AN EVOLVING SYNTHETIC BIOLOGICAL PLATFORM FOR ALKANE PRODUCTION



Joseph Kennerley

St Edmund Hall

Department of Engineering Science

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2022

Acknowledgements

The years I have spent in Oxford have been among the happiest and most fulfilling of my life. My experiences here, in this beautiful city and its magnificent university, have been transformative. It was here where I first felt as though anything was possible.

Bright times are lit by special people, and I want to give thanks to some of them here.

To my parents, David and Barbara, who have loved and lifted me since my first breath. Your unwavering generosity, kindness, and faith has given me strength and a treasured sense of peace throughout my life. Not once, in all my journeys, have I looked behind me for reassurance and found you missing. All that I build, I build in honour of you.

To my beloved wife, Raisha, who is always at my side and has encouraged me to strive in all that I do. You fill every corner of our shared moments with joy until they overflow. Thank you for all the happy memories made together in this special place, where we have found our home, and for all those yet to come.

To my teachers, Professors Wei Huang and Ian Thompson, for their trust and guidance. You first opened the door for me into Oxford, and I will always be grateful for that opportunity. I look forward to our continued work together into the future.

To Rob Habgood, for your mentorship in my early years at Oxford, and your continued friendship thereafter.

To Ben White, for welcoming me into Teddy Hall, and for your part in my proposal to Raisha on that cold November evening, in the garden behind St Peter-in-the-East.

To my friends across the ocean, in Australia, who I hope will forgive my long absence.

To all those in the Farmhouse, for your conversation, friendship, and support: Jack, Boon, Reza, Jochem, Yutong, Simona, Morgan, Laura, Nick, Fariza, Chengchen, Catherine, Weiming, Tong, Jiabao, Nahla, Hakmin, Calvin, Josh, Helen, Karen. I could not have hoped for a better community in which to work and thrive.

And to John McGeever and David Cavalier, head chefs during my time at Teddy Hall, whose works nourished the spirits of many Aularians.

Abstract

Fossil fuels are finite and polluting, but underpin almost every system and product we use in the modern world. Anthropogenic greenhouse gas emissions from fossil fuel combustion are the major driver of climate change, yet demand for these fuels continues to grow. Alleviating global reliance on fossil carbon, especially in the transport sector, demands a drop-in compatible substitute with comparable energy density and fuel chemistry to diesel, jet fuel, or gasoline. If derived from atmospheric carbon, such a fuel could be central to future low-carbon circular economies. Liquid biofuels, produced via industrial microbial fermentation, have similar characteristics to their petroleum counterparts with a much smaller carbon footprint.

Microbial biosynthesis of advanced biofuels is currently uneconomical, but has been improved by rational engineering efforts. The physiological complexity inherent in living systems often limits such approaches, but can be sidestepped by harnessing natural evolutionary processes to optimise cellular metabolism for production of valuable molecules.

This work presents a novel biosensor-based directed evolution system for improving biosynthesis of alkanes – straight chain hydrocarbons useful as biofuel – in the bacterium *Acinetobacter baylyi* ADP1: the "artificial evolution", or Artevo, system.

By placing a growth-essential gene under control of genetic regulators sensitive to alkanes, the Artevo system links alkane productivity of the host cell to fitness. High-producing cells grow faster and outcompete their neighbours, resulting in generational enrichment of high-performing alkane biosynthesis genes within a growing population.

The Artevo system can interrogate a small pool of diverse alkane biosynthesis genes and distinguish between them based on relative productivity, conferring a growth advantage to host cells with the highest alkane yields. This system consolidates screening, selection, and amplification of useful biosynthetic genes from potentially large libraries into a single platform, and could form the basis for future continuous directed evolution strategies as it is built in a naturally transformable and recombinant microbe.

Contents

CHAPTER 1 INTRODUCTION 1

Aims and objectives of this thesis 6

Chapter breakdown 6

CHAPTER 2 LITERATURE REVIEW 8

Part I: Energy, climate, motion, and metal 9

Climate change 10

The global energy landscape 11

Metals scarcity and land use 15

Energy density 18

The ideal sustainable fuel 20

Part II: Molecular machines and directed evolution 21

Microbial biosynthesis of fuels 22

Directed evolution 28

Research gap 33

CHAPTER 3 A NOVEL HOST FOR DIRECTED EVOLUTION 35

Design of the gene circuit 35

Diversity generation 37

The alkane biosensor 38

Chassis selection 39

Natural competence 41

Selecting the rescue gene 42

Constructing the host chassis 44

Methods and results 44

Characterisation of the *dapB* deletion mutant 50

Discussion 52

CHAPTER 4 BUILDING THE EVOLUTION CIRCUIT 54

Part I: Assembly and initial tuning 55
Assembly of the Artevo circuit 55
Initial tuning and upgrading of the Artevo circuit 61
Impact of culture vessel on apparent circuit leakiness 68
Part I. Discussion 71
Part II: Fine-tuning the Artevo circuit 74
Transcriptional interference for control of gene expression 74
Application of transcriptional interference in Artevo strains 79
Growth response of Artevo strains to exogenous alkane 79
Materials and methods 82
Part II. Discussion 86

CHAPTER 5 ALKANE BIOSYNTHESIS 89

Enzymes for alkane biosynthesis 89
Installation of alkane biosynthesis machinery 92
Growth response to native alkane production 95
Enrichment of high-performing gene sequences 97
Materials and methods 101
Discussion 103

CHAPTER 6 FUTURE WORK 109

Part I: Next steps 110
Part II: Long-term plans 114
In vivo diversity generation with variable mutation rate 119
In situ continuous directed evolution 120
Discussion 123

CHAPTER 7 CONCLUSION 125

REFERENCES 128

SUPPLEMENTARY INFORMATION 150

Foreword

Time is precious, and how we use it reflects what matters to us. I appreciate any time you, the reader, are willing to devote to my work. I put a lot of time into writing it, after all, because your experience when reading it matters to me. I want you to be able to enjoy what you read here, to feel excited by the things I've built and engaged by how they're presented. Because engaged and excited is how I felt when I made these discoveries and built these things. Not always, granted, but most of the time.

I was once told, by a fellow academic, that composing a thesis is "the hardest you will ever work for something nobody will ever fully read." I am not cynical enough to subscribe completely to this notion, but there's a truth here to acknowledge: academics, like most people, are busy. Everyone has lots to do and consider and decide all the time. You, dear reader, have priorities. This work may well not be first among them, and that's ok; it's an entirely understandable consequence of the modern human experience. Your time matters, and what you take away from the time you spend with my work is all the more important because of that.

There is a lot of information in this thesis, and this foreword aims to help you catch the most important bits, highlighting the landmarks to visit on your journey through these pages. First though, if you haven't done so already, please read the abstract. You may need some context for this next part.

A whistle-stop tour of this work begins, unsurprisingly, by reading the introduction in Chapter 1. This gives a broad overview of the field of synthetic biology and the bigger issues this project aims to help solve. Perhaps counterintuitively, it might then be best to consult *Figure 31* at the end of Chapter 6, which summarises how the final system I built in this project works. It should help provide some perspective for the rest of the information here. Then, come back and read the 'Research gap' section at the end of Chapter 2.

On your way through Chapter 3, make sure to consult the following figures: *Figure 6* shows the principle behind the system I've built and, later in Chapter 4, *Figure 19* adds extra details and features. If you're interested in how the system was constructed, see *Figure 13* in Chapter 3 and *Figure 20* in Chapter 4. *Figure 22*, at the end of Chapter 4, is a key figure highlighting what tuning efforts applied to this system were able to achieve.

Chapter 5 is a proof-of-concept test of the system, with *Figure 26* demonstrating that it should be able to do what it was designed to do. With all the previous information in mind, take another look at *Figure 31*, at the end of Chapter 6, before reading the conclusion in Chapter 7.

Like most theses, this one will probably not appeal to all audiences. Nevertheless, I hope a broad readership can find it both accessible and worthwhile. At the end of the day, it was written for everyone; academics, industrialists, experts, laymen, my examiners, my colleagues, friends, strangers, and you.

Thank you for reading.

Chapter 1

Introduction

The fourth industrial revolution, life's strange alchemy, and the greatest challenges of our time

Four great industrial revolutions mark major milestones in the history of humankind. The first three saw the rise of powered machinery, brought widespread electrification, and gave us our now-ubiquitous digital technologies. In each of these revolutions, astonishing technical innovation laid the foundation for a new kind of civilisation, bringing rapid economic growth and sweeping social change.[1] Now, new technologies emerging from the union of computing, engineering and biology are heralds of the next great age of change, where unprecedented control over biological processes promises to transform fundamentally how we live and how we care for Earth. This is the fourth industrial revolution.

Klaus Schwab, executive chairman of the World Economic Forum, characterises the fourth industrial revolution as "a fusion of technologies that blur the lines between the physical, digital, and biological spheres"[2], and notes that these new technologies could disrupt almost every industry in every country. Harvard University's George M. Church, known for his pioneering work in synthetic biology and genome engineering,

said in 2018 that "Moore's law is reaching a plateau, while biotechnology is going through super-exponential growth – where it's improving by factors of 10 per year in cost/benefit."[3] Biotechnology today is what digital technology was in the 1980s: beginning to revolutionise diverse sectors and undergoing explosive growth. To quote two colossal figures of the digital revolution: "If I were a teenager today, I would be hacking biology. If you want to change the world in some big way, that's where you should start – biological molecules." – Bill Gates[4], and "The biggest innovations of the 21st century will be at the intersection of biology and technology. A new era is beginning." – Steve Jobs[5]. Modern biotechnology, especially synthetic biology, is a paradigm shift in humanity's relationship with nature, promising tremendous impact in health, agriculture, materials, energy, and environmental management.

Living organisms, especially multicellular ones, are fantastically intricate and have an unparalleled ability to transform simple matter into complex systems. Consider a goat, for example; it has powerful optical and auditory sensors (eyes, ears, nose), versatile chemical harvesting and transport facilities (stomach, bloodstream), and is highly mobile across unworked terrain (legs). Life is such a sophisticated matter converter that a mother goat can construct another goat from only the plant matter she's consumed. Amazingly, every tiny piece of this new goat that harbours DNA contains all the information required to build, operate, and maintain the entire goat. Making goats – and consequently all their useful products like milk, hair, meat, and horn – out of plants seems like a tall order for traditional chemical engineering, but living systems operate in a different realm of complexity. The takeaway from this is simple but it underpins much of what is driving the fourth industrial revolution: biology is the world's most sophisticated manufacturing platform, able not only to self-maintain, adapt and reproduce, but also to convert diverse matter into valuable products under mild conditions.

Synthetic biologists work at the intersection of engineering and biology to design and construct new bio-based systems with extraordinary capabilities – from next-generation data storage[6] to custom microbes producing fine chemicals[7], textiles[8], and life-saving drugs[9] to early artificial life[10]. The UK government recognises that synthetic biology lies at the innovative heart of a £220 billion national bioeconomy[11], citing the field as one of the Eight Great Technologies which will be core to the UK's future industrial strategy[12].

Around the world, investment in synthetic biology companies reached $18 billion USD (~£13.8 billion GBP) in 2021, almost matching the entire amount raised for this field since 2009, despite a lean fourth quarter[13]. These companies are innovating within a

vast global bioeconomy, foundational to so many aspects of our lives; our food, our clothes, our medicines, and innumerable other products we rely on are all built with biology. A 2020 study by the McKinsey Global Institute examined a pipeline of 400 technologies and innovations across diverse bio-based sectors, almost all of which are feasible to implement today, and found their application alone could have direct economic impact of $4 trillion USD per year over the next one to two decades[14]. Such is the scale and importance of the biological systems that support our lives, and the potential for positive impact innovations in this space can bring.

For new biotechnologies, the future is bright: the path from laboratory to industry is well-travelled and investor faith is high. Biotechnology is increasingly recognised as being well-suited to tackle some of the greatest global problems of our age[15]: global nutrition, sustainable consumption and production, and climate change. Legendary naturalist Sir David Attenborough recently called this last issue humanity's "greatest threat in thousands of years."[16]

Fossil carbon is finite and polluting, but demand for its derivative products – plastics and other materials, oils and solvents, fuels, and many more – is rising across the world and outpacing decarbonisation efforts, especially in the energy sector[17]. This is despite an increasing share of renewables in the energy mix – due largely to the falling cost of solar and wind power[18] – and rising global awareness of climate change, which is due mostly to extreme weather events affecting more communities worldwide[19].

Sustainable alternatives to fossil materials are key to ecological preservation. In 2018, the year I began working on this research project and when much of these early chapters were first written, the Intergovernmental Panel on Climate Change (IPCC) released a special report detailing the immense environmental damage consequent to allowing anthropogenic global warming to exceed 1.5°C[20]. The IPCC predicted that this threshold may be breached as early as 2030 and highlighted the need for urgency, advising an "unprecedented shift in energy systems and transport" to keep climate change within safe levels[21]. Just three years later, in 2021, the IPCC released a new report predicting that "unless there are immediate, rapid and large-scale reductions in greenhouse gas emissions, limiting warming to close to 1.5°C or even 2°C will be beyond reach"[22]. Side by side, these two reports impart hard meaning; the solutions we develop to this global problem, technical and political, seem at present unable to keep pace with the worsening climate crisis. It is an uphill battle fought on a steepening hill. Historic human activity has created momentum within parts of our planet's climate that is irreversible in the foreseeable future, and which may overshadow short-term positive effects of climate action. Still, decisive and sustained reductions in greenhouse gas

emissions will limit long-term climate change, and these reductions begin with the primary driver of climate change: carbon dioxide (CO_2) produced during combustion of fossil fuels.

Transportation accounts for 24% of CO_2 released worldwide: three-quarters of these emissions come from land vehicles and the remainder from aviation and marine transport[23]. The transport sector is the fastest-growing contributor to global climate emissions and, while electric road vehicles are gaining popularity, aircraft and ships require fuels with high energy-to-weight ratios that batteries cannot currently provide. Increasing demand for air travel and the lack of a comparable electrical alternative to jet fuel could result in a sevenfold increase in CO_2 emissions from the aviation sector by mid-century[24]. The overwhelming majority of vehicles in the transport sector are powered by fossil hydrocarbons and electricity generated using fossil sources. The processing, distribution, storage, and engine infrastructure that handles liquid hydrocarbon fuels is so extensive and the properties of these fuels – their huge energy density, high stability, and carbon-based nature – are so attractive that it is difficult to envision our modern civilisation operating without fossil energy. Solving such a global problem demands a unity of vision and an alacrity uncharacteristic of historical energy transitions, which took decades even when adopting energies as versatile and effective as petroleum[25].

While renewable electricity is increasingly present on the international energy scene[18], and electric vehicles feature heavily in the zeitgeist of modern mobility, issues with energy storage efficiency[26], battery cost[27], infrastructure deployment, energy policy[28] and metals availability[29] leave plenty of room for other solutions. Managing climate change will require innovative new liquid chemical fuels which can deliver the many benefits of petroleum without the carbon footprint.

Biofuel is essentially the recyclable version of petroleum fuels, being synthesised from a reservoir of atmospheric rather than lithospheric carbon. Combustion products of biofuels re-enter the atmosphere and can again be reconcentrated, often by plant biomass, into useful carbon molecules. The overall process is net carbon neutral. Biofuels, whether derived from plant biomass or synthesised directly using renewable electricity and atmospheric carbon – such as in the case of solar fuels or electro fuels – are well established but not yet competitive with their petroleum counterparts. Biosynthetic pathways to produce a variety of useful fuel molecules do exist, but they are often slow or inefficient[30]. Optimising these pathways and their microbial hosts to produce high yields of fuel reliably would be a huge boon to the field of industrial biotechnology and the energy sector. Energy economies based on circular carbon

derived directly from the atmosphere would also help to depoliticise the global energy landscape, which is currently influenced by geographical distribution of fossil energy reserves. Technologies that can concentrate universally available atmospheric carbon back into useful liquid fuels would help insulate energy systems against geopolitical issues and create independence from oil-producing nations.

When approaching this optimisation problem with traditional engineering strategies, we encounter the inherent limits of rational design. Our ability to engineer any system, such as a microorganism, is limited by our knowledge of that system and living systems are notoriously complex. Myriad factors contribute to the behaviour of even simple living systems like bacteria, but there is a way to target all these factors at once and optimise a living system towards a given goal: directed evolution.

Directed evolution is a massively parallel black-box approach to the design-build-test cycle of traditional engineering. Frances Arnold, Sir Greg Winter, and George Smith shared the 2018 Nobel Prize in Chemistry for their work using evolutionary forces to shape enzymes and phages[31], so directed evolution is already an established and powerful tool. This thesis discusses the creation of a novel biosensor-based bacterial directed evolution system – the "artificial evolution" system, hereafter called the Artevo system – targeting the biosynthesis of alkanes, high-impact bioproducts which could be used as drop-in biofuels or as platform chemicals for industrial chemistry[32].

Earth's climate is unequivocally intertwined with human action and the choices we will make in the next decade[33]. Many of the competing solutions we see today will undoubtedly establish themselves and settle into their niches, giving rise to a versatile and sustainable system of mixed energy sources and reservoirs. But we are not on-track to deliver such a system in time to avert climate catastrophe[20]. A liquid chemical fuel, compatible with existing transport infrastructure and that can be deployed rapidly as a drop-in alternative to petroleum fuels, will be pivotal in tackling the challenge of climate change. Such a fuel is especially needed in these most crucial transitionary years leading up to 2050, before repairing the damage to our environment becomes unaffordable. This is what the directed evolution system developed in this project aims to help create: a living manufacturing platform that can synthesise valuable biofuels from atmospheric carbon on an industrial scale.

Aims and objectives of this thesis

The objectives of this project have been defined as follows:

i. **Design a directed evolution system** that couples growth rate of a microbial host to production of a valuable metabolite. Alkanes have been chosen as the metabolite-of-interest, due to their value as drop-in compatible biofuels.

ii. **Construct the gene circuitry** driving this directed evolution system in a next-generation bacterial chassis. *Acinetobacter baylyi* was chosen as the bacterial host due to its genetic malleability and robust metabolism.

iii. **Tune the directed evolution system** such that differences in local alkane concentration are detectable by the host cells and proportional to cell fitness. Higher alkane concentrations should translate to greater growth rates.

iv. **Demonstrate the directed evolution system** can distinguish between alkane biosynthesis modules with varying relative productivities. In a combined culture of diverse gene variants, the system should selectively enrich the most productive gene module, as cells hosting that module should have the highest growth rate.

v. **Apply the directed evolution system** to interrogate a mutagenized library of biosynthetic gene modules. Isolate and sequence enriched modules to facilitate structural analysis of high-performing gene variants.

The following section provides a breakdown of content in the chapters of this thesis, identifying which chapters address specific objectives.

Chapter breakdown

The following chapter, Chapter 2, examines the global energy landscape, looking at the primary energies driving transportation, issues with resource extraction and metals dependency, and the challenges inherent to transitioning away from fossil fuels. It also reviews the state of the art in directed evolution and microbial biotechnology, as it pertains to industrial biosynthesis of fuels and specialty chemicals. The information presented here on energy is broad and high-level, and should be accessible by most readers. Discussions regarding biological science are more technical and assume a

familiarity with cell theory, gene theory, evolutionary theory, and a basic understanding of synthetic biology. The goal of Chapter 2 is to present a review of the existing literature, provide context for this project, and clarify how the overall project direction was established and broken down into manageable goals.

From there, Chapter 3 describes the design decisions underpinning the practical work of this project and the construction of fundamental aspects of the Artevo system, specifically the host mutant bacterial strain. This chapter addresses *objective i.* in full, and the generation of a mutant bacterial host forms part of *objective ii.*

Chapter 4 details the building of a gene circuit to drive directed evolution in the Artevo system, and the approaches used to tune the system's behaviour. *Objective ii.* is completed in this chapter, and this chapter's second half covers *objective iii.* It follows the development of a basic proof of concept circuit into a system that is ready to use in directed evolution applications.

Chapter 5 showcases an exciting practical test of the Artevo system and discusses the extent of its current abilities, addressing *objective iv.* Here we see the Artevo system demonstrate a basic principle of natural selection that enables most directed evolution processes: survival of the fittest.

Chapter 6 provides a roadmap for future work on this project, including plans for further development and tuning of the Artevo system and its practical applications. This chapter provides plans for completing *objective v.* It also presents blueprints for additional technologies that build upon the Artevo system, in pursuit of the eventual construction of a consolidated continuous directed evolution platform.

Finally, Chapter 7 concludes this thesis with a reflection on the achievements of this work and its contributions to the field of synthetic biology. In summary, these are:

i. Design and development of an artificial evolution – "Artevo" – system, able to interrogate a small pool of diverse alkane biosynthesis genes and distinguish between them based on relative productivity.

ii. Tuning, based on transcriptional interference, of the biosensor-based directed evolution circuit underpinning the Artevo system, such that host cells produce a graded growth response proportional to local alkane concentration.

iii. Generation of a novel auxotrophic mutant in *Acinetobacter baylyi* to house the gene circuitry driving this Artevo system.

Chapter 2

Literature review

This chapter is in two parts. Part I presents an overview of the global energy landscape, discussing alternative fuels, resource extraction, and energy storage; each providing context for various approaches to microbial biofuel production. Part II examines the biotechnology behind the production of biofuels and reviews the use of directed evolution strategies for bioprocess improvement.

Energy transitions in our society are era-defining and multifaceted. Historical and ongoing energy transitions have been covered comprehensively by authors such as Professor Vaclav Smil of the University of Manitoba, one of the world's foremost thinkers on energy of all kinds. The discussion you will find in this chapter is comparatively limited in depth and makes no independent analyses of raw climate or energy data. The scope of this review is broad though, as I have endeavoured to collate encompassing and compelling support for the Artevo system and the class of molecule it aims to produce: alkanes.

The key points of this chapter can be summarised as follows: fossil fuels are finite and polluting but represent the major form of energy powering our modern world. Applications such as aviation and shipping demand liquid fuels with a high energy density and are consequently difficult to electrify and decarbonise. Atmospheric carbon

represents an abundant, geographically unbounded reservoir of material from which to construct energy-carrying molecules. Living systems, such as plants or carbon-fixing microbes, represent an efficient way of concentrating atmospheric carbon into useful forms, such as feedstock for subsequent refining and upgrading. Rational engineering of microbes for improved production of biofuels is effective but limited by the complexity and scalability of biological systems. Directed evolution allows us to overcome that limitation, expanding the capabilities of microbial systems towards efficient industrial biosynthesis of drop-in compatible biofuels.

Part I: Energy, climate, motion, and metal

Energy is the universal capital underpinning all processes. By its transformation from one form to another, all useful work is done. The ponderous orbits of planets, the colossal motion of ocean currents, the steady growing of a tree and the light of a candle: all are manifestations of some kind of energy conversion. Almost all forms of energy available on earth – excepting tidal, geothermal, and to some extent nuclear – are originally derived from solar fusion processes radiating light and heat. The photosynthetic capture of this solar energy by phytomass is foundational to all higher life on Earth: all animate power depends in some way on this transformation of solar to chemical energy. Ancient phytomass buried in the Earth's crust eventually becomes fossil material such as coal, petroleum, and natural gas. These are some of the most concentrated and versatile stores of energy ever to be discovered by humanity and great efforts have been made in human history to control them[25].

Humans are exceptional in their ability to harness extrasomatic energies to exceed the limits of animate power. Utilising the vast energy embodied in fossil matter has given mankind great constructive and destructive powers, and a huge influence over nature. Before its loss during the 2022 war in Ukraine, the AN-225 Mryia cargo aircraft – itself weighing 285 tonnes – was able to fly 189 tonnes of cargo over 3000 kilometres in a single day[34]. As we extract and combust fossil fuels to provide energy for such feats, we shatter ancient hydrocarbon molecules made of atmospheric carbon that was concentrated into biomass millions of years ago, dispersing this carbon into the environment once more. Human activity since the industrial revolution, driven by fossil energy, channels concentrated lithospheric reservoirs of carbon into the huge, dilute reservoir of our atmosphere. In this way, we harness the accumulated energy of millions of years of sunlight captured in chemical bonds yet place immense strain on our environment.

Climate change

The Intergovernmental Panel on Climate Change (IPCC) has reported an increase in global average atmospheric carbon dioxide (CO_2) concentration from ~280 parts per million (ppm) in the pre-industrial era to over 410ppm in 2021[22]. The heat-trapping effect of CO_2 and other greenhouse gases has caused an average global temperature rise of 1.09°C since the late 19th century[22], a rate of warming ten times faster than Earth experienced during previous natural temperature cycles[35]. In their 2018 special report, and again in the first part of their Sixth Assessment Report released in 2021, the IPCC explored the environmental consequences of allowing average global temperature rise to exceed 1.5°C. They warned that, without immediate and drastic reduction in humanity's CO_2 emissions, exceeding this warming limit before mid-century could become unavoidable[20,22]. The impacts of human-induced climate change are already far-reaching, with extreme weather wracking much of the globe in recent years; 2020 tied with 2016 as the warmest year on record[36,37]. 2018 saw an above-average number of tropical cyclones in all four Northern Hemisphere basins; significant drought in Eastern Australia; heatwaves affecting large areas of Europe, especially Germany, the Czech Republic, Poland, the Netherlands, Belgium, and France; and devastating wildfires in Athens, British Columbia, and California[38]. Economic losses from extreme weather events totalled $155 billion USD in 2018 and $350 billion USD in 2017.[39] In 2018, these climate-change-induced weather disasters claimed 5000 lives and saw 28.9 million people in need of humanitarian aid or emergency assistance.[40]

Beyond the human cost of climate change is extensive and lasting ecological harm. The Earth's oceans have absorbed much of the increased heat and CO_2, warming and acidifying and consequently damaging marine ecosystems[41,42]. Glacial retreat and melting sea ice are causing rising sea levels, threatening coastal habitats and destroying large parts of the cryosphere[43–45]. The human threat to biodiversity is multifaceted and includes pollution, overexploitation of natural resources, direct habitat loss due to human expansion and more, as well as the effects of climate change on the biosphere. A series of reports released in 2018 by the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) highlight the dangerous impact of human action on the health of our environment, suggesting that climate change is emerging as a significant driver of species loss[46–51]. The evidence for anthropogenic climate change is unequivocal – 99.9999% certainty that human CO_2 emissions are driving climate change[52] – yet global demand for fossil energy is growing and outpacing decarbonisation efforts[17]. Humanity's reliance on versatile yet finite and polluting fossil fuels is best summarised by Smil: "Modern civilisation has been created by the massive and increasing combustion of fossil fuels, but this practice is clearly limited by their

crustal abundance as well as the environmental consequences of burning coals and hydrocarbons. High-energy societies can ensure their survival only by an eventual transition to non-fossil sources."[25]

The global energy landscape

Smil is sceptical about the time it will take to complete this energy transition, noting both that the transition from wood to coal took over a century and that use of fossil fuels has actually been increasing in recent years[53]. A review of the current global energy landscape quickly reveals the vast inertia sustaining modern civilisation's dependence on fossil fuels. *F*igure *1*, generated from data sourced from the International Energy Agency (IEA), shows that fossil fuels currently supply over 80% of the world's total primary energy demand and highlights the change in global energy usage since 1971[54]. *Figure 2*, sourced directly from the World Energy Outlook 2021 report by the IEA, predicts that, under existing policy scenarios (called STEPS, the Stated Policies Scenario), fossil fuels will still make up more than half of the energy mix in 2040[55]. Assuming a delivery of announced pledges (the Announced Pledges Scenario, APS), the predicted global energy mix in 2040 more heavily favours renewables, though fossil fuels still dominate and carbon emissions are declining but non-zero. Aggressive policymaking towards internationally agreed objectives to limit climate change by becoming net-zero emitters of greenhouse gas (the net-zero emissions scenario, NZE) sees the fossil share of primary energy drop drastically by 2040, with net-zero carbon emissions being achieved by 2050. According to the 2021 IPCC report on climate change and the Emissions Gap Report 2018 released by the United Nations Environment Programme (UNEP), the world is not on track to meet emissions targets that would keep global warming below the critical threshold of 1.5°C[22,56].

Fossil fuel combustion emits far more CO_2 than any other primary energy source[55]. To curb these emissions, renewable energy transitions are underway across many sectors, but as of yet these efforts have not made major impact. The SARS-CoV-2 (COVID-19) pandemic, which began in 2020, was hugely disruptive for all sectors, especially transport, which saw a 10% decrease in CO2 emissions that year[57]. Global CO_2 emissions – which had been steadily climbing since 2016 – plunged in 2020, but have rebounded in 2021 by more than 4% with increasing demand for coal, oil, and gas[58]. This is the largest annual increase since the recovery from the global financial crisis in 2008, and brings global carbon emissions back to pre-COVID levels at 37.4 gigatonnes of CO_2 ($GtCO2$)[59]. Power generation and transport account for over two thirds of

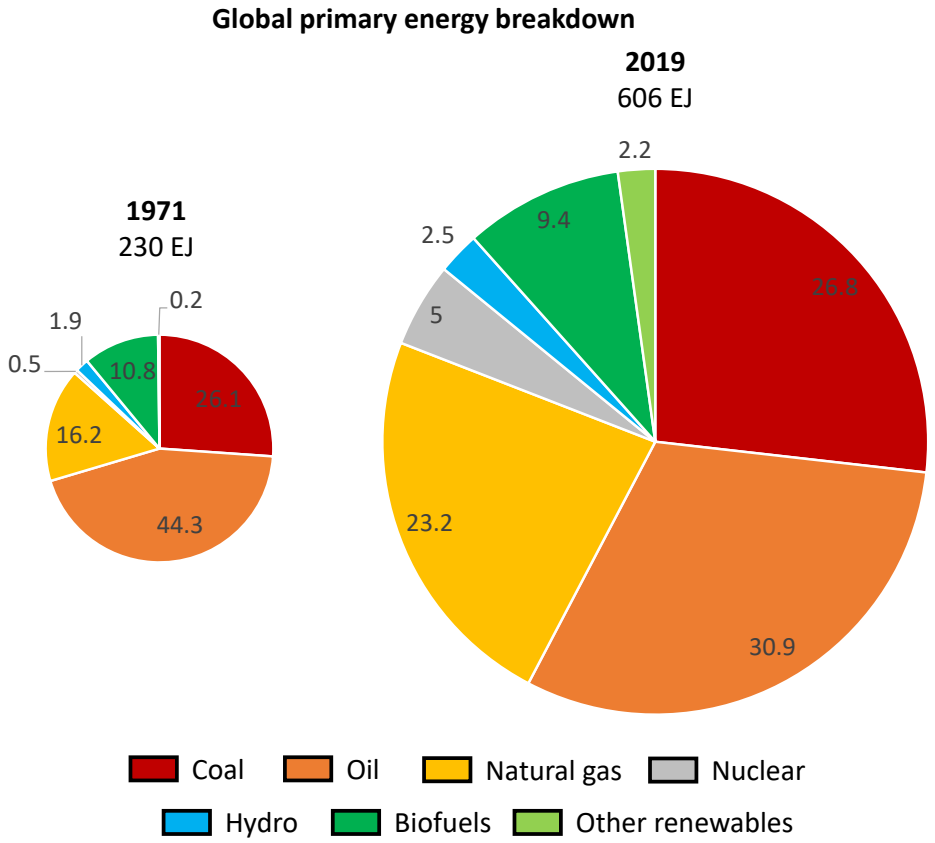


Figure 1. Breakdown of global primary energy demand (exajoules, EJ) in 1971 and 2019, as a percentage of total, by sector. Figure composed by the author using public data provided by the International Energy Agency (IEA) in 2021, available online[54], according to their terms and conditions. See permissions following the References section of this thesis.

emissions, as shown in *Figure 3*, and have been responsible for almost all global growth since 2010[58].

Alongside renewable alternatives to displace fossil fuels, carbon capture and storage (CCS) technology will be essential to reversing the impact of these emissions; emissions pathways that succeed in keeping global warming under 1.5°C demand net-zero global CO_2 emissions by 2050 and rely on CCS to remove 100-1000 $GtCO_2$ from the atmosphere over the 21st century[20].

A by-sector breakdown of CO_2 emissions is shown in *Figure 3*. Electricity and heat generation, by far the largest contributor, is primarily driven by coal and natural gas. Use of coal – the most carbon-intensive fossil fuel[60] – has increased overall since 2015 but is projected to grow only modestly by 2040, being displaced in the energy mix by natural gas or, to a lesser extent, renewables[18]. For electricity and heat generation, natural gas is preferable to coal because of its high power density, with the fuel-to-energy conversion efficiency of combined-cycle gas turbines exceeding 60%[61]. In contrast, solar and wind energy have much lower power densities[62] – requiring more

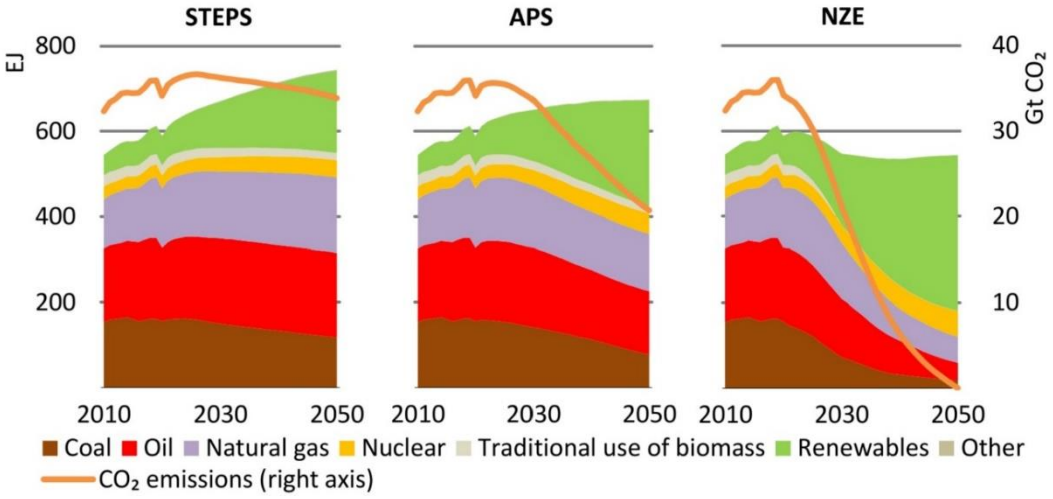


Figure 2. Total global primary energy demand (exajoules, EJ) and greenhouse gas emissions (gigatonnes of carbon dioxide equivalent, Gt CO_2) over time, from 2010. Predictions out to 2050 are based on three scenarios, as proposed by the International Energy Agency (IEA): (i) The states policies scenario (STEPS), which reflects current and announced energy policies from governments around the world; the announced pledges scenarios (APS), which assumes existing commitments on climate change limitation, air quality, and energy access made by governments will be met; the net zero emissions scenario (NZE), which aims for net zero carbon emissions globally by 2050. Figure republished directly from the IEA World Energy Outlook 2021[55], see permissions following the References section of this thesis.

land area per unit of electricity generated – but are nevertheless also displacing coal because of their much lower embodied carbon emissions and, importantly, their lower costs. On-shore wind and utility-scale solar are now significantly cheaper than any other form of energy if subsidiaries are discounted[63]. Despite this milestone in the renewables revolution, it is important to remember that solar, wind and hydro installations are currently built using heavy machinery powered by diesel fuel, carbon-intensive concrete is required for dams and turbine foundations, the blades of wind turbines are composites of fossil-derived plastic, and the steel towers that support them are forged using coal.

Transport has the highest reliance on fossil fuels of any sector. It is the second largest but fastest growing contributor to global CO_2 emissions and is characterised mainly by road transport, including civilian vehicles and road freight vehicles, which represents 75% of emissions from this sector. Road transport is experiencing growing electrification: electric vehicles (EVs) are becoming a popular option for light-duty civilian use but market penetration of EVs has been minor overall[64]. Under the most aggressive EV deployment policies, by 2030 there will be 220 million EVs on roads around the world: 130 million battery EVs and 90 million hybrids[64], 11% of a predicted global total of 2 billion vehicles[33]. Less ambitious policymaking in the next few years could see this figure drop to less than half of the best-case scenario. Either way, all remaining internal combustion engine (ICE) vehicles will still rely on fossil fuels, but

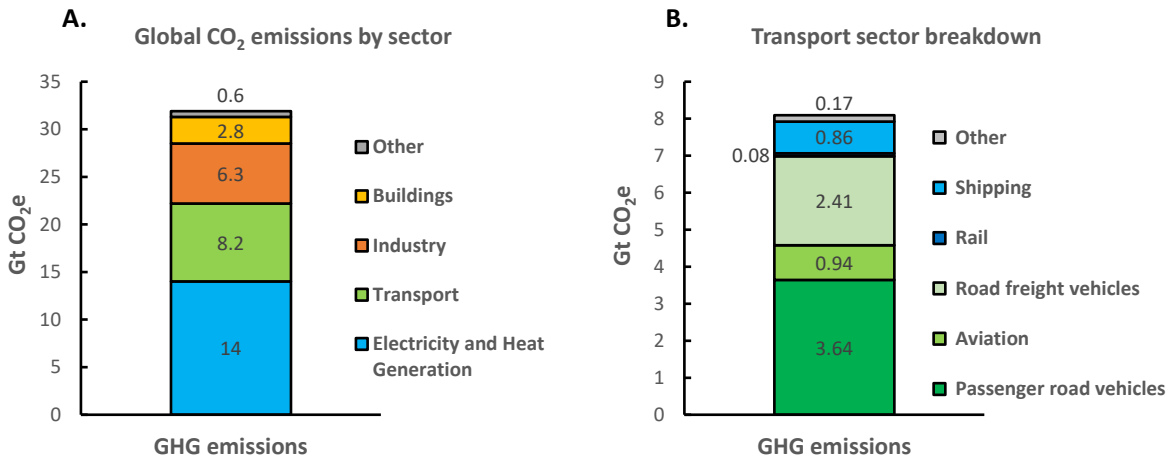


Figure 3. (A.) Share of anthropogenic greenhouse gas (GHG) emissions by sector (gigatonnes of CO_2 equivalent, Gt CO_2e). (B.) Breakdown of GHG emissions (Gt CO_2e) from the transport sector, share of emissions across various transport modes. Figure generated by the author using data provided by the IEA, accessible online[57], according to their terms and conditions. See permissions following the References section of this thesis.

improvements in fuel efficiency in new ICE vehicles reduces oil demand by 9 million barrels per day in 2040, three times as much as is displaced by the entire EV fleet at this time[18]. Producing these EVs will be carbon-intensive: the manufacture of one EV and its batteries contributes 50% more to global warming potential than production of a comparable ICE vehicle[65,66]. It is the carbon embodied in the fuel source that distinguishes ICE vehicles and EVs; if both electricity and liquid fuel were carbon-neutral, ICE vehicles would be the more environmentally friendly option.

Light road transport vehicles are nevertheless well placed to transition to electrical power due to their lower power requirements. Aeroplanes, ships, and trucks, on the other hand, require fuels with a higher energy-to-weight ratio than current batteries can provide, making these sectors difficult to decarbonise. Fossil fuel is so attractive to these sectors because liquid hydrocarbons are energy dense, so less of a vehicle's cargo volume needs to be devoted to carrying fuel. For comparison outside the transport sector, static installations providing baseload electricity, such as coal-fired power plants, are – when ignoring geographical factors – better suited for replacement by renewable alternatives, as they often have land area available to adopt the high-volume energy storage solutions accompanying most renewable energy technologies; that is, batteries[67]. While greener, these alternatives are still less effective and versatile than existing advanced fossil-based technologies, such as a combined-cycle gas turbine[68]. Energy storage is more crucial in the transport sector; planes, ships, and trucks would become far less effective modes of transport if a large proportion of their carrying capacity was devoted to their fuel, hence why these applications in particular are difficult to decarbonise. The issue of energy density is discussed in more detail later in this chapter.

Aviation and shipping have been growing even more rapidly than road transport, aviation emissions in particular having increased by 105% since 1990 and forecast to rise a further 300-700% by 2050[23,24]. These industries lack an energy dense battery based alternative to liquid hydrocarbon fuels[69,70], and it is here where sustainable, carbon-neutral substitutes like biofuel are most needed.

The net-zero emissions (NZE) scenario proposed by the IEA requires transport sector emissions to fall by 20% in the next eight years. Achievable, it is hoped, by encouraging the use of less carbon-intensive transport options, such as rail travel[55,57], and by broad adoption of electric road vehicles.

Purely electrical solutions to humanity's power problems rely on widespread electrification of infrastructure and typically require batteries for energy storage. The use of batteries for high-energy applications poses three interrelated challenges: resource availability and land use[29], battery cost[27], and energy storage efficiency[26]. These issues, and similar ones pertaining to biofuels, will be addressed sequentially in the following sections.

Metals scarcity and land use

Batteries achieve carbon-neutrality with metals-dependency, trading a reliance on fossil energy sources for a reliance on rare earths and other metals, specifically cobalt and lithium[71–73]. While rare earths – contrary to their name – and the most vital element for battery manufacture, lithium, are not particularly scarce, cobalt demand could exceed supply by 40 000 tonnes per year by 2040 if EV adoption rates are high[29]. Mining and processing these metals is carbon-intensive and does direct ecological damage, so increasing supply to account for growing battery demand may damage already fragile ecosystems with roles in climate maintenance[74–79]. Consider that processing a single tonne of rare earths produces approximately 12 000 cubic meters of waste gas – containing dust, hydrofluoric acid, sulfur dioxide, and sulfuric acid – 75 tonnes of toxic wastewater and one tonne of radioactive waste[80]. The severe polluting effects of these effluents on local human and animal populations, especially in areas with limited regulation of mining operations, is well-documented[74–76,78,79].

Mining operations also have a large geographical footprint. 90% of the world's supply of rare earths are produced in China and most of this comes from Bayan Obo, in Inner Mongolia, the largest deposit of rare earth metals discovered to date[77]. The main open pit mine at Bayan Obo spans fifty square kilometres and is a kilometre deep, and it feeds an 11 km^2 tailings dam. Even if this entire operation were powered by carbon-

free electricity, Bayan Obo would still have a geographical footprint of 100 square kilometres. Extracting resources from the lithosphere is a land-hungry endeavour and significantly impacts local environments by destroying habitats and generating toxic waste products. While mining rare earths is not an energy generation operation, this land requirement drawback is common to wind turbine installations, solar photovoltaic arrays, and hydropower dams[61], which also share a major point of inflexibility with mining: their location is dictated by geography. Mining operations are established where ore can be found in quantity and extracted under economically viable conditions. Similarly, renewable power generation requires certain geographical features: solar arrays require a high incidence of sunlight, wind turbines need regular winds, and hydropower needs water reservoirs on different elevations.

Biofuel crops

Biofuel crops serve to concentrate atmospheric CO_2 into a form that can be accessed and converted by microbes, but also create issues with land use. Photosynthetic capture of solar energy as phytomass is inefficient[81], requiring large tracts of land to produce sufficient biomass for refining, as is subsequent processing into useful forms of transport fuel. First generation biofuels are produced using sugar-rich biomass from food harvests, directly impacting food supply, or are grown on existing agricultural land. The US produces almost one million barrels per day of bioethanol, most of which is derived from corn[82]. Even if established cropland is repurposed to grow dedicated energy crops, there is an indirect land-use effect when food crops that would otherwise have been grown on this land are removed from the market, increasing the price of those foods and stimulating land conversion elsewhere. Agricultural economics can prompt a seemingly beneficial effort to grow biofuel crops but result in net positive GHG emissions. Some of these issues can be mitigated by leveraging a broad variety of crops to establish operations in diverse environments; second-generation biofuels utilise agricultural waste streams, lignocellulosic crops, or crops such as *Jatropha* or *Miscanthus* that grow on poor or degraded soils unsuitable for typical agriculture[83,84]. The land requirement for even second-generation biofuel crops is extensive[50,85], but adverse environmental impacts typically associated with large scale agriculture can be avoided by careful selection of energy crops that utilise sparse, arid areas low in biodiversity and that are otherwise undesirable for agriculture or habitation.

Despite the limitations of first-generation biofuels, some countries have managed to embed products such as bioethanol into their energy landscapes at scale. Brazil has been producing ethanol for blending with gasoline since the 1930s. In response to the 1973 oil crisis, Brazil further established a sugarcane-based ethanol fuel programme to

help alleviate reliance on imported fossil material[86]. The Brazilian government mandated a minimum blend of 10% ethanol in automotive gasoline, helped finance the modification of gasoline engines to support added ethanol, and incentivised agro-industrial ethanol production by fixing the price of both gasoline and bioethanol[87]. Brazil remains a world leader in sustainable bioethanol use, despite a significant drop in interest toward ethanol-fuelled cars in the 1990s, which was driven by the cheap price of oil, changes to bioethanol subsidies, and increased demand for sugar or arable land[88]. Today, biofuels nevertheless represent approximately 25% of transport fuels in Brazil[89], demonstrating that committed policies to stimulate the biofuels market can have lasting impacts. In the case of bioethanol, technical developments were required to support its widespread adoption, including the development of flex-fuel engines that can burn a range of ethanol-gasoline blends[90]. Such developments are necessary for any alternative fuels with significantly different chemical characteristics to petroleum fuels.

Advanced biofuel crops target desert and marine environments. *Agave* and *Opuntia*, while potentially less effective for traditional bioethanol production than corn or soybean but more effective for gas production, are highly-efficient energy crops that can grow in semi-arid regions and that require less pre-treatment before bioprocessing than lignocellulosic crops[91,92]. Similarly, both micro- and macro-algae are effective carbon sinks, representing a source of biomass and of useful oils that can utilise marine spaces for growth. Algacultures also require less surface area than leading biofuel crops, such as corn, to produce the same volume of fuel, all while filtering water and recycling other agricultural runoff[93,94].

Cost of energy storage

The second major issue with existing battery technology is the cost of energy storage per kilowatt hour (kWh). All renewable power generators need battery installations to effectively respond to electricity demand: storing excess energy when conditions are favourable – when the sun is shining, or the winds are high – and releasing energy they are not. Batteries facilitate the integration of renewable energy sources like wind and solar into the existing electricity grid, allowing them to continuously supply electricity and vary output throughout the day, much like a typical coal or natural gas combined-cycle power plant. South Australia's 100MW battery installation is a good example of battery storage technology improving utilisation of renewable energy assets[95], but lithium-ion batteries struggle to compete at scale, offering diminishing returns as more storage capacity is required[96].

Fossil power plants run continuously and provide far more power per unit land area than renewable plants because their energy storage medium – coal or natural gas – is so cheap and so energy dense. Their flexibility comes from the ability to tap this source on demand, independent of weather conditions. A feasibility study carried out in by the Clean Air Task Force on energy storage options for California – already committed to sourcing 50% of its power renewably by 2030 and aiming to fully decarbonise its energy sector by 2045 – found that, at large scales, renewable energy systems become dominated by the cost of storage[27,97]. Wholly renewable energy systems based on battery storage and flexible grids have both detractors[98] and proponents[99], but the abilities of such systems can be assessed by studying real-world examples that attempt to cater for different scales of demand. Germany's sweeping, transformative 'Energiewende' initiative equipped the country with an additional 90 gigawatts (GW) of renewable power capacity, equal to its entire existing energy requirements, but has reduced the country's reliance on fossil fuels by only 4%[53,56,100,101]. Australia's 129 megawatt-hour (MWh) battery, on the other hand, is competitive for its size at ~$63.5 million USD and set to recoup almost a third of its costs in a single year of operation[95]. To meet their ambitious renewable energy targets, Californians will need 36.3 million megawatt hours of energy storage, with an estimated cost of $360 billion USD[97]. This figure assumes the cost of Li-ion batteries will drop by two-thirds by 2045.

Energy density

The energy density of a Li-ion battery is lower than that of traditional biofuels, including dry wood, which has a similar energy density to sub-bituminous coal[25,102]. This is counterintuitive because cutting-edge photovoltaics are straining at the limits of their theoretical solar-to-electricity conversion efficiency – 29.1% for silicon solar cells – whereas the solar-to-chemical conversion efficiency of natural photosynthesis is only 3-6%[25,103]. It is the energy density of the storage medium that separates these artificial and natural technologies.

Storing solar energy in a growing tree, which also sequesters carbon, is slow and accessing this energy reservoir involves burning biomass and harnessing the resultant heat, entailing far more conversion losses than when drawing high-grade electricity straight from a battery. A carbon-based energy platform, however, represents a versatile system with dense, stable storage capacity. Compared to lithium, cobalt, and rare earths, carbon is abundant – the fourth-most abundant element in the universe – and available in vast, albeit dilute, quantities from the atmosphere. A robust system for tapping carbon flows is the basis of a circular carbon economy.

The advent of accessible electrical power revolutionised every sector it touched, but in recent years electricity has experienced its "most dramatic transformation since its creation more than a century ago"[18]. Electricity is a high-grade energy and electrical powertrains are often more efficient than their mechanical counterparts; it is the energy of choice for personal electronics and light industrial sectors. It is the capacity to store electricity in small volumes at low weight that presents the biggest problem opposing its total dominance in the energy landscape. There is probably a Li-ion battery in your pocket right now, but energy-intensive transport applications like aviation, shipping, and trucking require energy dense liquid fuels.

To get an idea of the scale of the problem, imagine that you – as a traveller – must physically carry in some form all the energy needed to fly you from one place to another whenever you board a plane. Lufthansa Group stated, in a 2017 sustainability report, that their fleet of aircraft had achieved landmark fuel efficiency, requiring on average only 3.68 litres of aircraft kerosene to fly one passenger 100 kilometres[104]. Carrying this 3.68 litres of fuel would be trivial, like carrying two large milk cartons totalling just over 3 kilograms. Alternatively, you could carry 66 litres of lithium-ion batteries, enough to fill a large duffel bag and weighing 137 kilograms. As of early 2022, the cartons of kerosene will cost you almost exactly £2 GBP ($2.66 USD) and the batteries will cost £5805 GBP ($7600 USD)[33,102,105]. The economic comparison is not entirely relative as batteries are an energy carrier and can be recharged whereas the kerosene is destroyed during combustion, but even with cycles lives in excess of several thousand charging cycles, the batteries are still not as economical for this purpose as the kerosene. Devoting 22 times the volume and over 40 times the weight of an aircraft's carrying capacity to fuel – even accounting for the increased efficiency of an electric powertrain, batteries still require approximately 16 times the volume jet fuel does to deliver an equivalent amount of energy – severely reduces its operational effectiveness and makes air transport economically unfeasible. While an economic window for battery-powered aircraft does seem to exist, drastic improvements in battery technology will be required before electrification of the aviation sector becomes a viable option[106].

While the energy density of Li-ion batteries has been increasing by 5-8% per year[102], battery technology faces a fundamental physical challenge to match the energy density of liquid chemical fuels. Some part of a battery will always be devoted to electrolyte spatially separating the anode and cathode, which does not contribute to the embodied energy of a battery. In liquid chemical fuels every unit of volume, all the way down to a molecular scale, constitutes harvestable energy released on combustion.

The ideal sustainable fuel

Considering the growing popularity of electricity in light industrial and transport sectors, projected metals availability, the limitations of battery-based energy storage, and how deeply entrenched fossil fuels are in modern society, a new sustainable fuel should have the following characteristics:

a. *Renewable:* It should be carbon based, ideally sequestering some carbon as part of its production cycle.

b. *Drop-in compatible:* It should complement the fuel chemistry of existing blendstocks and leverage existing and widespread fossil fuel transport, storage, and engine infrastructure for rapid deployment and adoption.

c. *Versatile:* It should be energy dense and stable. For transport, it should ideally mimic the characteristics of fossil fuel as closely as possible.

d. *Robust:* It should be possible to derive the fuel from a wide variety of feedstocks, such as direct capture of atmospheric carbon, bioelectrosynthesis of precursor chemicals, or simple biomass. Precursor crops for carbon concentration should be cultivatable on non-arable land low in biodiversity, unsuitable for other agriculture, and undesirable for human habitation.

e. *Targeted:* To achieve maximum impact, the fuel should target energy-intensive transport sectors with no viable non-fossil alternatives: aviation or maritime transport.

Advanced biofuels that are chemically comparable to their petroleum counterparts, yet renewably produced using atmospheric CO_2, represent a drop-in solution that can meet all these criteria. Early generation biofuels were expensive to manufacture and competed with food crops for arable land[85,107,108]. The choice of fuels was also mostly limited to biodiesel or bioethanol, which could be blended with motor gasoline or jet fuel but could not replace these products. It has not been possible so far to biologically manufacture, economically and at scale, a liquid fuel that can act as a direct substitute for fossil fuels. Gasoline, diesel, and jet fuel are characterised by performance specifications, not by chemical composition, so a replacement fuel will be comprised mainly of straight chain alka(e)nes, with some branched and cyclic alka(e)nes and some aromatics. Microbial biosynthesis pathways for all these components have been investigated, but as of yet yields still fall well short of the theoretical maximum.

Part II: Molecular machines and directed evolution

Biology is the world's most sophisticated manufacturing platform. Every cell is like a minute factory outfitted with the wonderfully refined molecular machinery of life; self-repairing, adaptive, and discerning. Harnessing biological processes for industrial use is an established practice, from brewing to biologics, and increasing control over cellular activities has broadened the range of useful products that can be made biologically, a range which already spans diverse sectors like healthcare, food, agriculture, chemicals, materials, and energy.

There are two approaches to engineering and design within biological systems: bottom-up design and top-down design. Bottom-up design involves assembling well-defined and characterised parts to create systems with predictable behaviour and has been the successful foundation of traditional engineering disciplines. Applying this philosophy to biological systems is a goal of synthetic biology: to create a variety of standardised genetic parts that have well-defined and reliable functions. Efforts thus far have resulted in the Registry of Standard Biological Parts[109], the BioBricks assembly standard[110], Synthetic Biology Open Language (SBOL)[111], and more, but revolutionary advancements would need to be made to truly decouple a standard biological part from its native genetic and cellular context and conform to key performance requirements: independence, reliability, tunability, orthogonality, and composability[112,113].

Bottom-up engineering and design is limited in biological contexts by the sheer physiological complexity of most living systems. Artificial organisms created so far, such as the JCVI-syn series of *Mycoplasma* with their synthetic genomes, have been rudimentary, but are nonetheless impressive examples of genetic engineering[114]. A fully characterised artificial microbe compatible with all standardised biological parts would revolutionise biotechnology, so bottom-up approaches to creating one are to be applauded, but top-down approaches to bioengineering are far more common today.

Top-down engineering builds upon and alters established systems to achieve a design goal. In synthetic biology, this involves manipulating living organisms – called chassis – that have already evolved and refined their fundamental vital processes. Working within the native biological context of a natural or laboratory adapted chassis organism means synthetic biologists can rely on robust underlying physiology, but changing this physiology is limited by the need for prior knowledge and the complex, often unpredictable impact of even small alterations. These limits can be sidestepped by leveraging a massively parallel black-box approach to the design-build-test cycle of traditional engineering: Darwinian evolution.

Microbial biosynthesis of fuels

The inventor of the diesel engine, Rudolph Diesel, envisioned that it could be run on vegetable oils to power agriculture in remote parts of the world, where refined petroleum fuels were not readily available. Cheap, chemically superior fossil derivatives limited the growth of liquid biodiesel until the late 1980s and early 1990s, when the transesterification process – proposed earlier in 1937 by Georges Chavanne – began to see widespread use in production of biodiesel from vegetable oils[25]. Despite steady advances in the quality of fuel produced, availability of alternative feedstocks, rising energy prices, and increased global awareness of climate change, biodiesel – and all liquid biofuels in general – have failed to significantly displace their fossil counterparts in all sectors. Even ethanol, derived predominantly from corn and now the most successful and widely commercialised liquid biofuel[82], sees limited use due to a lack of infrastructure and low market penetration of engines capable of handling high-ethanol fuel blends[115]. Biodiesel is more like fossil diesel in terms of energy density than ethanol is like gasoline, but it nevertheless struggles to compete against the cheaper and more entrenched alternative.

Ethanol is compatible with typical automotive engines only when blended in small amounts with regular gasoline. Ethanol, like other alcohols, is hygroscopic and oxygen-containing, so ethanol-blended fuels requires that engine components be produced from corrosion resistant materials. The different combustion characteristics of ethanol compared to gasoline also necessitates changes to spark plugs and fuel injection systems, fuel compression ratios, and the addition of auxiliary cold-start assistance technologies. Consequently, specially designed or adapted engines, often called 'flex fuel' engines, are required to handle ethanol blends exceeding 10% (v/v)[116]. A truly 'drop-in compatible' biofuel will have chemical characteristics identical or complementary to those of existing fuel blendstocks, such that it can be effectively blended with petroleum products or replace them entirely. Chemically, this means the constituent molecules of this drop-in compatible biofuel should resemble those found in the relevant petroleum counterpart; primarily paraffins (*n*-alkanes, generally with carbon length C4-C12 for gasoline, C9-C25 for diesel, and C5-C15 for jet fuel)[117], olefins (*n*-alkenes), isoparaffins (branched alkanes), naphthenes (cycloalkanes), and aromatics (mono- or polycyclic unsaturated hydrocarbons)[118]. *n*-alkanes are a major component of transport fuels and biologically-derived alkanes can directly displace fossil-derived alkanes without impacting fuel quality or requiring changes to engine infrastructure, hence why biosynthetic pathways for producing *n*-alkanes are attractive candidates for optimisation and upscaling.

As discussed in Part I of this chapter, liquid biofuels, especially those with fuel chemistries comparable to those of existing petroleum products, offer one of the only routes to decarbonisation of sectors for which electrification is not feasible; aviation and marine transport. Increasing global demand for air travel places particular urgency on the search for scalable drop-in alternatives to jet fuel, which can currently be produced most effectively by direct blending with, or chemical upgrading of, microbially-synthesised isoprenoid precursors[119]. Generally though, paraffinic hydrocarbons (*n*-alkanes) of medium to long chain length represent a flexible molecule for subsequent chemical upgrading or direct use in petroleum blends, especially diesel, and these molecules are commonly produced in microbial fermentations[120]. The cyanobacterial alkane biosynthesis pathways described in the following section are among the most well-characterised in microbial biofuel production, despite their relatively recent discovery[121].

Hydrocarbon-producing metabolic pathways

Microbial synthesis of alkanes, alkenes, alcohols and other oleochemicals has been researched since the 1960s and understandably well-reviewed in recent years[115,120,122,123]. While alcohols like ethanol and butanol (isobutanol and *n*-butanol) have been successfully produced in commercial bioprocesses for use as biofuel[124], *n*-alkanes and alkenes boast a higher energy density and better compatibility with existing liquid fuel infrastructure and consequently are pursued – albeit with less success – as a better alternative to petroleum fuels. Several pathways for alka(e)ne biosynthesis exist in nature and can be found in cyanobacteria, algae, yeasts, and fungi[125]; these pathways have been well reviewed by Arai *et al.* and are illustrated in *Figure 4*, republished directly from that work[126]. A selection of pathways is also summarised in *Table 1*.

Artificial alka(e)ne biosynthesis pathways have also been constructed in *E. coli*, such as a pathway converting free fatty acids to fatty aldehydes via a fatty acid reductase (FAR) complex encoded by the cyanobacterial *luxCDE* genes, which are then converted to alkanes via aldehyde deformylating oxygenase (ADO)[127,128]. More recently, the cyanobacterial acyl-ACP reductase (AAR)-ADO pathway expressed in *E. coli* has been used to achieve titres of 1.31 g/L of alkane in a fed-batch reactor[129], and a combined *in silico* modelling approach reports titres of 2.54 g/L under similar conditions[130]. As of 2022, this latter work by Fell *et al.* remains the highest reported titre of alkanes. Both approaches used glucose or glycerol as the sole carbon source and output mostly pentadecane (40-45% of total alka(e)ne) and heptadecane (50-55% of total alka(e)ne), though it has been shown that the predominant alkane produced can be predictably changed by manipulating the precursor fatty-acid pool[128]. It should be noted that high

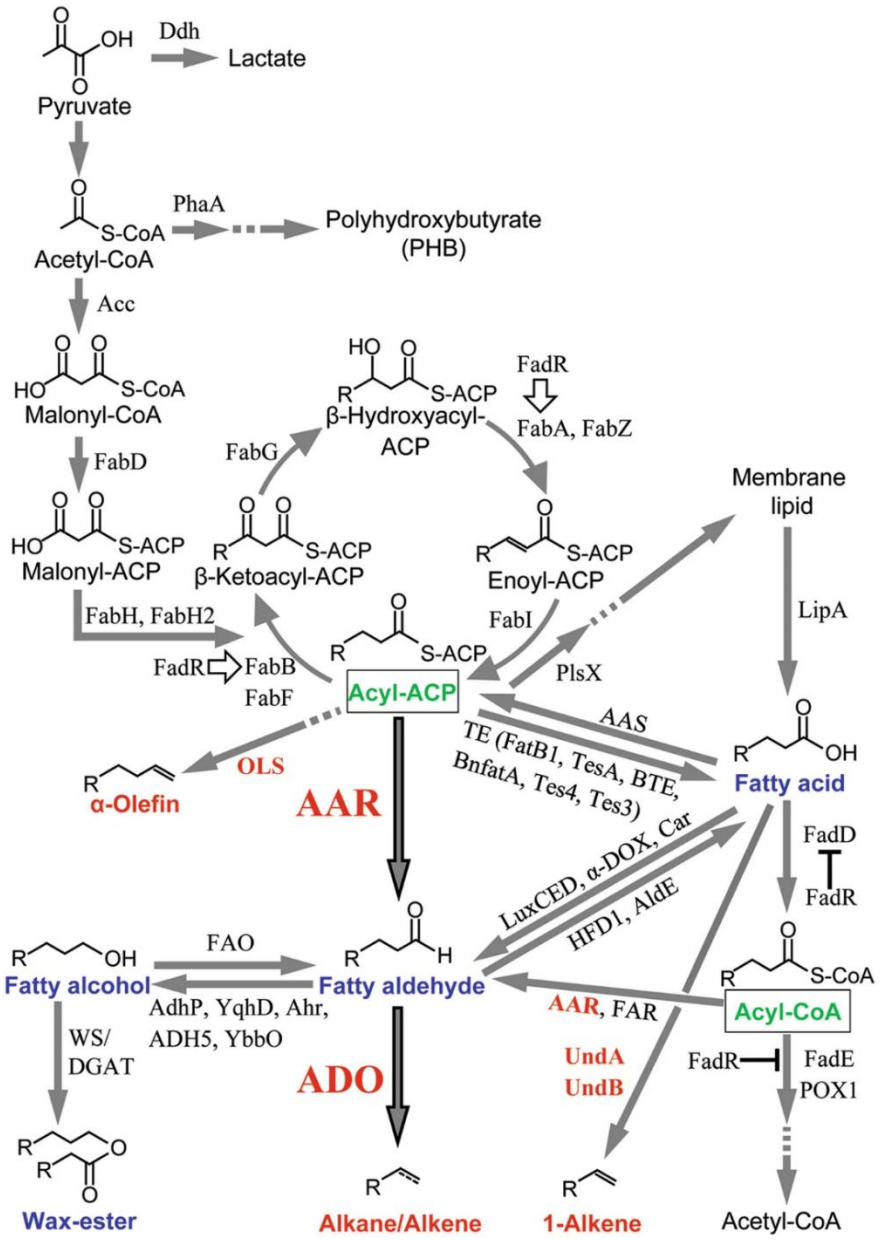


Figure 4. Prokaryotic alkane biosynthesis pathways. Republished directly from Arai *et al.*[125], see permissions following the References section of this thesis. Both the *luxCDE-ADO* pathway, converting fatty acid to fatty aldehyde and then alkane, and the AAR-ADO pathway, producing fatty aldehydes directly from acyl-ACP, are used in this project.

titres, while desirable, are not the sole consideration when optimising industrial bioprocesses; production rates and overall yields must be examined in concert to evaluate the viability of a process. These three measures are often grouped together and referred to as TRY (titres, rates, and yields).

Alkanes synthesised biologically typically have carbon numbers in the diesel fuel range (C9-25), but chain length has been adjusted using a mutant fatty acyl-ACP thioesterase to create short fatty acids which are converted to short-chain alkanes, C9-14, for gasoline (C4-12) or kerosene-type jet fuel (C5-15)[131].

Table 1. Microbial linear alka(e)ne biosynthesis pathways.

Pathway; key enzymes	**Intermediate metabolites → Product**	**Natural host**
Cyanobacterial alkane biosynthesis; AAR-ADO	acyl-ACP → fatty aldehyde → **alka(e)ne** (C13-17)	*Synecchococcus* sp. [121]
Head-to-head condensation; OleABCD protein family	acyl-CoA + acyl-CoA → **long-chain alkene** (C27-29)	*Micrococcus luteus* [132]
Terminal olefin synthesis; $OLeT_{JE}$	fatty acid → **1-alkene** (C15-17)	*Jeogalicoccus* sp. [133]
Artificial alkane biosynthesis; CAR/ADO-FAR	fatty acid → fatty aldehyde → **alka(e)ne** (C11-C17)	None; engineered host is *E. coli* [128,134]

Enzymes: AAR, acyl-ACP reductase; ADO, aldehyde deformylating oxygenase; $OLeT_{JE}$, alkane-producing cytochrome P450; CAR, carboxylic acid reductase; FAR, fatty acid reductase.

Petroleum fuels do not usually have a strictly prescribed chemical makeup. Instead, they are defined by performance characteristics – density, flash point, freezing point, autoignition temperature, specific energy and sometimes lubricity, hygroscopicity, conductivity, etc. – that can manifest from a potentially diverse combination of blendstocks. To match these performance specifications, biofuel substitutes will likely be comprised predominantly of straight-chain alkanes, with smaller fractions of alkenes and branched, cyclic, and aromatic compounds.

Unlike the fatty-acid-derived linear alka(e)nes produced via pathways in *Table 1*, branched or aromatic compounds are biologically derived from isoprenoids. Alternatively, Howard *et al.* have engineered *E. coli* to express β-ketoacyl-ACP synthase III (KASIII) and branched-chain α-keto acid dehydrogenase (BCKD) enzymes from *Bacillus subtilis*, enabling the host chassis to produce an internal pool of branched fatty acids in the C14-C16 range, which can be subsequently converted in the same host via the cyanobacterial *luxCDE* pathway to branched alkanes[128]. This work demonstrated that some control over product chain length (C_n) is achievable, and that manipulation of the fatty acid precursor pool in cells enables basic branched molecules to be produced via the cyanobacterial AAR-ADO pathway.

Isoprene (C5) itself represents a short-chain gasoline blendstock and monoterpenes (C10) are good candidates for use in jet fuel blends[122,135]. Farnesene (C15) has reportedly been produced at incredible titres of 130 g/L in yeast, grown on cane syrup over two weeks in a fed-batch configuration, by Amyris, the company responsible for engineering yeast to produce an artemisinin precursor for antimalarial drug production[136]. Amyris developed this production technology as part of their MegaBio project; originally slated for completion in 2019, the project aimed to produce (E)-β-farnesene from cellulosic sugars at $2 USD per litre but failed to reach project

milestones[137]. Remarkably, that work has received no further attention since its publication in 2016 and the conclusion of the project in 2019. Despite this extensive work in this area, industrial biosynthesis of alkanes is still limited[138], but they remain the major component of petroleum fuels and the key to creating true drop-in alternative biofuels.

Aldehyde deformylating oxygenase (ADO; formerly fatty aldehyde decarbonylase, FAD) is considered to be the bottleneck for efficient alkane biosynthesis in microorganisms[139]. Alkanes are highly reduced compounds, but enzymatic action of ADO requires oxygen, meaning alkane production must occur in a thermodynamically unfavourable environment. Alkanes are also a rare example of a biological product bearing no functional groups, making them difficult for a cell to produce. ADO removes formate from C_n fatty aldehydes to produce C_{n-1} alkanes, a difficult and unusual reaction to perform, hence the low catalytic efficiency of ADO[140]. While the catalytic mechanism of ADO has been studied and conditions for its activity optimised, turnover rate of ADO has proved difficult to improve[141,142]. ADO nevertheless remains the frontrunner in efforts to biologically synthesise alkanes, though enzymes like the cytochrome-P450 peroxygenase, $OLeT_{JE}$, have attracted attention for their ability to produce alkenes, which also have biofuel applications, directly from fatty acids, bypassing the creation of toxic aldehyde intermediates[143].

Prokaryotic chassis organisms

Most directed evolution processes, in fact most bacterial bioengineering and bioprocessing methods, are built around the workhorse of molecular biology: *Escherichia coli*, a gram-negative bacterium native to the mammalian gut. While foundational to the creation of recombinant DNA and consequently the field of modern biotechnology[144], *E. coli* is not always ideal as a chassis for genetic engineering, or as a platform for industrial use, because of its metabolic features and the saturated intellectual property space around it[145]. In synthetic biology, a chassis organism provides the biological machinery to support exogenous genetic elements, whether natural or artificial – the 'hardware' for engineered DNA 'software'. Good microbial chassis vary depending on application but, in general, should be fast-growing, resilient to harsh conditions, metabolically diverse, and genetically unconvoluted, providing a reliable, tractable framework in which to execute biological functions. A perfect chassis would display broad orthogonality to all standard biological parts, predictably executing any non-native cellular processes.

E. coli is the most widely studied prokaryote and has been the host organism for most bacterial genetic engineering over the past 60 years[146]. It is a legacy organism favoured for such purposes because it is well characterised – the functions of most genes are known – and used industrially for its high rates of protein production[147]. It has the largest available toolkit of genetic parts and regulators, DNA vectors, protocols, *in silico* models, and computational design programs of any organism used in synthetic biology today. Increasingly, as the sophistication of genetic circuits increases and potential applications of synthetic biology expand, the desire for a panel of well-characterised next-generation chassis organisms has led to the use of more versatile, more representative, or more resilient microbial species. These include the soil bacteria *Pseudomonas putida* and *Bacillus subtilis*, the photosynthetic bacteria *Rhodobacter sphaeroides* or the cyanobacteria *Synecchococcus elongatus*, extremophiles like *Geobacillus thermoglucosidasius*, the fast-growing *Vibrio natriegens*, and simple cells like *Mycoplasma genitalium*[148–152].

Acinetobacter baylyi is another soil bacterium that has been suggested as a next-generation synthetic biology chassis due to its robust and rapid growth – generation time of around 35 minutes in minimal media with succinate as a sole carbon source – its metabolic diversity, and its natural competence, making it useful in laboratory and industrial settings and highly amenable to genetic engineering[153–155]. *A. baylyi* will naturally incorporate into its genome DNA from the surrounding environment, and this recombination can be targeted to a specific locus based on sequence homology. This trait makes *A. baylyi* simple to transform and ideal for synthetic biology applications. *A. baylyi* has a wide substrate range including sugars, organic acids, aromatics, alcohols, and long-chain hydrocarbons, and the underlying genes are conveniently organised into clusters called 'catabolic islands'. A full, annotated genome sequence of 3.7 Mbp and basic toolkit elements are available for *A. baylyi*, as is a simple metabolic model[153–159].

A. baylyi boasts diverse anabolic pathways, including some for producing long-chain hydrocarbons; it is a natural producer of wax esters and triacylglycerols[160,161]. *A. baylyi* is a promising chassis for biorefining, especially biofuel production. As discussed in Part I of this chapter, an ideal drop-in biofuel would consist mainly of straight-chain alkanes, with some branched and cyclic alkanes and alkenes. Biosynthetic pathways for producing alkanes do exist, but they are slow or inefficient. Optimisation methods such as biosensor-based directed evolution might be able to improve these pathways though, as transcription-based biosensors for the detection of long-chain alkanes have been found[162]. The design goal would be a mutant strain of *A. baylyi* that can transform some renewable feedstock – such as bioelectrosynthesised acetate, or an extract from O*puntia* biomass – into alkane with industrially-relevant titres, rates, and yields.

Directed evolution

Evolutionary forces select based on biological function and are blind to underlying molecular mechanisms, so directed evolution represents a way of achieving functional design goals without comprehensive knowledge of the organism being evolved – with the caveat that organisms engineered in this way must create new generations on a time scale relevant to human interest; single-celled organisms are ideal. Harnessing evolutionary forces to achieve top-down design goals is a paradigm shift in engineering thought. It allows synthetic biologists to define the desired outputs of a biological system and then simultaneously alter – through random mutation – all underlying components at once, regardless of prior knowledge of those components. Screening and selection of beneficial mutations has so far been the bottleneck of directed evolution approaches, as the ability of synthetic biologists to create new DNA sequences and new mutants currently vastly exceeds their ability to evaluate them[163]. The development of new DNA barcoding techniques has enabled some high-throughput screening methods though, such as trackable multiplex recombineering (TRMR); the expression of over 4000 genes in *E. coli* was modified using barcoded oligonucleotides, with the barcodes facilitating detailed genotypic profiling following growth-based selection of mutant phenotypes[164].

Directed evolution involves creating a genetically diverse population of mutants and then screening and selecting for those mutants with desirable phenotypes. These phenotypes can arise from iterative diversification of one or multiple targeted genetic loci, or by random mutagenesis across the whole genome. Traditional random mutation strategies require a detectable phenotype to be linked to product formation and will only capture mutations within the coding region of interest where genetic variety is introduced[165]. Beneficial mutations occurring outside the region of interest can elude such screens, but a potential strategy for capturing these mutations is presented in Chapter 6. Most high-profile directed evolution technologies in use today focus on targeted evolution of genetic loci within an organism, and these technologies have been comprehensively reviewed in recent years by Raman *et al.*[166], Wang *et al.*[167] and by Simon, d'Oelsnitz, and Ellington[168,169]. Key technologies relevant to this project are briefly discussed below.

Multiplexed automated genome engineering (MAGE), a technique developed by George Church at Harvard's Wyss Institute, creates and screens a broad mutation space by simultaneously targeting multiple locations on the chromosome, generating combinatorial genetic diversity: 4.3 billion variants per day based on modification of lycopene biosynthesis across 24 sites[170]. MAGE was the first example of an automated

process for using exogenous oligonucleotides to modify genomes on a large and parallel scale. The approach uses a large pool of synthetic DNA to introduce sequence diversity in parallel across multiple genes in a metabolic pathway of interest. This captures otherwise neutral or deleterious mutations that would otherwise be lost but are potentially beneficial in combination. There is still the need for active intervention in the form of a screening step however, limiting exploration of an otherwise enormous sequence space.

Other high-throughput combinatorial recombination methods exist that build upon or were inspired by MAGE. Coselection-MAGE (CoS-MAGE) overcomes limitations in incorporation efficiency seen in the MAGE process by activating selectable markers with introduced oligonucleotides, which causes a local increase in incorporation efficiency of mutant sequences proximal to the marker[171]. Directed evolution with random genomic mutations (DIvERGE) uses iterative introduction by electroporation of overlapping mutant oligonucleotides which, when assembled, comprise the gene being targeted for evolution[172]. This process, called 'tiling', systematically optimises the combination of mutations present in the evolved gene, but may discover only a local performance maximum for certain gene products. These recombineering approaches were developed for use in bacterial chassis and are of limited use in eukaryotes due to their greatly reduced recombination efficiencies. The yeast oligo-mediated genome engineering (YOGE) system, also developed by the Church laboratory, nevertheless manages to achieve useful incorporation efficiencies in a *Saccharomyces cerevisiae* chassis engineered for reduced DNA mismatch repair functionality and overexpression of homologous recombination proteins[173]. Multiplexed CRISPR-based systems, such as CasPER[174], CREATE[175], and EvolvR[176], have proven effective at large-scale genome engineering in both prokaryotes and eukaryotes. SCRaMbLE, a recombinase-mediated combinatorial mutagenesis method developed specifically for the synthetic yeast genome project (Sc2.0), enables broad rearrangements within the synthetic yeast chromosome[177]. Iterative application of this technique to create diverse genotypes followed by selection for increased carotenoid production resulted in a 39-fold increase in carotenoid yield, demonstrating again the industrial applicability of this type of directed evolution approach[178].

Reliance on diverse exogenous oligonucleotides limits the extent of the mutation space that can be practically explored by the directed evolution methods outlined above. *In vivo* diversity generation is one of the fundamental advantages of a living system, and it facilitates continuous, consolidated diversification and selection without manual intervention. Random, genome-wide mutagenesis, while reflective of natural evolutionary processes, is slow, but introducing mutations via chemical or irradiative

technologies, or by accelerating basal mutation rate through suppression of DNA repair mechanisms, can often be deleterious to the host cell. Instead, *in vivo* mutagenesis can be targeted to specific genetic modules, leaving background mutation rates unaffected. The OrthoRep system consists of a plasmid which is exclusively replicated by an engineered error-prone DNA polymerase, which creates targeted *in vivo* random mutagenesis within a set of genes while leaving native, high-fidelity polymerases to replicate the host genome[179].

Viruses have also been used to achieve continuous or near-continuous directed evolution. Mechanistically distinct from, but functionally comparable to, exogenous-oligonucleotide-mediated directed evolution approaches, virus-mediated techniques currently lead the field in regard to throughput and have been applied to both bacterial[180] and mammalian systems[181], but are limited to evolving genes related to transcription or expression[182]. PACE (phage-assisted continuous evolution[183]), and its robotics-assisted high-throughput derivative PRANCE (phage- and robotics-assisted near-continuous evolution[184]), couple infectivity, and consequently growth and survival, of a virus to the capabilities of the target protein being evolved, which can include proteases, transcription factors, polymerases, and other DNA-binding proteins.

In silico methods leveraging machine learning may hold much potential for the directed evolution of enzymes, and the tools to support this approach have developed rapidly in recent years. In 2021, the AlphaFold tool, a neural network-based model developed by Alphabet's DeepMind, demonstrated an ability to predict the folded structure of certain proteins using their amino acid sequence as an input[185]. If quantifiable functional insights could be extracted from this structural information, indicative of the actual *in vivo* or *in vitro* performance of an enzyme, AlphaFold may help researchers leverage the unmatched processivity of *in silico* methods to enable the exploration of truly vast enzyme sequence spaces. Directed evolution of enzymes *in vivo* by linking function to a selectable phenotype is physically limited by high-throughput screening technology, but allows for the direct investigation of function without sequence information and for the evolution of enzymes within the broader cellular context.

Industrially, directed evolution is typically concerned with improved titres, rates, and yields (TRY) of some valuable metabolite. Directed evolution strategies aligning with this goal must either iteratively screen diverse populations for high-producing phenotypes or enable continuous directed evolution by coupling productivity to host fitness. The latter can be achieved by leveraging biosensory gene circuits that use some metabolite-of-interest to govern expression of another gene product.

Biosensor-based directed evolution

Allosteric transcription factors generate a regulatory genetic response in the presence of some ligand. Typically, binding at a distal site on the transcription factor causes conformational changes that alter its activity, enabling or preventing it from binding to a promoter sequence and, consequently, causing or suppressing gene expression. Such mechanisms are commonly used by microbes to monitor and respond to their environment, up- or down-regulating a huge variety of genes involved in metabolism, stress responses, intercellular signalling, and more. The *lac* operon of *Escherichia coli*, which functions to control lactose metabolism and transport in many enteric bacteria, is a classic example of this form of gene regulation, an understanding of which earned François Jacob and Jacques Monod a share of the Nobel Prize in Physiology in 1965[186].

Allosteric transcription factors that regulate gene expression in proportion to the concentration of some external metabolite, especially small molecules, are sometimes referred to as 'biosensors'. In the case of the *lac* operon, the regulatory protein LacI may be thought of as a biosensor for lactose (specifically for its derivative, allolactose, produced from lactose by the LacZ enzyme). These metabolite biosensors are central to engineering efforts across synthetic biology and are required for precise, controllable gene expression. Genetic systems of this type, inducible by small molecules such as isopropyl β-D-1-thiogalactopyranoside (IPTG, a molecular mimic of allolactose compatible with the *lac* operon), arabinose, anhydrotetracycline (aTc), salicylate, or even cationic dyes such as crystal violet, are cornerstones of molecular biology.

Many of these regulators are orthogonal across diverse host organisms and some have been engineered to respond to new ligands[187]. Often, especially in new microbial chassis, the relationship between ligand concentration and resulting changes in gene expression – called the 'transfer function' – is poorly understood or suboptimal, necessitating engineering of the biosensor system to improve specificity, reduce leakiness, or adjust dynamic range. It has even been possible to invert the native transfer function of some biosensors such that they repress, rather than activate, expression of particular genes[188]. As these inducible expression systems become more advanced, with synthetic biologists able to exert independent and fine control over a range of gene products, more complex interacting biological circuits can be layered atop one another, allowing sophisticated algorithms from control theory to be implemented in a biological context. In 2014, Brophy and Voigt published an excellent review of methods in genetic circuit design – still relevant now – which details the many types of logic gates that can be constructed using various biological parts[189].

Mannan *et al.* have also described, in great detail, fundamental design considerations for metabolite biosensors based on allosteric transcription factors, using the *lac* operon as a model system[190]. They investigated the relationship between promoter sequence, basal expression levels, achievable dynamic range, and characteristics of the transfer function, concluding that dynamic range of a biosensor can be adjusted directly by mutating either the promoter sequence or the binding affinity of the transcription factor to its target metabolite. While the latter approach requires structural insights regarding the transcription factor, it can also be useful both for reducing leakiness in biosensory circuits and improving the sensitivity of biosensory transcription factors.

Directed evolution methods based on biosensory genes aim to abolish the need for active screening by consolidating screening, selection, and amplification steps into a single, self-optimising system. Coupling detection of an intracellular metabolite, often the target product, to a measurable output, such as fluorescence or luminescence, functions as a screening step, with reporter genes being expressed in proportion to the prevalence of that metabolite. Linking detection of this valuable metabolite to a survival output translates product accumulation into evolutionary advantage; a selection step. High-producing phenotypes outcompete their peers and mutations benefiting production of desirable metabolites are fixed into the evolving population; an amplification step.

Previous biosensor-based evolution methods have coupled metabolite detection to factors including antibiotic resistance[191] or mutation rate[192], actuating the expression of rescue genes in response to valuable substrates to defend against antibiotic challenge or excessive mutation rate, via expression of an antibiotic efflux pump or repression of error-prone DNA polymerase, respectively[193]. Notably missing from previous attempts is the restoration of growth characteristics in auxotrophic mutant chassis in response to production of a valuable metabolite; such an approach would incentivise evolution via a 'carrot' rather than a 'stick'.

Existing biosensor-based evolution methods predominantly utilise pre-existing natural sensing machinery; a table of natural and engineered biosensors can be found in the paper by Rogers, Taylor, and Church 2016[193]. This limits the application of biosensor-based directed evolution to optimising pathways resulting in detectable metabolites, but new biosensors will continue to be mined from metagenomic sequences[194] or created artificially by alteration of existing sensors[195].

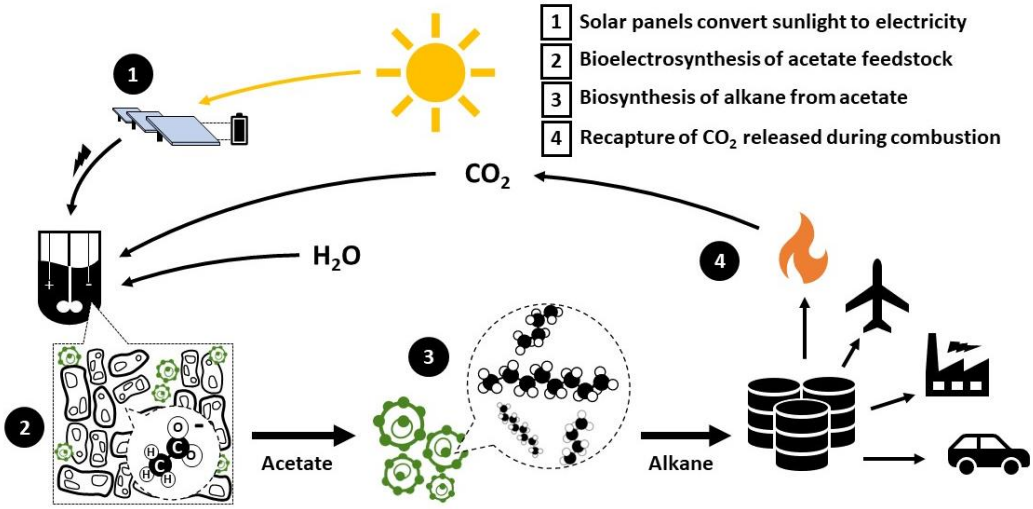


Figure 5. A circular carbon process for capture of solar energy in liquid fuels ('solar fuels'). Solar panels, which can be substituted with any other renewable electricity generation method, produce low-carbon electrical energy, which is used for the electrosynthesis of a precursor feedstock, such as formate, from atmospheric CO_2. This precursor feed is metabolically upgraded to a higher-energy carbon molecule, such as acetate. This secondary feedstock is again upgraded by microbes into useful hydrocarbon products, such as alkanes, for blending into liquid fuels. Upon combustion, these fuels release their embodied CO_2 back into the atmosphere for recapture, closing the loop on a low-carbon or carbon-neutral energy cycle. Electrosynthesis of precursor feedstocks may be replaced by direct carbon fixation by photosynthetic organisms, such as growth of energy crops to produce sugars for microbial conversion to alkanes.

Research gap

Examination of the academic and industrial research spaces surrounding advanced biofuels reveals a clear gap: there is great demand for sustainable biofuels that match the performance of petroleum fuels. Creating such biofuels will inevitably involve biosynthesis of alka(e)nes – linear, branched, and cyclic – and aromatics such as terpenes, as these compounds boast far more desirable properties than the alcohol-based biofuels that are widely produced today; more energy dense, less corrosive, more stable. Alkanes also represent an attractive target, useful both as chemical precursors for upgrading to higher value fuel molecules and as low-carbon blendstocks. Biosynthesised alkanes would represent a step along the path to creating a true drop-in replacement for petroleum fuels, rather than settling for a compromise between sustainability and utility.

Microbial production of alkanes and other advanced biofuel components has been well researched, but yields are uneconomically low. Metabolic engineering towards improved production rates has so far been dominated by top-down rational design

strategies, which are hindered by the huge physiological complexity of chassis organisms. Directed evolution represents an optimisation strategy free from such limitations but has not yet been applied to alkane biosynthesis or been applied in a naturally transformable auxotrophic bacterial chassis.

The doctoral research project described in this thesis, carried out at Begbroke Science Park under the supervision of Professors Wei Huang and Ian Thompson, bridges this research gap in the following ways:

a. This research contributes to the development of a sustainable source of advanced biofuels, such as that outlined in *Figure 5*, which depicts a circular carbon energy process that may help alleviate reliance on fossil fuels.

b. The biosensor-based directed evolution system – which I have called the artificial evolution, or "Artevo" system – described herein was built in a novel auxotrophic bacterial chassis developed specifically to host said system.

c. The synthetic genetic circuit driving the evolution process is chromosomally integrated, so the microbial chassis itself is self-adapting, self-screening, and self-selecting; the directed evolution process should proceed without iterative active intervention steps, representing an improvement over existing processes.

d. The Artevo system was built using a naturally competent bacterial chassis, which simplifies the evolution process and facilitates expanded applications in continuous directed evolution and in functional metagenomics.

e. The Artevo system has been shown to couple host growth rate to local alkane concentration, consequently demonstrating an ability to interrogate a diverse population of phenotypes for alkane production capacity and enrich the phenotype carrying the highest-performing alkane biosynthesis genes.

The remaining chapters in this thesis discuss the development and application of the Artevo system, and how insights gained during this project contribute to closing the research gap outlined above.

Chapter 3

A novel host for directed evolution

This chapter describes the opening stages of the project, beginning with design considerations for the greater artificial evolution system – which I will continue to abbreviate as Artevo – and choice of a suitable bacterial chassis to house it. It follows the design of the gene circuit underpinning the Artevo system, as well as generation and characterisation of the novel knockout mutant host, which was first characterised in this work. Following considerations on biofuel and synthetic biology put forward in Chapter 2, alkanes were selected for production by the Artevo system and *Acinetobacter baylyi* was chosen as the host chassis, for its useful combination of diverse metabolism, natural competence, and genetic malleability.

Design of the gene circuit

The genetic circuit detailed in *Figure 6* drives evolution of high-performing bacterial phenotypes by linking biosynthesis of a valuable metabolite to cell fitness. This link is fundamental to how the proposed Artevo system will operate. Core to the system is a biosensory gene which, for illustrative purposes, we will denote as *senS*, sensitive to some metabolite of interest, which we will call *M*. The biosensor *senS* encodes a

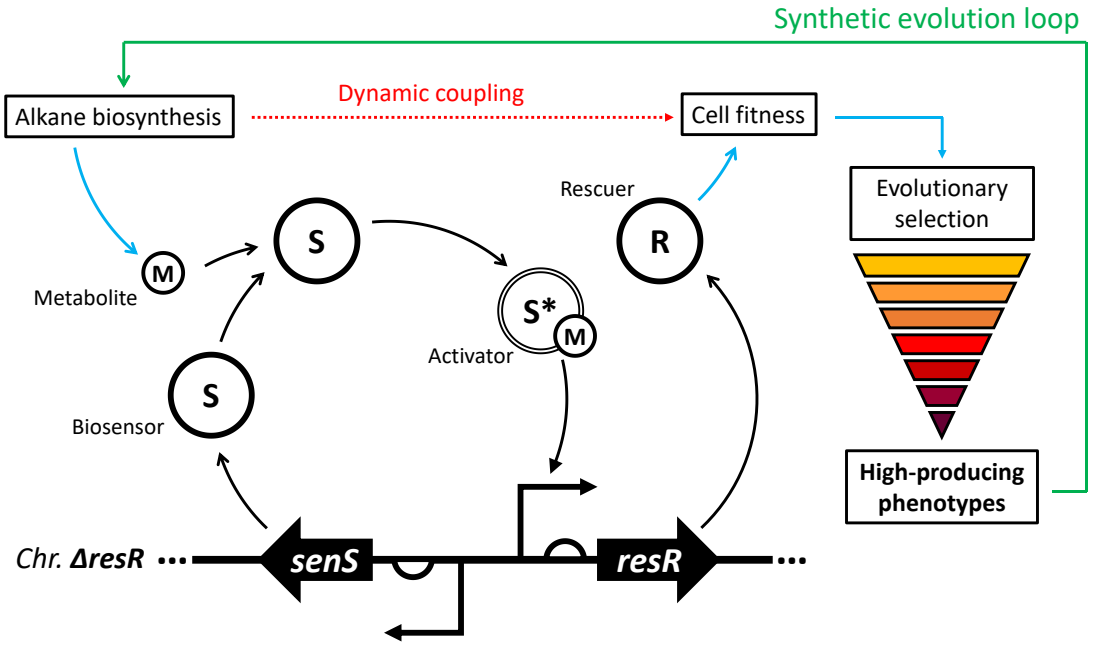


Figure 6. A genetic circuit for biosensor-based directed evolution of production of some valuable metabolite. The concentration of a metabolite-of-interest (M) is linked to expression of a rescue gene (*resR*) by a biosensor gene (*senS*). The rescue gene product (R) improves fitness of the host cell in proportion to its expression level, dynamically linking biosynthesis of metabolite (M) to host growth rate and causing generational enrichment of high-performing biosynthesis genes through evolutionary selection. In this case, the rescue gene *resR* is a growth-essential gene that has been deleted from its native genomic locus, creating an auxotrophic host (⊿*resR*), and placed under control of a promoter regulated by the biosensor *senS*. The metabolite (M) may be a secondary metabolite proportional to the target product (such as formate, produced during alkane biosynthesis by ADO), or, as in in this case, it may be the target product. In the context of this project, this circuit architecture is referred to as the 'artificial evolution' – or Artevo – circuit.

regulatory protein which, in the presence of metabolite *M*, activates expression of a gene that contributes favourably to host fitness. This gene is typically called a rescue gene and is denoted here as *resR*. The mechanism by which a rescue gene improves host fitness can vary; antibiotic resistance genes are broadly compatible with many strains used in synthetic biology and have been used to rescue growth in cells pressured by antibiotics[191]. Alternatively, the rescue gene may facilitate metabolism of an abundant carbon or energy source. In the Artevo system, a growth-essential gene is deleted from its typical genomic locus and placed under control of the biosensory regulator; specific genetic details can be found later in this chapter. In all cases, the concentration of metabolite *M* detectable by biosensory regulator proteins is dynamically linked to cell fitness. If a population of cells can all manufacture *M* intracellularly, those producing the greatest concentration of *M* should be relatively fitter than cells producing lower amounts of *M*. The high-producing phenotypes should grow faster and outcompete their neighbours, eventually dominating the population and enriching the genes contributing to effective production of *M*. Over successive generations, evolutionary

selection will cause genotype frequency in the population to shift towards that which facilitates maximum production of metabolite *M*, as expected when beneficial mutations emerge in any natural population.

Diversity generation

Creating a selective environment requires variation between individuals. This variation can arise from sequence diversity within the biosynthesis genes directly involved in production of metabolite *M*, or from mutations occurring elsewhere on the genome that are not deleterious to the general health of the cell. The prototype Artevo system I describe in this thesis is designed to handle intentional diversity generation only within genes directly responsible for biosynthesis of metabolite *M*, though it will also capture random beneficial mutations occurring across the genome. Biosynthesis genes may be native to the host chassis and diversity can be introduced directly at their genomic locus, though preserving the rest of the genome from mutation can be challenging. Diversity generation in an exogenous DNA library, which is then introduced via a plasmid or integrated into the host genome, is typically simpler. Chapter 6 expands upon various approaches to both local and genome-wide diversity generation, providing strategies and designs for a second-generation Artevo system where the driver module, consisting of the biosensor gene and rescue gene plus any tuning elements, is isolated from the host genome and processed by dedicated high-fidelity polymerases. Such a system should allow for genome-wide diversity generation without compromising constituent genes within the directed evolution circuit, though it may be difficult to effectively tune such that genomic mutation rates remain below the threshold for single-generation lethality.

The synthetic evolution loop depicted in *Figure 6* refers to the process of isolating high-performing sequences enriched by the Artevo system and using them as inputs for a subsequent round of diversity generation and directed evolution. Such a process is a classical example of directed evolution requiring manual intervention between cycles of selection, and Chapter 6 discusses how the Artevo system may be adapted toward continuous directed evolution, which requires no such intervention. This thesis focuses on the genetic circuitry at the core of the evolution loop, its scope bounded upstream by the diversity generation step and downstream by isolation and analysis of enriched variants. The overall synthetic evolution loop has not been optimised as part of this thesis, but the loop is presented to provide context for the gene circuits that have been built for the Artevo system.

The alkane biosensor

The biosensor-based directed evolution strategy outlined above is applicable to any metabolite, provided there exists:

a. A biosynthesis pathway for production of the metabolite.

b. A biosensory activator gene system that can detect the metabolite.

c. A rescue gene able to confer a growth advantage to the host proportional to gene expression.

d. A microbial chassis able to host all the relevant genes.

Note that, in this work, the *alkRM* system acts as an activator of rescue gene expression. Biosensory repressor systems would require either additional circuit architecture to invert their response or the use of a toxic, rather than beneficial, 'rescue' gene, the suppression of which translates into a growth advantage and which is not lethal at maximum expression levels. Repressor-based systems would be more susceptible to mutational escape and restoration of a high-growth phenotype, creating a type of 'cheater' cells, which are unproductive but exhibit normal growth.

Biosynthesis pathways are well characterised for many industrially useful chemicals, and the available suite of biosensors is being constantly expanded as novel sensors are discovered or derived from existing ones. As a proof-of-concept, the Artevo system developed in this project was designed to link alkane biosynthesis to growth rate via restoration of cell wall synthesis capacity in an auxotrophic host.

Having established alkanes as the target metabolite (*M*) for biosynthesis, we can start introducing specific components to the general circuit structure shown in *Figure 6*. The core circuit in the Artevo system utilises an alkane biosensor originally developed by Zhang *et al.*, under supervision by my supervisor Prof. Huang, in *Acinetobacter baylyi* for use in bioremediation[162]. Alkanes are detected by the constitutively expressed transcriptional regulatory protein AlkR, which regulates the alkane monooxygenase gene *alkM* to enable alkane degradation. The *alkRM* system is native to *A. baylyi*, functioning to enable these bacteria to detect and utilise environmental alkanes as a carbon source. It is thought to have 'activated-activator' biosensor architecture, with the AlkR biosensor protein becoming activated upon binding alkane, subsequently activating expression of *alkM*. AlkR is a member of the araC-XylS-like family of transcriptional regulators and binds medium- and long-chain alkanes in the C7-C36 range[196,197], but has some sensitivity for alkenes and a preference for C17 alkanes, as investigated in Chapter 4. AlkM is an integral-membrane hydrocarbon hydroxylase

which catalyses the oxygenation of alkane substrates into primary alcohols, and its deletion removes host ability to degrade alkanes[196]. The biosensor built by Zhang *et al.* replaces *alkM* at its native genomic locus with the cyanobacterial *luxCDABE* operon, which produces a bioluminescent signal in the presence of alkanes. The Zhang biosensor strain (genotype: ⊿*alkM*::*luxCDABE*) will luminesce in proportion to the concentration of exogenous alkanes in its environment, as shown in *Figure S.1*, and was originally intended for detection of oil spills in seawater.

Applying this biosensory circuit in the context of directed evolution, a rescue gene can be placed under the *alkM* promoter, meaning expression of the rescue gene will be regulated by AlkR. *alkM* or any analogues should be deleted if present in the host to prevent degradation of the target bioproduct. A bacterial chassis equipped with such a circuit should exhibit variable growth rate in response to changing concentrations of exogenous or intracellular alkane. It is assumed that intracellular alkanes will have a greater effect on expression of the rescue gene than exogenous alkanes, due to the localisation of biosensory machinery within the cell. Consequently, in the absence of exogenous alkane, growth will be restricted to cells producing intracellular alkane, either natively or via an introduced enzymatic pathway. Alkane exported from productive cells may be detected by another kind of 'cheater' cells, which derive a growth advantage from environmental alkanes but produce none themselves. If intracellular alkanes do indeed confer a greater growth advantage in the host than exogenous alkanes, as assumed, cheater phenotypes should not persist within the population as they will be continuously outcompeted. This assumption was eventually validated later in the project and is discussed in Chapter 5. Higher-producing phenotypes will have a survival advantage and the population will evolve towards higher alkane titres over successive generations. Inclusion of a reporter gene, such as *gfp*, under the *alkM* promoter should also allow the system to signal activation of the evolution circuit.

Chassis selection

The microbial chassis selected to house the directed evolution circuit in *Figure 6* represents the central living element of the Artevo system. Chapter 2 reviewed the use of common prokaryotes in synthetic biology, chiefly *Escherichia coli*, the workhorse of bacterial biotechnology. The existing *A. baylyi* ADP1 alkane biosensor strain, developed by Zhang *et al.*[162] and described in the previous section, prompted consideration of *A. baylyi* and other 'next-generation' bacterial chassis as a host for the directed evolution circuit.

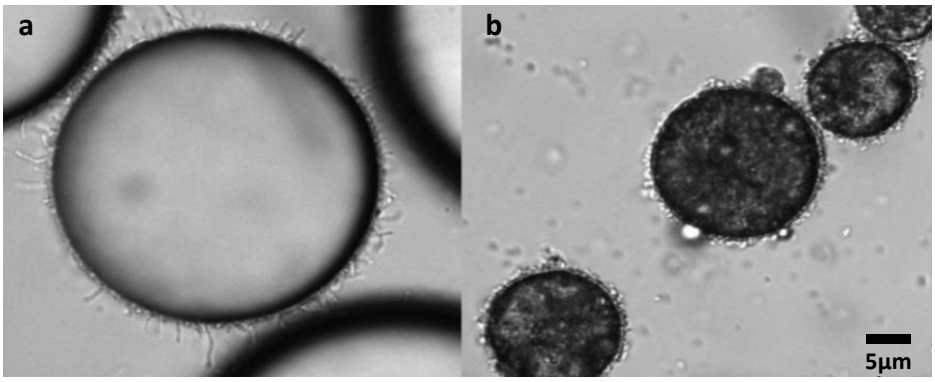


Figure 7. Bright-field microscopy of *Acinetobacter baylyi* ADP1, with mineral oil (a) and crude oil (b) droplets adhering to the cell surface. Images republished with permission from Zhang *et al*[161].

Acinetobacter baylyi ADP1, formerly *Acinetobacter calcoaceticus* – is a gram-negative, prototrophic, strictly-aerobic cocci exhibiting a wide substrate range and diverse metabolism, including native alkane detection and degradation machinery as well as synthesis pathways for long-chain hydrocarbons[198–200]. Usefully, it also releases emulsifying capsular polysaccharides that pseudosolubilise hydrophobic compounds in the culture media without added surfactants, enabling oil droplets to adhere to the cell surface and facilitating their uptake into the cell[153]. It is hardy, fast-growing, and easily cultivated on rich and minimal salts at mesophilic temperatures (optimum 30-37°C)[155]. *A. baylyi* ADP1 contains no known pathogenicity or virulence factors but is used as a model organism for the study of pathogenic *Acinetobacter* sp., including the close relative *Acinetobacter baumanii*, which is listed by the World Health Organisation (WHO) as a critical priority for development of new antimicrobials due to its multidrug resistance[201].

A. baylyi ADP1 is naturally competent during exponential growth phase and maintains native recombination machinery, making it highly amenable both to transformation with plasmids and to direct genomic edits. The latter are made through homologous recombination and involve simple addition of linear or circular double-stranded DNA (dsDNA) to growing culture. As a naturally competent and recombinant bacterium, *A. baylyi* ADP1 is even capable of *in vivo* assembly of plasmid constructs, enabling automated cloning workflows[202].

These features distinguish *A. baylyi* ADP1 as a promising next-generation chassis for synthetic biology. Consequently, the toolkit for this organism has been steadily growing, as discussed in Chapter 2, and it was significantly expanded in recent years when Biggs *et al.* published a promoter library and genomic insertion hotspot map[159].

Usefully, a streamlined version of ADP1 exhibiting improved transformability has been produced by removing transposons from the genome[203]. This chassis has been called *A. baylyi* ADP1-ISx by its creators, Suárez *et al.*, and was generously provided for use in

this project by Prof. Jeffrey E. Barrick at the University of Texas at Austin. The lack of transposable elements in the ADP1-ISx genome is likely to increase the long-term stability of genomic insertions[204] which, together with the heightened transformation efficiency of ADP1-ISx, make this strain ideal for use in continuous directed evolution.

For general commentary not directly discussing mobile genetic elements, where distinguishing between *A. baylyi* ADP1 and *A. baylyi* ADP1-ISx is irrelevant, this thesis conflates the two strains. Most figures present only experimental data obtained using the ADP1-ISx chassis, and strain nomenclature in the text builds upon the ISx designator (ISxJK001, ISxJK002, ISxJK003, etc). However, for many early experiments, identical workflows were carried out in parallel for both *A. baylyi* ADP1 and *A. baylyi* ADP1-ISx. The behaviour of ADP1 and ADP1-ISx is broadly similar for all experiments covered by this project, but data specific to the ADP1 chassis, where available and relevant, will be referenced in the supplementary information section accompanying this thesis.

Natural competence

For synthetic biologists, one of the most attractive features of *Acinetobacter baylyi* ADP1 is its growth phase-specific natural competence and capacity for natural homology-mediated recombination[205]. In nature, this ability is useful for horizontal acquisition of adaptive traits and sourcing templates for DNA repair[206]. In a laboratory context, these functions have been used to streamline the traditional molecular biology workflow and enable *in vivo* assembly (IVA) of DNA fragments, even in chassis where this capability is generally assumed to require recombination-enhancing gene edits[207]. In ADP1, simple addition of dsDNA to log-phase culture can be used to make precision edits to the host genome[208]. Interestingly, because the mechanism of DNA uptake is thought to be mediated by type-IV pili, transformation efficiency of ADP1 increases when cultures are grown on solid media compared with liquid media[209]. This quirk was used a few times in this project to facilitate difficult transformations and genomic integration of large cassettes, via the aptly named 'puddle transformation' method[210].

Genetic malleability is central to the success of this project, both in terms of making edits to the base chassis and subsequent continuous introduction of diverse DNA into the evolving population. Continuous directed evolution processes are often enabled by robotics, which automate the cycle of *in vitro* diversity generation, transformation, and amplification of high-performing sequences. Continuous directed evolution in a naturally competent chassis sidesteps the need for disruptive intervention during the transformation step, potentially facilitating *in situ* continuous directed evolution. The

immediate benefits of this approach are a reduction in complexity and reduced system perturbation, but naturally competent chassis may even be combined with *in vivo* diversity generation to consolidate the entire continuous directed evolution cycle into a single unit operation. This concept is discussed further in Chapter 6.

Selecting the rescue gene

Having established a target bioproduct and a chassis strain with relevant native biosensory genes, a suitable rescue gene could now be selected. For the Artevo system, it was decided that an essential gene in ADP1 should be deleted from its native genomic locus and moved to replace *alkM* at the *alkRM* locus.

Meso-diaminopimelate (*meso*-DAP) is an essential component of peptidoglycan in cell walls of all gram-negative bacteria and gram-positive bacilli. In gram-negative bacteria, a *meso*-DAP-alanine linkage between adjacent polysaccharide strands is essential for cross-linking of peptidoglycan moieties, without which the cell will be structurally compromised and undergo osmotic lysis[140,211]. *Meso*-DAP is also present as an intermediate in aspartate-derived biosynthesis of lysine, which is essential in protein production[212]. *Meso*-DAP formation occurs via three routes in bacteria[213]: the succinylase pathway – for which enzymes appear to be present in *A. baylyi* ADP1 – and the dehydrogenase and acetylase pathways, which ADP1 lacks, as illustrated in *Figure 8*.

The *dapB* gene encodes a 4-hydroxy-tetrahydropicolinate reductase which is required for biosynthesis of L-tetrahydropicolinate, the substrate that feeds into all three *meso*-DAP synthesis pathways. Strains lacking a functional *dapB* gene have been shown to be auxotrophic for both *meso*-DAP and lysine in *Escherichia coli*[214] and in *Pseudomonas* spp.[215,216]; *Figure 9* illustrates the metabolic context of this auxotrophy in ADP1.

⊿*dapB* mutants have broad research applications, having been exploited in *in vivo* expression technology (IVET)-based promoter traps to study plant-microbe interactions[217]; in a clinical context to discover new antimicrobials, especially antitubercular agents, due to the lack of the pathway in mammals[218–220]; in DNA delivery and DNA immunisation[221]; and in maintaining bacterial plasmids in animal experiments[222]. In the context of the Artevo system, *dapB* may be employed as a rescue gene to restore growth in an auxotrophic ⊿*dapB* chassis in response to production of alkanes. To demonstrate the feasibility of this approach, a viable ⊿*dapB* knockout mutant first had to be generated in *A. baylyi* ADP1. Such a mutant is absent from the comprehensive collection of single-gene mutants in ADP1 published by de Berardinis *et al.* in 2008, and was obtained for the first time in this work.

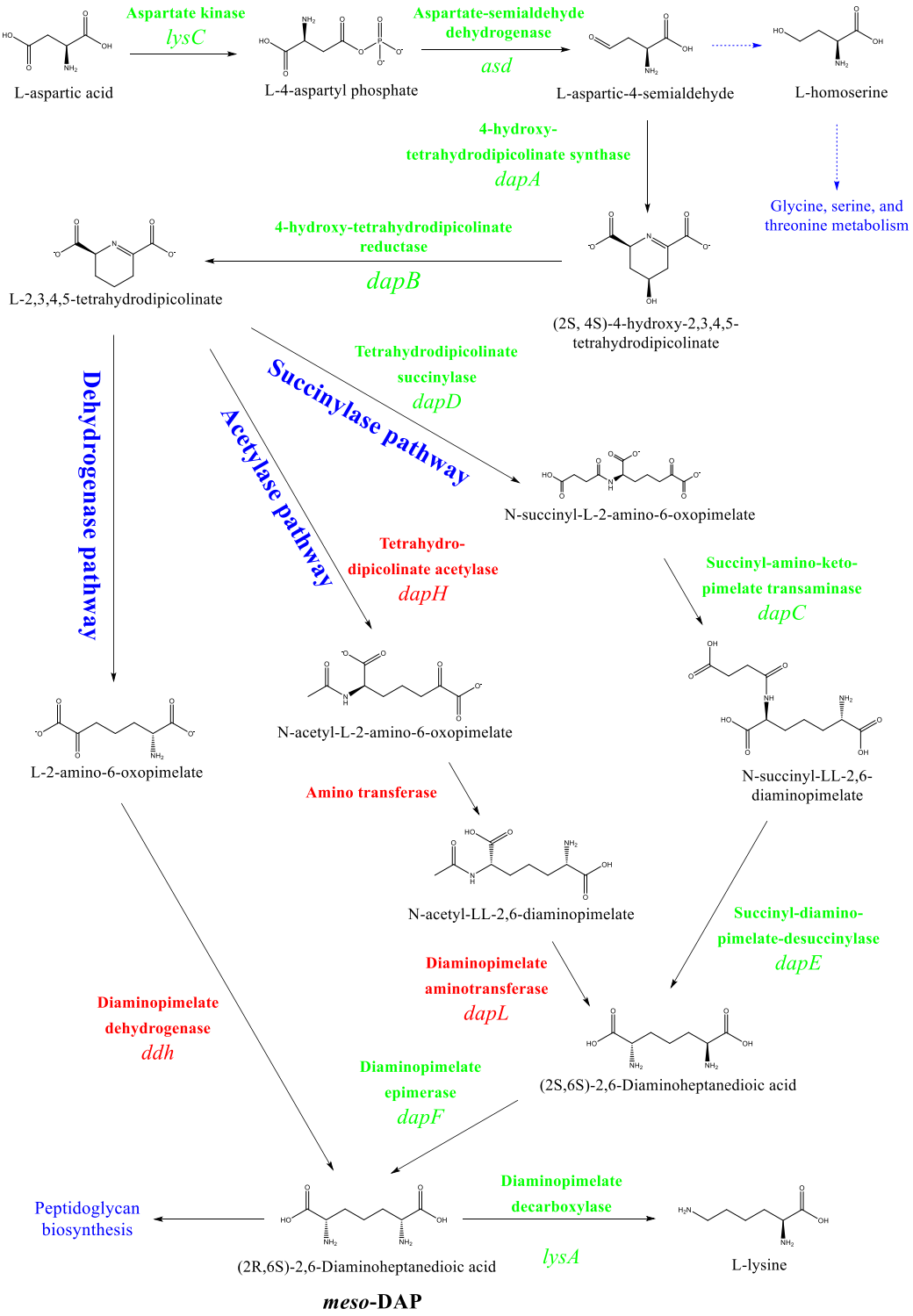


Figure 8. Meso-DAP biosynthesis pathways. The dehydrogenase pathway forms *meso*-DAP directly via diaminopimelate dehydrogenase (ddh) and is found in some *Bacillus* species. The succinylase and acetylase pathways are largely analogous, utilising N-succinyl or N-acetyl intermediates, respectively. The succinylase pathway is common in proteobacteria, such as *A. baylyi*, and actinobacteria. The acetylase pathway is limited to certain *Bacillus* species and is not fully characterised[212]. Genes depicted in green are present in *A. baylyi* ADP1.

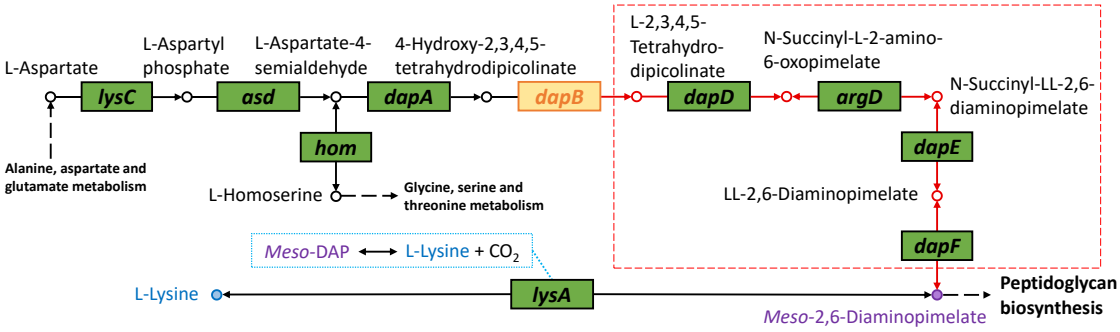


Figure 9. Metabolic context of lysine auxotrophy in *dapB* deletion mutants of *Acinetobacter baylyi* ADP1. Routing of metabolites through the genes bounded in red is disrupted by deletion of the *dapB* gene, shown in orange. The cell wall component *meso*-DAP must be obtained from supplemental L-lysine by reversible action of LysA protein.

Constructing the host chassis

The following sections describe the approach used to generate a lysine-auxotrophic mutant of *A. baylyi* ADP1, including identification of the *dapB* locus and deletion of *dapB* from the chromosome. For the directed evolution circuit in *Figure 6* to function as intended, the *dapB* gene – (EC 1.17.1.8 in ADP1, encoding DapB protein) – needed to be moved from its native genomic locus and used to replace the native ADP1 *alkM* gene, which is expressed in the presence of alkanes. The cloning strategy for deleting *dapB* is described in full below and illustrated, with results, in *Figure 10*.

Meso-DAP is derived from aspartate or lysine in ADP1. Deletion of the *dapB* gene eliminates carbon flux through the pathway bounded in red in *Figure 9*, so ⊿*dapB* mutants in ADP1 must utilise exogenous DAP or obtain it by conversion from lysine via reversible action of diaminopimelate decarboxylase (*lysA*). Enzymatic action of LysA protein is listed in the Kyoto Encyclopaedia of Genes and Genomes (KEGG) as being one-directional for conversion of DAP to lysine, but KEGG reaction data (R00451), and our own results, shows that it is reversible.

Methods and results

Strains and culture media

A list of important strains and plasmids used in this research is presented in *Table 2*. Unless specified in the description, strains were sourced as part of this work. For brevity, strains generated for tangential research, such as ⊿*lysA* deletion mutants or

mutants containing a *dapB-gfp* fusion protein, are omitted. All *Acinetobacter baylyi* strains were grown at 30°C and *Escherichia coli* strains were grown at 37°C. Liquid cultures were shaken at 200 rpm unless otherwise stated. Unless otherwise stated, chemicals used were analytical-grade reagents obtained from Sigma-Aldrich. Luria-Bertani (LB) medium or minimal medium (MM) as described by Huang *et al.*[223] (recipe detailed in supplementary information for Chapter 3, on page 151) was used for culturing bacteria as appropriate. MM with succinate as a carbon source (MMS) was prepared by adding succinate to MM to a final concentration of 20 mM. Where antibiotics are used, final concentrations are as follows unless specified: ampicillin at 100 μg/mL for *E. coli*, and kanamycin at 50 μg/mL for *E. coli* or 30 μg/mL for *A. baylyi*. Where required, culture media was supplemented with 2,6-diaminopimelic acid (DAP) and L-lysine at 800 μg/mL and 100 μg/ml, respectively, unless other concentrations are stated. When mineral oil or another alkane was mixed with culture media, the mixture was vigorously vortexed to combine the two immiscible phases, then immediately sonicated in a water bath at 42 kHz for 60 seconds to pseudosolubilise the hydrocarbon phase. Alkane solubility in aqueous culture media is extremely low, but this approach enabled reliable production of semi-stable emulsions with hydrocarbon phases that could be readily accessed by cells. Notably, higher concentrations of emulsified mineral oil increase the turbidity of media solutions, which can mask the starting inoculation density of cultures during growth curves. Experimentalists may correct for this by subtracting the optical density (OD600nm) of media-only controls measured periodically alongside test conditions. Initial OD600 of cultures should not be used as a constant measure of basal absorbance throughout the growth curve, as mineral oil emulsions have been observed to destabilise over time; phase separation occurs over 6-8 hours in 250mL culture flask (50mL working volume) and is faster in microtitre plate. In some experimental setups, where strains have very low growth rates and consistent inoculation densities across media conditions – such as that shown in *Figure 26(a)* later in Chapter 5 – it can be simpler not to adjust for higher basal absorbance of oil-containing media, which still allows for clear observation of growth phenomena.

DNA manipulation and strain construction

All enzymes for molecular cloning were obtained from New England Biolabs (NEB). General molecular cloning was conducted in NEB 5-α high-efficiency competent cells through either restriction cloning or through NEBuilder HiFi DNA Assembly. DNA amplification for assembly applications was carried out using Q5 high-fidelity DNA polymerase. DNA amplification for screening purposes was performed using DreamTaq DNA polymerase from ThermoFisher. DNA purification was performed using a PCR purification kit, QIAprep spin miniprep kit, and QIAquick gel extraction

kit, all from QIAGEN. All deletion cassettes were assembled in a pGEM vector backbone from Promega which cannot replicate in *A. baylyi* ADP1. PCR screens were visualised using gel electrophoresis (0.8-1.0% w/w agarose gels) and compared with a GeneRuler 1 kb DNA ladder from ThermoFisher.

Deletion of *dapB* from the ADP1 genome

The amino acid sequence of DapB from *E. coli* was used to search against the ADP1 proteome using the NCBI Basic Local Alignment Search Tool (BLASTp) suite[224]. The top hit (Query cover: 98%; Identity; 62%; E-value: 1e-103; Sequence ID: WP_004923088.1) was annotated as "MULTISPECIES: 4-hydroxy-tetrahydro-dipicolinate reductase [Acinetobacter]", and the coding region on the ADP1 chromosome was deduced to be the *dapB* gene. Primers to amplify 1 kB regions immediately upstream ($dapB^{up}$) and downstream ($dapB^{down}$) of *dapB* were designed and used to construct pGEM-$dapB^{up}$ and pGEM-$dapB^{down}$. The *dapB*-targeting vector pJK001 was then constructed by restriction cloning of $dapB^{up}$ under the BamHI and SpeI sites in pGEM-$dapB^{down}$. The *dapB* deletion vector pJK002, consisting of a *tdk*:Kan^R counterselection cassette from pYTK001-*tdk*.KanR flanked by $dapB^{up}$ and $dapB^{down}$ in a pGEM backbone, was then constructed PCR amplification of constituent fragments with overlap primers and subsequent HiFi assembly.

The *dapB* deletion cassette was amplified from pJK002 then treated with DpnI to remove residual template DNA. ADP1 was transformed by addition of approximately 1 μg of purified PCR amplicon (high-purity, linear dsDNA) and 70 μL of liquid cell culture (which has been grown overnight in LB) to 1 mL of fresh LB broth. The transformation mix was incubated overnight at 30°C / 200 rpm and then serial dilutions were plated onto LB agar supplemented with kanamycin, DAP, and lysine. This supplemental lysine was provided despite LB media already containing lysine[225]; the exact concentration of lysine in LB broth or agar was unknown, as LB is a complex, undefined media.

After approximately four days incubation at 30°C, small colonies appeared that were screened by colony PCR for replacement of the *dapB* CDS with the *tdk*:Kan^R cassette. A second colony PCR reaction utilising primers specific to sites internal to the *dapB* CDS was performed to confirm loss of *dapB* and verify it had not been retained elsewhere on the genome.

Table 2. Strains and plasmids used in this work. Key strains and plasmids are highlighted in gold.

Strains and plasmids	Description
***Acinetobacter baylyi* strains**	
ADP1 (BD413)	Wild-type (WT). Referred to in text as ADP1. Source: Zhang *et al.*[162]
ADP1-ISx	ADP1 with all six copies of IS*1236* deleted from the genome, showing reduced mutation rates and higher transformation efficiency. Genotype: ⊿*is1236*. Source: Suárez *et al.*[203]
ADPWH_alk	ADP1 alkane biosensor developed by Zhang *et. al.*[162] Luminesces in proportion to exogenous alkane concentration. Contains a modified *alkRM* promoter decoupling expression from growth phase. Genotype: ⊿*alkM*::*luxCDABE* $P_{alkRM.mod}$
ISxJK001	⊿*dapB* mutant of ADP1-ISx. In-frame replacement of entire *dapB* CDS with a BamHI restriction site. ADP1 counterpart is ADPJK001. Genotype: ⊿*is1236* ⊿*dapB*
ISxJK001c	$dapB^+$ complement of ISxJK001, with *dapB* reintroduced at original locus.
ADP1-ISx *alkRM*$^-$	ADP1-ISx with an in-frame deletion of the native *alkRM* promoter and the *alkM* gene, replaced by a *tdk*.Kan^R counterselector cassette.
ISxJK001 *alkRM*$^-$	As above, but in the ISxJK001 *dapB* deletion mutant.
ISxJK002	ISxJK001 chassis equipped with the basic Artevo circuit; a promoterless *dapB* and a *gfp* fluorescent reporter replace the native *alkM* gene. ADP1 counterpart is ADPJK002. Grows faster and fluoresces in the presence of exogenous alkanes. A high-performing (minimal leakiness) singlicate of this strain is denoted as 002*. Genotype: ⊿*is1236* ⊿*dapB* ⊿*alkM*::*dapB gfp*
ISxJK003	ISxJK001 chassis equipped with upgraded Artevo circuit architecture: modified *alkRM* promoter, brighter *sfgfp* replaces *gfp*. Grows and fluoresces in the presence of exogenous alkanes, with improved specific fluorescence but binary growth rather than graded. Genotype: ⊿*is1236* ⊿*dapB* ⊿*alkM*::*dapB sfgfp* $P_{alkRM.mod}$
ISxJK003 Pro1/A/B/D	As ISxJK003 but containing an insulated circuit structure with overexpression of *alkR* by a Pro-series promoter, from Davis *et al.*[226] Promoters and relative expression strengths are: ProD=1.00, Pro1=0.009, ProA=0.030, ProB=0.119.
ISxJK003 BCD9	As ISxJK003 but *alkR* is overexpressed by a strong bi-cistronic design (BCD) promoter, BCD9, from the Biggs *et al.* toolkit[159].
ISx *sfgfp* anti.Trc	A proof-of-concept ADP1-ISx strain containing the modified *alkRM* promoter. *alkM* has been replaced with *sfgfp*, which is opposed by an antisense Trc promoter placed downstream. Inducible expression of this promoter is facilitated by a *lacI* gene under the constitutive *lacIq* promoter, chromosomally integrated immediately downstream of the Trc promoter. Can also function as a fluorescent alkane biosensor. Genotype: ⊿*alkM*::*sfgfp*-$P_{Trc.anti}$ *lacI* $P_{alkRM.mod}$
ISxJK004 $LVA^{+/-}$	ISxJK001 chassis with an upgraded Artevo circuit; contains the modified *alkRM* promoter, *dapB* rescue gene with and without an attached ssrA degradation tag (denoted LVA^+ or LVA^- respectively), and a convergently-oriented inducible Trc promoter opposing *dapB* expression. Genotype: ⊿*is1236* ⊿*dapB* ⊿*alkM*::*dapB.LVA*-$P_{Trc.anti}$ $P_{alkRM.mod}$

Strains and plasmids	Description
ISxJK004-S1/2 pBAV1k *gfp lacI*	ISxJK004 LVA$^+$ (S1) or LVA$^-$ (S2), containing the plasmid pBAV1k *gfp lacI*. Allows for tuneable stringency of selection in ISxJK004 LVA$^+$. Exhibits a binary growth response proportional to exogenous alkane concentration.
ISxJK004 LVA$^+$ pTL2 [alk]	ISxJK004 LVA$^+$ transformed with one of three alkane biosynthesis plasmids: pTL2 seADO-AAR, pTL2 pmADO-RAMO, or pTL2 CER1-ACR1. Exhibits a varying growth response dependent on productivity of the alkane biosynthesis module carried by pTL2.
***Escherichia coli* strains**	
NEB-5α (DH5α)	High efficiency competent cells. Referred to in-text as DH5α. Source: New England Biolabs
DH5α pGEM *alkR sfgfp*-$P_{Trc.anti}$ *lacI*	A proof-of-concept strain for testing antisense regulation of gene expression. NEB-5α (DH5α) strain containing the plasmid pGEM *alkR sfgfp*-$P_{Trc.anti}$ *lacI*.
Plasmids	
pYTK001-tdkkan	Source of thymidine kinase (*tdk*) / kanamycin resistance (Kan^R)counter-selectable cassette. Source: Gift from the laboratory of Prof. J. E. Barrick, University of Texas at Austin.
pCK227	Source of superfolder green fluorescent protein (*sfgfp*) gene, from Kelly *et al.*[227]
pJK001	*dapB*-targeting cassette in pGEM backbone. *Amp^R ColE1*
pJK002	*tdk*.Kan^R cassette from pYTK001-tdkkan inserted in *dapB*-targeting cassette of pJK001. *Amp^R Kan^R ColE1*
pRH001	Basic Artevo circuit cassette in pGEM backbone. Promoterless *dapB* and *gfp* between homology arms facilitating allelic exchange of *alkM* CDS in ADP1. *Amp^R ColE1*
pRH002p	Precursor plasmid for installation of various upgraded circuits to make pRH002-005.
pRH002	As pRH001, but containing an upgraded Artevo circuit with *sfgfp* in place of *gfp*, and the modified *alkRM* promoter decoupling expression of *dapB* and *sfgfp* from growth phase.
pRH003-006	As pRH002 but *alkR* is overexpressed by a Pro-series promoter from Davis *et al.*[226] Promoters and relative expression strengths are: ProD=1.00, Pro1=0.009, ProA=0.030, ProB=0.119. Also contains insulating terminator sequences.
pRH002 ArES-008/9	As pRH002 but *alkR* is overexpressed by a strong bi-cistronic design (BCD) promoter, BCD9, from the Biggs *et al.* toolkit[159].
pWH1274 *luxCDE:ADO*	Plasmid containing an aTc-inducible *luxCDE:ADO* operon for alkane biosynthesis. Source: previous work by this lab[127]. *Tet^R ColE1*
pTL2 seADO-AAR pTL2 pmADO-RAMO pTL2 CER1-ACR1	pTL2 plasmid constructs containing one of three alkane biosynthesis modules: seADO-AAR, pmADO-RAMO, or CER1-ACR. Used to test the response of the Artevo circuit to natively-produced alkanes. A gift from the laboratory of Prof. V. Santala at Tampere University. More information on these strains can be found in *Table 4*, in Chapter 5.

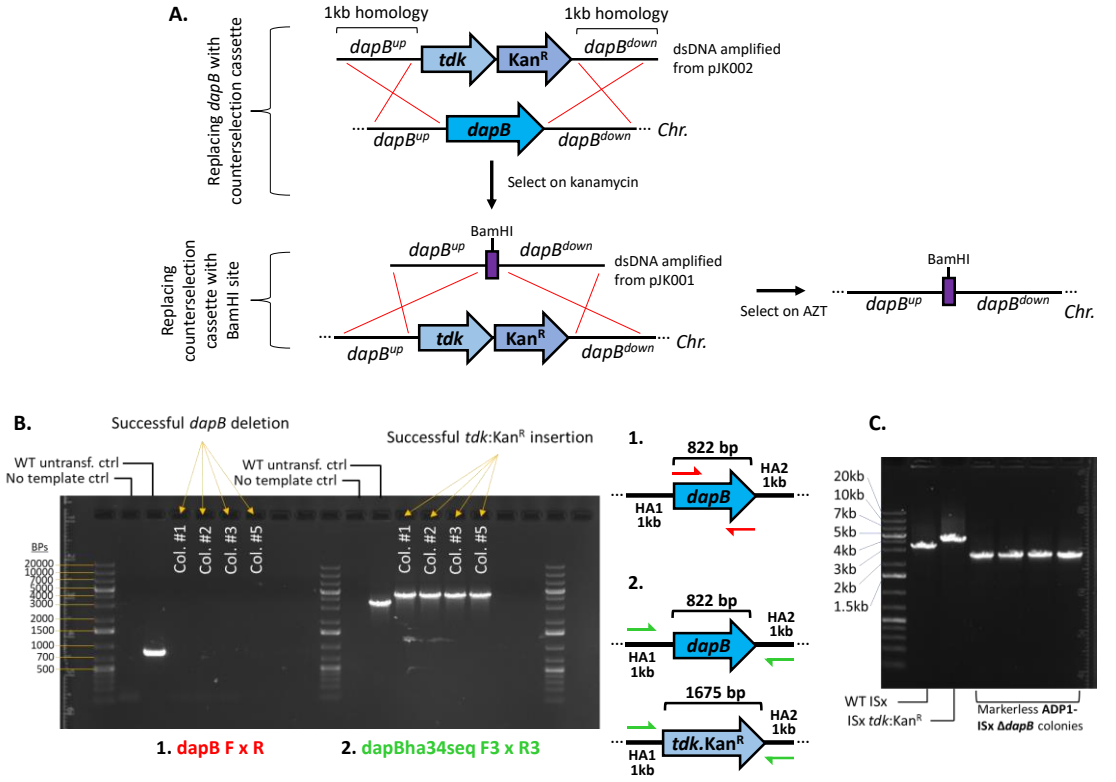


Figure 10. Cloning strategy and deletion of *dapB* gene from *Acinetobacter baylyi* ADP1. Part (a) illustrates the deletion strategy for *dapB*: natural homologous recombination of exogenous dsDNA replaces the *dapB* CDS with a *tdk*.Kan^R counterselection cassette. Selection on kanamycin eliminates unsuccessful transformants. Subsequent recombination with a second dsDNA cassette targeting the same locus replaces *tdk*.Kan^R with a BamHI restriction site. Plating on azidothymidine (AZT), or its analogue zidovudine (ZDV), selects transformants that have lost the *tdk*.Kan^R cassette. Part (b) shows a visualisation of colony PCR products used to screen for *dapB* deletion, using primers internal to the *dapB* gene sequence, and *tdk*.Kan^R cassette insertion. Part (c) Shows a colony PCR screening for successful removal of the *tdk*.Kan^R cassette and introduction of the BamHI site, creating a markerless ⊿*dapB* ADP1-ISx strain, which has been called ISxJK001 (or ADPJK001, for the transposon-containing counterpart).

To remove the *tdk*:Kan^R marker, colonies showing successful homologous recombination were transformed, using the same methodology as before, with the *dapB*-targeting cassette amplified from pJK001. Transformants exhibiting the ⊿*dapB* mutation and which had lost the *tdk*:Kan^R marker were selected by plating on LB agar supplemented with DAP, lysine, and 200 µg / mL zidovudine (ZDV) (a pharmaceutical version of azidothymidine, AZT). Presence of *tdk* converts zidovudine into a toxic intermediate that causes cell death. Marker loss was confirmed using colony PCR, as before. The resultant markerless ⊿*dapB* mutant strains obtained in both wild-type ADP1 and ADP1-ISx were named ADPJK001 and ISxJK001, respectively.

PCR amplicons of the *dapB* locus amplified from ⊿*dapB* ISx mutants (ISxJK001) were sequenced by Eurofins to confirm in-frame replacement of the *dapB* CDS with a BamHI site. One T>C SNP was identified ~100 bp upstream of the *dapB* deletion site

in a non-coding region. The sequence was verified in three replicates and the SNP occurs in all of them. The SNP occurs in a long run of thymine and may be an artefact of Taq polymerase amplification.

Finally, gDNA was extracted from the new ⊿*dapB* mutant and sent for genome sequencing with MicrobesNG in Birmingham. The assembled genome sequence confirms removal of *dapB* from the *A. baylyi* genome, with a BLAST search of the *dapB* sequence against sequenced genome producing no hits (see supplementary *Figure S.2*).

Characterisation of the *dapB* deletion mutant

dapB^{+} complement strains of ISxJK001, where *dapB* has been reintroduced into its original locus, were constructed to demonstrate that phenotypic changes observed in the knockout strains are due to the ⊿*dapB* mutation and are not process artefacts. Complements were obtained by transformation of ISxJK001 with the *dapB*^{up}-*dapB*-*dapB*^{down} region amplified from ADP1 gDNA. Transformants were selected by plating onto MMS agar and successful complements screened by colony PCR. 100% of screened colonies on MMS agar contained *dapB*, so *dapB* may be a useful alternative genetic marker for genetic engineering in a ⊿*dapB* background. *dapB*$^{+}$ complement strains were designated ISxJK001c.

Growth of ISxJK001, as well as ISxJK001c and wild-type (WT) ISx, was observed in various media conditions. Each strain was tested using three biological replicates (n=3). Overnight cultures of each strain were grown in LB and starting optical density at 600nm (OD600) was normalised by washing and resuspension in MM. Cells were then cultivated in a 96-well plate at 30°C for 24 hours, with OD600 measured every 10 minutes by a Tecan Spark plate reader. Each measurement was preceded by 20 seconds of double-orbital shaking. Each well contained either LB, unsupplemented MMS, MMS+lysine, MMS+DAP, or MMS+lysine+DAP, at the concentrations mentioned earlier. Wells were sealed with an oxygen-permeable membrane to allow aerobic growth while retaining sterility.

Supplementary *Figure S.3* shows that wild-type-like growth is restored in *dapB*$^{+}$ complement strains, both in LB and MMS, demonstrating that *dapB* may be used as a rescue gene in a biosensor-based directed evolution circuit. The ⊿*dapB* mutant ISxJK001 grows in LB media, though more slowly than wild-type or complement strains, potentially due to *meso*-DAP synthesis via *lysA* being slower than the *dapBDEF-argD* pathway. Wild-type and complement strains both grow in MMS without supplemental lysine or DAP, however ⊿*dapB* strains can only grow in MMS if lysine is

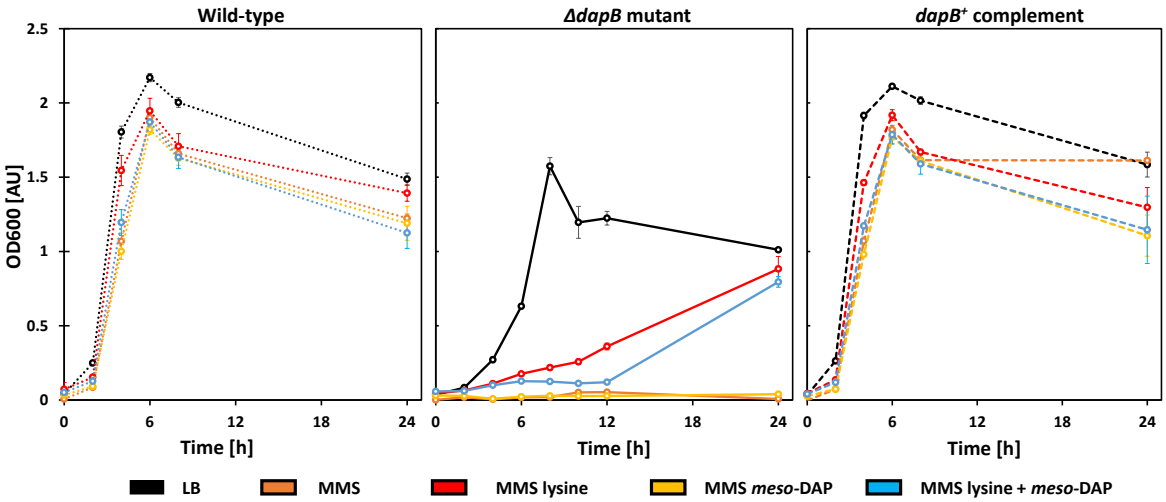


Figure 11. Characterisation of lysine auxotrophy in *dapB* deletion mutants of ADP1-ISx (designated as ISxJK001). Growth is restored in the presence of supplemental lysine (with or without *meso*-DAP) or when the *dapB* gene is reinstated (in a *dapB*$^+$ complement strain). As expected of a mutant strain with reduced capacity for cell wall synthesis, ISxJK001 exhibits slower growth compared to its wild-type and complement counterparts, even in rich (LB) media, which contains lysine.

present. Even if supplemental DAP is provided, ISxJK001 will not grow in MMS without lysine; possible reasons for this are discussed in the next section.

A repeat growth curve experiment was performed in 250mL culture flasks, rather than 96-well plate, to test strain behaviour at scale. This experiment, and all future experiments in this project that utilise 250mL culture flasks, used a working liquid volume of 50mL. The behaviour of some bacterial strains in 96-well plate may not be reflective of their growth at scale, because of difficulties in agitating such small (200µL) culture volumes by orbital shaking and in achieving effective aeration of the culture. This can cause biofilms or suspended aggregates to form in the wells, which may interfere unpredictably with the beam used to read optical density of the culture. Manual sampling of the same cultures, grown instead in culture flasks, is more labour intensive than using automated plate readers and comes at a loss of data resolution on the time axis, but often provides more accurate information about strain characteristics that may be ambiguous at smaller scales.

Using the same media conditions as those used in the earlier plate curve, sterile 250mL flasks were set up in triplicate and grown for 24 hours at 30°C and 150rpm. Flasks were inoculated from overnight pre-cultures of true biological replicates at an OD of 0.05. Flasks were sampled and OD measured every two hours for eight hours, then again at 24 hours. Resultant growth curves are shown in *Figure 11* and provide similar conclusions to the plate curve, albeit more pronounced.

Impact of varying supplemental lysine and DAP concentration

A second growth curve experiment, in 96-well plate, was performed as previously described to assess the impact of varying the concentration of supplemental lysine and DAP provided to ISxJK001. ISxJK001 cells were grown over 24 hours at varying concentrations of lysine, DAP, and a combination of lysine and DAP. Results in supplementary *Figure S.4* show no change in growth rate of ISxJK001 across the concentration ranges (Lysine: 25-100 μg/mL, DAP: 100-800 μg/mL) used in this experiment.

Discussion

Lysine auxotrophy in *dapB* deletion strains

∠I*dapB* mutants in ADP1-ISx cannot grow in minimal media without supplemental lysine, even if DAP is provided. This finding echoes results obtained by Gal *et al.*, who showed that a ∠I*dapB* mutant in *Pseudomonas fluorescens* was auxotrophic for both DAP and lysine, though they do not mention whether the mutant can grow in lysine alone[216]. It is postulated that DAP transport into *A. baylyi* ADP1-ISx is poor while lysine transport is good, so ∠I*dapB* mutants cannot utilise exogenous DAP and must instead obtain it by conversion from lysine via reversible action of diaminopimelate decarboxylase, as shown earlier in *Figure 9*. A previous study by Liu and Shaw generated a DAP auxotroph in *Pseudomonas syringae* and demonstrated that growth is restored in the presence of supplemental DAP alone; another study by Rediers *et al.* showed that DAP transport into *Pseudomonas stutzeri*, while present, appears inefficient whereas lysine is readily taken up by cells[215,228]. *E. coli* has also been shown to have the ability to transport DAP, with Allard *et al.* using DAP deprivation to select against a DAP-auxotrophic donor strain as a method of antibiotic-free post-conjugational counterselection[229]. It is not known why *A. baylyi* ADP1-ISx appears to lack the capacity for DAP transport; further investigation into the underlying mechanisms might provide wider insight into some metabolic features of ADP1-ISx. A *lysA* deletion mutant was obtained in ADP1-ISx, in the same manner described earlier for deletion of *dapB* from the chromosome, in pursuit of elucidating this transport mechanism. A ∠I*dapB* ∠I*lysA* double mutant in ADP1-ISx would also be required for this investigation but has not yet been obtained. This work, as it is tangential to the primary goal of the project, was set aside for the future.

The ⊿*dapB* mutant in *Acinetobacter baylyi* ADP1 was previously thought to be unobtainable based on the collection of single-gene mutants reported by Berardinis *et al.* in 2008[157]. Beyond its applications in this project, this mutant may aid in the discovery of new antitubercular drugs[220] and other antimicrobial agents[230], as well as facilitate research into *Acinetobacter baumannii*[231] – a multidrug-resistant pathogenic relative of *A. baylyi* that is often modelled using *A. baylyi* due to their extreme similarity.

Restoring *dapB* rescues growth in complement strains

Figure 11 shows restoration of wild-type-like growth when *dapB* is reintroduced to a ⊿*dapB* strain of *A. baylyi* ADP1-ISx, demonstrating the viability of this gene for use as a rescue gene in a directed evolution system. This also validated an underlying assumption made when identifying *dapB* on the *A. baylyi* ADP1-ISx genome using homology to *E. coli dapB*: that *dapB* in *A. baylyi* ADP1-ISx is functionally homologous to *dapB* in *E. coli*. Introducing *E. coli dapB* into the ⊿*dapB* strain of *A. baylyi* ADP1-ISx and observing growth restoration would conclude that these two homologues are functionally identical, but was considered unnecessary as the core function of *dapB* in the context of the Artevo system had already been established. It should also be noted, as it would become a thorny problem later in this work, that the simple conditions tested in the growth experiments described in this chapter – binary presence or absence of *dapB* – offer no insight into the activity of *dapB* across various expression levels. That is, whether graded expression of *dapB* would restore growth in a graded manner or a binary one. As we will see in the following chapters, achieving this graded response is central to the proper functioning of the Artevo system, and was a focus of much genetic tuning work. At this stage though, *dapB* had been validated as a useful rescue gene which, and could now be coupled to alkane production through the *alkRM* system, as will be described in Chapter 4.

Chapter 4

Building the evolution circuit

This chapter describes how the Artevo circuit, designs for which were shown in Chapter 3, was created. It follows assembly, installation, testing, upgrading, and tuning of the Artevo system, culminating in a proof-of-concept demonstration of cell fitness being linked to exogenous alkane concentration. The work is presented in two parts, the first with combined methods and results section followed by a discussion, the second with dedicated sections for materials and methods, results, and discussion. Part I details the construction and testing of prototype *A. baylyi* ADP1-ISx strains containing the Artevo circuit, and early upgrades made to the system in an attempt to reduce apparent leaky expression of the rescue gene, *dapB*. Part II focusses on concerted tuning of the Artevo circuit to achieve a graded growth response from host cells to exogenous alkanes in their environment. Some content in this second section is drawn directly from a paper on antisense regulation methods authored by myself, Dr Jack Chen, and Prof. Wei Huang, which is not yet published but currently available as a pre-print on bioRχiv[232].

Part I: Assembly and initial tuning

Assembly of the Artevo circuit

Obtaining an evolving strain for alkane biosynthesis is dependent on the reintroduction of a functional *dapB* gene under control of an alkane-inducible promoter, as shown in *Figure 12* and discussed in Chapter 3. The following sections outline the methods used to build this directed evolution circuitry and then install it on the genome of the markerless ⊿*dapB* mutant (ISxJK001) obtained in Chapter 3.

Cloning strategy

The plasmid pRH001 was constructed, containing a targeted exchange cassette to replace the entire CDS of alkane monooxygenase (*alkM*) with a promoterless copy of ADP1 *dapB* and a downstream reporter gene, *gfp* (*Figure 13*). This cassette consisted of the *dapB* gene followed by *gfp*, the pair flanked by 1kb regions of homology corresponding to those immediately up- and downstream of *alkM* on the ADP1-ISx genome, which includes the native *alkRM* promoter and a portion of *alkR*. The cassette was housed in a pGEM vector backbone which was amplified in *E. coli* DH5α. Fragments were amplified by overlap PCR using high-fidelity Q5 polymerase from NEB and primers from Sigma-Aldrich (now Merck), and then assembled via HiFi assembly using HiFi assembly mastermix, also from NEB.

The exchange cassette would replace *alkM* via natural homologous recombination in ADP1-ISx. This recombination would concomitantly remove host ability to metabolise alkane by removing *alkM*[233]. The *alkM* promoter is activated by AlkR protein in the presence of alkanes and has previously been exploited to express the *luxCDABE* bioluminescence operon in the biosensor strain ADPWH_alk[162].

Genomic installation of the Artevo circuit

The *dapB.gfp* exchange cassette was amplified by PCR from pRH001 and then used to transform ISxJK001 as described in Chapter 3 (page 46, in 'Deletion of *dapB* from the ADP1 genome'), except that transformed cells were recovered overnight in LB broth containing 0.01% (v/v) mineral oil and were washed and diluted in MMS before being plated onto MMS agar, also containing 0.01% (v/v) mineral oil. Selection of positive transformants, those which had taken up a functional copy of *dapB*, was stringent on

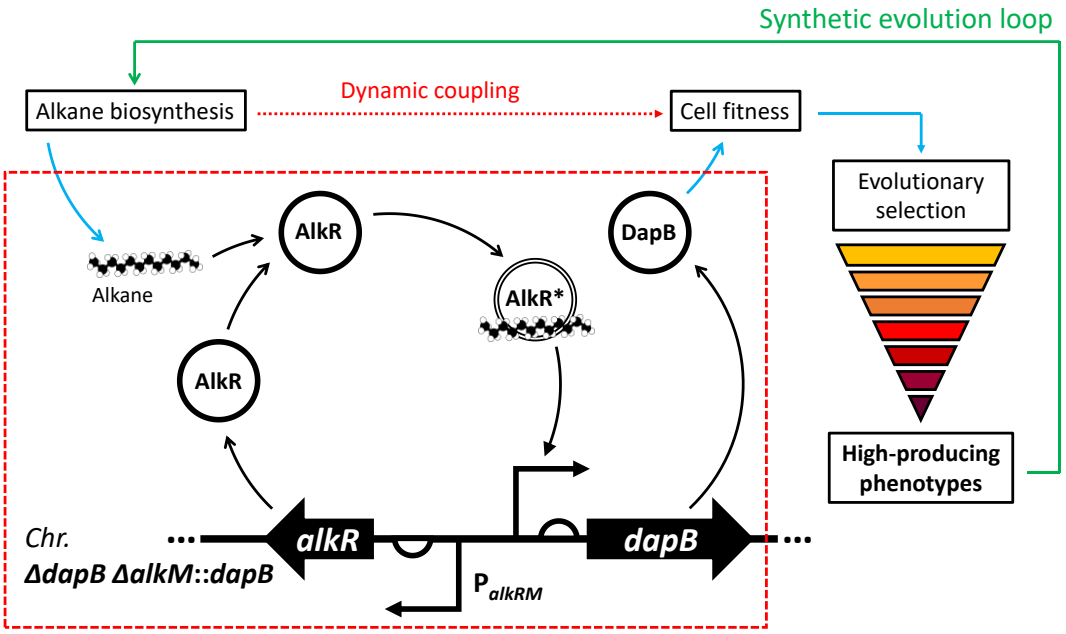


Figure 12. Artevo circuit (bounded in red) depicted with alkane biosensor gene (*alkR*) and rescue gene (*dapB*). This figure expands on the generic circuit structure depicted earlier in *Figure 6*.

MMS agar; 100% of colonies that appeared tested positive by colony PCR for homologous recombination of the Artevo circuit cassette at the *alkM* locus, and no colonies appeared in the absence of transforming DNA. *Figure 13* illustrates the cloning strategy and transformation results.

Digestion of colony PCR products from the above screen with XbaI cuts the amplicon in the intergenic region between *dapB* and *gfp*, producing a pair of fallout bands approximately 2 kb in length in the case of successful replacement of *alkM* with *dapB* and *gfp*. Successful transformants possessed a single copy of both *dapB* and *gfp* under control of the *alkM* promoter, P_{alkM}. This new strain, with genotype *⊿dapB ⊿alkM::dapB.gfp*, was named ISxJK002.

Streaking ISxJK002 onto MMS agar with and without 0.01% (v/v) mineral oil showed that the strain grew faster in the presence of mineral oil, suggesting that growth was being rescued by *dapB* expression in the presence of alkane, as per the intent of the circuit. Slower growth of ISxJK002 was observed on MMS-only agar plates, though the fact that any growth at all was observed suggested that the Artevo circuit was leaking and that *dapB* could be expressed in the absence of alkanes.

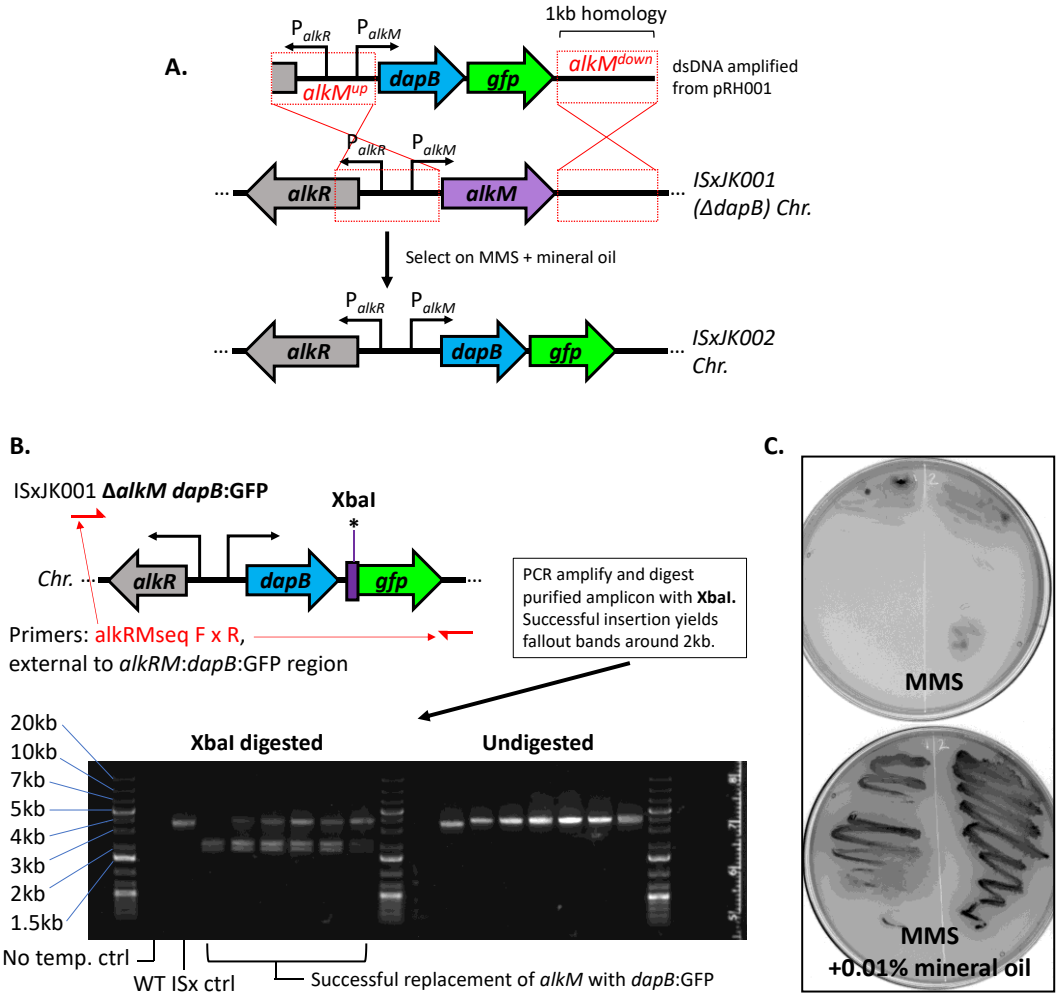


Figure 13. Installation of the basic Artevo circuit. Part (a) shows how homologous recombination with linear dsDNA exchanges the *alkM* CDS with *dapB* and *gfp* in markerless ⊿*dapB* mutants. Selection on MMS agar with mineral oil is stringent, as unsuccessful transformants lack a functional copy of *dapB*. Part (b) shows a colony PCR screen of transformants, highlighting the pair of fallout bands that indicate the presence of the *dapB gfp* module. Part (c) shows growth of ISxJK002 mutants on MMS agar with and without mineral oil, demonstrating that growth on solid media is more rapid in the presence of exogenous alkane.

Testing the early circuit

Growth (OD600) and fluorescence (GFP; excitation 395 nm, emission 509 nm, bandwidth 20 nm) of ISxJK002 in varying concentrations of mineral oil (0.001-0.1% v/v) was observed over 24 hours in a 96-well plate. Wells were inoculated in triplicate from overnight pre-cultures, which were grown in MMS+0.01% (v/v) mineral oil and washed in PBS prior to inoculation to remove oil and resuspended to normalise inoculum density.

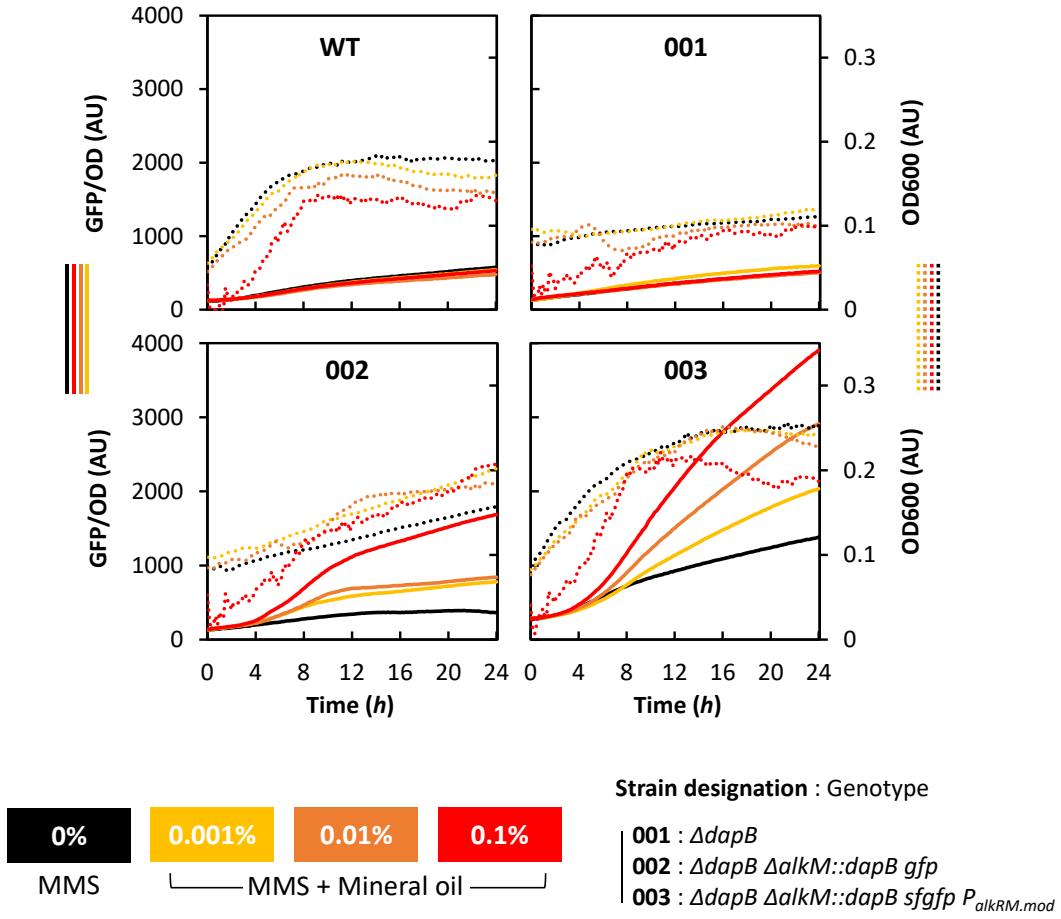


Figure 14. Comparison of growth (OD600) and specific fluorescence (GFP) of WT, 001, 002, and 003 strains (see *Table 2*), in varying concentrations of exogenous mineral oil (%, v/v). Specific fluorescence due to *gfp* expression (solid lines) is plotted on the left axis, OD600 (dotted lines) is plotted on the right axis. Data was gathered over 24 hours using a 96-well microtitre plate (n = 3).

Results, depicted in *Figure 14*, show specific fluorescence increasing with mineral oil concentration, demonstrating that *gfp* expression, and implicitly *dapB* expression, increases as extracellular alkane concentration rises. The rate of change of OD600 did not vary noticeably with mineral oil concentration, notably indicating growth even in MMS alone, and fluorescence was observed in 002 strains in MMS-only media. This supports previous qualitative observations on agar plates that suggest P_{alkM} is somewhat leaky, expressing *dapB* and *gfp* even in the absence of alkane. The 003 strain in *Figure 14* will be described in the following sections, but it can be noted as having similar growth characteristics to 002 but an improved spread of specific fluorescence curves.

As with characterisation of ⊿*dapB* mutants (001 strains) in Chapter 3, growth and fluorescence of ⊿*dapB* ⊿*alkM*::*dapB.gfp* strains (002 strains) was characterised in 250mL flasks as well as 96-well plate. Media conditions in the previous plate growth curve were replicated in flasks, which were manually sampled every two hours for the first ten

hours of growth, then again after 24 hours. *Figure 15* shows the resulting growth curve from this flask experiment. Growth rate of the 002 strain in flask again appeared binary, affected only by the presence or absence of mineral oil rather than by varying mineral oil concentrations, as in plate. A distinct graded response in specific fluorescence was observed though, with a maximum at 0.1% (v/v) mineral oil. Importantly, no errant growth was observed in the absence of mineral oil over the first ten hours of growth, suggesting apparent leakiness of the circuit seen in 96-well plate may be an artifact of those experimental conditions, rather than an issue with the circuit itself.

This conclusion was further cemented by fluorescent imaging (using a VersaDoc imaging system with filter conditions set to detect FITC) of ISxJK002 streaked onto MMS agar with and without mineral oil, as shown in part (b) of *Figure 15*. On solid media, growth of ISxJK002 in response to exogenous mineral oil appears tightly controlled over 24 hours, with colonies appearing in the absence of oil only on the second day of growth. These colonies did not fluoresce even when fully grown, compared with strong fluorescence observed in the presence of mineral oil.

Findings from these previous experiments suggest:

i. Expression of *dapB* and *gfp* from under the *alkM* promoter is proportional to exogenous alkane concentration.

ii. Expression from under the *alkM* promoter is leaky, and this leakiness is, or appears, pronounced in 96-well plate compared to 250mL flask growth.

iii. As the growth rate response of ISxJK002 to exogenous alkanes was binary, either full restoration of growth or none, the range of *dapB* expression over which meaningful, graded variation in growth rate may be observed is low. That is, DapB protein may have high activity, and a small amount of it is sufficient to restore growth to wild-type-like levels.

Sensitivity to different alkane species

Sensitivity of AlkR and the Artevo circuit to various hydrocarbons was tested by observing fluorescence of 002* strains in response to various chain lengths of alka(e)ne. Strains were grown in a 96 well plate over 24 hours in MMS with 0.01%, 0.001% or 0.0001% (v/v) of C12, C13, C14, C15, or C17 alkane, C12:1 alkene, or mineral oil. Fluorescence (GFP; excitation 395 nm, emission 509 nm, bandwidth 20 nm) was

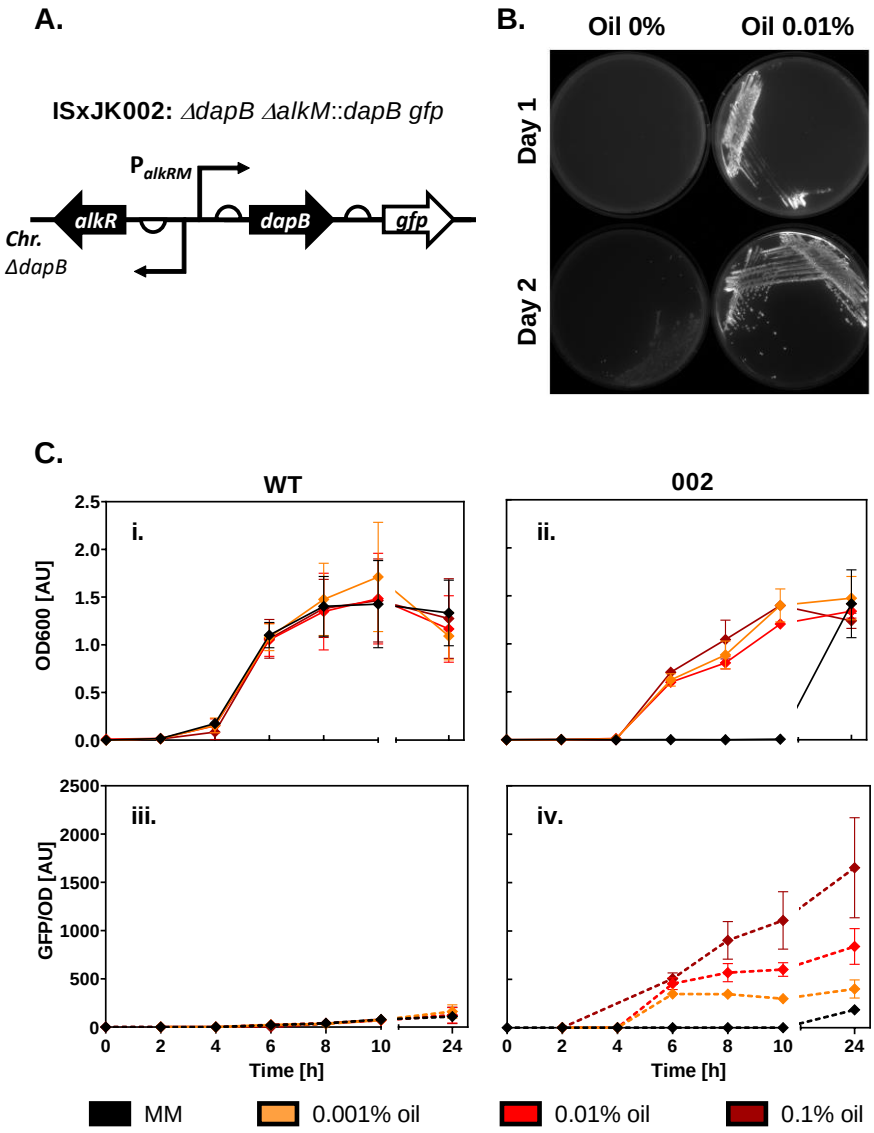


Figure 15. Part (a) shows the Artevo circuit structure for 002 strains (genotype: *ΔdapB ΔalkM::dapB gfp*). Part (b) shows the growth response of 002 strains to exogenous mineral oil on solid media: 002 strains grown on MMS agar with supplemental mineral oil fluoresce and grow faster than their non-oil counterparts. Part (c) shows the binary growth response of the untuned Artevo circuit to exogenous alkanes in liquid media, compared to the response of wild-type *A. baylyi* ADP1-ISx. Growth of 002 strains is either fully restored or limited (over the first ten hours of incubation) in the presence or absence of any mineral oil, respectively. The graded specific fluorescence response (part C iv.) indicates that expression of *dapB* and *gfp* from under the *alkRM* promoter is varying in proportion to exogenous alkane concentration, which implies that the entire dynamic range of *dapB* expression should be downshifted to elicit a useful graded growth response.

measured every 10 minutes as previously described. Results, shown in *Figure 16*, suggest that AlkR indiscriminately binds alkanes in the C12-C15 range but is more specific for heptadecane (C17) and for 1-dodecene (C12:1). The response of the system is also weaker at lower alkane concentrations; at 0.001% (v/v) alkane, C12-C15 alkanes no longer actuate a fluorescence response distinguishable from background media, and this is the case for 1-dodecene and heptadecane at 0.0001% (v/v). Adjusting specificity

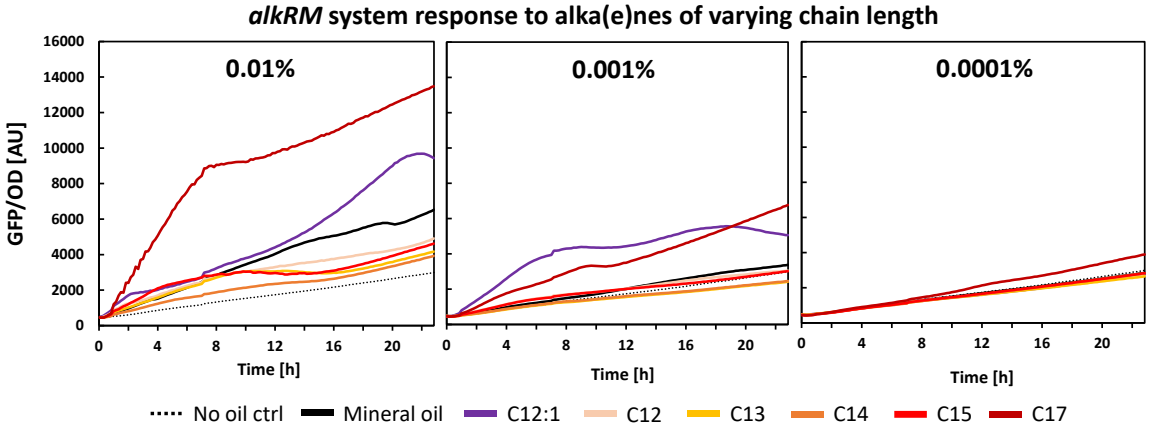


Figure 16. Response of .002* strains to alka(e)nes of varying chain length. Graphs show specific fluorescence (GFP/OD) signal from the Artevo circuit in response to varying concentrations (0.01%-0.0001%, v/v) of mineral oil, C12:1 alkene (1-dodecene), C12 alkane (dodecane), C13 alkane (tridecane), C14 alkane (tetradecane), C15 alkane (pentadecane), and C17 alkane (heptadecane). n = 3 biological replicates, error bars omitted for clarity.

of AlkR for different chain lengths may feature in future work of this project, to expand the potential pool of alkanes that can be produced for use as fuel blendstocks. AlkR could also potentially be engineered to detect branched, cyclic, and aromatic compounds which, combined with a suitable biosynthetic pathway, would enable the Artevo system to optimise production of diverse molecules useful in fuel blending.

Initial tuning and upgrading of the Artevo circuit

This section discusses early upgrades made to the Artevo circuit in an attempt to achieve a graded growth response from the host chassis in respect to exogenous mineral oil concentration. *Table 3* summarises the approaches taken or planned to tune the system, some of which are elaborated upon either below or in Part II of this chapter. For brevity, some unsuccessful approaches – specifically, direct chemical inhibition of DapB, translational interference via short antisense RNA, and mutation of DapB for activity reduction – are not described in further detail in this Chapter.

To summarise, decoupling the *alkRM* promoter from growth phase via introduction of three single nucleotide polymorphisms (SNPs), identified in earlier work by Zhang *et al.*[162], was the only approach to produce useful improvements to the system in the early stages of tuning. It was not until direct regulation of *dapB* via transcription interference was explored that the Artevo system began to function in a manner conducive to directed evolution. That is, with host growth rates being proportional to the concentration of exogenous alkanes.

Table 3. Summary of tuning strategies attempted during this project. Key strategies are highlighted in gold.

Strategy	Method	Results	Comments
Decoupling of *alkRM* expression from growth phase	Insertion of three SNPs in the *alk*R*M* promoter region known to decouple this promoter activity from growth phase[162].	Improved fluorescence response to exogenous mineral oil. No change in growth.	It should be noted that the *gfp* fluorescent reporter in .003 strains was replaced with *sfgfp*, a brighter analogue[234], for improved sensitivity of detection. It is the spread of the response curves across the mineral oil concentration gradient that is of interest however, rather than the absolute magnitude of the specific fluorescence observed.
Overexpression of *alkR*	Placed *alk*R under control of Pro-family promoters (ProD, ProA, ProB, Pro1) from Davis *et al.*[226]	No change in growth response.	Originally an attempt to reduce leaky expression of *dapB* by saturating P_{alkM} with AlkR protein unbound to alkane, which was thought to act as a repressor. This apparent leakiness was later found to be largely an artifact of culture scale, being far less prevalent in culture flask than in 96-well plate. AlkR, as an activator of P*alkM*, appears already to be quite effective.
Direct chemical inhibition of *dapB* activity	Chemical inhibitor of DapB, 2,6-PDC, added to culture media of ⊿*dapB* ⊿*alkM*::*dapB*.*gfp* strains.	No change in growth.	Scapin *et al.* reported direct inhibition of DapB *in vitro* by 2,6-pyridinedicarboxylate (PDC)[235]. *In vivo* response was not tested, so lack of inhibition suggested in this case may indicate ADP1-ISx does not have the capacity to transport 2,6-PDC.
Interference of *dapB* translation by short antisense RNA	Inducible transcription of 50bp of antisense *dapB* sequence from a plasmid.	No change in growth.	Intended to interfere with translation of *dapB* mRNA by binding of the short interfering sequence directly to the *dapB* transcript. Was unlikely to succeed in the first place as prokaryotic transcription and translation are typically simultaneous, but worth testing as a simple method for targeted control of gene expression.
Direct transcriptional interference of *dapB* expression	Convergently-oriented inducible promoter directly opposes *dapB* expression by RNA polymerase collision.	Tuneable reduction in expression of the controlled gene. Enabled a graded growth response to be achieved in the Artevo circuit via antisense regulation of *dapB* expression.	Effective strategy for adding a regulatory layer to biosensory gene circuits without manipulating the sense promoter. Functional at both plasmid and chromosomal levels.
Accelerated degradation of DapB protein	A C-terminal ssrA tag attached to DapB reduces protein half-life. The system enabling targeted ssrA-tag-mediated protein degradation has been extensively characterised in the literature[236–240].	Appears to be an effective strategy for clearing background levels of rescue gene and encouraging continual reliance on target metabolite for growth.	In the Artevo system, this strategy was postulated as a method of strengthening the link between alkanes and growth rate by limiting the persistence of rescue protein DapB in the system. Notably, this approach was only used in conjunction with direct transcriptional interference and would need to be investigated separately to fully characterise the benefits.
Direct activity reduction of DapB protein	Site-directed mutagenesis of DapB	Scapin *et al.* reports the following mutations in DapB and their consequent effect on its enzymatic activity[235]: H159A = 139-200 fold reduction K163A = 625-830 fold reduction	Not yet attempted. Antisense regulation was effective so protein engineering of DapB was not pursued, but remains an option for further system optimisation.

Decoupling rescue gene expression from growth phase

The luminescence response reported by Zhang *et al.* in their whole-cell ADP1 alkane bioreporter has excellent dynamics in response to exogenous alkanes. Notably, the *alkRM* promoter region in this strain contains three point mutations compared to wild-type ADP1, which appear to decouple this promoter activity from growth phase[162]. This feature was speculated to be useful in the context of the Artevo system, as decoupling rescue gene expression from growth phase would keep circuit behaviour constant across individuals in the population, possibly helping to reduce stochastic growth effects favouring suboptimal phenotypes.

002 strains (⊿*dapB* ⊿*alkM*::*dapB.gfp*) were upgraded first to include this modified promoter and to replace the *gfp* reporter with superfolder GFP (*sfgfp*), to improve the fluorescence signal[234]. The transformation cassette for this upgrade was assembled as previously described for the cassette used in construction of ISxJK002, with the modified *alkRM* promoter sequence ($P_{alkRM.mod}$) having been synthesised by Biomatik Corp. in the USA. Upgraded strains, designated '003' (ISxJK003, genotype: ⊿*dapB* ⊿*alkM*::*dapB.sfgfp* $P_{alkRM.mod}$), were transformed and selected as previously described for 002 strains.

The growth and fluorescence of ISxJK003 in response to exogenous alkane concentration was shown earlier in *Figure 14*, for comparison with ISxJK002. While the graded fluorescence response of ISxJK003 to exogenous alkanes is an improvement over ISxJK002, growth rate of ISxJK003 does not vary significantly across different alkane concentrations.

Overexpression of alkane biosensor *alkR*

Leakiness and dynamic range issues were also addressed by designing an upgraded Artevo circuit, shown in *Figure 17*. A strong, constitutive promoter drives production of AlkR protein, the inactive form of which is predicted to bind P_{alkM} and prevent transcription. Presence of alkane activates AlkR, forming AlkR*, which then competes with inactive AlkR to bind P_{alkM} and promote expression of *dapB*. Such a regulation mechanism appears in the SalR/P_{sal} system native to *A. baylyi* ADP1 and it was postulated that the AlkR/P_{alkM} system would behave similarly[241]. While not tested in this project, a planned experiment for investigating relative expression of *alkR* and *alkM* from under the *alkRM* promoter region is described supplementary *Figure S.9*.

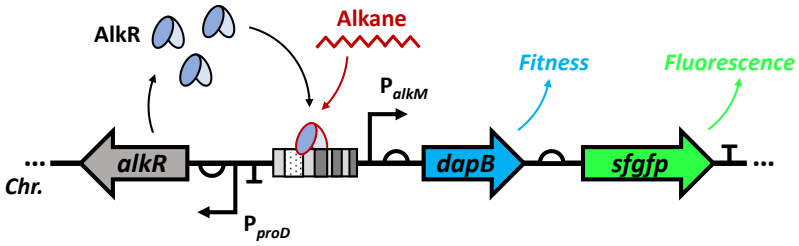


Figure 17. Insulated Artevo circuit structure with overexpression of *alk*R. Strong bi-direction terminators flank the modified *alkRM* promoter and *dapB sfgfp* cassette. Overexpression of *alk*R is facilitated by a strong constitutive promoter, either from the Pro series from Davis *et al.*[224] (promoters and relative strengths are: ProD=1.00, Pro1=0.009, ProA=0.030, ProB=0.119), or a bi-cistronic design promoter (BCD9) from the Biggs *et al.* toolkit[158]. Expression from under the modified *alkM* promoter is reported by superfolder green fluorescent protein (*sfgfp*). The *alkM* promoter was thought to be regulated by competitive binding of active (alkane-bound) and inactive (unbound) AlkR protein. High concentrations of inactive AlkR were theorised to suppress expression of *dapB* and *sfgfp* in the absence of alkane, reducing circuit leakiness.

Strong, bi-directional, terminator sequences were inserted in the intergenic region between *alk*R and *dapB* and also flanking the ends of the circuit, insulating each circuit element and preventing transcriptional read-through from upstream or downstream regions. The entire upgraded circuit module was synthesised by Biomatik Corp. and included a ProD promoter as the strong, constitutive promoter element[226]. The *alkRM* promoter region in this module included the three point mutations decoupling promoter activity from growth phase[162] mentioned in the previous section and used in 003 strains. The intergenic bi-directional terminator separating *alk*R and P_{ProD} from the *alkM* promoter region was drawn from the iGEM Registry of Standard Biological Parts[109] (part no. BBa_B0024). The terminator downstream of *sfgfp* was the rrnB T1 and T2 dual-region terminator, which is native to *E. coli* and broadly employed as an efficient terminator of transcription in various cloning vectors[242].

To test the effect of *alk*R expression under various promoters, a precursor plasmid backbone (pRH002p), which lacked promoter regions for *alk*R and *dapB.sfgfp*, was constructed by HiFi assembly as previous described for pRH001. pRH002 was constructed by reintroducing the modified *alkRM* promoter, containing three point mutations decoupling expression of genes under this promoter from growth phase[162]. The insulated Artevo circuit containing P_{proD} was cloned into pRH002p using BamHI and EcoRI restriction sites to produce pRH003, which had *alk*R under control of P_{proD}. The insulated gene module from Biomatik was also subjected to Q5 site-directed mutagenesis to change the sequence of the -35 region of P_{proD}, producing three additional modules with different *alk*R promoters of varying strength. Relative promoter strength is given in *Table 2* and was originally established in *E. coli* by Davis *et al.*[226] Each module was cloned into a pRH002p backbone as described before to produce pRH004-6.

Upgraded Artevo circuit cassettes were PCR amplified from their respective vectors and the dsDNA amplicons used to transform 001 strains (the original ⊿*dapB* knockout mutants) as previously described. Positive transformants were selected by plating onto MMS agar with 0.01% (v/v) mineral oil and screened by colony PCR.

To test the growth response of these new strains, a 96-well plate experiment was set up as previously described to monitor growth (OD600) and fluorescence (GFP) of each strain across varying concentrations of exogenous mineral oil. Biological triplicates (n=3) of each strain were grown overnight in LB media at 30°C and 200 rpm. Cells were washed once in MMS broth, OD600 was normalised, and growth curves established across a mineral oil gradient from 0.001% - 0.1% (v/v). Results, which do not indicate any improvement in growth dynamics using these upgraded Artevo circuits, are shown in supplementary *Figure S.11*.

Using a bicistronic design (BCD) expression system

Continuing the investigation into overexpression of *alkR* toward improved performance of the Artevo circuit, it was decided to test a bicistronic design (BCD) expression system that was developed in *E. coli* but has recently been validated in *A. baylyi* ADP1[159,243]. BCD promoters contain two ribosome binding sites (RBSs), the first of which initiates translation of a leader peptide. The final nucleotide of the STOP codon of this leader peptide overlaps the start codon of the gene being controlled by the BCD promoter. This promoter architecture is thought to leverage helicase activity of the translating ribosome complex to stabilise mRNA secondary structure, and consequently improve expression of the gene-of-interest[243].

As their orthogonality was unknown, it was suspected that the Pro series of promoters may have different performance in *A. baylyi* compared to *E. coli*, perhaps even being entirely unsuitable for use in the former chassis. Contrastingly, the functionality of BCD promoters in *A. baylyi* had been demonstrated and shown to produce strong expression of a gene-of-interest.

A gene cassette similar to that shown in *Figure 17*, except with the ProD promoter replaced with a BCD9 promoter as described by Biggs *et al.*[159], was synthesised by TWIST Bioscience in the USA and installed between targeting homology arms on existing pRH002 plasmid architecture, by HiFi assembly as previously described. The plasmid map can be seen in supplementary *Figure S.12*. The transforming cassette was amplified and purified as previously described, and used to transform ISxJK001, as described earlier for cassettes containing the ProD promoter. Again, as with the Pro series of promoters, growth and specific fluorescence curves were constructed for this

strain in 96-well plate, and the results can be seen in supplementary *Figure S.13*. Similar to the Pro series promoters, no improvements to growth dynamics were observed when using the BCD promoter to overexpress *alk*R.

Investigation of a high-performing singlicate

Surprisingly, 002 strains, those chassis equipped with the original, uninsulated Artevo circuit, had so far exhibited the most ideal growth behaviour in response to alkanes. This is observable in *Figure 14*, though not abundantly clear unless considered alongside statistical analysis. An analysis of variance (ANOVA) performed on triplicates of ISxJK002 identified that there was a significant difference in growth rate between groups grown under different conditions. A Tukey HSD post-hoc test was then performed to identify from which groups this significance arose, revealing a significant growth response in the presence of mineral oil at 0.001% ($p=0.010$) and 0.1% (0.005) (v/v) compared to MMS without mineral oil, at $t = 24$ hours. Growth on mineral oil at 0.01% (v/v) was not significant at $t = 24$ hours but may have been significant at an earlier time point. See supplementary *Figure S.14* for details.

While fluorescence signal varied with mineral oil concentration for both 002 and 003 strains, with the best dynamic range being observed in 003, growth appeared similar regardless of the amount of oil added (*Figure 14*). A single replicate of 002 used in this experiment seemed to show a more pronounced growth response in the presence of alkane than other replicates. This singlicate was isolated and cultured in 100 mL of MMS with and without 0.01% v/v mineral oil, in glass 250 mL flasks, and exhibited a distinct qualitative difference in growth over 24 hours. The cell mass, following centrifugation of culture volume, after 24 hours of growth in MMS+0.01% (v/v) mineral oil was far greater than that observed in MMS alone; see supplementary *Figure S.15*. This high-performing 002 strain was denoted as 002*.

Sequencing of the *alkRM* region of this high-performing 002* singlicate, compared to the two other 002 replicates, revealed a G→A point mutation at nucleotide 208 in the *alk*R CDS in 002 strains. This causes a nonconservative G70R missense mutation that may be a contributor to the perceived reduced performance of these strains compared to 002*, which retains the native *alk*R sequence. Supplementary *Figure S.16* shows the sequence and alignments in more detail.

Effect of primary inoculum on subsequent growth

Despite promising qualitative results and fluorescent signals suggesting expression from under $P_{alkRM.mod}$ varies with alkane concentration, variability of growth in 002 and

003 strains was high in growth curves performed in 96-well plates. This may be because positive transformants confirmed by PCR are not screened for behaviour when grown on alkanes, only for insertion of the relevant gene modules. A high-throughput screening method to isolate high-performing colonies from an otherwise broadly isogenic population may prove useful in overcoming this limitation, albeit time and resource intensive to establish.

Restreaking the high-performing isolate and its regular counterpart – 002* and 002 strains, respectively – from glycerol stocks revealed a difference in growth between cells originally grown on MMS and cells grown on MMS with mineral oil; see supplementary *Figure S.17*. Repeated cultivation on media without a source of alkane or lysine to rescue growth encourages leaky expression from the Artevo circuit, as cells able to access *dapB* in the absence of mineral oil are positively selected under these conditions.

These results are counterintuitive to what might be expected based on previous studies, which have observed *E. coli* accumulating *meso*-DAP for later use in cell-wall construction[244]. It would be reasonable to expect that culture conditions of primary inoculum impact perceived leakiness by contributing to intracellular accumulation of either alkanes, DapB protein, or *meso*-DAP; prior growth in mineral oil activates expression of *dapB* and may allow cells to accumulate an intracellular reservoir of DapB protein or *meso*-DAP, which can later be tapped to support growth when conditions change.

The primary inoculum for pre-culturing of 002* is MMS or LB broth plus 0.01% (v/v) mineral oil. Due to the ability of ADP1 to emulsify oil droplets, it was postulated that some of this oil may be retained during washes and promote apparent leakiness in secondary growth conditions.

To test this, ISxJK002 was grown in varying pre-culture medias and used to inoculate secondary media with and without mineral oil. Primary inoculum was either LB broth or MMS, both with or without 0.01% (v/v) mineral oil. Secondary culture media was MMS with or without 0.01% (v/v) mineral oil. Primary cultures were washed twice by repeated centrifugation and resuspension in PBS, with the final suspension being used to normalise OD of the inoculums and ensure consistent starting ODs of secondary cultures. Culture vessels were 50mL Falcon tubes (10mL culture volume) incubated horizontally on an orbital shaking platform, such that shaking conditions approximated wave mixing. Culture conditions were otherwise as previously described.

Supplementary *Figure S.10* illustrates that primary inoculum does impact secondary growth characteristics in .002 strains, with the presence of oil being the most important

factor and the richness of media being the second. That is, strains pre-cultured in rich media with oil exhibit the most rapid secondary growth, but strains pre-cultured in minimal media with oil exhibit more rapid secondary growth than strains pre-cultured in rich media without oil. It is not known whether this phenomenon is attributable to carryover of oil droplets adhering to the cell surface of ADP1-ISx or accumulation of an intracellular DAP reservoir during primary growth. If we assume that expression of the rescue gene *dapB* in the presence of oil enables host cells to generate a similar amount of DAP as they would obtain in rich media – the latter through intake of lysine present in LB media and reversible action of *lysA* to produce DAP – then these results suggest the dominant mechanism is oil retention. This persists despite multiple washes and resuspension in PBS prior to secondary inoculation. This effect seems counterintuitive as the oil reserve is extracellular, with droplets exposed on the surface of the cell and so presumably more susceptible to washing than intracellular DAP. Further investigation, including a direct measurement of intracellular DAP concentration, would be required to fully understand the mechanism involved, but this experiment highlights the importance of consistent primary culture conditions for consistent analysis of subsequent growth.

Impact of culture vessel on apparent circuit leakiness

The increased aerobicity and culture volume of a 250mL glass Erlenmeyer flask (or conical flask) is suspected to be the cause of reduced variation in OD and fluorescence, or reduced apparent leakiness of the Artevo circuit, observed in previous growth experiments performed using 96-well plates. This reduction in leakiness is evident when comparing *Figure 14* and *Figure 15*, shown earlier in this chapter, the latter of which highlights the binary growth response of 002* strains to exogenous alkanes. Reduced leakiness in larger culture volumes may be due to the improved agitation achievable in shake flasks compared to microtitre plates. *A. baylyi* grown in suspension culture sometimes aggregates, possibly to create a favourable microenvironment for growth in comparatively stressful conditions. Formation of these aggregates is favoured in microtitre plate, where limited agitation allows cells to settle, but constant shaking in conical flasks ensures a homogenous cell suspension and reduces errant growth of unproductive phenotypes, originally though to be caused by leaky *dapB* expression.

While considered far less likely, it is also possible that culture vessels made of plastics may secrete trace hydrocarbons into cell cultures and activate AlkR, causing some background leakiness of the Artevo circuit; this phenomenon has previously been reported in the literature[245,246]. Some portion of the basal expression of *dapB* from under

$P_{alkRM.mod}$ may be due to cross sensitivity of AlkR for the long hydrocarbon tails of other intracellular molecules, though given the minimal leakiness observed in Zhang *et al.'s* bioreporter strain, ADPWH_alk, this effect, if present at all, is likely inconsequential. Work by Lehtinen *et al.* using a twin-layer aldehyde/alkane biosensor in *A. baylyi* confirms that AlkR is not cross-sensitive for native metabolites, and appears also not to be cross-sensitive for aldehyde intermediates of the ADO-FAR pathway[247,248].

It was noticed early in this project that a fine, white precipitate would settle out of minimal media when left without agitation, and this was impacting the smoothness of OD data obtained in 96-well plates. This precipitate is thought to be calcium from the saturated $CaCl_2$ added during media preparation. Even under constant agitation, this precipitate would settle in the centre of a 200µL culture well and obscure the true OD of the sample. Settling of particulates – and aggregation of cells, as mentioned earlier – is a general drawback of small-volume culture vessels agitated by orbital shaking[249]. The precipitation issue was solved by selectively transferring the supernatant phase of MM stock to a fresh bottle following settling of precipitate, though differences in cell growth between flask and microtitre plate can only be addressed by experimental scale up.

Before the experiment in *Figure 15* was performed, to establish whether growth of 002 strains is influenced by culture vessel, the 002* high-performing singlicate was grown for 24 hours in flasks (30°C, 150 rpm) as previously described. Supplementary *Figure S.15* provides more details, and shows a clear increase in growth rate of 002* in the presence of both mineral oil and heptadecane when compared to the MMS-only control. Growth in MMS without alkane after 24 hours indicates leaky expression of *dapB*, as expected, though this effect is far less pronounced than that observed in microtitre plate experiments.

Growth in 250mL culture flasks appears to improve the behaviour of the Artevo system, reducing apparent leaky expression of *dapB*, however experimental workload is greatly increased and manual sampling does involve a loss of data resolution on the time axis. Sophisticated chemostats or their smaller laboratory analogues, such as the Chi.BIO reactors designed and built by our colleague Prof. Harrison Steel[250], allow for continuous mixing and high-resolution sampling of growing culture. Growth of Artevo strains was trialled in Chi.BIO reactors, as shown in supplementary *Figure S.18*, and this experiment did demonstrate increased growth rate of 002* strains in high-oil conditions compared to low-oil conditions. Concerns about separation of the hydrocarbon phase in the media storage vessel, potentially causing fluctuations in bulk hydrocarbon concentration in the reactor, were raised and will need to be addressed before using the system for further experiment. Mechanical strategies for maintaining immiscible phases

in chemostats do exist, but may be challenging to implement in automated, low-cost benchtop reactor systems[251].

Intracellular versus extracellular alkanes

In Chapter 3, it was assumed that 'cheater phenotypes' within an evolving population – unproductive cells that nevertheless utilise metabolite-of-interest produced by other cells in the population to obtain a fitness output for themselves – would, in the case of the Artevo system, be disadvantaged compared to productive phenotypes when alkanes are produced natively by host machinery. That is, it was postulated that production of intracellular alkanes would bestow a greater fitness advantages than that which might be gained using extracellular alkanes.

As an early test of this hypothesis, to check whether natively-produced alkanes could provide a growth advantage to cells hosting the Artevo circuit, an alkane biosynthesis plasmid was constructed and installed by natural transformation in ISxJK002*. The plasmid, called pWH1274 *lux*, was assembled as previously described by PCR amplification of fragments and HiFi assembly. pWH1274 *lux* contained the cyanobacterial *luxCDE* gene cluster and aldehyde deformylating oxygenase (ADO), both from *Synechococcus elongatus*, on a modified pWH1274 backbone with a kanamycin resistance marker[252]. *luxCDE:ADO* produces alkanes, predominantly C15 alkanes, according to the chemical pathway shown in *Figure 18*, as well as fatty aldehyde intermediates. Expression of the *luxCDE:ADO* genes was regulated by an anhydrotetracycline (aTc)-inducible *tet*R system. Empty pWH1274, with kanamycin resistance but no *luxCDE:ADO* cargo, was used as a control.

ISxJK002* and wild-type ADP1-ISx strains, each containing either pWH1274 *lux* or empty pWH1274, were grown in MMS media with and without supplemental mineral oil and containing aTc (10 nM). Both solid and liquid medias were tested. On solid media, after one week of growth, 002* pWH1274 *lux* strains fluoresced in the absence of supplemental alkanes (supplementary *Figure S.8*) suggesting alkane-induced expression of the *dapB.gfp* operon under P_{alkRM} was being driven by natively-produced intracellular alkanes. In liquid media, this same strain was able to grow – albeit minimally and with aggregate growth suggesting stressed cells – without supplemental alkanes, whereas its empty pWH1274 counterpart could not.

This result sparked a broad, but eventually fruitless, experimental campaign to detect alkanes produced in this system using the existing ADPWH_alk luminescent biosensor strain[162]. Raman analysis was also unable to detect alkanes in these strains, so it is unknown whether the observed growth and fluorescence in minimal media is indeed

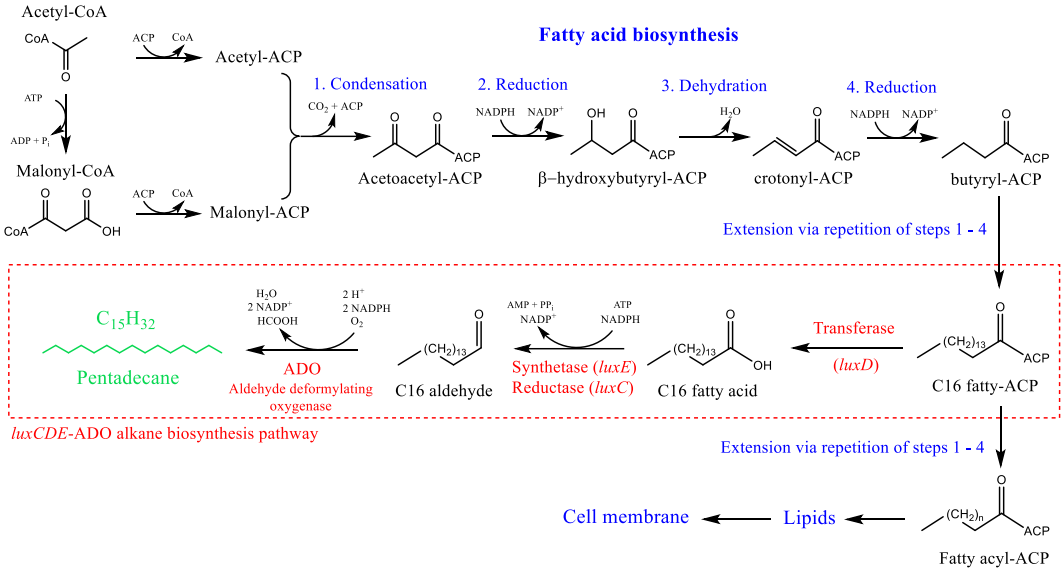


Figure 18. Alkane biosynthesis mediated by *luxCDE* and *ADO*. Blue reactions indicate native fatty acid biosynthesis in *A. baylyi* ADP1. Red reactions have been introduced using cyanobacterial *luxCDE* and *ADO* genes, which utilise C16 or C18 fatty-ACP to produce penta- or heptadecane[126]. The greater metabolic context of these pathways is illustrated in *Figure 4*.

driven by alkanes produced via *luxCDE:ADO*. Supplementary *Figure S.19* and *Figure S.20* provide more detail, but these results are inconclusive and tangential to the core results of this project. By the time tuning of 002* strains had been completed, which we will see in Part II of this Chapter, another cyanobacterial AAR-ADO pathway – which directly converts fatty acyl-ACP into hydrocarbon, avoiding formation of toxic aldehyde intermediates – had become the pathway of choice for alkane biosynthesis. This prompted the use of new, more reliable and better characterised plasmids for demonstrating the capabilities of the Artevo system.

Part I. Discussion

The novel *A. baylyi ⊿dapB* mutant, obtained in Chapter 3, was equipped with the first iteration of the Artevo circuit and a high-performing strain, denoted 002*, was isolated. Early results showed that 002* strains could distinguish between increasing concentrations of mineral oil in minimal growth media, but leaky expression of the rescue gene *dapB* from under the *alkRM* promoter (P_{alkRM}) was able to rescue growth even in the absence of oil. This leakiness was found to be largely an artifact of growth in microtitre plate and is greatly reduced when cultures are grown in 250mL flasks. At this scale, the growth response of 002* strains to alkanes is binary; growth is either minimal or fully restored in the presence or absence of alkanes, respectively. Graded expression of the rescue gene *dapB* was implicit, evidenced by specific fluorescence –

caused by expression of the *gfp* reporter immediately downstream of *dapB* – exhibiting a graded response proportional to exogenous alkane concentration. This suggests that the activity of DapB protein is high and the range of *dapB* expression over which meaningful variation in growth may be observed is low. While this work did not directly observe varying *dapB* expression levels, for example by tagging DapB protein and performing a Western Blot analysis, or by RT-qPCR targeting *dapB* transcripts, the variation in *dapB* expression implied by the response of the *gfp* reporter was considered sufficient to proceed to the next stage of the project. Later, an attempt at directly coupling fluorescence to *dapB* expression would be made by fusing the C-terminus of DapB to the N-terminus of GFP, but this fusion protein did not produce any detectable signal and the work is not detailed further in this thesis.

Initial promoter tuning efforts undertaken to adjust expression of *alkR* from under P_{alkRM} were unsuccessful: strong ProD promoters initially installed to control *alkR* did not reduce leaky expression of *dapB* from P_{alkRM}, but the activity of these promoters in *A. baylyi* ADP1 had not yet been demonstrated. The pre-print publication in 2019, by Biggs *et al.*, of a promoter toolkit for ADP1[159] inspired the use of a strong bicistronic design (BCD) promoter to overexpress *alkR*, but this also had no effect on the performance of 002* strains. Supplementary *Figure S.13* shows the increase in specific fluorescence expected from exchanging *gfp* in the 002 Artevo circuit for *sfgfp* in the upgraded circuit, but not the graded growth response to exogenous mineral oil hoped for after installing the strong BCD9 promoter for overexpression of *alkR*. Without having observed *alkR* overexpression directly, through SDS-PAGE or RT-qPCR for *alkR* transcripts, the relationship between AlkR protein prevalence and control of *dapB* cannot be concluded explicitly, but it is evident that strong promoters controlling *alkR* have no significant effect on growth of the host strain in response to exogenous mineral oil.

This suggests that overexpression of *alkR* is not an effective strategy for tuning the Artevo circuit, but could be further investigated in future work by deleting *alkR* from the 002* genome and introducing a plasmid where *alkR* is under control of an inducible promoter. In this way, AlkR expression can be adjusted by varying inducer concentration and the effect of AlkR on regulating leakiness of the Artevo circuit can be assessed. Further to this, an experiment to investigate the bi-directionality of P_{alkRM} and the mechanism of AlkR regulation of this promoter has been designed and is described in supplementary *Figure S.9*. The experiment replaces *alkR* and *alkM* on the ADP1 chromosome with red and green fluorescent reporters, respectively, with *alkR* expression instead being inducible from the aforementioned separate plasmid, and observes the relative fluorescence as *alkR* expression is varied. Insight from that

experiment should help future projects develop improved promoter tuning strategies regarding P_{alkRM}.

The Artevo circuit regulator protein, AlkR, has been shown to be able to distinguish between increasing concentrations of mineral oil and between straight-chain alkanes of varying length. It has not yet been established whether AlkR can effectively regulate the Artevo circuit in response to branched, cyclic, or aromatic hydrocarbons, which constitute several valuable blendstocks for products like aviation kerosene. Reed, Blazeck, and Alper demonstrated that the AlkSp transcriptional regulator from *Pseudomonas putida* could be evolved by successive rounds of error-prone PCR and fluorescence activated cell sorting (FACS) to exhibit increased sensitivity to short-chain alkanes such as hexane and pentane[253]. A similar approach might be taken to evolve AlkR and generate a versatile suite of sensors for use in the Artevo system.

Additionally, the half-life of the rescue protein, DapB, and the accumulation of alkanes or DAP in primary inoculum should be considered. As it has been shown that cultivation without mineral oil appears to encourage leaky expression of *dapB*, there may be some optimisation required regarding growth of primary inoculum, such as cultivating initially without mineral oil but limiting cultures to a single passage, and not allowing cells to remain in stationary phase for too long. Background clearing of DapB protein by targeted degradation – for example via addition of an ssrA tag such as AANDENYALVA, which has been shown to increase protein turnover rate[238] – should also reduce apparent leakiness in the system and create greater reliance on alkanes to induce continued expression of rescue gene. This strategy is employed in Part II of this Chapter, which begins below.

Part II: Fine-tuning the Artevo circuit

Transcriptional interference for control of gene expression

Transcription is a dynamic process affected by environmental factors and cellular context, though synthetic gene circuits commonly appear as simple gene blocks regulated by an inducible promoter. Such circuits are typically fit for purpose and avoid transcriptional interference between promoter elements, but offer only one handle for controlling gene expression. Previous transcriptomic studies have revealed antisense transcription to be a widespread regulatory mechanism on bacterial genomes[254–256], though synthetic gene circuits incorporating this mechanism are uncommon. Inducible antisense promoters represent a convenient, reliable handle for tuning gene expression without affecting the behaviour of the primary promoter or requiring laborious promoter engineering. Strategies for modulating gene expression by transcriptional interference were explored to adjust the expression of *dapB*, and consequently the growth response of the Artevo system, to alkanes. The goal was to produce a graded rather than binary growth response to exogenous alkanes, and simultaneously equip the Artevo system with a handle for adjustable stringency of selection; that is, adjustable thresholds of selection pressure which a given phenotype must overcome in order to proliferate. *Figure 19* provides an overview of the strategy, illustrating the principle of transcriptional interference and updating *Figure 6*, which was shown in Chapter 3.

The mechanisms of transcriptional interference have been well described by Shearwin *et al.*[257] and include: (i) 'roadblock'; the classical mechanism of inducible gene regulation, where a repressor protein binds near a promoter and directly blocks transcription machinery, (ii) 'competition'; where opposing promoters overlap and RNA polymerases (RNAPs) compete to bind, (iii) 'sitting duck'; a slow-processing RNAP is dislodged by a more highly processive RNAP complex, (iv) 'occlusion'; transient blocking of a promoter site by an RNAP elongating over it, and (v) 'collision'; two convergently transcribing RNAP complexes collide and are dislodged. The primary mechanism of transcriptional control explored in this work is RNAP collision, which is thought to be the dominant mechanism of transcriptional interference when convergently-oriented promoters are separated by more than 200bp[258]. Though the term 'collision' is commonly used and will be used here, studies of RNAP interactions by atomic force microscopy did not observe direct contact of actively transcribing RNAPs, instead postulating that the interference mechanism may be mediated by DNA supercoiling[259]. RNAP collision does not preclude translational interference by antisense RNA (asRNA), which binds to and disrupts translation of sense mRNA to

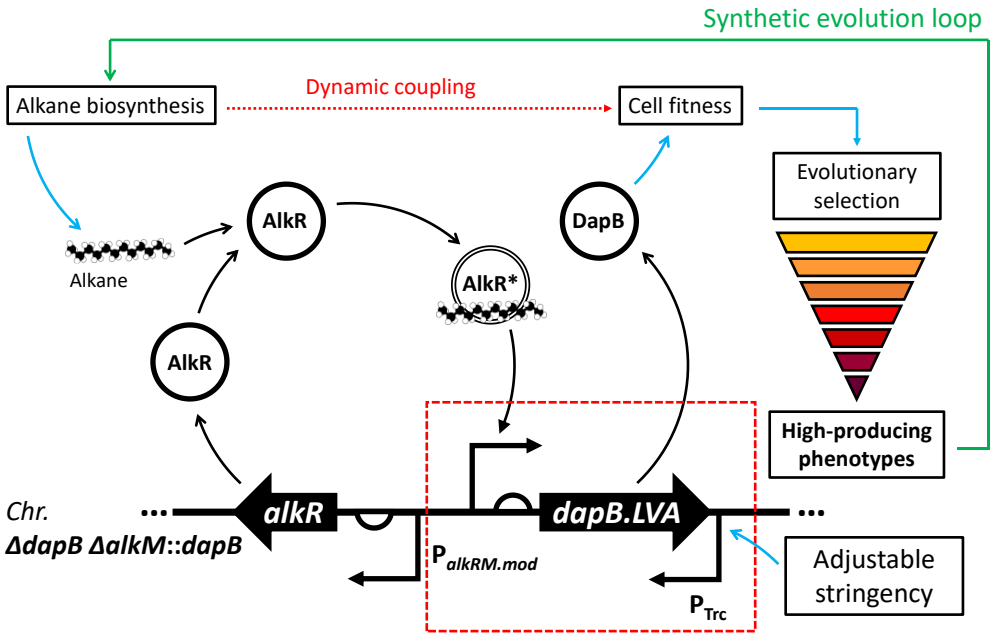


Figure 19. Tuneable stringency of selection within the Artevo system using direct transcriptional interference with *dapB* expression. The relevant system architecture is bounded in red. *dapB* expression is opposed by a convergently-oriented inducible promoter; RNAPs fired from this promoter collide with transcribing RNAP-ribosome complexes proceeding in the sense direction and potentially dislodge them, downregulating *dapB* without modifying the function of the *alkRM* biosensory circuit. An 'LVA' ssrA targeted degradation tag has also been attached to *dapB* to reduce the half-life of DapB protein, clearing it rapidly from the host and preventing accumulation of a growth-restoring DapB pool. This is theorised to increase host reliance on continual alkane production and maintain high selection pressure.

modulate gene expression, and has been shown to act in conjunction with direct transcriptional interference[260,261].

The tuneability of gene circuitry incorporating antisense promoters has been previously explored by varying relative promoter strength, interpromoter distance, and sequence content[262,263]. More recently, a 2019 study by Hoffman *et al.* demonstrated that increased RNAP processivity reduces the likelihood of transcriptional termination on impact with a convergent RNAP[264]. Subsequent work by O'Connor, Bordoy, and Chatterjee expanded on this and explored antitermination strategies that increase processivity of convergent RNAPs to better downregulate a gene-of-interest[265].

Downregulation of a gene-of-interest via collision of convergent RNAPs can be achieved either by direct action on the gene itself or indirect action on a regulatory gene[263]. Genetic engineers may choose to position antisense promoters to interfere with active transcription of a gene-of-interest or to interfere with transcription of activators required for expression of said gene. The dynamics of both systems have been explored, though the former is advantageous in situations where the response function

of the primary promoter must be preserved. In the case of the *alkRM* operon used in this project, transcriptional interference acting on the regulatory gene *alk*R may affect the biosensory capabilities of the system, so direct interference with the rescue gene under control of the *alkM* promoter is preferable. Additionally, due to the divergent orientation of *alk*R and *alkM*, a convergent promoter interfering with *alk*R would read through into *alkM*, necessitating further gene editing to insulate the *alk*R and *alkM* promoters from one another with terminators or changes in gene orientation. Preserving as much of the native genome as possible was seen as preferable in this case.

To demonstrate an antisense control system could achieve tight regulation of *dapB* in the Artevo system, a proof-of-concept circuit was constructed in *E. coli* DH5α and *Acinetobacter baylyi* ADP1-ISx. The response function of this circuit was validated at both plasmid and chromosome levels, showcasing the orthogonality of this approach. While native antisense mechanisms have been widely identified on bacterial genomes, no previous work has validated synthetic antisense control circuitry installed on a bacterial genome. This work establishes inducible antisense promoters as effective tools for dynamic threshold adjustment of biosensory gene circuits, while preserving the response function of the primary promoter. Tight regulation of gene expression is important throughout synthetic biology; achieving this by simple addition of a convergent inducible promoter is a useful strategy for any synthetic biologist to add to their toolkit.

Interestingly, but unrelated to transcriptional interference and of no real consequence to this project, when constructing the plasmids required for this work a 'looping out' effect was observed during PCR amplification of the relevant transformation cassette (see supplementary *Figure S.21*). This necessitated a rebuilding of the relevant constructs to avoid secondary structure formation between two terminator sequences. It is thought this effect is due to high-processivity of Q5 polymerase, causing 'skipping' of the polymerase over a hairpin structure created by interaction of the two terminator sequences when the template dsDNA has melted to ssDNA.

Context of the *alkRM* operon

As mentioned in Chapter 3, the *alkRM* operon native to *Acinetobacter baylyi* ADP1 has previously been used by Zhang *et al.* to construct a bioluminescent reporter strain for detection of alkanes (C7-C36) in seawater[162]. The *alkRM* promoter region sits between the alkane-sensitive regulatory protein *alk*R and the alkane monooxygenase *alkM*, appearing to be constitutive in the *alk*R (antisense) direction and regulated by AlkR protein in the *alkM* (sense) direction[197,233]. In a native cellular context, this operon

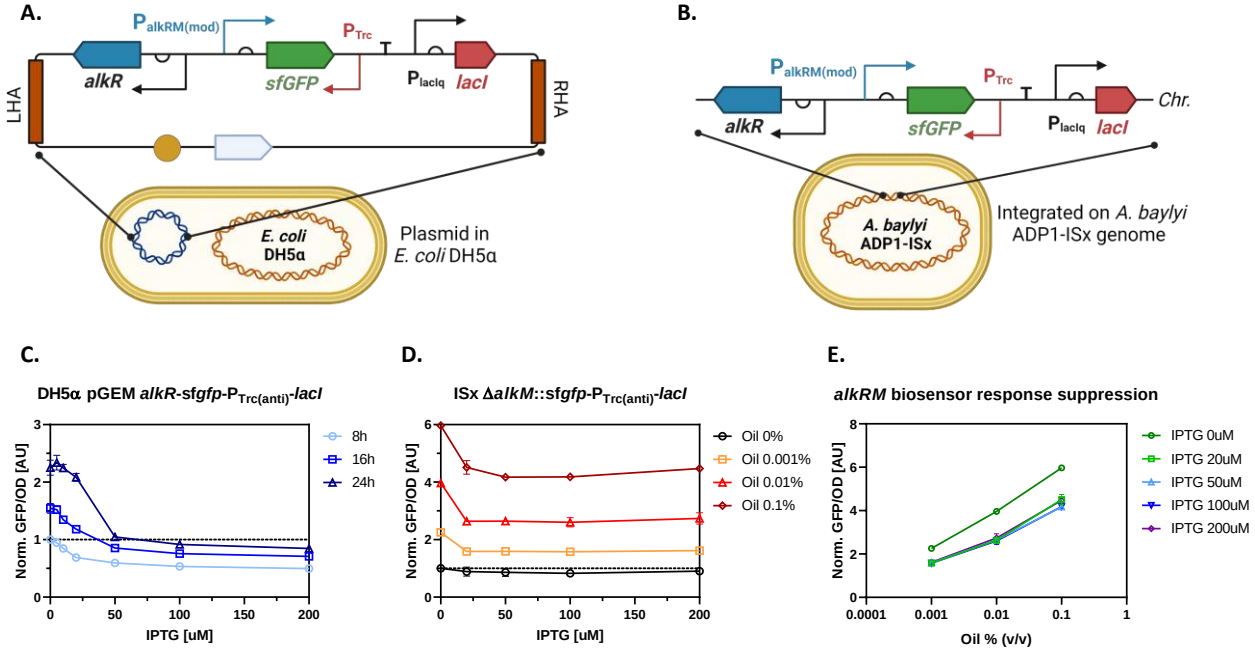


Figure 20. Direct transcriptional interference with *sfgfp* expression from under an alkane-inducible *alkRM* promoter, at both plasmid and chromosome levels. Part (a) shows the circuit structure housed on a pGEM vector backbone; the *alkRM* promoter (modified to contain three point mutations reported by Zhang *et al.* to decouple promoter activity from growth phase[161]), *alkR* regulatory gene, *sfgfp* reporter gene, and antisense control elements are flanked by 1kb of homology to the *alkRM* region on the *Acinetobacter baylyi* ADP1-ISx genome. Antisense regulatory elements consist of a convergently-oriented IPTG-inducible Trc promoter, placed downstream of *sfgfp*, and a constitutively-expressed *lacI* gene under P_{lacIq}. The plasmid is shown in the context of its host chassis, *Escherichia coli* DH5α. In DH5α, expression of *sfgfp* from under the *alkRM* promoter is constitutive, whereas in *A. baylyi* ADP1-ISx it is proportional to exogenous alkane concentration. Part (b) illustrates the same gene circuit integrated into the genome of *A. baylyi* ADP1-ISx and replacing the native *alkRM* region, functioning to replace the *alkM* gene with *sfgfp* and antisense control elements. Part (c) shows a response curve of DH5α containing the antisense control plasmid: specific fluorescence of the strain is reduced with increasing IPTG concentration. Part (d) shows the reduction in specific fluorescence observable in ADP1-ISx strains with the integrated antisense control circuit, across varying concentrations of exogenous mineral oil (%, v/v). Part (e) presents this same suppression of the *alkRM* biosensor in ADP1-ISx as a function of oil % (v/v), highlighting how the entire *alkRM* biosensor response can be downshifted using this antisense regulatory strategy. All measurements were taken from cultures grown in 96-well microtitre plates, according to previously described conditions and protocols (n = 3).

triggers expression of *alkM* in response to environmental alkanes, which ADP1 can utilise as a carbon source[196]. Zhang *et al.* generated the alkane biosensor strain by deleting *alkM* from its native genomic locus in ADP1 and replacing it with the cyanobacterial *luxCDABE* operon from *Synechoccocus elongatus*[121,266]. The resulting strain produces a luminescent signal proportional to the concentration of exogenous alkanes in its environment. The *alkRM* promoter can be altered with three single-nucleotide polymorphisms (SNPs) that decouple promoter activity from growth phase[162], facilitating a rapid and consistent bioluminescent response. In the following sections, references to the *alkRM* promoter (e.g. P_{alkRM}) concern this modified promoter containing these three SNPs. In earlier sections comparing the two, the modified promoter was denoted $P_{alkRM.mod}$.

In this work, the *luxCDABE* reporter element in the ADP1 biosensor strain was replaced with a fluorescent reporter, *sfgfp*, to generate the *alk*R-P_{alkRM}-*sfgfp*-$P_{Trc(as)}$ operon shown in *Figure 20(a)*. This simplified the circuit for testing with antisense regulation as the fluorescent signal is directly attributable to a single gene, *sfgfp*, and removed cross-sensitivity of the biosensor circuit for fatty aldehydes, which are a substrate for the luciferase encoded by *luxAB*.

The *alk*R-P_{alkRM}-*sfgfp*-$P_{Trc(as)}$ gene block, including a downstream *lacI* gene to facilitate inducibility of the antisense Trc promoter, modulates expression of *sfgfp* by using an antisense promoter that interferes directly with that gene. It was designed for installation on the *A. baylyi* ADP1 genome via that chassis' native homologous recombination mechanisms[155,209,267]. The pGEM vector containing this gene block was first transformed into *E. coli* DH5α cells for plasmid amplification, and the behaviour of the gene circuitry was tested in this chassis.

Reducing basal expression of *sfgfp* on plasmid DNA

E. coli DH5α cells containing the plasmid pGEM *alk*R-P_{alkRM}-*sfgfp*-$P_{Trc(as)}$ exhibit constitutive expression of *sfgfp* from under the *alkRM* promoter. Expression of *sfgfp* was directly opposed by an IPTG-inducible antisense Trc promoter, originally created in *E. coli* as a mutant of the Tac promoter and characterised in *A. baylyi* by Biggs *et al.* in their ADP1 toolkit[159,268]. Circuit architecture and normalised fluorescence data showing suppression of basal expression of *gfp* is shown in *Figure 20(c)*. A fold reduction in specific fluorescence of approximately 2-2.5 was observed across the induction range tested (IPTG 0-200μM), with the shape of the response function remaining constant across growth phases (induction curves generated at t = 8, 16, and 24 hours).

Chromosomal integration of an antisense promoter

A. baylyi ADP1-ISx is a transposon-free version of *A. baylyi* ADP1 generated by Suárez *et al.* that exhibits reduced mutation rates and increased transformability compared to wild-type ADP1[203]. ADP1-ISx is a useful chassis for genetic engineering as it features the natural competence of ADP1 but, due to the absence of transposable elements, genomic insertions are less likely to be inactivated. The *alk*R-P_{alkRM}-*sfgfp*-$P_{Trc(as)}$ circuit described above was inserted into the genome of *A. baylyi* ADP1-ISx at the native *alkRM* locus, such that the native *alkRM* promoter was replaced by the modified, growth-phase-decoupled version and the native alkane monoxygenase, *alkM*, was replaced by *sfgfp*. Downstream of *sfgfp* was the convergently-oriented inducible Trc promoter, followed by a terminator and a constitutively expressed *lacI* gene to express

the LacI regulator for Trc. Alkane-inducible expression of *sfgfp* was thus opposed by RNAPs fired from an IPTG-inducible antisense Trc promoter. Transcriptional interference from the antisense promoter was able to reduce specific fluorescence by 25-50% and, early during growth and at low alkane concentrations, overpower the sense promoter and return specific fluorescence roughly to baseline. Maximum reduction in fluorescence signal was observed between 20-50μM for each examined alkane concentration. Results are shown in *Figure 20*.

Application of transcriptional interference in Artevo strains

The proof-of-concept results shown in *Figure 20* demonstrate that a reduction in basal expression of *dapB*, and a tunable handle on stringency of selection, could be achieved in the Artevo system through inclusion of a convergent inducible promoter opposing P_{alkM}. This led to the construction of an upgraded Artevo circuit using the antisense Trc promoter architecture investigated here, plus an additional control element: an ssrA degradation tag on *dapB* that has been shown to shorten the half-life of proteins to which is it attached[236–240]. Targeted degradation of DapB was speculated to be useful for upkeep of alkane production in host Artevo strains, as the intracellular DapB pool would rapidly deplete in the absence of alkane.

Growth response of Artevo strains to exogenous alkane

Biofilm formation in microtitre plate

Acinetobacter baylyi ADP1-ISx ⊿*dapB* mutants (ISxJK001) that have been upgraded with the latest Artevo circuit, illustrated in *Figure 19*, are denoted ISxJK004 LVA^+ or ISxJK004 LVA^-, depending on whether the rescue gene *dapB* has a functional ssrA tag attached (abbreviated as LVA). The LVA^- phenotype was serendipitously generated when cloning efforts to produce ISxJK004 LVA^+ yielded some transformants with a frameshift mutation inactivating the LVA tag. ISxJK004 $LVA^{+/-}$ strains that have been further transformed with pBAV1k *gfp lacI* are hereafter called ISxJK004-S1 or S2, the former having the LVA^+ genotype and the latter LVA^-. ISxJK004 LVA^+ strains, at this stage, constitute the 'Artevo system'.

ISxJK004-S1 was grown in 96-well microtite plates, in MMS with varying combinatorial combinations of mineral oil and IPTG, to produce the growth curves shown in *Figure 21(a)*. These curves show each individual measurement of all three replicates, for each

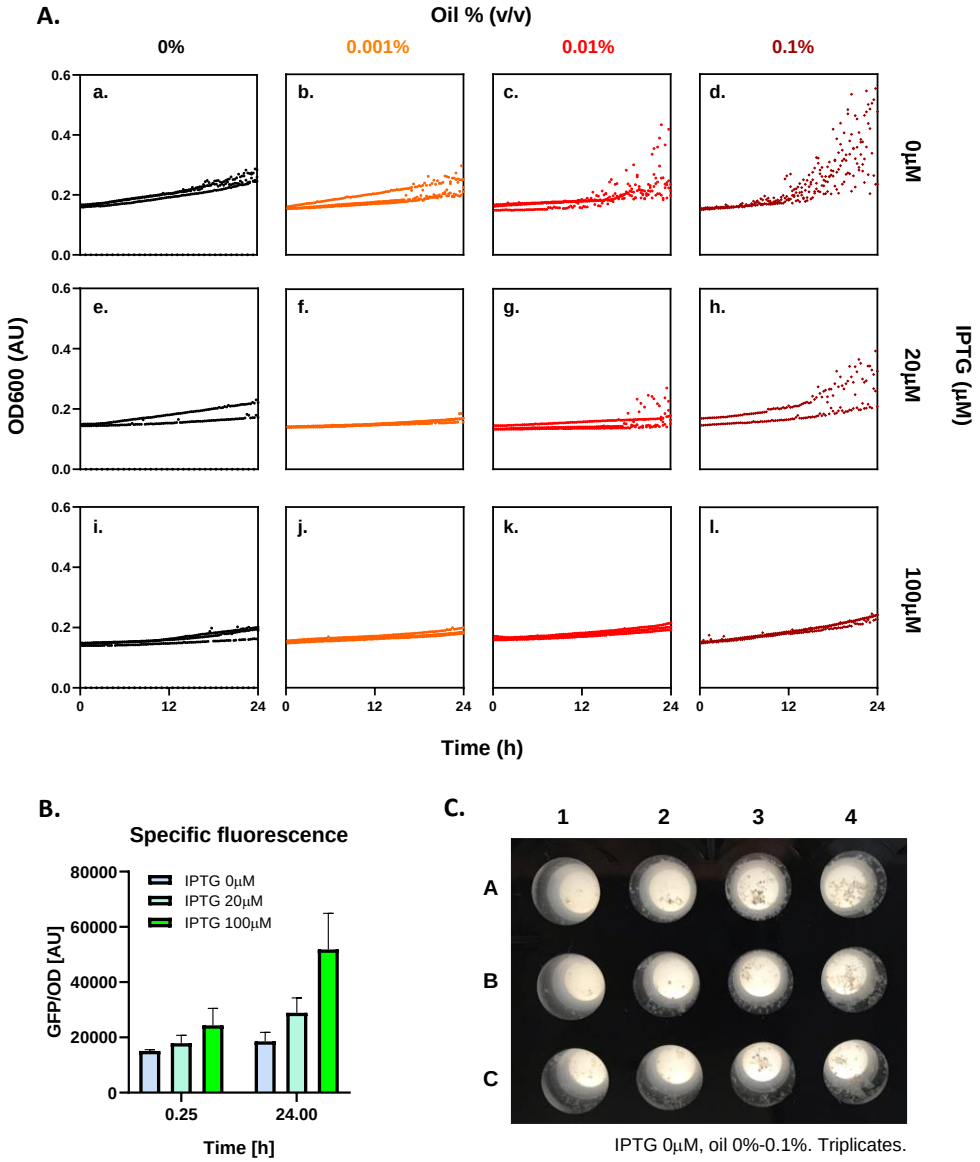


Figure 21. Growth of ISxJK004-S1 and biofilm formation in 96-well microtitre plates. Part (a) shows individual replicate measurements of OD600 (n=3) across culture wells with varying combinations (graphs a. to l.) of exogenous mineral oil (%, v/v) and IPTG (µM). Part (b) shows specific fluorescence of ISxJK004-S1 increasing over time and with IPTG concentration, indicating inducible expression of the T5 promoter on pBAV1k *gfp lacI* is functional and implying proper actuation of the antisense regulation mechanism acting on *dapB* expression. Part (c) is a qualitative comparison of biofilms observed within a subset of growth conditions (all oil concentrations, with no IPTG induction, n = 3), highlighting an apparent increase in growth rate that may be obscured when measuring optical density using a plate reader.

condition, at every time point, to illustrate the scattered nature of the data but also the qualitative changes in growth rate that can be observed across conditions. Examining the microtitre plate reveals extensive biofilm formation in the wells, shown in *Figure 21(c)*, but supports the conclusions suggested by the OD data, which are below. *Figure 21(b)* confirms expression of *gfp* from pBAV1k *gfp lacI*, verifying that IPTG-inducible promoters within the system should be functional.

It appears that growth ISxJK004-S1 is improved proportionally to exogenous mineral oil concentration, and suppressed by increasing concentrations of IPTG. This behaviour is consistent with expectations of how the Artevo system should behave given its new circuit architecture, which includes inducible antisense regulation of the rescue gene *dapB*.

Graded growth response in culture flasks

Growth of ISxJK004-S1 and S2 in 250mL culture flasks, again with varying concentrations of exogenous mineral oil and IPTG, produced clearer and more conclusive results compared to microtitre plate, as expected. *Figure 22* shows how increasing IPTG concentration suppresses growth in low-alkane conditions, improving system dynamics and causing a graded growth response to alkanes to manifest at 50μM IPTG. A significant difference in growth rate is observable between conditions (mineral oil = 0.01, 0.1, and 1.0%, v/v), though the range is not linearly proportional to alkane concentration; that is, a tenfold increase in exogenous alkane concentration does not produce a tenfold increase in growth rate. The Artevo system is also not leaky, with minimal errant growth in the absence of alkanes, even when *dapB* expression is unsuppressed. This result is the first direct observation of the graded growth response to exogenous alkanes which has been pursued throughout this chapter.

Both ISxJK004-S1 and S2 underwent preliminary tests in culture flask prior to the experiment which produced *Figure 22*. These early results, which enable a surface-level evaluation of the effects of the LVA degradation tag, can be found in supplementary *Figure S.22*, and include a qualitative comparison between growth of ISxJK004 strains in various concentrations of mineral oil. While not pronounced, the LVA tag does appear to have a positive effect on behaviour of the Artevo system, contributing to the desired graded growth response. Consequently, it was ensured that ISxJK004 strains tested with alkane biosynthesis plasmids in Chapter 5 were of the LVA^+ genotype.

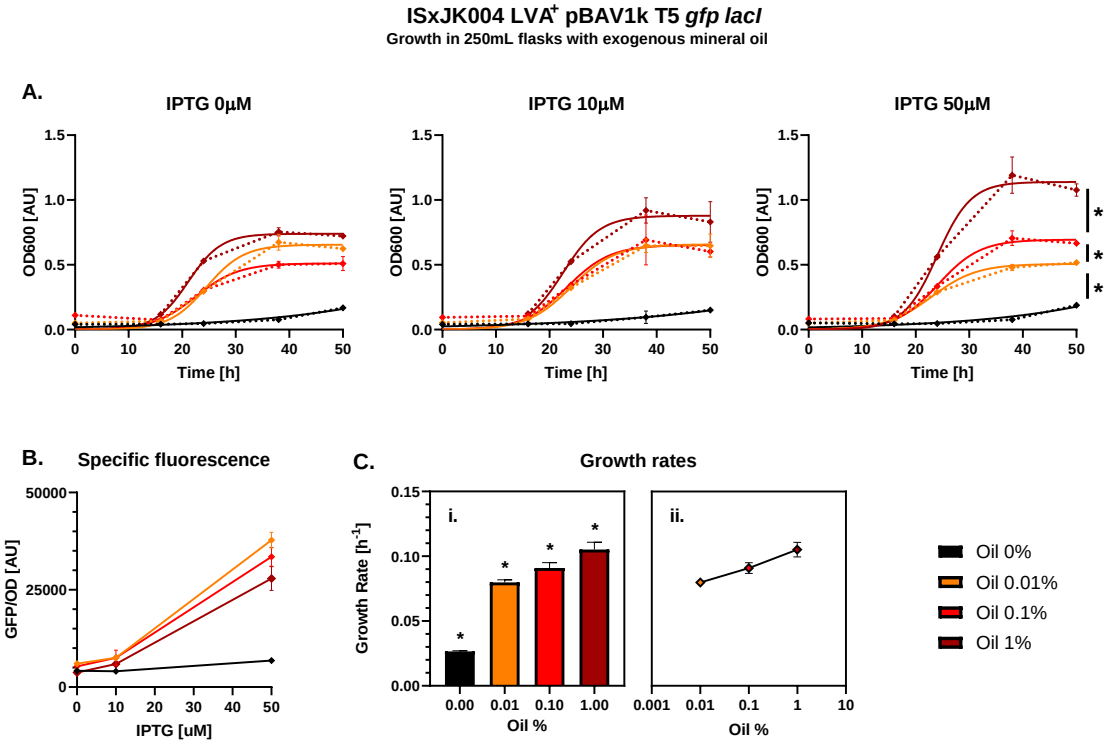


Figure 22. Graded growth response of Artevo strains (ISxJK004-S1) to exogenous mineral oil in 250mL culture flasks. Part (a) shows that increasing IPTG concentration improves the graded growth response to alkanes, with downregulation of the *dapB* rescue gene by transcriptional interference preventing errant growth of strains in low-oil conditions. Part (b) shows expression of *gfp* from pBAV1k *gfp lacI* increasing with IPTG concentration, confirming the function of LacI-regulated promoters within the system. Part (c) compares growth rates of ISxJK004-S1 in different media conditions, highlighting that growth rate varies significantly with increase concentration of exogenous mineral oil. (n = 3 biological replicates).

Materials and methods

Chemicals, media, growth conditions and equipment

All protocols were carried out with enzymes or mastermixes obtained from New England Biolabs (NEB), all primers were synthesised by Merck, and any gene blocks were synthesised by TWIST Bioscience. The pGEM-T plasmid backbone was originally obtained from Promega. QIAGEN or NEB kits and reagents were used for nucleic acid handling, such as gDNA extraction, plasmid extraction, HiFi assembly, and PCR purification. All chemicals were purchased from Sigma-Aldrich unless otherwise noted. For routine cell growth, bacteria were inoculated from a single colony and grown in Luria-Bertani (LB) broth with antibiotics at the following concentrations: kanamycin (50µg/mL for *E. coli* and 30µg/mL for *A. baylyi*), carbenicillin (100 µg/mL for *E. coli*, not used for *A. baylyi*). Cultures were incubated at 37°C (for *E. coli*) or 30°C (for *A.*

baylyi) and 200rpm. *E. coli* DH5α was routinely used for molecular cloning and plasmid maintenance. Final optimised constructs were tested in *E. coli* DH5α or in *A. baylyi* ADP1-ISx. Minimal media (MM) for *A. baylyi* ADP1-ISx was made according to the recipe previously described (supplementary information, page 151) with 20mM succinate as a carbon source. Filter-sterilised mineral oil was used to induce the *alkM* promoter in ADP1-ISx, and was introduced to the medium by making up a 1% (v/v) stock of mineral oil in MM, mixing vigorously by vortexing, and sonicating the mixture at 60Hz for 60 seconds to emulsify the oil droplets. This fresh stock solution of 1% mineral oil was diluted into the final minimal medium as required. IPTG (Isopropyl ß-D-1-thiogalactopyranoside) was added to the medium to the final desired concentration for induction of the Trc promoter. A Tecan Spark plate reader was used to monitor optical density (OD at 600nm) and fluorescence (excitation and emission wavelengths given below). GraphPad PRISM software was used for data visualisation and analysis.

Constructing the pGEM $P_{alkRM.mod}$-*sfgfp-lacI* vector

The transformation cassette for chromosomal insertion of the circuit shown in *Figure 20* was assembled on a pGEM vector backbone via NEB HiFi DNA assembly. The antisense control cassette and downstream *lacI* gene were synthesised as a gene block with 25bp overlaps at the 5' and 3' ends homologous to flanking regions on the *A. baylyi* ADP1-ISx genome. Upstream and downstream homology arms 1kb in length, with homology for the regions immediately up- and down-stream of *alkRM*, were amplified by PCR from *A. baylyi* ADP1-ISx gDNA. A pGEM vector backbone linearised via PCR amplification, both homology arms, and the antisense control circuit gene block were assembled by HiFi DNA assembly, and the resulting plasmid transformed into DH5α cells. Selection was performed on LB plates containing 100μg/mL carbenicillin. Positive transformants were screened using colony PCR and visualisation of amplification products by gel electrophoresis.

Genomic integration of antisense control cassette into ADP1-ISx

The PalkRM-*sfgfp-lacI* antisense control cassette was inserted into the *A. baylyi* ADP1-ISx genome at the native *alkRM* locus, using *tdk*.KanR counterselection, which has been described in detail by Suárez *et al.*[269]. The *tdk*.KanR counterselection cassette was a gift from the Barrick lab at UT Austin.

Firstly, an *alkM* targeting cassette was constructed as described for the pGEM PalkRM-*sfgfp-lacI* plasmid, above, except the homology arms flanked a *tdk*.KanR counterselector, which replaces *alkM* and its promoter on the ADP1-ISx genome. This *alkM* targeting

counterselector cassette was amplified by PCR and purified. *A. baylyi* ADP1-ISx was transformed using this purified, linear dsDNA, as described by the commonly-used protocol from the Barrick lab[270]. Positive transformants were selected on LB agar containing 30μg/mL kanamycin. Correct integration of the *tdk*.KanR counterselector at the *alkRM* locus, and deletion of the native *alkM* gene and its promoter, were verified by colony PCR. The counterselector was removed by repeating this process, transforming the ADP1-ISx Δ*alkM* strain with purified PalkRM-*sfgfp-lacI* cassette amplified from pGEM PalkRM-*sfgfp-lacI*. Loss of the *tdk*.KanR counterselector was selected for by plating transformants on LB agar containing 200μg/mL zidovudine, which is converted to a toxic metabolite in the presence of *tdk*. Insertion of the *sfgfp* antisense circuit was verified by colony PCR.

Characterisation of antisense-controlled circuits

E. coli DH5α containing pGEM P_{alkRM}-*sfgfp-lacI* was grown in triplicate in a 96-well plate at 37°C, in LB media containing 100μg/mL carbencillin and between 0-200uM IPTG. OD at 600nm and *sfgfp* fluorescence (excitation 480nm, emission 520nm) was measured every ten minutes, preceded by 15 seconds of double-orbital shaking at 360rpm. Induction curves were created using data from readings at 8, 16, and 24 hours.

Similarly, *A. baylyi* ADP1-ISx Δ*alkM*::*sfgfp*-$P_{Trc(anti)}$-*lacI* was grown in minimal media, formulated as described above with succinate as a carbon source, and supplemented with mineral oil at 0, 0.001, 0.01, and 0.1% (v/v), and with IPTG between 0-200uM. The same plate reader protocol for DH5α was used for ADP1-ISx, except ADP1-ISx was grown at 30C. An induction curve was created using data for each mineral oil concentration at t = 16 hours.

Genomic integration of upgraded Artevo circuit in ISxJK001

The upgraded Artevo circuit consisted of the modified *alkRM* promoter, the *dapB* rescue gene with a strong ssrA degradation tag (LVA), and an antisense, IPTG-inducible Trc promoter. This circuit was flanked by homology arms (1kB) targeting the *alkRM* locus, as previously described. The circuit was assembled by HiFi assembly into a pGEM-T vector, as previously described, except that the LVA tag and antisense Trc promoter were introduced using ssDNA oligos to 'bridge' between fragments in the assembly reaction. These ssDNA bridges were synthesised by Merck with 20-25 bp of homology to adjacent fragments, which stick to complementary ends of neighbouring fragments during assembly and are filled in, becoming dsDNA, by DNA polymerase. This method is outlined by NEB as a subset of their HiFi assembly protocol[271].

The upgraded circuit and homology arms, constituting the targeted transformation cassette, was PCR amplified from the pGEM vector and used to transform a ⊿*dapB* markerless mutant of *A. baylyi* ADP1-ISx by natural transformation, as previously described. The *alkRM* locus, including the *alkRM* promoter and downstream *alkM* gene, had previously been replaced in-frame in this mutant using a *tdk*.Kan^R counterselection cassette targeted to the locus using 1kb flanking homology arms, according to the counterselection method previously described. This mutant had the genotype ⊿*dapB* ⊿*alkM*::*tdk*.Kan^R. Transformants were selected on MMS agar containing ZDV and mineral oil, and screened using colony PCR.

Positive transformants were designated ISxJK004. Sequencing of the *alkRM* locus revealed two distinct genotypes produced by this transformation: one with an in-tact LVA tag on *dapB* and another with a frameshift mutation inactivating the tag. Serendipitously, this enabled subsequent comparison between the two strains to give some insight into the effects of the LVA tag, though a dedicated assessment would be required to fully elucidate its behaviour in the Artevo system. ISxJK004 strains were designated as LVA$^{+/-}$ depending on whether they contain a functional LVA tag.

The broad-host plasmid pBAV1k *gfp lacI*, obtained from addgene (plasmid #26702) and originally developed by Bryksin *et al.*[272], was transformed into ISxJK004 strains by natural transformation, to provide the necessary *lacI* regulatory gene facilitating inducible expression of the antisense Trc promoter on the Artevo circuit. Maintenance of this plasmid, and expression of *gfp* from under its *in situ* T5 promoter, had been previously validated in ADP1-ISx. ISxJK004 strains containing pBAV1k *gfp lacI* were named ISxJK004-S1 or S2, the former being of the LVA$^+$ phenotype and the latter LVA$^-$.

Growing ISxJK004 in the presence of exogenous alkanes

ISxJK004-S1 and S2 strains were initially grown in MMS containing 30μg/mL kanamycin (MMSK), in 250mL culture flasks (50mL working volume), across a mineral oil concentration gradient of 0.001,0.01, and 0.1% (v/v), and an IPTG gradient of 0, 50, and 100μM, according to the method previously described. Manual sampling of replicate flasks (n=3) was performed at timepoints indicated in supplementary *Figure S.22*. A high-performing singlicate of ISxJK004-S1 was isolated from this experiment and grown in 96-well microtitre plate under the same media conditions, with OD (600nm) and fluorescence (GFP; excitation 395 nm, emission 509 nm, bandwidth 20 nm) measured as previously described. Following observation of biofilm formation in microtitre plates, the high-performing ISxJK004-S1 singlicate was again grown in

250mL culture flasks, in MMSK, this time with combinatorial combinations of IPTG and mineral oil concentrations as follows: IPTG = 0, 10, 50 μM, mineral oil = 0, 0.01, 0.1, 1.0% (v/v). Replicates (n=3) were sampled at timepoints indicated in *Figure 22*.

All replicates in these experiments were true biological replicates, arising from distinct transformation events (that is, separate colonies on the transformation plate) and carried forward as separate lineages (that is, stored as separate glycerol stocks and inoculated into separate pre-cultures for inoculation of cultures in these experiments). Hence the high variability between replicates observed in these experiments. Replicate strains were pre-cultured in MMSK plus 0.01% (v/v) mineral oil, and washed three times in PBS. Pre-culture densities were normalised prior to inoculating secondary cultures.

Statistical analysis of data in *Figure 22* was performed using the GraphPad PRISM statistics suite. A two-way ANOVA was performed to compare absolute OD between all mineral oil conditions at t = 16 and t = 24 hours, for the IPTG = 50μM case. An ordinary one-way ANOVA was used to identify significant differences between growth rates across conditions, which were calculated using the linear region of a plot of ln(OD) against time. Complete growth curves for each data set were approximate using a Gompertz model.

Part II. Discussion

Tight control of gene expression is important when engineering biological systems. Bioengineers may have only a limited toolkit of tightly-regulated sense promoters – such as the Jungle Express system reported by Chen *et al.*[273] – compatible with their chassis, typically necessitating laborious discovery of new or altered promoters. This work suggests an alternative solution: the combination of two existing, individually imperfect promoter systems in an antisense configuration to achieve tight control of gene expression. These promoter systems may be either inducible or constitutive and have broad applicability within existing gene circuit architecture. Practical applications of antisense control include reduction of basal expression of toxic gene products, tuning of biosensory circuits where the sense promoter is regulated by a target metabolite, or threshold adjustment of genetic switches. This approach is particularly relevant in the context of the Artevo system, as it was discovered in this project that the tightly-regulated Jungle Express system does not function in *A. baylyi* ADP1-ISx, even when the *eilR* regulator is codon optimised for that chassis.

Early tuning attempts, discussed in Part I of this chapter and summarised in *Table 3*, focussed on eliminating apparent leaky expression of *dapB* within the Artevo circuit. While decoupling *alkRM* promoter activity from growth phase via introduction of three point mutations did improve system dynamics, the leakiness issue was solved not through genetic engineering methods, but simply by changing the experimental scale using larger culture vessels. Some leaky expression of *dapB* should not render the Artevo system ineffective, provided that increasing expression of *dapB* confers a proportional growth advantage to the host. That is, alkane-producing phenotypes must be fitter than phenotypes surviving solely due to basal expression of *dapB*, and fitness must be proportional to alkane productivity. Errant growth of unproductive phenotypes could also be limited by separating the evolving population into several independent replicates, diluting cheater phenotypes and favouring the exploration of multiple evolutionary paths. The growth response of Artevo strains to exogenous mineral oil was still binary in the early stages of this project, rather than graded, and a graded response is fundamental to the function of a directed evolution system.

Tuning efforts in Part II of this chapter focussed on reducing the dynamic range of *dapB* expression, to elicit a graded growth response from host cells in the presence of varying concentrations of exogenous alkane. This was accomplished via direct transcriptional interference with *dapB*, through a convergently-oriented inducible Trc promoter. *Figure 20* demonstrates this system operating at both plasmid and chromosome levels, as a proof-of-concept, and *Figure 22* highlights the success of the strategy in the Artevo system.

While the tuned Artevo system did achieve a graded growth response to exogenous alkanes, less than a twofold range of growth rates was observed across three orders of magnitude of alkane concentration. There is much room for improvement in this range, as currently even very high producing phenotypes would be only marginally fitter than lower producing ones. Optimisation efforts could target several aspects of the system which remain unexplored, such as varying the dynamic range of the antisense promoter regulating *dapB*, the strength of the ssrA degradation tag attached to *dapB*, or the catalytic activity of DapB protein itself. The ratio of growth rate increase to alkane productivity should be characterised in more detail ahead of further tuning attempts, specifically in regards to intracellular alkane concentration derived from installed alkane biosynthesis gene modules. Such modules are discussed in the next chapter, where intracellular alkanes manufactured by the host cell will also be seen to produce higher growth rates than even 1.0% (v/v) exogenous mineral oil.

Figure 22 also suggests a timeframe over which the system may be effectively utilised in noncontinuous directed evolution. By the end of the 50-hour growth curve, Artevo strains lacking exogenous mineral oil had begun to grow, albeit minimally. Cheater phenotypes that are capable of growth in the absence of alkanes may present a small risk to system operation, enriching suboptimal genotypes within the population and potentially obscuring high-performing alkane biosynthesis gene modules. As discussed earlier in this section though, this effect is unlikely to disrupt the function of the Artevo system if errant growth is minimal. Knowing the growth rate of this background cheater population could nevertheless inform future bioprocess optimisation, helping to establish boundaries for incubation periods or retention time, if the system is developed for use in a chemostat for *in situ* continuous directed evolution.

Artevo strains grown in 0% (v/v) mineral oil exhibited low specific fluorescence. It was expected that IPTG-inducible expression of *gfp* from pBAV1k *gfp lacI* would be consistent across all oil concentrations, as seen for 0.01, 0.1, and 1.0% (v/v) conditions. Reduced specific fluorescence in oil-free media may be due to the greatly extended lag phase in the absence of oil; expression of *gfp* may not have begun in earnest at that early stage of growth, whereas expression of *dapB* from under the modified *alkRM* promoter is independent of growth phase.

Higher oil concentrations were used in flask experiments compared to microtitre plates, as better mixing at flask scale appears to keep the pseudosolubilised hydrocarbon phase stable for longer than what is achievable in plates. Phase separation in microtitre plate can also impact OD readings by interfering with the beam as it passes through the culture volume, whereas manual sampling in flasks avoids this issue. The flask growth experiment in *Figure 22* also uses a different IPTG concentration gradient than that used in microtitre plate (*Figure 21*). This is because preliminary flask growth experiments for ISxJK004 LVA$^{+/-}$ strains (see supplementary *Figure S.22*), and proof-of-concept results from *Figure 20*, suggested maximal growth suppression would occur around 50 μM IPTG.

While further tuning could be performed to optimise the Artevo system response to exogenous alkanes, it is the response to intracellular alkanes, those produced by the host cell when equipped with alkane biosynthesis genes, which will ultimately drive the directed evolution system. It was decided to proceed with the Artevo strains as they were and install genes coding for alkane biosynthesis machinery, testing whether host fitness would be meaningfully affected by relative productivity of the installed gene modules.

Chapter 5

Alkane biosynthesis

This chapter builds upon the Artevo system established in Chapter 4 with strain ISxJK004 LVA$^+$. It discusses the addition of alkane biosynthesis machinery to ISxJK004 LVA$^+$, enabling host cells to natively produce alkanes and drive the directed evolution circuit using endogenous rather than exogenous alkane. The Artevo system should link this endogenous alkane production rate to host growth rate; cells that have been transformed with sequences encoding high-yielding alkane biosynthesis enzymes should grow faster than those containing low-yielding sequences. In this way, high-producing bacterial phenotypes dominate the population over generations and, consequently, the most productive underlying alkane biosynthesis sequences are enriched. Experiments described in this chapter tested the capacity of the Artevo system to distinguish between biosynthetic gene modules with relatively low, medium, and high alkane productivity, and enrich the highest-performing module over time.

Enzymes for alkane biosynthesis

In a 2018 paper, the Santala group at the Tampere University of Technology (now Tampere University), Finland, engineered *A. baylyi* ADP1 to metabolise acetate into

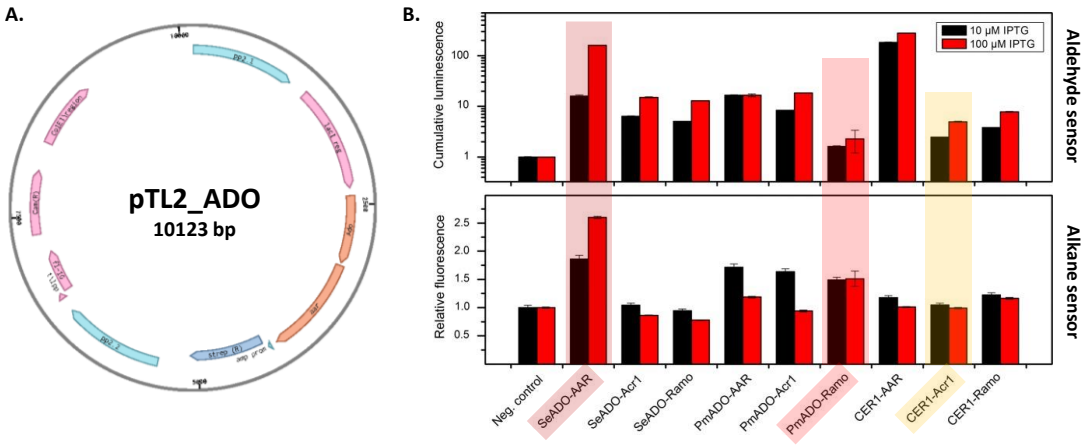


Figure 23. Varying relative strength of alkane biosynthesis modules. Part (a) shows the pTL2 plasmid which carries any given alkane biosynthesis module (depicted here with seADO-AAR). Part (b) shows the relative fluorescence and luminescence produced by each alkane biosynthesis module when expressed in the dual-layer biosensor developed by Lehtinen *et al.*[244] Part (b) of this figure is republished directly from a study of alkane biosynthesis from CO_2 in *A. baylyi* ADP1, by the Santala group at Tampere University[245]. See permissions on page 149.

alkanes using a two-stage bacterial process similar to the process targeted by our own group[248]. They first utilise the native Wood-Ljungdahl pathway in the acetogen *Acetobacterium woodii* to fix CO_2 and hydrogen into acetate. This acetate is fed to a strain of *A. baylyi* ADP1 that has been engineered for alkane production by transformation with a plasmid encoding a fatty acyl-CoA reductase (FAR) and an aldehyde deformylating oxygenase (ADO). The study tested several combinations of FAR and ADO, from a variety of microbial and plant sources, to determine the alkane yield from each enzyme pair. These enzymes are detailed in *Table 4* and results from the Santala study are republished in *Figure 23*. Alkane yield was quantified using a twin-layer biosensor circuit developed by the Santala group[247], based on the *alkRM* alkane biosensor originally reported by Prof. Huang in 2012[162] and which underpins the Artevo system in this project. The twin-layer system improves upon Prof. Huang's original bioreporter by using *luxAB* genes to produce a bioluminescence signal in response to aldehydes and *gfp* to produce a fluorescent signal in response to alkanes. This allows real-time and distinct monitoring of both intermediate fatty aldehydes and end product alkanes produced by FAR and ADO components, respectively, of an alkane biosynthesis system.

The Santala group kindly gifted me with three constructs used in their study, which had been verified in *A. baylyi* ADP1 and shown to produce relatively high, medium, and low levels of alkane. These constructs, pTL2 seADO-AAR, pTL2 pmADO-RAMO, and pTL2 CER1-ACR1, are described in *Table 4*. The pTL2 vector carrying each alkane biosynthesis cassette is a medium-copy plasmid originally derived from pET-28 and engineered to contain a gene conferring cross-resistance to spectinomycin and streptomycin[160]. Replication of this vector has been verified in *A. baylyi* ADP1[200].

Table 4. Alkane biosynthesis constructs, their constituent enzymes, source organisms, and relative productivity.

ENZYME	SOURCE	NOTES
Fatty acyl-CoA reductases (FARs)		
AAR	*Synechococcus elongatus*	Acyl-ACP reductase involved in alkane production. Studies suggest it is the most efficient cyanobacterial homologue of this gene[274].
Acr1	*Acinetobacter baylyi*	Endogenous to ADP1, involved in wax ester production[275].
Ramo	*Nevskia ramosa*	Putative dehydrogenase identified by sequence homology with Acr1[200]. Source organism is a bacterium found in the neuston layer of freshwater lakes and bogs.
Aldehyde deformylating oxygenases (ADOs)		
seADO	*Synechococcus elongatus*	Bottleneck of current efforts to microbially synthesise alkanes[139]. A nonheme di-iron enzyme catalysing conversion of C_n fatty aldehyde to C_{n-1} hydrocarbon with formate as the co-product[141].
pmADO	*Prochlorococcus marinus*	Another cyanobacterial ADO, similar to seADO. Originally used in a landmark study by Schirmer *et al.* for microbial biosynthesis of alkanes[121].
CER1	*Arabidopsis thaliana*	Has been used in previous studies for heterologous alkane production, as an alternative to cyanobacterial ADOs. Produces very long chain alkanes[276].

CONSTRUCT	ADO	FAR	RELATIVE ALKANE PRODUCTIVITY
pTL2 seADO-AAR	seADO	AAR	**High**
pTL2 pmADO-RAMO	pmADO	RAMO	**Medium**
pTL2 CER1-ACR1	CER1	ACR1	**Low**

Notably, the pTL2 vector cargo – an alkane biosynthesis module in this case – is regulated by an IPTG-inducible T5 promoter, the dynamics of which have been demonstrated in this project by expression of *gfp* from the plasmid pBAV1k T5 *gfp lacI*. This means, in the ISxJK004 LVA$^+$ strain equipped with a pTL2 plasmid, IPTG will induce both expression of the ADO-FAR module on pTL2 and firing of the antisense Trc promoter opposing *dapB*.LVA on the host genome. To avoid additional engineering of the pTL2 vectors received from the Santala group, and because the *A. baylyi* chassis has a limited toolkit of reliable inducible promoters, it was decided that the dual activity of IPTG could be balanced by inducing with a concentration that had been shown both to promote strong expression of the ADO-FAR module in pTL2 and to produce a graded growth response to alkanes in ISxJK004 LVA$^+$. *Figure 22* in

Chapter 4 suggests that such a concentration would be 50uM IPTG, half the maximum inducer concentration used by the Santala group in their study.

Each pTL2 construct carries one of three distinct alkane biosynthesis modules:

i. seADO-AAR, which is a purely cyanobacterial pathway derived from *synechococcus elongatus*.

ii. pmADO-RAMO, which contains an ADO from *Prochlorococcus marinus* and an FAR from a freshwater neuston bacterium *Nevskia ramosa*.

iii. CER1-ACR1, which has an ADO from the thale cress plant *Arabidopsis thaliana* and a FAR native to *A. baylyi* ADP1 involved in wax ester biosynthesis.

Figure 23 (also verified in *Figure 24*) shows that each module produces a high, medium, or low amount of alkane. By transforming these pTL2 constructs into the tuned Artevo system, it was possible to test whether the system could distinguish between these sequences based on their relative alkane production rate and, in a combined culture containing all three constructs, enrich the highest-producing one over time.

Installation of alkane biosynthesis machinery

Alkane yield from biosynthesis modules in pTL2 plasmids was first tested in the ADP1 alkane biosensor strain developed by Zhang *et al.*[162], which produces a bioluminescent signal in response to alkanes, as previously discussed. The alkane biosensor strain was transformed separately with each pTL2 plasmid and luminescence of these strains monitored following induction with IPTG. *Figure 24* shows specific luminescence increasing between modules CER1-ACR1, pmADO-RAMO, and seADO-AAR, in that order, confirming results from Lehtinen *et al.*[248].

The tuned Artevo strain, ISxJK004 LVA$^+$, was transformed separately with each pTL2 construct to produce three new strains: ISxJK004 LVA$^+$ pTL2 seADO-AAR, ISxJK004 LVA$^+$ pTL2 pmADO-RAMO, and ISxJK004 LVA$^+$ pTL2 CER1-ACR1. Positive transformants were selected on LB agar and subsequently pre-cultured in LB broth, both containing spectinomycin, for glycerol stocks. These pre-cultures were used to inoculate a variety of secondary media to test whether strains containing a high-yielding alkane biosynthesis module (seADO-AAR) would grow faster than those containing medium- (pmADO-RAMO) and low- (CER1-ACR1) yielding modules, implying the alkane biosynthesis machinery had been successfully installed and was driving the Artevo circuit.

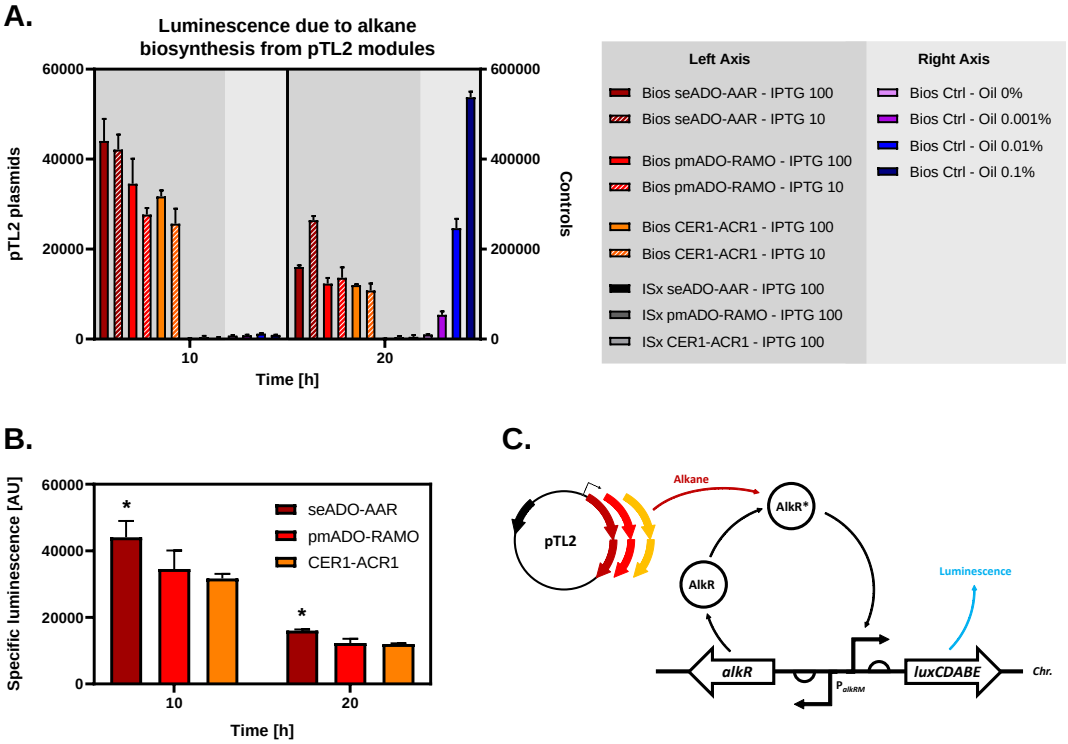


Figure 24. Testing of relative alkane productivity of biosynthesis modules: seADO-AAR, pmADO-RAMO, and CER1-ACR1, all in pTL2 plasmid backbones. Part (a) compares specific luminescence, after 10 and 20 hours of growth, from the ADPWH_alk biosensor strain transformed with one of three pTL2 alkane biosynthesis plasmids, each containing one of the modules mentioned above. Strains were grown in MMS media and induced with 10 μM or 100 μM IPTG, and compared to ADPWH_alk controls grown in MMS + supplemental mineral oil (0-0.1%, v/v). Wild-type ADP1-ISx strains transformed with each pTL2 plasmid were also included as controls. ADPWH_alk and ISx control strains containing alkane biosynthesis modules are shown in the dark grey portion of the plot, corresponding to the left axis, and ADPWH_alk control strains grown in supplemental mineral oil are shown in the light grey portion of the plot, corresponding to the right axis. This separation was made because the peak luminescence response of ADPWH_alk control strains at t = 20 hours is an order of magnitude greater than that of ADPWH_alk strains carrying an alkane biosynthesis module. Part (b) focusses on the 100 μM condition from part (a), highlighting that expected relative alkane productivity translates to a difference in specific luminescence, suggesting modules are performing as described by Lehtinen *et al.*[245] Part (c) illustrates the circuit structure of ADPWH_alk interacting with a pTL2 plasmid to produce a luminescent signal.

Figure 25 illustrates a subset of conditions from this experiment. The experiment began with all ISxJK004 LVA$^+$ pTL2 strains, and ISxJK004 LVA$^+$ pBAV1k T5 *gfp lacI* controls from the experiment depicted in *Figure 22*, grown in either the presence (at 0.01% v/v) or absence (0% v/v) of mineral oil. Additionally, each of these conditions was replicated with either 0μM, 50μM, or 100μM of IPTG to induce both expression of the ADO-FAR genes from the pTL2 plasmid and antisense interference with *dapB*. The lag phase of ISxJK004 LVA$^+$ pTL2 strains was far greater than expected, and most of these conditions and replicates were discounted after no growth was observed within 72 hours. Nine key conditions, covering ISxJK004 LVA$^+$ pTL2 strains growing without inducer, with inducer, and with both inducer and supplemental mineral oil, were retained to investigate any extremely slow growth; these are shown in *Figure 25*.

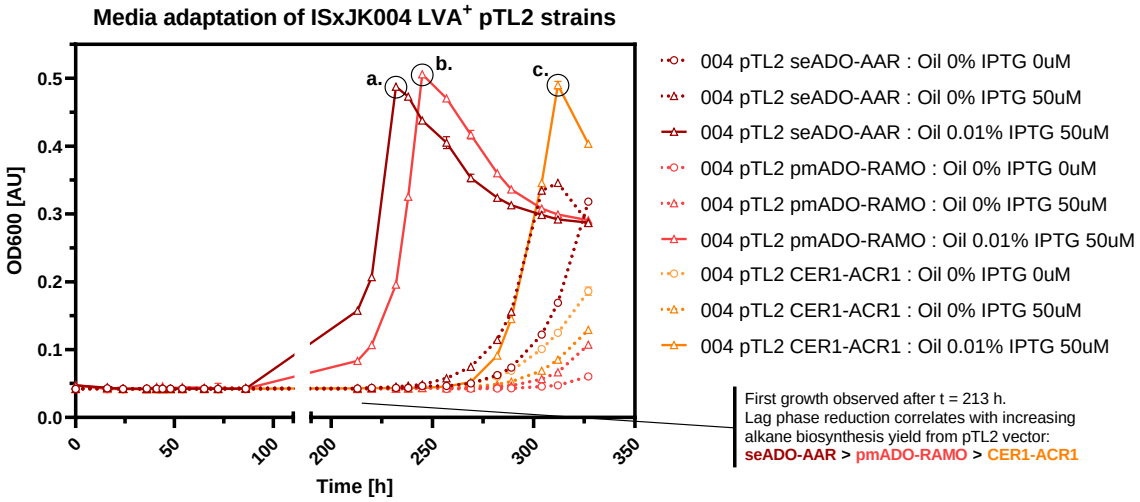


Figure 25. Early experiment to test ISxJK004 LVA+ strains transformed with pTL2 plasmids, in various media conditions. Stages of this experiment are detailed in the accompanying text.

Growth of the new ISxJK004 LVA+ pTL2 strains was first observed after 213 hours, in media containing supplemental mineral oil at 0.01% (v/v). Growth in pTL2-containing strains correlated with expected relative alkane yields from the respective biosynthesis modules carried by each plasmid – seADO-AAR, pmADO-RAMO, and CER1-ACR1 – with high-producing modules seeming to confer a greater lag phase reduction in the host chassis. This pattern appears first in media containing supplemental mineral oil and IPTG but is later observable in media containing only IPTG inducer. ISxJK004 LVA+ pTL2 strains growing with 50µM IPTG and 0.01% (v/v) supplemental mineral oil, marked as a., b., and c. in *Figure 25*, were streaked onto minimal agar for screening of individual colonies via colony PCR. Primer probes targeting both *dapB* genomic insertion at the *alkM* locus in .004 strains, and the specific ADO-FAR module on each pTL2 plasmid, positively identified each of these strains. Subsequent sequencing of the latter amplicons confirmed the presence of each alkane biosynthesis module in the ISxJK004 LVA+ host. New glycerol stocks of ISxJK004 LVA+ pTL2 seADO-AAR, ISxJK004 LVA+ pTL2 pmADO-RAMO, and ISxJK004 LVA+ pTL2 CER1-ACR1 were made from these positively identified clonal colonies that had now adapted to growth in minimal media.

Having separately installed the three pTL2 plasmids into ISxJK004 LVA+ strains, a proof-of-concept experiment could be performed to investigate the fitness advantage gained by hosts carrying high-producing alkane biosynthesis genes.

Growth response to native alkane production

In obtaining the ISxJK004 LVA$^+$ pTL2 strains, it was observed that the growth advantage expected in hosts containing a relatively high-producing alkane biosynthesis module appeared to manifest as a reduction in lag phase. As described in the previous section, ISxJK004 LVA$^+$ strains containing pTL2 seADO-AAR were the first to grow, followed by those containing pTL2 pmADO-RAMO and, finally, pTL2 CER1-ACR1. This lag phase reduction appeared again when these strains were tested in a preliminary plate experiment prior to attempting a full-scale growth experiment in 250mL flasks; see supplementary *Figure S.23*. While the issues with using 96-well microtitre plates for growth of strains in this project has been discussed in previous chapters, 24-well plates are a useful middle-ground between plate and flask formats that allow for high-time-resolution data to be automatically obtained prior to committing to laborious manual sampling experiments at flask scale. This preliminary examination of pTL2 strains in 24-well plate revealed a similar growth pattern to that observed in *Figure 25*, with seADO-AAR, pmADO-RAMO, and CER1-ACR1 phenotypes being distinguished both by length of lag phase and peak culture density. Notably, growth in media both containing and lacking supplemental mineral oil was observed much sooner than in the previous experiment. A combined culture of all three pTL2-containing strains exhibited the shortest lag phase but the same peak culture density as ISxJK004 LVA$^+$ pTL2 seADO-AAR.

The three media conditions retained in the previous growth experiment – minimal media containing neither inducer nor supplemental mineral oil, minimal media containing only inducer, and minimal media containing both inducer and oil – were used to test the growth of pTL2 strains in 250mL flasks. Further, a combined culture of all three pTL2 strains was tested in these same media conditions and sampled periodically for subsequent investigation of enrichment of alkane biosynthesis modules, which is described later in this chapter.

Figure 26(a) shows raw OD600 data manually sampled from cultures of each ISxJK004 LVA$^+$ pTL2 strain, as well as a combined culture of all three strains. Similar to previous experiments, supplemental mineral oil accelerated the growth of each pTL2 strain, with growth of those strains in oil-containing media being the only growth observable within the timeframe of this experiment. While the CER1-ACR1 strain began growing first, the growth rate of this strain was lower than seADO-AAR and pmADO-RAMO strains, which both quickly overtook CER1-ACR1. Applying the Gompertz model to this data provided three well-fitting curves, shown in *Figure 26(b)*, which illustrate the

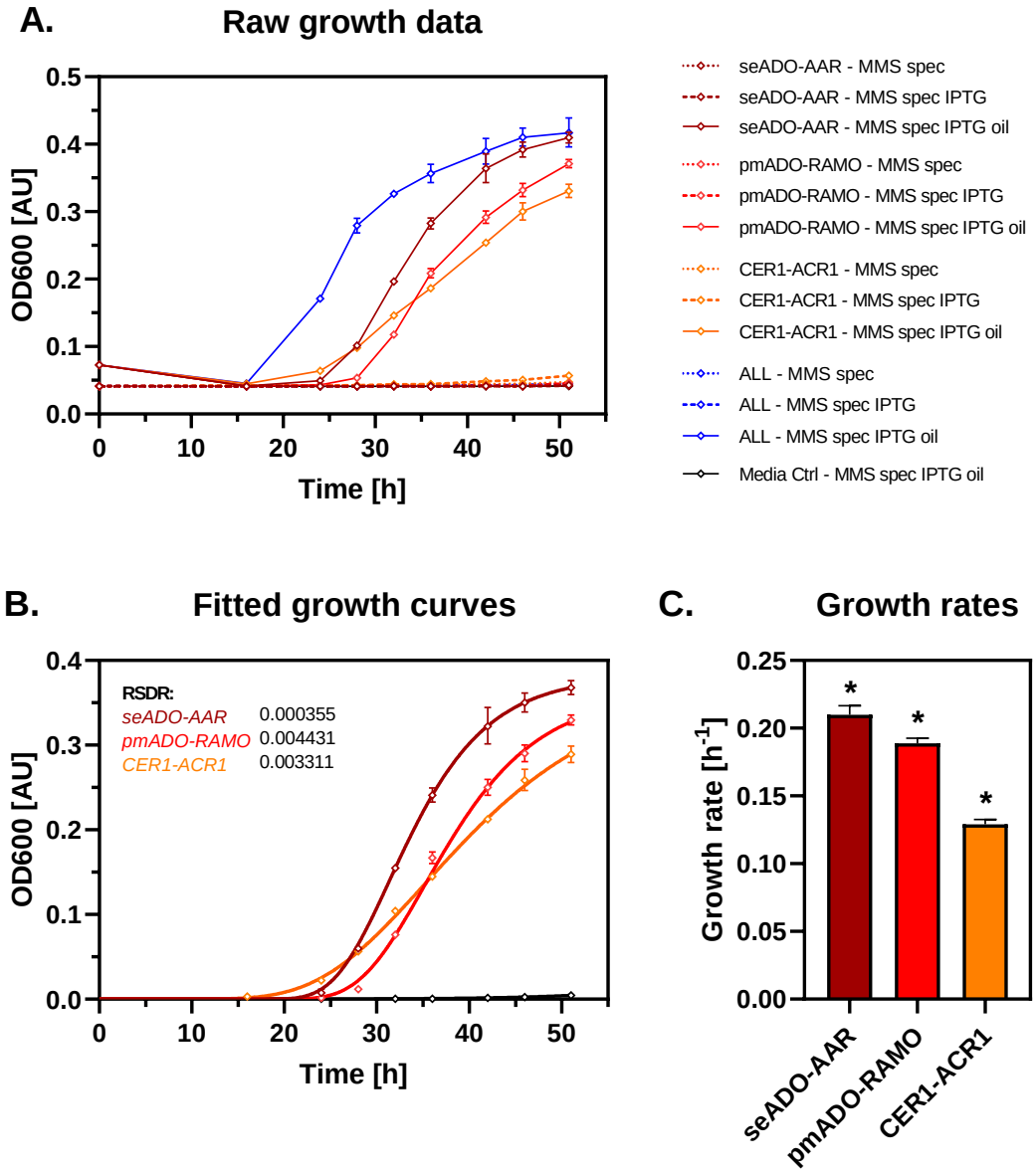


Figure 26. Proof-of-concept demonstration of the Artevo system. Artevo strains (ISxJK004 LVA+), transformed with a pTL2 plasmid containing one of three alkane biosynthesis modules, exhibit varying growth rates proportional to the productivity of the alkane biosynthesis module they are carrying. Part (a) shows raw OD600 data from manual sampling of ISxJK004 LVA+ cultures, in 250mL flasks (n = 3 pseudo biological replicates). Growth was only observable in media conditions containing a mineral oil background (0.01% v/v), represented by solid lines, though growth rate did vary between strains depending on their constituent alkane biosynthesis module. The blue line indicates a combined culture of all three ISxJK004 LVA+ pTL2 strains. Part (b) again shows growth curves for ISxJK004 LVA+ pTL2 strains, after data had been processed and fitted to a Gompertz curve. Robust Standard Deviation of Residuals (RSDR) for each curve is shown on the graph. Part (c) compares growth rates (derived from their respective Gompertz curves) of Artevo strains expressing each alkane biosynthesis module.

relative growth of each pTL2 strain and demonstrate a significant difference in fitness between them. *Figure 26(c)* compares growth rates obtained from each fitted curve, again supporting the pattern observed in previous experiments and highlighting the result the Artevo system has been aiming to achieve: a significant difference in growth rate dependent on alkane productivity of the host chassis.

These results demonstrate that, in the Artevo system strain ISxJK004 LVA$^+$, relative differences in alkane production rate will translate into relative differences in host fitness, manifesting as a difference in growth rate.

Generational enrichment of seADO-AAR in the combined culture is implicit but has not yet been directly observed. Experiments described in the following section aimed to quantify the enrichment of this high-performing alkane biosynthesis module compared to its lower-performing counterparts: pmADO-RAMO and CER1-ACR1.

Enrichment of high-performing gene sequences

The previous section showed that a population of ISxJK004 LVA$^+$ cells can compare the effectiveness of three heterologous alkane biosynthesis modules contained within sub-populations – seADO-AAR, pmADO-RAMO, and CER1-ACR1 – and produce a graded growth rate response within these populations proportional to the alkane yield from each module. The practical goal of this system is for high-producing phenotypes in a mixed population to outgrow low-producing phenotypes, enriching one sub-population of cells over all others. The prevalence of whatever alkane biosynthesis module is being carried by this sub-population will also increase and can be later examined to identify specific mutations contributing to improved alkane productivity.

Enrichment of specific sequences within a combined culture of ISxJK004 LVA$^+$ containing equal starting amounts of each pTL2 plasmid – the blue line in *Figure 26* – should be observable by several methods, two of which were attempted as part of this project and detailed here. Results from these approaches were inconclusive and troubleshooting efforts were limited by project time constraints, but the methodology described below should be effective in directly demonstrating enrichment if the issues identified can be addressed in future work.

The first method used to measure sequence enrichment was an intercalating dye quantitative PCR (qPCR) assay, illustrated qualitatively in *Figure 27* and quantitatively in *Figure 28*. Primer probes specific to each module – seADO-AAR, pmADO-RAMO, and CER1-ACR1 – were used to compare the relative number of copies of each module within a crude culture of ISxJK004 LVA$^+$ containing all pTL2 plasmids. The cycle threshold (C_t) of each amplification reaction was compared to that of a control gene, *dapB*, of which only one copy should exist per cell, at both time zero (t_1) and after 51 hours of growth (t_2). The expected result was an equal relative C_t value for all three modules at t_1, immediately following inoculation of the combined culture with an equal number of cells from each pre-culture containing either seADO-AAR, pmADO-

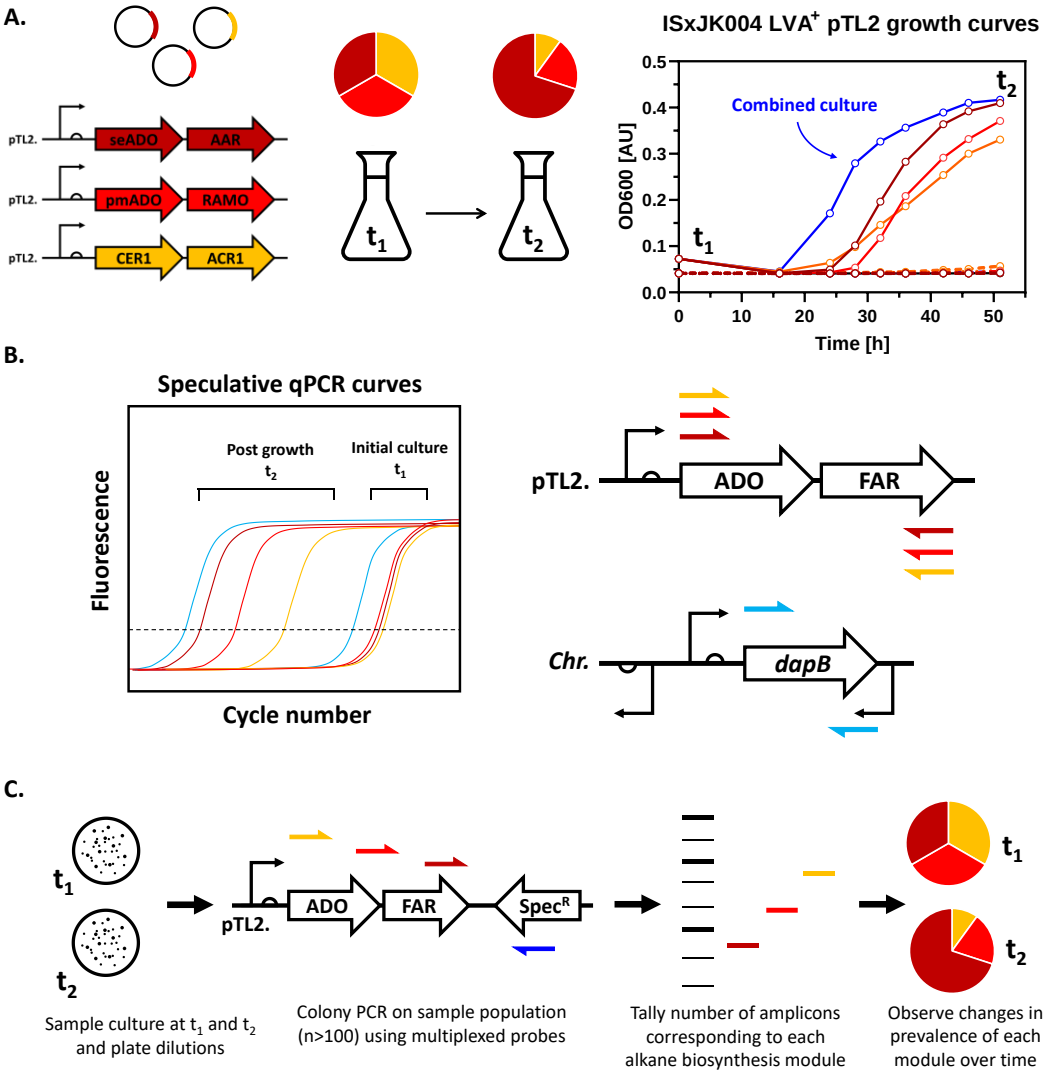


Figure 27. Proposed methods for measuring enrichment of high-performing alkane biosynthesis modules by the Artevo system. Part (a) depicts a combined culture of ISxJK004 LVA+ pTL2 [seADO-AAR / pmADO-RAMO / CER1-ACR1] strains – the blue line in *Figure 26(a)* – being grown and sampled at the start (t_1 = 0 hours) and end (t_2 = 51 hours) of a period of growth, with a qualitative depiction of the expected trend in genotype enrichment. Part (b) shows qPCR probes targeting each ADO-FAR module present in the combined population, and speculative qPCR curves showing expected results of this experiment: a cluster of curves at t_1 indicating equal prevalence of each module, and a spreading out of curves, each with lower C_t values, at t_2 indicating an absolute increase in abundance of all modules and a relative increase in prevalence of seADO-AAR over pmADO-RAMO and CER1-ACR1. Blue lines in this speculative qPCR amplification plot represent a reference gene, *dapB* in this case, of which one copy exists per cell. *Figure 28* shows actual qPCR curves obtained from this experiment. Part (c) illustrates the workflow for a multiplexed colony PCR assay performed on starting and enriched cultures of pTL2-containing Artevo strains. Distinct amplicons, corresponding to a specific ADO-FAR module, could be visualised by gel electrophoresis, enabling the number of host cells containing each module in a sample population to be counted.

RAMO, or CER1-ACR1. At t_2, this relative C_t value should increase in the reaction probing for seADO-AAR compared to pmADO-RAMO and CER1-ACR1, reflecting a proportionally greater increase in the prevalence of pTL2 seADO-AAR within the combined population. Comparing C_t values of each module relative to *dapB*, between

t_1 and t_2 samples, there should be a greater C_t downshift observable for seADO-AAR than for pmADO-RAMO and CER1-ACR1.

When calculating primer efficiencies for each primer pair, according to the method described by Kralik and Ricchi[277] and using the built-in standard curve function in Bio-Rad's CFX Maestro qPCR analysis software suite, the efficiencies of primers targeting seADO-AAR, pmADO-RAMO, and CER1-ACR1 modules were found to be above 100%, approximately 100%, and below 100%, respectively (*Figure 28*). This produced a biased result which, if only examining the sample of combined culture taken after 51 hours of growth, may appear to suggest enrichment of seADO-AAR. Knowing the efficiencies of available primers precluded an accurate comparative analysis of module prevalence between t_1 and t_2 samples, and knowing that screening a batch of new potential probes for suitable pairs would be time-consuming and possibly fruitless, it was decided to attempt the enrichment analysis using conventional colony PCR and gel visualisation.

Figure 27(b) outlines the process of using multiplexed PCR analysis to screen clonal colonies obtained from t_1 = 0h and t_2 = 51h samples of combined ISxJK004 LVA^+ pTL2 culture. While this approach is more laborious and less comprehensive than a qPCR analysis, being limited to a smaller sample size, it was thought that it would be more reliable and easier to troubleshoot. Unfortunately, the multiplex reaction appeared biased toward pmADO-RAMO, producing a bright band corresponding to this module even when an equimolar mix of all pTL2 plasmids was used as a control template. As with the qPCR analysis described above, these issues with primer pairs prohibited this approach from yielding insightful data.

Both conventional and quantitative PCR approaches described here were performed in the limited remaining experimental time toward the end of this project, restricting troubleshooting efforts and leaving a direct observation of sequence enrichment to future projects inheriting the Artevo system.

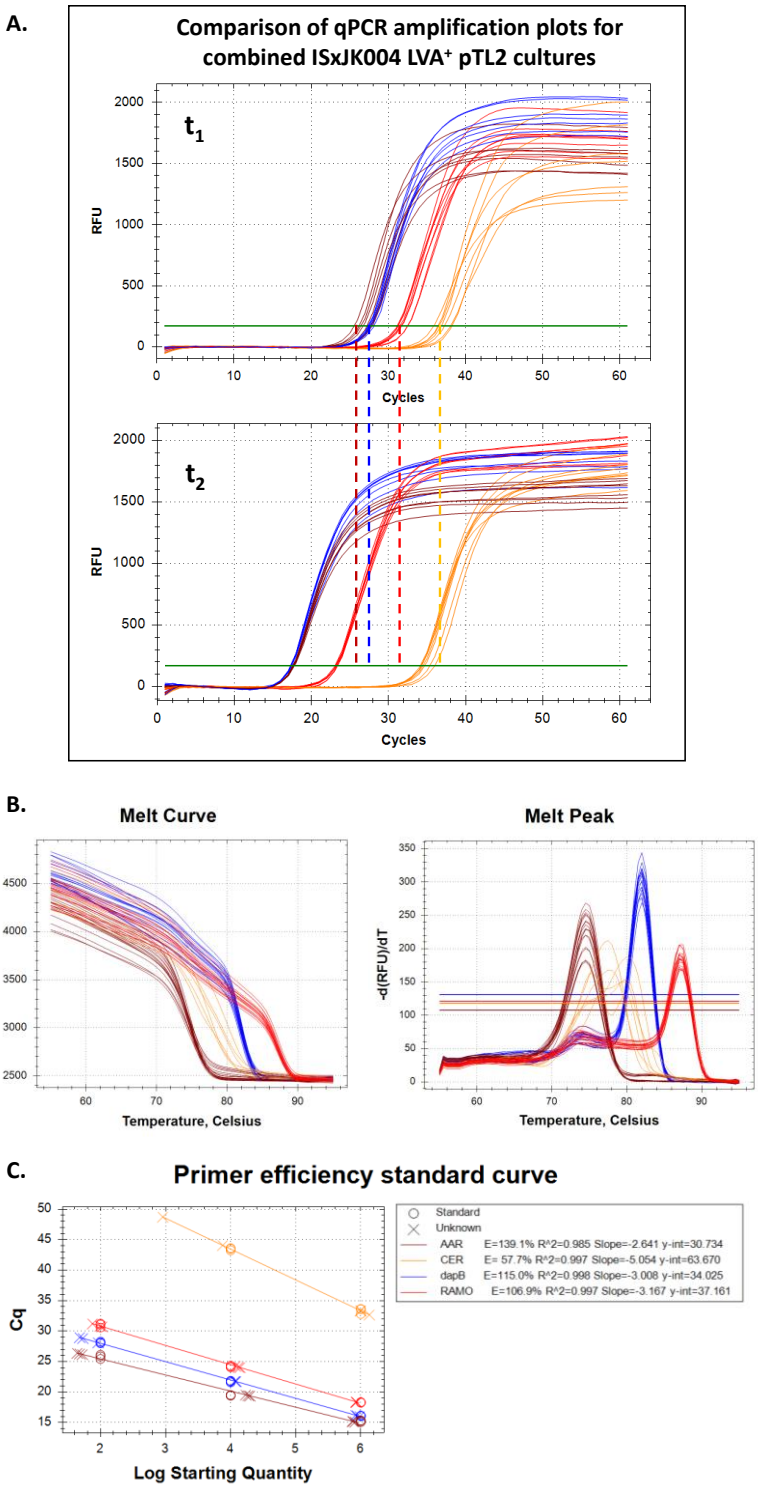


Figure 28. Part (a) shows a comparison between qPCR amplification plots generated by probing samples of combined ISxJK004 LVA$^+$ pTL2 culture, containing equal numbers of cells carrying each ADO-FAR module, as described in *Figure 27(b)*. Vertical dotted lines highlight the apparent C_t shift for each module between t_1 and t_2 (dark red = seADO-AAR, red = pmADO-RAMO, yellow = CER1-ACR1, blue = *dapB*). Part (b) shows melt data for each set of primer probes, illustrating the issues with probes targetting seADO-AAR and CER1-ACR1. Part (c) shows the standard curve generated for each probe set and calculated primer efficiencies.

Materials and methods

Chemicals, media, growth conditions, and equipment

Strains, reagents, media, growth conditions, nucleic acid handling kits, equipment, and other general microbial handling materials were all as previously described in earlier chapters. The only additional materials and equipment required to carry out work described in this chapter was a Bio-Rad CFX96 real-time PCR detection system and Bio-Rad iQ Supermix.

Alkane biosynthesis plasmids

pTL2 plasmids carrying the ADO-FAR alkane biosynthesis modules seADO-AAR, pmADO-RAMO, and CER1-ACR1 were generously gifted by Professor Ville Santala at Tampere University, Finland. An example construct map for pTL2 seADO-AAR can be found in *Figure 23*. Primers for conventional PCR and qPCR analysis of these plasmids were obtained from Merck.

Installing alkane biosynthesis machinery

Transformation of ISxJK004 LVA$^+$ with pTL2 plasmids was performed by natural transformation as previously described, with transformants selected on LB agar supplemented with spectinomycin. Positive transformants were re-streaked to remove background transforming DNA and screened using colony PCR, with primer probes specific for each ADO-FAR module. Positive transformants of ISxJK004 LVA$^+$ containing each ADO-FAR module were pre-cultured in LB broth with spectinomycin before being washed three times in PBS through consecutive centrifugation and resuspension, then normalised to an optical density (OD 600nm) of 1.00. These normalised pre-cultures were used to inoculate, in triplicate, sterile 250mL flasks containing 50mL of MMS media with spectinomycin, IPTG at 0, 50, and 100 μM, and mineral oil at 0 or 0.01% (v/v), prepared by sonication as previously described. A full list of strains and media conditions can be found in *Figure 25*.

A subset of flasks was incubated at 150rpm and 30°C for 327 hours, with the first instance of bacterial growth being observed in retained singlicates after 213 hours, as depicted in *Figure 25*. ISxJK004 LVA$^+$ pTL2 strains containing each ADO-FAR module were re-streaked onto MMS agar containing spectinomycin and clonal colonies probed using colony PCR to confirm the presence of each ADO-FAR module and the genomic

integration of *dapB* at the native *alkM* locus. Positively identified transformants were carried forward into the next experiments.

Preliminary growth analysis of alkane-producing Artevo strains

Newly-obtained Artevo strains – ISxJK004 LVA$^+$ pTL2 [seADO-AAR, pmADO-RAMO, or CER1-ACR1] – were pre-cultured for 48 hours in MMS media containing 50 μg/mL spectinomycin and mineral oil at 0.01% (v/v). These pre-cultures were washed three times in PBS as previously described and normalised to OD1. Normalised pre-cultures were used to inoculate 1mL of MMS media containing spectinomycin, IPTG at 0 or 50μM, and mineral oil at 0 or 0.01% (v/v), in duplicate in a clear 24-well plate. Media conditions were inoculated either with pure culture containing only one ADO-FAR module (orange, red, and brown lines) or a pre-mixed sample of all three cultures (blue line); see supplementary *Figure S.23*. The plate was sealed with a Breathe-Easy oxygen-permeable membrane from Sigma-Aldrich and incubated in a Tecan Spark plate reader at 30°C with constant shaking at 150rpm. Growth in each well was monitored by measuring absorbance (OD600nm) of each culture every ten minutes.

Flask-scale growth analysis of alkane-producing Artevo strains

ISxJK004 LVA$^+$ strains containing each pTL2 plasmid were pre-cultured in MMS media containing spectinomycin and mineral oil at 0.01% (v/v). Pre-cultures were washed and normalised as previously described and used to inoculate 50 mL of MMS media in sterile 250 mL flasks, in triplicate. Media contained spectinomycin, IPTG at either 0 or 50 μM, and mineral oil at 0 or 0.01% (v/v). Flasks were incubated at 30°C and 150 rpm. Flasks were sampled periodically, as illustrated in *Figure 26*, and OD600 measured to produce a growth curve. A Gompertz curve was fitted to raw absorbance data for each strain and media combination to produce *Figure 26(b)*, and growth rates were calculated manually for each individual replicate as previously described, using the linear portion of a plot of ln(OD) versus time.

Analysis of enrichment of high-performing ADO-FAR modules

Prevalence of each ADO-FAR module in a combined culture of ISxJK004 LVA$^+$ pTL2 strains at t_1 and t_2 was probed using qPCR and direct colony PCR. qPCR was performed in the format of an intercalating dye assay, using Bio-Rad iQ Supermix and probes specific for seADO-AAR, pmADO-RAMO, or CER1-ACR1, with all amplicons being approximately 100bp in size. The PCR programme was a typical 3-step reaction programme with melt curve, as recommended by the manufacturer. Primer efficiencies were checked using the method described by Kralik and Ricchi[277].

Multiplex colony PCR was performed on a small subset of test colonies prior to full-scale analysis, using forward primers specific to sections of either AAR, RAMO, or CER, and a reverse primer specific to *spec*R on the pTL2 vector backbone, as shown in *Figure 27*. Serial dilutions of ISxJK004 LVA$^+$ containing all three pTL2 plasmids, sampled at $t_1 = 0$ hours and $t_2 = 51$ hours, were plated onto LB agar with spectinomycin until individual colonies were observable. A subset of colonies was picked into colony PCR mastermix (made using DreamTaq Green PCR Mastermix from Thermo-Fisher) containing the multiplexed probes. Primer pairing and internal complementarity of the multiplexed probes was checked using the Primer3 web suite[278].

Discussion

In Chapter 4, the Artevo strain (ISxJK004 LVA$^+$) was shown to exhibit a graded growth response to exogenous alkane concentration, with adjustable stringency. Here, transformation of the Artevo strain with pTL2 plasmids carrying various ADO-FAR alkane biosynthesis modules allowed that strain to produce endogenous alkanes to drive the host growth response. This response is proportional to the alkane productivity of each ADO-FAR module – seADO-AAR, pmADO-RAMO, and CER1-ACR1 – which was verified first by Lehtinen *et al.*[248], who produced the plasmids for this study, and then again in this work using the alkane biosensor constructed by Zhang *et al.*[162] (see *Figure 24*). Growth rate variation between Artevo strains containing each pTL2 plasmid was small but significant, implying that enrichment of the most productive module – seADO-AAR – would occur in a combined culture of all three. Enrichment of this module was not directly observable by qPCR or multiplex colony PCR because of experimental issues, but the methodology for investigating enrichment has been detailed and will be completed in future work.

The relationship between alkane productivity and growth rate in the Artevo strains is difficult to quantify because no chemical analysis of these cultures has yet been performed to quantify specific alkane yields from each module. Lehtinen *et al.* report specific alkane yields only for the highest-performing ADO-FAR module, seADO-AAR, with fed-batch cultivation on acetate producing 21 $\mu g.L^{-1}.h^{-1}$ in a 1L benchtop bioreactor[248]. The difference in feedstock (acetate, versus succinate in this work) and scale (1L stirred bioreactor versus 250mL shaken flask) means alkane productivity data for seADO-AAR may not be directly applicable to this work and should be reanalysed in the context of the Artevo system; see the section on short-term future work in Chapter 6.

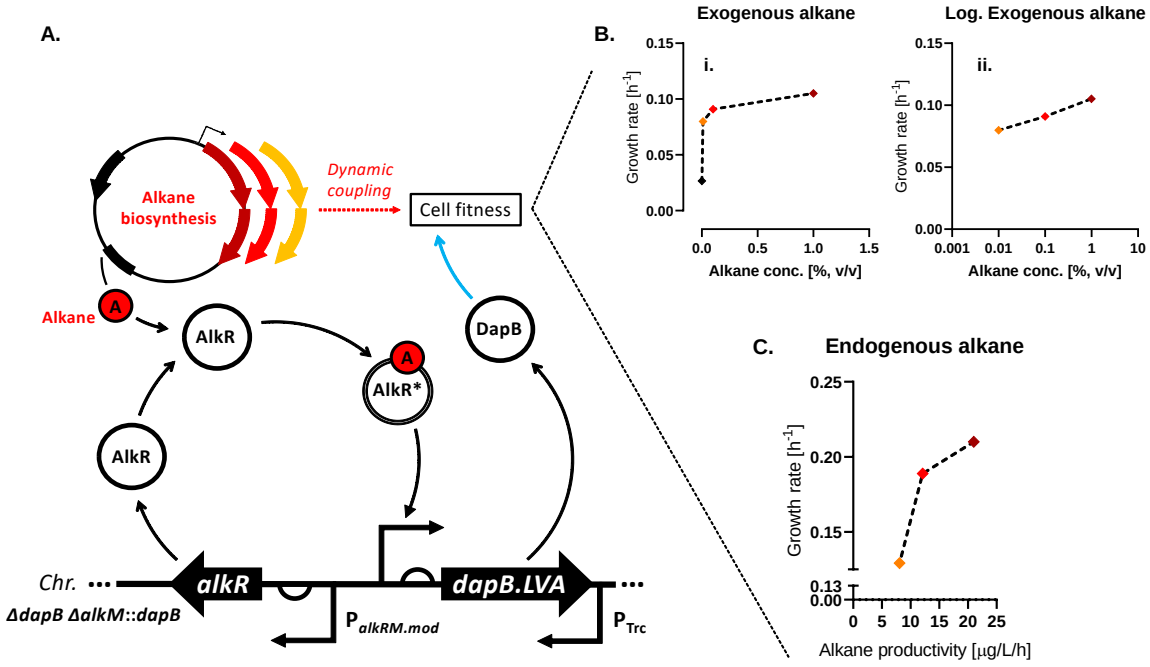


Figure 29. Calculated relationship between growth rate of Artevo strains (ISxJK004 LVA+) and either exogenous alkane concentration or endogenous alkane productivity. Part (a) shows the Artevo circuit linking productivity of alkane biosynthesis gene modules to a survival output, as detailed in earlier chapters. Part (b) depicts growth rate of this Artevo strain as a function of exogenous mineral oil concentration, on linear (i.) and logarithmic (ii.) scales. Part (c) estimates the alkane productivity of each ADO-FAR module used in this work, based on the absolute productivity reported for seADO-AAR and the relative fluorescence reported for each module by Lehtinen *et al*[245], and plots those productivities against growth rate of Artevo strains carrying each module.

For interest, a simple comparison was performed assuming consistent behaviour of modules across culture conditions and a linear relationship between intracellular or extracellular alkane concentration and growth rate in Artevo strains; this comparison is shown in *Figure 29*. The limitations of such a crude mathematical analysis in describing what is a complex, multifactorial phenomenon must be acknowledged, as should the small size of the data set supporting the analysis, though some effects can be inferred. The absolute difference in growth rate of Artevo strains driven by exogenous mineral oil (*Figure 22*) and by endogenous mineral oil (*Figure 26*) suggests that intracellular alkanes have a greater impact on growth than extracellular alkanes, validating an early assumption made when considering the potential for 'cheater' phenotypes to arise in this system. This is somewhat contrary to what is suggested by *Figure 24*, which shows the luminescence signal from the Zhang *et al.* alkane biosensor strain peaking higher in the presence of supplemental mineral oil than when the same strain is equipped with any of the pTL2 plasmids. The response rate from these biosensor strains is more rapid in the presence of natively produced alkanes however, peaking after 10 rather than 20 hours, which may explain why producing alkanes endogenously in Artevo strains results in a higher growth rate than what is observed when alkanes are supplied exogenously

to the same strain. This phenomenon may be attributable to transport of long-chain hydrocarbons across the cell membrane; high levels of exogenous alkane provide a greater driving force for diffusion and higher maximum concentration of intracellular alkanes after equilibration, but accumulation of this intracellular pool is limited by the rate of diffusion, which is not the case when alkanes are manufactured endogenously. Deletion of the integral-membrane hydrocarbon hydroxylase *alkM*, which has been carried out in both the Zhang *et al.* biosensor and in Artevo strains in this work, has been shown to abolish alkane utilisation as a carbon source in ADP1[233], though it is unclear whether this gene plays an active role in transporting alkanes into the cell comparable to that of type-IV pilus-related systems or secreted proteins with emulsifying activity, which pseudosolubilise alkanes in aqueous solution[279–281].

Supplemental mineral oil in the experiments that produced *Figure 25* and *Figure 26* was added to provide an initial boost to growth rate that would be constant across all three pTL2 Artevo strains, as it was theorised that initial conditions where expression of the rescue gene *dapB* is driven only by endogenous alkane may be too strenuous for the cells and prohibit growth. Supplemental mineral oil should provide a 'background' growth rate for all strains, with further relative fitness gains from native alkane production manifesting as a growth rate improvement above this baseline. Local environmental alkane concentration will be constant across pTL2 Artevo strains, regardless of the ADO-FAR module they carry, but intracellular alkane concentration will be greater in strains carrying high-performing modules. That is, the positive selection pressure is the same on all strains in the absence of native alkane production, with any internal biosynthesis of alkanes providing additional impetus for growth.

The lag phase reduction observed between experiments in *Figure 25* and *Figure 26* (and supplementary *Figure S.23*) is likely due to adaptation of the pTL2 strains to minimal media, these strains having initially been generated on rich media immediately prior to inoculating the first flask experiment. In addition to the lag phase being far longer in *Figure 25* than in *Figure 26*, the pattern of growth in both experiments is slightly different; in the latter experiment, the Artevo strain containing pTL2 CER1-ACR1 exhibits a shorter lag phase than strains containing seADO-AAR or pmADO-RAMO, though growth rates of each strain still conform to the expected pattern. While normalising inoculation ratios across cultures using optical density or cell counts typically offers an acceptable level of precision, the Artevo system is, at this stage, sensitive to initial conditions and reliant upon differences in absolute alkane productivity in ADO-FAR modules, which may themselves be low. It is encouraging to observe that, in this case, the Artevo system is robust enough to produce a growth

rate difference that can overcome small biases in initial conditions, selecting the seADO-AAR module over both pmADO-RAMO and CER1-ACR1.

The higher apparent error in *Figure 22* compared to *Figure 26* may be a product of local microenvironments with differing alkane concentrations, stemming from the limited solubility of mineral oil in aqueous solution. More likely though, it is due to replicates in *Figure 26* being only pseudo-biological, closer to technical replicates, having been derived from a single clonal stock and pre-culture split into separate culture flasks. The experiment that produced *Figure 22* utilised true biological replicates from distinct lineages, stemming from separate transformation events.

Artevo strains containing pTL2 plasmids were initially tested using a 24-well plate to leverage the automatability to this experimental format. Key inflection points, such as initiation of log phase, peak culture density, and the start of death phase, may be missed when manually sampling larger culture vessels. While the drawbacks of 96-well plates have been previously discussed and demonstrated (*Figure 21*), especially in relation to the Artevo system, 24-well plates represent an acceptable middle-ground between plate and flask, allowing for improved agitation under continuous shaking and frequent, automated measurements. The automation aspect facilitates collection of much larger data sets that can balance mechanical drawbacks of using a small-scale system; for example, area scans, where multiple, spatially-separated measurements are made across the well, can be used to account for potential non-homogeneity in the culture volume. Reduced well numbers limit either media conditions or replicates and, in the case of *Figure S.23*, necessitated a sample size of just two, causing untenable error in some conditions towards the end of the experiment. Manual sampling from culture flasks provides data more representative of the true behaviour of the Artevo system in its final form, which should be in a chemostat, but requires data points be fit to a logistic growth curve. The Gompertz model used to fit data in *Figure 22* and *Figure 26* has been extensively used in bacterial growth analyses[282] and is appropriate here as growth in this system is thought to be slower initially, when the intracellular suite of alkane biosynthesis machinery is still being established.

Earlier in this project, during initial tuning attempts, the first prototype Artevo strain (002*) was transformed with an alkane biosynthesis module to observe whether there would be any response at all to native alkane production (supplementary *Figure S.8*). The modules chosen were the cyanobacterial *luxCDE* operon and ADO, both from *Synechococcus elongatus*. This approach has been described in detail in Chapter 4. This module elicited a binary growth response in the host 002* strain, and it was planned to reattempt native alkane biosynthesis using this module after the Artevo strain had been

sufficiently tuned. By the time tuning was completed, the *luxCDE* operon had fallen somewhat out of favour for microbial production of alkanes, with ADO more commonly being used alongside other fatty acyl-CoA reductases, with encouraging success[115]. The ADO-FAR modules used in this work were by then available in a format compatible with, and verified in, *A. baylyi* ADP1, removing the need to characterise their functionality prior to use. Testing the Artevo system with a variety of genetically distinct alkane biosynthesis modules, as opposed to the same module expressed at different levels, is more representation of how the final system is intended to operate; selecting and enriching specific genes from a diverse library of sequences.

Attempts to directly observe this enrichment, summarised in *Figure 27*, were unsuccessful because of experimental issues and limited time for troubleshooting. The qPCR approach was hampered by unsuitable primer probe efficiencies, meaning prevalence of each module could not accurately be compared using this method. Qualitatively, *Figure 28(a)* does appear to show an increase in absolute copy number (indicated by reduced C_t values) of seADO-AAR, pmADO-RAMO, and CER1-ACR1, as well as the *dapB* reference gene, from t_1 to t_2, as well as relative enrichment of seADO-AAR over the other ADO-FAR modules. Evident amplification bias, clear in the t_1 sample – which was inoculated equally with each pTL2-containing Artevo strain – precludes a meaningful mathematical comparison of module prevalence between t_1 and t_2 cultures. A larger probe library has been obtained and will be screened in future work to find a suitable combination of probes with comparable efficiencies, which can then be used to compare relative prevalence of seADO-AAR, pmADO-RAMO, and CER1-ACR1 at various times in a combined culture of ISxJK004 LVA^+ pTL2 strains.

An intercalating dye assay was chosen over a 5' nuclease assay because of extended production lead times for hydrolysis probes required for the latter; materials and primers for the dye assay were readily available. The dye assay is non-specific for distinct templates and so cannot be multiplexed, meaning a large duplication of labour is required to test for all three modules. A multiplexed approach was attempted in the form of direct colony PCR, see *Figure 27(c)*, though initial results on a small subset of test colonies suggested bias in the reaction toward pmADO-RAMO, necessitating a redesign of the primer probes in future work.

The experiment may also be adjusted to allow dynamic monitoring of pTL2 plasmid prevalence in combined culture using fluorescent reporter genes. Each pTL2 plasmid could be equipped with a different fluorescent reporter, such as *gfp*, *rfp*, and *yfp*, which could be monitored *in situ* in a batch reaction, for example using the previously mentioned benchtop Chi.BIO reactor system. Alternatively, the combined culture

could be periodically sampled and the changing distribution of fluorescent phenotypes within the population analysed by flow cytometry. Employing fluorescent reporters in this way is limited by the number of chassis-compatible, optically-distinct probes available and the detection range of sensors in the chemostat or flow cytometer, which places hard limits on the size of the pool of alkane biosynthesis modules which can be investigated. Reporter genes may also inconsistently burden host strains and skew the growth response, though this effect is likely to be minimal, if present at all. If modules can be appropriately marked, Raman analysis at the single cell level may also prove useful for detecting which module is present in individual cells. As single cell Raman analysis allows for investigation of intracellular concentrations of various molecules, it should also enable a comparison of alkane yields between cells.

While enrichment of seADO-AAR is, at this stage, implicit based on relative growth rates of each pTL2 strain, I am confident it will be directly observable by qPCR after experimental troubleshooting, and is not a misbehaviour of the Artevo system. I hope this troubleshooting can be performed by inheritors of the Artevo system, whose work will expand upon the achievements of this project and eventually utilise the Artevo system in new and exciting ways. The following chapter provides a guide for experimentalists hoping to further develop the Artevo system, and speculates on the future of this work.

Chapter 6

Future work

This chapter is future-facing. It discusses immediate future work to continue developing the directed evolution technology described in this project and then looks ahead to the future of the overall field. The first half of the chapter presents specific sequential steps to further develop the Artevo system, aiming either to generate additional data supporting existing conclusions, further tune aspects of Artevo system, or continue practical testing of the system. The second half of the section presents broad and directional suggestions for future work, listed in no particular order of priority, including possible strategies for expanding the capabilities of the Artevo system, exploring new applications, or designing auxiliary systems to improve the evolution process. Some later ideas are explored in greater detail following a discussion of library generation in for use in directed evolution, with gene circuit designs being included for a potential genome-wide directed evolution system, involving *in vivo* diversity generation, that captures beneficial mutations occurring outside the alkane biosynthesis module. A discussion of continuous directed evolution and handles for future tuning of the Artevo system closes out this chapter.

Part I: Next steps

Enrichment analysis

Both forms of enrichment verification described in Chapter 5 – colony PCR and qPCR on combined cultures of ISxJK004 LVA$^+$ pTL2 [seADO-AAR/pmADO-RAMO/CER1-ACR1] – should be completed as an immediate next step to continue this project. While the observed difference in growth rate between ISxJK004 LVA$^+$ containing seADO-AAR, pmADO-RAMO, and CER1-ACR1 modules implies enrichment would occur in combined culture, direct observation of enrichment of the seADO-AAR module over pmADO-RAMO and CER1-ACR1 would further validate the system.

For the colony PCR approach, more primer combinations should be trialled to obtain a working multiplexed probe mixture that can produce distinct, unique bands for each alkane biosynthesis module. The *in silico* analysis tool – Primer3[278] – used to estimate primer behaviour in the previous attempt at multiplexed colony PCR did not produce successful results in the first instance, but should nevertheless be used again to avoid glaringly inappropriate primer combinations.

Similarly, for the qPCR approach, a library of primer pairs will need to be tested to obtain three sets of probes with comparable amplification efficiencies. Many standard protocols exist for qPCR that can be followed, but Illumina offer a comprehensive and practical guide[283] that can be useful alongside the more detailed methods described by Kralik and Ricchi[277].

Direct detection of biosynthesised alkanes

Alkane productivity in Artevo strains containing pTL2 – that is, ISxJK004 LVA$^+$ pTL2 [seADO-AAR, pmADO-RAMO, or CER1-ACR1] – should be quantified by direct analytical methods to confirm the relative alkane yields from each module and the specific hydrocarbon species being produced. A gas chromatography service has been previously employed by this group to quantify alkane production from *E. coli*, though inconsistencies in the hexane extraction process were suspected to obfuscate what may have been very low (μg/mL range) alkane yields. Until recently – after the conclusion of the experimental aspects of this thesis – in-house alkane detection by gas chromatography was impossible, as the laboratory lacked a chromatography column suitable for this work.

Use of the Zhang *et al.* ADPWH_alk biosensor[162], which has been extensively referenced in this work, is similarly limited, requiring extraction of alkanes from host cells if they are not excreted during production. Cell lysis by sonication, heating, or other methods, followed by separation of cell debris and growth of ADPWH_alk on recovered supernatant, was trialled in this work for analysis of produced alkanes; alkane yields were either too low to produce a measurable response from the biosensor or the signal was masked by cross-sensitivity of the biosensor to other molecules liberated during cell lysis. Raman spectroscopy, a technology available in the Huang laboratories, may be useful for alkane detection, especially as it can be used for single-cell analysis to quantify intracellular metabolite concentrations[284].

Further tuning and characterisation of the Artevo system

The Artevo system response to intracellular alkanes should be further characterised and tuned to improve system dynamics, especially the degree of growth improvement observed per unit alkane produced. The fold-change in fluorescence, indicative of alkane concentration, observed between seADO-AAR, pmADO-RAMO, and CER1-ACR1 modules by the Santala group using their twin-layer biosensor[248] is much greater than the fold-change in growth rate observed when these modules are installed in ISxJK004 LVA^+ (see *Figure 23* and *Figure 26*). The growth rate difference is significant but small, only ~10% between ISxJK004 LVA^+ containing pTL2 seADO-AAR or pTL2 pmADO-RAMO. Ideally, an almost twofold increase in alkane productivity – as expected between seADO-AAR and pmADO-RAMO modules, for example – would produce a larger difference in growth rate. A numerically comparable response, such as a twofold increase in alkane production causing a doubling of growth rate, could be targeted, although the response curve would necessarily plateau as cell growth becomes limited by other factors. Achieving this effect may require both extending the dynamic range of rescue gene expression, through tuning of the relative transcriptional interaction between sense and antisense promoters regulating *dapB*, and potentially altering the activity of DapB protein itself to reduce its restorative impact on growth rate. As mentioned in Chapter 4, site-directed mutagenesis of DapB by Scapin *et al.*[235] successfully reduced the catalytic activity of this protein by 135-200 fold (H159A/Q) and 625-830 fold (K163A/C/Q).

It is difficult to perform an insightful mathematical analysis of the relationship between alkane production and growth rate in Artevo strains, based only on the limited data available. With growth rate data for only three biosynthesis modules, the resolution of any response curves generated (see *Figure 29*) is too low to make convincing conclusions about the shape of the curve. Further characterisation of the system would be insightful

here; by testing a wider range of Artevo strains transformed with more pTL2 plasmids, containing all the modules investigated by the Santala group, a more representative response curve could be generated that would inform future tuning efforts.

To continue tuning the Artevo system effectively, the expression system governing antisense transcriptional interference of the rescue gene *dapB*, or the expression system regulating alkane biosynthesis genes on the pTL2 vector, must be replaced. Limitations on time and available parts necessitated using cross-sensitive IPTG-inducible promoters on both the ISxJK004 LVA$^+$ genome and the pTL2 vector. These handles on downregulation of *dapB* and upregulation of the alkane biosynthesis module must be decoupled to fully explore the capabilities of the system in its current form. Alternative inducible expression systems with suitable profiles exist for use in *A. baylyi*, including the anhydrotetracycline-inducible TetR system[156] and the salicylate-inducible SalR system[241].

Further tuning efforts should include extending the dynamic range of *dapB* expression, consequently achieving more precise control over the growth response of Artevo strains in the presence of alkanes. As discussed in Chapter 4, transcriptional interference as a strategy for regulating gene expression has not yet been fully explored in the context of the Artevo system, though promising results have been obtained from just this shallow exploration. There is much space to improve the antisense control of *dapB* expression, which could be refined using a combination of strategies found in the literature, including adjusting relative promoter strength, interpromoter distance, promoter sequence content, and relative processivity of the opposing RNAP-ribosome complexes[261–265]. Complete suppression of *dapB* via transcriptional interference may be achievable by increasing the strength of the antisense promoter, though the Trc promoter used in this thesis is among the strongest available for *A. baylyi* that has been characterised, so this strategy may be limited without additional promoter engineering.

Finally, in preparation for *in situ* use of the Artevo system with a full-sized diverse plasmid library, Artevo strains should be transformed *in situ* with an equimolar mixture of pTL2 plasmids carrying seADO-AAR, pmADO-RAMO, and CER1-ACR1 modules, then grown in selective conditions to observe whether enrichment of seADO-AAR is still achievable under these conditions. Transformation of many modules simultaneously during practical application of the Artevo system will necessarily reduce the starting prevalence of individual modules within the population, potentially impacting the precision and efficiency with which high-performing modules are amplified.

Library generation and screening of diverse modules

The pTL2 plasmids used in this project, containing the seADO-AAR, pmADO-RAMO, and CER1-ACR1 alkane biosynthesis modules, were chosen to demonstrate the Artevo system as a proof-of-concept. In practice, the Artevo system must interrogate a diverse library of sequences encoding randomly mutated alkane biosynthesis genes, and these libraries would typically be constructed *in vitro* and transformed into the Artevo strain (the possibility of *in vivo* diversity generation is discussion later in this chapter).

Diverse genetic libraries must be constructed and characterised systematically to avoid bias or duplication of variants, and will in general adhere to the following experimental workflow:

(i) Selection and characterisation of the vector and expression systems based on size of the intended insert[285].

(ii) Synthesis of diverse gene blocks by hybrid biological or purely chemical methods, such as error-prone PCR[286] or site saturation mutagenesis[287], or column- or array-based oligonucleotide synthesis using phosphoramidite chemistry[288].

(iii) Cloning of diverse cargo sequences into the vector system, by Gibson assembly[289] or another high-efficiency method[290,291].

(iv) Transformation of the assembled vector pool into the host platform.

Library generation itself is a significant undertaking, reflective of the massive underlying sequence space the library aims to partially cover. Generating a library of randomly mutated alkane biosynthesis modules, using the seADO-AAR module as a starting point, should be the next step in this project in parallel to further tuning and characterisation of the Artevo system. Those two workstreams can then be joined and the Artevo system used to isolate high-performing mutants from this library.

Advanced Artevo strains developed from further tuning of ISxJK004 LVA$^+$ should be transformed with pooled variants and the dominant genotype identified after a period of growth. Strategies such as sequence barcoding can aid in later investigation of specific sequences[292], which could then be used as a starting point for subsequent rounds of directed evolution.

Investigation of high-performing modules

Alkane biosynthesis modules enriched by the Artevo system should be investigated for underlying causative mutations, particularly structural mutations within ADO and AAR. ADO is a notorious bottleneck in alkane biosynthesis and it is hoped that a directed evolution approach will improve its activity toward its thermodynamic limit. Structurally, AAR has been reported to form an acyl tunnel upon substrate binding, guiding the ligand toward the ADO-AAR interface, which itself opens during association of these two enzymes[293]. While Gao *et al.* identified the importance of reversible binding between ADO and AAR – so as not to limit AAR activity with the low turnover of ADO[293] – if the individual activity of ADO can be increased by directed evolution, then the stable AAR-ligand-ADO complex may evolve toward becoming a hybrid enzyme. It would be interesting to try and produce such an enzyme, or just investigate its viability, using the Artevo system, by tethering ADO and FAR via a flexible GS linker and using that structure as a starting point for directed evolution. Building on the work of Andre *et al.*, further hybrid enzymes might be created by fusing catalase to ADO, which clears the local microenvironment of inhibitory hydrogen peroxide (H_2O_2)[294].

If promoters and ribosome binding sites are included in the region of diversity in the mutant library, modules that favourably adjust the relative expression of ADO and AAR may also be selected by the Artevo system, providing insight into optimal resource allocation within the cell. As will be mentioned in the upcoming section, it may also be of interest to tag ADO and AAR for compartmentalisation, sequestering some or all alkane biosynthesis activity from the bacterial cytoplasm. Enrichment of such tagged proteins by the Artevo system may provide insight into optimum cellular microenvironments for alkane biosynthesis and help inform future optimisation of bioprocess conditions.

Part II: Long-term plans

Growth of Artevo strains on various feedstocks

Options for upstream fixation of CO_2 into substrates utilisable by *A. baylyi* are not experimentally explored in this project. Instead, succinate is used throughout as a favourable carbon source to support growth of the Artevo strains. Practically, when moving toward a fully carbon-neutral system, feedstocks such as bioelectrosynthesised acetate, precursor compounds produced by photosynthetic marine microalgae, or liquid

extracts from CAM plants such as *Opuntia ficus-indica* may be preferable. The last of these could be assessed readily in future work, as other members of the Huang lab work with *Opuntia* sp. and the relevant biomass is available. If Artevo strains can utilise *Opuntia* biomass for growth, this may represent a link to practical industrial applications of the Artevo system beyond sequence discovery; a basis for a versatile biorefinery with ready access to a vast atmospheric carbon reservoir.

Additionally, if an Artevo strain optimised for alkane production could be grown in coculture with CO_2-fixing microbes, the resulting microbial consortium would represent a powerful living catalyst for conversion of atmospheric carbon to energy; an integrated biological platform for sustainable manufacturing of carbon-neutral liquid biofuels. Of primary interest is photosynthetic marine algae, as *A. baylyi* is saline-resistant and can grow in seawater[162], reducing the terrestrial footprint of the systems involved in energy- and carbon-gathering activities and maximising power density of the produced biofuels.

Biosensing and synthesis of non-linear hydrocarbons

AlkR protein has been observed to induce *alkM* in the presence of medium-chain-length alkanes, heptane to undecane, and long-chain alkanes, dodecane to octadecane, as well as 1-dodecene[196]. Future work in this project should test the sensitivity and specificity of AlkR for various branched, cyclic, and aromatic hydrocarbons, in pursuit of expanding the pool of valuable products that may be targeted by the Artevo system in something close to its current form. It may also be worthwhile to engineer or evolve the AlkR protein for improved detection of these alternative hydrocarbons; producing fuels from this system compatible with aviation turbine engines will require effective biosynthesis of branched and cyclic hydrocarbons. In parallel, biosynthetic pathways for production of these alternative hydrocarbon species, or useful precursor molecules, should be characterised in *Acinetobacter baylyi* ADP1-ISx.

Evolving *alkR* to detect diverse valuable metabolites

Directed evolution of AlkR protein could be performed in pursuit of increased sensitivity and specificity to other potentially valuable hydrocarbon products or related molecules, including terpenes and isoprenes. There is precedent for this strategy; sensitivity of the AlkSp regulator system from *P. putida* has previously been evolved to exhibit increased sensitivity to short-chain alkanes[253]. A similar approach could be used for AlkR regulation in the Artevo circuit or, if the circuit can be properly tuned, the system itself could be used to evolve specific traits in AlkR protein.

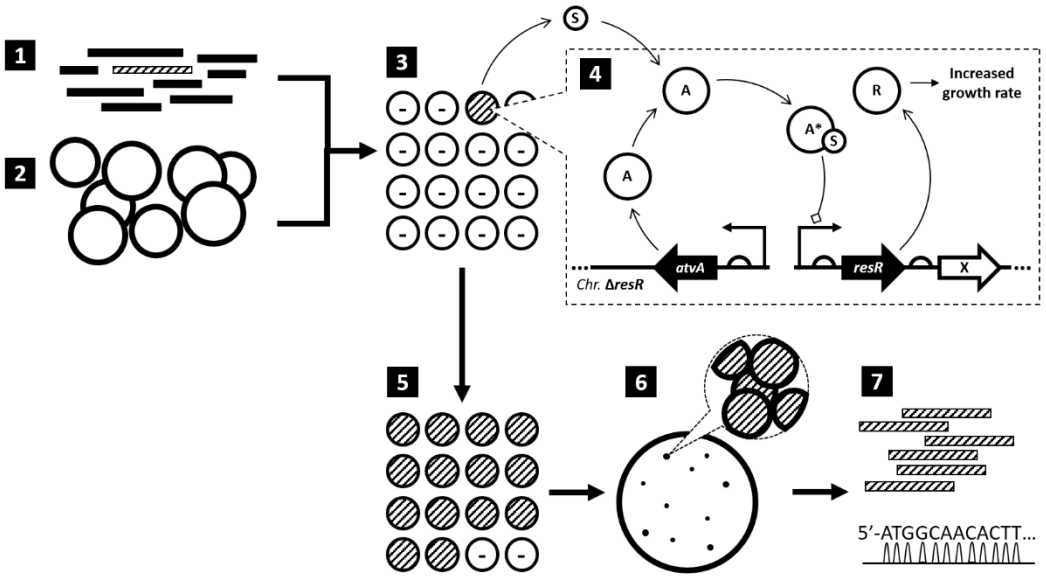


Figure 30. Use of the Artevo system for functional metagenomics and bioprospecting. A metagenomic DNA library (1), that harbours some unknown DNA of interest encoding a useful protein, is introduced to cells equipped with the Artevo circuit (2), shown here in generalised format (*atvA* = activator gene, *res*R = rescue gene, X = reporter gene, S = substrate, A = regulator protein, R = rescue protein). The resulting population (3) contains some cells expressing the DNA of interest and producing desired chemical (S), which activates regulator proteins (A) expressed by the Artevo circuit (4). Cells producing (S) can express rescue gene (*res*R) and gain a growth advantage, eventually dominating the population after many generations (5). This population, consisting mostly of cells that have incorporating the DNA sequence of interest, can then be screened (6) and the DNA isolated and sequenced (7). In this way, useful DNA sequences may be isolated and amplified from metagenomic samples.

Carbon flux manipulation

Pathways potentially competing with alkane biosynthesis, such as wax ester synthesis in *A. baylyi*, could be removed to direct carbon flux towards desired hydrocarbon metabolites. Further carbon flux manipulation could include over- or under-expressing key pathway enzymes to minimise concentrations of toxic metabolites, such as fatty aldehydes. While the recombinant deletion strategy used throughout this project is suitable for this, setting up a CRISPR system in *A. baylyi* chassis would potentially allow parallel silencing or removal of multiple genes for a streamlined optimisation workflow. A novel *A. baylyi* ⊿*dapB* mutant equipped with CRISPR machinery would also represent a versatile chassis for synthetic biology beyond the applications in this project.

Life cycle and power density analysis

The potential power density (energy yield per unit land area required for production), cost, life cycle, and wider impacts of fuels produced using strains derived from the

Artevo system should be considered in greater detail and compared to other biofuels. Various upstream carbon-fixation mechanisms and possible feed streams that may be leveraged (*Opuntia* sp. biomass, marine microalgae, bioelectrosynthesised acetate, existing energy crops, refined sugars, waste streams, industrial effluent streams, etc) should also be assessed to paint a clearer picture of the true carbon neutrality of microbially derived alkanes. Such an analysis either requires extensive development of the underlying Artevo system to provide input data on yields and optimal growth conditions, or a reverse approach whereby boundaries for system performance can be set using net carbon neutrality as the goal.

Use of the Artevo system in functional metagenomics

Effective tools for functional metagenomics are currently lacking[295]. If correctly tuned, the Artevo system may have applications as a bioprospecting tool, as illustrated in *Figure 30*. The system may be able to isolate and enrich useful new DNA sequences from diverse environments and access immense ranges of novel drug candidates and industrially useful enzymes.

Attenuation of Artevo strains to high oil concentrations

Mineral oil concentrations of 1.00% (v/v) or above far exceed what is currently biologically achievable and may inhibit cellular processes, limiting the peak biosynthetic performance selectable by the Artevo system. Industrially relevant synthesis of alkanes may require greater tolerance, so attenuation of Artevo strains to high oil concentrations should be performed by repeated growth in oil-rich substrate. An examination of oleaginous bacteria for features facilitating their growth in high-oil substrates may inform these attenuation efforts[123,296].

Compartmentalisation of alkane biosynthesis

Spatial separation of activities within a cell may help alleviate issues with toxicity of intermediary metabolites and unfavourable reaction conditions for certain enzymes, such as ADO. Previous work in this area has identified the genes necessary to construct intracellular microcompartments and tag proteins for transfer to these locations, potentially representing a way to create ideal microenvironments for each individual reaction in a biosynthetic pathway[297]. This may help overcome the thermodynamic and metabolic disadvantages involved in synthesising highly-reduced molecules such as long-chain alkanes.

Evolution of optimal operating conditions for ADO

ADO operates optimally in microaerobic conditions and requires iron[141], but operation of the Artevo system within low-iron or fully aerobic conditions may enrich mutant ADO compatible with operating conditions that are more industrially applicable. The Artevo system should enrich whichever mutant sequences are most active within the greater system context, which includes the type of feedstock, nutrient availability, osmolarity of the growth media, dissolved oxygen content, pH, temperature, and more. The background environment also impacts growth of the host Artevo strain, and consequently enrichment of the module within. While the limits of the system have not yet been fully established, it may not be impossible to evolve enzymes like ADO in unexpected directions, requiring different metallic cofactors or acquiring improved thermostability, for example.

Alternative hosts for alkane biosynthesis

Alternative hosts for eventual application of modules discovered by the Artevo system should be considered. While *Acinetobacter baylyi* is a robust environmental microbe exhibiting industrially favourable characteristics – as well as the natural competence central to *in situ* continuous directed evolution, which will be discussed later in this chapter – other microbes may be better suited as eventual hosts for alkane biosynthesis modules discovered by the Artevo system. In particular, oleaginous yeasts such as *Yarrowia lipolytica* have been used to produce high titres (in the gram per litre range) of hydrocarbon[123,298–300]. It should be noted that modules evolved in the Artevo system will necessarily be optimised for use in *Acinetobacter baylyi*, and their use in alternative chassis will be limited by the orthogonality of the module.

As a counterpoint, even strains that are not naturally predisposed toward metabolising hydrocarbons, such as *Escherichia coli*, can produce titres in the gram per litre range. Fell *et al.* reported 2.54 g/L of long-chain alkane and 12.5 g/L of fatty alcohol produced in a fed-batch bioreactor, the highest reported yields to date[130]. The approach used by Fell *et al.* leverages *in silico* modelling to inform rational adjustments to carbon flux within the cell, mainly within the pentose phosphate pathway and lower half of glycolysis. The effectiveness of *in silico*-based approaches is of course dependent on the completeness of the model, whereas *in vivo* approaches are complete by their very nature, being wholly housed within a functional organism. The success of Fell *et al.* is perhaps indicative of the value of working within a well-characterised chassis, such as *E. coli*, but probably more so the value of hybrid approaches to bioprocess improvement, combining some black-box discovery platform – be it a computer model or an *in vivo* directed evolution system – and rational engineering.

In vivo diversity generation with variable mutation rate

A. baylyi's close pathogenic relative, *Acinetobacter baumannii*, can enter a 'hypermutator' state for rapid generation of genetic diversity in response to antibiotic challenge[301]. If this ability can be reproduced in the Artevo system, it may represent a way to generate sequence diversity *in vivo*, directly within the evolving microbial population, rather than upstream during the library generation process. This may also practically accelerate the evolution process by eliminating sequential rounds of *in vitro* diversity generation. Using the *alkRM* biosensory system, it may be feasible to link intracellular alkane concentration to a mutational actuator such that mutation rate increases when alkane production is low and stabilises when it is high. A similar system (feedback regulated evolution of phenotype, FREP) for evolving production of tyrosine and isoprenoids in *E. coli* has previously been reported by Chou and Keasling[192]. Actuators in *A. baylyi* could include expression of an error-prone DNA polymerase encoded by *dinB*[302], or inactivation of mismatch repair gene *mutS*, which causes the hypermutator phenotype in *A. baumannii*[301]. While approaches that increase basal mutation rate across the whole genome are well established[303], they have the drawback of potentially disrupting rescue and biosensor genes driving the enrichment process in Artevo strains, so a strategy for preserving core genes in the evolution circuit would need to be developed. These specific, high-priority, regions of DNA that are critical for system function – those that maintain evolutionary pressure towards enhanced alkane production – must be insulated from general increases to mutation rate or otherwise continuously repaired by reintegration of fresh sequences, the latter of which may be possible due to the natural competence of *A. baylyi* but is impractical.

In the prototype version of the Artevo system built during this project, diversity of alkane biosynthesis sequences is intended to be generated upstream, *in vitro*, during library construction, using established methods[304]. It was postulated early in the project that xenonucleic acids (XNAs) – synthetic nucleic acid analogues with a different sugar backbone to naturally occurring nucleic acids, and that can be incompatible with natural polymerases but recognised by engineered polymerases[305] – may be a useful tool in fully segregating genetic elements driving the evolution process from mutations benefitting alkane biosynthesis. The latter may occur within the alkane biosynthesis modules themselves or elsewhere on the genome, having an indirect benefit such as adjusting carbon flux to minimise pools of toxic intermediates. Practically, XNA technology is not advanced enough to facilitate this approach and natural polymerases are typically promiscuous enough to readily accept modified nucleotides[306], though examples of

some polymerases requiring engineering to incorporate XNA substrates lends legitimacy to the potential of the strategy[307,308], if not its current feasibility.

It may be possible instead to maintain the driver module of the Artevo system – the gene block containing rescue and biosensor genes as well as regulatory elements – separate to the genome on a partially orthogonal plasmid which encodes its own high-fidelity polymerase compatible only with origin sequences on said plasmid. Inspiration for this method comes from reversing the approach of the continuous directed evolution platform, OrthoRep, which stably mutates select genes on an orthogonal plasmid in yeast above otherwise deleterious genomic error thresholds[309]. A collaborative effort in 2020 between Professors Chang C. Liu, who developed the original OrthoRep system, and Frances Arnold saw the system used to expand the substrate profile of the tryptophan synthase β-subunit of *Thermotoga maritima*, towards discovery of novel, industrially-useful biocatalysis[179]. Unlike the OrthoRep system, which uses a stable host to rapidly mutate genes carried on a plasmid, I propose relevant alkane biosynthesis modules be genomically integrated into an Artevo chassis, then whole-genome mutation rates be elevated by operating mutational actuators on the genome, such as overexpression of *dinB* or suppression of *mutS*. The Artevo circuit itself would be housed on an orthogonal vector processed only by high-fidelity polymerases, keeping mutation rates of that plasmid low while potentially beneficial mutations are generated across the entire genome of the host organism, not just within the alkane biosynthesis module. The benefit of this approach is that it captures mutations occurring outside the locus of alkane biosynthesis genes, but a drawback is the possible elevation of mutation rates above the threshold for single-generation extinction.

In situ continuous directed evolution

Continuous directed evolution systems are designed to consolidate translation, screening, selection, and mutation into a single biological chassis and, ideally, a single unit operation. Continuous directed evolution using the Artevo system was originally envisaged as the continual addition of raw DNA into an *A. baylyi* population held in exponential phase, and thus naturally transformable, within a chemostat. The system would be able to evolve some phenotype, at the level of a set of genes or a whole organism, over many integrated cycles of mutation and selection, with minimal manual intervention. Some far-future form of the Artevo system would perhaps resemble a chemical process unit, with 'genetic diversity' as a process variable, just like

'temperature' or 'pH'. This variable could be adjusted by changing the composition or flow rate of an input stream consisting of naked DNA, which the cells in the reactor – essentially a living catalyst being evolved *in situ* toward a single process goal – would take up and potentially integrate into their genomes. Practical issues – the sheer cost of synthesising such a DNA stream, the long-term stability of the population, limited transformation efficiencies, loss of rare valuable variants in a large population, and the technology required for active in-line monitoring of something as complex as 'genetic diversity' – notwithstanding, the image helps us visualise what continuous directed evolution using the Artevo system could be and explore what auxiliary processes might be useful to support it.

One proof-of-concept experiment, which could be performed to develop the Artevo system toward continuous *in situ* directed evolution, is outlined below:

i. Wild type *Acinetobacter baylyi* ADP1-ISx is naturally resistant to ampicillin. A minimum inhibitory concentration (MIC) assay could be performed to establish the range of ampicillin concentrations over which growth of *A. baylyi* is reduced but not completely halted. Crucially, the concentration of ampicillin should reduce growth enough that expression of an ampicillin resistance cassette would act as a rescue gene, restoring growth to that of the wild type.

ii. Wild type *A. baylyi* could then be grown in a chemostat against the background ampicillin concentration established by the MIC assay. Using the Chi.Bio benchtop reactor platform developed by Steel *et al.* in collaboration with our laboratory[250], the bacterial population could be held in exponential phase, when cells become naturally competent and recombinant.

iii. This culture could then be fed with raw dsDNA encoding an ampicillin resistance cassette and a fluorescence reporter gene, flanked by 0.5-1 kb of DNA homologous to one of the integration hotspots on the *A. baylyi* genome identified by Biggs *et al.*[159]. Some portion of the population in the chemostat should take up this DNA and integrate it into their genomes, producing an *A. baylyi* phenotype with improved ampicillin resistance.

iv. The ampicillin resistance cassette should be enriched within the population as the phenotype will display improved growth compared to any untransformed neighbours. Increasing prevalence of the cassette could be monitored via the specific fluorescence signal of the population.

This experiment would demonstrate that a naturally competent host chassis can be transformed and selected *in situ*. The experiment could be multiplexed with multiple additional insertion cassettes or repeated using a previously transformed population, provided multiple rescue genes are available, to investigate the limits of such a system in continuous directed evolution applications.

This continuously growing culture is likely to encounter bioprocess issues, such as biofilm formation, which limit its practical usefulness, but advances in bioprocess equipment have enabled increasingly sophisticated modern continuous directed evolution approaches, facilitating autonomous cycles of mutation, selection, and enrichment. Partitioned continuous culture devices have helped ameliorate biofilm-related issues[310], as has the implementation of workflow parallelisation enabled by cheaper and more user-friendly robotic liquid handling systems[184,311,312] and miniaturised chemostats[313–315]. The PRANCE (phage- and robotics- assisted near continuous evolution) system developed by DeBenedictis *et al.* applies high-throughput liquid handling robotics to PACE (phage-assisted continuous evolution) – an established directed evolution technique[183] – offering improved scalability and real-time monitoring of molecular fitness rather than just growth[184].

Effective *in situ* gene delivery is required for working continuous directed evolution systems lacking *in vivo* diversity generation capabilities, including the Artevo system, as host basal mutation rates are typically insufficient for achieving novel functionalities in a useful timeframe[316]. Scalability of *in situ* delivery strategies is limited by the practicality of large-scale synthesis of transforming DNA and fixation of transformants into large populations, which is why gene circuits that accelerate basal mutation rates – like those described in the previous section – or chemical mutagens, are popular technologies for continuous directed evolution. Systems such as the Church group's MAGE (multiplex automated genome engineering) use a site-specific mutagenesis approach, leveraging automated liquid handling systems to achieve the high throughputs necessary for success[170]. While naturally competent chassis, such as *A. baylyi*, do not require manual intervention to transform, scalable electroporation systems have enabled near-continuous delivery of diverse DNA to host strains without this characteristic[317].

Surveying the landscape of directed evolution processes in use today highlights a variety of auxiliary systems that facilitate, accelerate, automate, or otherwise streamline some aspect of the process, be it screening, selection, or diversity generation. These systems function to enhance the effect of engineered genetic circuitry directing the evolution process, or sometimes simply to harness and accelerate natural evolution toward industrial goals.

Discussion

The Artevo system described in this thesis has real potential for improvement and a demonstrable basis for success. Several next steps to continue the project have been described in this chapter. Of immediate interest is further characterisation of the system response to intracellular alkane production, to quantify the relationship between productivity of an alkane biosynthesis module and change in host growth rate. This information will aid efforts to control and tune the growth advantage received by a host cell per unit of alkane, over time, produced by that cell.

First, there must be a decoupling of promoters governing antisense regulation of the rescue gene *dapB* and expression of the alkane biosynthesis genes, both of which are currently induced by IPTG. Alternative expression systems induced by anhydrotetracycline (*tetR*) or salicylate (*salR*) may be used to replace either of the existing IPTG-inducible promoters. Transcriptional interference elements regulating the rescue gene may also be reconfigured such that the interfering antisense promoter represses the regulator, *alkR*, indirectly downregulating *dapB*. Such a configuration was explored in previous work by this author and colleague Dr Jack Chen[232]. Currently, the circuit architecture resembles that described by Brophy and Voigt[263], with direct transcriptional interference across the rescue gene. The strength of antisense regulation could be further adjusted using various strategies described in Chapter 4, including varying the distance between interfering promoters or improving processivity of the RNAP-ribosome complex[262–265]. Direct mutation of DapB protein to reduce its activity[235] may also be useful in adjusting response thresholds within the system. Overall, there are many handles available for further tuning of the Artevo system and for investigating the limits of system performance; that is, the sensitivity with which the Artevo system can interrogate a library of diverse gene modules and selectively enrich specific sequences.

Figure 31 is a visual summary of the Artevo system, from the level of genetic circuitry through to cell culture to enrichment of high-performing gene modules. While latter stages of the process depicted in *Figure 31* have not yet been rigorously demonstrated, the illustration builds on what has been achieved in this project and captures a vision for the potential of the system, which should help to guide future work described in this chapter.

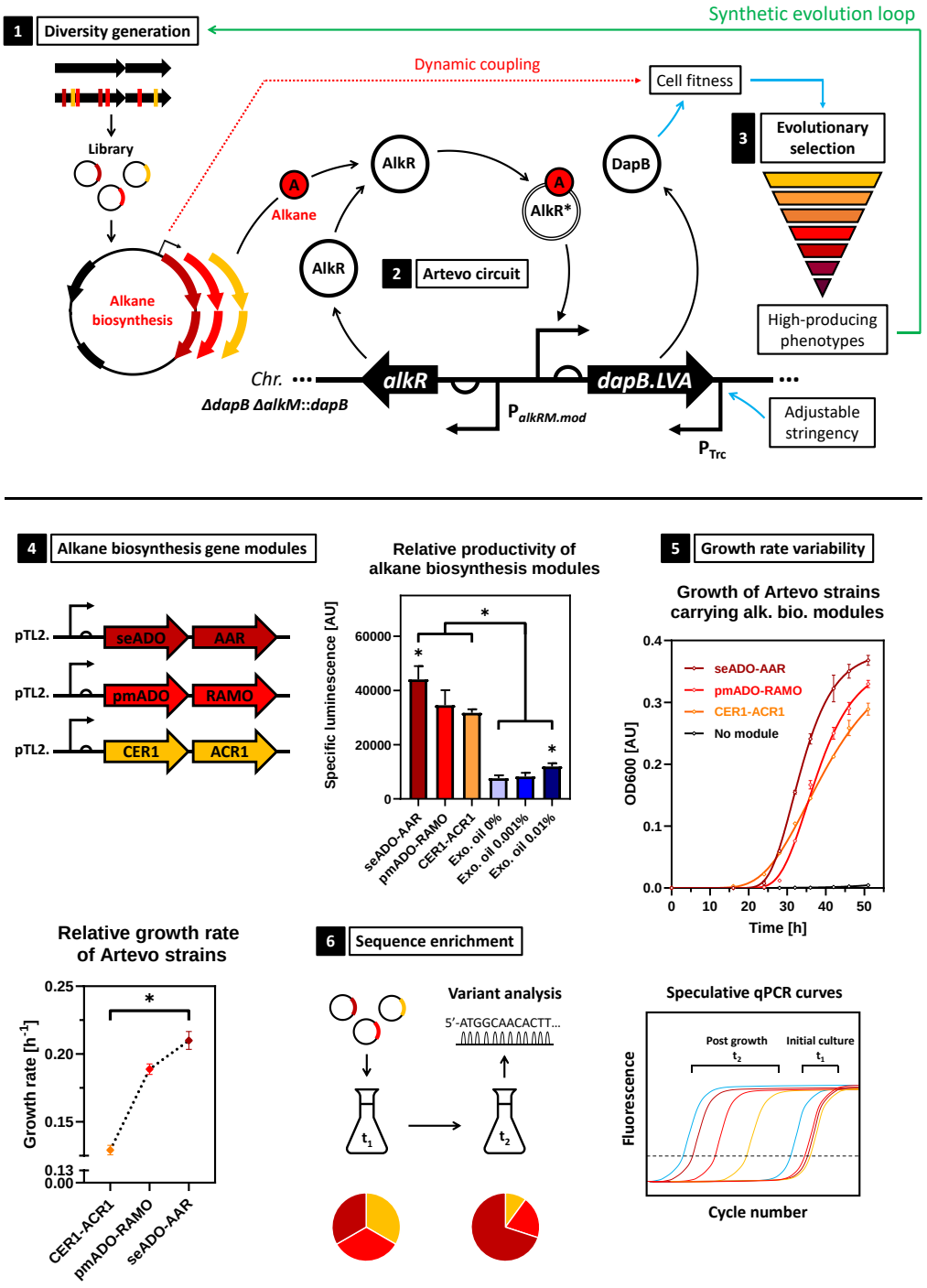


Figure 31. Full Artevo system diagram. [1]. Mutagenesis methods create a diverse library of variant alkane biosynthesis sequences. In this project, three ADO-FAR modules with differing relative alkane productivities are used as a proof-of-concept to test the Artevo system. [2]. The Artevo circuit; a biosensor-based directed evolution circuit linking growth rate of a bacterial host (*Acinetobacter baylyi*) to its alkane biosynthesis capacity. [3]. Generational enrichment of high-performing phenotypes amplifies valuable sequence variants. [4]. The proof-of-concept gene modules used in this project to test the Artevo system; seADO-AAR, pmADO-RAMO, and CER1-ACR1. Relative productivity of these modules has been demonstrated in *A. baylyi* by Lehtinen *et al.*[245] and again here in the ADPWH_alk biosensor, which luminesces in proportion to exogenous (blue) and endogenous (red) alkanes[161]. [5]. Graded growth response of the Artevo strain proportional to productivity of the ADO-FAR module. [6]. Speculative illustration of enrichment of high-performing gene modules in a combined culture of Artevo strains. Illustration of predicted C_t downshift in qPCR curves, indicative of increased prevalence of high-producing module (seADO-AAR) over lower-producing modules (pmADO-RAMO and CER1-ACR1) in a combined culture after prolonged growth.

Chapter 7

Conclusion

Directed evolution strategies operate within colossal solution spaces, vast landscapes of sequence diversity that are themselves limited iterations on part of some overarching system; gentle foothills of combinatorial complexity in the shadow of towering mountains. These daunting landscapes need not be explored in their entirety however, as improvements to many systems are not predicated on complete knowledge; we don't always need to know everything, we often just need to know enough to make a difference. Generating and processing sufficient sequence space to achieve this still challenges even automated, high-throughput directed evolution methods, hence the attraction of a system that can examine and selectively enrich diverse phenotypes within a growing population of cells, from among billions of variants. This approach mirrors natural evolutionary processes and is fundamental to the system developed in this project: a novel biosensor-based prokaryotic platform technology for directed evolution of alkane biosynthesis genes. This system has been referred to as 'artificial evolution', or the Artevo system.

This project has culminated in a proof-of-concept demonstration of the Artevo system, which has been shown to couple endogenous alkane productivity to host fitness, driving generational enrichment of high-producing phenotypes within a growing population. The system is designed to interrogate a diverse library of genetic modules

conferring alkane biosynthesis capacity, screening and selecting between a vast number of variants within a single culture vessel. Growth rate of an auxotrophic bacterial host is proportional to expression of a rescue gene, which is linked to local alkane concentration by a biosensory regulator protein. Stringency of selection can also be tuned using an inducible antisense promoter that opposes rescue gene expression. The current system requires upstream *in vitro* diversity generation and has been tested using a pool of three distinct alkane biosynthesis modules, the relative productivity of which has been shown to elicit a graded growth rate response in host Artevo strains. Strategies for incorporating *in vivo* diversity generation have also been outlined, in pursuit of a system capable of *in situ* continuous directed evolution.

This work contributes to the literature a novel ⊿*dapB* markerless mutant of *Acinetobacter baylyi* ADP1 and its transposon-free counterpart ADP1-ISx. It has produced a manuscript, currently available on bioRχiv, on the use of transcriptional interference for regulating gene expression. The prototype Artevo system built and tested during this project is also ready for further development, and strategies for improving it have been provided to guide inheritors of the technology. Beyond the Artevo system, I have been fortunate enough to be part of other projects within the groups of Profs. Huang and Thompson, having advised and assisted with wax ester synthesis work in *A. baylyi* and contributed to a publication on ultrasound-mediated DNA delivery to bacterial biofilms[318]. Once the enrichment analysis experiments outlined in Chapters 5 and 6 are completed – hopefully within this year – either by myself or by whoever inherits this system, I will condense the story of this project into a manuscript for publication. Continuity of this work is important to me and I am happy to be able to pursue the next stage of my career in Oxford, so that I may maintain strong ties with Prof. Huang and remain involved with future development of the Artevo system. More generally, I have also been responsible for training and support of new members of the Huang lab since August 2020, in lieu of dedicated postdoctoral staff. It has been a great pleasure to manage the laboratory and work alongside such a friendly, diverse, and talented group of people.

Climate change has, for some years now, been a tangible phenomenon, with extreme weather events becoming more common and affecting the lives of more people across the globe. World-changing upheaval resulting from a climate catastrophe is no longer exclusive to a faraway future or a prehistoric past; human choices have accelerated the negative effects of climate change to a rate relevant to human experience. The silver lining is an increased global awareness and appreciation of climate change, which necessarily precedes any meaningful or concerted action from international communities. The point of no return for our climate will be reached in my lifetime and

it is only thanks to this increased awareness that we have the chance to stop and choose whether to cross it. Otherwise, it may well pass us unnoticed. Earth will undoubtedly survive whatever great ecological shifts are brought about by climate change, and humanity, clever and tenacious, almost certainly will as well, but it is difficult to know exactly what our planet and our societies will look like when that transformative survival process is complete. It is easy though to recognise the irreplaceable natural treasures we stand to lose. Unprecedented technical innovation, social adaptation, and global co-operation is needed to preserve humanity's priceless relationship with our only home.

It has been a privilege to contribute in some small part to the global research efforts directed toward protecting our climate. For all that this work has taught me about the true scale of the climate problem and pervasiveness of fossil energy, I nevertheless am hopeful, and I look forward to the future with all the challenges and triumphs it may bring.

References

1. Smil, V. *Transforming the 20th Century - Technical Innovations and their Challenges*. (Oxford University Press, 2006).
2. Schwab, K. *The Fourth Industrial Revolution. World Economic Forum* (Crown Publishing Group, 2017).
3. Hutson, M. Reversed aging, pig organs, and the future of humankind. *Medium - Future Human* https://medium.com/s/futurehuman/reversed-aging-pig-organs-and-the-future-of-humankind-50f1cdb1e014 (2018).
4. Levy, S. Geek Power: Steven Levy revisits tech titans, hackers, idealists. *WIRED* https://www.wired.com/2010/04/ff-hackers/ (2010).
5. Isaacson, W. *Steve Jobs: The Exclusive Biography*. (Simon & Schuster, 2011).
6. Church, G. M., Gao, Y. & Kosuri, S. Next-generation digital information storage in DNA. *Science (1979)* **337**, 1628 (2012).
7. Gingko Bioworks. Biology by design. https://www.ginkgobioworks.com/our-work/ (2018).
8. Bolt Threads. Microsilk. https://boltthreads.com/technology/microsilk/ (2018).
9. Paddon, C. J. & Keasling, J. D. Semi-synthetic artemisinin: A model for the use of synthetic biology in pharmaceutical development. *Nature Reviews Microbiology* **12**, 355–367 (2014).

10. Venter, J. C. What is life? A 21st century perspective. *70th Anniversary of Schroedinger's Lecture - Trinity College* https://www.edge.org/conversation/what-is-life (2012).

11. HM Government. *Growing the bioeconomy: National strategy 2018-2030*. (2018).

12. SynBio Leadership Council. *Biodesign for the bioeconomy: UK SynBio Strategic Plan '16*. (2016).

13. Bunger, M. & Upton, L. *4Q 2021 Synthetic Biology Venture Investment Report*. (2022).

14. Chui, M., Evers, M., Manyika, J., Zheng, A. & Nisbet, T. *The Bio Revolution: Innovations transforming economies, societies, and our lives*. (2020).

15. EuropaBio. *Industrial Biotechnology—Contributing Towards Achieving the UN Global Sustainable Development Goals. Industrial Biotech - Europe's key asset in meeting the UN's Sustainable Development Goals (SDGs)* (2018) doi:10.1089/ind.2018.29138.ebi.

16. McGrath, M. Sir David Attenborough: Climate change "our greatest threat." *BBC News - Science & Environment* (2018).

17. Jackson, R. B. *et al.* Global energy growth is outpacing decarbonization. *Environmental Research Letters* **13**, 120401 (2018).

18. International Energy Agency. *World Energy Outlook 2018*. *IEA Publications* (2018) doi:10.1049/ep.1977.0180.

19. Manning, C., Clayton, S., Whitmarsh, L. & Capstick, S. *Psychology and Climate Change: Human perceptions, impacts, and responses*. (Academic Press, 2018).

20. Masson-Delmotte, V. *et al. Global warming of 1.5C. An IPCC special report on the impacts of global warming: Summary for policymakers*. (2018).

21. Watts, J. We have 12 years to limit climate change catastrophe, warns UN. *The Guardian* https://www.theguardian.com/environment/2018/oct/08/global-warming-must-not-exceed-15c-warns-landmark-un-report (2018).

22. Allan, R. P. *et al. Climate Change 2021 - The physical science basis*. (2021).

23. International Energy Agency. *CO2 emissions from fuel combustion*. (2017).

24. International Civil Aviation Organization. *Global aviation CO2 emissions projections to 2050*. (2012).

25. Smil, V. *Energy and Civilization - A History*. (MIT Press, 2017).

26. Battery University. Cost of Mobile and Renewable Power. *Batteries as a power source* https://batteryuniversity.com/index.php/learn/article/bu_1006_cost_of_mobile_power (2017).

27. Temple, J. The $2.5 trillion reason we can't rely on batteries to clean up the grid. *MIT Technology Review* (2018).

28. Condon, M., Revesz, R. L. & Unel, B. *Managing the Future of Energy Storage Implications for Greenhouse Gas Emissions*. (2018).

29. Olivetti, E. A., Ceder, G., Gaustad, G. G. & Fu, X. Lithium-Ion Battery Supply Chain Considerations: Analysis of Potential Bottlenecks in Critical Metals. *Joule* **1**, 229–243 (2017).

30. Fast, A. G. & Papoutsakis, E. T. Stoichiometric and energetic analyses of non-photosynthetic CO2-fixation pathways to support synthetic biology strategies for production of fuels and chemicals. *Current Opinion in Chemical Engineering* **1**, 380–395 (2012).

31. NobelPrize.org. Press Release: The Nobel Prize in Chemistry 2018. *Nobel Media AB* https://www.nobelprize.org/prizes/chemistry/2018/press-release/ (2018).

32. Kang, M. K. & Nielsen, J. Biobased production of alkanes and alkenes through metabolic engineering of microorganisms. *Journal of Industrial Microbiology and Biotechnology* **44**, 613–622 (2017).

33. Davis, S. C. & Boundy, R. G. *Transportation Energy Data Book*. (Oak Ridge National Laboratory, 2019). doi:10.1007/978-94-011-2174-3_2.

34. Bexten, R. An-225 sets record cargo lift. *Airports International* https://airportsinternational.keypublishing.com/2009/09/18/an225-sets-record-cargo-lift/6769/ (2009).

35. North, G. R. *et al. Surface Temperature Reconstructions*. vol. 1 (2006).

36. NASA. GISS Surface Temperature Analysis (GISTEMP v3). https://data.giss.nasa.gov/gistemp/ (2021).

37. Brown, K. 2020 Tied for Warmest Year on Record, NASA Analysis Shows. *NASA Earth* https://www.nasa.gov/press-release/2020-tied-for-warmest-year-on-record-nasa-analysis-shows (2021).

38. World Meteorological Organisation. *The State of the Global Climate in 2018*. (2018).

39. Jérôme, D. & Haegeli, J. *Preliminary sigma estimates for 2018: global insured losses of USD 79 billion are fourth highest on sigma records*. (2018).

40. Centre for Research on the Epidemiology of Disasters. International Disaster Database. *EM-DAT* https://www.emdat.be/ (2019).

41. Levitus, S.; Antonov, J.; Boyer, T.; Baranova, O.; Garcia, H.; Locarnini, R.; Mishonov, A.; Reagan, J.; Seidov, D.; Yarosh, E.; Zweng, M. NCEI ocean heat content, temperature anomalies, salinity anomalies, thermosteric sea level anomalies, halosteric sea level anomalies, and total steric sea level anomalies from 1955 to present calculated from in situ oceanographic subsurface profile data. *National Centres for Environmental Information Dataset* https://catalog.data.gov/dataset/ncei-ocean-heat-content-temperature-anomalies-salinity-anomalies-thermosteric-sea-level-anomali (2017) doi:10.7289/V53F4MVP.

42. Mostofa, K. M. G. *et al.* Reviews and Syntheses: Ocean acidification and its potential impacts on marine ecosystems. *Biogeosciences Discussions* **12**, 10939–10983 (2016).

43. Center, N. S. & I. D. State of the cryosphere. *State of the Cryosphere: Sea Ice* https://nsidc.org/cryosphere/sotc/sea_ice.html (2018).

44. ICSU. World glacier monitoring service. *Online* https://wgms.ch/ (2019).

45. Nerem, R. S. *et al.* Climate-change–driven accelerated sea-level rise detected in the altimeter era. *PNAS* **115**, (2018).

46. IPBES. *Biodiversity and Ecosystem Services for Africa*. (2018) doi:10.1016/B978-0-12-384719-5.00349-X.

47. Diaz, S. & Duffy, J. *Biodiversity and Ecosystem Services for the Americas*. (2018) doi:10.1016/B978-0-12-384719-5.00349-X.

48. Clemens, P. V., Mark, S., Markus, R. & Markus, B. *Biodiversity and Ecosystem Services for Asia and the Pacific*. (2018).

49. Clemens, P. V., Mark, S., Markus, R. & Markus, B. *Biodiversity and Ecosystem Services for Europe and Central Asia*. (2018).

50. IPBES. *The IPBES assessment report on land degradation and restoration*. (2018).

51. Ferrier, S. *et al. IPBES: The methodological assessment report on Scenarios and Models of Biodiversity and Ecosystem Services*. (2016).

52. Santer, B. D. *et al.* Celebrating the anniversary of three key events in climate change science. *Nature Clim. Change* **9**, submitted (2018).

53. Voosen, P. The realist. *Science (1979)* **359**, 1320–1324 (2018).

54. International Energy Agency. Total Primary Energy Supply by Fuel, 1971 and 2019. *World Energy Balances, Overview* https://www.iea.org/data-and-statistics/charts/total-primary-energy-supply-by-fuel-1971-and-2019 (2019).

55. Cozzi, L. (International E. A. & Gould, T. (International E. A. *World Energy Outlook 2021*. www.iea.org/weo (2021).

56. Olhoff, A. *et al. Emissions Gap Report 2018*. (2018).

57. McBain, S., Teter, J. & Tattini, J. *Tracking Transport*. https://www.iea.org/reports/tracking-transport-2021 (2021).

58. International Energy Agency. Climate Change. *CO2 Emissions* https://www.iea.org/topics/climate-change (2021).

59. Friedlingstein, P. *et al.* Global Carbon Budget 2021. *Earth Syst. Sci. Data Discuss. [preprint]* (2021).

60. US Energy Information Administration. Carbon dioxide emissions coefficients. *Environmental Information* https://www.eia.gov/environment/emissions/co2_vol_mass.php (2016).

61. Smil, V. Power Density Primer (Parts 1-5): Understanding the Spatial Dimension of the Unfolding Transition to Renewable Electricity Generation. *Atlantic* (2010).

62. Smil, V. *Power Density: A key to understanding energy sources and uses*. (MIT Press, 2016).

63. Lazard. *Levelized Cost of Energy Analysis*. (2017).

64. International Energy Agency. *Global EV outlook 2018*. (2018).

65. International Energy Agency. *The role of critical minerals in clean energy transitions*. (2021).

66. Bieker, G. *A global comparison of the life-cycle greenhouse gas emissions of combustion engine and electric passenger cars*. (2021).

67. Smil, V. *Grand Transitions: How the modern world was made*. (Oxford University Press, 2021). doi:10.1093/oso/9780190060664.001.0001.

68. Smil, V. *Natural Gas: Fuel for the 21st century*. (Wiley-Blackwell, 2015).

69. Jimï¿½nez-Dï¿½az, L., Caballero, A., Pï¿½rez-Hernï¿½ndez, N. & Segura, A. Microbial alkane production for jet fuel industry: motivation, state of the art and perspectives. *Microbial Biotechnology* **10**, 103–124 (2017).

70. Davis, S. J. *et al.* Net-zero emissions energy systems. *Science (1979)* **360**, 1–9 (2018).

71. Clark, J. Green chemistry for a circular economy. Preprint at (2018).

72. Ballinger, B. *et al.* The vulnerability of electric vehicle deployment to critical mineral supply. *Applied Energy* **255**, 113844 (2019).

73. Ballinger, B. *et al.* The vulnerability of electric-vehicle and wind-turbine supply chains to the supply of rare-earth elements in a 2-degree scenario. *Sustainable Production and Consumption* **22**, 68–76 (2020).

74. Kaiman, J. Rare earth mining in China: the bleak social and environmental costs. *The Guardian* (2014).

75. Paul, J. & Campbell, G. *Investigating rare earth element mind development in EPA Region 8 and potential environmental impacts*. (2011).

76. Office of Solid Waste. *Human health and environmental damages from mining and mineral processing wastes*. (1995).

77. Hu, F.-F., Fan, H.-R., Yang, K.-F., Wang, K.-Y. & Liu, S. The giant Bayan Obo REE-Nb-Fe deposit, China: Controversy and ore genesis. *Geoscience Frontiers* **7**, 335–344 (2015).

78. Schreiber, A. *et al.* Environmental Impacts of Rare Earth Mining and Separation Based on Eudialyte: A New European Way. *Resources* **5**, 32 (2016).

79. Ali, S. Social and Environmental Impact of the Rare Earth Industries. *Resources* **3**, 123–134 (2014).

80. NASA. Rare earth in Bayan Obo. *Visible Earth* https://visibleearth.nasa.gov/view.php?id=77723 (2012).

81. Zhu, X.-G., Long, S. P. & Ort, D. R. What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Current Opinion in Biotechnology* **19**, 153–159 (2008).

82. Nalley, S. & LaRose, A. *Annual US Energy Outlook 2022 (AEO2022)*. (2022).

83. de Jong, K. P. *et al.* The frontiers of energy. *Nature Energy* **1**, 15020 (2016).

84. Wray, A. *Crops grown for bioenergy in England and the UK*. (2017).

85. Taheripour, F., Zhao, X. & Tyner, W. E. The impact of considering land intensification and updated data on biofuels land use change and emissions estimates. *Biotechnology for Biofuels* **10**, 1–16 (2017).

86. Amorim, H. v. & Lopes, M. L. Ethanol production in a petroleum dependent world: the Brazilian experience. *Sugar Journal* **67**, 11–14 (2005).

87. Goldemberg, J. The Brazilian biofuels industry. *Biotechnology for Biofuels* **1**, 6 (2008).

88. Potter, N. I. How Brazil achieved energy independence and the lessons the United States should learn from Brazil's experience. *Washington University Global Studies Law Review* **7**, 331–351 (2008).

89. da Costa, A. O., dos Santos, L. B. O., Henriques, R. M. & Araujo, R. B. *Implementation of bioenergy in Brazil - 2021 update*. (2021).

90. Rossi, L. M. *et al.* Ethanol from Sugarcane and the Brazilian Biomass-Based Energy and Chemicals Sector. *ACS Sustainable Chemistry & Engineering* **9**, 4293–4295 (2021).

91. Santos, T. do N. *et al.* Potential for biofuels from the biomass of prickly pear cladodes: Challenges for bioethanol and biogas production in dry areas. *Biomass and Bioenergy* **85**, 215–222 (2016).

92. Yang, L. *et al.* Biomass characterization of Agave and Opuntia as potential biofuel feedstocks. *Biomass and Bioenergy* **76**, 43–53 (2015).

93. Adeniyi, O. M., Azimov, U. & Burluka, A. Algae biofuel: Current status and future applications. *Renewable and Sustainable Energy Reviews* **90**, 316–335 (2018).

94. Carneiro, M. L. N. M. *et al.* Potential of biofuels from algae: Comparison with fossil fuels, ethanol and biodiesel in Europe and Brazil through life cycle assessment (LCA). *Renewable and Sustainable Energy Reviews* **73**, 632–653 (2017).

95. Wahlquist, C. South Australia's Tesla battery on track to make back a third of cost in a year. *The Guardian* (2018).

96. de Sisternes, F. J., Jenkins, J. D. & Botterud, A. The value of energy storage in decarbonizing the electricity sector. *Applied Energy* **175**, 368–379 (2016).

97. Brick, S. *Charting Pathways to Deep Decarbonization: Challenges for Analysts, Policymakers, Advocates and the Public*. (2018).

98. Heard, B. P., Brook, B. W., Wigley, T. M. L. & Bradshaw, C. J. A. Burden of proof: A comprehensive review of the feasibility of 100% renewable-electricity systems. *Renewable and Sustainable Energy Reviews* **76**, 1122–1133 (2017).

99. Brown, T. W. *et al.* Response to 'Burden of proof: A comprehensive review of the feasibility of 100% renewable-electricity systems.' *Renewable and Sustainable Energy Reviews* **92**, 834–847 (2018).

100. Schwirplies, C. Fair pay for green energy. *Nature Energy* **3**, 822–823 (2018).

101. Teräväinen, T. Visions of energy futures. *Nature Energy* **3**, 923–924 (2018).

102. Li, M., Lu, J., Chen, Z. & Amine, K. 30 Years of Lithium-Ion Batteries. *Advanced Materials* **30**, 1–24 (2018).

103. Kawasaki, H. *et al.* Silicon heterojunction solar cell with interdigitated back contacts for a photoconversion efficiency over 26%. *Nature Energy* **2**, 17032 (2017).

104. Lufthansa. *Balance: Sustainability Report Issue 2018*. (2018).

105. International Air Transport Association. Jet Fuel Price Monitor. *Fuel Price Analysis* https://www.iata.org/en/publications/economics/fuel-monitor/ (2022).

106. Ploetner, K. O., Schmidt, M., Baranowski, D., Isikveren, A. T. & Hornung, M. Operating cost estimation for electric-powered transport aircraft. in *Aviation Technology, Integration, and Operations Conference* (2013).

107. Tao, L., Milbrandt, A., Zhang, Y. & Wang, W. C. Techno-economic and resource analysis of hydroprocessed renewable jet fuel. *Biotechnology for Biofuels* **10**, 1–16 (2017).

108. Dominković, D. F., Bačeković, I., Pedersen, A. S. & Krajačić, G. The future of transportation in sustainable energy systems: Opportunities and barriers in a clean energy transition. *Renewable and Sustainable Energy Reviews* **82**, 1823–1838 (2018).

109. iGEM Foundation. Registry of standard biological parts. http://parts.igem.org/Main_Page (2019).

110. Fischer, M. *et al. BioBrick Public Agreement*. (2009).

111. Sauro, H. M. *et al.* Sharing Structure and Function in Biological Design with SBOL 2.0. *ACS Synthetic Biology* **5**, 498–506 (2016).

112. Lucks, J. B., Qi, L., Whitaker, W. R. & Arkin, A. P. Toward scalable parts families for predictable design of biological circuits. *Current Opinion in Microbiology* **11**, 567–573 (2008).

113. Decoene, T. *et al.* Standardization in synthetic biology: an engineering discipline coming of age. *Critical Reviews in Biotechnology* **38**, 647–656 (2018).

114. Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science (1979)* **351**, (2016).

115. Keasling, J. *et al.* Microbial production of advanced biofuels. *Nature Reviews Microbiology* **19**, 701–715 (2021).

116. Azhaganathan, G. & Bragadeshwaran, A. Critical review on recent progress of ethanol fuelled flex-fuel engine characteristics. *International Journal of Energy Research* **46**, 5646–5677 (2022).

117. Hsu, C. S. & Robinson, P. R. *Springer Handbook of Petroleum Technology*. (Springer International Publishing, 2017). doi:10.1007/978-3-319-49347-3.

118. Fahim, M. A., Alsahhaf, T. A. & Elkilani, A. Refinery Feedstocks and Products. in *Fundamentals of Petroleum Refining* 11–31 (Elsevier, 2010). doi:10.1016/B978-0-444-52785-1.00002-4.

119. Baral, N. R. *et al.* Techno-economic analysis and life-cycle greenhouse gas mitigation cost of five routes to bio-jet fuel blendstocks. *Energy and Environmental Science* **12**, 807–824 (2019).

120. Liu, Y. *et al.* Biofuels for a sustainable future. *Cell* **184**, 1636–1647 (2021).

121. Schirmer, A., Rude, M. A., Li, X., Popova, E. & del Cardayre, S. B. Microbial biosynthesis of alkanes. *Science (1979)* **329**, 559–562 (2010).

122. Zhou, Y. J., Kerkhoven, E. J. & Nielsen, J. Barriers and opportunities in bio-based production of hydrocarbons. *Nature Energy* **3**, 925–935 (2018).

123. Patel, A. *et al.* An overview of potential oleaginous microorganisms and their role in biodiesel and omega-3 fatty acid-based industries. *Microorganisms* **8**, (2020).

124. Liao, J. C., Mi, L., Pontrelli, S. & Luo, S. Fuelling the future: Microbial engineering for the production of sustainable biofuels. *Nature Reviews Microbiology* **14**, 288–304 (2016).

125. Wang, W. & Lu, X. Microbial Synthesis of Alka(e)nes. *Frontiers in Bioengineering and Biotechnology* **1**, 1–5 (2013).

126. Arai, M., Hayashi, Y. & Kudo, H. Cyanobacterial Enzymes for Bioalkane Production. in *Synthetic Biology of Cyanobacteria* (eds. Zhang, W. & Song, X.) vol. 1080 119–154 (Springer, 2018).

127. Huang, W., Hunter, N., Davidson, P. & Canniffe, D. WO2013128184 - Alkane Production via LUX pathway. 77 (2013).

128. Howard, T. P. *et al.* Synthesis of customized petroleum-replica fuel molecules by targeted modification of free fatty acid pools in Escherichia coli. *Proceedings of the National Academy of Sciences* **110**, 7636–7641 (2013).

129. Cao, Y. X. *et al.* Heterologous biosynthesis and manipulation of alkanes in Escherichia coli. *Metabolic Engineering* **38**, 19–28 (2016).

130. Fell, D. A. *et al.* Model-assisted metabolic engineering of Escherichia coli for long chain alkane and alcohol production. *Metabolic Engineering* **46**, 1–12 (2018).

131. Choi, Y. J. & Lee, S. Y. Microbial production of short-chain alkanes. *Nature* **502**, 571–574 (2013).

132. Beller, H. R., Goh, E. B. & Keasling, J. D. Genes involved in long-chain alkene biosynthesis in micrococcus Luteus. *Applied and Environmental Microbiology* **76**, 1212–1223 (2010).

133. Brubaker, S. *et al.* Terminal Olefin (1-Alkene) Biosynthesis by a Novel P450 Fatty Acid Decarboxylase from Jeotgalicoccus Species. *Applied and Environmental Microbiology* **77**, 1718–1727 (2011).

134. Akhtar, M. K., Turner, N. J. & Jones, P. R. Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. *Proceedings of the National Academy of Sciences* **110**, 87–92 (2012).

135. Walls, L. E. & Rios-Solis, L. Sustainable Production of Microbial Isoprenoid Derived Advanced Biojet Fuels Using Different Generation Feedstocks: A Review. *Frontiers in Bioengineering and Biotechnology* **8**, 1–19 (2020).

136. Meadows, A. L. *et al.* Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature* **537**, 694–697 (2016).

137. Mitrovich, Q. *MegaBio: An integrated process for production of farnesene, a versatile platform chemical, from domestic lignocellulosic feedstock.* (2019).

138. Marella, E. R., Holkenbrink, C., Siewers, V. & Borodina, I. Engineering microbial fatty acid metabolism for biofuels and biochemicals. *Current Opinion in Biotechnology* **50**, 39–46 (2018).

139. Marsh, E. N. G. & Waugh, M. W. Aldehyde Decarbonylases: Enigmatic Enzymes of Hydrocarbon Biosynthesis. *ACS Catalysis* **3**, 2515–2521 (2013).

140. Wang, Q. *et al.* Identification of residues important for the activity of aldehyde-deformylating oxygenase through investigation into the structure-activity relationship. *BMC Biotechnology* **17**, 1–9 (2017).

141. Wang, C., Zhao, C., Hu, L. & Chen, H. Calculated Mechanism of Cyanobacterial Aldehyde-Deformylating Oxygenase: Asymmetric Aldehyde Activation by a Symmetric Diiron Cofactor. *Journal of Physical Chemistry Letters* **7**, 4427–4432 (2016).

142. Patrikainen, P., Carbonell, V., Thiel, K., Aro, E. M. & Kallio, P. Comparison of orthologous cyanobacterial aldehyde deformylating oxygenases in the production of volatile C3-C7 alkanes in engineered E. coli. *Metabolic Engineering Communications* **5**, 9–18 (2017).

143. Blankley, R. T. *et al.* Production of alkenes and novel secondary products by P450 OleT JE using novel H2O2-generating fusion protein systems. *FEBS Letters* **591**, 737–750 (2017).

144. Russo, E. The birth of biotechnology. *Nature* **421**, 456–457 (2003).

145. Liu, H. & Deutschbauer, A. M. Rapidly moving new bacteria to model-organism status. *Current Opinion in Biotechnology* **51**, 116–122 (2018).

146. Madigan, M. T., Bender, K. S., Buckley, D. H., Sattley, W. M. & Stahl, D. A. *Brock Biology of Microorganisms.* (Pearson, 2017).

147. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology* **5**, 1–17 (2014).

148. Nikel, P. I., Chavarría, M., Danchin, A. & de Lorenzo, V. From dirt to industrial applications: Pseudomonas putida as a Synthetic Biology chassis for hosting harsh biochemical reactions. *Current Opinion in Chemical Biology* **34**, 20–29 (2016).

149. Adams, B. L. The Next Generation of Synthetic Biology Chassis: Moving Synthetic Biology from the Laboratory to the Field. *ACS Synthetic Biology* **5**, 1328–1330 (2016).

150. Foley, P. L. & Shuler, M. L. Considerations for the Design and Construction of a Synthetic Platform Cell for Biotechnological Applications. *Biotechnology and Bioengineering* **105**, 26–36 (2010).

151. Wu, F. *et al.* Design and Reconstruction of Regulatory Parts for Fast-frowing Vibrio natriegens Synthetic Biology. *ACS Synthetic Biology* **9**, 2399–2409 (2020).

152. Fisher, A. K., Freedman, B. G., Bevan, D. R. & Senger, R. S. A review of metabolic and enzymatic engineering strategies for designing and optimizing performance of microbial cell factories. *CSBJ* **11**, 91–99 (2014).

153. Barbe, V. *et al.* Unique features revealed by the genome sequence of Acinetobacter sp. ADP1, a versatile and naturally transformation competent bacterium. *Nucleic Acids Research* **32**, 5766–5779 (2004).

154. de Berardinis, V., Durot, M., Weissenbach, J. & Salanoubat, M. Acinetobacter baylyi ADP1 as a model for metabolic system biology. *Current Opinion in Microbiology* **12**, 568–576 (2009).

155. Metzgar, D. *et al.* Acinetobacter sp. ADP1: An ideal model organism for genetic analysis and genome engineering. *Nucleic Acids Research* **32**, 5780–5790 (2004).

156. Murin, C. D., Segal, K., Bryksin, A. & Matsumura, I. Expression vectors for Acinetobacter baylyi ADP1. *Applied and Environmental Microbiology* **78**, 280–283 (2012).

157. De Berardinis, V. *et al.* A complete collection of single-gene deletion mutants of Acinetobacter baylyi ADP1. *Molecular Systems Biology* **4**, (2008).

158. Durot, M. *et al.* Iterative reconstruction of a global metabolic model of Acinetobacter baylyi ADP1 using high-throughput growth phenotype and gene essentiality data. *BMC Systems Biology* **2**, (2008).

159. Biggs, B. W. *et al.* Development of a genetic toolset for the highly engineerable and metabolically versatile Acinetobacter baylyi ADP1. *Nucleic Acids Research* **48**, 5169–5182 (2020).

160. Santala, S. *et al.* Improved Triacylglycerol Production in Acinetobacter baylyi ADP1 by Metabolic Engineering. *Microbial Cell Factories* **10**, 1–10 (2011).

161. Luo, J., Efimova, E., Losoi, P., Santala, V. & Santala, S. Wax ester production in nitrogen-rich conditions by metabolically engineered Acinetobacter baylyi ADP1. *Metabolic Engineering Communications* **10**, (2020).

162. Zhang, D. *et al.* Whole-cell bacterial bioreporter for actively searching and sensing of alkanes and oil spills. *Microbial Biotechnology* **5**, 87–97 (2012).

163. Rogers, J. K. & Church, G. M. Multiplexed Engineering in Biology. *Trends in Biotechnology* **34**, 198–206 (2016).

164. Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B. A. & Gill, R. T. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nature Biotechnology* **28**, 856–862 (2010).

165. Schallmey, M., Frunzke, J., Eggeling, L. & Marienhagen, J. Looking for the pick of the bunch: High-throughput screening of producing microorganisms with biosensors. *Current Opinion in Biotechnology* **26**, 148–154 (2014).

166. Raman, S., Rogers, J. K., Taylor, N. D. & Church, G. M. Evolution-guided optimization of biosynthetic pathways. *Proceedings of the National Academy of Sciences* **111**, 17803–17808 (2014).

167. Wang, Y. *et al.* Directed Evolution: Methodologies and Applications. *Chemical Reviews* **121**, 12384–12444 (2021).

168. d'Oelsnitz, S. & Ellington, A. Continuous directed evolution for strain and protein engineering. *Current Opinion in Biotechnology* **53**, 158–163 (2018).

169. Simon, A. J., d'Oelsnitz, S. & Ellington, A. D. Synthetic evolution. *Nature Biotechnology* **37**, (2019).

170. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).

171. Wang, H. H. *et al.* Genome-scale promoter engineering by coselection MAGE. *Nature Methods* **9**, 591–593 (2012).

172. Nyerges, Á. *et al.* Directed evolution of multiple genomic loci allows the prediction of antibiotic resistance. *Proc Natl Acad Sci U S A* **115**, E5726–E5735 (2018).

173. Dicarlo, J. E. *et al.* Yeast Oligo-Mediated Genome Engineering (YOGE). *ACS Synthetic Biology* **2**, 741–749 (2013).

174. Jakočiūnas, T., Pedersen, L. E., Lis, A. V., Jensen, M. K. & Keasling, J. D. CasPER, a method for directed evolution in genomic contexts using mutagenesis and CRISPR/Cas9. *Metabolic Engineering* **48**, 288–296 (2018).

175. Garst, A. D. *et al.* Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nature Biotechnology* **35**, 48–55 (2017).

176. Halperin, S. O. *et al.* CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature* **560**, 248–252 (2018).

177. Dymond, J. S. *et al.* Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471–476 (2011).

178. Jia, B. *et al.* Precise control of SCRaMbLE in synthetic haploid and diploid yeast. *Nature Communications* **9**, (2018).

179. Rix, G. *et al.* Scalable continuous evolution for the generation of diverse enzyme variants encompassing promiscuous activities. *Nature Communications* **11**, 1–11 (2020).

180. Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A. & Liu, D. R. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nature Chemical Biology* **10**, 216–222 (2014).

181. Tschulena, U., Peterson, K. R., Gonzalez, B., Fedosyuk, H. & Barbas, C. F. Positive selection of DNA-protein interactions in mammalian cells through phenotypic coupling with retrovirus production. *Nature Structural and Molecular Biology* **16**, 1195–1199 (2009).

182. Brödel, A. K., Isalan, M. & Jaramillo, A. Engineering of biomolecules by bacteriophage directed evolution. *Current Opinion in Biotechnology* **51**, 32–38 (2018).

183. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).

184. DeBenedictis, E. A. *et al.* Systematic molecular evolution enables robust biomolecule discovery. *Nature Methods* **19**, 55–64 (2022).

185. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

186. NobelPrize.org. The Nobel Prize in Physiology or Medicine 1965. *Nobel Prize Outreach AB* https://www.nobelprize.org/prizes/medicine/1965/summary/ (2022).

187. Taylor, N. D. *et al.* Engineering an allosteric transcription factor to respond to new ligands. *Nature Methods* **13**, 177–183 (2016).

188. Snoek, T. *et al.* Evolution-guided engineering of small-molecule biosensors. *Nucleic Acids Research* **48**, 1–14 (2020).

189. Brophy, J. A. N. & Voigt, C. A. Principles of genetic circuit design. *Nature Methods* **11**, 508–520 (2014).

190. Mannan, A. A., Liu, D. & Zhang, F. Fundamental Design Principles for Transcription-Factor-Based Metabolite Biosensors. *ACS Synthetic Biology* **6**, 1851–1859 (2017).

191. Dietrich, J. A., Shis, D. L., Alikhani, A. & Keasling, J. D. Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis. *ACS Synthetic Biology* **2**, 47–58 (2013).

192. Chou, H. H. & Keasling, J. D. Programming adaptive control to evolve increased metabolite production. *Nature Communications* **4**, 1–8 (2013).

193. Rogers, J. K., Taylor, N. D. & Church, G. M. Biosensor-based engineering of biosynthetic pathways. *Current Opinion in Biotechnology* **42**, 84–91 (2016).

194. Wilson, M. C. & Piel, J. Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chemistry and Biology* **20**, 636–647 (2013).

195. Paepe, B. De, Peters, G., Coussement, P. & Maertens, J. Tailor-made transcriptional biosensors for optimizing microbial cell factories. *Journal of Industrial Microbiology & Biotechnology* **44**, 623–645 (2017).

196. Ratajczak, A., Geißdörfer, W. & Hillen, W. 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**, 5822–5827 (1998).

197. Throne-Holst, M., Wentzel, A., Ellingsen, T. E., Kotlar, H. K. & Zotchev, S. B. Identification of novel genes involved in long-chain n-alkane degradation by Acinetobacter sp. strain DSM 17874. *Applied and Environmental Microbiology* **73**, 3327–3332 (2007).

198. Santala, S., Efimova, E., Koskinen, P., Karp, M. T. & Santala, V. Rewiring the wax ester production pathway of acinetobacter baylyi ADP1. *ACS Synthetic Biology* **3**, 145–151 (2014).

199. Lehtinen, T. *et al.* Production of long chain alkyl esters from carbon dioxide and electricity by a two-stage bacterial process. *Bioresource Technology* **243**, 30–36 (2017).

200. Lehtinen, T., Efimova, E., Santala, S. & Santala, V. Improved fatty aldehyde and wax ester production by overexpression of fatty acyl-CoA reductases. *Microbial Cell Factories* **17**, 1–10 (2018).

201. Tacconelli, E. *et al. Global Priority List Of Antibiotic-Resistant Bacteria To Guide Research, Discovery And Development Of New Antibiotics*. *World Health Organization (WHO)* (2017) doi:10.1016/S1473-3099(09)70222-1.

202. Jiang, X. *et al.* Automating Cloning by Natural Transformation. *ACS Synthetic Biology* **9**, 3228–3235 (2020).

203. Suárez, G. A., Renda, B. A., Dasgupta, A. & Barrick, J. E. Reduced mutation rate and increased transformability of transposon-free Acinetobacter baylyi ADP1-ISx. *Evolutionary and Genomic Microbiology* **83**, 1–17 (2017).

204. Fan, C. *et al.* Defensive Function of Transposable Elements in Bacteria. *ACS Synthetic Biology* **8**, 2141–2151 (2019).

205. Utnes, A. L. G. *et al.* Growth phase-specific evolutionary benefits of natural transformation in Acinetobacter baylyi. *ISME Journal* **9**, 2221–2231 (2015).

206. Hülter, N. *et al.* Costs and benefits of natural transformation in Acinetobacter baylyi. *BMC Microbiology* **17**, 1–10 (2017).

207. Watson, J. F. & García-Nafría, J. In vivo DNA assembly using common laboratory bacteria: A re-emerging tool to simplify molecular cloning. *Journal of Biological Chemistry* **294**, 15271–15281 (2019).

208. De Vries, J. & Wackernagel, W. Integration of foreign DNA during natural transformation of Acinetobacter sp. by homology-facilitated illegitimate recombination. *Proc Natl Acad Sci U S A* **99**, 2094–2099 (2002).

209. Leong, C. G. *et al.* The role of core and accessory type IV pilus genes in natural transformation and twitching motility in the bacterium Acinetobacter baylyi. *PLoS ONE* **12**, 1–25 (2017).

210. Renda, B. A. Protocols: Acinetobacter baylyi Puddle Transformation. https://barricklab.org/twiki/bin/view/Lab/ProtocolsAcinetobacterPuddleTransformation (2016).

211. Royet, J. & Dziarski, R. Peptidoglycan recognition proteins: Pleiotropic sensors and effectors of antimicrobial defences. *Nature Reviews Microbiology* **5**, 264–277 (2007).

212. Born, T. L. & Blanchard, J. S. Structure/function studies on enzymes in the diaminopimelate pathway of bacterial cell wall biosynthesis. *Current Opinion in Chemical Biology* **3**, 607–613 (1999).

213. Wehrmann, A., Phillipp, B., Sahm, H. & Eggeling, L. Different Modes of Diaminopimelate Synthesis and Their Role in Cell Wall Integrity: a Study with Corynebacterium glutamicum. *Journal of Bacteriology* **180**, 3159–3165 (1998).

214. Bukhari, A. I. & Taylor, A. L. Genetic analysis of diaminopimelic acid- and lysine-requiring mutants of Escherichia coli. *Journal of Bacteriology* **105**, 844–854 (1971).

215. De Mot, R. *et al.* Development and Application of a dapB-Based In Vivo Expression Technology System To Study Colonization of Rice by the Endophytic Nitrogen-Fixing Bacterium Pseudomonas stutzeri A15. *Applied and Environmental Microbiology* **69**, 6864–6874 (2003).

216. Gal, M., Preston, G. M., Massey, R. C., Spiers, A. J. & Rainey, P. B. Genes encoding a cellulosic polymer contribute toward the ecological success of Pseudomonas fluorescens SBW25 on plant surfaces. *Molecular Ecology* **12**, 3109–3121 (2003).

217. Silby, M. W., Silby, M. W., Levy, S. B. & Levy, S. B. Use of In-Vivo Expression Technology To Identify Genes Important in Growth and Survival of Pseudomonas fluorescens Pf0-1 in soil: Discovery of expressed sequences with novel genetic organization. *Society* **186**, 7411–7419 (2004).

218. Hutton, C. A., Perugini, M. A. & Gerrard, J. A. Inhibition of lysine biosynthesis: an evolving antibiotic strategy. *Molecular BioSystems* **3**, 458–465 (2007).

219. Hutton, C. A., Southwood, T. J. & Turner, J. J. Inhibitors of lysine biosynthesis as antibacterial agents. *Mini-reviews in Medicinal Chemistry* **3**, 115–127 (2003).

220. Paiva, A. M. *et al.* Inhibitors of dihydrodipicolinate reductase, a key enzyme of the diaminopimelate pathway of Mycobacterium tuberculosis. *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology* **1545**, 67–77 (2001).

221. Sizemore, D. R., Branstrom, A. A. & Sadoff, J. C. Attenuated Shigella as a DNA delivery vehicle for DNA-mediated immunization. *Science (1979)* **207**, 299–303 (1995).

222. Curtiss, R., Galan, J. E., Nakayama, K. & Kelly, S. M. Stabilization of recombinant avirulent vaccine strains in vivo. *Research in Microbiology* **7**, 797–805 (1990).

223. Huang, W. E. *et al.* Chromosomally located gene fusions constructed in Acinetobacter sp. ADP1 for the detection of salicylate. *Environmental Microbiology* **7**, 1339–1348 (2005).

224. NCBI. Basic Local Alignment Search Tool. https://blast.ncbi.nlm.nih.gov/Blast.cgi (2022).

225. Sezonov, G., Joseleau-Petit, D. & D'Ari, R. Escherichia coli physiology in Luria-Bertani broth. *Journal of Bacteriology* **189**, 8746–8749 (2007).

226. Davis, J. H., Rubin, A. J. & Sauer, R. T. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Research* **39**, 1131–1141 (2011).

227. Harris, A. W. K. *et al.* Synthetic negative feedback circuits using engineered small RNAs. *Nucleic Acids Research* **46**, 9875–9889 (2018).

228. Liu, L. & Shaw, P. D. Characterization of dapB, a gene required by Pseudomonas syringae pv. tabaci BR2.024 for lysine and tabtoxinine-β-lactam biosynthesis. *Journal of Bacteriology* **179**, 507–513 (1997).

229. Allard, N. *et al.* A diamonipimelic acid auxotrophic Escherichia coli donor provides improved counterselection following intergenic conjugation with Actinomycetes. *Canadian Journal of Microbiology* **61**, 565–574 (2015).

230. Cox, R. J. The DAP pathway to lysine as a target for antimicrobial agents. *Natural Product Reports* **13**, 29–43 (1996).

231. Vallenet, D. *et al.* Comparative analysis of acinetobacters: Three genomes for three lifestyles. *PLoS ONE* **3**, (2008).

232. Kennerley, J., Chen, J. X. & Huang, W. E. Inducible antisense promoters enable tuning of gene circuits on plasmids, bacterial chromosomes, and in minicells. *BioRxiv - Preprint* (2022).

233. Ratajczak, A., Geißdörfer, W. & Hillen, W. Alkane hydroxylase from Acinetobacter sp. strain ADP1 is encoded by alkM and belongs to a new family of bacterial integral-membrane hydrocarbon hydroxylases. *Appl Environ Microbiol* **64**, 1175–1179 (1998).

234. Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology* **24**, 79–88 (2006).

235. Scapin, G., Reddy, S. G., Zheng, R. & Blanchard, J. S. Three-dimensional structure of Escherichia coli dihydrodipicolinate reductase in complex with NADH and the inhibitor 2,6-pyridinedicarboxylate. *Biochemistry* **36**, 15081–15088 (1997).

236. Fei, X., Bell, T. A., Barkow, S. R., Baker, T. A. & Sauer, R. T. Structural basis of clpxp recognition and unfolding of ssra-tagged substrates. *Elife* **9**, 1–39 (2020).

237. Karzai, A. W., Roche, E. D. & Sauer, R. T. *The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue*. http://structbio.nature.com (2000).

238. Purcell, O., Grierson, C. S., Bernardo, M. D. & Savery, N. J. Temperature dependence of ssrA-tag mediated protein degradation. *Journal of Biological Engineering* vol. 6 Preprint at https://doi.org/10.1186/1754-1611-6-10 (2012).

239. Gottesman, S., Roche, E., Zhou, Y. & Sauer, R. T. *The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system*. www.genesdev.org (1998).

240. Farrell, C. M., Grossman, A. D. & Sauer, R. T. Cytoplasmic degradation of ssrA-tagged proteins. *Molecular Microbiology* **57**, 1750–1761 (2005).

241. Chen, J. X. *et al.* Development of Aspirin-Inducible Biosensors in Escherichia coli and SimCells. *Appl. Environ. Microbiol.* **85**, e02959-18 (2019).

242. Orosz, A., Boros, I. & Venetianer, P. *Analysis of the complex transcription termination region of the Escherichia coli rrnB gene*. *Eur. J. Biochem* vol. 201 (1991).

243. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods* **10**, 354–360 (2013).

244. Wientjes, F. B., Pas, E., Taschner, P. E. M. & Woldringh, C. L. Kinetics of uptake and incorporation of meso-diaminopimelic acid in different Escherichia coli strains. *Journal of Bacteriology* **164**, 331–337 (1985).

245. Wackett, L. P., Frias, J. A., Seffernick, J. L., Sukovich, D. J. & Cameron, S. M. Genomic and biochemical studies demonstrating the absence of an alkane-producing phenotype in Vibrio furnissii M1. *Applied and Environmental Microbiology* **73**, 7192–7198 (2007).

246. Schallschmidt, K. *et al.* Investigation of cell culture volatilomes using solid phase micro extraction: Options and pitfalls exemplified with adenocarcinoma cell lines. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* **1006**, 158–166 (2015).

247. Lehtinen, T., Santala, V. & Santala, S. Twin-layer biosensor for real-time monitoring of alkane metabolism. *FEMS Microbiology Letters* **364**, 1–7 (2017).

248. Lehtinen, T., Virtanen, H., Santala, S. & Santala, V. Production of alkanes from CO2 by engineered bacteria. *Biotechnology for Biofuels* **11**, 228 (2018).

249. Durauer, A., Hobiger, S., Walther, C. & Jungbauer, A. Mixing at the microscale: Power input in shaken microtitre plates. *Biotechnology Journal* 1539–1549 (2016).

250. Steel, H., Habgood, R., Kelly, C. & Papachristodoulou, A. In situ characterisation and manipulation of biological systems with Chi.Bio. *PLoS Biology* **18**, 1–12 (2020).

251. Bucheli-Witschel, M. & Egli, T. Growth of Hydrocarbon-Degrading Bacteria in Continuous Culture. in *Handbook of Hydrocarbon and Lipid Microbiology* 3529–3541 (Springer Berlin Heidelberg, 2010). doi:10.1007/978-3-540-77587-4_276.

252. Hunger, M., Schmucker, R., Kishan, V. & Hillen, W. Analysis and nucleotide sequence of an origin of an origin of DNA replication in Acinetobacter calcoaceticus and its use for Escherichia coli shuttle plasmids. *Gene* **87**, 45–51 (1990).

253. Reed, B., Blazeck, J. & Alper, H. Evolution of an alkane-inducible biosensor for increased responsiveness to short-chain alkanes. *Journal of Biotechnology* **158**, 75–79 (2012).

254. Ju, X., Li, D. & Liu, S. Full-length RNA profiling reveals pervasive bidirectional transcription terminators in bacteria. *Nature Microbiology* **4**, 1907–1918 (2019).

255. Dar, D. & Sorek, R. High-resolution RNA 3-ends mapping of bacterial Rho-dependent transcripts. *Nucleic Acids Research* **46**, 6797–6805 (2018).

256. Sedlyarova, N. *et al.* sRNA-Mediated Control of Transcription Termination in E. coli. *Cell* **167**, 111-121.e13 (2016).

257. Shearwin, K. E., Callen, B. P. & Egan, J. B. Transcriptional interference - A crash course. *Trends in Genetics* **21**, 339–345 (2005).

258. Sneppen, K. *et al.* A mathematical model for transcriptional interference by RNA polymerase traffic in Escherichia coli. *Journal of Molecular Biology* **346**, 399–409 (2005).

259. Crampton, N., Bonass, W. A., Kirkham, J., Rivetti, C. & Thomson, N. H. Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. *Nucleic Acids Research* **34**, 5416–5425 (2006).

260. Giangrossi, M. *et al.* A novel antisense RNA regulates at transcriptional level the virulence gene icsA of Shigella flexneri. *Nucleic Acids Research* **38**, 3362–3375 (2010).

261. Chatterjee, A. *et al.* Convergent transcription confers a bistable switch in Enterococcus faecalis conjugation. *Proc Natl Acad Sci U S A* **108**, 9721–9726 (2011).

262. Bordoy, A. E., Varanasi, U. S., Courtney, C. M. & Chatterjee, A. Transcriptional Interference in Convergent Promoters as a Means for Tunable Gene Expression. *ACS Synthetic Biology* **5**, 1331–1341 (2016).

263. Brophy, J. A. & Voigt, C. A. Antisense transcription as a tool to tune gene expression. *Molecular Systems Biology* **12**, 854 (2016).

264. Hoffmann, S. A., Hao, N., Shearwin, K. E. & Arndt, K. M. Characterizing Transcriptional Interference between Converging Genes in Bacteria. *ACS Synthetic Biology* **8**, 466–473 (2019).

265. O'Connor, N. J., Bordoy, A. E. & Chatterjee, A. Engineering Transcriptional Interference through RNA Polymerase Processivity Control. *ACS Synthetic Biology* **10**, 737–748 (2021).

266. Winson, M. K. *et al.* Engineering the luxCDABE genes from Photorhabdus luminescens to provide a bioluminescent reporter for constitutive and promoter probe plasmids and mini-Tn5 constructs. *FEMS Microbiology Letters* **163**, 193–202 (1998).

267. Bacher, J. M., Metzgar, D. & De Crécy-Lagard, V. Rapid evolution of diminished transformability in Acinetobacter baylyi. *Journal of Bacteriology* **188**, 8534–8542 (2006).

268. Brosius, J., Erfle, M. & Storella, J. Spacing of the -10 and -35 regions in the tac promoter. Effect on its in vivo activity. *Journal of Biological Chemistry* **260**, 3539–3541 (1985).

269. Suárez, G. A. *et al.* Rapid and assured genetic engineering methods applied to Acinetobacter baylyi ADP1 genome streamlining. *Nucleic Acids Res* **48**, 4585–4600 (2020).

270. Suarez, G., Renda, B., Zhang, X., Gifford, I. & Barrick, J. E. Transforming A. baylyi ADP1. https://barricklab.org/twiki/bin/view/Lab/ProtocolsAcinetobacterTransformation (2020).

271. New England Biolabs. NEBuilder HiFi DNA assembly: Bridging dsDNA with a ssDNA oligo. *NEB Protocols* https://international.neb.com/tools-and-resources/video-library/nebuilder-hifi-dna-assembly-bridging-dsdna-with-a-ssdna-oligo (2022).

272. Bryksin, A. v. & Matsumura, I. Rational design of a plasmid origin that replicates efficiently in both gram-positive and gram-negative bacteria. *PLoS ONE* **5**, (2010).

273. Chen, J. C. *et al.* Jungle Express is a versatile repressor system for tight transcriptional control. *Nature Communications* **9**, 1–13 (2018).

274. Kudo, H., Nawa, R., Hayashi, Y. & Arai, M. Comparison of aldehyde-producing activities of cyanobacterial acyl-(acyl carrier protein) reductases. *Biotechnology for Biofuels* **9**, 1–15 (2016).

275. Reiser, S. & Somerville, C. Isolation of mutants of Acinetobacter calcoaceticus deficient in wax ester synthesis and complementation of one mutation with a gene encoding a fatty acyl coenzyme A reductase. *Journal of Bacteriology* **179**, 2969–2975 (1997).

276. Bernard, A. *et al.* 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**, 3106–3118 (2012).

277. Kralik, P. & Ricchi, M. A basic guide to real time PCR in microbial diagnostics: Definitions, parameters, and everything. *Frontiers in Microbiology* **8**, 1–9 (2017).

278. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Research* **40**, e115–e115 (2012).

279. Asperger, O. & Kleber, H. Metabolism of alkanes by Acinetobacter. in *The biology of Acinetobacter* (eds. Towner, K., Bergogne-Berezin, E. & Fewson, C.) 323–350 (Springer, 1991). doi:https://doi.org/10.1007/978-1-4899-3553-3_22.

280. Van Beilen, J. B., Penninga, D. & Witholt, B. Topology of the membrane-bound alkane hydroxylase of Pseudomonas oleovorans. *Journal of Biological Chemistry* **267**, 9194–9201 (1992).

281. Parche, S., Geißdörfer, W. & Hillen, W. Identification and characterization of xcpR encoding a subunit of the general secretory pathway necessary for dodecane degradation in Acinetobacter calcoaceticus ADP1. *Journal of Bacteriology* **179**, 4631–4634 (1997).

282. Tjørve, K. M. C. & Tjørve, E. The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family. *PLOS ONE* **12**, e0178691 (2017).

283. Illumina. qPCR Quantification Protocol Guide. Preprint at (2009).

284. Wang, D. *et al.* Advances in single cell Raman spectroscopy technologies for biological and environmental applications. *Current Opinion in Biotechnology* **64**, 218–229 (2020).

285. Simon, C. & Daniel, R. Construction of Small-Insert and Large-Insert Metagenomic Libraries. in *Methods in Molecular Biology: Metagenomics - Methods and Protocols* (eds. Streit, W. R. & Daniel, R.) 1–12 (Humana Press, 2017). doi:10.1007/978-1-4939-6691-2.

286. Copp, J. N., Hanson-Manful, P., Ackerley, D. F. & Patrick, W. M. Error Prone PCR and Effective Generation of Gene Variant Libraries for Directed Evolution. in *Methods in Molecular Biology: Directed Evolution Library Creation - Methods and Protocols* (eds. Gillam, E. M. J., Copp, J. N. & Ackerley, D. F.) 3–22 (Humana Press, 2014). doi:10.1007/978-1-4939-1053-3.

287. Patwardhan, R. P. *et al.* High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nature Biotechnology* **27**, 1173–1175 (2009).

288. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: Technologies and applications. *Nature Methods* **11**, 499–507 (2014).

289. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343–345 (2009).

290. Marsischky, G. & LaBaer, J. Many paths to many clones: A comparative look at high-throughput cloning methods. *Genome Research* **14**, 2020–2028 (2004).

291. Marchuk, D., Drumm, M., Saulino, A. & Collins, F. S. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Research* **19**, 1154 (1991).

292. Davidsson, M. *et al.* A novel process of viral vector barcoding and library preparation enables high-diversity library generation and recombination-free paired-end sequencing. *Scientific Reports* **6**, 1–18 (2016).

293. Gao, Y. *et al.* Structural insights into catalytic mechanism and product delivery of cyanobacterial acyl-acyl carrier protein reductase. *Nature Communications* **11**, 1–11 (2020).

294. Andre, C., Kim, S. W., Yu, X.-H. & Shanklin, J. 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* **110**, 3191–3196 (2013).

295. Lam, K. N., Cheng, J., Engel, K., Neufeld, J. D. & Charles, T. C. Current and future resources for functional metagenomics. *Frontiers in Microbiology* **6**, 1196 (2015).

296. Kim, H. M., Chae, T. U., Choi, S. Y., Kim, W. J. & Lee, S. Y. Engineering of an oleaginous bacterium for the production of fatty acids and fuels. *Nature Chemical Biology* **15**, 721–729 (2019).

297. Lee, M. J., Palmer, D. J. & Warren, M. J. Biotechnological Advances in Bacterial Microcompartment Technology. *Trends in Biotechnology* **37**, 325–336 (2019).

298. Sharma, A. & Yazdani, S. S. Microbial engineering to produce fatty alcohols and alkanes. *J Ind Microbiol Biotechnol* **48**, (2021).

299. Li, J. *et al.* Synthesis of high-titer alka(e)nes in Yarrowia lipolytica is enabled by a discovered mechanism. *Nature Communications* **11**, (2020).

300. Xu, P., Qiao, K., Ahn, W. S. & Stephanopoulos, G. Engineering *Yarrowia lipolytica* as a platform for synthesis of drop-in transportation fuels and oleochemicals. *Proceedings of the National Academy of Sciences* **113**, 10848–10853 (2016).

301. Hammerstrom, T. G., Beabout, K., Clements, T. P., Saxer, G. & Shamoo, Y. Acinetobacter baumannii repeatedly evolves a hypermutator phenotype in response to tigecycline that effectively surveys evolutionary trajectories to resistance. *PLoS ONE* **10**, 1–24 (2015).

302. Tegova, R., Tover, A., Tarassova, K., Tark, M. & Kivisaar, M. Involvement of error-prone DNA polymerase IV in stationary-phase mutagenesis in Pseudomonas putida. *Journal of Bacteriology* **186**, 2735–2744 (2004).

303. Greener, A., Callahan, M. & Jerpseth, B. An Efficient Random Mutagenesis Technique Using an E. coli Mutator Strain. *Applied Biochemistry and Biotechnology - Part B Molecular Biotechnology* **7**, 189–195 (1997).

304. Tee, K. L. & Wong, T. S. Polishing the craft of genetic diversity creation in directed evolution. *Biotechnology Advances* **31**, 1707–1721 (2013).

305. Duffy, K., Arangundy-franklin, S. & Holliger, P. Modified nucleic acids : replication , evolution , and next-generation therapeutics. *BMC Biology* **18**, 1–14 (2020).

306. Liu, C. *et al.* Phosphonomethyl Oligonucleotides as Backbone-Modified Artificial Genetic Polymers. *J Am Chem Soc* **140**, 6690–6699 (2018).

307. Houlihan, G., Arangundy-Franklin, S. & Holliger, P. Engineering and application of polymerases for synthetic genetics. *Current Opinion in Biotechnology* **48**, 168–179 (2017).

308. Eremeeva, E. & Herdewijn, P. Enzymatic synthesis using polymerases of modified nucleic acids and genes. in *Enzymatic and Chemical Synthesis of Nucleic Acid Derivatives* (ed. Fernandez-Lucas, J.) 159–194 (2018). doi:10.1002/9783527812103.ch7.

309. Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A. & Liu, C. C. Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell* **175**, 1946-1957.e13 (2018).

310. De Crécy, E. *et al.* Development of a novel continuous culture device for experimental evolution of bacterial populations. *Applied Microbiology and Biotechnology* **77**, 489–496 (2007).

311. Tegally, H., San, J. E., Giandhari, J. & de Oliveira, T. Unlocking the efficiency of genomics laboratories with robotic liquid-handling. *BMC Genomics* **21**, 1–15 (2020).

312. Horinouchi, T., Minamoto, T., Suzuki, S., Shimizu, H. & Furusawa, C. Development of an Automated Culture System for Laboratory Evolution. *Journal of Laboratory Automation* **19**, 478–482 (2014).

313. Takahashi, C. N., Miller, A. W., Ekness, F., Dunham, M. J. & Klavins, E. A low cost, customizable turbidostat for use in synthetic circuit characterization. *ACS Synthetic Biology* **4**, 32–38 (2015).

314. Hoffmann, S. A., Wohltat, C., Müller, K. M. & Arndt, K. M. A user-friendly, low-cost turbidostat with versatile growth rate estimation based on an extended Kalman filter. *PLoS ONE* **12**, 1–15 (2017).

315. Matteau, D., Baby, V., Pelletier, S. & Rodrigue, S. A small-volume, low-cost, and versatile continuous culture device. *PLoS ONE* **10**, 1–13 (2015).

316. Hershberg, R. Mutation—the engine of evolution: Studying mutation and its role in the evolution of bacteria. *Cold Spring Harbor Perspectives in Biology* **7**, 1–14 (2015).

317. Madison, A. C. *et al.* Scalable Device for Automated Microbial Electroporation in a Digital Microfluidic Platform. *ACS Synthetic Biology* **6**, 1701–1709 (2017).

318. Ng, C. K. *et al.* Genetic engineering biofilms in situ using ultrasound-mediated DNA delivery. *Microbial Biotechnology* **14**, 1580–1593 (2021).

319. Bauchop, T. & Elsden, S. R. The growth of micro-organisms in relation to their energy supply. *Microbiology (N Y)* **23**, 457–469 (1960).

Permissions

Figure 2 has been republished from the World Energy Outlook 2021, produced by the International Energy Agency (IEA), in accordance with the IEA's copyright terms and conditions, which can be found at https://www.iea.org/terms. These terms stipulate that an insubstantial amount of material may be reproduced in public work, such as this thesis, provided proper attribution is made and that figures are reproduced in their entirety, with only stylistic modification to their content. Additionally, as part of these terms, it should be noted that this thesis is a work derived by Joseph Kennerley (the author) which makes use of IEA material, and that the author is solely liable and responsible for this derived work. The derived work is not endorsed by the IEA in any manner.

Figure 4 has been republished from Arai *et al.* 2018, without modification, under an agreement between Springer Nature (the publisher and copyright owner) and Joseph Kennerley (the author of this thesis), granting a personal, non-exclusive, non-transferable, world-wide license to reproduce the licensed material for the purpose of presentation in this thesis. Details of this license are provided below:

License Content Title:	Cyanobacterial Enzymes for Bioalkane Production (Chapter 6 of *Synthetic Biology of Cyanobacteria*, 2018)
License Content Authors:	Munehito Arai, Yuuki Hayashi, Hisashi Kudo
License Content	Publisher: Springer Nature
License Number:	5357541012447
Licensed Content Format:	Figures/tables/illustrations (quantity, 1)

Figure 7, both parts (a) and (b), has been republished from Zhang *et al.* 2012, with minor edits, and with written permission from the corresponding author, Prof. Wei Huang.

Part (b) of *Figure 23* has been republished from Lehtinen *et al.* 2018, with minor edits, in this thesis under the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction of this material in any medium, contingent upon proper attribution to the authors and reference to the source. A link to this license is provided here: http://creativecommons.org/licenses/by/4.0/

Supplementary Information

For Chapter 3

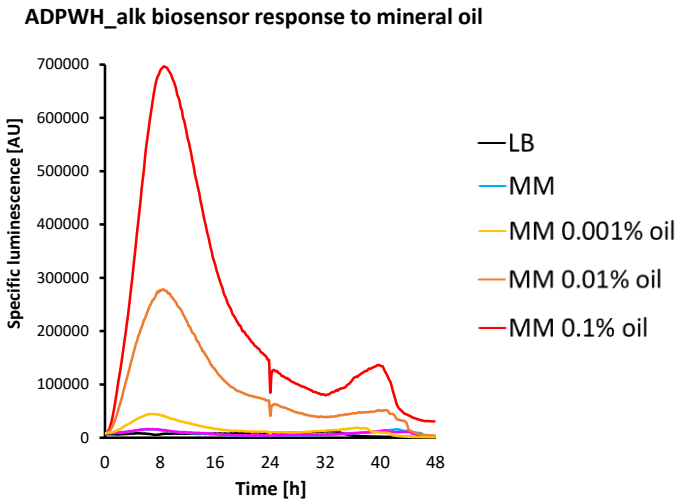


Figure S.1. ADPWH_alk biosensor (*Acinetobacter baylyi* ADP1 ⊿*alkM*::*luxCDABE*) luminescence in varying concentrations of exogenous mineral oil (%, v/v), in ADP1 minimal media (MM) with 20mM succinate as a carbon source. Data from 48 hours growth in 96-well microtitre plate. The ADPWH_alk biosensor strain was constructed by Zhang *et al.*[162] and provided for use in the Artevo project by Prof. Wei Huang. Data in this figure was obtained as part of this thesis to validate the behaviour of ADPWH_alk.

Minimal medium (modified)

For culturing of *Acinetobacter baylyi* ADP1. The following recipe makes 1 L of medium.

Sodium phosphate, dibasic	Na_2HPO_4	2.5 g
Potassium phosphate, monobasic	KH_2PO_4	2.5 g
Ammonium chloride	NH_4Cl	1.0 g
Magnesium sulfate heptahydrate	$MgSO_4.7H_2O$	0.1 g

Dissolve in 950mL of purified water, then add:

Saturated calcium chloride solution	$CaCL_2$	10 μL
Saturated iron sulfate solution	$FeSO_4$	10 μL

Supplement this with 1 mL of the following trace minerals solution, which is made by dissolving the components listed below in 1 L of purified water, as per Bauchop and Elsden[319]:

Magnesium sulfate	$MgSO_4$	10.75 g
Iron sulfate heptahydrate	$FeSO_4.7H_2O$	4.5 g
Calcium carbonate	$CaCO_3$	2.0 g
Zinc sulfate heptahydrate	$ZnSO_4.7H_2O$	1.44 g
Manganese sulfate tetrahydrate	$MnSO_4.4H_2O$	1.12 g
Copper sulfate pentahydrate	$CuSO_4.5H_2O$	0.25 g
Cobalt sulfate heptahydrate	$CoSO_4.7H_2O$	0.28 g
Boric acid	H_3BO_3	0.06 g
Hydrochloric acid (concentrated)	HCl	51.3 mL

Adjust pH to 6.8 using HCl solution and make up volume to 1 L with purified water. Autoclave medium prior to use.

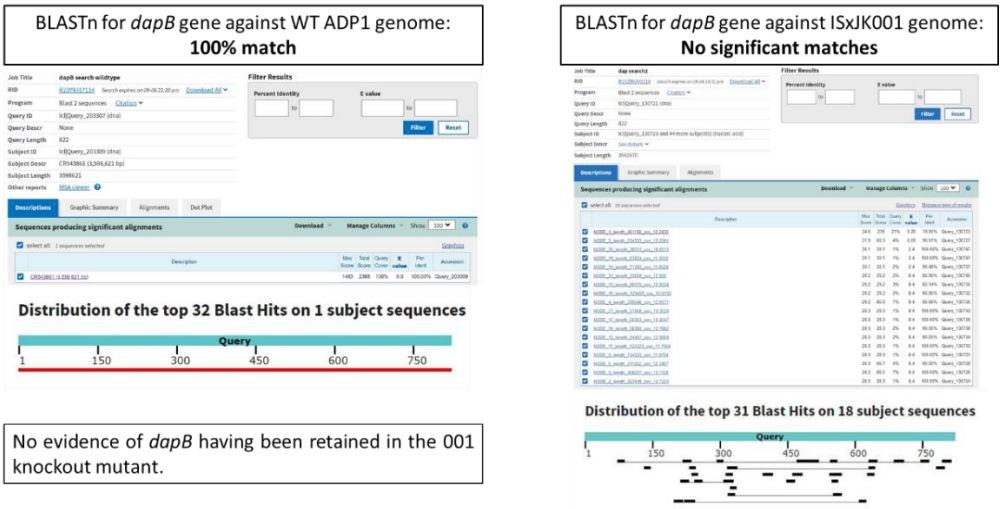


Figure S.2. BLASTn of *dapB* gene sequence against genome sequencing results from a novel *dapB* deletion mutant in *Acinetobacter baylyi* ADP1 and ADP1-ISx.

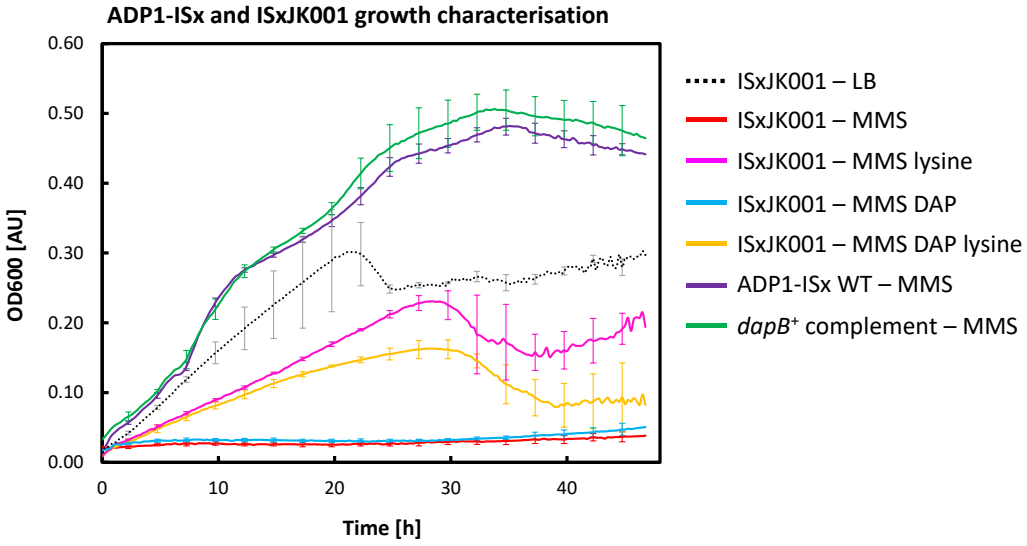


Figure S.3. Restoration of growth of ISxJK001 strains in the presence of supplemental lysine. Growth rescue was not observed in the presence of supplemental *meso*-DAP alone. Strains were grown in 96-well microtitre plate over 48 hours. Error shown is one standard deviation. n = 3 biological replicates with distinct transformation lineages.

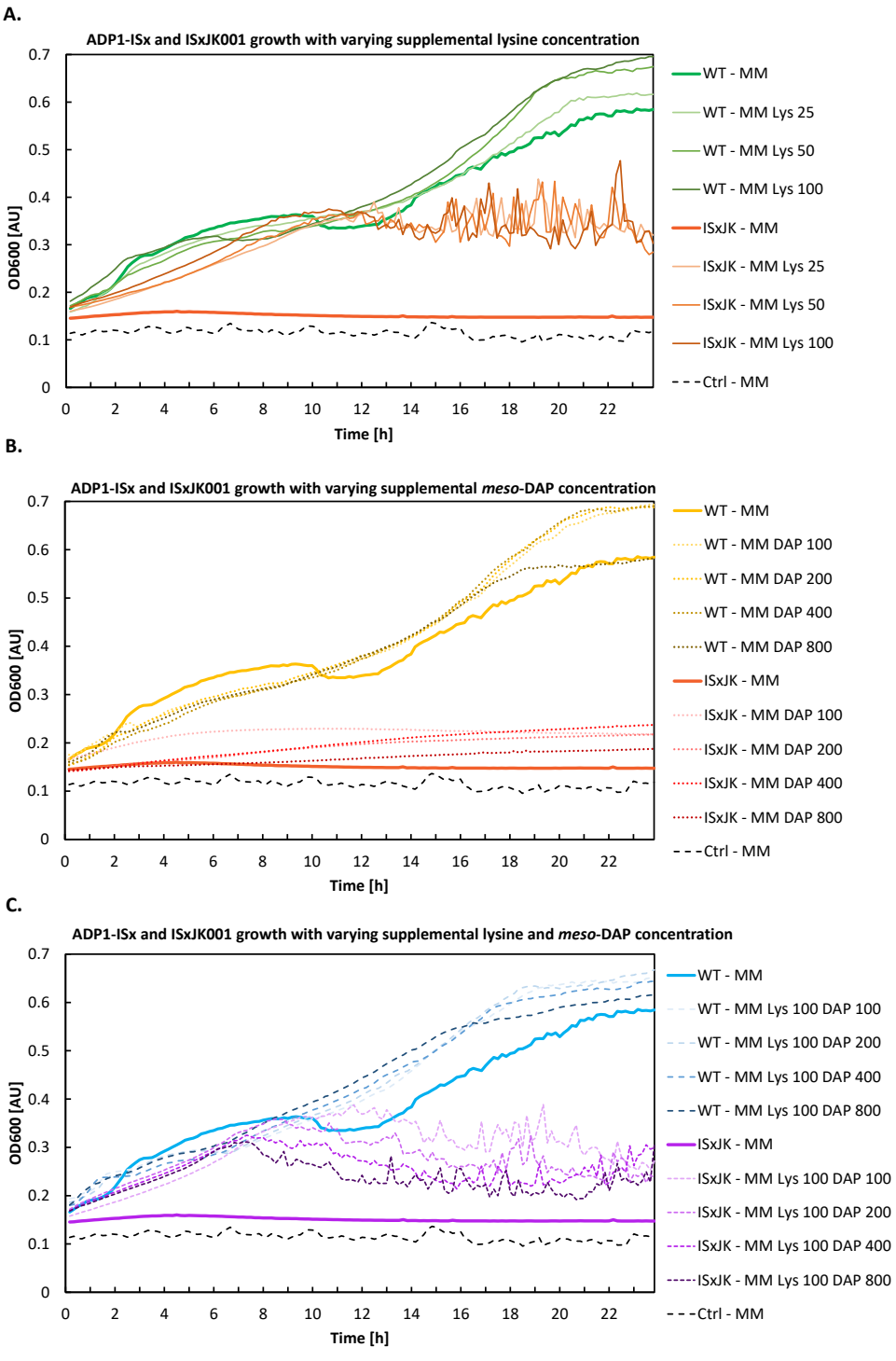


Figure S.4. Impact of varying lysine (lys) (A.), *meso*-DAP (DAP) (B.), and combine lysine and *meso*-DAP (C.) concentration (in µg/mL) on growth of WT ADP1-ISx and ISxJK001 mutants. ISxJK001 auxotrophs appear to exhaust available lysine reserves after approximately 8 hours growth. n = 3 biological replicates.

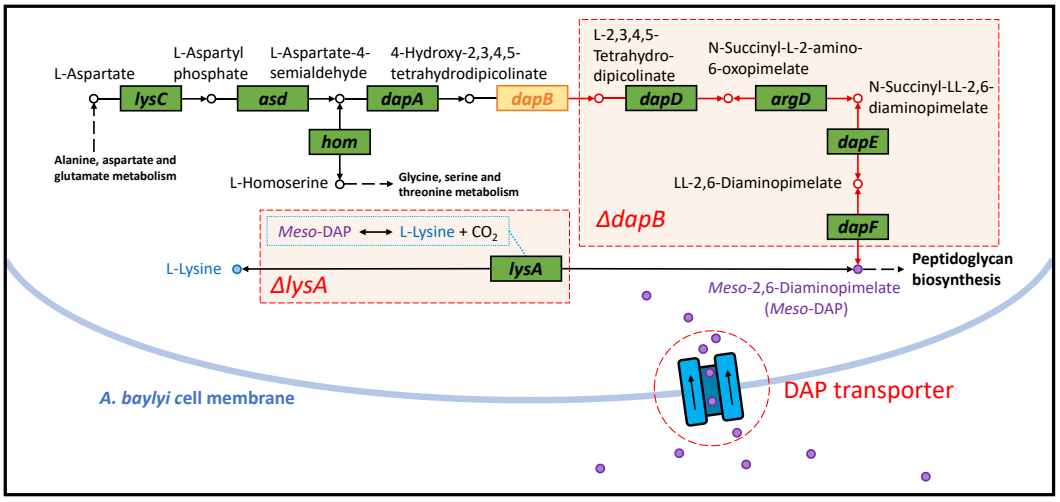


Figure S.5. Investigation of *meso*-DAP transport in *A. baylyi* ADP1. Red boxes indicate the metabolic pathways cut off by deletion of *lysA* and *dapB* genes. Examination of growth restoration in the presence of either supplemental lysine or *meso*-DAP in ⊥*dapB* mutants (which have been shown to be auxotrophic for lysine), ⊥*lysA* mutants (which have been obtained – see *Figure S.6*, below – but may have an uncharacterised metabolic route to lysine), and double mutants (⊥*lysA* ⊥*dapB*; not yet obtained), should provide insights into *meso*-DAP transport mechanisms in ADP1.

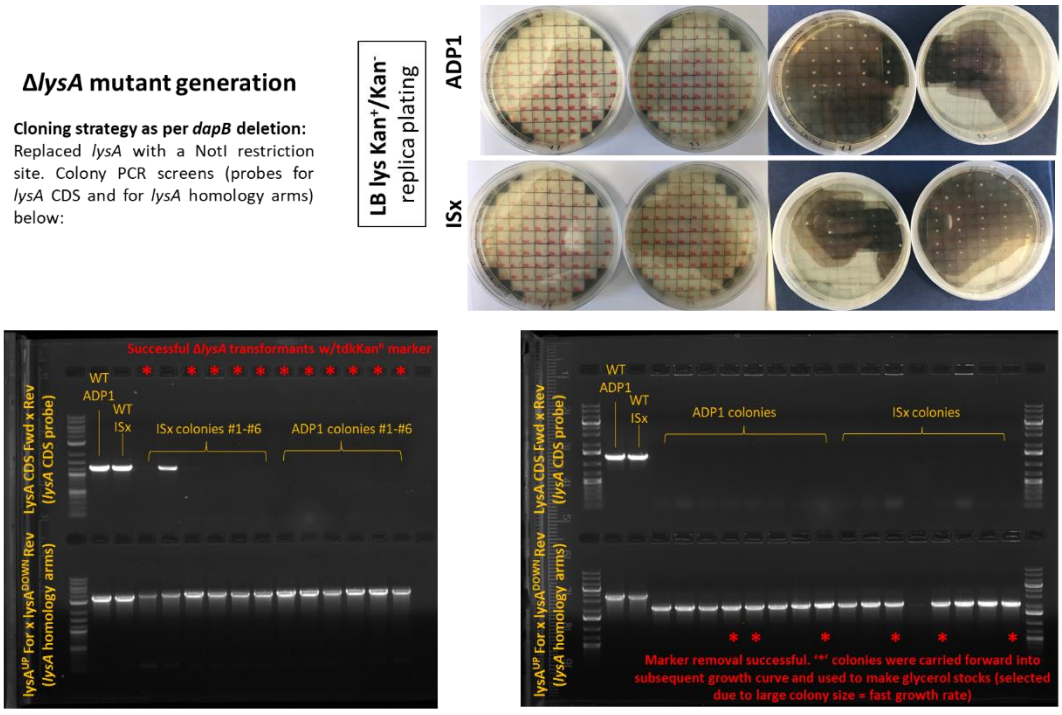


Figure S.6. Generating a ⊥*lysA* mutant in *A. baylyi* ADP1-ISx. The method used was homology-mediated targeted recombination with *tdk*.Kan^R counterselection, as described for the novel ADP1-ISx ⊥*dapB* mutant in Chapter 3.

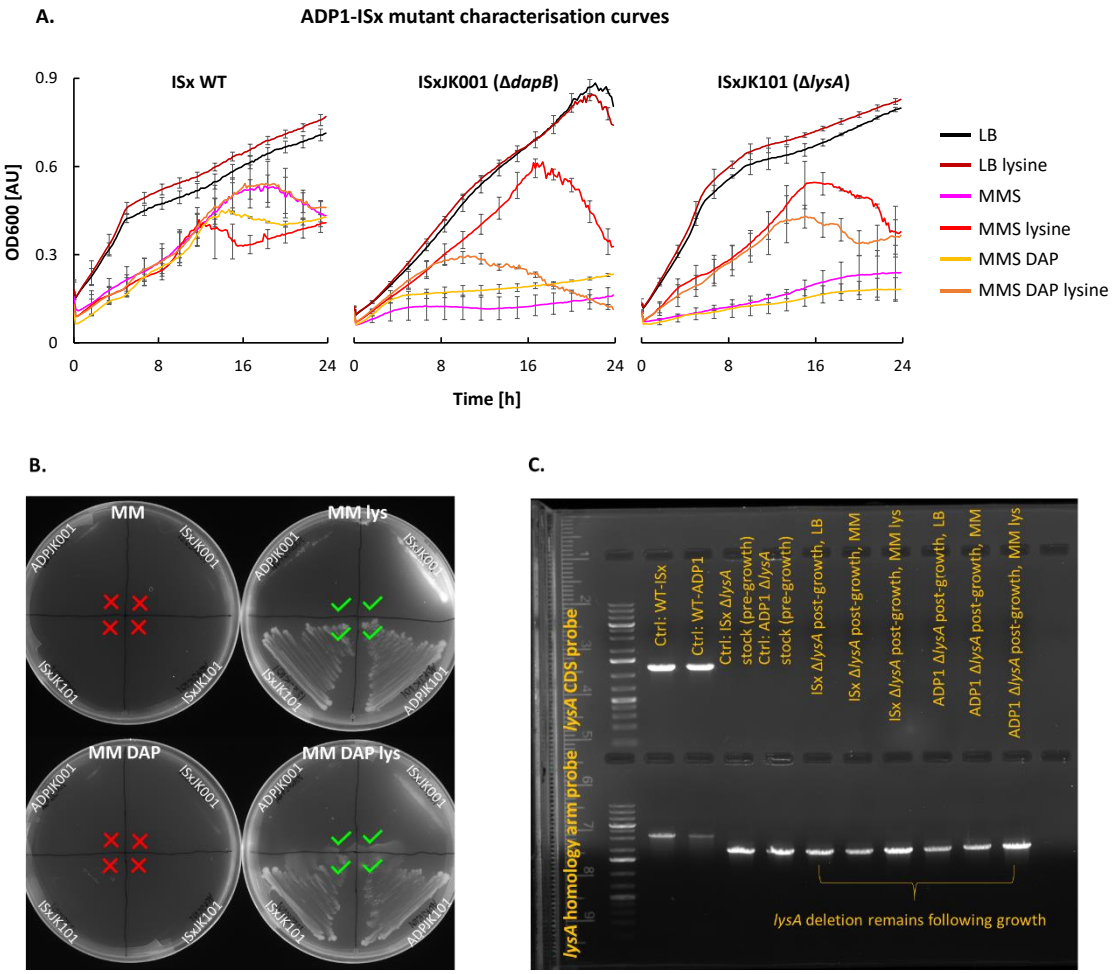


Figure S.7. Characterisation of ⊿*lysA* mutants (101) and comparison with ⊿*dapB* mutants (001), in liquid and on solid media. Part (a) shows growth in microtitre plate of WT, 001, and 101 strains, in the presence of supplemental lysine and *meso*-DAP. Part (b) shows growth recovery on solid agar. Part (c) shows results of a colony PCR screen of mutants to demonstrate *lysA* has not been reacquired during growth.

For Chapter 4

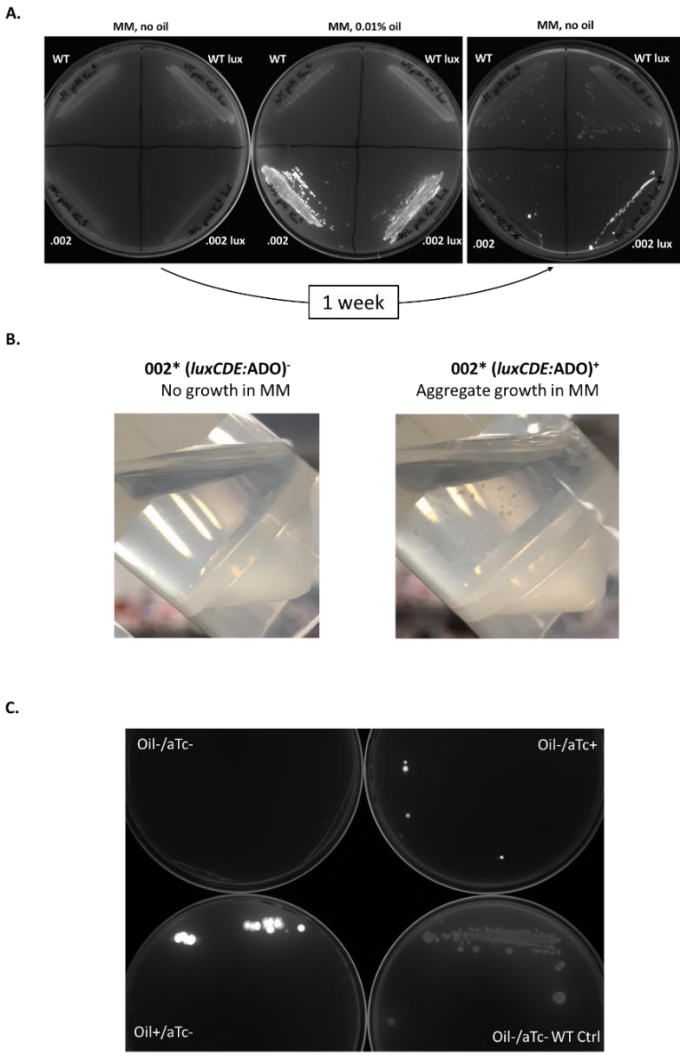


Figure S.8. First attempts at incorporating endogenous alkane production into the early Artevo system. Part (a) shows growth of WT ADP1-ISx and ISxJK002, both with and without the alkane biosynthesis plasmid pWH1274 *luxCDE:ADO*, on solid minimal agar with and without supplemental mineral oil. MM = minimal media (carbon source in all cases here is 20mM succinate). Growth and fluorescence of .002 pWH1274 *luxCDE:ADO* strains is observable after 24 hours in the presence of exogenous alkane, with growth greatly reduced in the absence of any supplemental oil. Some growth and fluorescence can be seen in .002 pWH1274 *luxCDE:ADO* after 1 week of growth, possibly indicating that expression of the alkane biosynthesis module is very low or that production of alkanes is otherwise limited in these conditions. The lack of comparable fluorescence in .002 strains without pWH1274 *luxCDE:ADO* implies this effect is not due to leakiness in the Artevo circuit. Part (b) shows aggregate growth of ISxJK002 in liquid MM when the *luxCDE:ADO* module is present. Part (c) shows restreaks on minimal agar, with and without mineral oil (0.01%, v/v) and aTc inducer (200nM), from glycerol stocks of a high-performing replicate ISxJK002 pWH1274 *luxCDE:ADO* strain. Plates imaged using a FITC filter to capture *gfp* fluorescence. Observed fluorescence pattern suggests alkane biosynthesis may be occurring: no growth or fluorescence in the absence of both supplemental oil and inducer; some growth and fluorescence observed when inducer, but no oil, is present; higher growth and fluorescence in the presence of exogenous mineral oil but no inducer.

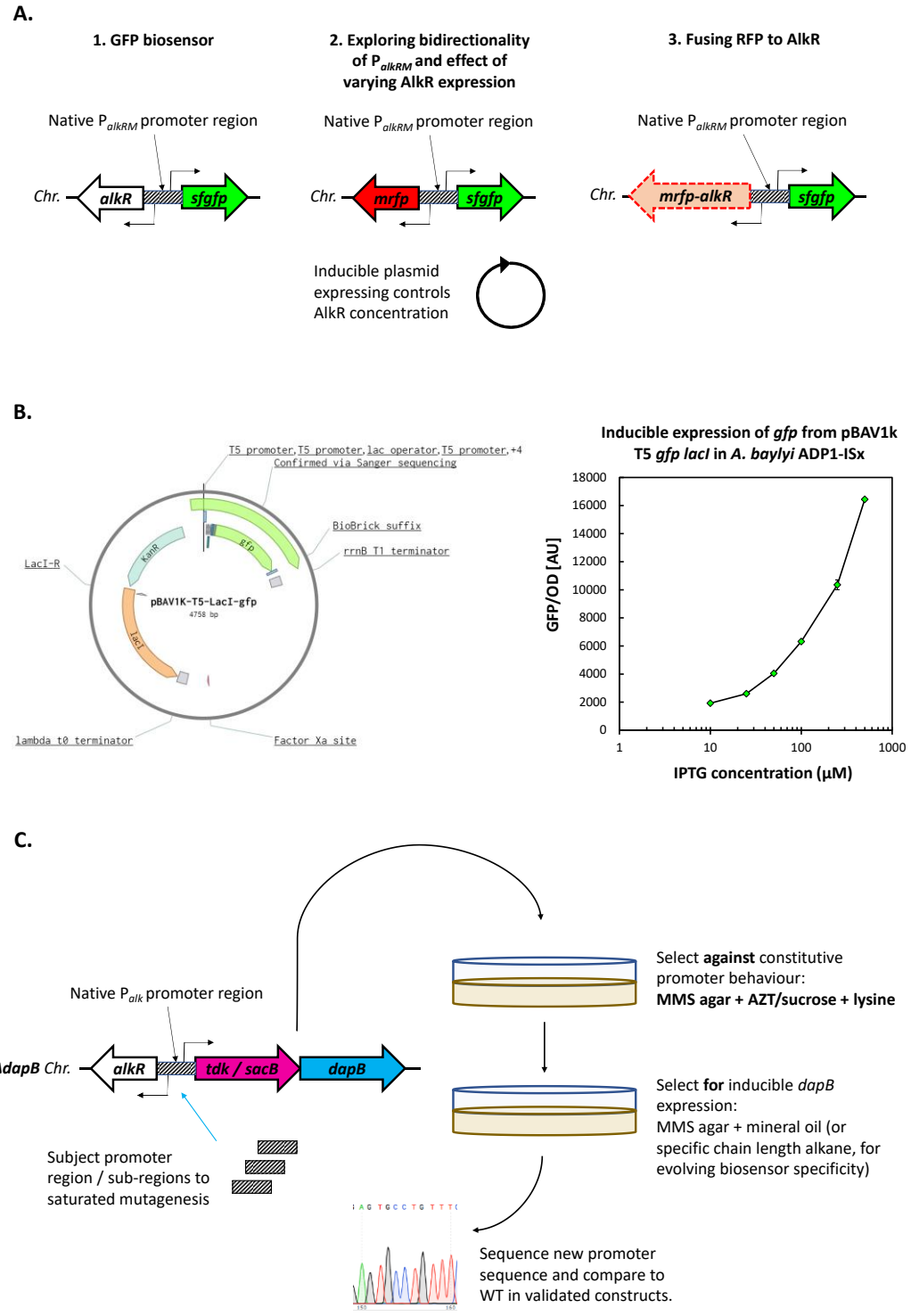


Figure S.9. Strategy for exploring bi-directionality of *alk*RM promoter region in *A. baylyi* ADP1 and for potential subsequent promoter engineering. Part (a), replacing *alk*R and *alk*M with *mrfp* and *sfgfp*, and placing *alk*R under control on an inducible promoter on a separate plasmid. *alk*R expression could then be varied and relative expression of *mrfp* and *sfgfp* tracked. Part (b), pBAV1k T5 *gfp* plasmid constructed as a proof-of-concept of inducible behaviour in ADP1. Jungle Express system originally intended for this purpose was found to be non-functional in ADP1, even with codon optimisation of the *eil*R regulator. Part (c) shows a hypothetical strategy for achieving tighter control in the *alk*RM biosensor system through directed evolution; cycles of mutagenesis followed by positive and negative selection, then sequencing of high-performing mutants. This strategy was developed jointly with Dr Rob Habgood.

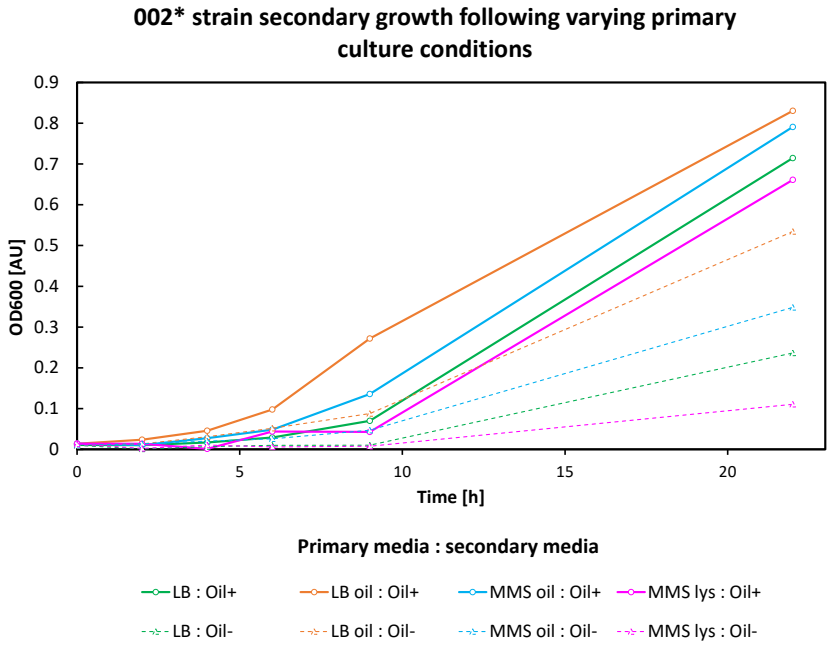


Figure S.10. Effect of primary inoculum on secondary growth in .002 strains. Prior growth in mineral oil appears to improve secondary growth, regardless of whether oil is present in the secondary media. Oil in secondary media improves growth in all cases, as expected.

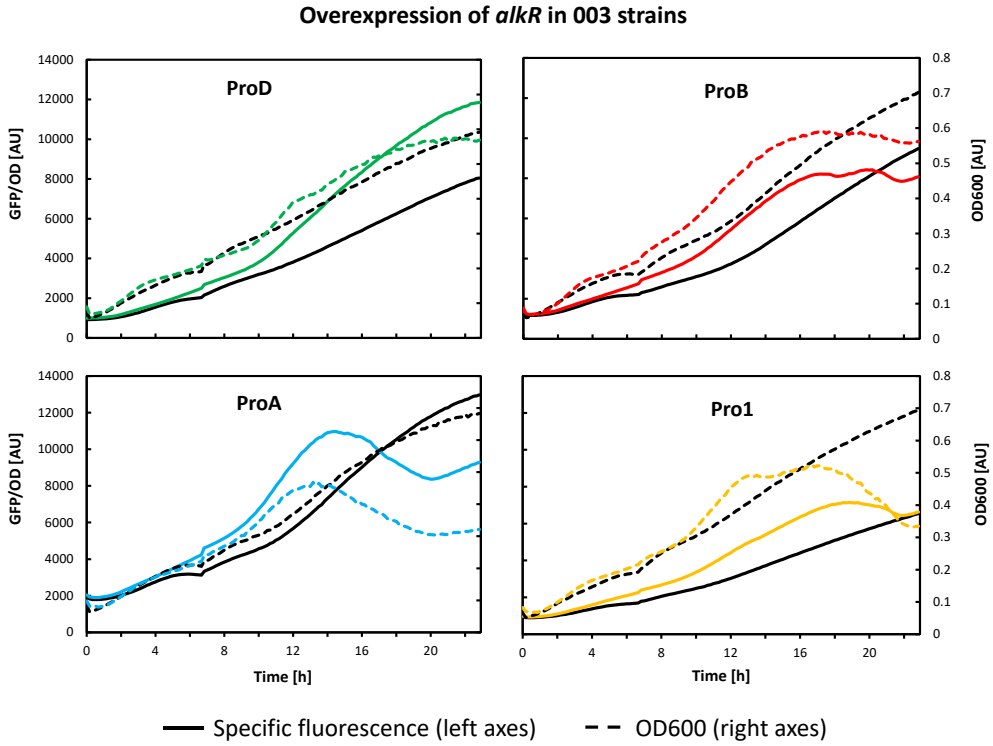


Figure S.11. Effect of overexpression of *alkR* on growth (right axes, solid lines) and fluorescence (left axes, dashed lines) response to alkanes in 003 strains. *alkR* was overexpressed using Pro-series promoters (see *Table 2* in Chapter 3). Black lines (dashed and solid) represent strains grown in the absence of mineral oil. Coloured lines (dashed and solid) represent strains grown with 0.01% (v/v) supplemental mineral oil. n = 3 biological replicates, error bars omitted for clarity.

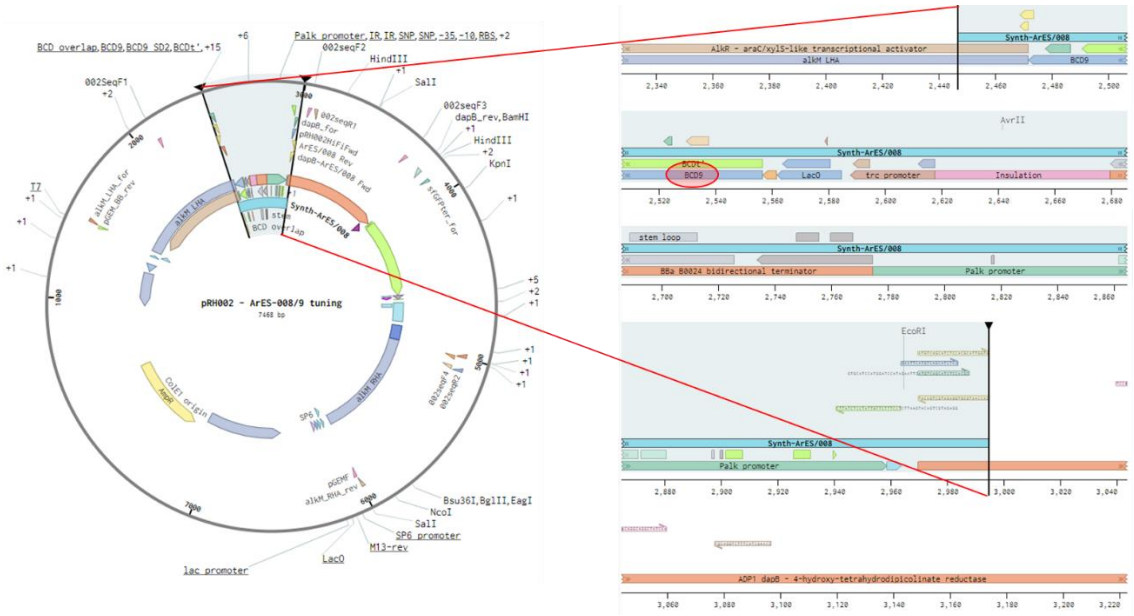


Figure S.12. pRH002 plasmid carrying the insulated Artevo circuit. Overexpression of *alkR* is facilitated by a strong, constitutive bi-cistronic design (BCD) promoter, BCD9, from the Biggs *et al.* toolkit[159]. The fluorescent reporter, *gfp*, found in other Artevo circuits, has been replaced here with superfolder *gfp* (*sfgfp*) to improve signal strength.

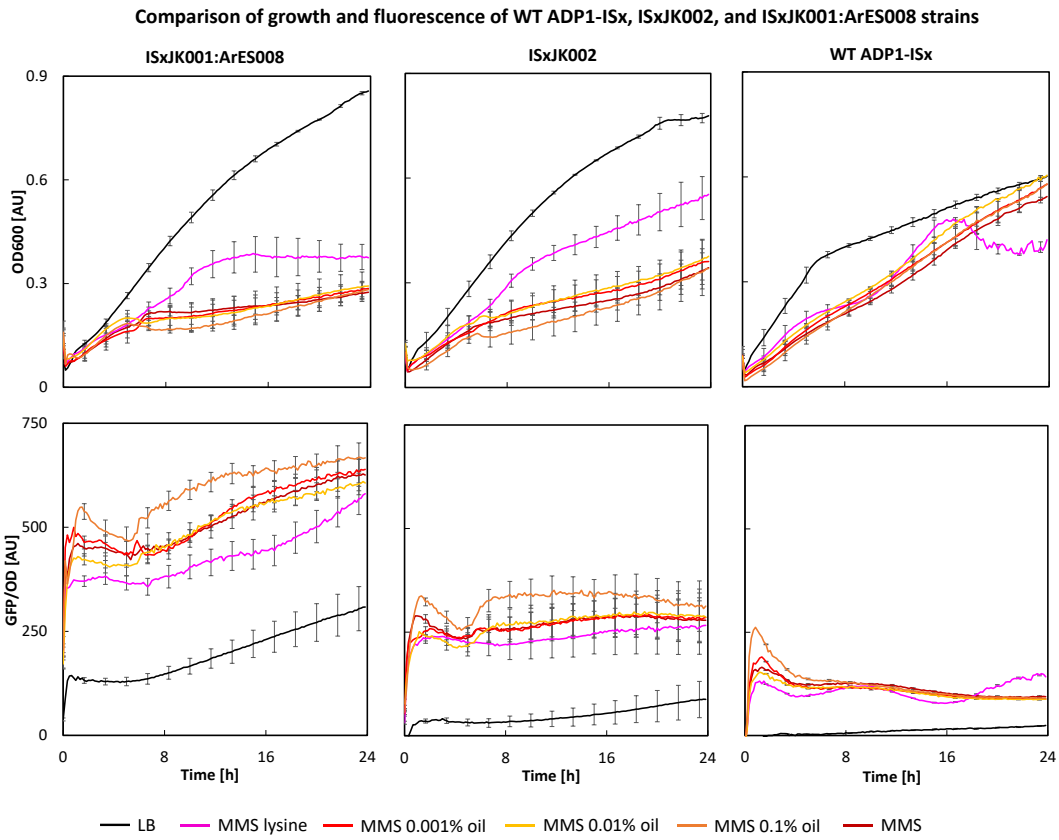
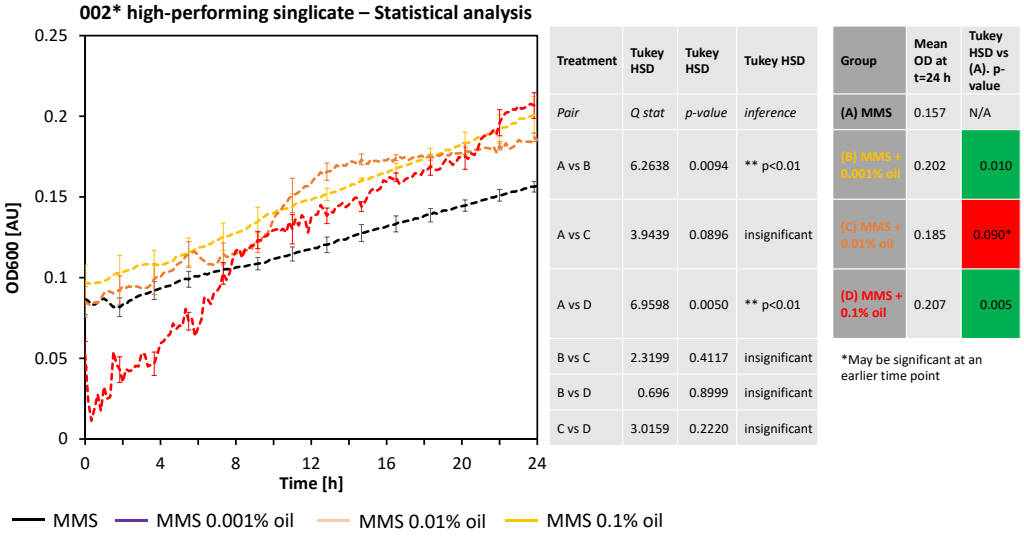


Figure S.13. Characterisation of growth and specific fluorescence, in various media, of the ISxJK001:ArES008 strain (which contain the insulated Artevo circuit shown in *Figure S.12*). Overexpression of *alkR* is facilitated by the bi-cistronic design (BCD) promoter BCD9. Growth and fluorescence are compared to existing ISxJK002 strains and WT ADP1-ISx. Cultures grown in 96-well microtitre plate, as previously described. Error shown is one standard deviation (n = 3 biological replicates).



Treatment	Tukey HSD	Tukey HSD	Tukey HSD
Pair	*Q stat*	*p-value*	*inference*
A vs B	6.2638	0.0094	** p<0.01
A vs C	3.9439	0.0896	insignificant
A vs D	6.9598	0.0050	** p<0.01
B vs C	2.3199	0.4117	insignificant
B vs D	0.696	0.8999	insignificant
C vs D	3.0159	0.2220	insignificant

Group	Mean OD at t=24 h	Tukey HSD vs (A). p-value
(A) MMS	0.157	N/A
(B) MMS + 0.001% oil	0.202	0.010
(C) MMS + 0.01% oil	0.185	0.090*
(D) MMS + 0.1% oil	0.207	0.005

*May be significant at an earlier time point

Figure S.14. Statistical analysis of growth in the 002* strain (high-performing singlicate) in the presence of varying concentrations of mineral oil. Error bars are +/- one standard deviation (n = 3). Significance evaluated based on mean OD at t = 24h.

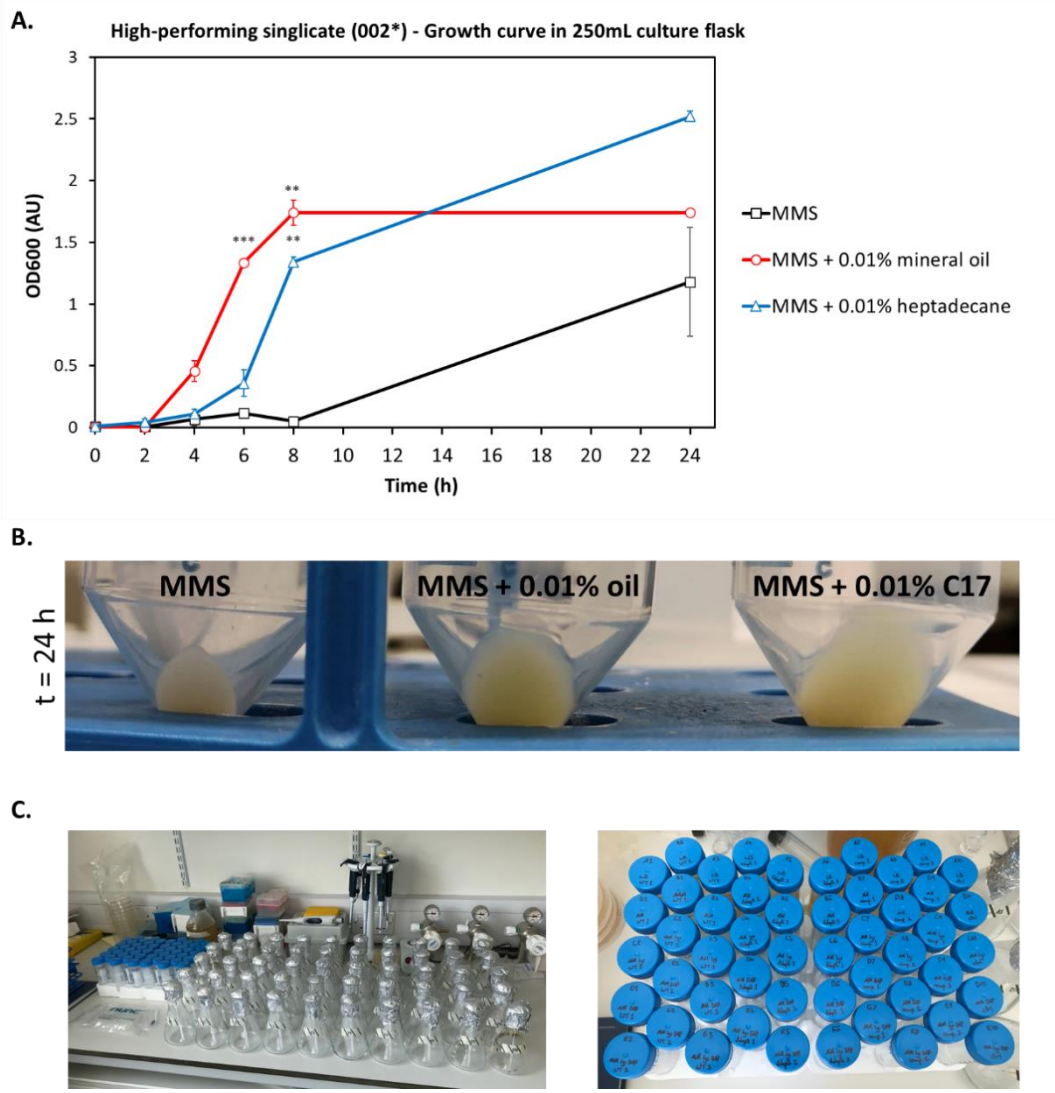


Figure S.15. Effect of culture vessel on apparent leakiness in the Artevo circuit. Parts (a) shows growth of the high-performing singlicate (002*) *⊿dapB ⊿alkM::dapB gfp* strain in MMS media with and without supplemental mineral oil or heptadecane. Strains were grown aerobically in culture flasks (n=3) in MMS, MMS + 0.01% (v/v) mineral oil, and MMS + 0.01% (v/v) heptadecane, for 24 hours (30°C, 150 rpm). OD600 was measured every 2 hours for 8 hours, then again at 24 hours. One replicate of each culture condition was centrifuged to produce the qualitative comparison of cell pellets shown in part (b). A F-test was performed to test equivalence of variances, then a T-test was performed to test for significant difference between each hydrocarbon condition against the MMS-only control (** = $p \leq 0.01$, *** = $p \leq 0.001$). Part (c) illustrates the increase in experimental workload inherent to upscaling a combinatorial investigation of growth conditions in culture flask, compared to microtitre plate. Given the time required to manually sample a large number of culture flasks, there may be error introduced to the resultant growth curves between those flasks sampled first and those sampled last.

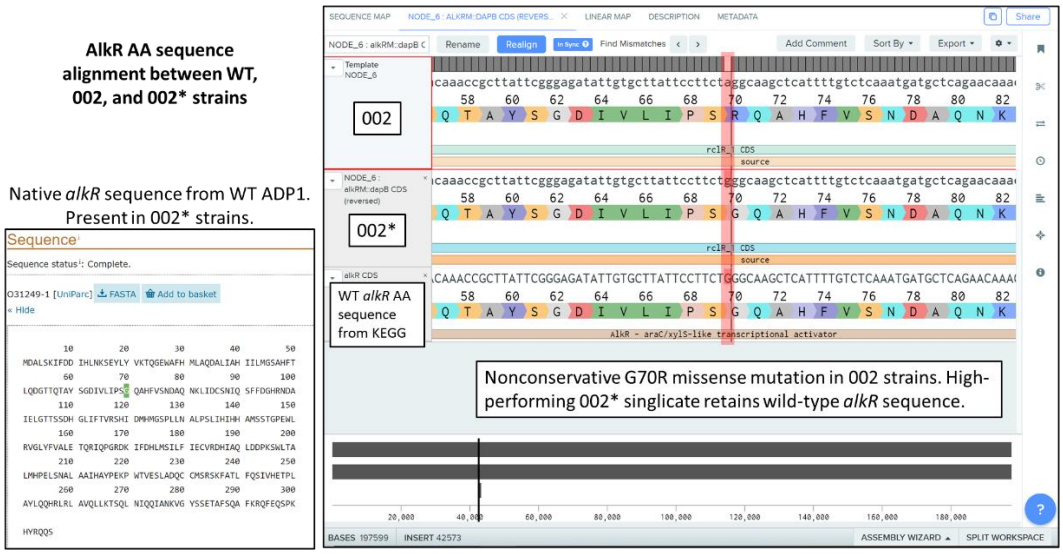


Figure S.16. Nonconservative p.G70R point mutation (g.208 G>A) in *alkR* CDS observed in low-performing 002 strains. The high-performing singlicate strain (002*) maintains the wild-type *alkR* sequence.

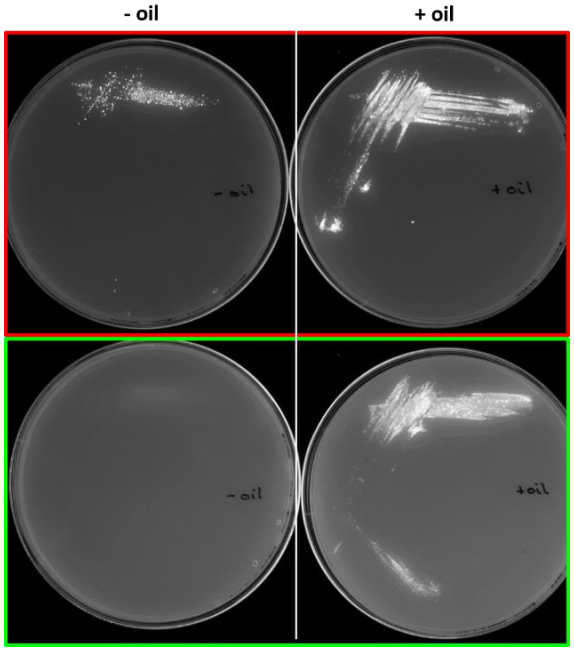


Figure S.17. Qualitative illustration of growth response to mineral oil in 002 strains with varying culture condition lineages. These plates suggest that differences in culture conditions of the primary inoculum may affect leakiness of the Artevo circuit. All plates are MMS agar, with or without 0.1% mineral oil (v/v). Upper plates (red border) are streaked with 002 strains from a glycerol stock originally grown in MMS without mineral oil or supplemental lysine; growth under those conditions would rely on leaky expression of *dapB* within the Artevo circuit. Lower plates (green border) are similarly streaked with 002 strains from a glycerol stock that have originally been grown in MMS + 0.01% (v/v) mineral oil; leakiness in this case appears to be reduced in subsequent culture, with reduced growth and fluorescence observable in the absence of activating mineral oil.

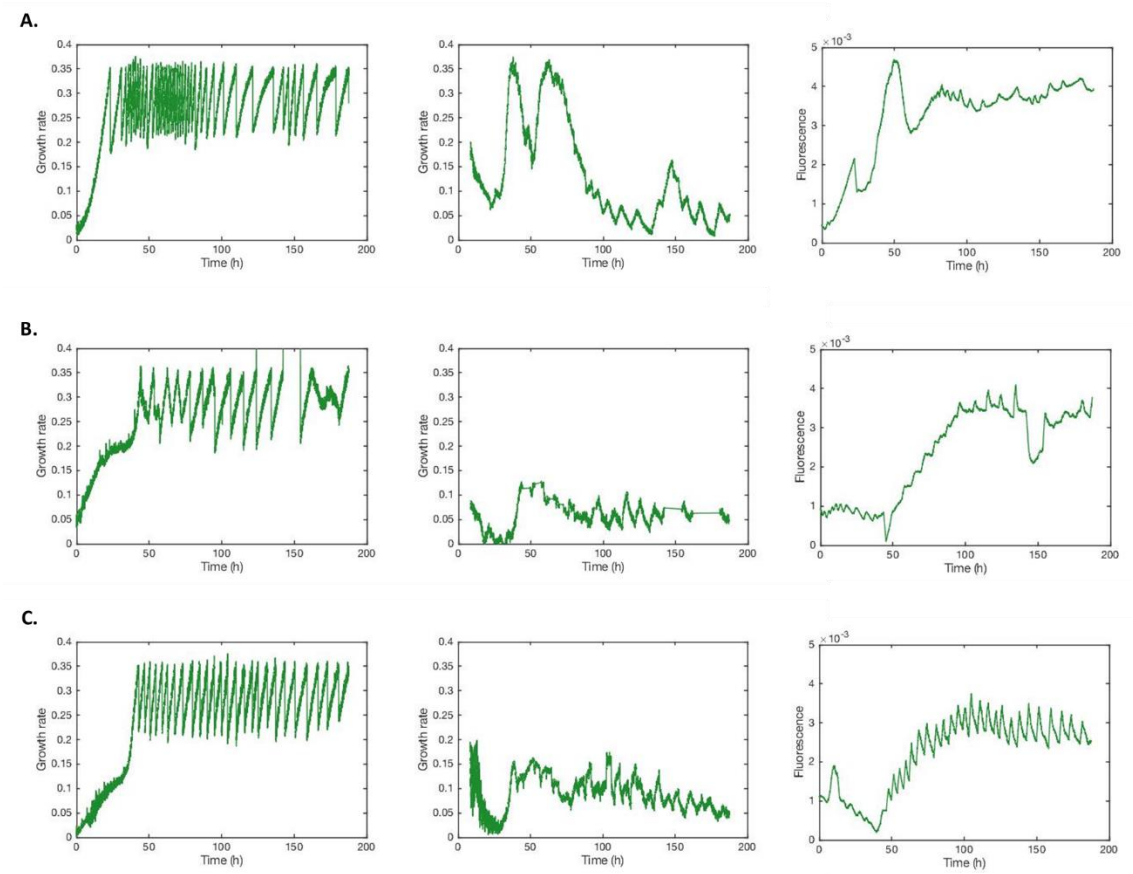


Figure S.18. Growth of *A. baylyi* ⊿*dapB* ⊿*alkM*::*dapB gfp* (002* strain) in Chi.BIO chemostat, in high oil (0.1% v/v), low oil (0.001% v/v), and no oil conditions. Sawtooth pattern is characteristic of Chi.BIO operation; see Steel *et al.* 2020 for detailed information about these reactor units[250].

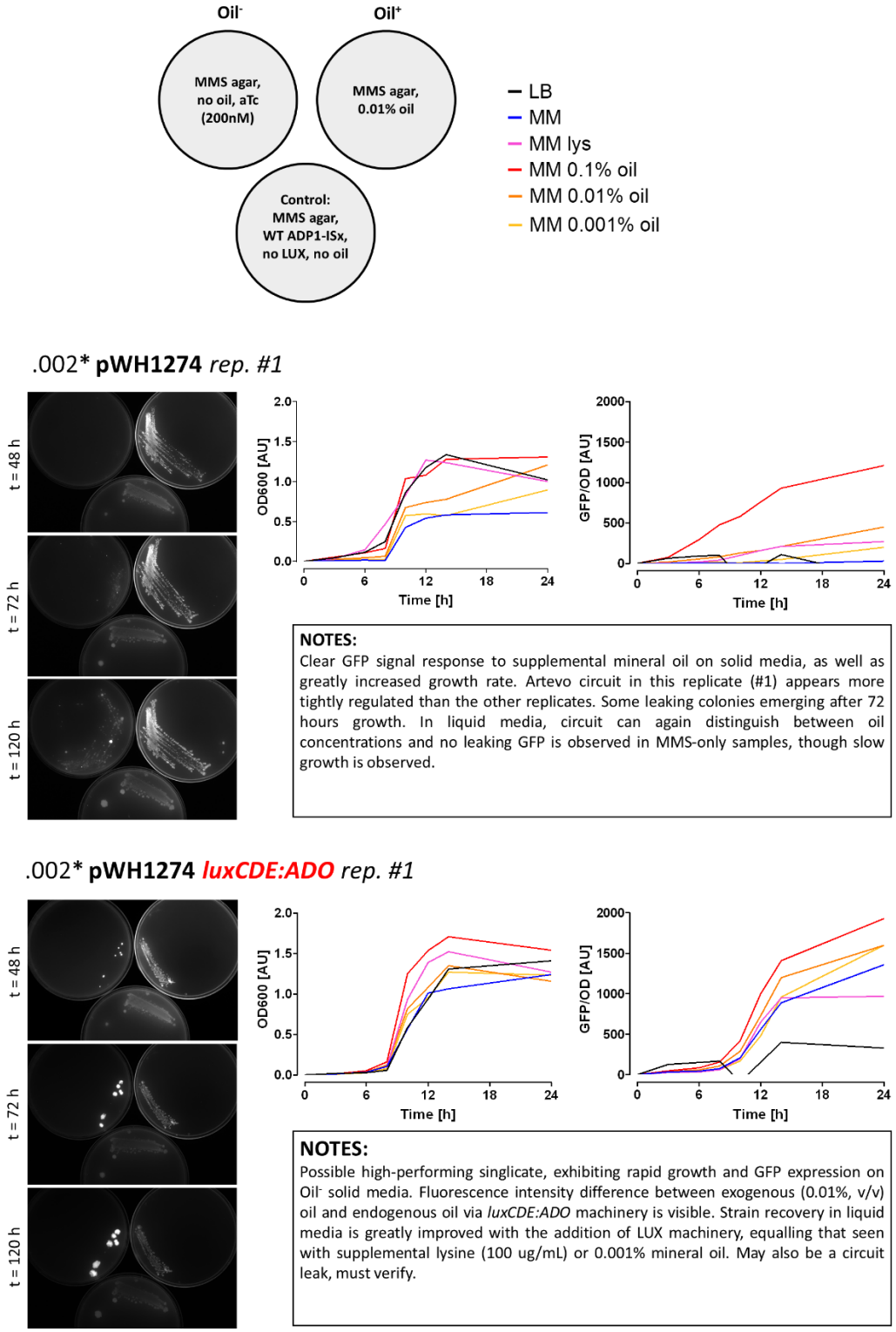


Figure S.19. Subset of results and summary of approach to investigating alkane production from pWH1274 *luxCDE:ADO* in 002* strains. MMS plates (layout at top of figure., alongside liquid culture media conditions) were streaked from a bank of glycerol stocks of replicates (derived from distinct transformation events) of ISxJK002 pWH1274 (empty vector control, Tet^R marker) and ISxJK002* pWH1274 *luxCDE:ADO* strains, to find high-performing singlicates of the latter. These plates were imaged as previously described to observe growth and fluorescence on solid media. The same replicates were cultured in 50mL Falcon tubes and manually sampled to monitor growth and fluorescence in liquid MMS media.

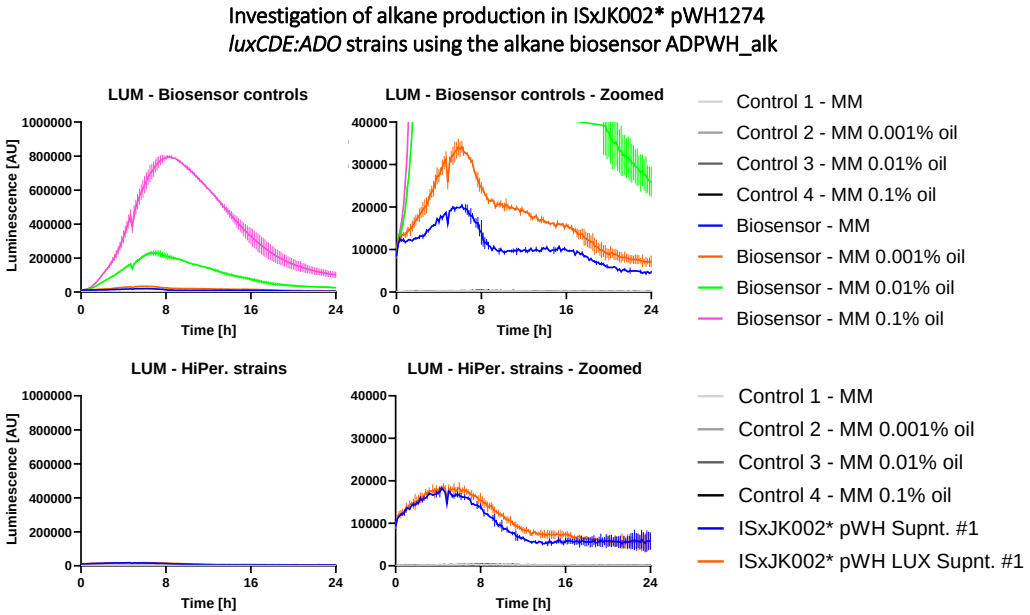


Figure S.20. Analysis of supernatant from ISxJK002* pWH1274 *luxCDE:ADO* cultures using the ADPWH_alk alkane biosensor strain. ISxJK002* pWH1274 *luxCDE:ADO* strains do not appear to be producing detectable levels of alkane, or are retaining produced alkanes in the cytoplasm. Similar results have been obtained when the ADPWH_alk biosensor is applied to lysate of these same cultures, obtained both by heating and by sonication, though this does not preclude the first theory; that alkane yields from these strains is very low. Graphs show specific luminescence over time across two scales, for control conditions (with supplemental mineral oil) and for a high-performing replicate (HiPer.) of ISxJK002* pWH1274 *luxCDE:ADO*.

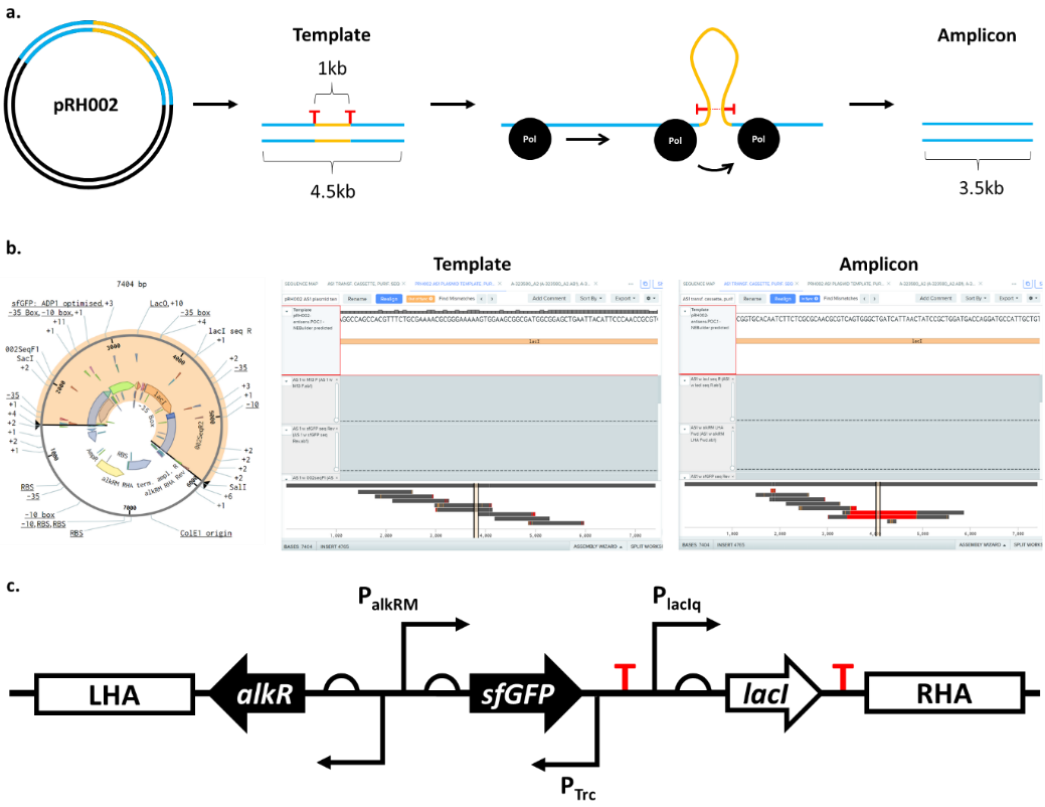


Figure S.21. (a) Illustration of suspected 'looping out' mechanism of gene excision during PCR amplification. In this case, it is suspected that homology between two similar but non-identical terminator sequences separated by the 1kb lacI gene caused a secondary loop to form in ssDNA during PCR. Highly-processive Q5 polymerase appears to have skipped across the secondary structure, reassociating with the template and continuing to produce an intact amplicon missing lacI. (b) Sequencing data confirming the presence of *lacI* in the template plasmid and the absence of *lacI* in the amplicon. (c) Design of the antisense control system transformation cassette highlighting the terminators responsible for causing 'looping out' of *lacI*. This phenomenon was not observable again when the terminator downstream of *lacI* was removed during subsequent cloning attempts.

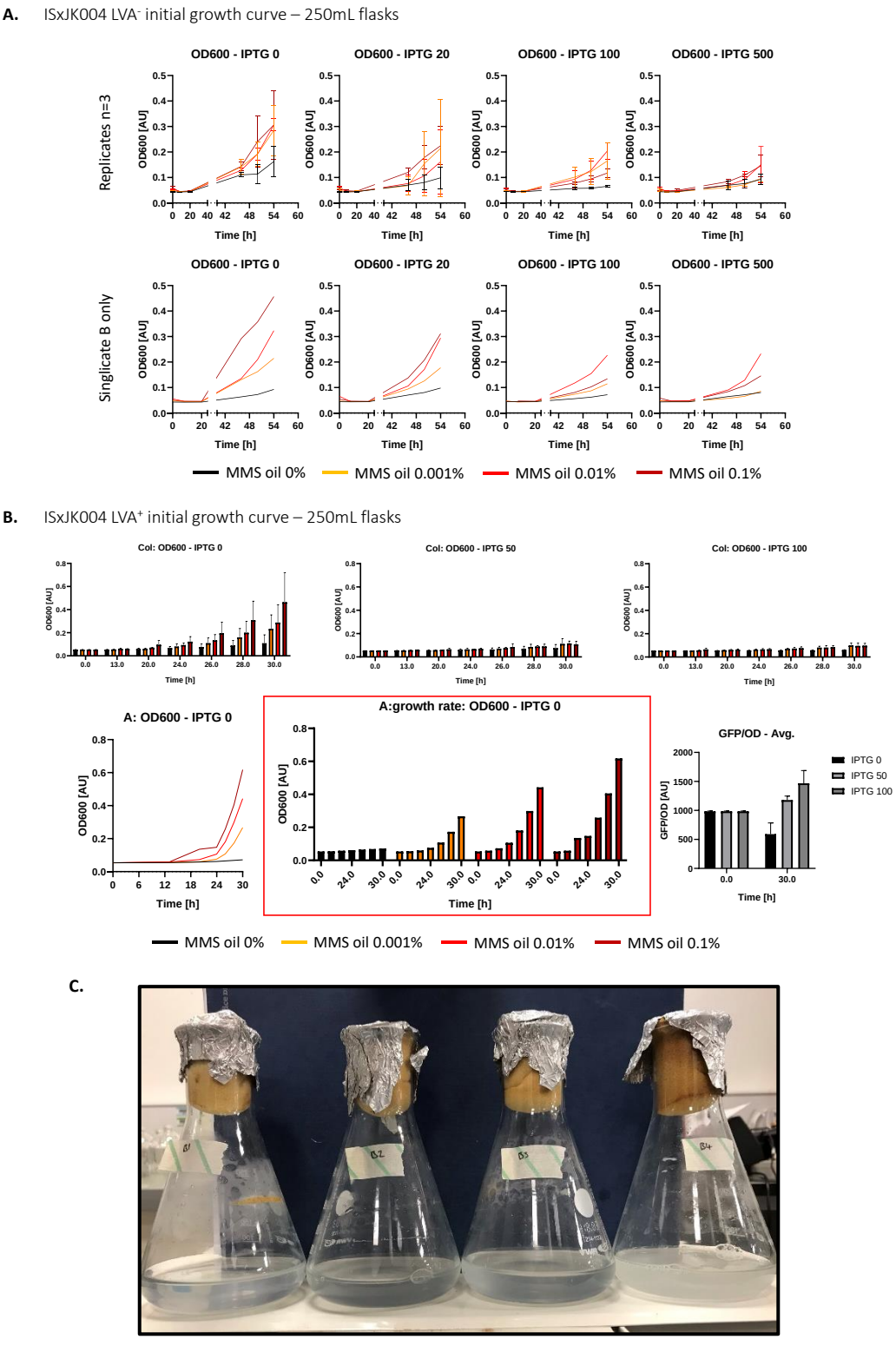


Figure S.22. Early comparison between ISxJK004 LVA+ and LVA- phenotypes. Part (a) shows LVA- strains in varying concentrations of mineral oil, with growth limited by induction of an antisense Trc promoter opposing *dapB*. Part (b) shows the same for LVA+ strains, presented so as to highlight the growth rate difference observed for strains grown in each mineral oil condition. The 'LVA' ssrA degradation tag on *dapB* appears to slightly improve dynamics of the Artevo system in response to alkanes, as predicted in Chapter 4. Part (c) shows a qualitative illustration of observed growth rate difference in culture flask, with higher culture densities being observable in proportion to exogenous alkane concentration.

For Chapter 5

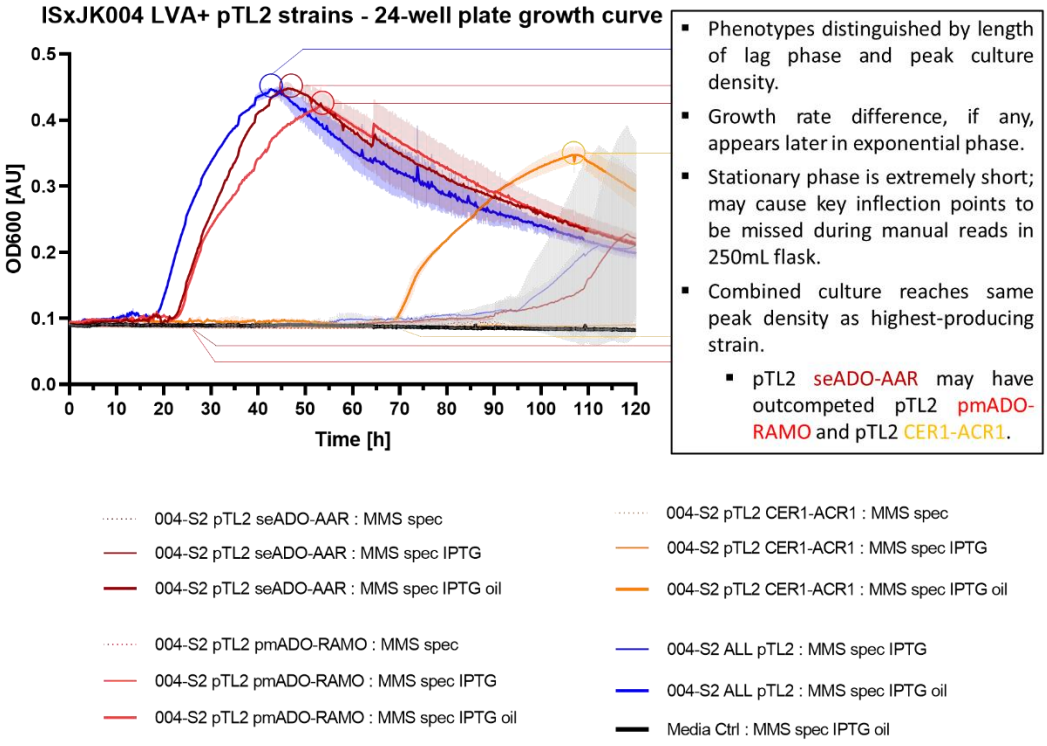


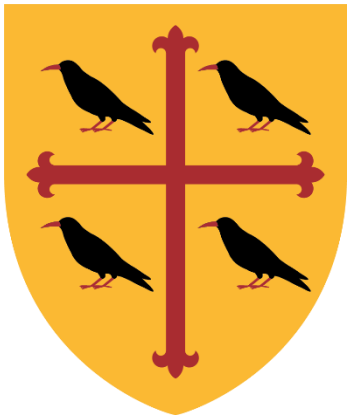
Figure S.23. Initial growth rate investigation of ISxJK004 LVA^{+} pTL2 strains, in 24-well plate. Error shown (shaded areas) is one standard deviation (n = 2).

"He had been eight years upon a project for extracting sunbeams out of cucumbers, which were to be put into vials hermetically sealed, and let out to warm the air in raw, inclement summers."

Jonathan Swift

Gulliver's Travels. 1726.

END





Joseph Kennerley

St Edmund Hall

University of Oxford

2022