was able to grow on fructose and glycerol. In contrast to sorbitol these carbon sources were expected to be selective since their catabolism does not directly lead to NAD^+ reduction. This mutant, dubbed 'M1', demonstrated rapid growth on fructose and glycerol as well as sorbitol (Figure 6.10A). Whole genome sequencing of M1 confirmed the deletion of *mdh, maeA, sthA, pntAB* and *lpd*, and confirmed the integrity of the *gapA::gapN* gene replacement. However, sequencing revealed an E183A mutation in *nuoF* (NADH:ubiquinone oxidoreductase), which is the cofactor-interacting subunit of the NADH dehydrogenase complex I (NDH-1) involved in the first step of NADH-driven oxidative phosphorylation. The same mutation was identified in previous work studying an NADPH-overproducing *E. coli* in selective conditions (Auriol et al., 2011). The authors linked the $nuoF^{E183A}$ mutant to a broadening of the cofactor specificity of *nuoF* to include NADPH as well as NADH. This mutation therefore allowed oxidative phosphorylation to operate using NADPH as opposed to just NADH. This is a significant alteration of *E. coli* redox metabolism, as it essentially removes the requirement for NADH flux to drive oxidative phosphorylation. Considering the NADH-deficient cellular context of NADHaux, it is clear why this mutation was selected for, as it now allows oxidative ATP flux without requiring NADH (Figure 6.10B and C).

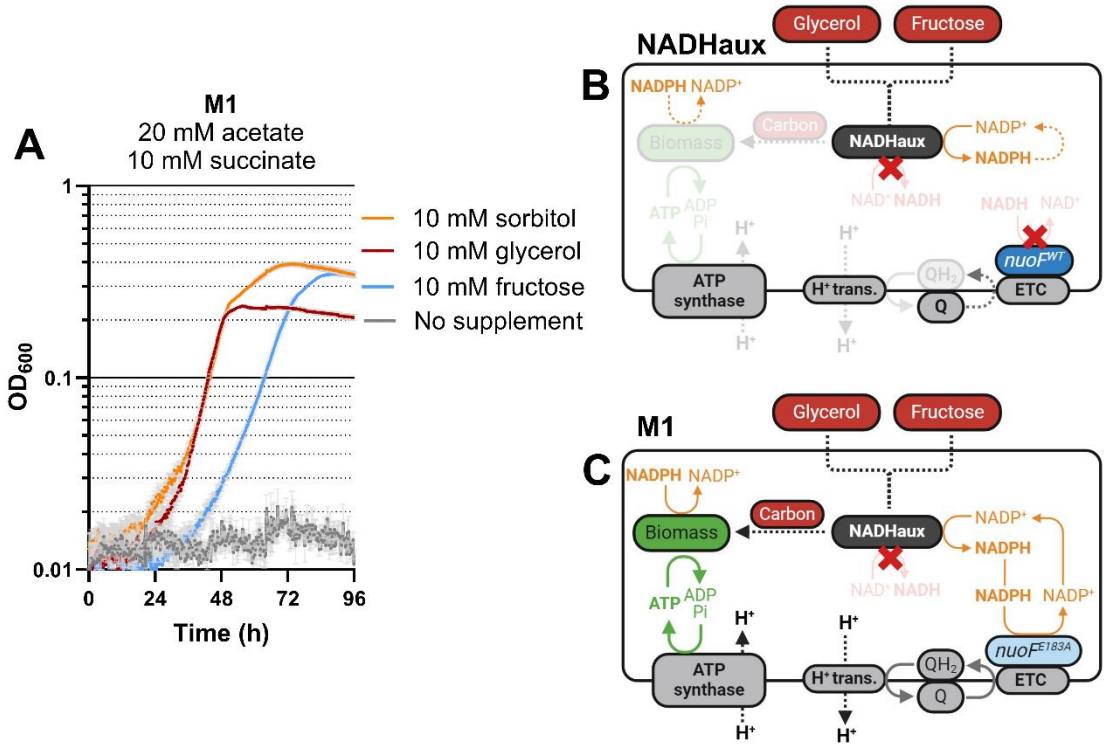


Figure 6.10. Growth characterization of *nuoF*-mutant M1. **(A)** M1 was grown using acetate, succinate and either sorbitol, glycerol, or fructose, resulting in growth on all glycolytic carbon sources. **(B)** Schematic representation of the deficiency in oxidative phosphorylation in the NADHaux mutant. Since NADH is unavailable, electrons cannot enter the electron transport chain (ETC) via NADH dehydrogenase complex I. This leads to a deficiency in proton translocation and concomitant ATP synthase flux, resulting in growth cessation. **(C)** Schematic representation of how $NADP^+$-dependent *nuoF*^{E183A} results in growth rescue of NADHaux. The NADPH generated in central metabolism can now be used to drive oxidative phosphorylation, resulting in ATP flux and biomass formation. The source of NADH-generating flux in M1 as required for secondary metabolite biosynthesis is unclear. Note that the diagram does not indicate the correct stoichiometries involved in oxidative phosphorylation.

We identified two other mutations in M1 that were probably linked to the overproduction of α-ketoglutarate in NADHaux, caused by the inactivation of the αKGDH complex. The first was a D229G mutation in citrate synthase (*gltA*), located in the binding pocket linked to α-ketoglutarate-mediated allosteric inhibition of GltA. Mutations in this region have been shown to strongly decrease α-ketoglutarate binding affinity (Anderson & Duckworth, 1988). Another change occurred in the intergenic region between *kgtP* and *rrfG* (T>C, -321/+2), the α-KG:proton symporter and 5S ribosomal RNA of the rrnG operon, respectively. Although this mutation's exact effect is unknown, a change in the expression level of the α-KG symporter would not be unexpected.

6.5 Discussion

This study was based on previous work investigating the Δ*lpd* mutant, a strain that displayed auxotrophy to all forms of reducing power when grown on acetate. However, *Δlpd* could not be used to discriminate between NADH- or NADPH-generating reactions nor could it be grown selectively using glycolytic carbon sources. In this work we extended the Δ*lpd* mutant into NADHaux, a strain displaying auxotrophy specifically to NADH and one that can be grown selectively using glycolytic substrates. It can therefore be used for engineering enzymes or pathways involved in sugar catabolism or other metabolic pathways requiring high flux from upper glycolysis. We demonstrated that NADHaux growth can be rescued through a variety of NADH-generating reactions, be it endogenous (beta oxidation, sorbitol catabolism) or exogenous (formate oxidation). We propose therefore that NADHaux can be used as a flexible platform strain for the selection of specifically NADH-generating enzymes and pathways.

Besides NADHaux, the characterization of ΔLSPMA_GC shed light on the substrate specificity of *gapC* from *Clostridium acetobutylicum*. There is conflicting data regarding this specificity, several studies (including the first characterization of *gapC*) reported that GapC is at least partially NAD^+/NADH specific (Iddar et al., 2002; Liu et al., 2015; Schreiber et al., 1999), while other studies reported increased yields of NADPH-dependent pathways when *gapC* is expressed (Huang et al., 2011; Jiang et al., 2013; Martinez et al., 2008), although these two opposing findings are not mutually exclusive. Indeed, from the data gathered in our study taken together with previous work using *gapC* in a redox cofactor auxotroph (Zhang et al., 2018), we conclude that *gapC* cannot be stringently $NADP^+$/NADPH specific and must have a level of affinity for NAD^+/NADH that is physiologically relevant. The development of a strictly $NADP^+$/NADPH-specific GAPDH would provide metabolic engineers with a tool to improve solely NADPH yields through glycolysis, without having to resort to non-phosphorylating GapN enzymes which, while highly $NADP^+$/NADPH-specific, reduce the net ATP yield of glycolysis.

The NADHaux strain also displayed $NADP^+$ auxotrophy when grown on glycolytic substrates, an effect we attribute to the NADPH-generating glycolytic pathway and a lack of

transhydrogenase activity. Normal growth of NADHaux on rich medium containing NADH-generating sorbitol was made possible by addition of metabolites that stimulated NADPH-oxidation. This requirement for an NADPH sink has been previously observed in studies describing strains that overproduce NADPH (Auriol et al., 2011; Kramer et al., 2020; Maxel et al., 2020; Zhang et al., 2018). Interestingly, this means that NADHaux represents a dual-use enzyme engineering platform as it can exhibit both NADH and $NADP^+$ auxotrophy depending on the specific culturing conditions used. This may make it an interesting strain for engineering enzymes or pathways that produce NADH, consume NADPH, or both simultaneously.

One common observation across our work was that succinate significantly improved the growth of NADH deficient strains, including ΔLSPMA_GC and NADHaux. There may be several reasons for this, the first being that the inactivation of the αKGDH complex results in a decrease in succinyl-CoA flux. This metabolite is required for the biosynthesis of several essential metabolites, including methionine and lysine. Previous work has shown that knockouts of *sucAB* or *sucCD,* but not *sucABCD*, can be made while still retaining cell viability, although these mutants displayed longer lag phases (Yu et al., 2006). Since a ***Δlpd*** mutant essentially represents a ***ΔsucAB*** mutant in terms of TCA cycle flux, it is possible that succinate supplementation provided the mutants with a higher flux towards succinyl-CoA, accelerating growth. Alternatively, succinate supplementation could be contributing to the cell's energy state through the activity of succinate:quinone oxidoreductase (*sdhABCD*). This complex is active in aerobic conditions, and catalyses the oxidation of succinate to fumurate, with the concomitant reduction of ubiquionine to ubiquinol in the electron transport chain (Maher et al., 2018). The additional flux through this reaction as a result of succinate supplementation could be contributing to ATP synthesis. In addition, due to the deletion of *mdh* oxaloacetate-forming flux must be generated either through the alternative membrane-bound malate:quinone-oxidoreductase (*mqo*) or via the CO_2-fixing PEP carboxylase. The *mqo* enzyme is known to be able to carry aerobic flux and its transcription is induced by the presence of acetate (Millard et al., 2023; Van der Rest et al., 2000). Flux through *mqo* results in the reduction of ubiquinone

to ubiquinol, and it is possible that succinate supplementation would cause a higher flux through this electron-generating reaction due to it entering the TCA cycle at the C_4 branch. Finally, although not directly related to succinate supplementation, pyruvate oxidase (*poxB*) activity is also ubiquinone-linked. This means that several potentially high flux reactions in NADHaux directly transfer electrons to the electron transport chain without requiring the NAD^+/NADH redox cofactor. Succinate supplementation may contribute to this flux, enabling strains such as NADHaux to reach a higher ATP yield per reducing equivalent being generated through e.g. formate- or sorbitol oxidation, resulting in improved growth. Further study would be required to identify exactly what effect succinate supplementation has on cell metabolism of these NADH deficient strains.

The whole genome sequencing results of mutant M1 provided evidence of a selective pressure to decrease the NADPH/$NADP^+$ ratio and increase ATP synthesis. We found an E183A mutation in *nuoF*, which alters its cofactor specificity to accept NADPH as well as NADH (Auriol et al., 2011; Morina et al., 2011). This change has a significant impact on the redox metabolism of NADHaux, since it meant that a) oxidative phosphorylation could be operated using NADPH, removing the requirement for NADH to generate ATP oxidatively, and b) more rapid oxidation of NADPH could occur enabling higher flux operation of *gapN*-dependent glycolysis. It seemed that NADHaux growth was limited not only by NADH availability, but also by the NADPH oxidation rate. In fact, we did not detect any mutations in oxidoreductases that would enhance NADH generation in the M1 mutant, suggesting that background NADH-generating reactions were sufficient to rescue NADHaux carrying the mutated NADPH-dependent *nuoF*^{E183A} (and the two αKG-related mutations).

Related to this, we do note that several native NADH-generating pathways remained intact in NADHaux, such as pathways for the biosynthesis of several amino acids (e.g. tyrosine, histidine, threonine, serine), cofactors (pyroxidal-5-phosphate), beta oxidation, as well as other secondary metabolic pathways (the GABA shunt). The 3-phosphoglycerate dehydrogenase (*serA*) in particular is a probable candidate for NADH flux since it branches

off directly from the relatively high-flux glycolytic pathway. Although *serA* is highly allosterically regulated by its terminal product serine (Al-Rabiee et al., 1996), flux through *serA* was increased upon expression of an NADH oxidase as a result of heightened glycolytic flux (Holm et al., 2010). It is possible that the flux through these secondary metabolite biosynthesis pathways is sufficient to generate a 'baseline' ratio of NAD^+/NADH that does not require additional NADH release directly from substrate catabolism, at least not when oxidative phosphorylation is driven by NADPH. The observed growth of M1 on all tested glycolytic carbon sources provided evidence for this, and we therefore conclude that NADH-generating flux through central NADH-generating pathways is no longer required if oxidative phosphorylation is uncoupled from NADH. We do note that the growth of M1 was achieved with supplementation of acetate and succinate along with glycolytic carbon sources. Considering the significant effect that succinate addition had on formate-dependent growth of NADHaux, and on sorbitol-dependent growth of ΔLSPMA_GC, removing these additional carbon sources could provide new insights into the growth of M1. More detailed modelling, mutant generation, growth assays and possibly carbon labelling experiments could elucidate the exact metabolic mechanisms behind the growth and redox state of M1.

While remaining NADH-generating pathways represent potential risks for the evolution of 'cheater' mutants, they also represent potential target pathways for growth-coupled engineering efforts. Engineering the production of secondary metabolites such as amino acids often requires complex combinatorial pathway engineering and targeted mutagenesis due to the highly regulated multi-enzyme pathways that are involved (Juminaga et al., 2012; Wu et al., 2020). With a growth-coupled approach using a cofactor auxotroph such as the one described in this work, high throughput combinatorial gene expression optimization and directed evolution could be performed without requiring the measurement of product synthesis by discrete variants.

6.6 Conclusion

Enzyme engineering platforms are required to optimise biocatalysts for industrial use. Redox cofactors are ubiquitous in Nature and involved in many enzymatic reactions. In this work, an *E. coli* strain was developed that displayed auxotrophy for the reduced state of the NAD^+/NADH cofactor, named NADHaux. This strain can be used to growth-couple NADH-generating enzymes or pathways, providing a platform for their high-throughput engineering. This work also provided new insights into the extremes of redox metabolism in *E. coli,* by identification of a mutant lacking all canonical NADH-generating pathways while using NADPH to drive oxidative phosphorylation.

Acknowledgements

I would like to thank Yutong Yin, Reza Rohani and Simona Della Valle for their insightful discussions during the experimental work involved in this study. This work was supported by the BBSRC (grant BB/M011224/1) and the Oxford Interdisciplinary Bioscience Doctoral Training Partnership. Figure schematics were created using Biorender.com.

6.7 References

Al-Rabiee, R., Zhang, Y., & Grant, G. A. (1996). The mechanism of velocity modulated allosteric regulation in D-3-phosphoglycerate dehydrogenase. Site-directed mutagenesis of effector binding site residues. *Journal of Biological Chemistry*, *271*(38), 23235–23238. https://doi.org/10.1074/jbc.271.38.23235

Alter, T. B., & Ebert, B. E. (2019). Determination of growth-coupling strategies and their underlying principles. *BMC Bioinformatics 2019 20:1*, *20*(1), 1–17. https://doi.org/10.1186/S12859-019-2946-7

Anderson, D. H., & Duckworth, H. W. (1988). In vitro mutagenesis of Escherichia coli citrate synthase to clarify the locations of ligand binding sites. *Journal of Biological Chemistry*, *263*(5), 2163–2169. https://doi.org/10.1016/S0021-9258(18)69186-0

Auriol, C., Bestel-Corre, G., Claude, J. B., Soucaille, P., & Meynial-Salles, I. (2011). Stress-induced evolution of Escherichia coli points to original concepts in respiratory cofactor selectivity. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(4), 1278–1283. https://doi.org/10.1073/PNAS.1010431108

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, *2*, 2006.0008. https://doi.org/10.1038/MSB4100050

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., & Milo, R. (2011). The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, *50*(21), 4402–4410. https://doi.org/10.1021/BI2002289

Chen, J., Wang, Y., Zheng, P., & Sun, J. (2022). Engineering synthetic auxotrophs for growth-coupled directed protein evolution. *Trends in Biotechnology*. https://doi.org/10.1016/J.TIBTECH.2022.01.010

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/BIOINFORMATICS/BTY560

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

De Graef, M. R., Alexeeva, S., Snoep, J. L., & Teixeira De Mattos, M. J. (1999). The Steady-State Internal Redox State (NADH/NAD) Reflects the External Redox State and Is Correlated with Catabolic Adaptation in Escherichia coli. *Journal of Bacteriology*, *181*(8), 2351. https://doi.org/10.1128/JB.181.8.2351-2357.1999

Deatherage, D. E., & Barrick, J. E. (2014). Identification of mutations in laboratory evolved microbes from next-generation sequencing data using breseq. *Methods in Molecular Biology (Clifton, N.J.)*, *1151*, 165. https://doi.org/10.1007/978-1-4939-0554-6_12

Della Seta, F., Boschi-Muller, S., Vignais, M. L., & Branlant, G. (1997). Characterization of Escherichia coli strains with gapA and gapB genes deleted. *Journal of Bacteriology*, *179*(16), 5218–5221. https://doi.org/10.1128/JB.179.16.5218-5221.1997

Ebrahim, A., Lerman, J. A., Palsson, B. O., & Hyduke, D. R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology 2013 7:1*, *7*(1), 1–6. https://doi.org/10.1186/1752-0509-7-74

Friedrich, T. (1998). The NADH:ubiquinone oxidoreductase (complex I) from Escherichia coli. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1364*(2), 134–146. https://doi.org/10.1016/S0005-2728(98)00024-3

Goodall, E. C. A., Robinson, A., Johnston, I. G., Jabbari, S., Turner, K. A., Cunningham, A. F., Lund, P. A., Cole, J. A., & Henderson, I. R. (2018). The Essential Genome of Escherichia coli K-12. *MBio*, *9*(1). https://doi.org/10.1128/MBIO.02096-17

Grenier, F., Matteau, D., Baby, V., & Rodrigue, S. (2014). Complete Genome Sequence of Escherichia coli BW25113. *Genome Announcements*, *2*(5), 1038–1052. https://doi.org/10.1128/GENOMEA.01038-14

Holm, A. K., Blank, L. M., Oldiges, M., Schmid, A., Solem, C., Jensen, P. R., & Vemuri, G. N. (2010). Metabolic and Transcriptional Response to Cofactor Perturbations in Escherichia coli. *The Journal of Biological Chemistry*, *285*(23), 17498. https://doi.org/10.1074/JBC.M109.095570

Huang, M., Wang, Y., Liu, J., & Mao, Z. (2011). Multiple Strategies for Metabolic Engineering of Escherichia coli for Efficient Production of Coenzyme Q10. *Chinese Journal of Chemical Engineering*, 19(2), 316–326. https://doi.org/10.1016/S1004-9541(11)60171-7

Iacometti, C., Marx, K., Hönick, M., Biletskaia, V., Schulz-Mirbach, H., Dronsella, B., Satanowski, A., Delmas, V. A., Berger, A., Dubois, I., Bouzon, M., Döring, V., Noor, E., Bar-Even, A., & Lindner, S. N. (2022). Activating Silent Glycolysis Bypasses in Escherichia coli. *BioDesign Research*, *2022*. https://doi.org/10.34133/2022/9859643

Iddar, A., Valverde, F., Serrano, A., & Soukri, A. (2002). Expression, purification, and characterization of recombinant nonphosphorylating NADP-dependent glyceraldehyde-3-phosphate dehydrogenase from Clostridium acetobutylicum. *Protein Expression and Purification*, 25(3), 519–526. https://doi.org/10.1016/S1046-5928(02)00032-3

Iverson, S. V., Haddock, T. L., Beal, J., & Densmore, D. M. (2016). CIDAR MoClo: Improved MoClo Assembly Standard and New *E. coli* Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synthetic Biology*, *5*(1), 99–103. https://doi.org/10.1021/ACSSYNBIO.5B00124

Jensen, S. I., Lennen, R. M., Herrgård, M. J., & 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), 17874. https://doi.org/10.1038/srep17874

Jiang, L. Y., Zhang, Y. Y., Li, Z., & Liu, J. Z. (2013). Metabolic engineering of Corynebacterium glutamicum for increasing the production of l-ornithine by increasing NADPH availability. *Journal of Industrial Microbiology and Biotechnology*, 40(10), 1143–1151. https://doi.org/10.1007/S10295-013-1306-2

Juminaga, D., Baidoo, E. E. K., Redding-Johanson, A. M., Batth, T. S., Burd, H., Mukhopadhyay, A., Petzold, C. J., & Keasling, J. D. (2012). Modular engineering of L-tyrosine production in Escherichia coli. *Applied and Environmental Microbiology*, *78*(1), 89–98. https://doi.org/10.1128/AEM.06017-11

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/NAR/28.1.27

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O., & Lewis, N. E. (2016). BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, *44*(D1), D515–D522. https://doi.org/10.1093/NAR/GKV1049

Kramer, L., Le, X., Rodriguez, M., Wilson, M. A., Guo, J., & Niu, W. (2020). Engineering Carboxylic Acid Reductase (CAR) through a Whole-Cell Growth-Coupled NADPH Recycling Strategy. *ACS Synthetic Biology*, *9*(7), 1632–1637. https://doi.org/10.1021/ACSSYNBIO.0C00290

Li, M., Ho, P. Y., Yao, S., & Shimizu, K. (2006). Effect of lpdA gene knockout on the metabolism in Escherichia coli based on enzyme activities, intracellular metabolite concentrations and metabolic flux analysis by 13C-labeling experiments. *Journal of Biotechnology*, *122*(2), 254–266. https://doi.org/10.1016/J.JBIOTEC.2005.09.016

Lin, P. P., Jaeger, A. J., Wu, T. Y., Xu, S. C., Lee, A. S., Gao, F., Chen, P. W., & Liao, J. C. (2018). Construction and evolution of an Escherichia coli strain relying on nonoxidative glycolysis for sugar catabolism. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(14), 3538–3546. https://doi.org/10.1073/PNAS.1802191115

Liu, J., Qi, H., Wang, C., & Wen, J. (2015). Model-driven intracellular redox status modulation for increasing isobutanol production in Escherichia coli. *Biotechnology for Biofuels*, 8(1), 1–15. https://doi.org/10.1186/S13068-015-0291-2

Lindner, S. N., Calzadiaz Ramirez, L., Krü, J. L., Yishai, O., Belkhelfa, S., He, H., Bouzon, M., Dö, V., & Bar-Even, A. (2018). *NADPH-Auxotrophic E. coli: A Sensor Strain for Testing in Vivo Regeneration of NADPH*. https://doi.org/10.1021/acssynbio.8b00313

Longwell, C. K., Labanieh, L., & Cochran, J. R. (2017). High-throughput screening technologies for enzyme engineering. *Current Opinion in Biotechnology*, *48*, 196–202. https://doi.org/10.1016/J.COPBIO.2017.05.012

Maher, M. J., Herath, A. S., Udagedara, S. R., Dougan, D. A., & Truscott, K. N. (2018). Crystal structure of bacterial succinate:quinone oxidoreductase flavoprotein SdhA in complex with its assembly factor SdhE. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(12), 2982–2987. https://doi.org/10.1073/PNAS.1800195115

Martínez, I., Zhu, J., Lin, H., Bennett, G. N., & 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. https://doi.org/10.1016/j.ymben.2008.09.001

Maxel, S., Aspacio, D., King, E., Zhang, L., Acosta, A. P., & Li, H. (2020). A Growth-Based, High-Throughput Selection Platform Enables Remodeling of 4-Hydroxybenzoate Hydroxylase Active Site. *ACS Catalysis*, *10*(12), 6969–6974. https://doi.org/10.1021/ACSCATAL.0C01892

Millard, P., Enjalbert, B., Uttenweiler-Joseph, S., Portais, J. C., & Létisse, F. (2021). Control and regulation of acetate overflow in escherichia coli. *ELife*, *10*. https://doi.org/10.7554/ELIFE.63661

Millard, P., Gosselin-Monplaisir, T., Uttenweiler-Joseph, S., & Enjalbert, B. (2023). Acetate is a beneficial nutrient for *E. coli* at low glycolytic flux . *The EMBO Journal*, *42*(15). https://doi.org/10.15252/EMBJ.2022113079

Morina, K., Schulte, M., Hubrich, F., Dörner, K., Steimle, S., Stolpe, S., & Friedrich, T. (2011). Engineering the respiratory complex I to energy-converting NADPH:Ubiquinone oxidoreductase. *Journal of Biological Chemistry*, *286*(40), 34627–34634. https://doi.org/10.1074/jbc.M111.274571

Nielsen, J. R., Weusthuis, R. A., & Huang, W. E. (2023). Growth-coupled enzyme engineering through manipulation of redox cofactor regeneration. *Biotechnology Advances*, 108102. https://doi.org/10.1016/J.BIOTECHADV.2023.108102

Novotny, M. J., Reizer, J., Esch, F., Saier, M. H., Grenier, F. C., Shiell, R., & Saier, M. H. (1984). Purification and properties of D-mannitol-1-phosphate dehydrogenase and D-glucitol-6-phosphate dehydrogenase from Escherichia coli. *Journal of Bacteriology*, *159*(3), 986–990. https://doi.org/10.1128/JB.159.3.986-990.1984

Orsi, E., Claassens, N. J., Nikel, P. I., & Lindner, S. N. (2021). Growth-coupled selection of synthetic modules to accelerate cell factory development. *Nature Communications 2021 12:1*, *12*(1), 1–5. https://doi.org/10.1038/s41467-021-25665-6

Orth, J. D., Fleming, R. M. T., & Palsson, B. Ø. (2010). Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide . *EcoSal Plus*, *4*(1). https://doi.org/10.1128/ECOSALPLUS.10.2.1

Sauer, U., Canonaco, F., Heri, S., Perrenoud, A., & Fischer, E. (2004). The Soluble and Membrane-bound Transhydrogenases UdhA and PntAB Have Divergent Functions in NADPH Metabolism of Escherichia coli*. *Journal of Biological Chemistry*, *279*(8), 6613–6619. https://doi.org/10.1074/JBC.M311657200

Schreiber, W., & Dürre, P. (1999). The glyceraldehyde-3-phosphate dehydrogenase of *Clostridium acetobutylicum*: Isolation and purification of the enzyme, and sequencing and localization of the gap gene within a cluster of other glycolytic genes. *Microbiology*, 145(8), 1839–1847. https://doi.org/10.1099/13500872-145-8-1839

Sellés Vidal, L., Murray, J. W., & Heap, J. T. (2021). Versatile selective evolutionary pressure using synthetic defect in universal metabolism. *Nature Communications 2021 12:1*, *12*(1), 1–15. https://doi.org/10.1038/s41467-021-27266-9

Spaans, S. K., Weusthuis, R. A., van der Oost, J., & Kengen, S. W. M. (2015). NADPH-generating systems in bacteria and archaea. In *Frontiers in Microbiology* (Vol. 6, Issue JUL, p. 742). Frontiers Research Foundation. https://doi.org/10.3389/fmicb.2015.00742

Takeno, S., Hori, K., Ohtani, S., Mimura, A., Mitsuhashi, S., & Ikeda, M. (2016). l-Lysine production independent of the oxidative pentose phosphate pathway by Corynebacterium glutamicum with the Streptococcus mutans gapN gene. *Metabolic Engineering*, *37*, 1–10. https://doi.org/10.1016/J.YMBEN.2016.03.007

Van der Rest, M. E., Frank, C., & Molenaar, D. (2000). Functions of the membrane-associated and cytoplasmic malate dehydrogenases in the citric acid cycle of Escherichia coli. *Journal of Bacteriology*, *182*(24), 6892–6899. https://doi.org/10.1128/JB.182.24.6892-6899.2000

Von Kamp, A., & Klamt, S. (2017). Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms. *Nature Communications 2017 8:1*, *8*(1), 1–10. https://doi.org/10.1038/ncomms15956

Wenk, S., Schann, K., He, H., Rainaldi, V., Kim, S., Lindner, S. N., & Bar-Even, A. (2020). An “energy-auxotroph” *Escherichia coli* provides an in vivo platform for assessing NADH regeneration systems. *Biotechnology and Bioengineering*, *117*(11), 3422–3434. https://doi.org/10.1002/bit.27490

Wu, H., Tian, D., Fan, X., Fan, W., Zhang, Y., Jiang, S., Wen, C., Ma, Q., Chen, N., & Xie, X. (2020). Highly Efficient Production of l-Histidine from Glucose by Metabolically Engineered Escherichia coli. *ACS Synthetic Biology*, *9*(7), 1813–1822. https://doi.org/10.1021/ACSSYNBIO.0C00163

Xiao, H., Bao, Z., & Zhao, H. (2015). High throughput screening and selection methods for directed enzyme evolution. *Industrial and Engineering Chemistry Research*, *54*(16), 4011–4020. https://doi.org/10.1021/IE503060A

Yu, B. J., Sung, B. H., Lee, J. Y., Son, S. H., Kim, M. S., & Kim, S. C. (2006). sucAB and sucCD are mutually essential genes in Escherichia coli. *FEMS Microbiology Letters*, *254*(2), 245–250. https://doi.org/10.1111/J.1574-6968.2005.00026.X

Zhang, L., King, E., Luo, R., & Li, H. (2018). Development of a High-Throughput, in Vivo Selection Platform for NADPH-Dependent Reactions Based on Redox Balance Principles. *ACS Synthetic Biology*, *7*(7), 1715–1721. https://doi.org/10.1021/acssynbio.8b00179

6.8 Supplementary data

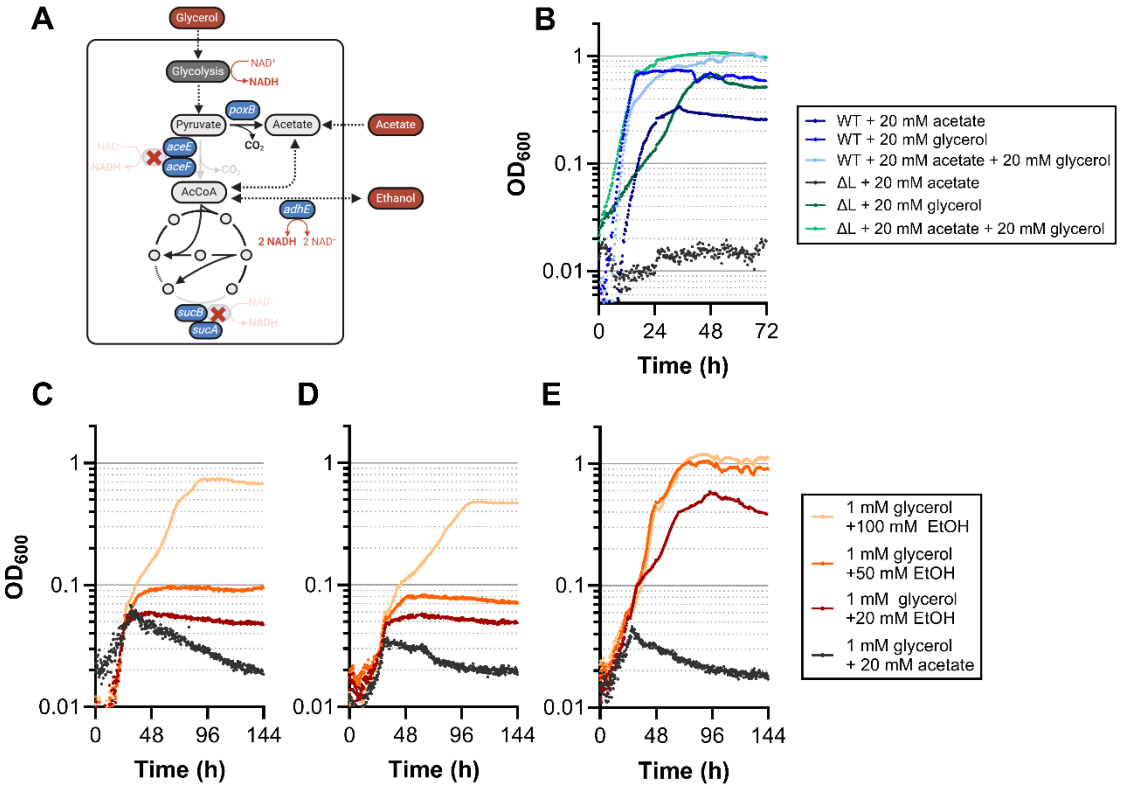


Figure S6.1. Ethanol growth-rescue of the Δ*lpd* mutant. **A)** Schematic representing central metabolic pathways with the effects of the Δ*lpd* mutation highlighted. **B)** Control growth assays involving *E. coli* BW25113 wildtype and the *E. coli* BW25113 Δ*lpd* mutant (ΔL), the latter of which is unable to grow on solely acetate. **C-F)** Growth of ΔL expressing the bifunctional alcohol dehydrogenase *adhE* under constitutive control of Anderson promoters **C)** J23116, **D)** J23107 or **E)** J23100. Growth assays were carried out using 1 mM glycerol and either 20 mM acetate or 20-, 50-, or 100-mM ethanol supplementation. Growth curves were averages of triplicate measurements, error bars representing standard deviations were removed for visual clarity since they did not exceed a 5% variance.

Table S6.1. Primers used in this study

Primer code	Sequence	Purpose
JRN075	ATGTTCCGGCAAACGAAA	Check *lpd* locus, rev
JRN114	ggaaaggtaaattgcagacg	Check *lpd* locus, fwd
JRN079	TCGTCCTGGCTCACCG	Check *sthA* locus, fwd
JRN080	GGGGATCAATTGGCTTAC	Check *sthA* locus, rev
JRN081	CAAAATTGCGCTTGAAC	Check *pntAB* locus, fwd
JRN082	TAACGGAGTAACATTTAGC	Check *pntAB* locus, rev
JRN137	gatctcgacgaaatggc	Check *gapA* locus, fwd
JRN138	gacgcagtcctttaacttc	Check *gapA* locus, rev
JRN162	catctctgctctggagac	Check *mdh* locus, fwd
JRN163	gaacggtagggtatattgtc	Check *mdh* locus, rev
JRN083	GTTCTGCATAGCAGGTG	Check *maeA* locus, fwd

JRN084	CTGCTTTATACTTGAGGCC	Check *maeA* locus, rev
JRN152	AAAATTGCAAAGTAAttgtgtaggctggagctgcttc	Amplify pKD3 for gapC insertion, fwd
JRN153	GGCGATTTTCGCCATtcgctcaagacgtgtaatgctg	Amplify pKD3 for gapC insertion, rev
JRN154	TTTGACATCAAATAAttgtgtaggctggagctgcttc	Amplify pKD3 for gapN insertion, fwd
JRN155	GTACTGTTTAGTCATtcgctcaagacgtgtaatgctg	Amplify pKD3 for gapN insertion, rev
JRN235	AGTGACGGCCTCAGCAGAGGCCGTCAGGGTTACAGAGCTTTCAGGATTGC tgtgtaggctggagctgcttc	Amplify cmR marker with *pntAB* flanks fwd, pKD3 binding
JRN236	CTGACACAGGCAAACCATCATCAATAAAACCGATGGAAGGGAATATCATGcatatgaatatcctccttag	Amplify cmR marker with *pntAB* flanks rev, pKD3 binding
JRN126	caagaatggatggccatttcgataaagttttaaaacaggcggtttaaacctgtgtaggctggagctgcttc	Amplify cmR marker with *sthA* flanks fwd, pKD3 binding
JRN127	tgttaccattctgttgcttttatgtataagaacaggtaagccctaccatgcatatgaatatcctccttag	Amplify cmR marker with *sthA* flanks rev, pKD3 binding
JRN156	attttacaggcaacctttattcactaacaaatagctggtggaatatatgGCGAAAATCGCCATTAATGG	Amplify *gapC* marker with *gapA* flanks fwd, gapC binding
JRN157	attttacaggcaacctttattcactaacaaatagctggtggaatatatgACTAAACAGTACAAAAATTATG	Amplify *gapN* marker with *gapA* flanks fwd, gapN binding
JRN129	CTCTTTTTAGATCACAGTGTCATCTCAACTTATTTGGAGATGTGAGCGATcatatgaatatcctccttag	Amplify *cmR* marker with *gapA* flanks rev, pKD3 binding
JRN160	gctccggtttttttattatccgctaatcaattacttattaacgaactcttctgtgtaggctggagctgcttc	Amplify cmR marker with *mdh* flanks fwd, pKD3 binding
JRN161	cagcggagcaacatatcttagtttatcaatataataaggagtttaggatgcatatgaatatcctccttag	Amplify cmR marker with *mdh* flanks rev, pKD3 binding
JRN348	CCCGGTAGCCTTCACTACCGGGCGCAGGCTTAGATGGAGGTACGGCGGTAtgtgtaggctggagctgcttc	Amplify cmR marker with *maeA* flanks fwd, pKD3 binding
JRN349	GGCCGACGCCCTGGCGGTAAAGCAAAGACGATAAAAGCCCCCCAGGGATGcatatgaatatcctcctta g	Amplify cmR marker with *maeA* flanks rev, pKD3 binding

Table S6.2. Synthesised DNA fragments used in this work

Pseudomonas sp. 101 formate dehydrogenase (*fdh*)	TAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAA GAAGGAGATATACATATGCATCATCACCATCACCACGCTAAAGTTCTGTGCGTTCTGTACGACG ACCCGGTTGACGGTTACCCGAAAACCTACGCTCGTGACGACCTGCCGAAAATCGACCACTACC CGGGTGGTCAGACCCTGCCGACCCCGAAAGCTATCGACTTCACCCCGGGTCAGCTGCTGGGT TCTGTTTCTGGTGAACTGGGTCTGCGTAAATACCTGGAATCTAACGGTCACACCCTGGTTGTTA CCTCTGACAAAGACGGTCCGGACTCTGTTTTCGAACGTGAACTGGTTGACGCTGACGTTGTTAT CTCTCAGCCGTTCTGGCCGGCTTACCTGACCCCGGAACGTATCGCTAAAGCTAAAAACCTGAA ACTGGCTCTGACCGCTGGTATCGGTTCTGACCACGTTGACCTGCAATCTGCTATCGACCGTAA CGTTACCGTTGCTGAAGTTACCTACTGCAACTCTATCTCTGTTGCTGAACACGTTGTTATGATGA TCCTGTCTCTGGTTCGTAACTACCTGCCGTCTCACGAATGGGCTCGTAAAGGTGGTTGGAACAT AGCTGACTGCGTAAGCCACGCTTACGACCTGGAAGCTATGCACGTTGGTACCGTTGCTGCTGG TCGTATCGGTCTGGCTGTTCTGCGTCGTCTGGCTCCGTTCGACGTTCACCTGCACTACACCGA CCGTCACCGTCTGCCGGAATCTGTTGAAAAAGAACTGAACCTGACCTGGCACGCTACCCGTGA AGACATGTACCCGGTTTGCGACGTTGTTACCCTGAACTGCCCGCTGCACCCGGAAACCGAACA CATGATCAACGACGAAACCCTGAAACTGTTCAAACGTGGTGCTTACATCGTTAACACCGCTCGT GGTAAACTGTGCGACCGTGACGCTGTTGCTCGTGCTCTGGAATCTGGTCGTCTGGCTGGTTAT GCGGGTGACGTGTGGTTCCCCCAGCCGGCTCCGAAAGACCACCCGTGGCGTACCATGCCGTA CAACGGTATGACCCCGCACATCTCTGGTACCACCCTGACCGCTCAGGCTCGTTACGCTGCTGG TACCCGTGAAATCCTGGAATGCTTCTTCGAAGGTCGTCCGATCCGTGACGAATACCTGATCGTT

	CAGGGTGGTGCTCTGGCTGGTACCGGTGCTCACTCTTACTCTAAAGGTAACGCTACCGGTGGT TCTGAAGAAGCTGCTAAATTCAAAAAAGCTGTTTAACTAGCATAACCCCTTGGGGCCTCTAAAC GGGTCTTGAGGGGTTTTTTG
Streptococcus mutans gapN	ATGACTAAACAGTACAAAAATTATGTAAATGGCGAATGGAAGCTGAGCGAGAACGAAATCAAAA TATACGAACCGGCTTCGGGTGCCGAATTAGGCTCAGTTCCGGCGATGTCCACCGAAGAAGTAG ATTACGTGTATGCGTCGGCTAAGAAAGCCCAACCAGCATGGCGTTCACTGAGTTACATTGAAC GCGCCGCATATCTCCATAAAGTGGCGGATATTCTGATGCGCGATAAGGAAAAGATTGGTGCAG TGCTGTCCAAAGAGGTGGCAAAGGGCTATAAATCAGCGGTTTCTGAAGTTGTAAGGACGGCAG AAATCATCAACTATGCCGCTGAAGAGGGTCTCCGGATGGAAGGTGAAGTGCTGGAAGGAGGC AGCTTCGAAGCTGCGAGCAAAAAGAAAATTGCCGTAGTGCGCCGCGAACCGGTCGGCCTTGT CCTGGCCATCAGCCCGTTTAATTACCCAGTTAACCTGGCGGGTAGTAAAATTGCACCTGCGTTG ATTGCGGGGAATGTCATTGCCTTCAAACCGCCGACCCAAGGCTCTATTTCGGGCCTGCTGCTG GCCGAAGCCTTTGCGGAAGCAGGCCTGCCCGCAGGTGTGTTTAACACGATTACCGGCCGTGG CTCGGAAATTGGCGATTATATCGTCGAACACCAGGCCGTCAATTTCATTAACTTCACGGGGAGC ACCGGTATTGGCGAACGTATCGGCAAAATGGCGGGAATGCGTCCGATTATGCTGGAGTTAGGT GGTAAAGACTCGGCCATTGTGCTCGAGGACGCGGATTTAGAACTAACCGCGAAAAACATCATC GCGGGGGCATTCGGTTATAGCGGCCAACGTTGTACCGCGGTCAAGCGCGTGTTGGTGATGGA GAGTGTTGCTGATGAGTTGGTTGAGAAAATTCGCGAGAAAGTGCTGGCCCTGACCATTGGCAA TCCAGAAGACGATGCAGATATTACTCCCTTAATCGATACTAAGTCTGCCGACTATGTTGAGGGA CTGATTAATGATGCGAACGACAAAGGGGCGGCGGCGCTGACTGAGATCAAGCGAGAAGGTAA CCTTATCTGCCCGATCTTATTTGACAAAGTGACCACCGATATGCGCCTTGCCTGGGAAGAACCC TTTGGCCCGGTACTCCCTATTATTCGTGTTACCTCCGTCGAGGAAGCGATTGAGATTAGCAACA AAAGCGAATACGGCCTTCAGGCCAGCATCTTTACAAACGATTTCCCTCGCGCTTTTGGTATAGC CGAACAGTTGGAAGTGGGAACAGTTCATATTAATAACAAAACGCAGCGTGGGACAGATAATTTC CCGTTTCTAGGCGCGAAAAAAAGTGGCGCCGGAATACAGGGAGTTAAGTACTCAATTGAAGCA ATGACCACGGTGAAAAGCGTAGTGTTTGACATCAAATAA
Clostridium acetobutylicum gapC	ATGGCGAAAATCGCCATTAATGGCTTTGGCCGTATTGGCCGATTGGCCTTACGTCGGATTCTG GAGGTACCTGGTCTGGAAGTAGTGGCCATCAACGATTTAACCGATGCCAAAATGCTGGCGCAT CTGTTCAAATATGATTCGAGCCAGGGTCGTTTCAATGGTGAAATTGAGGTCAAAGAAGGCGCGT TCGTGGTGAACGGCAAAGAGGTTAAAGTGTTCGCGGAGGCGGATCCAGAAAAGTTGCCATGG GGTGATCTCGGAATTGACGTGGTTCTGGAATGCACAGGCTTTTTTACTAAAAAAGAAAAAGCTG AAGCGCATGTTCGCGCAGGAGCTAAGAAAGTAGTTATCTCGGCCCCGGCAGGAAACGATCTGA AGACCATCGTCTTTAATGTAAATAATGAAGATCTTGATGGCACGGAAACCGTGATTAGTGGGGC ATCCTGTACTACGAACTGCCTCGCTCCGATGGCGAAAGTGCTTAATGACAAATTTGGCATTGAA AAAGGCTTCATGACCACCATTCACGCGTTTACAAACGACCAAAACACCCTGGACGGCCCGCAC CGCAAAGGTGACCTACGTCGCGCGAGGGCCGCAGCCGTGTCGATTATCCCGAATTCTACTGG GGCGGCGAAAGCGATTAGTCAGGTTATACCCGATTTAGCGGGTAAACTGGATGGGAACGCCCA GCGCGTGCCTGTCCCGACCGGCAGCATTACCGAATTAGTTTCTGTGTTGAAAAAGAAAGTGAC CGTGGAAGAAATCAACGCCGCAATGAAGGAGGCGGCTGACGAGAGCTTTGGCTATACAGAAG ATCCGATCGTCTCAGCAGATGTCGTAGGTATTAACTACGGAAGCCTGTTCGATGCCACCCTGA CGAAGATCGTTGACGTGAATGGTAGCCAGCTTGTTAAAACGGCCGCCTGGTATGATAACGAAA TGTCATACACTTCCCAACTGGTCCGCACGCTGGCATACTTTGCGAAAATTGCAAAGTAA

7. Construction of an acetyl-CoA auxotroph reveals unexpected flexibility in aerobic metabolism connecting glycolysis to the tricarboxylic acid cycle

7.1 Abstract

Acetyl-CoA lies at the core of life's metabolism. Characterizing all available pathways for its biosynthesis therefore contributes to an important understanding of canonical- and underground metabolism involving this central metabolite. In this work, we attempted to generate a strain displaying acetyl-CoA auxotrophy when grown aerobically on glycolytic carbon sources. These efforts revealed that non-canonical metabolic routes can contribute to growth in conditions lacking canonical acetyl-CoA forming enzymes, for example via the activity of pyruvate formate-lyase and pyruvate flavodoxin oxidoreductase. Further characterization of these mutants also revealed intriguing effects of L-threonine supplementation. Taken together, this study highlights the metabolic flexibility of *E. coli* by its ability to adapt to drastic changes in canonical carbon metabolism. These insights can aid with the construction of robust cell factories, and with the development of synthetic auxotrophic strains for pathway- and enzyme engineering.

7.2 Introduction

Acetyl-Coenzyme A (AcCoA) is an ancient central metabolite found in all domains of life and it lies at the foundation of autotrophic growth in early lifeforms (Ljungdahl, 2009; Martin, 2020). Perhaps more familiarly, it functions as entry point to the tricarboxylic acid (TCA) cycle and is the starting point for fatty acid biosynthesis. In addition, AcCoA can act in a regulatory capacity via protein acetylation and allosteric interactions, acting as an indicator and effector related to the global metabolic state of the cell (Pietrocola et al., 2015; Shi & Tu, 2015). Since AcCoA underpins many metabolic processes, having a full understanding of the metabolic pathways involved in acetyl-CoA formation is desirable.

In *E. coli*, acetyl-CoA is natively formed through three main pathways: a) as a terminal product of glycolysis, b) through activation of acetate or c) through the catabolism of fatty acids (Ku et al., 2020). The exact pathway used for AcCoA formation from glycolysis is reliant on oxygen availability. In aerobic environments, pyruvate is mostly decarboxylated to AcCoA through the pyruvate dehydrogenase complex (PDHc) consisting of *aceE, aceF* and *lpd* gene products, with concomitant release of NADH. The quinone-reducing pyruvate oxidase enzyme (*poxB*) is also expressed in aerobic conditions to facilitate stationary acetate flux, achieved through decarboxylation of pyruvate directly to acetate. In anaerobic conditions, pyruvate-formate lyase (*pflB*) is activated by *pflA* and removes a formate group from pyruvate to yield AcCoA. Acetate can be activated to AcCoA through the irreversible acetyl-CoA synthase (*acs*) reaction or the reversible acetate kinase (*ackA*) and phosphate acetyltransferase (*pta*) pathway. Finally, even-chain fatty acids can be consumed via beta oxidation (*fadD, fadE*, *fadB*, *fadA*), releasing one AcCoA per cycle of the pathway.

Apart from these well-studied pathways, other *E. coli* enzymes that are annotated to directly produce AcCoA include 2-ketobutyrate ligase (*kbl*), pyruvate flavodoxin-oxidoreductase (*ydbK*), pyruvate formate lyase homolog *tdcE*, and acetyl-CoA acetyltransferase (*atoB*) (Lin et al., 2015; Nakayama et al., 2013; Sawers et al., 1998). Other pathways that generate acetate, acetaldehyde, or acetyl phosphate can be used to establish indirect pathways towards AcCoA biosynthesis via the previously mentioned acetate activation pathways.

In this work, we attempted to generate a strain auxotrophic for AcCoA when grown on glycolytic substrates. We found that fructose and to a lesser extent glucose were able to support growth in strains with all canonical direct pathways between pyruvate and acetyl-CoA knocked-out (*lpd*, *poxB*, *ydbK*, *pflB*). We also show that *ydbK* and *pflB* both individually contribute to growth in a Δ*lpd* Δ*poxB* background, shedding new light on enzymes not often associated with aerobic AcCoA-forming flux. The additional removal of *kbl* also did not lead to a full cessation of growth on fructose, indicating that other pathways must be contributing to

acetyl-CoA flux. We concluded that there is more flexibility than previously thought in the array of enzymes contributing to aerobic metabolism connecting glycolysis to the TCA cycle.

The use of glucose and relatively short cultivation times (24-48h) is common practice for characterizing strain behaviour. However, this work suggests that these characterization efforts can result in oversimplifications of *E. coli* central metabolism. For the purposes of designing robust and predictable microbial chassis strains, especially for growth-coupling and/or bioproduction purposes, establishing 'non-leaky' and predictable connections between metabolic nodes of interest is essential. In summary, this study proposes that the use of longer cultivation times and non-glucose carbon sources could help establish predictable and robust growth behaviour in a wide variety of conditions.

7.3 Materials and methods

7.3.1 Strains, plasmids, media, and growth conditions

Routine growth of BW25113 and derivates was achieved using solid or liquid LB medium. Strains were routinely grown at 37°C (shaken at 250 rpm, where relevant), unless they were carrying pSIJ8 for gene deletions at which point they were maintained at 30°C. Antibiotics were added where relevant at the following concentrations: carbenicillin (carb, 100 µg/mL), kanamycin (kan, 50 µg/mL) or chloramphenicol (cm, 25 µg/mL). All chemicals used in this work, including carbon sources D-fructose, D-glucose, sodium acetate, L-arabinose, L-rhamnose, L-threonine, were of >99.5% grade (Sigma-Aldrich).

To pre-culture strains for plate reader growth assays, they were first streaked out on LB agar without antibiotics and grown O/N at 37°C. Where relevant, the presence of gene deletions was confirmed by colony PCR (see Table S7.1 for primer sequences) and correct clones were inoculated as biological triplicates in LB medium supplemented with relevant filter-sterilised carbon sources (10-40 mM D-fructose, 0.5-10 mM D-glucose, 20 mM sodium acetate, 20 mM L-threonine, as indicated in the text). These were grown O/N at 37°C, shaking at 250 rpm. Grown LB cultures were diluted 1:100 fold in M9 minimal medium supplemented with carbon

sources as used in the plate growth assays. M9 minimal medium was composed as follows: 47.8 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 18.7 mM NH_4Cl, 2 mM $MgSO_4$, 100 µM $CaCl_2$, 134 µM ethylene diamine tetrachloroacetic acid (EDTA), 31 µM $FeCl_3 \cdot 6H_2O$, 6.2 µM $ZnCl_2$, 0.76 µM $CuCl_2 \cdot 2H_2O$, 0.42 µM $CoCl_2 \cdot 2H_2O$, 1.62 µM H_3BO_3, and 0.081 µM $MnCl_2 \cdot 4H_2O$, adjusted to pH 7.0.

Table 7.1. Plasmids used in this study.

Plasmid name	Contents	Reference
pSIJ8	araC-P_{ara}-α-β-γ rhaRS-P_{rha}-FLP, amp^R	(Jensen et al., 2015)
pKD3	FRT-cm^R-FRT, oriR6K, amp^R	(Datsenko & Wanner, 2000)

Table 7.2. Strains used and constructed in this study.

Strain name	Genotype	References
BW25113	*F- LAM- rrnB3 ΔlacZ4787 hsdR514 Δ(araBAD)567 Δ(rhaBAD)568 rph-1*	DSMZ, Germany
ΔL	BW25113 Δ*lpd::kanR*	Chapter 6
ΔLX	ΔL Δ*poxB::FRT*	This study
ΔLXP	ΔLX Δ*pflB::FRT*	This study
ΔLXY	ΔLX Δ*ydbK::FRT*	This study
ΔLXYP	ΔLXY Δ*pflB::FRT*	This study
ΔLXYPD	ΔLXYP Δ*deoC::cmR*	This study
ΔLXYPK	ΔLXYP Δ*kbl::FRT*	This study

7.3.2 Gene deletions

Lambda red recombination was used for gene deletions, as outlined in previous work (Datsenko & Wanner, 2000; Jensen et al., 2015). Briefly, dsDNA cassettes were prepared by PCR amplification (DreamTaq, ThermoFisher) of the FRT-flanked cm^R marker from plasmid pKD3 using oligos with 50 bp overhangs targeting either side of the gene locus (see Table S7.1 for all targeting primers) (Baba et al., 2006; Datsenko & Wanner, 2000). *E. coli* BW25113 carrying pSIJ8 was grown to an OD_{600} of ~0.3 at 30°C, after which the culture was induced with 15 mM L-arabinose (final concentration) for 45 minutes. Cells were then cooled in an ice slurry for 30 minutes, spun down (1500*g*, 10 mins) and washed twice in ice-cold sterile deionised water. Finally, cells were resuspended in 300 µL ice-cold sterile 10% glycerol. Electrocompetent cells (50 µL) were mixed with 3.5 µL dsDNA cassette (>250 ng/µL in DI

water), the mixture transferred to pre-chilled 0.1 cm electroporation cuvettes, and a 1.8 kV pulse was applied. Cells were immediately recovered for 2.5 hours in LB, supplemented with 20 mM glycerol, 20 mM acetate, 10 mM sorbitol, 10 mM succinate, 1 mM L-arabinose and 100 µg/mL carbenicillin. Cells were subsequently plated on LB agar supplemented identically to the recovery medium, with an additional 17 µg/mL chloramphenicol to select for marker presence and incubated at 30°C. Correct clones were confirmed by colony PCR (see Table S7.1 for primer sequences), and these were inoculated and re-streaked at least once to ensure clonality. To remove integrated cm^R marker, 0.5 mL of an O/N LB culture supplemented identically to the electroporation recovery medium (but lacking L-arabinose) was inoculated into 35 mL of LB (also with carbon source supplements), and once its OD_{600} reached ~0.3 flippase expression was induced by addition of 50 mM sterile L-rhamnose. Induction was maintained for 4 hours and subsequently diluted 10^{-5}- and 10^{-6}-fold and plated on LB agar with carbenicillin. Correct clones were identified by colony PCR and re-streaked as described earlier. When needed, plasmid pSIJ8 was cured by an overnight cultivation at 42ºC.

7.3.3 Microtiter plate growth assays

Microtiter plate experiments were performed using a Tecan Spark plate reader with SparkControl software (v3.1 SP1) (Tecan, USA). Clear flat-bottom 96-well plates were used filled with 200 µL of M9 minimal medium supplemented with carbon sources as indicated in the text (Greiner, USA). M9 minimal medium was inoculated to an OD_{600} of 0.01 with pre-cultures washed three times in sterile 1x M9 salts. To avoid contamination and evaporation, plates were covered with a Breathe-Easy Sealing Membrane (Sigma-Aldrich) and cultivation was performed using a small humidity cassette (Tecan, USA) re-filled every 24h. Growth assay were performed at 37ºC and OD_{600} measurements were taken every 15 minutes, using the following four-step shaking profile: 1) double orbital, 3 mm amplitude, 90 rpm frequency, 225 seconds 2) orbital shaking, 3 mm amplitude, 180 rpm, 225 seconds 3) double orbital, 2.5 mm amplitude, 108 rpm, 225 seconds, 4) orbital, 2.5 mm amplitude, 216 rpm, 224 seconds. Growth rates were calculated through a log conversion of the blank medium-corrected OD

values, followed by determining the maximum slope of these values over time using a 6-12 hour sliding window per replicate.

7.4 Results

7.4.1 Growth of *E. coli* BW25113 without pyruvate dehydrogenase- and oxidase activity

In our previous work describing construction of an NADH-auxotrophic *E. coli* (see Chapter 6), we relied on the deletion of *lpd* to inactive the NADH-producing PDHc and αKGDH reactions (Figure 7.1). Other work investigating Δ*lpd* mutants has shown that *poxB* can be used as an alternative carbon route towards AcCoA synthesis in *E. coli* BW25113 Δ*lpd* (M. Li et al., 2006), although *E. coli* MG1655 was reported to not exhibit this behaviour (Wenk et al., 2020). A double knock-out of these genes (Δ*lpd* Δ*poxB*) was therefore expected to be lethal as the connection between glycolysis and the TCA cycle was effectively severed, which was corroborated in a study using *E. coli* BW25113 (M. Li et al., 2006).

To confirm this observation ourselves, we used BW25113 *Δlpd* (ΔL) to construct a double mutant BW25113 Δ*lpd* Δ*poxB* (ΔLX) and attempted to grow both using acetate, glucose, and fructose as carbon sources (Figure 7.2A, B, C). As expected, the *lpd* deletion prohibited ΔL and ΔLX to grow using acetate as a sole carbon source. Both strains carrying deletions grew rapidly on either glucose or fructose when supplemented with acetate, since acetate can be activated to AcCoA. However, we also observed growth of these strains using 10 mM of solely glucose or fructose.

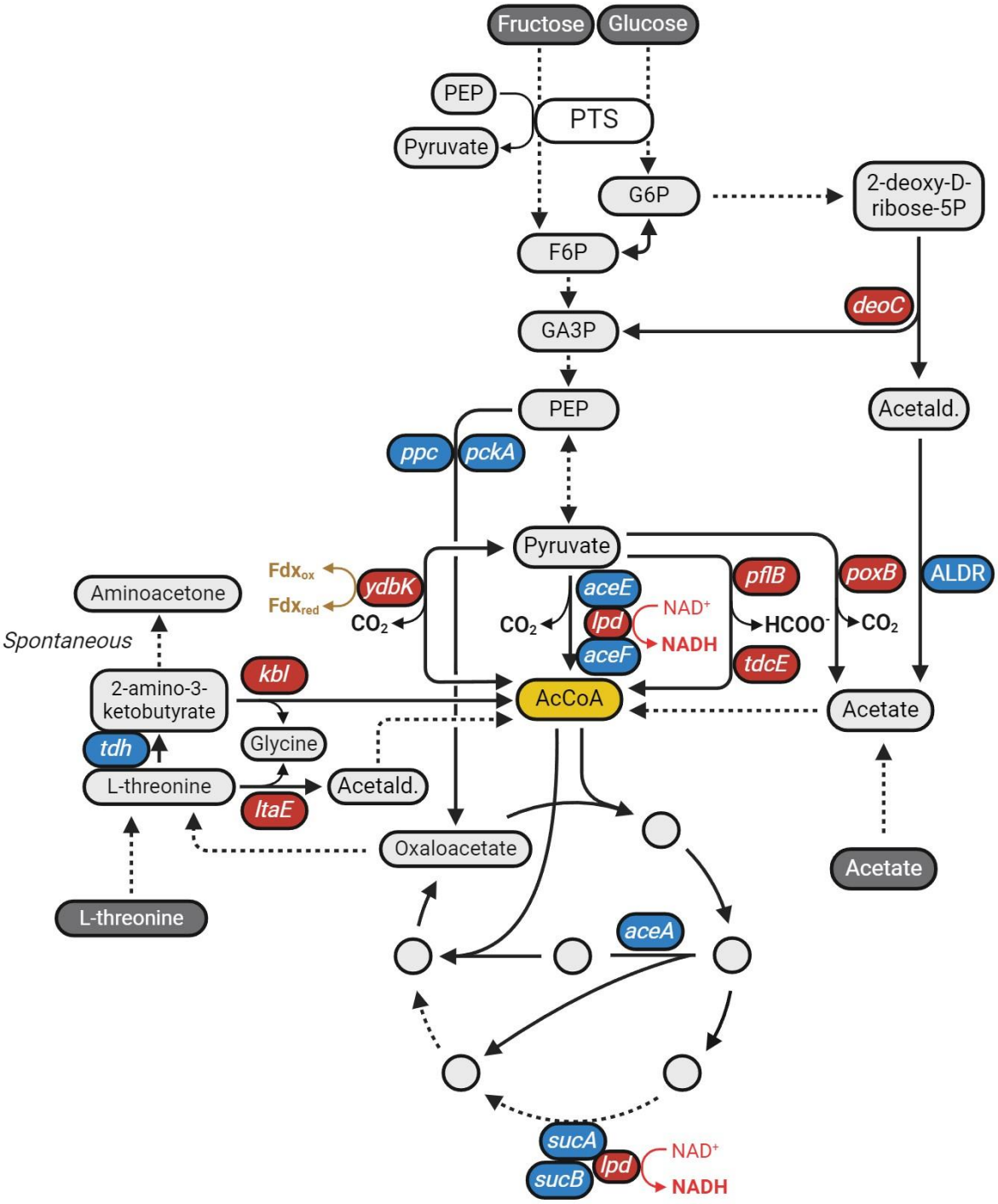


Figure 7.1. Schematic of *E. coli* central metabolism and pathways leading towards acetyl-CoA formation. Gene names in blue boxes represent central metabolic enzymes of interest. Genes names in red boxes represent enzymes identified as being (potentially) involved in AcCoA biosynthesis of which several were investigated in this work.

The ΔL strain grew using fructose within 24h of initiating the growth assay, while ΔL grown on glucose displayed a 48h lag phase. A similar pattern was observed in ΔLX, indicating that the additional *poxB* knockout was not sufficient to block aerobic carbon flux from glycolysis to the TCA cycle. In addition, the growth behaviours of ΔL and ΔLX in their respective sugar conditions was highly similar (Figure 7.2B and C). This indicated that the deletion of *poxB* did not have a significant effect on growth behaviour, leading us to conclude that it was not the pathway responsible for AcCoA-forming flux in these conditions. From these results it was also apparent that the utilisation of glucose as a sole carbon source for ΔL and ΔLX was causing an extended lag phase.

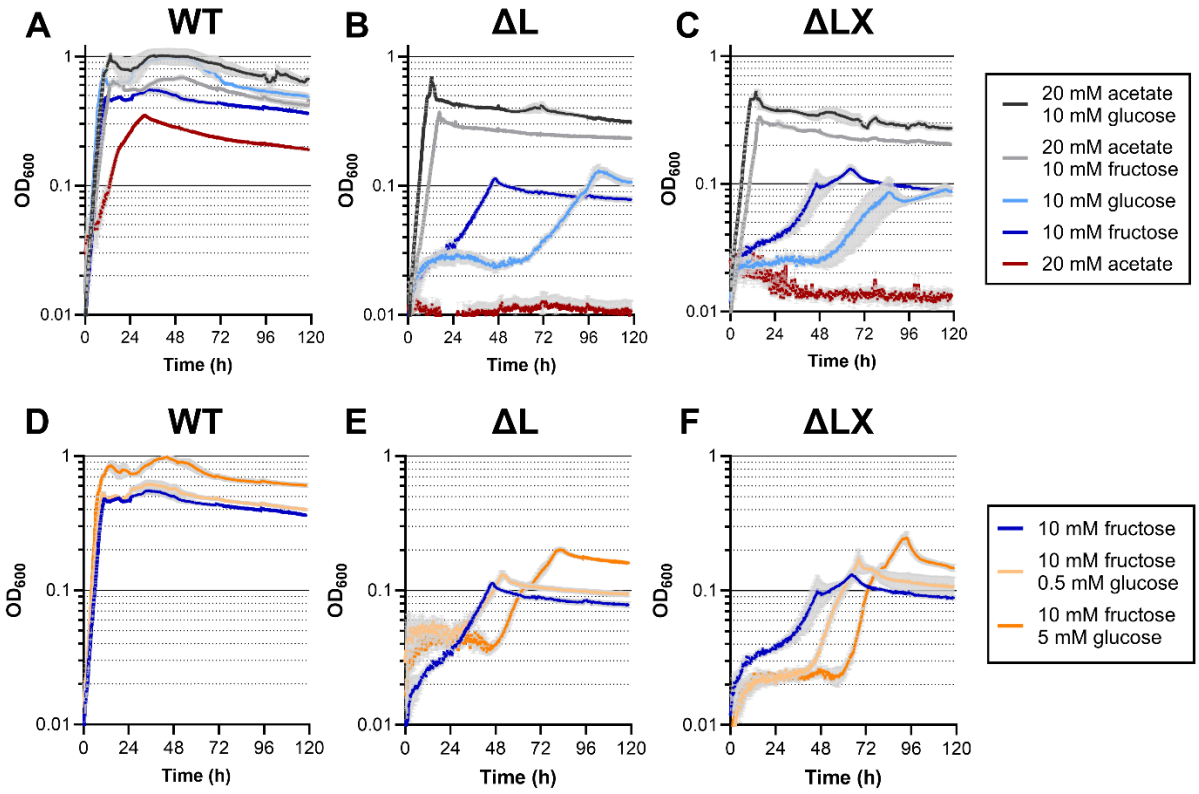


Figure 7.2. Growth characterization of Δ*lpd* and Δ*lpd* Δ*poxB* mutants. Growth of BW25113 (WT - **A, D**), BW25113 *Δlpd* (ΔL - **B, E**), BW25113 *Δlpd ΔpoxB* (ΔLX - **C, F**) using carbon sources as indicated in the respective legends found on the same row. All datapoints represent averages of biological triplicates, and errors bars represent their standard deviations.

To assess this lag phase effect of glucose further, we simultaneously provided glucose and fructose to WT, ΔL and ΔLX strains (Figure 7.2D, E, F), and observed that supplementation

of 10 mM fructose with 0.5- or 5 mM glucose had no effect on lag phase or growth rate in the WT strain. For the ΔL mutant, the addition of 0.5 mM glucose in addition to 10 mM fructose had no effect, while adding 5 mM glucose increased the lag phase by ~24 hours. This growth repressive effect was more pronounced in ΔLX, where the addition of 0.5 mM glucose stunted growth for around 36h, and for around 60h after addition of 5 mM glucose (Figure 7.2F). Final OD_{600} values increased in glucose-supplemented conditions, as expected when more carbon is available.

The differences in lag phase between ΔL and ΔLX when using either fructose or glucose, and when fructose-grown strains were supplemented with glucose, were notable since both sugars enter at a nearly identical points in central metabolism (Figure 7.1). They are both actively imported via the phosphotransferase system (PTS) which utilises phosphoenolpyruvate (PEP) as a phosphate donor, producing pyruvate and importing a hexose. Since both sugars are imported this way, this ruled out possible effects of the required PEP to pyruvate flux.

7.4.2 Pyruvate-formate lyase and pyruvate-flavodoxin oxidoreductase contribute to aerobic growth

After confirming that growth was still possible in a ΔLX background, we sought to investigate other metabolic routes towards AcCoA. Two potentially responsible direct pathways between pyruvate and AcCoA included pyruvate-formate lyase (*pflB*) and pyruvate-flavodoxin oxidoreductase (PFOR, *ydbK*) (Figure 7.1). The former is a well-studied glycyl radical enzyme active during fermentative metabolism of *E. coli*, while the latter is considerably less characterised. Several studies have identified the substrates and corresponding affinities of a PFOR enzyme in *E. coli*, which is presumed to be encoded by *ydbK* (Nakayama et al., 2013). Usually, PFOR enzymes are found to be strictly anaerobic enzymes but *ydbK* was recently found to carry significant aerobic carbon flux in stationary phase *E. coli* (S. Li et al., 2021). To identify possible contributions of *pflB* and *ydbK* to aerobic growth, we generated separate knockouts in a ΔLX background, forming Δ*lpd* Δ*poxB* Δ*ydbK* (ΔLXY) and Δ*lpd* Δ*poxB* Δ*pflB*

(ΔLXP), in addition to a quadruple knockout Δ*lpd* Δ*poxB* Δ*ydbK* Δ*pflB* (ΔLXYP) and assessed their growth in various conditions (Figure 7.3A-E).

Compared to ΔLX, the additional *pflB* or *ydbK* knockout reduced maximum growth rate significantly by ~30%, respectively, when grown on 40 mM fructose (Figure 7.3G). Deletion of all four genes (ΔLXYP) resulted in a more significant reduction in growth rate of ~75% compared to LX, and ~60% compared to ΔLXP and ΔLXY. These findings suggested that both *pflB* and *ydbK* were contributing to aerobic growth rate in a ΔLX background. The effect of knocking both genes out on growth was additive, further indicating that both enzymes were individually contributing to growth.

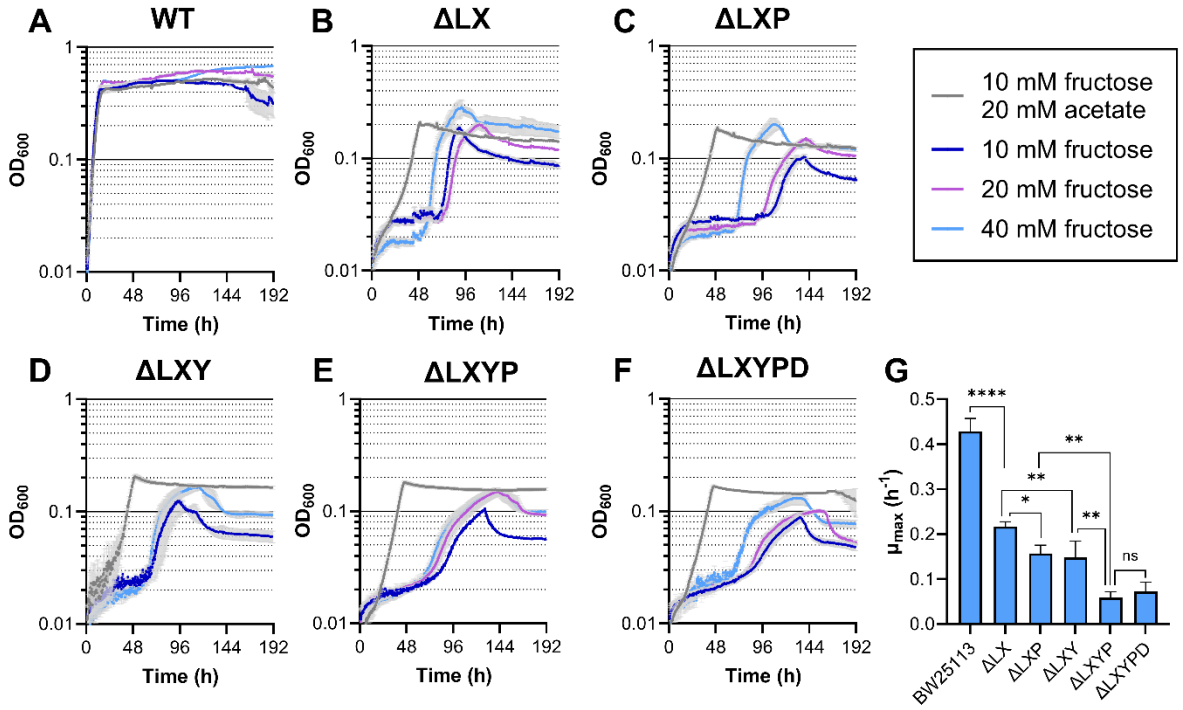


Figure 7.3. Growth characterization of strains lacking acetyl-CoA generating enzymes. Assays were carried out using various concentrations of fructose as glycolytic carbon source, supplemented with 20 mM acetate. Assayed strains included **(A)** wildtype BW25113 (WT), **(B)** BW25113 *Δlpd ΔpoxB* (LX) **(C)** BW25113 *Δlpd ΔpoxB ΔpflB* (LXP), **(D)** BW25113 *Δlpd ΔpoxB ΔydbK* (LXY), **(E)** BW25113 *Δlpd ΔpoxB ΔpflB ΔydbK* (LXYP) and **(F)** BW25113 *Δlpd ΔpoxB ΔpflB ΔydbK ΔdeoC::cmR* (LXYPD). **(G)** Average μ_{max} values of each strain grown with 40 mM fructose. Datapoints represent the average of biological triplicate measurements, and errors bars represent their standard deviations. The 20 mM fructose condition of strain ΔLXY was omitted due to contamination. Where indicated, statistical analysis was performed by comparing means using a one-way ANOVA and post-hoc Tukey's multiple comparison test. P-value significance indicated are as follows: ns (p-value >0.05); * (p-value <0.05), ** (p-value <0.01), *** (p-value <0.001) or **** (p-value <0.0001).

In our assessment of potential AcCoA-forming pathways we identified deoxyribose-phosphate aldolase (*deoC*) to be a potential source of acetaldehyde, which itself can be a source of AcCoA (Valentin-Hansen et al., 1982). We knocked-out *deoC* in a ΔLXYP background, forming ΔLXYPD, but did not observe significantly lower growth rates than ΔLXYP when grown on 40 mM fructose (Figure 7.3G). There were significant differences between the two strains in terms of final OD_{600} using 20- or 40 mM fructose, but this was not the case when using 10 mM fructose. Hence, we concluded that *deoC* was not responsible for the majority of AcCoA flux in strains with a ΔLXYP background.

Taken together, the observed growth of the ΔLXYP(D) mutant on fructose as a sole carbon source indicated that other pathways must be responsible for aerobic carbon flux between glycolysis and the TCA cycle.

7.4.3 The activity of 2-amino-3-ketobutyrate ligase contributes to aerobic growth

Another AcCoA-generating metabolic route occurs via threonine degradation, in which L-threonine is cleaved to form glycine and acetyl-CoA by 2-amino-3-ketobutyrate ligase (*kbl*). Threonine can be generated either from serine via the activity of the glycine cleavage system and the low-specificity L-threonine aldolase *ltaE*, or formed from aspartate which in turn requires the TCA cycle intermediate oxaloacetate (Figure 7.1) (Lin et al., 2015). The *ltaE*-dependent formation of L-threonine requires acetaldehyde, which would itself be a source of AcCoA. We therefore expected that L-threonine biosynthesis in ΔLXYP occurred via the carboxylation of PEP to oxaloacetate, with subsequent formation of L-aspartate and conversion to L-threonine. The terminal Kbl enzyme could then generate AcCoA via its aforementioned cleavage reaction, which prompted us to delete this gene.

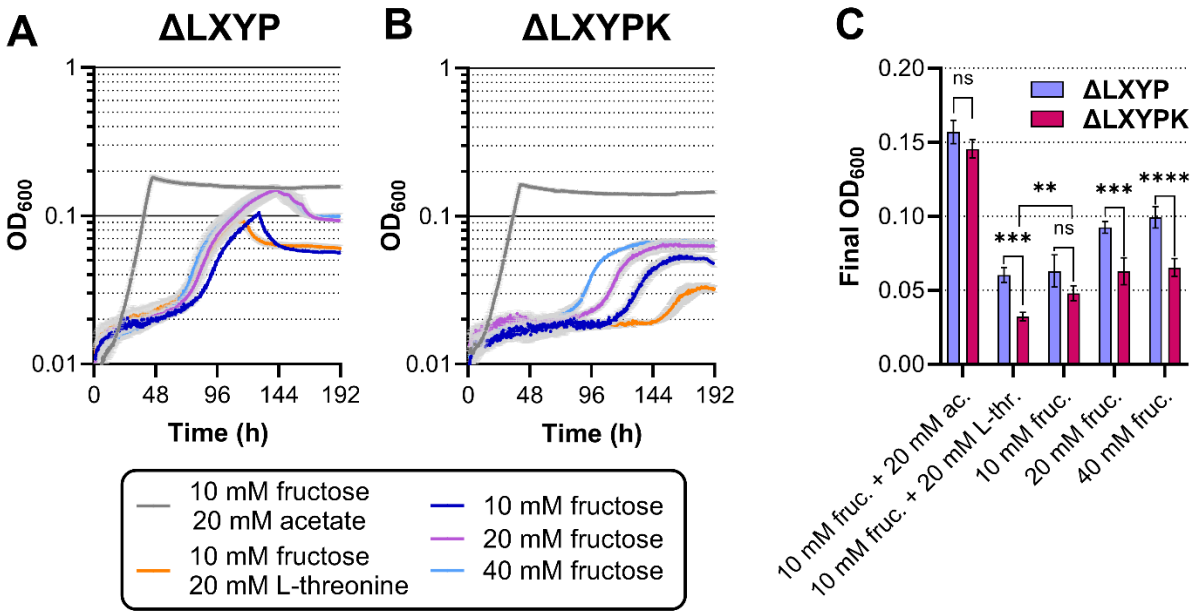


Figure 7.4. The effect of a 2-amino-3-ketobutyrate ligase (*kbl*) deletion on growth of ΔLXYP. (A) Growth of BW25113 *Δlpd ΔpoxB ΔpflB ΔydbK* (LXYP) or **(B)** BW25113 *Δlpd ΔpoxB ΔpflB ΔydbK Δkbl* (LXYPK) using different concentrations of fructose, including supplementation with 20 mM acetate or 20 mM L-threonine where indicated. Note that the fructose- and fructose/acetate data of ΔLXYP were identical to those in Figure 7.3. **(C)** Comparison of final OD_{600} values achieved by ΔLXYP and ΔLXYPK per carbon source(s) used. Datapoints represent the average of biological triplicate measurements, and errors bars represent their standard deviations. Where indicated, statistical analysis was performed by comparing means using a two-way ANOVA and post-hoc Šídák's multiple comparison test. P-value significance indicated are as follows: ns (p-value >0.05); * (p-value <0.05), ** (p-value <0.01), *** (p-value <0.001) or **** (p-value <0.0001).

The deletion of *kbl* in a ΔLXYP background (ΔLXYPK) resulted in growth after a 72- to 144-hour lag phase when grown using solely fructose or fructose and L-threonine (Figure 7.4B). The duration of the lag phase was dependent on the fructose concentration, with higher concentrations reducing the lag phase, although no effect of fructose concentration on growth rate was observed. In contrast, ΔLXYP initiated exponential growth at approximately the same time irrespective of fructose concentration, indicating that this lag phase effect was caused by the additional *kbl* deletion (Figure 7.4A, B). Interestingly, the addition of 20 mM L-threonine to 10 mM fructose extended the lag phase of ΔLXYPK by another 48h compared to supplementation of only 10 mM fructose. Not only did it extend the lag phase, but supplementation of L-threonine also reduced final OD_{600} significantly compared to when it was lacking in the medium (Figure 7.4C). This deleterious effect of L-threonine supplementation

was only observed in ΔLXYPK, and not ΔLXYP, indicating that additional deletion of *kbl* could be the cause. We note that L-threonine supplementation in ΔLXYP was expected to lead to a fructose + acetate phenotype, since L-threonine could technically be catabolised to acetyl-CoA via *tdh* and *kbl* activity (Figure 7.1). In contrast, we measured no significant difference in final OD_{600} between ΔLXYP cultures with or without L-threonine supplementation, suggesting that this AcCoA-production route was not active in the tested conditions.

The significant reductions of final OD_{600} and longer lag phases compared to ΔLXYP indicated that *kbl* was contributing to growth in a ΔLXYP background. Even so, L-threonine supplementation in strains with intact *kbl* did not result in increased biomass formation, raising the question of the exact function of *kbl* in ΔLXYP. Nonetheless, the growth of ΔLXYPK on solely fructose indicated that yet more pathways must be capable of providing sufficient AcCoA.

7.4.4 Other metabolic routes towards AcCoA

After observing growth in a ΔLXYPK background, we initiated work on deleting the *ltaE* and *tdcE* genes. As described earlier, the LtaE enzyme can cleave L-threonine into L-glycine and acetaldehyde, the latter of which can be converted to AcCoA via the activity of endogenous enzymes such as AdhE. However, the longer lag phase and lower final OD_{600} we observed when supplementing ΔLXYPK with L-threonine suggested that L-threonine was detrimental to growth. This is the opposite of what we expected, namely that L-threonine would considerably improve growth since it would function as a source of AcCoA. We successfully constructed ΔLXYP ltaE::cmR (ΔLXYPA) and ΔLXYPK ΔltaE (ΔLXYPKA) but due to time restraints were unable to assess their growth.

A final candidate gene was *tdcE*, the oxobutanoate/pyruvate-formate lyase gene homologous to *pflB*. This gene is situated within the *tdcABCDE* operon, an L-threonine degradation operon. Previous work has shown that *tdcE* can perform pyruvate-formate lyase activity, and as such it would represent another direct pathway from pyruvate to AcCoA. In addition, *tdcE* is known

to experience cAMP-CRP mediated catabolite repression, which corroborates the observation that growth on glucose hindered in a ΔLX background. As mentioned earlier, the detrimental effect of L-threonine supplementation suggested that other metabolic routes other than L-threonine degradation were forming AcCoA in ΔLXYPK, which could possibly be through TdcE activity. We attempted to delete *tdcE* in the ΔLXYPKA background but were unable to isolate successful *tdcE* knockout mutants.

7.5 Discussion & conclusion

Acetyl-CoA is a core central metabolite represented in all trees of life, positioned at the intersection of glycolysis, fatty acid metabolism and the TCA cycle. In *E. coli*, acetyl-CoA also acts as the branching point in overflow metabolism. By studying strains with the canonical AcCoA biosynthesis pathways removed, novel 'underground' metabolic pathways or compensatory enzyme activity could potentially be uncovered, deepening our understanding of the metabolic connections between various parts of central metabolism in *E. coli* and potentially other microbes.

In this work, we found that the removal of canonical pyruvate dehydrogenase- and oxidase activity (Δ*lpd* and Δ*poxB*, respectively) was insufficient to render *E. coli* auxotrophic for AcCoA when grown on solely glycolytic substrates. The growth of these strains differed depending on the glycolytic carbon source; use of glucose as a sole C-source resulted in significantly longer lag phases compared to fructose, which we also observed when glucose was supplemented in fructose-containing medium. This indicated that transcriptional regulatory effects of glucose addition were hindering growth in this mutant strain. The Δ*lpd* deletion causes an inactivation of the αKGDHc which was known to result in an accumulation of αKG (M. Li et al., 2006). The accumulation of αKG can reduce glucose uptake rates via competitive inhibition of PtsI (or EI), a sugar non-specific subunit of the PTS. The authors investigated this using a Δ*sucAB* mutant, which is metabolically identical to Δ*lpd* in the context of TCA cycle flux (Doucette et al., 2011). While this would be an elegant explanation for the observed growth behaviour using glucose,

EI is also essential for fructose uptake via the PTS (Luo et al., 2014). Hence, we would expect to observe this effect using fructose as well, which we did not.

There are also other differences in the regulatory effects that either sugar has on central metabolism. The catabolism of glucose, and much less so with fructose, induces catabolite repression. The presence of glucose causes a rapid reduction in the levels of intracellular cAMP, inactivating the cAMP receptor protein (CRP) thereby reducing transcription of CRP-dependent enzymes (Deutscher, 2008; Görke & Stülke, 2008). One of these pathways is the glyoxylate shunt, initiated by the activity of isocitrate lyase (*aceA*). Glucose represses *aceBAK* expression, and strains with a deletion in *lpd* must utilise the glyoxylate shunt for growth due to the inactivation of the αKGDHc (Resnik et al., 1996). Furthermore, pyruvate is an allosteric activator of IclR, the direct regulator of the *aceBAK* operon (Lorca et al., 2007). In conditions where pyruvate (temporarily) accumulates, which is likely in ΔL or ΔLX strains, IclR is expected to more strongly repress *aceBAK* transcription which reduces carbon flux through the glyoxylate shunt. Although repression of *aceBAK* might explain the increased lag phase on glucose we did not observe any such lag phase when ΔL or ΔLX was grown on a combination of acetate and glucose, a condition that is still expected to require the glyoxylate shunt for carbon assimilation. Although the transcription of the *aceBAK* operon is induced by acetate, previous work has shown that when supplemented with glucose, *aceBAK* expression is strongly repressed regardless of the presence of acetate (Spencer et al., 2007). This suggested that the longer lag phase observed on solely glucose compared to fructose was due to transcriptional effects of genes other than just *aceBAK*. It has been found that fructose-grown *E. coli* exhibit higher global transcription of both glycolytic- and TCA cycle genes than glucose-grown cultures (Kim et al., 2018). Perhaps this stronger transcriptional response to fructose resulted in a decrease of its lag phase despite the existing metabolic barriers imposed by the gene deletions.

This study also reported that separate knockouts of *ydbK* and *pflB* in a ΔLX background resulted in significant decreases in growth rate when provided 40 mM of fructose, and a

combined ΔLXYP knockout even more so. This is surprising since both *pflB* and *ydbK* are not widely reported to be responsible for carrying glycolytic carbon flux in aerobically grown *E. coli*. Only recently have studies found that *ydbK* activity in *E. coli* can carry significant aerobic pyruvate-AcCoA flux potentially as a mechanism to balance $NADP^+$/NADPH pools through flavodoxin/ferredoxin reductase activity, but only in stationary phase (S. Li et al., 2021). In addition, *pflB* was shown to display activity in microaerobic conditions (Zhu et al., 2007).

Since our work showed that both *ydbK* and *pflB* contributed to aerobic exponential growth within the metabolic context of ΔLX, there may be mechanisms acting on both the regulatory- and protein level to enable their expression at sufficient levels and to confer them with higher oxygen tolerances than reported from *in vitro* studies. However, these conclusions are subject to the cultivation conditions having been truly aerobic. This should be confirmed by direct dissolved oxygen measurements or cultivation in oxygen-controlled chemostats. If *ydbK* and *pflB* are indeed more O_2-tolerant *in vivo* than previously concluded, studying these *in vivo* mechanisms may enable the exploitation of these proteins to greater effect. For example, the *ydbK* enzyme may be of interest due its requirement for flavodoxin/ferredoxin as an electron acceptor or donor. These PFOR enzymes are often reversible *in vivo* due to the low potential of the flavodoxin/ferredoxin electron donors, possibly providing a novel route for direct CO_2 fixation in *E. coli* at the pyruvate-AcCoA node in addition to the existing formate-requiring *pflB*-mediated pathway (Zelcbuch et al., 2016).

The degradation of L-threonine via the *tdh-kbl* pathway enables AcCoA and glycine synthesis (Ku et al., 2020; Lin et al., 2015; Yishai et al., 2017). In our work, we were able to demonstrate that *kbl* contributed to aerobic growth in a ΔLXYP background. However, the exact mechanism behind this behaviour was not clear since supplementation of L-threonine to ΔLXYP cultures did not result in any growth improvements. If *kbl* was indeed responsible for AcCoA flux via L-threonine degradation, the supplementation of this amino acid was expected to result in significant growth improvements. The fact that we did not observe improved growth in both ΔLXYP and ΔLXYPK after supplementing L-threonine also implied degradation via low-

specificity L-threonine aldolase (*ltaE*) was not carrying significant AcCoA-forming flux. One explanation for this could be the expression levels of the genes. Indeed, studies utilising this pathway resort to overexpression of *tdh-kbl* to establish pathway flux (Lin et al., 2015; Yishai et al., 2017). Another reason could have to do with the deletion of *lpd* in all ΔL and derived strains. If *kbl* were the main AcCoA-forming pathway the coproduct of the reaction, glycine, would accumulate. A known pathway for glycine catabolism is the glycine cleavage system, forming serine through the activity of the *gcvTHP* operon, *lpd* and *glyA* (Stauffer, 2004; Steiert et al., 1990). With the deletion of *lpd* this pathway cannot function, potentially causing glycine accumulation and reduced- or arrested *kbl* flux as a result.

Finally, we considered the deletion of *tdcE*, a homolog to the pyruvate formate lyase structural protein *pflB*. This enzyme can replace *pflB* activity in glucose-fed anaerobic cultures and is known to directly interact with the formate-translocating FocA channel (Falke et al., 2016; Sawers et al., 1998). Since we found that *pflB* activity can contribute to growth in aerobic conditions, the *tdcE* homolog may also play a role. It would therefore be of interest to generate and characterise Δ*tdcE* strains.

In summary, this work demonstrated that *E. coli* exhibits a high degree of metabolic flexibility regarding aerobic AcCoA-forming flux. Two enzymes not usually associated with aerobic pyruvate to AcCoA flux, *ydbK* and *pflB,* were found to contribute to exponential growth when canonical aerobic routes (PDHc and *poxB*) were unavailable. The additional removal of *kbl* in this quadruple ΔLXYP background resulted in even poorer growth, although the measured biomass formation means there are other AcCoA-forming pathways still available to *E. coli*. The use of carbon labelling, additional gene knockouts, overexpression studies and the use of adaptive- and directed evolution could uncover novel routes towards AcCoA formation in *E. coli*. These pathways may offer benefits in terms of yield or efficiency compared to canonical ones, potentially enabling more efficient bioprocesses.

Acknowledgements

I would like to thank Yutong Yin for our insightful discussions. This work was supported by the BBSRC (grant BB/M011224/1) and the Oxford Interdisciplinary Bioscience Doctoral Training Partnership. Figure schematics were created with Biorender.com.

7.6 References

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, *2*, 2006.0008. https://doi.org/10.1038/MSB4100050

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

Deutscher, J. (2008). The mechanisms of carbon catabolite repression in bacteria. *Current Opinion in Microbiology*, *11*(2), 87–93. https://doi.org/10.1016/J.MIB.2008.02.007

Doucette, C. D., Schwab, D. J., Wingreen, N. S., & Rabinowitz, J. D. (2011). α-ketoglutarate coordinates carbon and nitrogen utilization via Enzyme I inhibition. *Nature Chemical Biology*, *7*(12), 894. https://doi.org/10.1038/NCHEMBIO.685

Falke, D., Doberenz, C., Hunger, D., & Sawers, R. G. (2016). The glycyl-radical enzyme 2-ketobutyrate formate-lyase, TdcE, interacts specifically with the formate-translocating FNT-channel protein FocA. *Biochemistry and Biophysics Reports*, *6*, 185–189. https://doi.org/10.1016/J.BBREP.2016.04.005

Görke, B., & Stülke, J. (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nature Reviews Microbiology 2008 6:8*, *6*(8), 613–624. https://doi.org/10.1038/nrmicro1932

Jensen, S. I., Lennen, R. M., Herrgård, M. J., & 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), 17874. https://doi.org/10.1038/srep17874

Kim, D., Seo, S. W., Gao, Y., Nam, H., Guzman, G. I., Cho, B. K., & Palsson, B. O. (2018). Systems assessment of transcriptional regulation on central carbon metabolism by Cra and CRP. *Nucleic Acids Research*, *46*(6), 2901–2917. https://doi.org/10.1093/NAR/GKY069

Ku, J. T., Chen, A. Y., & Lan, E. I. (2020). Metabolic Engineering Design Strategies for Increasing Acetyl-CoA Flux. *Metabolites*, *10*(4). https://doi.org/10.3390/METABO10040166

Li, M., Ho, P. Y., Yao, S., & Shimizu, K. (2006). Effect of lpdA gene knockout on the metabolism in Escherichia coli based on enzyme activities, intracellular metabolite concentrations and metabolic flux analysis by 13C-labeling experiments. *Journal of Biotechnology*, *122*(2), 254–266. https://doi.org/10.1016/J.JBIOTEC.2005.09.016

Li, S., Ye, Z., Moreb, E. A., Hennigan, J. N., Castellanos, D. B., Yang, T., & Lynch, M. D. (2021). Dynamic control over feedback regulatory mechanisms improves NADPH flux and xylitol biosynthesis in engineered *E. coli*. *Metabolic Engineering*, *64*, 26–40. https://doi.org/10.1016/J.YMBEN.2021.01.005

Lin, Z., Zhang, Y., Yuan, Q., Liu, Q., Li, Y., Wang, Z., Ma, H., Chen, T., & Zhao, X. (2015). Metabolic engineering of Escherichia coli for poly(3-hydroxybutyrate) production via threonine bypass. *Microbial Cell Factories*, *14*(1), 1–12. https://doi.org/10.1186/S12934-015-0369-3

Ljungdahl, L. G. (2009). A life with acetogens, thermophiles, and cellulolytic anaerobes. *Annual Review of Microbiology*, *63*(Volume 63, 2009), 1–25. https://doi.org/10.1146/ANNUREV.MICRO.091208.073617

Lorca, G. L., Ezersky, A., Lunin, V. V., Walker, J. R., Altamentova, S., Evdokimova, E., Vedadi, M., Bochkarev, A., & Savchenko, A. (2007). Glyoxylate and Pyruvate Are Antagonistic Effectors of the Escherichia coli IclR Transcriptional Regulator. *Journal of Biological Chemistry*, *282*(22), 16476–16491. https://doi.org/10.1074/JBC.M610838200

Luo, Y., Zhang, T., & Wu, H. (2014). The transport and mediation mechanisms of the common sugars in Escherichia coli. *Biotechnology Advances*, *32*(5), 905–919. https://doi.org/10.1016/J.BIOTECHADV.2014.04.009

Martin, W. F. (2020). Older Than Genes: The Acetyl CoA Pathway and Origins. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/FMICB.2020.00817

Nakayama, T., Yonekura, S. I., Yonei, S., & Zhang-Akiyama, Q. M. (2013). Escherichia coli pyruvate:flavodoxin oxidoreductase, YdbK - regulation of expression and biological roles in protection against oxidative stress. *Genes & Genetic Systems*, *88*(3), 175–188. https://doi.org/10.1266/GGS.88.175

Pietrocola, F., Galluzzi, L., Bravo-San Pedro, J. M., Madeo, F., & Kroemer, G. (2015). Acetyl Coenzyme A: A Central Metabolite and Second Messenger. *Cell Metabolism*, *21*(6), 805–821. https://doi.org/10.1016/J.CMET.2015.05.014

Resnik, E., Pan, B., Ramani, N., Freundlich, M., & Laporte, D. C. (1996). Integration Host Factor Amplifies the Induction of the aceBAK Operon of Escherichia coli by Relieving IclR Repression. *JOURNAL OF BACTERIOLOGY*, *178*(9), 2715–2717.

Sawers, G., Heßlinger, C., Muller, N., & Kaiser, M. (1998). The glycyl radical enzyme TdcE can replace pyruvate formate-lyase in glucose fermentation. *Journal of Bacteriology*, *180*(14), 3509–3516. https://doi.org/10.1128/JB.180.14.3509-3516.1998

Shi, L., & Tu, B. P. (2015). Acetyl-CoA and the Regulation of Metabolism: Mechanisms and Consequences. *Current Opinion in Cell Biology*, *33*, 125. https://doi.org/10.1016/J.CEB.2015.02.003

Spencer, C. C., Bertrand, M., Travisano, M., & Doebeli, M. (2007). Adaptive Diversification in Genes That Regulate Resource Use in Escherichia coli. *PLOS Genetics*, *3*(1), e15. https://doi.org/10.1371/JOURNAL.PGEN.0030015

Stauffer, G. V. (2004). Regulation of Serine, Glycine, and One-Carbon Biosynthesis. *EcoSal Plus*, *1*(1). https://doi.org/10.1128/ECOSALPLUS.3.6.1.2

Steiert, P. S., Stauffer, L. T., & Stauffer, G. V. (1990). The lpd gene product functions as the L protein in the Escherichia coli glycine cleavage enzyme system. *Journal of Bacteriology*, *172*(10), 6142–6144. https://doi.org/10.1128/JB.172.10.6142-6144.1990

VALENTIN-HANSEN, P., BOËTIUS, F., HAMMER-JESPERSEN, K., & SVENDSEN, I. (1982). The primary structure of Escherichia coli K12 2-deoxyribose 5-phosphate aldolase. Nucleotide sequence of the deoC gene and the amino acid sequence of the enzyme. *European Journal of Biochemistry*, *125*(3), 561–566. https://doi.org/10.1111/J.1432-1033.1982.TB06719.X

Wenk, S., Schann, K., He, H., Rainaldi, V., Kim, S., Lindner, S. N., & Bar-Even, A. (2020). An “energy-auxotroph” *Escherichia coli* provides an in vivo platform for assessing NADH regeneration systems. *Biotechnology and Bioengineering*, *117*(11), 3422–3434. https://doi.org/10.1002/bit.27490

Yishai, O., Goldbach, L., Tenenboim, H., Lindner, S. N., & Bar-Even, A. (2017). Engineered Assimilation of Exogenous and Endogenous Formate in Escherichia coli. *ACS Synthetic Biology*, *6*(9), 1722–1731. https://doi.org/10.1021/ACSSYNBIO.7B00086

Zelcbuch, L., Lindner, S. N., Zegman, Y., Slutskin, I. V., Antonovsky, N., Gleizer, S., Milo, R., & Bar-Even, A. (2016). Pyruvate formate-lyase enables efficient growth of Escherichia coli on acetate and formate. *Biochemistry*, *55*(17), 2423–2426. https://doi.org/10.1021/acs.biochem.6b00184

Zhu, J., Shalel-Levanon, S., Bennett, G., & San, K. Y. (2007). The YfiD protein contributes to the pyruvate formate-lyase flux in an Escherichia coli arcA mutant strain. *Biotechnology and Bioengineering*, *97*(1), 138–143. https://doi.org/10.1002/BIT.21219

7.7 Supplementary data

Table S7.1. Primers used in this study

Name	Sequence	Purpose
JRN075	GAATCCCCAGGAGCTTAC	Check *lpd* locus
JRN114	Ggaaaggtaaattgcagacg	Check *lpd* locus
JRN135	Tctccgtaaacgtcgtcc	Check *poxB* locus
JRN136	gatttggttctcgcataatc	Check *poxB* locus
JRN354	GATAATAAGTCTGGTGAATG	Check *ydbK* locus
JRN355	CTCTGTTTGATCAAAACA	Check *ydbK* locus
JRN089	GAAGGCTACGTCGAGTCTG	Check *pflB* locus
JRN356	TTGGTTGGGTTGACATACTG	Check *pflB* locus
JRN613	ggattttaatcagcagatcg	Check *kbl* locus
JRN614	attatgcccagccaaaac	Check *kbl* locus
JRN112	gaaagacgacgggtatgaccgccggagataaatatatagaggtcatg atgtgtgtaggctggagctgcttc	Amplifying *cmR* from pKD3 with *lpd* flanks
JRN113	aattgccggatgttccggcaaacgaaaaattacttcttcttcgctttcggc atatgaatatcctccttag	Amplifying *cmR* from pKD3 with *lpd* flanks
JRN237	CCTTATTATGACGGGAAATGCCACCCTTTTTACC TTAGCCAGTTTGTTTTtgtgtaggctggagctgcttc	Amplifying *cmR* from pKD3 with *poxB* flanks
JRN238	GATGAACTAAACTTGTTACCGTTATCACATTCAG GAGATGGAGAACCATGcatatgaatatcctccttag	Amplifying *cmR* from pKD3 with *poxB* flanks
JRN342	TAGAATTGGATAATCCTTATCCAGAGCATTTAATCGGT GTTGCTTTTTTCtgtgtaggctggagctgcttc	Amplifying *cmR* from pKD3 with *ydbK* flanks
JRN343	CGCAGCAATTTCGTGCGCCCCTCATTTGCGCAATGTA AGGGTGTCATATGcatatgaatatcctccttag	Amplifying *cmR* from pKD3 with *ydbK* flanks
JRN352	TTTTACTGTACGATTTCAGTCAAATCTAATTACAT AGATTGAGTGAAGGTtgtgtaggctggagctgcttc	Amplifying *cmR* from pKD3 with *pflB* flanks
JRN353	CGAAGTACGCAGTAAATAAAAAATCCACTTAAGA AGGTAGGTGTTACATGcatatgaatatcctccttag	Amplifying *cmR* from pKD3 with *pflB* flanks
JRN611	agtttggataacgctttcatctcacatcctcaggcgataacgcccagttgt gtgtaggctggagctgcttc	Amplifying *cmR* from pKD3 with *kbl* flanks
JRN612	aagtttgggtaatatgtgctggaatttgccctgtctggagaatcgcaatgc atatgaatatcctccttag	Amplifying *cmR* from pKD3 with *kbl* flanks

8. General discussion

The rapidly changing climate requires the undivided attention of scientists and policymakers to enable a rapid transition towards energy sources with a reduced carbon footprint. While leaps forward have been made in the production and use of biofuels such as bioethanol and biodiesel, these are running into limitations regarding land use, blending limits, and production costs. Next generation sustainable fuels are therefore in high demand, and the direct production of drop-in fuels using bioprocesses holds much promise in this area.

As discussed in Chapter 2, a limitation to these biological processes is the current poor performance of enzymatic pathways responsible for synthesizing hydrocarbon products. Enzyme- and metabolic engineering efforts have led to considerable improvements in product synthesis, but ultimately these are still unsuitable for commercial exploitation. A core aim of this thesis was to establish new methods linking the production of hydrocarbon molecules to microbial growth, enabling growth-based engineering of sustainable fuel production pathways.

The work presented in this thesis covers a broad range of topics. Therefore, to facilitate and guide this general discussion, the original thesis scope and objectives have been replicated throughout. This thesis aimed to address several of the challenges involved in microbial hydrocarbon biosynthesis:

i. By demonstrating that the biosynthesis of hydrocarbons in *E. coli* is possible using alternative carbon sources to glucose.

The development of a cloning toolbox was first required before initiating bioproduction work (Chapter 3). This was done to ensure any production studies would not be limited by the available enzyme expression solutions. To do so, a medium-copy MoClo-compatible vector kit was constructed based on the MoClo cloning standards (Iverson et al., 2016; Weber et al., 2011). This set of plasmids was used throughout the experimental work of this thesis and demonstrates how standardised and modular cloning removes significant time constraints

related to experimental synthetic biology work. These plasmids are publicly available and therefore represent a contribution to the synthetic biology community.

By using this plasmid kit the expression levels of hydrocarbon-forming enzymes could be optimised in *E. coli* (Chapter 4). Several novel approaches for optimizing hydrocarbon biosynthesis were also attempted, including the deletion of a thioesterase gene and the supplementation of KCl, both of which increased hydrocarbon titre. Despite considerable optimization efforts, the hydrocarbon titres achieved in this study were below those reported in most works, demonstrating that the replicability and robustness of microbial hydrocarbon synthesis is not guaranteed.

However, the core aim of the work presented in chapter 4 was to demonstrate the viability of using non-sugar carbon sources for hydrocarbon synthesis. The fact that no studies using *E. coli* have attempted to utilise carbon sources besides glucose points to the low technology readiness level of microbial hydrocarbon synthesis. Current commercial processes for biofuel production use crop-derived substrates, but support for these 1st generation biofuels is dwindling due to concerns over arable land use. To establish commercially viable alternatives to 1st generation biofuels, efficient use must be made of alternative carbon sources, preferably from waste streams or compounds directly derived from CO_2. Acetate can be produced by acetogenic fermentation or through electrocatalysis, while ethanol can be produced sustainably via fermentation of sugars from waste biomass or via gas-fermenting microbes (Köpke et al., 2010). The work presented in chapter 4 provided evidence that both acetate and ethanol are viable carbon sources for microbial hydrocarbon synthesis, performing equally or better than glucose in terms of product yield per substrate. Although glucose performed better in metrics related to biomass accumulation and growth rate, the optimization of C_2 carbon source usage could result in further improvements.

It would be prudent to develop next-generation fuel production in conjunction with next-generation carbon source utilisation. This is in contrast to the usual bioproduction pipeline, where product synthesis is first proven and optimised using an ideal carbon source such as

glucose, initially resulting in promising titres and yields. However, when the use of non-ideal but sustainable carbon sources is required, this optimization process must be reinitiated. This costs resources and time and may even result in an infeasible process. The use of growth-coupled approaches as detailed in this thesis may offer an efficient solution to this problem, since it allows the simultaneous coupling of both substrate catabolism and product synthesis to growth. Such an approach would therefore be ideal for engineering all aspects involved in the conversion of novel substrates to product, as opposed to separately optimizing substrate catabolic pathways and product synthesis pathways and integrating these later with ensuing difficulty.

ii. By designing novel strategies to enable the growth-coupling of alkane biosynthesis enzymes, strictly linking fuel production to microbial growth.

The growth-coupled production of hydrocarbons represents an understudied field with considerable potential. As expanded on in detail in chapter 2, using directed evolution to engineer hydrocarbon biosynthesis pathways presents several challenges mostly related to the physicochemical properties of these molecules. The use of growth-based selection is a powerful approach for directed evolution but requires that hydrocarbon biosynthesis confers a fitness benefit to the host. Engineering microbes to exhibit this behaviour requires a comprehensive use of the synthetic biology toolkit, including modelling and strain design, genetic engineering, heterologous gene expression, and the construction of synthetic auxotrophs.

The growth-coupling strategies considered in this thesis focussed on the byproducts or cofactors of the alkane biosynthesis pathway. These included the consumption of NADPH required for fatty acid- and alkane biosynthesis, as well as the production of formate by ADO. The latter could be tied to cell metabolism by enabling formate to provide reducing power in the form of NADH, or to contribute to essential cofactor synthesis via its fixation to 5,10-mTHF.

Using redox cofactors as selective molecules was shown to be theoretically functional as a way to growth-couple alkane biosynthesis via ADO activity. Coupling formate production to NAD^+ reduction and cell respiration was found to ensure more stringent growth-coupling compared to using NADPH oxidation as the selective principle. Due to these findings the NADHaux strain described in Chapter 6 was developed, but due to time restrictions the strain was not assessed in the context of alkane biosynthesis growth-coupling.

Another approach explored the use of strains auxotrophic for the essential 5,10-mTHF cofactor, named $C_1(G)S_{aux}$. An introduced heterologous metabolic pathway can reinstate 5,10-mTHF biosynthesis via formate utilisation, which ensures growth is formate dependent. We demonstrated that introducing the formate-producing AAR-ADO pathway can rescue growth of such a strain. This represented the first description of a system where alkane biosynthesis was directly growth-coupled.

The dynamic range of an enzyme selection platform is a crucial property of the system. We found that the $C_1(G)S_{aux}$ strain was able to differentiate between supplemental formate concentrations of 0.1 – 1 mM in terms of growth rate. The highest alkane biosynthesis titres recorded in *E. coli* are around 2.5 g/L, which represents approximately 10-12 mM of alkanes depending on the chain length of the product (Fatma et al., 2018). Due to the 1:1 ratio of alkane:formate production by ADO, this translates to 10-12 mM formate as well. As such, $C_1(G)S_{aux}$ would not be able to discern between extremely high producers, although we acknowledge that the formate production rate is probably more important than its final titre.

To ensure there is always a selective pressure for formate production, we can turn to the NADHaux strain developed in this thesis. We found that NADHaux supplemented with fructose, acetate and succinate displayed a dynamic range in terms of growth rate and biomass accumulation between 1-20 mM formate. When omitting succinate NADHaux exhibited a reduced sensitivity, requiring formate concentrations between 20-80 mM to establish growth. Therefore, using a combination of the $C_1(G)S_{aux}$ and NADHaux strains enables selective pressure for formate concentrations over approximately 3 orders of

magnitude. This ensures that ADO can always be under selective pressure, allowing its engineering over a vast activity space (Figure 8.2).

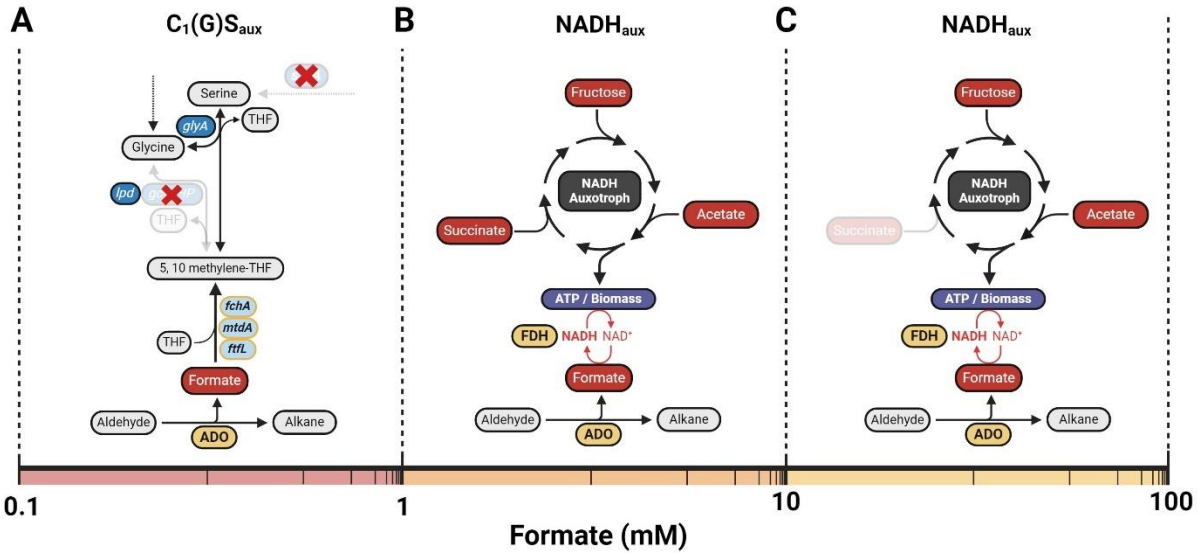


Figure 8.2. The dynamic ranges of $C_1(G)S_{aux}$ and NADHaux ensure a selective pressure for formate biosynthesis over three orders of magnitude. The $C_1(G)S_{aux}$ strain shows growth rate variants between 0.1-1 mM formate. NADHaux supplemented with fructose, acetate, and succinate displayed increasing growth rates between 1-20 mM formate, while omitting succinate reduced this sensitivity to ~20-80 mM formate. Created with Biorender.com.

iii. By engineering a microbial enzyme engineering strain capable of selecting for the activity of an alkane-producing enzyme.

The engineering of biocatalysts through exploitation of evolutionary principles represents a powerful approach to confer them with improved or novel traits. Identifying desirable variants from large mutant pools remains a challenge however, requiring a method that allows either the rapid direct measurement of each individual variant, or a process that enriches the population with only desirable variants. Both of these processes fall under the umbrella of directed evolution (DE). The construction of synthetic auxotrophic strains is gaining popularity in the field of enzyme- and pathway engineering (J. Chen et al., 2022; Orsi et al., 2021). By introducing a certain metabolic deficiency, enzymes that rescue this deficiency effectively will result in growth of the cell, resulting in growth-coupling of enzyme activity.

To bring the modelling results from chapter 5 into practice, an enzyme engineering strain incapable of generating reducing power in the form of NADH was constructed, named

NADHaux. This strain represented a redox cofactor auxotrophic strain, of which several have now been engineered (Nielsen et al., 2023). The generation of an NADH-specific auxotrophic variant has not been reported meaning that this strain is a novel addition to the repertoire.

As discussed earlier, this strain could potentially be used to engineer the ADO or the AAR-ADO pathway. However, besides alkane biosynthesis many other enzymatic activities are viable targets for this strain. Due to its selectivity for NADH, cofactor specificity towards NAD^+ could be engineered in a range of oxidoreductases, as well as providing a way to select for entire NADH-generating metabolic pathways. The fact that NADHaux generates high levels of NADPH could also be exploited to engineer pathways that require the consumption of NAPDH.

iv. By providing fundamental insights into the metabolic plasticity of *E. coli*.

Over the course of the research completed for this thesis it became clear that *E. coli* often behaves contrary to expectations. These divergences from canonical behaviour were often found when significant selective pressures were applied. For example, the growth of *E. coli* using ethanol as a carbon source revealed that the native *adhE* enzyme could support aerobic growth (chapter 4). This was in contradiction to a widely held understanding that the bifunctional aldehyde/alcohol dehydrogenases from *E. coli* is oxygen sensitive *in vivo* (Holland-Staley et al., 2000; Membrillo-Hernández et al., 2000). A similar finding included that the expression of pyruvate-flavodoxin oxidoreductase *ydbK* and pyruvate formate lyase *pflB* contributed to aerobic growth when canonical aerobic metabolic pathways were removed (chapter 7), as well as the observation of formate production in ethanol-grown *E. coli* in aerobic shake flasks, presumed to be from *pflB* activity (chapter 4). While a mechanistic explanation cannot be given for these observations, the metabolic context of these enzymes was usually highly unusual. For example, the deletion of *lpd* and *poxB* (chapter 7) represents a drastic change to cell metabolism for the catabolism of sugars, possibly resulting in unusual regulatory effects and stress responses. Similarly, *E. coli* natively expresses *adhE* to ferment sugars, although in this work ethanol – which can fluidise and permeabilise cell membranes (Dombek & Ingram, 1984) – is used as the sole carbon source instead. The accumulation of (toxic)

intermediates and redox cofactors, expression of stress response proteins, plus the regulatory rewiring required to establish growth in non-native conditions, may similarly lead to non-canonical activities of certain enzymes. Since the degree of oxygen tolerance is a common theme amongst these observations, further work should include oxygen concentration measurements to enable robust conclusions to be drawn.

Another confounding result was related to the NADHaux strain developed in chapter 6. We discovered a variant, M1, that was able to grow lacking any central metabolic pathway for NADH generation. We identified a mutation in the cofactor-interacting subunit of the NADH dehydrogenase complex I (NDH-I), *nuoF*^{E183A}, which is known to broaden the substrate specificity of NDH-I to include NADPH. The ability of M1 to grow rapidly without an obvious source of NADH flux was surprising and represented another example of the metabolic plasticity of *E. coli* when faced with considerable changes to its metabolism.

9. Concluding remarks

Over the last two decades considerable efforts have been made to establish biological drop-in fuel production. Despite significant advances in the field, process performance generally remains unsuitable for commercial exploitation. This thesis reports the successful use of non-sugar carbon sources for hydrocarbon production, as well as demonstrating the viability of strategies enabling the coupling of microbial growth to fuel production. The ideas and methods developed in this thesis are expected to contribute to the development of microbial fuel production by presenting novel approaches to solving some of the long-standing limitations of the field.

10. References

Ambaye, T. G., Vaccari, M., Bonilla-Petriciolet, A., Prasad, S., van Hullebusch, E. D., & Rtimi, S. (2021). Emerging technologies for biofuel production: A critical review on recent progress, challenges and perspectives. *Journal of Environmental Management*, *290*, 112627. https://doi.org/10.1016/J.JENVMAN.2021.112627

Andre, C., Kim, S. W., Yu, X. H., & Shanklin, J. (2013). Fusing catalase to an alkane-producing enzyme maintains enzymatic activity by converting the inhibitory byproduct H2O2 to the cosubstrate O2. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(8), 3191–3196. https://doi.org/10.1073/pnas.1218769110

Arnold, F. H. (1998). When blind is better: Protein design by evolution. Nature Biotechnology 1998 16:7, 16(7), 617–618. https://doi.org/10.1038/nbt0798-617

Bao, L., Li, J. J., Jia, C., Li, M., & Lu, X. (2016). Structure-oriented substrate specificity engineering of aldehyde-deformylating oxygenase towards aldehydes carbon chain length. *Biotechnology for Biofuels*, *9*(1), 185. https://doi.org/10.1186/s13068-016-0596-9

Bartek, T., Blombach, B., Zönnchen, E., Makus, P., Lang, S., Eikmanns, B. J., & Oldiges, M. (2010). Importance of NADPH supply for improved L-Valine formation in Corynebacterium glutamicum. *Biotechnology Progress*, *26*(2), 361–371. https://doi.org/10.1002/BTPR.345

Beller, H. R., Goh, E. B., & Keasling, J. D. (2010). Genes involved in long-chain alkene biosynthesis in micrococcus Luteus. *Applied and Environmental Microbiology*, *76*(4), 1212–1223. https://doi.org/10.1128/AEM.02312-09

Blankschien, M. D., Clomburg, J. M., & Gonzalez, R. (2010). Metabolic engineering of Escherichia coli for the production of succinate from glycerol. *Metabolic Engineering*, *12*(5), 409–419. https://doi.org/10.1016/J.YMBEN.2010.06.002

Bogorad, I. W., Lin, T. S., & Liao, J. C. (2013). Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature 2013 502:7473*, *502*(7473), 693–697. https://doi.org/10.1038/nature12575

Broecker, W. S. (1975). Climatic change: Are we on the brink of a pronounced global warming? *Science*, *189*(4201), 460–463. https://doi.org/10.1126/SCIENCE.189.4201.460

Bruder, S., Moldenhauer, E. J., Lemke, R. D., Ledesma-Amaro, R., & Kabisch, J. (2019). Drop-in biofuel production using fatty acid photodecarboxylase from Chlorella variabilis in the oleaginous yeast Yarrowia lipolytica. *Biotechnology for Biofuels 2019 12:1*, *12*(1), 1–13. https://doi.org/10.1186/S13068-019-1542-4

Buer, B. C., Paul, B., Das, D., Stuckey, J. A., & Marsh, E. N. G. (2014). Insights into substrate and metal binding from the crystal structure of cyanobacterial aldehyde deformylating oxygenase with substrate bound. *ACS Chemical Biology*, *9*(11), 2584–2593. https://doi.org/10.1021/cb500343j

Buijs, N. A., Zhou, Y. J., Siewers, V., & Nielsen, J. (2015). Long-chain alkane production by the yeast Saccharomyces cerevisiae. *Biotechnology and Bioengineering*, *112*(6), 1275–1279. https://doi.org/10.1002/BIT.25522

Butler, N. D., Anderson, S. R., Dickey, R. M., Nain, P., & Kunjapur, A. M. (2023). Combinatorial gene inactivation of aldehyde dehydrogenases mitigates aldehyde oxidation catalyzed by *E. coli* resting cells. *Metabolic Engineering*, *77*, 294–305. https://doi.org/10.1016/J.YMBEN.2023.04.014

Calzadiaz-Ramirez, L., Calvó-Tusell, C., Stoffel, G. M. M., Lindner, S. N., Osuna, S., Erb, T. J., Garcia-Borràs, M., Bar-Even, A., & Acevedo-Rocha, C. G. (2020). In Vivo Selection for Formate Dehydrogenases with High Efficiency and Specificity toward NADP + . *ACS Catalysis*, *10*(14), 7512–7525. https://doi.org/10.1021/acscatal.0c01487

Cavelius, P., Engelhart-Straub, S., Mehlmer, N., Lercher, J., Awad, D., & Brück, T. (2023). The potential of biofuels from first to fourth generation. *PLOS Biology*, *21*(3). https://doi.org/10.1371/JOURNAL.PBIO.3002063

Chang, M., Shimba, K., Hayashi, Y., & Arai, M. (2020). Electrostatic interactions at the interface of two enzymes are essential for two-step alkane biosynthesis in cyanobacteria. *Bioscience, Biotechnology, and Biochemistry*, *84*(2), 228–237. https://doi.org/10.1080/09168451.2019.1677142

Chemler, J. A., Fowler, Z. L., McHugh, K. P., & Koffas, M. A. G. (2010). Improving NADPH availability for natural product biosynthesis in Escherichia coli by metabolic engineering. *Metabolic Engineering*, *12*(2), 96–104. https://doi.org/10.1016/J.YMBEN.2009.07.003

Chen, J., Wang, Y., Zheng, P., & Sun, J. (2022). Engineering synthetic auxotrophs for growth-coupled directed protein evolution. *Trends in Biotechnology*. https://doi.org/10.1016/J.TIBTECH.2022.01.010

Chen, Z., & Liu, D. (2016). Toward glycerol biorefinery: metabolic engineering for the production of biofuels and chemicals from glycerol. *Biotechnology for Biofuels*, *9*(1), 1–15. https://doi.org/10.1186/S13068-016-0625-8

Chin, J. W., & Cirino, P. C. (2011). Improved NADPH supply for xylitol production by engineered Escherichia coli with glycolytic mutations. *Biotechnology Progress*, *27*(2), 333–341. https://doi.org/10.1002/BTPR.559

Claassens, N. J., Satanowski, A., Bysani, V. R., Dronsella, B., Orsi, E., Rainaldi, V., Yilmaz, S., Wenk, S., & Lindner, S. N. (2022). Engineering the Reductive Glycine Pathway: A Promising Synthetic Metabolism Approach for C1-Assimilation. *Advances in Biochemical Engineering/Biotechnology*, *180*, 299–350. https://doi.org/10.1007/10_2021_181

CO2 Emissions in 2022 – Analysis - IEA. (2023). https://www.iea.org/reports/co2-emissions-in-2022

Cooper, C. F. (1978). What Might Man-Induced Climate Change Mean? *Foreign Affairs*, *56*(3), 500. https://doi.org/10.2307/20039915

Crepin, L., Lombard, E., Guillouet, S., Crépin, L., & Guillouet, S. E. (2016). *Metabolic engineering of Cupriavidus necator for heterotrophic and autotrophic alka(e)ne production*. *37*, 92–101. https://doi.org/10.1016/j.ymben.2016.05.002ï

Cronan, J. E. (2021). The Escherichia coli FadR transcription factor: Too much of a good thing? *Molecular Microbiology*, *115*(6), 1080–1085. https://doi.org/10.1111/MMI.14663

Davis, M. S., Solbiati, J., & Cronan, J. E. (2000). Overproduction of acetyl-CoA carboxylase activity increases the rate of fatty acid biosynthesis in Escherichia coli. *Journal of Biological Chemistry*, *275*(37), 28593–28598. https://doi.org/10.1074/jbc.M004756200

Della Valle, S., Tu, W., & Huang, W. E. (2024). Construction of microbial platform chassis for CO2 utilisation. *Current Opinion in Systems Biology*, *37*, 100489. https://doi.org/10.1016/J.COISB.2023.100489

Dombek, K. M., & Ingram, L. O. (1984). Effects of ethanol on the Escherichia coli plasma membrane. *Journal of Bacteriology*, 157(1), 233. https://doi.org/10.1128/JB.157.1.233-239.1984

Fatma, Z., Hartman, H., Poolman, M. G., Fell, D. A., Srivastava, S., Shakeel, T., & Yazdani, S. S. (2018). Model-assisted metabolic engineering of Escherichia coli for long chain alkane and alcohol production. *Metabolic Engineering*, *46*, 1–12. https://doi.org/10.1016/j.ymben.2018.01.002

Fatma, Z., Jawed, K., Mattam, A. J., & Yazdani, S. S. (2016). Identification of long chain specific aldehyde reductase and its use in enhanced fatty alcohol production in *E. coli*. *Metabolic Engineering*, *37*, 35–45. https://doi.org/10.1016/J.YMBEN.2016.04.003

Gao, Y., Zhang, H., Fan, M., Jia, C., Shi, L., Pan, X., Cao, P., Zhao, X., Chang, W., & Li, M. (2020). Structural insights into catalytic mechanism and product delivery of cyanobacterial acyl-acyl carrier protein reductase. *Nature Communications*, *11*(1), 1–11. https://doi.org/10.1038/s41467-020-15268-y

Gong, G., Wu, B., Liu, L., Li, J., Zhu, Q., He, M., & Hu, G. (2022). Metabolic engineering using acetate as a promising building block for the production of bio-based chemicals. *Engineering Microbiology*, *2*(4), 100036. https://doi.org/10.1016/J.ENGMIC.2022.100036

Han, J., Asano, K., Matsumoto, T., Yamada, R., & Ogino, H. (2023). Engineering acyl-ACP reductase with fusion tags enhances alka(e)ne synthesis in Escherichia coli. *Enzyme and Microbial Technology*, *168*, 110262. https://doi.org/10.1016/J.ENZMICTEC.2023.110262

Hoang, A. T., Sirohi, R., Pandey, A., Nižetić, S., Lam, S. S., Chen, W. H., Luque, R., Thomas, S., Arıcı, M., & Pham, V. V. (2022). Biofuel production from microalgae: challenges and chances. *Phytochemistry Reviews 2022 22:4*, *22*(4), 1089–1126. https://doi.org/10.1007/S11101-022-09819-Y

Holland-Staley, C. A., Lee, K., Clark, D. P., & Cunningham, P. R. (2000). Aerobic activity of Escherichia coli alcohol dehydrogenase is determined by a single amino acid. *Journal of Bacteriology*, *182*(21), 6049–6054. https://doi.org/10.1128/JB.182.21.6049-6054.2000

Iacometti, C., Marx, K., Hönick, M., Biletskaia, V., Schulz-Mirbach, H., Dronsella, B., Satanowski, A., Delmas, V. A., Berger, A., Dubois, I., Bouzon, M., Döring, V., Noor, E., Bar-Even, A., & Lindner, S. N. (2022). Activating Silent Glycolysis Bypasses in Escherichia coli. *BioDesign Research*, *2022*. https://doi.org/10.34133/2022/9859643

IEA. (2021). *Energy Statistics 2021*. https://www.iea.org/data-and-statistics/data-tools/energy-statistics-data-browser

Intergovernmental Panel On Climate Change, Climate Change 2014, Impacts, Adaptations, and Vulnerability WG II. (2014).

IPCC. (2023). *AR6 Synthesis Report: Climate Change 2023*. https://www.ipcc.ch/report/ar6/syr/

Iverson, S. V., Haddock, T. L., Beal, J., & Densmore, D. M. (2016). CIDAR MoClo: Improved MoClo Assembly Standard and New E. coli Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synthetic Biology*, *5*(1), 99–103. https://doi.org/10.1021/ACSSYNBIO.5B00124

Janßen, H. J., & Steinbüchel, A. (2014). Fatty acid synthesis in Escherichia coli and its applications towards the production of fatty acid based biofuels. *Biotechnology for Biofuels 2014 7:1*, *7*(1), 1–26. https://doi.org/10.1186/1754-6834-7-7

Jaroensuk, J., Intasian, P., Kiattisewee, C., Munkajohnpon, P., Chunthaboon, P., Buttranon, S., Trisrivirat, D., Wongnate, T., Maenpuen, S., Tinikul, R., & Chaiyen, P. (2019). *Addition of formate dehydrogenase increases the production of renewable alkane from an engineered metabolic pathway*. https://doi.org/10.1074/jbc.RA119.008246

Jeswani, H. K., Chilvers, A., & Azapagic, A. (2020). Environmental sustainability of biofuels: a review. *Proceedings. Mathematical, Physical, and Engineering Sciences*, *476*(2243). https://doi.org/10.1098/RSPA.2020.0351

Jia, C., Li, M., Li, J., Zhang, J., Zhang, H., Cao, P., Pan, X., Lu, X., & Chang, W. (2014). Structural insights into the catalytic mechanism of aldehyde-deformylating oxygenases. *Protein and Cell*, *6*(1), 55–67. https://doi.org/10.1007/s13238-014-0108-2

Johannes, T. W., Woodyer, R. D., & Zhao, H. (2007). Efficient regeneration of NADPH using an engineered phosphite dehydrogenase. *Biotechnology and Bioengineering*, *96*(1), 18–26. https://doi.org/10.1002/BIT.21168

Kang, M.-K., & Nielsen, J. (2016). Biobased production of alkanes and alkenes through metabolic engineering of microorganisms. *Journal of Industrial Microbiology & Biotechnology 2016 44:4*, *44*(4), 613–622. https://doi.org/10.1007/S10295-016-1814-Y

Khara, B., Menon, N., Levy, C., Mansell, D., Das, D., Marsh, E. N. G., Leys, D., & Scrutton, N. S. (2013). Production of Propane and Other Short-Chain Alkanes by Structure-Based Engineering of Ligand Specificity in Aldehyde-Deformylating Oxygenase. *ChemBioChem*, *14*(10), 1204–1208. https://doi.org/10.1002/cbic.201300307

Kim, H. M., Chae, T. U., Choi, S. Y., Kim, W. J., & Lee, S. Y. (2019). Engineering of an oleaginous bacterium for the production of fatty acids and fuels. *Nature Chemical Biology 2019 15:7*, *15*(7), 721–729. https://doi.org/10.1038/s41589-019-0295-5

Kim, S., Lindner, S. N., Aslan, S., Yishai, O., Wenk, S., Schann, K., & Bar-Even, A. (2020). Growth of *E. coli* on formate and methanol via the reductive glycine pathway. *Nature Chemical Biology 2020 16:5*, *16*(5), 538–545. https://doi.org/10.1038/s41589-020-0473-5

Köpke, M., Held, C., Hujer, S., Liesegang, H., Wiezer, A., Wollherr, A., Ehrenreich, A., Liebl, W., Gottschalk, G., & Dürre, P. (2010). Clostridium ljungdahlii represents a microbial production platform based on syngas. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(29), 13087–13092. https://doi.org/10.1073/PNAS.1004716107

Kudo, H., Hayashi, Y., & Arai, M. (2019). Identification of non-conserved residues essential for improving the hydrocarbon-producing activity of cyanobacterial aldehyde-deformylating oxygenase. *Biotechnology for Biofuels*, *12*(1), 1–17. https://doi.org/10.1186/s13068-019-1409-8

Kunjapur, A. M., Tarasova, Y., & Prather, K. L. J. (2014). *Synthesis and Accumulation of Aromatic Aldehydes in an Engineered Strain of Escherichia coli*. https://doi.org/10.1021/ja506664a

Kurosawa, K., Boccazzi, P., de Almeida, N. M., & Sinskey, A. J. (2010). High-cell-density batch fermentation of Rhodococcus opacus PD630 using a high glucose concentration for triacylglycerol production. *Journal of Biotechnology*, *147*(3–4), 212–218. https://doi.org/10.1016/J.JBIOTEC.2010.04.003

Lee, S. Y., Kim, H. U., Chae, T. U., Cho, J. S., Kim, J. W., Shin, J. H., Kim, D. I., Ko, Y. S., Jang, W. D., & Jang, Y. S. (2019). A comprehensive metabolic map for production of bio-based chemicals. *Nature Catalysis 2019 2:1*, *2*(1), 18–33. https://doi.org/10.1038/s41929-018-0212-4

Lee, W.-H., Kim, M.-D., Jin, Y.-S., & Seo, J.-H. (2013). Engineering of NADPH regenerators in Escherichia coli for enhanced biotransformation. *Applied Microbiology and Biotechnology 2013 97:7*, *97*(7), 2761–2772. https://doi.org/10.1007/S00253-013-4750-Z

Lehtinen, T., Virtanen, H., Santala, S., & Santala, V. (2018). Production of alkanes from CO2 by engineered bacteria. *Biotechnology for Biofuels*, *11*(1), 228. https://doi.org/10.1186/s13068-018-1229-2

Li, J., Ma, Y., Liu, N., Eser, B. E., Guo, Z., Jensen, P. R., & Stephanopoulos, G. (2020). Synthesis of high-titer alka(e)nes in Yarrowia lipolytica is enabled by a discovered mechanism. *Nature Communications 2020 11:1*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-19995-0

Liew, F. E., Nogle, R., Abdalla, T., Rasor, B. J., Canter, C., Jensen, R. O., Wang, L., Strutz, J., Chirania, P., De Tissera, S., Mueller, A. P., Ruan, Z., Gao, A., Tran, L., Engle, N. L., Bromley, J. C., Daniell, J., Conrado, R., Tschaplinski, T. J., … Köpke, M. (2022). Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale. *Nature Biotechnology 2022 40:3*, *40*(3), 335–344. https://doi.org/10.1038/s41587-021-01195-w

Lin, F., Das, D., Lin, X. N., & Marsh, E. N. G. (2013). Aldehyde-forming fatty acyl-CoA reductase from cyanobacteria: expression, purification and characterization of the recombinant enzyme. *The FEBS Journal*, *280*(19), 4773–4781. https://doi.org/10.1111/FEBS.12443

Lin, P. P., Jaeger, A. J., Wu, T. Y., Xu, S. C., Lee, A. S., Gao, F., Chen, P. W., & Liao, J. C. (2018). Construction and evolution of an Escherichia coli strain relying on nonoxidative glycolysis for sugar catabolism. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(14), 3538–3546. https://doi.org/10.1073/PNAS.1802191115

Liu, Y., Chen, J., Khusnutdinova, A. N., Correia, K., Diep, P., Batyrova, K. A., Nemr, K., Flick, R., Stogios, P., Yakunin, A. F., & Mahadevan, R. (2020). A novel C-terminal degron identified in bacterial aldehyde decarbonylases using directed evolution. *Biotechnology for Biofuels*, *13*(1), 114. https://doi.org/10.1186/s13068-020-01753-5

Mazhin, S. A., Khankeh, H., Farrokhi, M., Aminizadeh, M., & Poursadeqiyan, M. (2020). Migration health crisis associated with climate change: A systematic review. *Journal of Education and Health Promotion*, *9*(1). https://doi.org/10.4103/JEHP.JEHP_4_20

Meegoda, J. N., Li, B., Patel, K., & Wang, L. B. (2018). A Review of the Processes, Parameters, and Optimization of Anaerobic Digestion. *International Journal of Environmental Research and Public Health*, *15*(10). https://doi.org/10.3390/IJERPH15102224

Mehrer, C. R., Hernández Lozada, N. J., Lai, R.-Y., & Pfleger, B. F. (2017). Production of Fatty Acids and Derivatives by Metabolic Engineering of Bacteria. *Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids: Production of Fuels and Chemicals*, 1–24. https://doi.org/10.1007/978-3-319-31421-1_385-2

Membrillo-Hernández, J., Echave, P., Cabiscol, E., Tamarit, J., Ros, J., & Lin, E. C. C. (2000). Evolution of the adhE gene product of Escherichia coli from a functional reductase to a dehydrogenase: Genetic and biochemical studies of the mutant proteins. *Journal of Biological Chemistry*, *275*(43), 33869–33875. https://doi.org/10.1074/jbc.M005464200

Mendez-Perez, D., Begemann, M. B., & Pfleger, B. F. (2011). Modular synthase-encoding gene involved in α-olefin biosynthesis in Synechococcus sp. strain pcC 7002. *Applied and Environmental Microbiology*, *77*(12), 4264–4267. https://doi.org/10.1128/AEM.00467-11

Meurer, A., & Kern, J. (2021). Fischer–Tropsch Synthesis as the Key for Decentralized Sustainable Kerosene Production. *Energies 2021, Vol. 14, Page 1836*, *14*(7), 1836. https://doi.org/10.3390/EN14071836

Mumm, R. H., Goldsmith, P. D., Rausch, K. D., & Stein, H. H. (2015). Land usage attributed to corn ethanol production in the united states: Sensitivity to technological advances in corn grain yield, ethanol conversion, and co-product utilization. *Fuel Production from Non-Food Biomass: Corn Stover*, *7*(1), 245–280. https://doi.org/10.1201/b18437

Muñoz, I., Flury, K., Jungbluth, N., Rigarlsford, G., Milà i Canals, L., King, H. (2014). Life cycle assessment of bio-based ethanol produced from different agricultural feedstocks. Int J Life Cycle Assess 19, 109–119. https://doi.org/10.1007/s11367-013-0613-1

Nielsen, J. R., Weusthuis, R. A., & Huang, W. E. (2023). Growth-coupled enzyme engineering through manipulation of redox cofactor regeneration. *Biotechnology Advances*, 108102. https://doi.org/10.1016/J.BIOTECHADV.2023.108102

O'Neill, B. C., Oppenheimer, M., Warren, R., Hallegatte, S., Kopp, R. E., Pörtner, H. O., Scholes, R., Birkmann, J., Foden, W., Licker, R., MacH, K. J., Marbaix, P., Mastrandrea, M. D., Price, J., Takahashi, K., Van Ypersele, J. P., & Yohe, G. (2017). IPCC reasons for concern regarding climate change risks. *Nature Climate Change*, *7*(1), 28–37. https://doi.org/10.1038/nclimate3179

Orsi, E., Claassens, N. J., Nikel, P. I., & Lindner, S. N. (2021). Growth-coupled selection of synthetic modules to accelerate cell factory development. *Nature Communications 2021 12:1*, *12*(1), 1–5. https://doi.org/10.1038/s41467-021-25665-6

Rahman, Z., Sung, B. H., Yi, J. Y., Bui, L. M., Lee, J. H., & Kim, S. C. (2014). Enhanced production of n-alkanes in escherichia coli by spatial organization of biosynthetic pathway enzymes. *Journal of Biotechnology*, *192*(Part A), 187–191. https://doi.org/10.1016/j.jbiotec.2014.10.014

Rajakovich, L. J., Nørgaard, H., Warui, D. M., Chang, W. C., Li, N., Booker, S. J., Krebs, C., Bollinger, J. M., & Pandelia, M. E. (2015). Rapid Reduction of the Diferric-Peroxyhemiacetal Intermediate in Aldehyde-Deformylating Oxygenase by a Cyanobacterial Ferredoxin: Evidence for a Free-Radical Mechanism. *Journal of the American Chemical Society*, *137*(36), 11695–11709. https://doi.org/10.1021/jacs.5b06345

Rodriguez, G. M., & Atsumi, S. (2014). Toward aldehyde and alkane production by removing aldehyde reductase activity in Escherichia coli. *Metabolic Engineering*, *25*, 227–237. https://doi.org/10.1016/j.ymben.2014.07.012

Rude, M. A., Baron, T. S., Brubaker, S., Alibhai, M., Del Cardayre, S. B., & Schirmer, A. (2011). Terminal olefin (1-alkene) biosynthesis by a novel P450 fatty acid decarboxylase from Jeotgalicoccus species. *Applied and Environmental Microbiology*, *77*(5), 1718–1727. https://doi.org/10.1128/AEM.02580-10

Rui, Z., Harris, N. C., Zhu, X., Huang, W., & Zhang, W. (2015). Discovery of a Family of Desaturase-Like Enzymes for 1-Alkene Biosynthesis. *ACS Catalysis*, *5*(12), 7091–7094. https://doi.org/10.1021/ACSCATAL.5B01842

Rui, Z., Li, X., Zhu, X., Liu, J., Domigan, B., Barr, I., Cate, J. H. D., & Zhang, W. (2014). Microbial biosynthesis of medium-chain 1-alkenes by a nonheme iron oxidase. *Proceedings of the National Academy of Sciences*, *111*(51), 18237–18242. https://doi.org/10.1073/PNAS.1419701112

Schirmer, A., Rude, M. A., Li, X., Popova, E., & Del Cardayre, S. B. (2010). Microbial biosynthesis of alkanes. *Science*, *329*(5991), 559–562. https://doi.org/10.1126/science.1187936

Shafer, W. D., Gnanamani, M. K., Graham, U. M., Yang, J., Masuku, C. M., Jacobs, G., & Davis, B. H. (2019). Fischer–Tropsch: Product Selectivity–The Fingerprint of Synthetic Fuels. *Catalysts 2019, Vol. 9, Page 259*, *9*(3), 259. https://doi.org/10.3390/CATAL9030259

Shakeel, T., Gupta, M., Fatma, Z., Kumar, R., Kumar, R., Singh, R., Sharma, M., Jade, D., Gupta, D., Fatma, T., & Yazdani, S. S. (2018). A consensus-guided approach yields a heat-stable alkaneproducing enzyme and identifies residues promoting thermostability. *Journal of Biological Chemistry*, *293*(24), 9148–9161. https://doi.org/10.1074/jbc.RA117.000639

Sharma, A., Shakeel, T., Gupta, M., Rajacharya, G. H., & Yazdani, S. S. (2021). Biophysical and structural studies reveal marginal stability of a crucial hydrocarbon biosynthetic enzyme acyl ACP reductase. *Scientific Reports 2021 11:1*, *11*(1), 1–21. https://doi.org/10.1038/s41598-021-91232-0

Singh, V. K., Jha, S., Rana, P., Soni, R., Lalnunpuii, R., Singh, P. K., Sinha, R. P., & Singh, G. (2024). Cyanobacteria as a Biocatalyst for Sustainable Production of Biofuels and Chemicals. *Energies 2024, Vol. 17, Page 408*, *17*(2), 408. https://doi.org/10.3390/EN17020408

Song, M., Zhang, X., Chen, Y., Zhang, Q., Chen, L., Liu, J., & Ma, L. (2023). Hydroprocessing of lipids: An effective production process for sustainable aviation fuel. *Energy*, *283*, 129107. https://doi.org/10.1016/J.ENERGY.2023.129107

Song, X., Yu, H., & Zhu, K. (2015). Improving alkane synthesis in Escherichia coli via metabolic engineering. *Applied Microbiology and Biotechnology 2015 100:2*, *100*(2), 757–767. https://doi.org/10.1007/S00253-015-7026-Y

Sorigué, D., Légeret, B., Cuiné, S., Blangy, S., Moulin, S., Billon, E., Richaud, P., Brugière, S., Couté, Y., Nurizzo, D., Müller, P., Brettel, K., Pignol, D., Arnoux, P., Li-Beisson, Y., Peltier, G., & Beisson, F. (2017). An algal photoenzyme converts fatty acids to hydrocarbons. *Science*, *357*(6354), 903–907. https://doi.org/10.1126/science.aan6349

Sun, S., Ding, Y., Liu, M., Xian, M., & Zhao, G. (2020). Comparison of Glucose, Acetate and Ethanol as Carbon Resource for Production of Poly(3-Hydroxybutyrate) and Other Acetyl-CoA Derivatives. *Frontiers in Bioengineering and Biotechnology*, *8*. https://doi.org/10.3389/FBIOE.2020.00833

Valdivia, M., Galan, J. L., Laffarga, J., & Ramos, J. (2016). Biofuels 2020: Biorefineries based on lignocellulosic materials. *Microbial Biotechnology*, *9*(5), 585. https://doi.org/10.1111/1751-7915.12387

Wang, Q., Bao, L., Jia, C., Li, M., Li, J. J., & Lu, X. (2017). Identification of residues important for the activity of aldehyde-deformylating oxygenase through investigation into the structure-activity relationship. *BMC Biotechnology*, *17*(1), 1–9. https://doi.org/10.1186/s12896-017-0351-8

Wang, Q., Huang, X., Zhang, J., Lu, X., Li, S., & Li, J. J. (2014). Engineering self-sufficient aldehyde deformylating oxygenases fused to alternative electron transfer systems for efficient conversion of aldehydes into alkanes. *Chemical Communications*, *50*(33), 4299–4301. https://doi.org/10.1039/c4cc00591k

Warui, D. M., Pandelia, M.-E., Rajakovich, L. J., Krebs, C., Bollinger, J. M., & Booker, S. J. (2015). Efficient Delivery of Long-Chain Fatty Aldehydes from the *Nostoc punctiforme* Acyl–Acyl Carrier Protein Reductase to Its Cognate Aldehyde-Deformylating Oxygenase. *Biochemistry*, *54*(4), 1006–1015. https://doi.org/10.1021/bi500847u

Weber, E., Engler, C., Gruetzner, R., Werner, S., & Marillonnet, S. (2011). A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLOS ONE*, *6*(2), e16765. https://doi.org/10.1371/JOURNAL.PONE.0016765

Xiberras, J., Klein, M., & Nevoigt, E. (2019). Glycerol as a substrate for Saccharomyces cerevisiae based bioprocesses – Knowledge gaps regarding the central carbon catabolism of this 'non-fermentable' carbon source. *Biotechnology Advances*, *37*(6), 107378. https://doi.org/10.1016/J.BIOTECHADV.2019.03.017

Xie, S., Li, Z., Luo, S., & Zhang, W. (2024). Bioethanol to jet fuel: Current status, challenges, and perspectives. *Renewable and Sustainable Energy Reviews*, *192*, 114240. https://doi.org/10.1016/J.RSER.2023.114240

Xu, P., Qiao, K., Ahn, W. S., & Stephanopoulos, G. (2016). Engineering Yarrowia lipolytica as a platform for synthesis of drop-in transportation fuels and oleochemicals. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(39), 10848–10853. https://doi.org/10.1073/PNAS.1607295113

Yishai, O., Lindner, S. N., Gonzalez de la Cruz, J., Tenenboim, H., & Bar-Even, A. (2016). The formate bio-economy. *Current Opinion in Chemical Biology*, *35*, 1–9. https://doi.org/10.1016/J.CBPA.2016.07.005

Youngquist, J. T., Schumacher, M. H., Rose, J. P., Raines, T. C., Politz, M. C., Copeland, M. F., & Pfleger, B. F. (2013). Production of medium chain length fatty alcohols from glucose in Escherichia coli. *Metabolic Engineering*, *20*, 177–186. https://doi.org/10.1016/J.YMBEN.2013.10.006

Zelcbuch, L., Lindner, S. N., Zegman, Y., Slutskin, I. V., Antonovsky, N., Gleizer, S., Milo, R., & Bar-Even, A. (2016). Pyruvate formate-lyase enables efficient growth of Escherichia coli on acetate and formate. *Biochemistry*, *55*(17), 2423–2426. https://doi.org/10.1021/acs.biochem.6b00184

Zhang, F., Ouellet, M., Batth, T. S., Adams, P. D., Petzold, C. J., Mukhopadhyay, A., & Keasling, J. D. (2012). Enhancing fatty acid production by the expression of the regulatory transcription factor FadR. *Metabolic Engineering*, *14*(6), 653–660. https://doi.org/10.1016/J.YMBEN.2012.08.009

Zhou, Y. J., Kerkhoven, E. J., & Nielsen, J. (2018). Barriers and opportunities in bio-based production of hydrocarbons. *Nature Energy 2018 3:11*, *3*(11), 925–935. https://doi.org/10.1038/s41560-018-0197-x