Metabolic Engineering for Butanol Yield Enhancement in *Clostridium acetobutylicum*

Mr. Ryan J. Hope, BSc (Hons), MSc

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

March 2021

Dedicated to Caitlin.

Abstract

Clostridium acetobutylicum is the model solventogenic saccharolytic *Clostridium* spp. representing a group of bacteria which exclusively produce acetone and n-butanol along with the common solvent, ethanol; known as the ABE pathway. There is broad utility for n-butanol, particularly as a transport fuel but also as an industrial solvent and as a platform chemical. Hydrogen is also a major product of this organism by way of reduction of protons via ferredoxin coupled hydrogenase activity, where electron flux to this product is mediated by the oxidation of organic metabolic intermediates by the enzymes pyruvate ferredoxin oxidoreductase (PFOR) and the electron bifurcating activity of butyryl-CoA dehydrogenase (BCD). The role of BCD was explored utilising homologous recombination in-frame deletion methods, however, the apparent essentiality of the gene resulted in maintenance of the vector and the target gene in the genome, likely as a result of a random vector integration event. Replacing BCD with trans-2-enoyl-CoA reductase (TER) presents a metabolic engineering opportunity by subversion of electron flux to ferredoxin, and ultimately hydrogen gas production, furthermore, it allows us to investigate the importance of the bifurcating role of BCD. Hypothetically, successful replacement of BCD with TER should result in an alcohologenic fermentation, as the cells attempt to maintain redox cofactor homeostasis. The expression of TER resulted in a significant improvement in solvent productivity. Nevertheless, the electron bifurcating activity of BCD appears to be an essential metabolic function for *C. acetobutylicum*, and DNA-seq data from a mutant strain obtained from a third party suggests that this is due to the role of hydrogenase in maintaining the proton motive force - in which case a complementary mutation interrupting the function of the proton powered flagella will ultimately facilitate the replacement of BCD with TER.

A prototypic lactose inducible orthogonal expression system was applied in order to maximise the flux to butanol in the TER expressing parent strain. A control study using a strain expressing the lactose binding transcriptional activator and the TcdR sigma factor produced an altered phenotype where enhanced solvent production was observed and a computational approach was used to try to identify

TcdR promotor binding sites in the *C. acetobutylicum* genome offering some insight as to the cause of the adjusted phenotype and a new regulator of solventogenesis is proposed.

Declaration

Unless otherwise acknowledged, the work presented in this thesis is my own. No part has been submitted for another degree in the University of Nottingham or any other institute of learning.

Ryan J. Hope

March 2021

Chapter 1: Introduction

1.1 The price of oil

Since the mid 19th century, the supply of energy fuelling the global economy has been reliant on geographically heterogenous, finite supplies of fossil fuels, consisting of oil, natural gas, and coal. The supply and price of oil, in particular, has been a major influence on the macroeconomic growth and stability of developed nations. The phenomenon of oil price shocks (a drastic increase in the price of oil) has been a precipitating factor of economic recessions in Western industrialised nations, which has been empirically suggested by Hamilton (1983; 1996; 2003). It is largely as a result of the evolution of modern monetary policy: a shift towards flexible labour markets and the reduced contribution from oil have these effects been mitigated (Blanchard & Riggi, 2013). This is an important consideration for those interested in the benefits of bio-renewables.

The requirement to control oil price in order to ensure a healthy economy in the West has also driven an expensive and fraught interventionist approach to foreign policy. One of the most bold and historically important examples of this includes the removal of the first democratically elected Prime Minister of Iran - Muhammad Musaddiq. The Iranian Prime Minister was removed from power in a 1953 coup d'etat ordered by President Dwight D. Eisenhower and Prime Minister Winston Churchill, after insisting that the largely British owned Anglo-Iranian oil company, should be transferred over to the Iranian state. Following the coup, the Anglo-Iranian oil company perhaps more fittingly changed their name to British Petroleum (BP) and power was returned to the monarchic Shah. The Shah's autocratic reign was propped up with the help of the American government until he was finally overthrown in the bloody 1979 Iranian Revolution, when the state governance transitioned to a Shiite Islamic Republic (Kinzer, 2003). Conflict with this important OPEC member continues to this day.

Another pertinent example of interventionism began with the Iraqi invasion of Kuwait in 1991 which was swiftly met by a US lead western coalition response known as Operation Desert Storm, or the Gulf War (figure 1). Leaders of western nations were nervous of Saddam's ambitions, also having reason to believe - given that he had forces stationed 250 miles from Saudi oil fields - that he further intended to invade Saudi Arabia. Had these two invasions been successful, Saddam would have controlled 44% of the worlds oil supply, allowing him to singlehandedly control the worlds oil prices (Yetiv, 2008). Under these circumstances, even the most determined pacifist should have difficulty with the moral justification of abstaining from the invasion of Iraq, and this therefore represents a perfect example of how the dependancy on foreign oil is forcing western liberal democracies to become embroiled in an ever perpetuating middle eastern conflict.

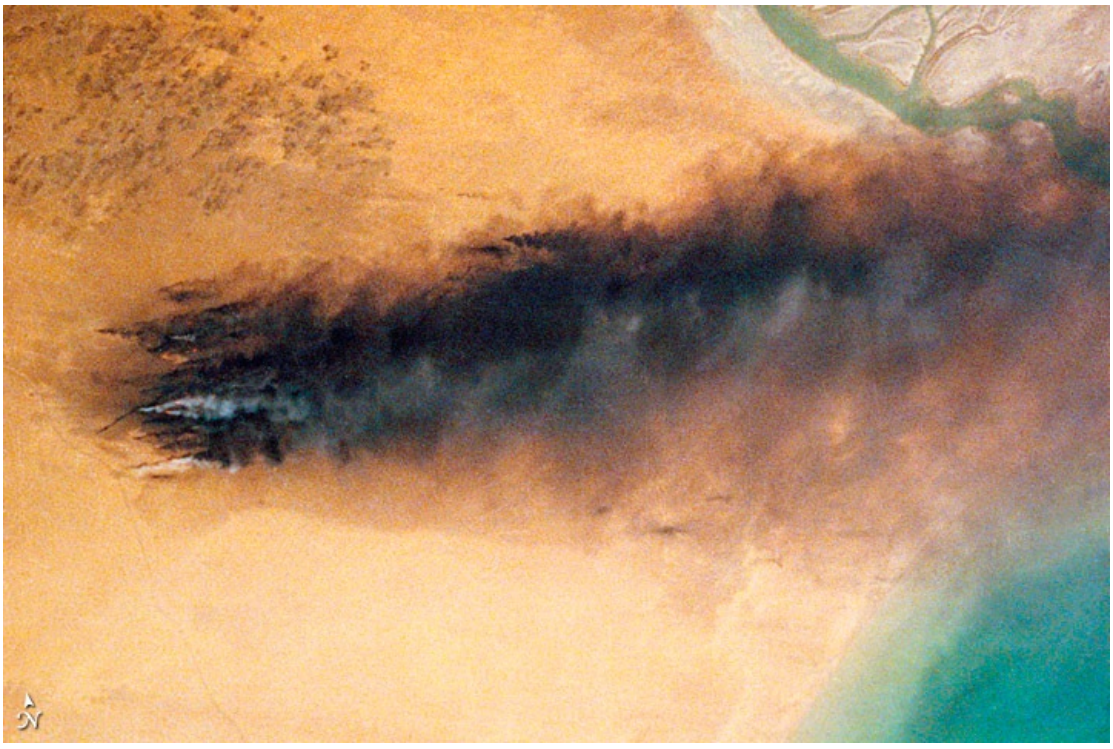


Figure 1: The image captured during the 1991 Gulf War shows Kuwaiti oil pumps set alight by Saddam Hussain's forces in retaliation to the American-led Operation Desert Storm. Taken from NASA's Earth Observatory image archive (NASA, 1991).

At 302.81 billion barrels, Venezuela has the greatest share of proven oil reserves (OPEC, 2019) and a more diverse range of extra-regional actors seek to gain influence in oil rich countries in the modern era. Vladimir Putin has made no effort to be discreet about establishing strong relations with the Venezuelan government, which includes selling and donating arms, using Russia's seat on the security council, establishing new banks and an oil-backed crypto-currency named "Petro" help to protect Venezuela from US and EU sanctions, as well as military training and high profile state to state visits. Russia has also made a great effort to prop up the floundering Venezuelan oil industry by making advanced purchases of oil before it is pumped and providing bail outs to the state owned Petróleos de Venezuela, S.A. (Farah & Babineau, 2019).

China also seeks to gain a slice of the Venezuelan reserves, and have continued lending to the struggling Maduro government, with an agreement to receive re-payment in oil, however, the western backed opposition to Maduro has pledged to tear up any such agreements made with China, which could certainly result in tensions between the West and China in the future. Historically, military-military training programmes have been used as a strategy gain and maintain international influence by the US for years, although, this has been replicated by China and Russia, and in 2015, China trained more Latin American military officers than the United States, the numbers have grown year on year ever since (Farah & Babineau, 2019) .

As the governments of West seek to re-establish themselves in this now crowded landscape, the struggle for influence in Venezuela, or indeed any oil rich nation, is certain to put strain on future relations between these nuclear armed groups.

1.2 Peak oil and the response to restricted supply

The potential for conflict to arise over oil ought to be expected to increase year on year after we reach the point known as “peak oil” (figure 2) anticipated in 2020 (GEA, 2012) when global production output begins to fall the global demand for oil will increasingly go unmet. As the economy and the global population grows as does the demand for energy, however, in the absence of renewable sources, inadequate supply will limit economic growth, which is tightly associated with oil supply/price (Hamilton, 1983; 1996; 2003).

The 1973 oil crisis was an event that saw the price of oil quadruple in a financial quarter year and was the result of an OPEC agreed oil embargo. The following year a global recession occurred which has been shown to be a direct consequence of the embargo (Balassa, 1985). The impact of the price shock was felt the world over and resulted in the US stock market halving in value (Alpanda & Peralta-Alva, 2010). The impact on the economy is usually felt through the effect of a reduced supply of petroleum and industrial petrochemicals, and through reduced demand in many secondary goods markets, with automobile sales being an important example (Lee & Ni, 2002).

Governments responded to these economic shocks broadly by adopting diplomatic and military interventionist foreign policy and through modifications to domestic economic policy. The immediate response by some governments of industrialised nations was to place a price cap on domestically produced petroleum in an attempt to mitigate the price effects of the restricted supply of foreign imported oil. This resulted in shortages at the refuelling stations, further impacting the economy (Baumeister & Kilian, 2016).

Countries with a powerful military such as the United States adapted to the two oil crises of the 1970s by developing relationships with OPEC members in the

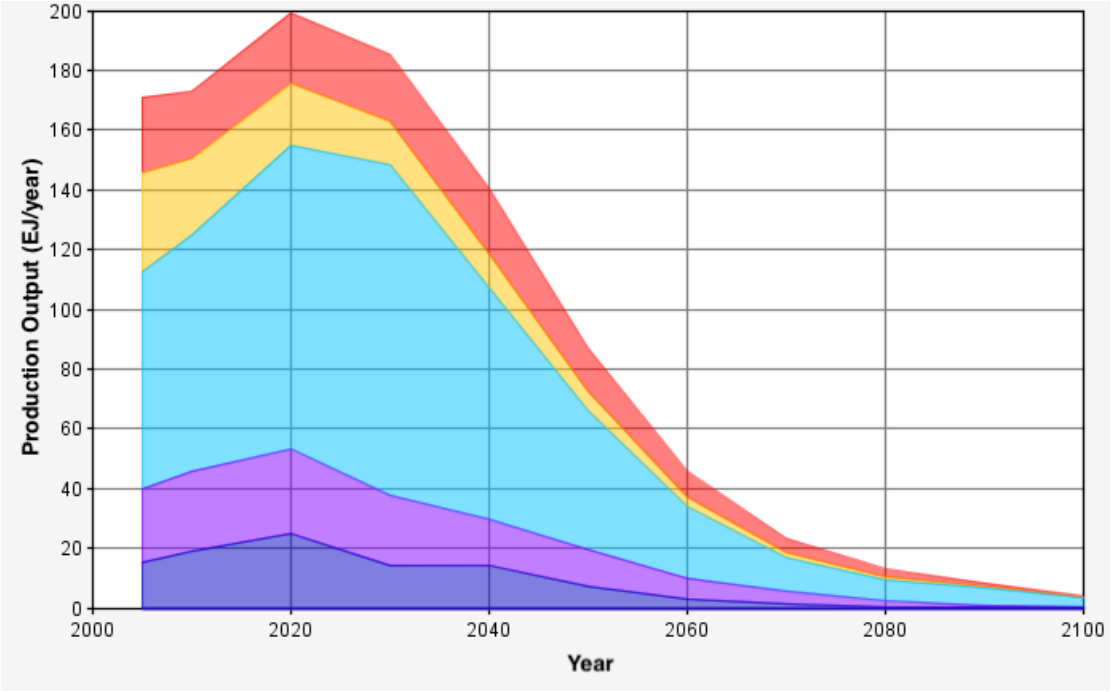


Figure 2: Projected production outputs from various oil producing groups. Peak oil is expected to occur in the year 2020: ▪, Reforming economies (Central and Eastern Europe and the Former Soviet Union undergoing economic reform); ▪, OECD90 (Countries that were OECD members in 1990); ▪, Middle East and Africa; ▪, Latin America and the Caribbean; ▪, Asia. (Graph generated using the GEA projection database, GEA, 2012).

Middle-East governed by brutal and vulnerable autocratic regimes, providing support through arms sales, donations, and direct military support ensuring a favourable response in terms of controlling oil price and maintaining output (Jones, 2012).

Other countries responded domestically and have adapted by building resistance into their economy by various means, such as the previously mentioned project of movement towards flexible labour markets (Blanchard & Riggi, 2013).

Some responses have been more progressive. Denmark serves as a primary example of an early adopter of alternative energy systems, and since 1973 they have implemented expansive combined heat and power conservation programmes. This has been so effective that primary fuel consumption remained stable for a period of over 30 years, despite a 70% increase in GDP. Moreover, this included a shift to a 14% share of energy provided from renewable sources (Lund, 2007).

In 1975 Brazil responded to the oil crisis with the launch of the “Proalcool” program which aimed to supplement and perhaps eventually replace petroleum with ethanol (Lopes *et al.,* 2016). Many other countries launched a similar program, however, they were dropped after a dramatic fall in oil prices in the 1980s. Brazil persisted, nevertheless, and at this point have over 40 years of experience of developing production strains, industrial bioprocesses, and infrastructure and as such are much better placed to deal with the present problem of peak oil (Hofstrand, 2015), temporarily achieving energy independence in 2006 by supplementing its domestic oil supply with 20% ethanol (Potter, 2008).

1.3 Greenhouse gas emissions and climate change

The Intergovernmental Panel on Climate Change (IPCC) was jointly established by the World Meteorological Organisation (WMO) and The United Nations Environment Programme (UNEP) in 1988 to provide governments with authoritative and objective technical analysis, subsequently their assessment reports have become standard works of reference for developing climate change mitigation policies.

Reporting on carbon and other biogeochemical cycles, they state that carbon dioxide concentrations have increased by 40% since pre-industrial times, primarily from fossil fuel use but also as a result of changes in net land use related emissions (figure 3). The effect of increased atmospheric carbon dioxide is strongly inferable to the increase in global average temperature, and when comparing the increase to paleoclimate reconstructive models, the period between 1983 and 2012 was likely to be the warmest 30-year period in the last 1400 years. However, ocean warming accounts for 90% of the energy stored in the climate system, with the upper 75m of the ocean increasing by 0.11^{o}C, this may sound low but when calculating the total energy increase in the upper ocean system - the top 700m (60% of the aforementioned 90%) this represents massive increase of 10^{22} J (IPCC, 2014).

The ecological impact of elevated CO_2 together with ocean deoxygenation is an important issue which is often neglected in the public conversation about climate change. Approximately 30% of the increased anthropogenic carbon dioxide has been absorbed by the oceans resulting in their acidification (IPCC, 2013). The report from the Third Symposium on the Ocean in a High CO_2 World outlined the fact that CO_2 induced acidification together with other stressors such as elevated temperature, deoxygenation, pollution and overexploitation would lead to large “dead zones” in the worlds oceans (IGBP, IOC, SCOR, 2013). This will have an obvious impact on

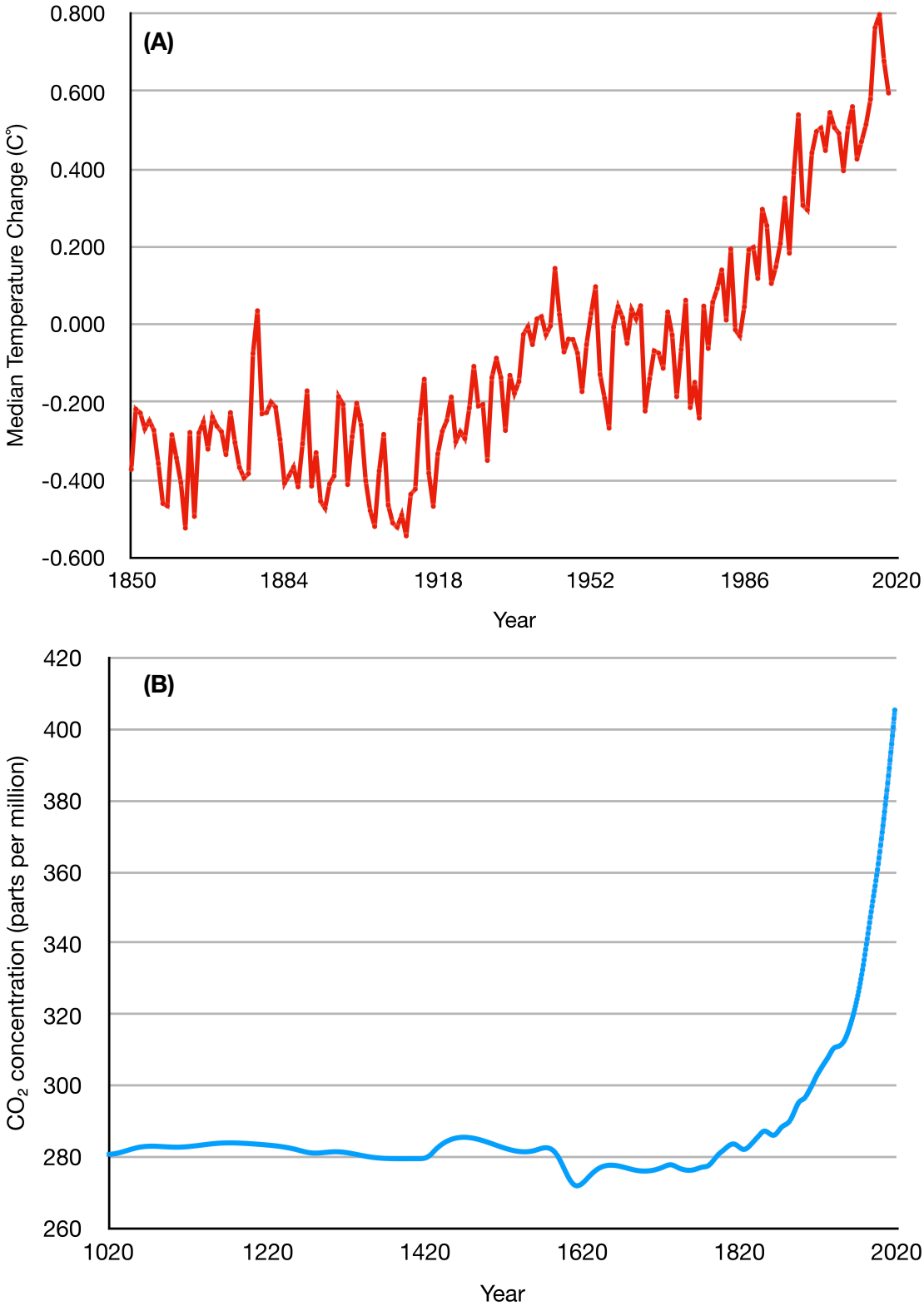


Figure 3: The relationship between anthropogenic emissions and global warming is carefully outlined by the IPCC (2014). The correlation between atmospheric CO2 and median temperature change is visible in the two graphs above. Graph A shows global median temperature change since 1850. Graph B shows atmospheric CO_2 (Ritchie & Roser, 2019).

natural fish stocks, as well as more complicated ecological consequences. Further to this, elevated CO_2 together with river runoff mediated eutrophication also threatens toxification of farmed fish from harmful neurotoxin producing cyanobacteria (Hallegraeff, 2010).

Looking to the cryosphere, the influence of warming has shown further worrying trends, with glaciers (excluding those on the periphery of ice sheets) losing mass at a rate of 226 Gt $yr^{-1}$and the Antarctic ice sheet losing ice at an increasing rate, measuring 30 Gt yr^{-1} between 1992-2002 and rising to 147 Gt yr^{-1} between 2002-2011. The most significant losses in the polar regions have come from the Arctic, where Arctic sea ice losses have been in the range of 9.4%-13.6%, measured per decade between 1979 and 2012. The resultant geographical impact of polar melting has been on sea levels, where since the mid-19th century, the mean rate of sea level rise has been greater than that of the previous two millennia (IPCC, 2014).

The predicted impact of these factors span from the pernicious to the devastating. While demand for food increases as the population continues to expand, additional pressure will come from climate related changes to availability of arable land and also from extreme weather shocks such as droughts and flooding having a major impact on crop yields (Wheeler, 2013). Changes in climate can also be expected to have certain ecological effects. How much consideration should given to whether or not it matters if certain species are unable to flourish under new climate conditions might be left to the moral philosophers. However, it is of no doubt that adjustments in host parasite interactions are the practical concern of everybody currently living in temperate climates with relatively benign ecosystems, who might soon have to get to grips with the idea that what once were considered “tropical diseases” must now be reclassified as “emerging diseases” (Altizer, 2013).

1.4 Industrial bioenergy and biotechnology for sustainable production of fuels and fine chemicals

Microbial species represent a vastly diverse array of biological systems which are able to sustain themselves through an extensive range of intricately regulated metabolic processes. Many of these processes produce metabolites which have been found to have some practical application and the organisms responsible have been implemented in large industrial fermentation processes producing fuels, industrial solvents, and pharmaceutical products.

Fermentation of sugars to ethanol and CO_2 carried out by *Saccharomyces cerevisiae* is probably the oldest and most well-known fermentation process which has been used for beer, wine, and bread making, dating as far back as the Ancient Egyptians in 1400 BC (El-Mansi *et al*., 2007). Industrial biotechnology later emerged in the early 20th century when ethanol from yeast fermentation had found a use as a biofuel used to blend into petroleum up to 25% by volume (Antoni *et al.,* 2007). Products produced by industrial fermentation began to diversify, a trend accelerated by disrupted supply chains in the First World War preventing conventional manufacture of industrial chemicals. One particularly innovating example involved glycerol manufacture, which was needed for explosives production. This was the result of a modified yeast fermentation with the addition of sodium bisulphate to the culture medium. This results in a disruption of redox homeostasis, diverting carbon flux from ethanol to glycerol production in order to oxidise the accumulated NADH (El-Mansi *et al.,* 2007).

During the 1940s the discovery of antibiotics presented another significant leap forward in biotechnology. *Penicillium spp.* were exploited for their antibiotic producing properties, and the subsequent research development in an effort to improved yields produced the first closed, stirred tank bioreactors; allowing the control of contamination, and the first laboratory evolution methods; involving

mutagenesis and screening, culminating in an eventual yield increase from 1g L^{-1} to 20 g L^{-1} of penicillin (El-Mansi *et al.,* 2007).

The production of acetone and butanol using saccharolytic solvent producing *Clostridium spp* was another pioneering industrial bioprocess which came to be known as the Weizmann process, in reference to one of the key scientists involved in the early development which took place in Manchester. The process was industrialised rapidly following the outbreak of the First World War in order to supply acetone for the manufacture of cordite (used as a smokeless gunpowder) and a manufacturing plant at Kings Lynn reached production yields averaging 0.9 tonnes of acetone per week from a maize based fermentation (Weizmann & Rosenfeld, 1937; Jones & Woods, 1986).

The Weizmann process was also exploited for butanol production needed in the production of a nitrocellulose-lacquer used for finishing car bodies by E. I. du Pont de Nemours & Co but the process eventually became uneconomical by the 1960s for both acetone and butanol as a result of the abundance of organic solvents derived from cheaper petrochemical sources (Antoni *et al.,* 2007). Industrial ethanol production from yeast carbohydrate fermentations for use as a fuel in combustion engines persisted, however, largely as a mitigating response to oil price shocks, the primary example being Brazil's Proalcool programme, as discussed previously.

Interest in biofuels has undergone a renaissance period recently with many research funding agencies and start-up biotech companies seeking to improve and commercialise biofuel technology for an increasingly dynamic energy market incentivised by government commitments to limit fossil fuel emissions, and volatility in the oil markets.

This has brought the Weizmann process back into focus, however, by now butanol has been identified as a superior biofuel. Butanol has several properties which make it preferable to ethanol: it has higher energy density, low hygroscopicity, greater compatibility with existing petrol combustion engines, and low corrosiveness (Dürre, 2011). In 2003, a joint venture between BP and DuPont known as Butamax (https://www.butamax.com) has sought to scale up and commercialise isobutanol production (Green, 2011), as have the company Gevo (NASDAQ:GEVO)

(established 2005) who operate a production plant in Luverne, Minnesota producing 6.8 million litres of isobutanol per year and have secured contracts to supply aviation fuel and isobutanol for terrestrial vehicles (https://finance.yahoo.com/gevo). Green Biologics (established 2003) have also achieved similar success and have commercialised a modern *Clostridium spp.* based process in Little Falls, Minnesota producing butanol and acetone for the industrial solvents market (https://greenbiologics.com).

Although unpredictability in the trading price of crude oil can place significant strain on commercial start up ventures, impeding their growth into the fuels market; during the course of this period of renewed interest, sustained academic advancements in the fields of genetic and metabolic engineering have been made. This has seen thorough extension in the utility of both the classical and an additional new range of industrially significant microbes with the development of many novel robust bacterial industrial chassis (Calero & Nikel, 2019).

1.5 *Clostridium acetobutylicum* as an industrial chemical platform host and as a model organism

1.5.1 *Clostridium acetobutylicum* physiology and metabolism

The genus *Clostridium* comprises a broadly diverse range of species which are Gram-positive, anaerobic, spore forming bacteria (Cato & Stackebrandt, 1989). A number of the mesophilic, saccharolytic, butyric acid producing organisms are also capable of producing butanol (Jones & Woods, 1986). Classically, *Clostridium acetobutylicum*, *Clostridium beijerinckii, Clostridium saccharobutylicum*, and *Clostridium saccharoperbutylacetonicum* are the species most well known for production of butanol. The species *C. acetobutylicum* ATCC 824 is the best understood in the literature and is generally regarded as the model solvent producing *Clostridium* (Dürre, 2011; Green 2011). Furthermore, *Clostridium pasteurianum* has also long been known for its butanol producing qualities and is also capable of fermenting glycerol as its sole carbon source (Johnson *et al.,* 2016). A recent comparative genome analysis has expanded this collection of solvent producing organisms to include *Clostridium puniceum, Clostridium roseum*, *Clostridium felsineum,* and a currently unnamed *Clostridium* sp. (Poehlein *et al.,* 2017).

C. acetobutylicum has a genome comprising a 3.941 Mb chromosome and a 192 Kb megaplasmid with a GC content measuring 30.7 % (Nolling *et al.,* 2001). A heterotroph, it is typically grown on sugar or starch based complex feedstocks in the plant (Green, 2011) but specifically it is known to grow on the hexoses glucose, mannose, galactose, and fructose, the pentoses arabinose and xylose, as well as the disaccharides sucrose, lactose, maltose, and cellobiose (Servinsky *et al.,* 2010).

In batch fermentations *C. acetobutylicum* grows in a typically biphasic manner. During exponential growth, sugars are converted to acetic and butyric acids, and is known as the acidogenic phase. This is followed by the solventogenic phase,

characterised by a partial or complete re-assimilation of acids and the production of the neutral solvents acetone, butanol, and ethanol, the so called the ABE pathway. Throughout the fermentation, hydrogen and carbon dioxide gas are also produced (Millat & Winzer, 2017), however, hydrogen production during solventogenesis is drastically decreased (Jones & Woods, 1986).

In general, sugars are transported across the membrane via a Phosphoenolpyruvate (PEP)-dependent phosphotransferase system (PTS). PTS systems are unlike other concentrative transport systems in that they catalyse the phosphorylation of the sugar substrate as they enter the cell. Clostridia, as is common with other anaerobes, are particularly dependant on PTS systems for the transport of hexose sugars - pentoses are likely transported by proton symporters (Mitchell, 2016)

Following entry via the PTS system, phosphorylated glucose proceeds through glycolysis. Glycolysis generates 2 moles of ATP, 2 moles of NADH, and 2 moles of pyruvate per mole of glucose. Pyruvate enters the ABE pathway via the activity of pyruvate ferredoxin oxidoreductase (PFOR) and for every 2 moles of pyruvate a mole of fully reduced ferredoxin is generated along with 2 moles of acetyl-CoA and 2 moles of carbon dioxide (Uyeda & Rabinowitz, 1971). During acidogenesis, acetyl-CoA and butyryl-CoA is converted to acetic acid or butyric acid via the activity of a separate pair of phosphotransferases and kinases resulting in the generation of 1 mole of ATP per mole of acid by substrate level phosphorylation (SLP) (Hartmanis & Gatenbeck, 1984). Acetic acid production generates 2 moles of ATP per mole of glucose, however, this pathway does not consume NADH and is therefore unbalanced with glycolysis. To compensate for this, some NAD^+ is regenerated by reduction of H^+ to hydrogen gas utilising reduced ferredoxin. Butyrate production, on the other hand, is balanced with glycolysis, requiring 2 moles of NADH, however, a third molecule of NADH is required for hydrogen production via ferredoxin and is supplied, in part, by the electron bifurcating activity of butyryl-CoA dehydrogenase (Li *et al.,* 2008). Butyric acid synthesis requires 2 moles of acetyl-CoA, the implication being that SLP via butyric acid synthesis is less efficient route for ATP production than acetic acid formation.

BCD and PFOR serve particularly interesting roles in the pathway as both are responsible for the reduction of the low potential election acceptor ferredoxin. PFOR is

an ancient enzyme that existed prior to the divergence of archaea and eukaryotes, its significance and mechanism is well understood as the oxidation of pyruvate to acetyl-CoA has been a focal interest in biochemistry research throughout the 20th century. PFOR is widely distributed among bacteria, and is also present in some anaerobic protozoa. In *Clostridium* spp PFOR directs organic carbon into the phosphoroclastic ABE pathway, converting pyruvate to acetyl-CoA and generating low potential electrons ($E^{o\prime}$ -540 mV) for the reduction of ferredoxin ($E^{o\prime}$ -420 mV) in the process. PFOR requires the incorporation of thiamine pyrophosphate as a cofactor which undergoes a deprotonation of the thiazolium ring via the glutamic acid in position 64, generating a highly nucleophilic ylide intermediate making it a powerful electron donor capable of reducing ferredoxin via three internal [4Fe-4S] clusters (Ragsdale, 2003).

By comparison, the flavin based electron bifurcation (FBEB) mechanism by which BCD is able to reduce ferredoxin has only been recently elucidated and is regarded as a newly discovered mechanism of energy conservation, first being described in 2008 (Herrmann *et al*., 2008; Li *et al*., 2008). Although, quinone based electron bifurcation (QBEB) was known to occur in the mitochondrial q-cycle it was once thought peculiar, however, now it seems that energy conservation by electron bifurcation is likely a general and ancient principle for reducing very low potential redox components (Nitschke & Russell, 2012). FBEB depends on a two electron donor - NADH in the case of BCD - first donating both electrons to a flavin adenine dinucleotide (FAD) producing a reduction potential low enough to transfer a single electron to ferredoxin, the second electron is shuttled through a second FAD to a third where it is stored until a second bifurcation results in a fully reduced ferredoxin and the reduction of the high potential double bond of crotonyl-CoA ($E^{o\prime}$ -10 mV).

During acidogenesis, internal pH is maintained at pH 6 while external pH drops to 4.5 providing a ΔpH of 1.5 (Huang *et al.,* 1985), this is occurs in part through low potential hydrogenase reactions with electrons supplied from reduced ferredoxin - capable of reducing these low potential couples ($H_2/2H^+$ couple is E_0' = -420 mV) (Girbal *et al.,* 1994; Gottschalk 1986) and in part by F_1 F_0 ATPase H^+ extrusion (Jones & Woods, 1986) . When external pH reaches the pK_a of the external carboxylic acids, the

uncharged free acid forms begin to make a net movement back into the cell across the membrane where they re-enter the pathway via the activity of coenzyme-A transferase (CTFA/B), and the external pH is lowered (Kell *et al*., 1981; Jones & Woods, 1986).

This point marks the end of acidogenesis and the commencement of solvent production, the cells fall into stationary phase as the ATP lost from SLP becomes limiting to growth. The duel activity of CTFA/B results in one mole of acetyl-CoA or butyrl-CoA produced per mole of respective re-assimilated carboxylic acid, which is then converted to acetone as catalysed by acetoacetyl-CoA decarboxylase (ADC), also liberating 1 mole of CO_2 in the process (Andersch *et al.,* 1983). The intermediates butyryl-CoA and acetyl-CoA are then reduced to butanol and ethanol by a range of various aldehyde and alcohol dehydrogenases, namely, ADHE1, ADHE2, BDHA, BDHB, BDHC (see figure 4 for pathway diagram) (Yoo *et al.,* 2015). Since ethanol and butanol production is balanced with glycolysis, during solventogenesis reduced ferredoxin is also oxidised via the activity of ferredoxin NAD^+ reductase (FdNR) and ferredoxin $NADP^+$ reductase (FdNPR) (Girbal & Soucaille, 1994.)

The switch from acidogenesis to solventogenesis has been shown to be responsive to external pH. Seminal work on this physiological response showed that solventogenesis could be induced by adjustment of the external pH to 5.0, with optimum solvent production observed at pH 4.3 (Bahl *et al*., 1982). Millat *et al.,* (2013) demonstrates that cells in chemostat culture controlled at pH ≥5.3 produces cells displaying the acidogenic phenotype and pH ≤5.0 results in solvent producing phenotype and pH 5.1 produces an intermediary phenotype for the organism.

The solvent producing phenotype can also be induced at pH values normally associated with acidogenesis via the manipulation of the intracellular redox co-factor balance. Several approaches have achieved this, firstly Meyer *et al*., (1986) demonstrated that in a continuous culture controlled at pH 5.2 an acidogenic fermentation could be switched to alcohol production without acetone (alcohologenesis) through CO sparging, a result of the inhibition of hydrogenase activity. Later, Vasconcelos *et al*., (1993) used a combination of two carbon sources: glucose and the more reduced glycerol, to shift fermentations toward alcohologenesis

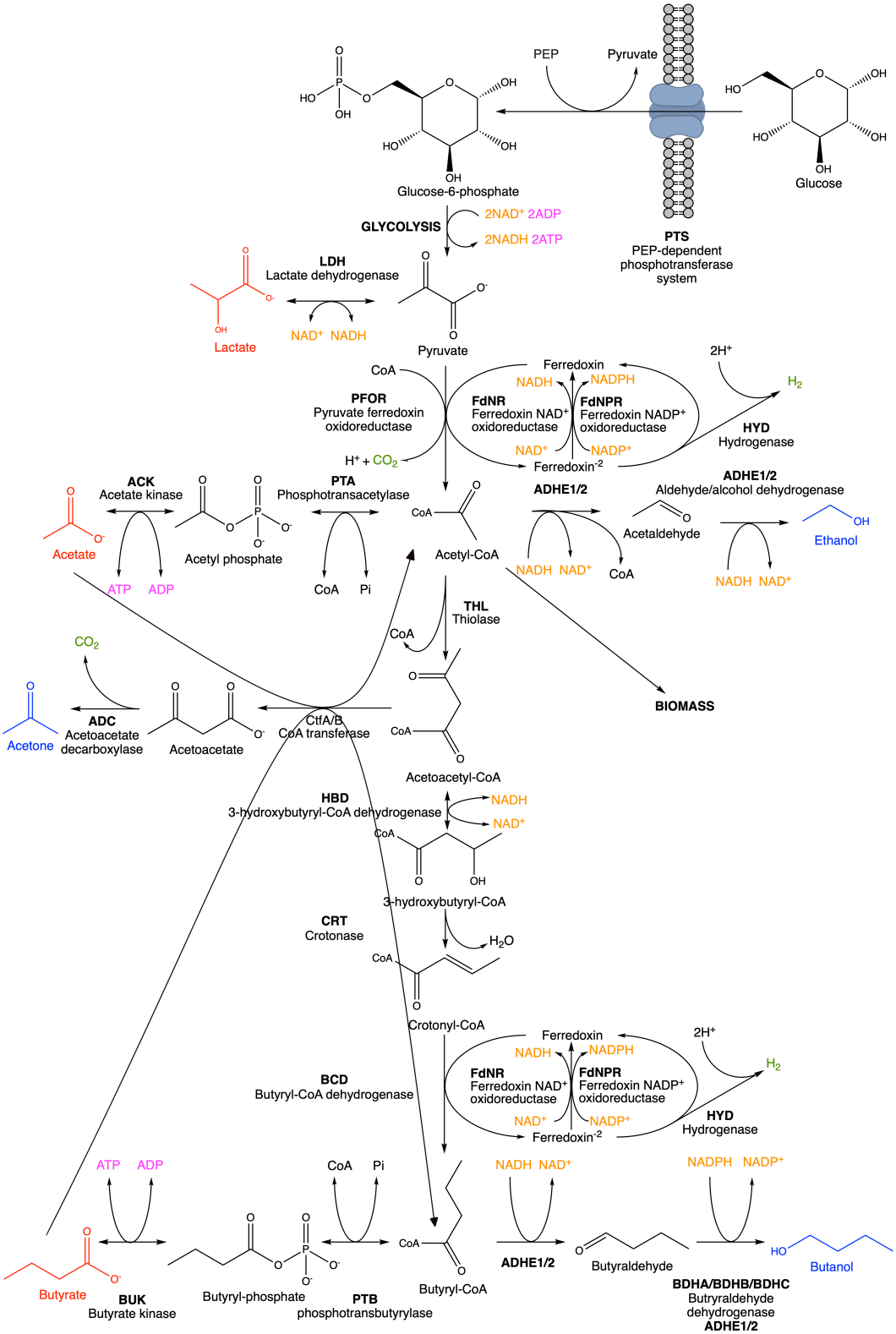


Figure 4: The acetone, butanol and ethanol pathway of *C. acetobutylicum*. Metabolites are coloured accordingly: Acids are shown in red, solvents in blue, and gasses in green, redox co-factors are labelled in amber, ADP/ATP in magenta, and metabolic intermediates are shown in black. Arrows indicate reactions and are labelled in black with the associated enzyme.

in continuous cultures maintained at neutral pH, also finding that coenzyme A-transferase was poorly expressed. Finally, these results were produced again by Peguin and Soucaille (1996) at pH 5.5 using iron limitation and supplementation with methyl viologen as an artificial electron transport chain bypassed hydrogenase, favouring FdNPR activity, resulting in elevated butanol and ethanol synthesis. Fermentations were able to produce ethanol and butanol without acetone, and again, poor expression of coenzyme A-transferase was observed. Peguin *et al*., (1994) also published results showing that intracellular NADH concentrations were increased in the presence of methyl viologen.

C. acetobutylicum also expresses a lactic acid dehydrogenase, however, very little lactic acid is produced under normal fermentation conditions, although, lactic acid has been shown to serve as a carbon source in fermentations containing an additional sugar based carbon source (Bahl *et al*., 1986). A TCA cycle is present, also, and small amounts of acetyl-CoA enter this pathway generating additional, negligible, amounts of NADH. Furthermore, small amounts of acetoin is also produced from a pathway stemming from the pyruvate node (Senger & Papoutsakis, 2008; Yoo *et al.,* 2015).

The sporulation process is initiated as an acid toxicity survival strategy, along with solventogenesis at the end of the acidogenic phase in order to prepare the cells to enter a resilient dormant state in the form of an endospore which is completed with the end of solventogenesis (Jones & Woods, 1986; Al-Hinai *et al.,* 2015).

1.5.2 Regulatory systems governing the transition

Much work has been undertaken to unpick the genetic regulatory systems governing the transition through acidogenesis to solventogenesis and sporulation. This progression is governed by an especially complex regulatory network where Spo0A is the master regulator (Dürre *et al.,* 2002), yet the full extent of the network regulating the transition through the phases remains poorly understood, hampering engineering projects seeking to improve butanol synthesis.

In the Firmicute *Bacillus subtilis*, Spo0A works in conjunction with a network of other regulatory proteins to ensure that sporulation is only initiated after other physiological responses to intracellular conditions or the environment have been rendered null (Perego, 1998; Hoch, 2002; Piggot & Hilbert, 2004). Spo0A is able to up-regulate or down-regulate depending on the placement of the "0A box". When the 0A box is upstream of the promotor and distal to the coding sequence Spo0A positively regulates, when it is downstream or overlapping with the promotor, proximal to the coding sequence, it negatively regulates (Strauch *et al.,* 1990; Fawcett *et al.,* 2000; Molle *et al.,* 2003).

The genes *adhE1*, *ctfA*, *ctfB* are expressed as an operon known as the *sol* operon which reside on the pSOL megaplasmid, *adc* is directly adjacent from the *sol* operon and is transcribed in the opposite direction in *C. acetobutylicum*, however, in other solventogenic saccharolytic *Clostridium* spp. *adc* is expressed as part of this operon (figure 5). The pSOL megaplasmid also harbours the *adhE2* gene, however, this gene is not in the region of *sol* operon.

The *C. acetobutylicum sol* operon and *adc* possess the same 0A boxes (TGNCGAA) in their 5' regulatory regions and are activated upon phosphorylation of Spo0A by a group of orphan kinases (Steiner *et al.,* 2011) which are thought to be expressed under regulation of σ^H (Al-Hinai *et al.,* 2015). Spo0A is directly involved in the transition from acidogenic metabolism to solventogenesis and solventogenic *Clostridium* spp. can be found with 0A boxes in the promotor regions of solventogenic genes including the *sol* operon, *bdhA*, *bdhB* and *adc*.

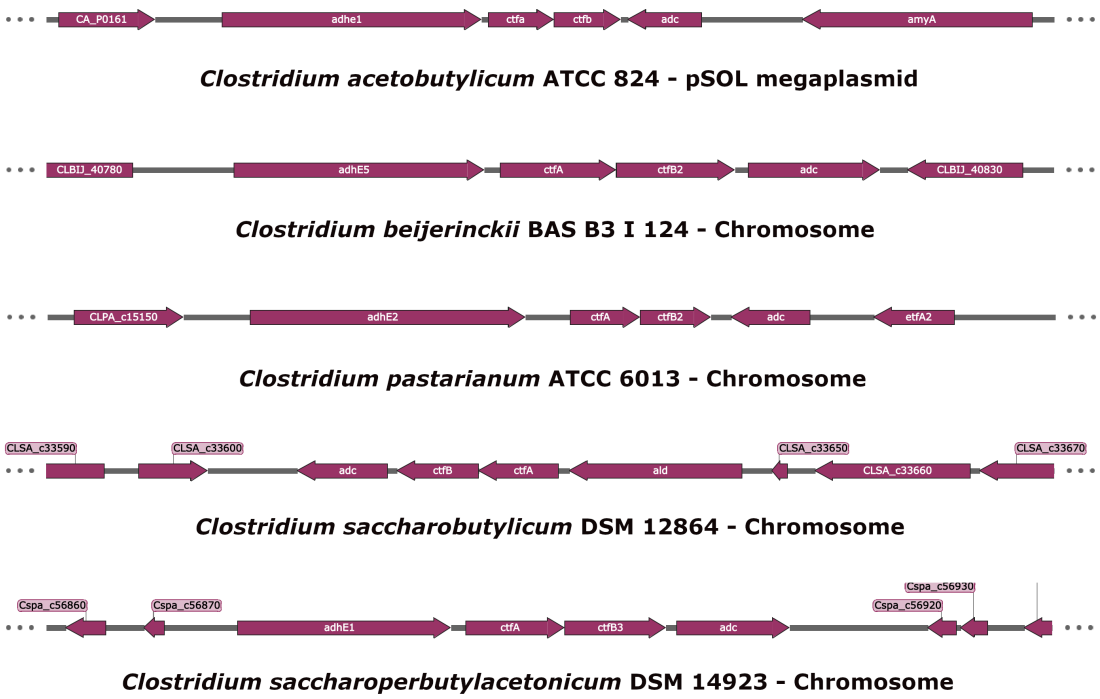


Figure 5: Comparison of the genomic arrangement of the conserved *sol* operon in saccharolytic, solventogenic *Clostridium* spp. Note that the *sol* operon is coded on the pSOL megaplasmid whereas all other species code these genes on the chromosome.

In the *C. beijerinckii* and *C. acetobutylicum* mutagenised 0A boxes upstream of the solventogenic *adc* gene results in impaired expression, whereas mutagenised 0A upstream of the acidogenic *ptb-buk* operon results in continued expression into stationary phase, consistent with the duel capacity to up or down-regulate as observed in *B. subtilis* (Ravagnani *et al*., 2000). A *spo0A* deletion mutant obtained in *C. acetobutylicum* also demonstrated an impaired transition to solventogenesis (Harris *et al.,* 2002).

Like *B. subtilis,* the master regulator Spo0A in *C. acetobutylicum* works in concert with a number of other regulatory systems. Some of which have been characterised.

The redox responsive transcription factor, Rex negatively regulates genes in the C4 intermediate pathway (*hbd, crt, etfAB, bcd*), *ptb-buk*, *adhE2* and *ldh,* allowing

the maintenance of redox homeostasis by electron sequestration into carbon rather than a response to environmental pH.

CodY is poorly understood, however, it is involved in repression of solventogenesis during exponential growth (Zimmerman, 2013). In *B. subtilis* CodY is known to down-regulate the synthesis of branched chain amino acids - the most reduced class of amino acid (Shivers *et al*., 2004). This might suggest the role of CodY is to ensure NADH is prioritised for some other reductive pathway.

The carbon catabolite repression protein CcpA works in the typical manner, where in the presence of glucose, genes required for sugar catabolism are not expressed. However, forward genetics approaches have also demonstrated that this transcription factor is also positively regulating solvent formation and negatively regulating acid butyrate formation (Ren *et al.,* 2010; Ren *et al*., 2012).

The *solB* gene codes for a small non-coding RNA found upstream of adjacent to the *sol* operon. Binding to the *ctfA* portion of the *sol* operon transcript, it has a regulatory function with complex effects, where up-regulation of transcription results in attenuation of solventogenesis, and deletion abolishes solvent formation (Jones *et al.,* 2018).

Quorum sensing has been found to also play an important role in triggering solvent formation. An RNPP type quorum sensing system has been discovered in *C. acetobutylicum* and a quorum sensing system designated QssB has been shown to play a role in activation of solventogenesis, where the over-expression of a repressor (QsrB) significantly dampens solvent production, however, the addition of a synthesised quorum sensing peptide (QspB) can restore solvent production (Kotte *et al.,* 2020).

1.5.4 Molecular tools for genetic engineering in *Clostridium acetobutylicum*

Genetic engineering in *Clostridium* has long been hampered by poor transformation efficiencies and a lack of molecular tools and molecular parts. However, in the last 10 years, this has been largely rectified and an assorted suite of tools are now at the disposal of those wishing to exploit these industrially valuable organisms.

One of the key challenges to overcome when wishing to engineer any organism is the issue of transformation. This problem was solved relatively early for *C. acetobutylicum* with an electroporation protocol published by Mermelstein & Papoutsakis in 1993 which employed the *B. subtilis* phage ϕ3T I methylase in order to overcome the issue of the *C. acetobutylicum* CA_C824I type II restriction enzyme targeting the GCNGC restriction sequence.

The development of the pMTL-80000 modular plasmid system established a standardised approach to rapid cloning of a variety of plasmid origins, promotors and antibiotic resistance selection markers established within the *Clostridium* research community. The modular system contained four regularly arranged modules flanked with corresponding restriction sites in BioBrick format (Heap *et al.,* 2009). This appears to be the only to reported standardised library of DNA parts for synthetic biology in *Clostridium* spp.

In a 2017 review, Gyulev *et al*., collated an updated and comprehensive list of DNA parts with demonstrated efficacy for molecular cloning in *Clostridium* spp. Importantly, several flavin binding fluorescent reporters were catalogued, derived from the LOV (Light-, Oxygen- or Voltage-sensing) domains of several bacterial proteins. These represent important alternatives to green fluorescent protein (GFP) as they are capable of maturing and fluorescing in anaerobic cells, in general enzymatic reporters are relied upon which require time consuming enzymatic assays.

Gyulev *et al*., (2017) also identified a broad expansion of the range of promotors (including constitutive, inducible and fermentation phase specific), insulators, ribosome binding sites, terminators, and other regulatory control components available for the genetic engineering of solventogenic *Clostridium* spp.

More recently, Mordaka & Heap (2018) produced a synthetic library of constitutive promotors of varying strengths based on mutations introduced to the *C. acetobutylicum* P_{thl} promotor. This work together with the collation of parts reported by Gyulev *et al*., (2017) would serve as a good foundation for a new standardised library ideally using the MoClo standardisation system (Moore *et al.,* 2020) which is optimised for automated high-throughput cloning.

Initial attempts to develop methods to prevent the expression of genes of interest relied on single crossover recombination to inactivate genes as a result of vector insertion (Green & Bennet, 1996; Green *et al*., 1996b), however, single crossover insertions are inherently unstable. Later came a more reliable method utilising targeted mobile group II introns which were able to stably prevent the expression of a gene as a result of insertional activation. This technology was made available as ClosTron (Heap *et al.,* 2007; Heap *et al.,* 2010) which has been widely adopted for investigation and engineering of *C. acetobutylicum* and many other species of *Clostridium* (Nguyen *et al.,* 2016; Zhang *et al.,* 2015; Liu *et al.,* 2015; Hoenicke *et al.,* 2014; Wietzke & Bahl*,* 2012; Jia *et al.,* 2012; Cooksley *et al*., 2012; Lehmann *et al*., 2012; Lehmann *et al*., 2012b; Lehmann & Lutke-Eversloh, 2011; Dong *et al.,* 2011). A similar, commercially available system from Sigma is TargeTron which, as with ClosTron, is based on the work of Karberg *et al*., (2001)

The development of allele coupled exchange for *Clostridium* spp. allowed for in-frame insertion or deletion of genes of interest and utilised auxotrophic parent strains and counter selection markers in order to achieve precise markerless, genetic manipulations.

Heap *et al*., (2012) describes a method for selectable insertion of heterologous genes at the *pyrE* locus. This first requires the generation of a disrupted *pyrE* gene

coding for orotate phosphoribosyltransferase, necessary for pyrimidine synthesis resulting in uracil autotrophy. Plasmid transformation is selected for using *catP* on a thiamphenicol and uracil supplemented medium allowing for delivery of a gene or operon to the *pyrE* locus, which relies on sequences of homology flanking the gene of interest recombining with target sequences at the *pyrE* locus. The larger right flanking homology arm recombines first, with its larger length providing increased opportunities for base pairing. This ensures correct orientation of the gene, the smaller left flanking homology arm complements the disrupted *pyrE* gene conferring uracil prototrophy. Uracil prototrophy allows for selection of single crossovers when colonies are transferred to minimal medium lacking uracil but supplemented with thiamphenicol. Stable double crossovers are obtained after strains are re-streaked several times onto plates lacking both uracil and thiamphenicol while plasmid loss is monitored by streaking strains in parallel onto thiamphenicol supplemented plates, where absence of growth indicates excision of the *catP* containing vector. This strategy has been broadly applied across multiple *Clostridium* spp. for a range of applications including expression of hydrolases for the breakdown of complex carbohydrates in *C. acetobutylicum* (Kovacs *et al*., 2013; Willson *et al*., 2016), delivery of therapeutic genes to bacteria as cancer targeting delivery vehicles in *Clostridium sporogenes* (Heap *et al*., 2014), insertion of a lactose inducible sigma factor based gene expression system controlling a *mariner* transposon in *Clostridium pastarianum* (Zhang *et al*., 2015) and for enhanced reproducibility of *Clostridium difficile* NAP1/B1/027 epidemic strain R20291 virulence in a hamster infection model by delivery and expression of the *ermB* erythromycin resistance gene (Kelly *et al*., 2016).

The use of counter selection markers in a disrupted *pyrE* background strain may also be applied in order to obtain in-frame deletion strains utilising adjacent homology arms of equal length on the plasmid vector targeting flanking regions of the gene of interest in the host genome. Crossovers of both homology arms transfers the gene to the vector excising it from the genomic locus. The gene is finally cured from the cell by passaging the strain on selective medium containing 5-

fluoroorotic acid (FOA) which is catalysed to the toxic fluorodeoxyuridine by the activity of orotate phosphoribosyltransferase expressed by a *pyrE* gene located on the vector. To ensure the plasmid is not maintained, this mechanism is coupled with a weak origin of replication. Following this, uracil prototrophy is restored by correction of the disrupted *pyrE* gene (Ng *et al.,* 2013) requiring a knock in plasmid as described previously.

This in-frame deletion method has been improved upon by utilisation of the *codA* counter selection marker which does not require a parent strain with a disrupted *pyrE* gene (Ehsaan *et al.,* 2015). Similarly to the previously described method, following transfer of the target gene, maintenance of the plasmid vector is selected against on media containing 5-fluorocytosine which is catalysed to the highly toxic 5-fluorouracil by cytosine deaminase, coded for by the *codA* gene present on the vector.

Clustered regularly interspaced repeats (CRISPR) and associated Cas proteins are the key components of antiviral adaptive immune systems in prokaryotes. The process of adaptation includes capture of foreign genetic sequences and incorporation into the CRISPR array. CRISPR array transcripts serve as CRISPR RNAs (crRNAs) binding to Cas nucleases allowing the targeting of a specific nucleic acid sequence through base paring for the crRNA (Barrangou *et al*., 2007; Wiedenheft *et al*., 2012). Jinek *et al*., (2012) demonstrated that crRNAs could be synthesised to target specific nucleotide sequences using Cas9 from *Streptococcus pyogenes* to direct sequence specific DNA cleavage. This discovery was followed by the revolutionary development of these two components into a suite of tools to for analysis and engineering of genetic systems across all six kingdoms, as well as a range of therapeutic applications (Knott & Doudna *et al.,* 2018).

Serving as a selection tool for gene knock out CRISPR/Cas systems are, in principle, able to very strongly select against strains which harbour target genes following transformation with plasmid vectors with flanking homology sequences. In a cell expressing Cas9 nuclease and an RNA sequence with homology to the target gene referred to as a single guide RNA (sgRNA). The sgRNA replaces the role of the

crRNA sequence allowing Cas9 to bind with the target gene. Following target binding, Cas9 nuclease will introduce lethal double strand breaks to the chromosomes of cells which have not excised the target gene to the vector, and will allow for the maintenance of cells which have successfully generated in-frame deletions by homologous recombination.

This general approach has been adapted for a variety of engineering applications in *Clostridium* spp. Methods using CRISPR/Cas9 tools normally need to circumvent the inherent toxicity of *S. pyogenes* Cas9 nuclease when delivered both without sgRNA but particularly with sgRNA, as double strand breaks occurring before target excision results in cell death. Bruder *et al*., (2016) and Wang *et al*., (2016) achieved very low transformation efficiency genome edits using vectors containing homology sequences, sgRNA, and *cas9* for *C. acetobutylicum (*0.2 ± 0.0 CFU μg^{-1} DNA) and *C. beijerinckii* (0.38 ± 0.05 CFU μg^{-1} DNA).

Pyne *et al*., (2016) observed similar results attributed to the toxic effects of heterologous Cas9 activity resulting in poor transformation and editing efficiencies, however, transformation efficiency was improved when the vector delivered an sgRNA cassette targeting an endogenous Cas3 causing editing efficiencies to rise to 100% compared to 25% for the heterologous *S. pyogenes* Cas9.

Nickase systems rely on a mutant version of Cas9 (Cas9n) in which one of the two catalytically active domains has been inactivated by amino acid substitutions (Cas9 D10A or Cas9 H840A) resulting in the ability to only cut one strand. Originally this was utilised with two guide RNAs targeting different sequences in the same target gene to enhance specificity and to minimise off target effects (cleavage at an unintended site due to sgRNA mismatching) (Shen *et al*., 2014). Cas9n has since been used to generate knock out mutants in *Clostridium cellulolyticum* with sgRNA in order to circumvent the undesirable potency of Cas9/sgRNA combinations (Xu *et al*., 2015) when delivered with a single guide RNA double strand breaks do not occur, nevertheless, nicks delivered to the target gene are able to produce a selective pressure strong enough to isolate in-frame deletion mutants from a mixed population of wild-type and single-crossover strains. A similar approach has been

replicated in *C. beijerinckii* and *C. acetobutylicum* with editing efficiencies varying from 6.7% to 100% and 18.8% to 100%, respectively (Li *et al*., 2016).

Wasels *et al*., (2017) employed a two plasmid system approach in *C. acetobutylicum*, placing the native *cas9* on one plasmid under the control of the inducible anhydrotetracycline promotor, with a second plasmid included the sequences for the editing templates and gRNA. The plasmids were integrated sequentially and once strains were obtained harbouring both plasmids *cas9* was induced to allow for selection of strains which had undergone homologous recombination. This method had an also had an editing efficiency of 100% and was successfully used for deletions, insertions, and substitutions.

Minton *et al.,* (2016) describe a comprehensive range of solvent producing *Clostridium spp.* industrial chassis to be used in conjunction with the aforementioned tools to facilitate genetic/metabolic engineering endeavours, with the exception of the CRISPR systems, which are yet to prove to be broadly robust and reproducible.

1.5.5 Previous metabolic and bioprocess engineering efforts to improve butanol productivity and yield in *Clostridium acetobutylicum*

The opportunity to improve or alter solvent productivities and yield through metabolic engineering is has been capitalised upon since the development of genetic engineering tools in the saccharolytic, solvent producing *Clostridium* spp.

The earliest report of metabolic engineering to increase solvent productivity in this group of bacteria was published by Green *et al*., (1996) who used integrative plasmids to target the disruption the acetate and butyrate pathway genes *pta* and *buk* in an attempt to redirect carbon flow towards solvent production, in the case of the butyrate pathway, this approach was successful, resulting in improved butanol yields; where acetate was targeted no increase in butanol was observed.

The first use of Group II introns to engineer *C. acetobutylicum* was reported by Jiang *et al*., (2009) who used TargeTron to disrupt the *adc* gene coding for acetoacetate decarboxylase necessary for the production of acetone. Disruption of this gene drastically reduced to amount of acetone produced to <1 g L^{-1} resulting in, essentially, an alcohologenic fermentation, however, this did not result in increased butanol and ethanol yields as is observed when alcohologenesis is induced by addition of methyl viologen (Peguin and Soucaille, 1996). In unbuffered medium, the *adc* disruption mutant produced less than half the amount of butanol than the wild type 2018p strain, possibly owing in part to pH stress, caused by the fact that acid re-assimilation is dependant on acetone synthesis. Calcium carbonate buffered medium resulted in an improvement to butanol yields but the *adc* disruption strain still only produced approximately 80% of the wild type strain. Ethanol synthesis was positively impacted in the buffered medium, where the disrupted mutant produced approximately 30% more than the 2019p wild type strain in comparable medium.

ClosTron was first utilised for metabolic engineering in *C. acetobutylicum* by Lehmann and Lütke-Eversloh (2011) to alter the product profile by targeting introns to the *hbd* gene, preventing the synthesis of C4 products butyrate and butanol, this resulted in the overproduction of ethanol and acetate. Hydrogen production was also

drastically reduced in this strain. This was followed by Lehmann *et al*., (2012) also with an attempt to improve butanol synthesis by deletion of competing pathways, targeting acetate and acetone biosynthesis. Again this did not produce the desired effect, hinting that improvement to butanol yields would necessitate a more sophisticated approach than deletion of pathways competing for carbon flux. Targeting the acetone pathway again succeeded in largely eliminating acetone production and also compromised butanol yields, although in this study ethanol yield was also reduced. This might be attributed to the fact that Jiang *et al*., (2009) used *C. acetobutylicum* EA 2018 as a background and Lehmann *et al*., (2012) used the ATC 824 strain. Targeting acetate synthesis via a *pta* disruption had little effect on acetate yield however the physiology was broadly impacted, albeit butanol and ethanol yields were negatively affected, as was acetone.

Acetate synthesis was also targeted by disruption of the *C. acetobutylicum* ATCC 824 *ack* gene in a study by Kuit *et al*., (2012), the second of the two genes required for this pathway. Acetate synthesis was reduced to 20% of the wild type strain during acidogenesis, the authors also reported a 63% increase in ethanol yield, a 16% increase in butanol and additionally, a 60% increase in acetoin was measured.

ClosTron was also used by Cooksley *et al*., (2012) in a thorough reverse genetics exploration the ABE pathway of *C. acetobutylicum,* generating 10 gene disruption mutants. An *adc* disruption strain was consistent with Lehmann *et al*., (2012) displaying negligible amounts of acetone, reduced butanol formation, and a very small amount of ethanol. An intron was also targeted to *ptb* which had a positive impact on butanol productivity and yield, where the ::*ptb* disruption strain reached the maximum butanol yield in around 30 hours, compared to around 45 hours for the wild type. Maximum yield was approximately 30 mM greater than the wild type. This contrasted earlier reports from Lehmann *et al.,* (2012b) who measured a 75% reduction in butanol yield. In this study a previously uncharacterised gene was investigated - designated *cap0059*. This was hypothesised to be an alcohol dehydrogenase, however, a disruption mutant for this gene resulted in improved metrics for solvents, with a slight yield improvement for butanol, and a 3-4 fold increase in yield for ethanol. This strain also gave improved

productivities for all solvents, and glucose was fully consumed within half the time observed for the wild type strain.

Jang *et al*., (2012) employed multiplex genetic manipulations after using metabolic flux balance analysis to establish the most productive route of the two pathways for butanol synthesis. The analysis determined that direct synthesis of butanol via reduction of butyryl-CoA was more efficient than re-assimilation of butyric acid. With this, both of the acid forming pathways were disrupted, targeting *pta* and *buk* using a TargeTron designed group II intron. Expressing a mutant *adhE1*^{D485G} resulted in a strain yielding 18.9 g L^{-1} butanol in batch fermentations, with a substrate to product conversion rate of 0.71 mol butanol/mol glucose. The strain's productivity was then optimised using biochemical engineering methods, circulating fermentation broth through an adsorbent sepabead column for *in-situ* product recovery. This increased the substrate to product conversion rate to 0.76 and achieving a high productivity of 1.32 g L^{-1} h^{-1}.

Given the unsuitability of acetone as a fuel Lee *et al.,* (2012) and Dusséaux *et al*., (2013) both saw the opportunity to convert acetone to isopropanol in order to produce an alcohol fuel blend as the final fermentation product of the *C. acetobutylicum* fermentation. In both studies, the butyrate pathway was disrupted to maximise carbon flux to acetate during the acidogenic phase, ensuring maximum precursor was available for acetone to isopropanol conversion during solventogenic stage. Lee *et al.,* (2012) then engineered the expression of isopropanol alcohol dehydrogenase from *C. beijerinckii* NRRL B-593 (*adh*$_{B-593}$) under the solventogenic phase *adc* promotor, with a separate operon also under P_{adc} expressing *ctfA*/*B* and *adc* to re-assimilate acetate while generating acetone - this approach generated 20.4 g L^{-1} total alcohol in a batch fermentation. Following this, gas stripping of the volatile solvents was employed to increase yield to 35.6 g L^{-1} in 45 hours. Dusséaux *et al*., (2013) expressed *adh*$_{B-593}$ with *ctfA*/*B* and *adc* as a single operon under the constitutive P_{thl} batch fermentation optimised at pH 4.8 achieved a similar but slightly higher yield of 21 g L^{-1}. These studies successfully demonstrated that the ABE fermentation could be transformed into an IBE fermentation with little to no waste products.

The most recent advancement demonstrating the economic utility of combined metabolic and process engineering approaches was reported by Nguyen *et al*., (2018). Here, advanced genetic engineering technology was implemented, making use of in-frame deletion by homologous recombination to delete the lactate, acetone and butyrate pathways. Additionally, *rex* was also deleted to prevent repression under low NADH conditions allowing *thlA*, *crt*, *bcd*, *etfA*, *etfB*, and *adhE2*, to be expressed constitutively. *E. coli atoB* was used in to replace of *thlA* using homologous recombination targeting the original locus, as this heterologous thiolase has higher catalytic activity, is less sensitive to CoA inhibition (Mann & Lütke-Eversloh, 2013), and is not sensitive to redox switch (Kim *et al*., 2015). The *hbd* gene from *Clostridium kluyveri* utilises NADPH as a redox cofactor, instead of NADH, given that the NADPH/$NADP^+$ ratio is at least 70 times higher than the NADH/NAD^+ ratio, this also replaced the native *C. acetobutylicum hbd* at its original locus using homologous recombination. With this strain, a continuous process was able to stably maintain 9-10 g L^{-1} for over 45 days. An *in-situ* product recovery system was employed to further optimise yields and productivity. In this case, low pressure distillation significantly lowers the boiling point of butanol which normally boils at 117.7°C under normal atmospheric pressure. This allows all three of the fermentation products to be removed from the broth while allowing cells to recycle back to the bioreactor. This combined metabolic and bioprocess engineering approach resulted in a productivity of 14 g L^{-1} h^{-1}, which is economically comparable to competitive wet corn milling continuous ethanol processes.

A number of *in-situ* product recovery strategies have been described, they are important because they benefit productivity in two ways. Firstly, they remove toxic products (in the case of ABE solvents) from the bioreactor, preserving optimum conditions for product synthesis, and secondly, it shifts the thermodynamic equilibrium in favour of the product forming route. As demonstrated by Jang *et al*., (2012) and Lee *et al*., (2012), *in-situ* recovery can be used to improve productivities of *Clostridium* spp. solvent biosynthesis. Given that butanol has a high boiling point, Jang *et al*., (2012)

opted for a resin adsorption technique using sepabeads SP850 constructed from styrene-divinylbenzene. Nielsen *et al*., (2009) demonstrated that styrene-divinylbenzene resins adsorbed butanol from aqueous solutions with the greatest affinity when compared to other commercially available adsorbents. This works on the principle that butanol is less polar than water, and will therefore adsorb to the hydrophobic surface of the resin, allowing the soluble components of the medium and the water to pass though uninhibited, recirculating back into the bioreactor. Adsorbed butanol can then be efficiently recovered using thermal treatment. Lee *et al*., (2012) focused on the volatility of the isopropanol and ethanol in the IBE fermentation and employed gas stripping, which relies on sparging nitrogen thorough the broth as a carrier for the volatile solvents, with subsequent capturing of the vaporised solvents using a condenser fitted to the bioreactor gas exit vent.

Toxicity can also be mitigated at the cellular level and is generally approached using either directed evolution or rational engineering. Tomas *et al*., (2003) demonstrated that over-expression of the *groESL* operon - a modulator for the CIRCE regulon including *hsp*09 which codes for a heat shock protein - resulted in increased productivity, yield, and solvent tolerance.

Xu *et al*., (2015) evolved a strain designated *C. acetobutylicum* JB200 in a fibrous bed reactor, which was able to reach 21 g L^{-1} butanol compared to 12.6 g L^{-1} produced by the parent strain. Comparative genomics identified a SNV resulting in a frameshift in the gene *cac3319* - a putative histidine kinase - resulting in a premature stop codon. This was replicated in the ATCC 55025 parent strain using a ClosTron targeting *cac3319,* producing a strain with a 44.4% butanol yield improvement and a 90% improvement to productivity. The authors hypothesised that since *cac3319* is associated with Spo0A activation that tolerance was enhanced as a consequence of the disrupted *cac3319*. Indeed, when subjected to increasing concentrations of butanol, both strains ::*cac3319* and JB200 were able to continue growing comparatively better than the wild type when exposed to increasing butanol stress.

Chapter 2: Materials and Methods

2.1 Microbial cultivation & media

2.1.1 Strains, plasmids, and primers

Strain	Features	Associated Plasmids	Reference
***E. coli* 5 α**	*E. coli* DH 5α derivative cloning strain.	-	NEB Cat#: C2987I
***E. coli* TOP10 pAN2**	*E. coli* 5α harbouring the pAN2 plasmid for methylation of plasmids to be transformed into *C. acetobutylicum*.	pAN2	Heap *et al.,* (2007)
***C. acetobutylicum* ATCC 824**	Wild type strain obtained from ATCC.	-	-
C. acetobutylicum* ATCC 824 Δ*pyrE	Partial deletion of *pyrE* gene resulting uracil auxotrophy	pMTL-JH12	Heap *et al*., (2012)
C. acetobutylicum* ATCC 824 Δ*pyrE* Δ*bcd	Deletion of *bcd* at native chromosomal locus using . Retention of gene in the chromosome integrated vector.	pMTL8-150Δ *bcd*	Ng *et al*., (2013)
C. acetobutylicum* ::*ptb	ClosTron insertion to the *ptb* at 87|88.bp	pMTL007*ptb*	Lehmann *et al.,* (2012b)
C. acetobutylicum* + *ter	*T. denticola ter* inserted downstream of *thlA* with *ermB* as a marker - transcribed from *thlA* promotor	pMTL8-WOK37	This study
***C. acetobutylicum* + *ter*_FLAG**	*T. denticola ter*_FLAG inserted downstream of *thlA* with *ermB* as a marker - transcribed from *thlA* promotor	pMTL8-WOK40	This study
C. acetobutylicum* Δ*pyrE* + *ter	*T. denticola ter*_FLAG inserted downstream of *thlA* with *ermB* as a marker - transcribed from *thlA* promotor	pMTL8-WOK40	This study
***C. acetobutylicum* Δ*pyrE* + *ter*_FLAG**	*T. denticola ter*_FLAG inserted downstream of *thlA* with *ermB* as a marker - transcribed from *thlA* promotor.	pMTL8-WOK40	This study
C. acetobutylicum* + *ter*_FLAG Δ*bcd/ etfAB	Deletion of *bcd* + *etfAB* at native chromosomal locus using. Retention of gene in the chromosome integrated vector. *C. acetobutylicum* + *ter*_FLAG is the parent strain.	pMTL80150_ RJH21Δbcd/ etfAB	This study
***C. acetobutylicum* + JH16 -ve control**	-ve control inserting *ermB* marker alone - transcribed from *thlA* promotor	pMTL8-JH16	Heap *et al*., (2012)
***C. acetobutylicum* + *ter*_FLAG + *bgaR_tcdR* + ButOH**	C4 operon genes (*thlA*, *hbd*, *crt*, *adhE2*) inserted at *pyrE* locus under control of the lactose inducible Ptcdb/TcdR expression system. *C. acetobutylicum* Δ*pyrE* + *ter*_FLAG is the parent strain.	pMTL85140-RJH22 (pbgaL_bglR _tcdR_ptcdB _CacButOH)	This study.

Strain	Features	Associated Plasmids	Reference
***C. acetobutylicum bgaR_tcdR*_ctrl**	Lactose inducible Ptcdb/TcdR expression system inserted at *pyrE*	pMTL-HZ7	This study.

Table 1: List of strains used in this study with associated engineering plasmids and references.

Plasmid	Features	Reference
pMTL007C-E2	Vector for ClosTron delivery.	Heap *et al*., (2010)
pMTL007C-E3::*bcd*496|497	ClosTron plasmid targeting *bcd* at 496|497 insertion site.	Heap *et al*., (2010)
pMTL-ME6	*pyrE* correction vector.	Ehsaan *et al*., (2016)
pMTL8-150Δ*bcd*	In-frame deletion vector with pyrE counter selection marker targeting *bcd*.	Ng *et al.,* (2013)
pMTL8-KZ146	Double counter selection marker based in-frame deletion targeting *bcd.*	This study
pMTL8-WOK37	*Treponema denticola* codon optimised *ter* inserted into pMTL8-JH16 parent vector.	Heap *et al*., (2012)
pMTL8-WOK40	*Treponema denticola* codon optimised and FLAG tagged *ter* inserted into pMTL8-JH16 parent vector.	Heap *et al*., (2012)
pMTL8-JH16	*thlA* downstream gene insertion vector also used as a -ve control.	Heap *et al*., (2012)
pMTL80150_RJH21Δ*bcd/etfAB*	In-frame deletion vector with codA counter selection marker targeting *bcd*.	Ehsaan *et al*., (2016)
pMTL83140-RJH40(pthl_cas9n-pJ23119_gRNA)	Cas9 nickase under P_{thl}, gRNA under strong synthetic promotor P_{J23119}. Targets *bcd* with flanking homology arms and gRNA.	Ingle *et al*., (2019)
pMTL83140-RJH40(PfacOID_cas9-ParaE_gRNA-PlacIQ_lacI)	*S. pyogenes* Cas9 codon optimised for *C. acetobutylicum* under P_{facOID}, *lacI* under P_{lacIQ}, gRNA under P_{araE}. Targets *bcd* with flanking homology arms and gRNA.	Kovacs *et al*., (2013); Wilson *et al*., (2016).
pMTL83140-RJH40(PfacOID_cas9-pJ23119_gRNA-Pthl_lacI)	*S. pyogenes* Cas9 codon optimised for *C. acetobutylicum* under P_{facOID}, *lacI* under P_{thl}, gRNA under P_{araE}. Targets *bcd* with flanking homology arms and gRNA.	Kovacs *et al*., (2013); Wilson *et al*., (2016).

Plasmid	Features	Reference
pMTL83140-RJH41(pthl_cas9n-pJ23119_gRNA)	Cas9 nickase under P_{thl}, gRNA under strong synthetic promotor P_{J23119}. Targets *pta* with flanking homology arms and gRNA.	Ingle *et al*., (2019)
pMTL83140-RJH41(PfacOID_cas9-ParaE_gRNA-PlaclQ_lacI)	*S. pyogenes* Cas9 codon optimised for *C. acetobutylicum* under P_{facOID}, *lacI* under P_{laclQ}, gRNA under P_{araE}. Targets *pta* with flanking homology arms and gRNA.	Kovacs *et al*., (2013); Wilson *et al*., (2016).
pMTL83140-RJH41(PfacOID_cas9-pJ23119_gRNA-Pthl_lacI)	*S. pyogenes* Cas9 codon optimised for *C. acetobutylicum* under P_{facOID}, *lacI* under P_{thl}, gRNA under P_{araE}. Targets *pta* with flanking homology arms and gRNA.	Kovacs *et al*., (2013); Wilson *et al*., (2016).
pMTL83140-RJH40(riboswitch_cas9)	*S. pyogenes* Cas9 under *Clostridium botulinum* P_{fdx} with a downstream synthetic theophylline riboswitch inhibiting ribosome binding. gRNA under strong synthetic promotor P_{J23119}. Targets *bcd* with flanking homology arms and gRNA.	Canadas *et al*., (2019)
pMTL83140-RJH41(riboswitch_cas9)	*S. pyogenes* Cas9 under *Clostridium botulinum* P_{fdx} with a downstream synthetic theophylline riboswitch inhibiting ribosome binding. gRNA under strong synthetic promotor P_{J23119}. Targets *pta* with flanking homology arms and gRNA.	Canadas *et al*., (2019)
pMTL83140-Placlq_lacI/FacOID_cas9	Plasmid backbone and P_{laclQ} controlled lacI /P_{FacOID} controlled *cas9* for assay of *cas9* toxicity. Targeting components are not present.	Kovacs *et al*., (2013); Wilson *et al*., (2016).
pMTL83140-Pthl_lacI/FacOID_cas9	Plasmid backbone and P_{thl} controlled lacI /P_{FacOID} controlled *cas9* for assay of *cas9* toxicity. Targeting components are not present.	Kovacs *et al*., (2013); Wilson *et al*., (2016).
pMTL83140-SBRC_cas9n	Plasmid backbone and Cas9 nickase under P_{thl} for assay of *cas9* toxicity. Targeting components are not present..	Ingle *et al*., (2019)
pMTL83140-riboswitch_cas9-ve	*S. pyogenes* Cas9 under *Clostridium botulinum* P_{fdx} with a downstream synthetic theophylline riboswitch inhibiting ribosome binding. For assay of *cas9* toxicity. Targeting components are not present..	Canadas *et al*., (2019)
pMTL8-5141	Control vector for *cas9* tox assay.	Heap *et al* (2009)
pMTL-HZ7	Vector for insertion of lactose inducible BglR controlled tcdR based expression system at the *pyrE* locus.	This study

Plasmid	Features	Reference
pMTL85140-RJH22 (PbgaL_bglR_tcdR_PtcdB_CacButOH)	Lactose inducible BglR controlled tcdR for expression of *C. acetobutylicum thlA*, *hbd*, *crt*, *etfA*/*B*, *adhE2*.	This study

Table 2: Plasmids used in this study and associated references.

Associated Template Sequence Identity	Primer Name	Primer Sequence	Purpose
pMTL8-KZ146	bcd_LHA_R2	tataCGCCGGCgcatataaacttacctcctatct	Generate RHA frag for 3' overhang extension with *pyrE*/*catP* marker. 5' MreI restriction site.
	bcd_LHA_F	tcgcATGCGCATagtgaaagccttctaacac	Generate RHA frag for 3' overhang extension with *pyrE*/*catP* marker. 5' FspI restriction site.
	catP_pyrE_R2	tataCGTCGACGgttgtaatcgatcgtatctgagatagctgTTAGGGTAACAAAAAACAC	Generate *pyrE*/*catP* marker frag for 5' overhang extension with LHA and 3' overhang extension with RHA.
	catP_pyrE_F2	tataATGCGCATcagctatctcagatacgatcgattacaaagAGTGGGCAAGTTGAAAAATTC	Generate *pyrE*/*catP* marker frag for 5' overhang extension with LHA and 3' overhang extension with RHA.
	bcd_RHA_R	atataCGCGCGCGtaatttaaggaggttaagag	Generate RHA frag for 3' overhang extension with *pyrE*/*catP* marker. 5' SalI restriction site.
	bcd_RHA_F	atataCGCGCGCGtaatttaaggaggttaagag	Generate RHA frag for 3' overhang extension with *pyrE*/*catP* marker. 5' MauBI restriction site.
	bcd_CT_ch_F	ACTGAGCCAAATGCAGGAAC	For identification of a group II insert using ClosTron - binds to *bcd*.

Associated Template Sequence Identity	Primer Name	Primer Sequence	Purpose
pMTL007C-E3::*bcd*496|497	bcd_CT_ch_R	CTTTTCCTTCTTTACCAATCATGT	For identification of a group II insert using ClosTron - binds to *bcd*.
	EBS_universal	CGAAATTAGAAACTTGCGTTCAGTAAAC	Intron specific primer - pairs with bcd_CT_ch_F.
***C. acetobutylicum* ATCC 824 Δ*pyrE* Δ*bcd* genomic DNA.**	bcdNotIFw1	gcggccgcttatctaaaaatttttcctgaaataactaatttctgaac	Amplifies *bcd* CDS - contains NotI as this primer was originally intended for the generation of a complementation strain. Eventually used for detection of vector integration event with excised fragment - demonstrated by amplification of *bcd*.
	bcdRv1	atggattttaatttaacaagagaacaag	Used for detection of vector integration event with excised fragment - demonstrated by amplification of *bcd*.
	bcd_chr_F	gtttattgcaagaagtgaataaaagccgagattagt	Primer pair flanks 1 kB upstream and downstream of *bcd* to produce a 2 kB fragment in the case of a deletion and a 3.1 kB fragment otherwise
	bcd_chr_R	ctaaaaccttatctgctccatgagataataaatccttt	Primer pair flanks 1 kB upstream and downstream of *bcd* to produce a 2 kB fragment in the case of a deletion and a 3.1 kB fragment otherwise
	Csp-pyrE-HpaI-sF1	AATATTGTTAACTAAGGAGAAGATATAAATGAGTAATATAAATGTTATAG	Amplifies *C. sporogenes pyrE* CDS. Originally a cloning primer. Repurposed for identification of vector integration event.

Associated Template Sequence Identity	Primer Name	Primer Sequence	Purpose
	Csp-pyrE-HpaI-sR1	AATATTGTTAACTTATTTTGTTCTCTACTACCTGGTTTTACAAAAGGT	Amplifies *C. sporogenes pyrE* CDS. Originally a cloning primer. Repurposed for identification of vector integration event.
Genomic DNA from strains expressing *ter* at the *thlA* locus	JH16ACEInsert_Fw	atcataaagagttacaacctttaacattcttttaatattcagattcaa	For confirmation of inserts at the thlA locus using the pMTL8-JH16 integration vector.
	JH16ACEInsert_Rv	gcggacaaggaacagcaatattgcta	For confirmation of inserts at the thlA locus using the pMTL8-JH16 integration vector.
pMTL80150_RJH21Δ*bcd/etfAB*	RJH21_Frag1.REV	tagataggaggtaagtttatCGCGCGCGgataaataaaaagaat	Generation of LHA for pMTL80150_RJH21Δ *bcd/etfAB* constructed by Gibson assembly.
	RJH21_Frag1.FOR	aacatcgtagaaatacggtgttttttgttaccctaagtttCGTCGACGtattatatataaaaaatcccc	Generation of LHA for pMTL80150_RJH21Δ *bcd/etfAB* constructed by Gibson assembly.
	RJH21_Frag2.FOR	tttttatttatcCGCGCGCGataaacttacctcctatctatttttgaagcc	Generation of RHA for pMTL80150_RJH21Δ *bcd/etfAB* constructed by Gibson assembly.
	RJH21_Frag2.REV	gttaagggattttggtcatgagattatcaaaaaggagtttaagaagtggaataaaagccgagattagtacggtaat	Generation of RHA for pMTL80150_RJH21Δ *bcd/etfAB* constructed by Gibson assembly.
	bcd-4_gRNA(J23119)XbaI.F	tcgattCTAGAttgacagctagctcagtcctaggtataatactagtcgaagcaagagcttacatgagttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttg	Pairs with “bcd-4_gRNA(J23119).R” to generate a 200 bp sgRNA transcription cassette. Amplified via a 34 bp overhang.

Associated Template Sequence Identity	Primer Name	Primer Sequence	Purpose
pMTL83140-RJH40(pthl_cas9n-pJ23119_gRNA)/ pMTL83140-RJH40(PfacOID_cas9-ParaE_gRNA-PlaclQ_lacl)/ pMTL83140-RJH40(PfacOID_cas9-pJ23119_gRNA-Pthl_lacl)/ pMTL83140-RJH40(riboswitch_cas9)	bcd-4_gRNA(J23119).R	tgaacagcgatcgcataaaaataagaagcctgcaaatgcaggcttcttatttttataaaaaaagcaccgactcggtgccacttttcaagttgatacggactagccttattttaacttg	Pairs with "bcd-4_gRNA(J23119)XbaI.F" to generate a 200 bp sgRNA transcription cassette. Amplified via a 34 bp overhang.
	SOE(gRNA).F1	tcgattTCTAGAttgacagctagctcag	Amplifies gRNA cassette with overhang extension overhangs. Introduces XbaI for restriction digest drop into CRISPR vector.
	SOE40.R1	ggaggtaagtttattgaacagcgatcgcataaaaataagaagc	Amplifies gRNA cassette with overhang extension overhangs
	SOE40.F2	gcgatcgctgttcaataaacttactcctatctatttttgaagcct	Amplifies LHA with overhang extension overhangs
	bcdHASOE2.R	aatggggatttttatatataataaagaagtgaataaaagccg	Amplifies LHA with overhang extension overhangs
	bcdHASOE3ii.R	taatctcggcttttattcacttcttattatatataaaaaatccccatttgataatgg	Amplifies RHA with overhang extension overhangs
	bcdHASOE1.F2	gaatgcggcgcgccgataaataaaaagaattatttaaagcttattatgccaaa	Amplifies RHA with overhang extension overhangs. Introduces AscI for restriction digest drop into CRISPR vector.
pMTL83140-RJH40(PfacOID_cas9-pJ23119_gRNA-Pthl_lacl)/ pMTL83140-RJH41(PfacOID_cas9-pJ23119_gRNA-	pthl+PmeI.F	gctaatGTTTAAACttttaacaaaatatattgataaaaataataatagtgggtataattaagt	Introduces P_{thl} control over *lacI* by overhang extension. 5' PmeI restriction site.
	lacI+pthl.R	cgttactggtttcactctaactaacctcctaaattttgatacggggtaac	Introduces P_{thl} control over *lacI* by overhang extension.
	pthl+LacI.F	aggaggttagttagagtgaaaccagtaacgttatacgatgtcg	Introduces P_{thl} control over *lacI* by overhang extension.

Associated Template Sequence Identity	Primer Name	Primer Sequence	Purpose
pJ23119_gRNA-Pthl_lacI)	lacI+PmeI.R	gctaatgtttaaactcactgcccgctttccagt	Introduces P_{thl} control over *lacI* by overhang extension. 3' PmeI restriction site.
pMTL83140-RJH41(pthl_cas9n-pJ23119_gRNA)/ pMTL83140-RJH41(PfacOID_cas9-ParaE_gRNA-PlacIQ_lacI)/ pMTL83140-RJH41(PfacOID_cas9-pJ23119_gRNA-Pthl_lacI)/ pMTL83140-RJH41(riboswitch_cas9)	PJ23119+pta_gRNA+XbaI.F	tcgattCTAGAttgacagctagctcagtcctaggtataatactagtactgttgttcaggctcaaaggttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttg	Pairs with "J23119+pta_gRNA+AsiSI.R" to generate a 200 bp sgRNA transcription cassette. Amplified via a 34 bp overhang.
	J23119+pta_gRNA+AsiS.R	tgaacagcgatcgcataaaaataagaagcctgcaaatgcaggcttcttatttttataaaaaaagcaccgactcggtgccacttttcaagttgatacggactagccttattttaacttg	Pairs with "PJ23119+pta_gRNA+XbaI.F" to generate a 200 bp sgRNA transcription cassette. Amplified via a 34 bp overhang.
	41SOE.F1	tcgattCTAGAttgacagctagctcag	Amplifies gRNA cassette with overhang extension overhangs. Introduces XbaI for restriction digest drop into CRISPR vector.
	41SOE.R1	gaagggagaaaatatgaacagcgatcgcataaaaataagaagc	Amplifies gRNA cassette with overhang extension overhangs
	41SOE.F2	gcgatcgctgttcatattttctcccttcataatgccaataattatattaacc	Amplifies LHA with overhang extension overhangs
	ptaHASOE2.R	tttagtattcactggtggggctattaatattatccg	Amplifies LHA with overhang extension overhangs
	ptaHASOE3.R	cggataatattaatagccccaccagtgaat	Amplifies RHA with overhang extension overhangs
	ptaHASOE1.F2	gaatgcggcgcgccggaggatttttatgaaaaacttagttattaact	Amplifies RHA with overhang extension overhangs. Introduces AscI for restriction digest drop into CRISPR vector.

Associated Template Sequence Identity	Primer Name	Primer Sequence	Purpose
pMTL85140-RJH22 (pbgaL_bglR_tcdR_ptcdB_CacButOH)	thl.REV	aaaatattcaataaaaatattatagtaaaggagaaaatcaaatgaaagaagttgtaatagctagtgcagtaaga	Gibson assembly of butanol synthesis operon for insertion into pMTL-HZ7 lac inducible TcdR vector.
	thl.FOR	ataacacatacctttttcatctagcacttttctagcaatattgctgttcc	
	hbd.REV	tattgctagaaaagtgctagatgaaaaaggtatgtgttataggtgcagg	
	hbd.FOR	atgacattgtttagttccatttattttgaataatcgtagaaacctttcctgattttcttcc	
	crt.REV	tctacgattattcaaaataaatggaactaaacaatgtcatccttgaaaagg	
	crt.FOR	ttttgatttgtaactttcatctatctatttttgaagccttcaattttctttttctctatgaaag	
	adhE2.REV	aaggcttcaaaaatagatagatgaaagttacaaatcaaaaagaactaaaacaaaagctaaa	
	adhE2.FOR	agtgccaagcttgcatgCCATGgacgcgtgacgtcgaaacacattcttttatttttaaaatgattttatatagatatccttaagttcac	

Table 3: Custom primers used in this study.

Module Type	Module		Primer	Sequence	Position
Gram + Replicon	3	pCB102/repH	83XXX-LF	CCAACACATCAAGCCGTTAG	AscI + 301
			83XXX-LR	ACGGCTTGATGTGTTGGTAG	AscI + 317c
			83XXX-RF	CTACGTCCAAAGCCGTTTCC	FseI - 226
			83XXX-RR	GGAAACGGCTTTGGACGTAG	FseI - 207c
	5	pIM13	85XXX-LF	ACGTCAATGCCGAGCGAAAG	AscI + 118
			85XXX-LR	TTCCGACGCTTATTCGCTTC	AscI + 56c
			85XXX-RF	TACTAATGAGAGGCGACGAC	FseI - 146
			85XXX-RR	TGCCTCTTGCTCAAAGTTCC	FseI - 76c
Marker	1	catP	8X1XX-LF	CCTTGTACCTACAGCATGAC	FseI + 160
			8X1XX-LR	GCTGTAGGTACAAGGTACAC	FseI + 174c
			8X1XX-RF	CGCAGTATGTGACGGATTTC	PmeI - 147
			8X1XX-RR	CGAAGGTTGACCACGGTATC	PmeI - 266c
	2	ermB	8X2XX-LF	AAAGCCATGCGTCTGACATC	FseI + 456
			8X2XX-LR	ATGTCAGACGCATGGCTTTC	FseI + 474c
			8X2XX-RF	CGCCATACCACAGATGTTCC	PmeI - 267
			8X2XX-RR	CTGGAACATCTGTGGTATGG	PmeI - 246c
Gram - Replicon	1	p15a	8XX1X-LF	AACCGAGGTAACTGGCTTGG	PmeI + 116
			8XX1X-LR	CAAGAGATTACGCGCAGACC	PmeI + 48
			8XX1X-RF	AGTCGAACGACCGAGCGTAG	SbfI - 36
			8XX1X-RR	TACGCTCGGTCGTTCGACTG	SbfI - 18c
	4	ColE1	8XX4X-LF	GTTCCACTGAGCGTCAGACC	PmeI + 50

Module Type	Module		Primer	Sequence	Position
			8XX4X-LR	GTCTGACGCTCAGTGGAACG	PmeI + 68
			8XX4X-RF	AGCGGAAGAGCGCCCAATAC	SbfI - 34
			8XX4X-RR	CTGCGTATTGGGCGCTCTTC	SbfI - 101c

Table 4: Screening and sequencing primers for use with the pMTL80000 system components.

2.1.2 Microbiological cultivation media

Agarose may be added to broth media at 1% w/v to make solid agar plates.

Component	Mass (g L-1)
NaCl	10
Yeast extract	5
Tryptone	10
Sterilise by autoclaving at 121˚C for 15 minutes.	

Table 5: Composition and prep instructions for Luria bertani (LB) medium.

Component	Mass (g L-1)
Tryptone	20
Yeast extract	10
NaCl	20
Glucose (add as a 50% solution after sterilisation)	5
pH to 7.4 using NaOH/HCl. Sterilise by autoclaving at 121˚C for 15 minutes.	

Table 6: Composition and prep instructions for 2×YTG medium.

Component	Mass (g L-1)
KH_2PO_4	0.75
K_2HPO_4	0.75
$MgSO_4$ ($7H_2O$)	0.4
$MnSO_4$ (H_2O)	0.01
$FeSO_4$ (H_2O)	0.01
Asparagine	2
Yeast extract	5
Cysteine	0.5
Glucose (add as a 50% solution after sterilisation)	10
pH to 6.2-6.6 using NaOH/HCl. Sterilise by autoclaving at 121˚C for 15 minutes.	

Table 7: Composition and prep instructions for CGM medium.

Component	Mass (g L-1)
Yeast extract	13
Peptone	10
Soluble starch	1
Sodium chloride	5
Sodium acetate	3
Cysteine hydrochloride	0.5
Agar	0.5
Glucose (add as a 50% solution after sterilisation)	5
pH to 6.8 ± 0.2 using NaOH/HCl. Sterilise by autoclaving at 121˚C for 15 minutes.	

Table 8: Composition and prep instructions for RCM. Ordered as a premix from Oxoid Cat # CM0149 (Hirsch & Grinstead 1954).

Component	Mass (g L-1)
Tryptone	20
Yeast extract	5
NaCl	0.5
KCl	0.186
MgCl2 (6 H_2O)	2.033
MgSO4 (7 H_2O)	2.645
Glucose	3.603

Table 9: SOC Ordered in pre-made 5 mL batches from Invitrogen Cat#: 15544-054

Component	Supplier	Sterilisation	Stock Concentration	Component Volume (mL)	Final Concentration
Casamino Acids	Sigma N4767	Autoclave	Dissolved in ELGA water to 800 mL	800	4 g L^{-1}
$MgSO_4$	Sigma M2773				200 mg L^{-1}
Glucose	Sigma D9434	Sterile filter (0.2 μm)	50% W/V	120	60 g L^{-1}
K_2HPO_4	Sigma P2222	Sterile filter (0.2 μm)	100×	10	0.5g L^{-1}
KH_2PO_4	Sigma P5655				0.5g L^{-1}
$MnSO_4$	Sigma M8179	Sterile filter (0.2 μm)	100×	7.58	7.58 mg L^{-1}
$FeSO_4$	Sigma F8263	Sterile filter (0.2 μm)	100×	10	10 mg L^{-1}
4-aminobenzoic acid	Sigma A9878	Sterile filter (0.2 μm)	1000×	1	1 μg L^{-1}
Thiamine-HCl	Sigma T4625				1 μg L^{-1}
Biotin	Sigma B4501				2 ng L^{-1}
$CaCO_3$	Sigma	Autoclave	50×		5 g L^{-1}
ELGA Water	Millipore	Autoclave	-	Add to 1L	
Agar (plates only)	Oxoid (LP0011)	Autoclave	-	Added directly to Autoclaved Casamino acids/ Manganese component.	15 g L^{-1}

Table 10: Clostridial Basal Medium (CBM) - A critical medium as is primarily used to assay physiology of engineered strains. Components are prepared as detailed above and added to Casamino acid/MgSO4 component in the order listed to avoid oxidation and precipitation of the iron. (O'Brien & Morris 1971)

2.1.3: General storage and handling of *Clostridium acetobutylicum*

All *Clostridium* microbiological manipulations were executed in an anaerobic cabinet (80% nitrogen, 10% hydrogen 10% CO_2 -Don Whitely Scientific) at 37˚C unless otherwise stated. Wild type *Clostridium acetobutylicum* ATCC 824 was streaked (cryo-cultures) or spread plated (spore stocks - briefly pasteurised at 80˚C for 3 minutes) onto an anaerobic CGM agar plate, having been incubated in a 37˚C anaerobic cabinet for at least four hours. *Clostridium acetobutylicum* ATCC 824 stocks were either stored as a cell suspension in 12.5% glycerol 2×YTG at -80˚C or as a spore stock prepared from CGM plates incubated for five days before suspending a 10 μL loop in 5mL PBS and subsequently aliquoting in 250 μL volumes. All liquid media was used anoxically by placing in anaerobic cabinet overnight ≥16 hours - up to 500 mL at a time. Plates were incubated for ≥ 1 hour.

2.1.4: General storage and handling of *E. coli*

E. coli manipulations were carried out aseptically at the bench using a bunsen and incubations would typically occur at 37˚C, with liquid cultures incubated on a shaking platform for aeration. Some cloning procedures deemed to be sensitive would be carried out at either 30˚C or 18˚C to minimise the occurrence of mutations resulting from any toxic effects of a gene's expression. *E. coli* is stored at -80˚C after diluting an overnight rich media culture with a 50% glycerol solution.

2.2 Molecular methods

2.2.1 Analytical PCR

PCRs used to screen for recombinant DNA insertions in either chromosomal or plasmid DNA, or knock out mutations in chromosomal DNA, either *Taq* polymerase (NEB) or LongAmp (NEB) polymerase was used.

Where expected amplicons <5 Kb NEB *Taq* was employed under the following conditions. A 25 μL reaction volume contained 2.5 μL 10× standard *Taq* reaction buffer 0.5μL 10 mM dNTPs (NEB), 0.5 μL 10μM forward primer, 0.5 μL 10μM reverse primer. DNA template was added at 50-200 ng for genomic DNA and 5-20 ng for plasmid DNA in the cases where secondary structure of primers or single stranded DNA (e.g. terminators) were suspected to be interfering with the reaction DMSO was added at 0.5 μL - 5 μL per reaction mix. Nuclease free water (ELGA) was added to a final volume of 25μL. For *Taq* polymerase reactions, thermocycling proceeded as follows: An initial denaturation at 95˚C for 30 seconds followed by 35 cycles of 95˚C for 15 seconds for denaturation, annealing was performed at 45-68˚C depending on primer Tm, where annealing was performed at 3˚C below (annealing temperatures calculated by SnapGene), and held for 30 seconds. Extension was carried out at 68˚C for 1 minute per kb. Cycling was followed by a final extension of 10 minutes at 68˚C.

For amplicons >5 kb NEB LongAmp® was employed. A 25μL reaction contained 5μL 5× LongAmp *Taq* reaction buffer, 0.75μL 10 mM dNTPs, 1μL 10μM forward primer, 1μL 10μM reverse primer. DNA template was added at 50-200 ng for genomic DNA and 5-20 ng for plasmid DNA. 1μL LongAmp *Taq* polymerase was added and nuclease free water (ELGA) to a final volume of 25μL. Reaction thermocycling was as follows. An initial denaturation of 94˚C lasted 30 seconds followed by 35 cycles of 94˚C denaturation for 15 seconds, 45-65˚C annealing for 30 seconds, temperature was dependant on primer Tm (calculated by SnapGene), and was set to 3˚C below primer Tm. Extension time was 50 seconds per kb at 65˚C. Cycling was followed by a final extension of 10 minutes at 65˚C.

Where analysis of correct construct was conducted directly from an *E. coli* colony without plasmid purification (colony PCR), a toothpick was used to add a small amount *E. coli* cells directly to the PCR reaction mix, and the initial denaturation was set to two minutes.

2.2.2: PCR for cloning

PCR fragments for subsequent downstream use in cloning were generated using NEB Phusion® polymerase (NEB M0530). Typically a 50 μL reaction volume contained10 μL 5× Phusion® HF buffer, 1 μL 10 mM dNTPs, 2.5 μL 1 mM forward primer, 2.5 μL 1 mM reverse primer, 50-200 ng of template DNA where genomic DNA was used and 5-10 ng for plasmid DNA. 0.5 μL Phusion® polymerase and nuclease free water (ELGA) to a final volume of 50μL. Reaction thermocycling proceeded as follows. An initial denaturation at 98˚C was followed by 35 cycles of 98˚C denaturation for 10 seconds, 45-72˚C annealing for 30 seconds, temperature was dependant on primer Tm (calculated by SnapGene), and was set to 3˚C above primer Tm, lasting 30 seconds. Extension time was 40 seconds per kb at 72˚C. Thermocycling was followed by a 10 minute final extension time.

In cases where multiple fragments were assembled using overhang extension PCR primers to amplify the fragments were designed to contain overhangs complimentary to the intended adjacent sequences with a Tm of 60˚C. Multiple fragments were purified by agarose gel electrophoresis (see section 2.2.3) and assembled in a PCR reaction with an annealing temperature of 60˚C and a forward and reverse primer priming the terminal 5' and 3' ends of the sequence and containing digest sites to produce a single fragment from the separate components which is subsequently inserted into a vector using restriction cloning (described in section 2.2.4 & 2.2.5). Primer design for overhang extension PCR was facilitated by the SnapGene software overhang extension primer design tool.

2.2.3 Agarose gel electrophoresis

0.8% (w/v) agarose was dissolved in TAE buffer (40mM tris-acetate, 1mM EDTA - pH 8.0) facilitated by heating in a microwave. SYBR Safe DNA stain (Life Technologies - S33102) was added to the molten agarose at 1µL mL^{-1} and cast in an appropriate

casting tray. Samples loaded with the addition of 6× loading dye blue (NEB-B7021S) Log 2 ladder was used as a DNA standard length marker (NEB-N3200L) or alternatively 1 kb ladder was used (NEB-N3232L). Samples ran at 100V for 50 minutes. Images were produced using UV radiation photography using a Biorad Gel Doc XR+.

2.2.4 Restriction digests

Restriction enzymes (FseI, AscI, PmeI, XbaI, XhoI, MauBI, AsiSI, NotI, NheI, SbfI - All sourced from NEB or Thermo Fisher) were used either to generate complimentary overhangs for insertion of PCR fragments (generated according to method described in 2.1.2) into multiple cloning sites, for exchange of modules according to the pMTL8000 vector cloning system, or to build restriction maps for confirmation of correct plasmid assembly. The following protocol was typically followed. A 50 μL reaction consisted of the following components 1μL of restriction enzyme (10 units) 1 μg of DNA, 5μL 10× “NEBuffer”, nuclease free water was added to 50μL and the mixture was incubated at 37˚C. Vectors or inserts were purified using gel electrophoresis using an 0.5%-1% agarose gel

2.2.5 DNA ligation

Ligations were performed on digested DNA for the formation of circular DNA - typically a plasmid. T4 ligase from NEB was used according to the following protocol. a 20 μL reaction consisted of 2 μL NEB DNA ligase buffer. In the case of inserting a single DNA fragment into a larger plasmid vector a 3:1 molar ratio was used. For the ligation of linear DNA for inverse PCR a 10-1 - 10-6 serial dilution was performed on fragmented genomic DNA before adding 10 μL to the reaction mix. Where multiple DNA fragments were assembled into a plasmid construct - equimolar amounts were used. 1 μL T4 ligase was added to the reaction mix after adding nuclease free water to give a 20 μL final volume. Ligations were incubated overnight at 16˚C.

Blunt ended ligations were carried out where PmeI was used as an insertion or to close empty vectors for creation of negative control plasmids. Antarctic phosphatase (NEB M0289S) was used to dephosphorylate vectors to prevent self-ligation. T4 polymerase (NEB M0203S) was used to blunt restriction digested overhangs.

2.2.6 Antibiotics and the appropriate concentrations for use.

Resistance marker	Antibiotic	Solvent	Stock Conc (mg mL^{-1})	Final Conc' *E. coli* (μg mL^{-1})	Final Conc *C. aceto'* (μg mL^{-1})
catP	Chloramphenicol	Ethanol	25	12.5 (broth) 25 (plate)	N/A
catP	Thiamphenicol	DMSO	15	N/A	15
ermB	Erthromycin	Ethanol	50	500	10
neo	Kanamycin	RO Water	50	50	N/A
bla	Ampicillin	RO Water	50	100	N/A
tetR	Tetracycline	Ethanol	50	5	N/A

Table 11: Antibiotics and dosing for selection markers.

2.2.7 Commercial DNA preparative kits

DNA Preperation Method	Manufacturer	Cat Number
Genomic DNA (Single Prep)	Sigma	NA2100
Genomic DNA (96 well block)	Machery-Nagel	740741.2
Mini Prep Plasmid DNA	Jena Bioscience	PP-204L
Midi Prep Plasmid DNA	Qiagen	12843
PCR/DNA Reaction Product	Jena Bioscience	PP-201L
Gel Extraction	Jena Bioscience	PP-202L

Table 12: DNA preparation kits used in this study.

2.2.8: Gibson assembly

Construct and primer design was aided by SnapGene® using the Gibson assembly tool to create primers with an annealing temp of 60˚C and 20 bp overhangs. Fragments for assembly were generated using the cloning PCR method described in 2.1.2. A Gibson master mix was prepared starting with a 5× isothermal reaction buffer containing the following components: 25% PEG-8000, 500 mM Tris-HCl pH 7.5, 50 mM $MgCl_2$, 50 mM DTT, 1mM each of the four dNTPs and 5 mM NAD. A final 2× Gibson reaction mix was made by adding 320 μL, 0.64 μL T5 Exonuclease (NEB M0363S), 20 μL Phusion® polymerase and 160 μL Taq ligase (NEB M0208L), and ELGA water up to 1.2 mL. 2× reaction mix was aliquoted to 15 μL and stored at -20˚C. Reactions were prepared by adding equimolar amounts of fragments aiming for total DNA in the range of 20-200 ng. After fragments were added, the total reaction mix was made to equal 30 μL with ELGA water before incubation at 50˚C for 1 hour (Gibson, 2009). The reaction mix was subsequently added to a 50μL aliquot of *E. coli* 5α (NEB C2987H) chemically competent cells and transformed according to the method described in section 2.2.9.

2.2.9: Preparation and transformation of *E. coli* chemically competent cells

Two strains of *E. coli* were used to amplify plasmid DNA. *E. coli* 5α (NEB) was used for cloning, and for methylation of DNA to be transformed into clostridia, the strain *E. coli* TOP 10 pAN2 was used . This strain carries the pAN2 plasmid (Heap, 2010) bearing the phage Φ3TI methyltransferase gene taken from *Bacillus subtilis* to confer resistance to the *C. acetobutylicum* ATCC824 type II restriction endonuclease Cac824I (Mermelstein and Papoutsakis, 1993) and is maintained by tetracycline resistance in media containing tetracycline at the concentration of 10 μg mL^{-1}.

A loop of -80˚C frozen *E. coli* glycerol stock (final concentration 25 % glycerol) was used to inoculate a 5 mL LB culture which was grown overnight, shaking at

37˚C. This culture was then inoculated into 100 mL LB broth to 0.05 OD_{600nm}, and grown at 37˚C in a shaking incubator until OD_{600nm} = 0.5-0.6. Cells were then chilled on ice, centrifuged at 4K × *g* for 10 minutes in a 4˚C refrigerated centrifuge. The pellet was resuspended in 10 mL ice cold (0.1 M) $MgCl_2$ and incubated on ice for 20 minutes before centrifuging as before, the cells were then resuspended in 5 mL 0.1 M $CaCl_2$ and incubated again for 20 minutes on ice before centrifuging as described before. Cells were then resuspended in 5 mL 0.1 M $CaCl_2$,15% glycerol, aliquoted at 100 μL, and frozen at -20˚C.

The cells were transformed by addition of 0.1-1 ng of plasmid DNA either from plasmid preps, ligations or Gibson assemblies followed by subsequent incubation on ice for 10-30 minutes. Cells were then heat shocked at 42˚C for 45 seconds before transferring to ice for 2 minutes. For recovery, 400 μL of SOC (Invitrogen Cat#: 15544-054) was added and the cell suspension was incubated at 37˚C for 45-90 minutes. Where transformants contained new constructs, cells were plated onto LB + antibiotic selection (see table x). Where cells were transformed for amplification of confirmed plasmid or methylation, cells were transferred to liquid media with appropriate antibiotic.

2.2.10 Synthetic DNA design and synthesis

Synthetic DNA was ordered from either Biomatik or Thermo Scientific Gene Art. Sequences were codon optimised for *Clostridium acetobutylicum*.

2.2.11 Phenol chloroform extraction of genomic DNA

1 mL of overnight culture of *C. acetobutylicum* (see section 2.1.3) was centrifuged at 10K *g* for 2 minutes and the supernatant was discarded. The pellet was resuspended in 180 μL lysis buffer (10 mg ml^{-1} lysozyme [Sigma: L6876] in PBS [Sigma: P4417]), followed by addition of 5 μL RNAse A (Sigma: R4642) and was then incubated for 30 minutes at 37˚C. Then the addition of 25 µl proteinase K, 85 µl sterile dH2O and 110

µl 10% SDS was added followed by mixing by inversion and incubation at 65˚C for 30 minutes. After incubation, 400 μL of ice cold phenol:chloroform:isoamyl alcohol (Sigma: P3803) was added followed by mixing by vigorous inversion for 30 seconds. The resulting emulsion was transferred to a phase lock heavy tube (5Prime/VWR: 733-2478) before centrifuging for 3 minutes at ≥12K × *g*. Following separation of the emulsion, the upper aqueous layer was transferred to a 1.5 mL eppendorf tube containing 40µl of 3M NaAc and 800µl of >99% ethanol which was mixed by gentle inversion 3-4 times before incubation at -80˚C for 30 minutes to precipitate DNA. Centrifugation for 15 minutes at ≥20K × *g* for 15 minutes pelleted the precipitated DNA, and the supernatant was discarded. To wash the pellet, 1 mL of 70% ethanol was added before inverting and centrifuging for 3 minutes as before. The supernatant was again discarded and the remaining DNA was air dried for 45 minutes in a vacuum fitted desiccation jar. The final DNA sample was dissolved in 50 μL sterile ELGA water.

2.2.12 DNA concentration measurement

For typical manipulations DNA was measured using a Thermo Scientific Nano Drop Lite spectrophotometer measuring absorbance at 260 nm. Purity was estimated using an A260/280 ratio calculation with an optimum measurement considered to equal 1.8. For Illumina MiSeq NGS sequencing or where highly accurate measurements were required the Thermo Fisher Scientific Qubit™ fluorometric quantitation system was utilised.

2.3 Genetic engineering and physiological characterisation of *Clostridium acetobutylicum*

2.3.1: Preparation and electroporation transformation of *Clostridium acetobutylicum* for genome editing.

Overnight cultures of *C. acetobutylicum* were prepared from a single colony prepared from a spore stock (see section 2.1.3) suspended from the CGM plate into 1 mL anoxic 2×YTG (liquid media became anoxic by incubating overnight in the anaerobic cabinet) after 1-2 hours incubation, the suspension was used to serially dilute 9 mL YTG down to 10^{-4} which was left to incubate overnight. A visual inspection was used to select a tube with growth in early to mid-exponential phase (OD_{600nm} 0.5-1.5). This tube was then decanted into 100mL (per transformation) 2×YTG pH 5.4 into an appropriately sized conical flask (filled to 60% max volume). The flask was incubated until cells reached OD_{600nm} 0.2-0.25 at which point cells were chilled on ice/water for 10 mins agitating occasionally before harvesting by centrifugation at 4000 × *g* for 10 minutes at 4˚C in anaerobic centrifuge tubes. Manipulations continued on ice - the supernatant was discarded and cell pellets were washed twice by re-suspending in 50 mL ice cold electroporation buffer (EPB: 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 270 mM sucrose) per 100 mL starting culture used and centrifuging as before. The final cell pellet was re-suspended in 600μL EPB per 100 mL starting culture.

Plasmid DNA generated in the *E. coli* Top 10 pAN2 bearing strain was used to transform *C. acetobutylicum.* 600 μL of the cell suspension was transferred to a 4 mm spaced electroporation cuvette on ice containing 5-20 μg plasmid DNA, and incubated for 2 minutes. Then the cells were pulsed using a Biorad Gene Pulser Xcell® programmed with the following method settings: 2000 V, 25 μF capacitance, ∞ resistance, 4mm cuvette. After pulsing cells were then immediately transferred to 4 mL anaerobic 2 × YTG and incubated for 3-4 hours. After incubation 0.1 μL of

benzonase was added to cell suspensions which were then inverted 3-4 times before centrifugation at 5000g for 2 minutes. The supernatant was discarded and cells resuspended in 450 μL 2 × YTG, the suspension was plated onto RCM + antibiotic (either 15 μg mL^{-1} thiamphenicol or 10 μg mL^{-1} erythromycin depending on resistance marker)

2.3.2: Flask fermentations for physiological characterisation

A single colony was resuspended in 1 mL CBM + calcium carbonate and incubated for 1-2 hours incubation, the suspension was used to serially dilute 9 mL down to 10^{-4} which was left to incubate overnight. An OD reading was taken to identify a culture at OD_{600nm} 0.4-0.75. Triplicate 150 mL flasks containing 50 mL CBMS were inoculated to OD_{600nm} 0.05 (±0.01). OD measurements for growth were taken every 90 minutes or 2 hours for the first 10-12 hours and then every 24 hours thereafter until 96-120 hours, depending on the size of the experiment. pH was measured if it was deemed necessary and 2 × 2 mL samples were taken for subsequent offline analytical chemistry analysis. The culture samples were centrifuged at ≥10 k*g* for 2 minutes and the supernatant was stored at -20˚C until needed.

2.3.3: Bioreactor fermentations for physiological characterisation

Fermentations were carried out using the Infors Multifors fermentation system with a 300 mL working volume of CBM. Fermentations were pH controlled at 4.9 ± 0.05, temperature was maintained at 35˚C. Sampling and growth monitoring was performed as described for flask fermentations.

2.3.4 Measurement of extracellular metabolites and glucose by HPLC

300 μL supernatant samples from fermentation broth were diluted using 300 μL mobile phase buffer (0.005 M sulphuric acid) +100mM valeric acid giving a final concentration of 50 mM valeric acid as an internal standard. The samples were clarified by filtering through a 0.2 μm syringe filter before running on a Thermo Scientific Dionex UltiMate 3000 System. An injection volume of 20 μL was ran at a flow rate of 0.5 mL minute-1 and the column temperature was 35˚C. The detection system were refractive index (35˚C) and diode array (UV-VIS 210nm). Where solvents, acids and glucose were resolved the HPLC was fitted with a Bio-Rad Aminex HPX-87H 300 mm × 7.8 mm × 9 μm column with a runtime of 55 minutes. Where glucose data only was needed a Rezex ROA-Organic Acid H^+ (8%) 150 mm × 4.6 mm × 8 μm column was used with a run time of 30 minutes. Standards of mixed 250 mM concentrations of acetic acid butyric acid ethanol butanol and acetone dissolved in ELGA water were diluted 1:1 down to 1 mM. A separate set of glucose standards was constructed according to the same 1:1 dilution regime. All standards were HPLC grade purchased from Sigma

2.3.5 Measurement of extracellular metabolites by GC

Solvents and acids were extracted using addition of 5 μL 10 M sulphuric acid to 500 μL culture supernatant followed by 500μL propyl-propionate. The samples were vortexed for 30 seconds and centrifuged for 30 seconds at ≥10K *g*. 300 μL upper organic layer was removed to a GC vial and sealed with a vinyl GC cap. The samples were analysed on a Thermo Focus GC fitted with a TR-FFAP 30 m × 0.25 mm × 0.25 µm column. Hydrogen was used as a carrier gas with a flow rate of 0.8 mL min^{-1} and the column oven was maintained at 240˚C with a split ratio of 50. A 1 μL sample was injected onto the column which ran for 6 minutes.

2.3.6 Genotyping stability assay

Cells from the triplicate fermenters for wild type control and the *ter* expressing strain were sampled at 48 hours and serially diluted down to 10^{-6} in PBS, 100 μL of each dilution was plated onto CGM, CGM^{Tm} and CGM^{Em}. Plates producing colonies in the range of 30-300 were selected for counting, all of which were produced by the 10^{-3} dilution. CGM only was used as a positive control, and as expected both cultures produced colonies on this media. The CGM^{Em} plates were used to test genotype stability for the *ter* knock in which was inserted along with the *ermB* gene as a selection marker. This resulted in no growth for the wild type strain and the *ter* strain demonstrated similar growth to on the CGM positive control media, indicating that the knock in locus was stable throughout the population. Finally, CGM^{Tm} was used as a negative control, to ensure the experimental results were not due to contamination with a different plasmid harbouring strain with *catP* selection, or due to an error in plasmid curing at an earlier stage, a negative result showed that this was not the case for either scenario.

2.4 Bioinformatics

2.4.1 Software, algorithms and packages used in this study

Software/ Algorithm	Organisation	Webpage URL
BLAST/BLASTX/ BLASTP	NCBI	https://blast.ncbi.nlm.nih.gov/Blast.cgi
ClustalΩ	EMBL	https://www.ebi.ac.uk/Tools/msa/clustalo/
SnapGene	GSL Biotech	https://www.snapgene.com
CLC Genomics Workbench	Qiagen	https://www.qiagenbioinformatics.com/products/clc-genomics-workbench/
ClosTron	Independant	http://clostron.com
Biopython	Independent/Open Source	https://biopython.org/wiki/Documentation
Benchling	Benchling	https://www.benchling.com
PyMOL	Schrödinger	https://pymol.org
TMHMM	DTU	http://www.cbs.dtu.dk/services/TMHMM/
TMPred	ExPASy	https://embnet.vital-it.ch/software/TMPRED_form.html
SWISS-MODEL	EXPASY	https://swissmodel.expasy.org
I-TASSER	Independent/Open Source	https://zhanglab.ccmb.med.umich.edu/I-TASSER/
SymmDock	Independent/Open Source	http://bioinfo3d.cs.tau.ac.il/SymmDock/
DPBind	Independent/Open Source	http://lcg.rit.albany.edu/dp-bind/

Table 13: Software and algorithms used in this study with URLs for access.

2.4.2 Biopython Search Algorithm For Identification Of TcdR Promotor Sequences

```
#script needs a qualifying section which selects promotors only
occurring in intergenic regions with a downstream atg i.e.
correct orientated CDS#

from Bio import SeqIO
from Bio.Seq import Seq

# Use python strings here. No Seq() as above

forseq = "TTTACA" revseq = "NTNNTN"
filename = "/Users/ryanjhope/Documents/PhD/DNA_Sequences/Genome/
Campylobacter_jejuni/NC_002163.1.fasta"

revcomp = {'A':'T', 'C':'G', 'G':'C', 'T':'A',
           'R':'Y', 'Y':'R', 'S':'S', 'W':'W',
           'K':'M', 'M':'K', 'B':'V', 'V':'B',
           'D':'H', 'H':'D', 'N':'N'}
redund = {'A':['A'], 'C':['C'], 'G':['G'], 'T':['T'],
          'R':['A','G'], 'Y':['C','T'],
          'S':['C','G'], 'W':['A','T'],
          'K':['G','T'], 'M':['A','C'],
          'B':['C','G','T'], 'V':['A','C','G'],
          'D':['A','G','T'], 'H':['A','C','T'],
          'N':['A','C','G','T']}

def revcom(seq):
    return "".join(revcomp.get(base, base) for base in
reversed(seq))

def seqmatch(query, reference):
    """Checks if the query sequence (perhaps a promoter with
redundant bases)
    matches the reference. Both query and reference must be the
same length.

    Returns true if there is a match (redundancy allowed),
returns false
    otherwise.
    """
    assert len(query) == len(reference)
    for i in range(len(query)):
        if reference[i] in redund[query[i]]:
            pass
        else:
            return False
    return True

forseq = forseq.upper()
revseq = revseq.upper()
forcom = revcom(forseq)
revcom = revcom(revseq)
print("seq\tstart\tstop\tgap\tdir\tseq")
```

```
for record in SeqIO.parse(filename, "fasta"):
    for i in range(len(record.seq)):
        for v1, v2, direc in [ (forseq, revseq, "F"), (revcom,
forcom, "R") ]:
            if seqmatch( v1, record.seq[i:i+len(v1)] ):
                for j in range(17,18):
                    if seqmatch( v2,
record.seq[i+len(v1)+j:i+len(v2)+j+len(v2)] ):
                        print("{}\t{}\t{}\t{}\t{}\t{}".format(
                            record.id, i,
                            i + len(v1)+j+len(v2)-1,
                            j, direc,
                            record.seq[i: i+len(v1) + j +
len(v2)] ))
```

This script was written in collaboration with Darrin T. Schultz, a PhD student at University of California - Santa Cruz.

Chapter 3: The Use of Established and Novel Molecular Methods to Obtain a Δ*bcd* Mutant in *Clostridium acetobutylicum*

3.1 Introduction

3.1.1 Electron bifurcation and energy conservation in the firmicutes

Flavin based electron bifurcation (FBEB) constitutes as a newly discovered, third mode of energy conservation in nature, in addition to substrate level phosphorylation (SLP) and electron transport phosphorylation (ETP), and is likely the most evolutionary ancient of the three (Schuchmann & Mueller, 2014). The first example of this phenomenon was demonstrated by Li *et al.,* (2008) who determined that the activity of butyryl-CoA dehydrogenase from cell extracts of *C. kluyveri* was ferredoxin dependant, yielding hydrogen gas as a by product of the reaction. Previously, it had been assumed that *C. kluyveri* cell extracts produced hydrogen using acetyl-CoA as a substrate and NADH as an electron donor in a ferredoxin dependant manner. However, the reduction of H^+ to H_2 using NADH as an electron donor is thermodynamically unfavourable (Li *et al.,* 2008):

$$NADH + H^+ \rightarrow H_2 + NAD^+ \ (\Delta G^{\circ\prime} = 20 \text{ kJ mol}^{-1})$$

Herrmann *et al.,* (2008) initially proposed that this problem could be overcome using the following hypothesis: Electron transfer flavoprotein butyryl-CoA dehydrogenase complexes from anaerobic bacteria are able to bifurcate two electrons donated by NADH (E_0' -320 mV), the first is shuttled up to a more positive electron acceptor, likely an FAD coenzyme of butyryl-CoA dehydrogenase. The second electron is transported to the low potential acceptor, ferredoxin (E_0' ≤-420 mV). A second round of bifurcation results in

the reduction of crotonyl-CoA to butyryl-CoA and the complete reduction of ferredoxin (Fd^{2-}). Given that FAD can exist stably in either one of three states, quinone (oxidised), semi-quinone (reduced), or hydroquinone (fully reduced) it would serve as the ideal prosthetic group. Therefore, the exergonic reduction of high potential crotonyl-CoA is able to power the endergonic reduction of low potential ferredoxin with equivalents from higher potential NADH:

$$2\ NADH + crotonyl\text{-}CoA + Fd \leftrightarrows 2\ NAD^{+} + butyryl\text{-}CoA + Fd^{2-}$$

$$(G^{\circ\prime}\ \text{-}40\ kJ\ mol^{-1})$$

The advantage of this system in butyrate producing *Clostridium spp* is that the reduction of ferredoxin and subsequent production of H_2 via an [Fe-Fe] hydrogenase A (HydA) allows for regeneration of the NAD^{+} redox cofactor, affording a higher yield of acetate and ATP by SLP. The acetate synthesis pathway does not consume any of the NADH generated by oxidation of glucose in the glycolysis pathway (Jones & Woods, 1986), however, the acetate pathway produces 2 moles of ATP and 2 moles of acetate per mole of glucose and is therefore more efficient than the butyrate producing pathway, which only produces 1 mole of butyrate and 1 mole of ATP per mole of glucose. Without hydrogen production via reduced ferredoxin, production of acetate would result in a redox cofactor imbalance, rendering acetate production inviable, and the cells would depend entirely on the less efficient butyrate pathway to maintain redox balance. The electron flux towards the production of H_2 from reduction equivalents donated by NADH permits the production of acetate (Herrmann *et al.,* 2008):

$$\text{Glucose} + 0.6\ H_2O \rightarrow 2\ CO_2 + 0.6\ \text{acetate} + 0.7\ \text{butyrate} + 1.3\ H^+ + 2.6\ H_2$$

$$\Delta G^{\circ\prime} = -251\ \text{kJ mol}^{-1};$$

$$3.3\ \text{ATP mol glucose}^{-1};\ 251/3.3 = -76\ \text{kJ mol}^{-1}\ \text{ATP}^{-1}$$

Hydrogenase is well adapted to this, under standard conditions (pH 7, 25˚C) the redox potential of the $H_2/2H^+$ couple is $E_0' = -420$ mV, which is about the same as ferredoxin at $E_0' = -410$ mV. This means that the reduction of hydrogen ions can still proceed even in conditions of high H_2 concentration (Gottschalk, 1986).

Electron bifurcation was first demonstrated in *C. kluyveri,* and the BCD/EtfAB system represents one the best studied flavin based bifurcation systems (FBEB) (Herrmann *et al.,* 2008; Li *et al.,* 2008; Buckel and Thauer, 2013; Aboulnaga *et al.,* 2013; Demmer *et al.,* 2017), however, since its discovery many more FBEB systems have been characterised - summarised in section 3.1.2 - (Weghoff *et al.,* 2015; Bertsch *et al.,* 2013; Demmer *et al.,* 2018; Kaster *et al.,* 2011; Huang *et al.,* 2012; Wang *et al.,* 2010; Wang *et al.,* 2013; Schuchmann *et al.,* 2012).

SDS-PAGE and size exclusion chromatography allowed the determination of the quaternary structure of the FBEB complexes, Li *et al.,* (2008) and found that the BCD/Etf complex contained 4 α BCD subunits 2 β EtfB subunits and 2 γ EtfC subunits. The purified complex was found to have activity for crotonyl-CoA in the presence of ferredoxin and NADH which led to the proposition of new a functional model by Buckel and Thauer, (2013).

Buckel and Thauer, (2013) gave the following description: a hydride from NADH ($E_0' = -320$ mV) reduces the $E_0' = -280$ mV γ-FAD to γ-$FADH^-$ which

transfers one electron to the higher potential β-FAD of the β subunit, and is subsequently shuttled to the higher potential α-FAD of BCD to produce a stable α-FADH• low potential (E'~180 mV) semiquinone. Upon transfer, the γ subunit is left with a highly reactive unpaired electron γ-FADH• (<500 mV) which is able to reduce oxidised ferredoxin with a reduction potential of E_0' = -400 mV (Thamer *et al.,* 2003). A second round of reduction starting again with the transfer of electrons from NADH to the γ-FAD allows for the complete reduction of ferredoxin resulting in an E_0' = -500 mV ferredoxin carrying two electrons, after the first electron is shuttled down to the α-FADH• forming α-$FADH^-$ which then goes on to reduce the high potential unsaturated β-carbon bond of crotonyl-CoA, E_0' = -10 mV. The midpoint potentials of free flavins in water have been determined to be -172 mV for E_1 and -238 mV for E_2, however, the reduction potentials of flavins are understood to be significantly modulated by their protein environment (Zhou & Swenson, 1996). It is therefore, feasible that the γ, α and β subunits are arranged so that reduction potentials of their flavin ligands are complimentary to their required function, allowing the flux of electrons from NADH up to the high potential crotonyl-CoA while also allowing for the bifurcation of elections to the low potential oxidised ferredoxin in this energy conserving manner.

Further work has led to greater insight into this newly discovered energy conserving system. Work by Chowdhury *et al.,* (2016); Chowdhury *et al.,* (2015); Chowdhury *et al.,* (2014); and Demmer *et al.,* (2017) allowed for the elucidation the mechanism of the BCD/EtfAB complex, leading to an updated model described by Demmer *et al.,* (2017). The research allowed the postulation of the molecular mechanics of the bifurcation system using structural protein crystallographic data obtained from the *Clostridium difficile*

BCD/EtfA/EtfB system, this was compared to earlier structural work on the EtfA/EtfB proteins from the butyrate producer *Acidaminococcus fermentans* (Chowdhury *et al.,* 2014).

The structural data contrasted the observations of Li *et al.,* (2008) concerning the quaternary structure of the *C. kluyveri* system, and shows a central BCD tetramer surrounded by 4 EtfA/EtfB complexes. The EtfA/EtfB complex consists of 3 functional domains. Domain I is the NADH binding domain, and represents the entire peptide chain of EtfB. EtfA is made up of two domains; domain II is the mobile domain containing the α-FAD, responsible for the transfer of a single electron from β-FAD and domain III is the fixed region of EtfA anchoring the protein to the EtfB (see figure 6) (Demmer *et al*., 2017).

Interpretation of the structural studies determined that the bifurcating activity was a result of the mobile α-FAD bound to domain II of EtfA. The crystal structure of *C. difficile* BCD/EtfA/B showed an 8 Å distance between α-FAD and δ-FAD of BCD. Comparing to *A. fermentans* EtfAB by superimposition of the protein structures showed that the α-FAD domain of EtfA and the β-FAD domain of EtfB were 14 Å apart. This led the authors to infer that the bifurcation was a result of a 80˚ swing of domain II after β-$FADH^-$ hydroquinone formation, following reduction from a hydride received from NADH, the mobile domain II α-FAD, the default state of which, the authors propose, is a stable semiquinone $FAD^{\bullet -}$, receives an electron from the β-$FADH^-$ forming an α-FADH• hydroquinone, which then swings 80˚ to donate the electron to the δ-FAD of BCD. The resulting unstable low potential β-FADH• is able to reduce the low potential ferredoxin. Repetition of this

process results in the formation of 1 mole of crotonyl-CoA and 1 mole of reduced ferredoxin from 2 moles of NADH.

Figure 6: The crystal structure of the *C. difficile* Bcd/EtfAB complex. EtfB NADH binding Domain I shown in blue (NADH not shown), Domains II and III of EtfA are shown in magenta and green, respectively. Bcd is shown in orange. FAD ligands are coloured yellow for each respective polypeptide. Coenzyme-A persulfide, at the substrate binding site in substitute of crotonyl-CoA is shown in red. The image shows the C. difficile Bcd/EtfAB complex was crystallised in the dehydrogenase conducting "D-like state" with the α-FAD of domain II 14 Å distance away from the δ-FAD of Bcd. The red arrow indicates the turn made by the mobile domain II which allows the complex to switch back to the FBEB conducting "B-like state" where the α-FAD domain of EtfA and the β-FAD domain of EtfB were 14 Å apart, allowing for the reduction of the α-FAD, after hydride transfer to the β-FAD domain. Adapted from PDB file 5OL2 using PyMOL.

3.1.2 Other energy conserving election distribution systems

Several other energy conserving bifurcating enzyme systems have also been elucidated (table 14), and an additional “confurcating” system, where two separate electron donors contribute exergonically and endergonically to reduce a single acceptor.

Weghoff *et al.,* (2014) identified a novel mode of lactate metabolism in *A. woodii*, which used a bifurcating system for conversion of lactate to pyruvate. The system comprises of a lactate dehydrogenase (Ldh) and an electron transfer flavoprotein (Etf). The high redox potential pair of lactate and

Enzyme system	Electron donor	Substrate	Product	Reference
Lactate dehydrogenase	Lactate/Fd^{2-}	Lactate/Fd^{2-}/ NAD^+	Pyruvate/ NADH/Fd^{ox}	Weghoff *et al*, (2014).
Caffeyl-CoA reductase	NADH	Caffeyl-CoA	Fd^{2-}, Hydrocaffeyl-CoA	Bertsch *et al*, (2012). Demmer *et al*, (2018).
Methyltransferase	H_2	CO_2/Fd^{ox}/ Heterodisulfide CoM-S-S-CoB	Fd^2/ Formylmethanofuran/ Coenzyme M (CoM-SH)/ Coenzyme B $(CoB\text{-}SH)^-$	Kaster *et al*, (2010)
$NADP^+$ oxidoreductase	NADH/Fd^{2-}	$NADP^+$	NADPH	Huang *et al*, (2012). Wang *et al*, (2010).
Formate dehydrogenase	Formate	Formate/Fd^{ox}/ NAD^+	CO_2/Fd^{2-}/ NADH/H^+	Wang *et al*, (2013)
[FeFe]-hydrogenase	H_2	H_2/Fd^{ox}/NAD^+	Fd^{2-}/NADH/H^+	Schuchmann *et al* (2012)

Table 14: Summary table detailing examples of electron bifurcating energy conserving systems.

pyruvate E_0' = -190 mV cannot directly use NAD^+ E_0' = -320 mV as an electron acceptor in the oxidation of lactate. Instead reduced ferredoxin is used in an exergonic reaction as an electron donor (E_0' ≈ -500 mV) facilitated by flavin based electron confurcation, receiving electrons from the second donor source, lactate to drive the endergonic reduction of NAD^+.

Bertsch *et al.,* (2013) and Demmer *et al.,* (2018) have identified and solved the structure and molecular mechanism of an electron bifurcating caffeyl-CoA reductase from *A. woodii*. Similar to the BCD system from butyrate producing organisms, caffeyl-CoA reductase works in complex (CarCDE) with two electron transfer flavoproteins (CarD and CarE), which all contain an FAD prothetic group, to reduce a metabolic intermediate and ferredoxin. Additionally, the CarCDE complex also contains a ferredoxin like domain with two [4Fe-4S] clusters, N-terminally fused to CarE.

The methanogenoic archea, *Methanothermobacter marburgensis* couples reduction of ferredoxin and a heterodisulfide in an electron bifurcating system (Kaster *et al.,* 2011). The first step in methanogenesis involves the endergonic reduction of CO_2 to the organic carbon compound formylmethanofuran using ferredoxin as an electron donor. The generation of reduced ferredoxin is a result of the bifurcating activity of the MvhADG/HdrABC complex, which bifurcates electrons from H_2 serving as the donor to endergonic reduction of ferredoxin, coupled with the exergonic reduction of heterodisufide CoM-S-S-CoB. The second recipient molecule in the bifurcation reaction is methylated coenzyme M, which reduces to free coenzyme M and methane as a metabolic end product.

In addition to the bifurcating activity of BCD, a second flavin based bifurcating system operating in *C. kluyveri* was subsequently reported by

Wang *et al.,* in 2010. Again an enzyme complexed with flavin based electron transport proteins, also containing iron-sulphur clusters, was responsible for the exergonic reduction of $NADP^+$, receiving electrons from reduced ferredoxin and coupled with an endergonic reduction of $NADP^+$ using NADH as the donor molecule. A homologous system has also been demonstrated in *Moorella thermoacetica* (Huang *et al.,* 2012) and the protein sequences for NfnAB have been found in many *Clostridium spp.* genomes with the notable exception of *C. acetobutylicum.*

3.1.3 Forward genetics for characterisation of the central carbon pathway in *Clostridium*

As discussed in section 1.5.4, molecular tools for investigation of *Clostridium* have advanced significantly in the last decade opening up the opportunity to comprehensively characterise the catabolic pathway of the model organism *C. acetobutylicum*. Cooksley *et al*., (2012) obtained a range of strains displaying altered physiology using ClosTron (Heap *et al*., 2010) gene disruption introns, however, there was difficulty in obtaining a strain with dysfunctional *thlA* (thiolase) and *ptb* (phosphotransbutyrylase), both representing disruptions to the C4 product pathway (butyrate or butanol).

Previously, however, Lehmann & Lütke-Eversloh (2011) had successfully obtained an intron insertion in *hbd* (3-hydroxybutyrate-CoA dehydrogenase) using ClosTron resulting in a strain producing only acetate, acetone and ethanol, completely absent of C4 product.

The only two remaining genes from the ABE pathway that had not had their function confirmed using forward genetic techniques were *crt* and *bcd.* In principle, given that a *hbd* knockout was obtainable, a knockout of either *crt* or *bcd* should also be achievable. Of particular interest is the *bcd* knock out,

since *C. acetobutylicum* possesses a gene: *ter* coding for the enzyme trans-2-enoyl-CoA dehydrogenase (TER) (Hu *et al.,* 2013) which catalysis the conversion of crotonyl-CoA to butyryl-CoA but in the typical non-bifurcating, ferredoxin independent manner, consuming 1 M of NADH per mole of butyryl-CoA produced. Normally this enzyme functions as part of the fatty acid synthesis pathway, however, it is of interest to observe the effects TER activity might have in the absence of *bcd,* given that electron flux is not split between ferredoxin and hydrogen production and reduction of the crotonyl-CoA - thus representing a conservation of reducing equivalents needed for butanol and ethanol production.

More recently homologous recombination knockout systems have been developed for use in *Clostridium* which allows the study of genes arranged in operons, as are *crt*, *bcd*, *etfA*, *etfB*, and *hbd* without the interference with of downstream polar effects. The knockouts produced by homologous recombination produce in-frame deletions which alleviate the necessity for selection markers (figure 7).

3.1.3 Aim of the study

This study aims to establish whether *bcd*, as with *hbd* may be disrupted using ClosTron technology. An in-frame deletion of *bcd* is also sought in order to demonstrate either the ability of native *ter* to compensate for the absence of *bcd* (figure 7) or otherwise the apparent necessity of *bcd*. In the case of obtaining a Δ*bcd* strain phenotype characterisation will be performed with the expectation to observe either an increase in butanol or butyrate in the case of active TER where surplus NADH cofactor results in increased productivity of reduced product or otherwise the abolishment of C4 products, in the case of insufficiency or inactive TER. Otherwise a phenotype representing the continuum between these two possibilities might be demonstrated.

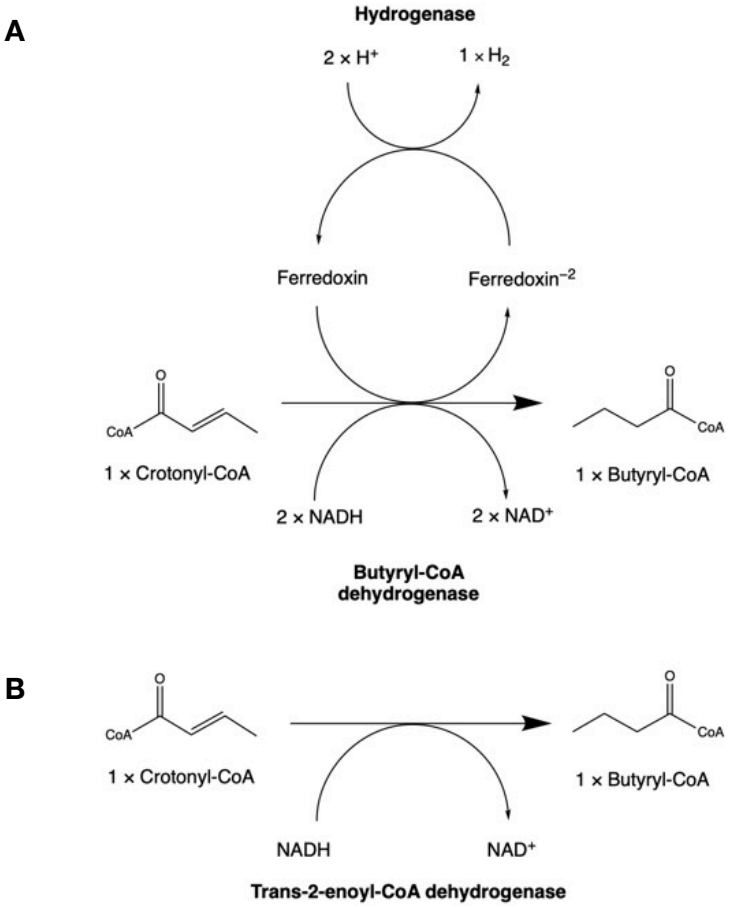


Figure 7: Comparison of the reactions carried out by BCD (A) vs TER (B).

3.3 Results and discussion

3.3.1 Preliminary physiological characterisation of a *Clostridium acetobutylicum* Δ*bcd* Δ*pyrE* strain and the subsequent attempt to restore uracil prototrophy

A novel mutant Δ*bcd* strain was generated using an adapted version of the allele exchange system described by Ng *et al.,* (2013) (figure 8) by Dr. Krzysztof Gizynski and Dr. Ying Zhang, where the Gram positive replicon was removed so that transformants would be selected for as single cross over integrants containing a chromosomal copy of *catP*. An initial solvent fermentation assay was performed in the Δ*pyrE* background strain using CBMS + 50 mg mL^{-1} uracil to provide a preliminary picture of the strain's phenotype. Genomic DNA extractions were performed on the fermentation broths at 72 hours and on the culture inoculum confirming that the strain produced a 2 kb PCR product consistent with a double cross-over in-frame deletion of *bcd*, demonstrating that this genotype was stable during the fermentation (figure 9). The *ΔpyrE* Δ*bcd* strain's physiology was measured in CBM minimal media and samples were taken over 100 hours to measure growth and end metabolites, a *ΔpyrE* and an ATCC 824 strain were also ran as controls in uracil supplemented CBM which were comparable. The ABE solvents and acids were measured using GC and some interesting differences in the phenotype are noted (figure 10). The fermentation profile in the *ΔpyrE* Δ*bcd* strain did not show the abolition of C4 products, however, altered carbon flux was observed with a reduction in butyrate productivity during the acidogenic phase and increased acetate, acetone and ethanol productivity, with acetate maximum yield increasing by

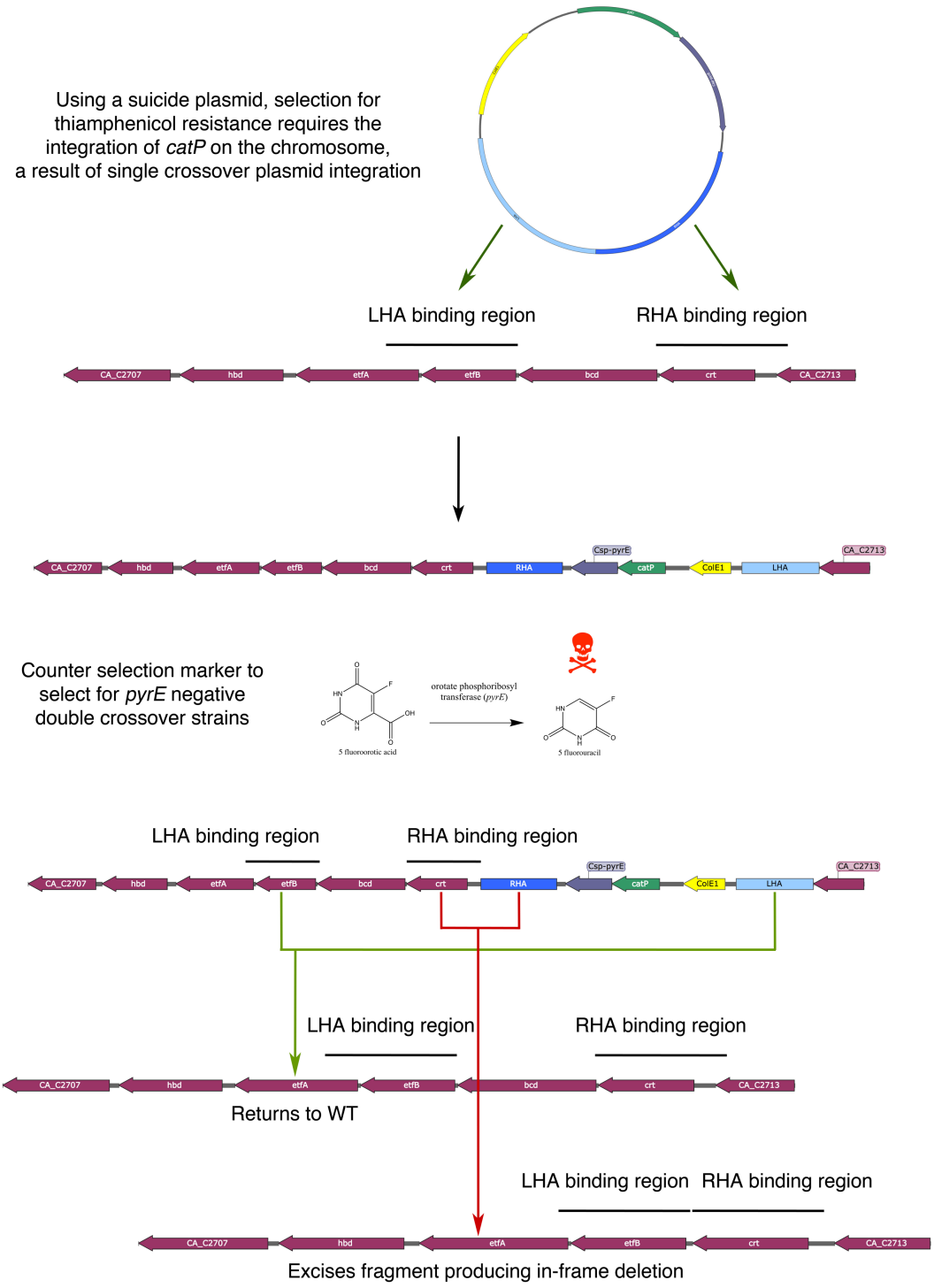


Figure 8: The general scheme for producing an in-frame deletion using the *pyrE* counter selection marker. Homologous recombination between the left or right regions of homology results in cointegrates of the plasmid and chromosome. Addition of 5-fluoroorotic acid to growth media results in counter selection for double cross-over strains due to PyrE activity generating the toxic product 5-fluorouracil, either giving rise to an in-frame deletion, or the reversion back to the wild-type genotype.

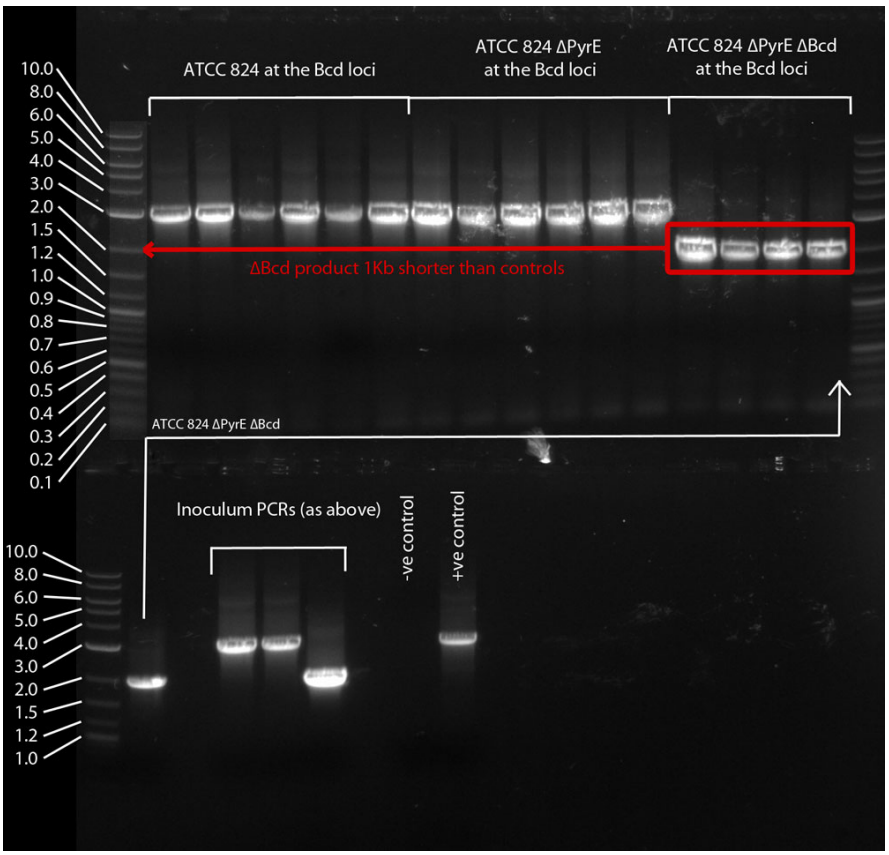


Figure 9: PCR results generated using primers flanking the 1kb *bcd* gene 1kb upstream and 1kb downstream at the chromosomal locus. Duplicate samples taken from fermentation flasks at 72 hours. Samples taken from inoculum represent) hours time point. Indicates strains stable throughout physiology assay. In-frame deletions at the chromosomal locus generate a 2 kB PCR product, control strains generate a 3 kb PCR product. +ve PCR control uses a previously prepared genomic DNA sample.

<10 mM and acetone final yield increasing, more significantly by 60 mM. This was not, however, reflected in the production of butanol, where productivity and total yield were comparable to wild type.

Judging by these fermentation results it seems apparent that TER activity is indeed compensating for the deletion of *bcd* given that C4 product was made but impaired flux through the C4 pathway resulting in diversion of intermediates through the C2 and C3 pathways causing the increase in production of acetate, ethanol, and acetone. Given that hydrogenase activity

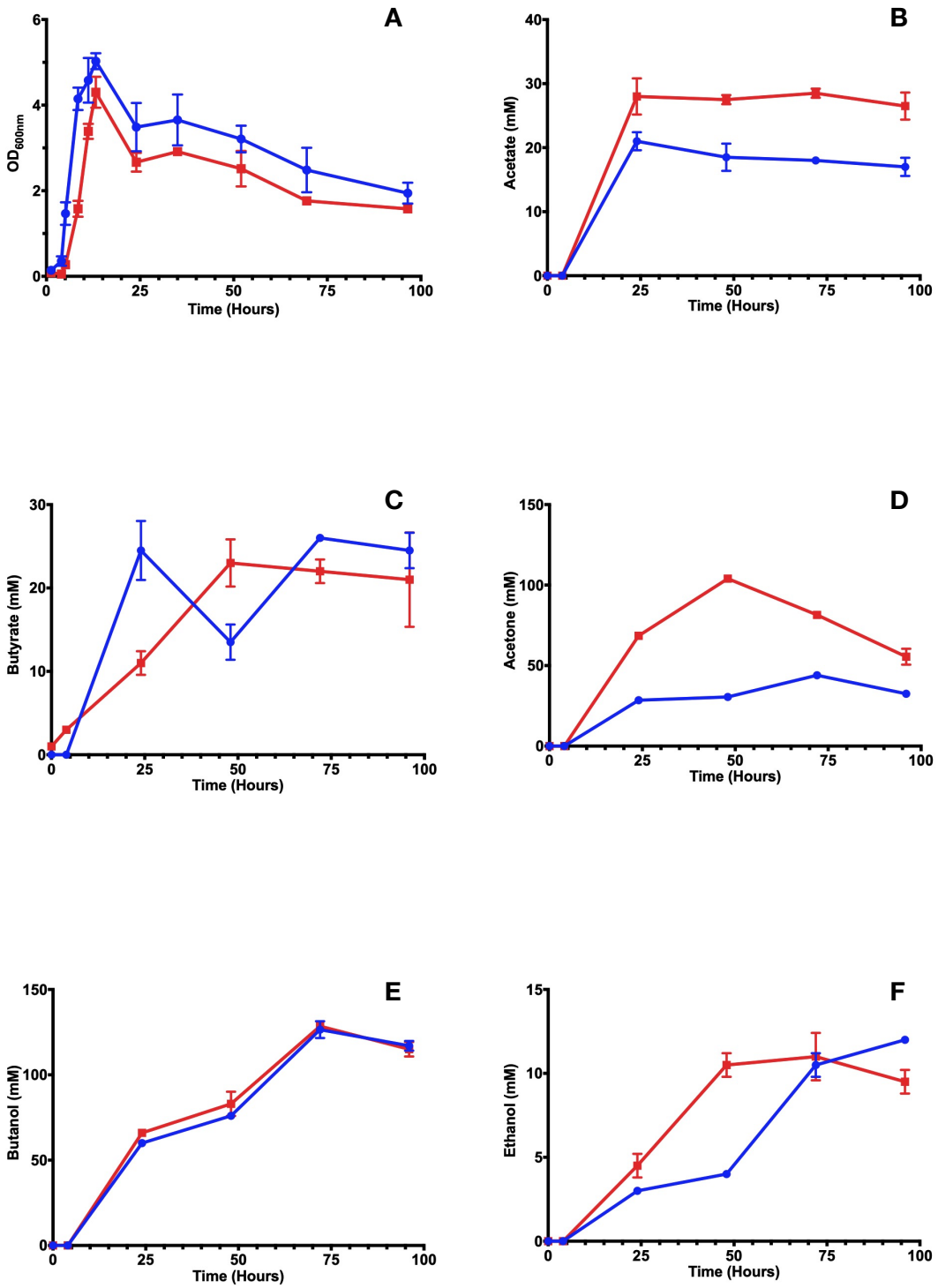


Figure 10: Physiological characterisation of end metabolites produced by the putative Δ*bcd* mutant in the Δ*pyrE* background compared to *C. acetobutylicum* ATCC 824 Δ*pyrE* Both fermentations were supplemented with 20 mg L-1 uracil. *C. acetobutylicum* ATCC 824 Δ*pyrE* control ●; the *C. acetobutylicum* ATCC 824 Δ*bcd* Δ*pyrE* strain ▪. Error bars to indicate standard deviation where *n* = 3. Replicates represent a single strain grown as three independent cultures.

is thought to permit acetate production during acidogenesis by the balanced recycling of NADH, this ought to be unexpected. Furthermore, increased acetone was observed, again, this is not an NADH consuming pathway and therefore a reduction in acetone ought to be expected during solventogenesis with an increase in ethanol in order to compensate for the NADH imbalance.

It is interesting to note that there is an apparent reduced rate of butyrate synthesis while observing no significant difference in butanol production. Given that acetone production is CTFA/B dependant the apparent reduced productivity in butyrate synthesis might be better understood as early re-assimilation contributing to the increase in acetone production. However, without a change in butanol production this would suggest that butyrate synthesis was cycling through the butyryl-CoA node via re-assimilation permitting the excess flux of carbon to acetone. The greater measured standard deviation in this physiological parameter relative to butanol might be owed to the fact that there is a greater amount of metabolic activity in the butyrate pathway, with the direct synthesis pathway and assimilation pathway active simultaneously, although, the increased measured standard deviation might also be owed to the volatility of this compound.

Additionally, there is an overall increase in products with an additional 75 mM of acetone and approximately 6 mM of ethanol with no relative drop in butanol. The only drop in product appears to be biomass where lower OD was measured throughout the fermentation. This may possibly be explained by substrate consumption rates, which were not measured in this fermentation.

Having generated interesting preliminary phenotype data in the Δ*pyrE* background strain the next step was to repair Δ*pyrE,* recovering uracil

prototrophy so that the mutant strain could be characterised against a wild type control, without concern for possible influence of the Δ*pyrE* background. The repair of Δ*pyrE* proceeded with a transformation requiring the plasmid pMTL-ME6 and transformants selected for using the *catP* selection marker on CGM + 15 μg mL^{-1} thiamphenicol (CGM^{Tm}).

Transformation plates had >300 colonies, 8 of which were subcultured onto fresh CBM/CBM^{Tm} to select for single cross over uracil prototrophy. The cells were passaged on both selective and non-selective minimal media until evidence for repaired *pyrE* double cross-over phenotypes was demonstrated by uracil prototrophy; plasmid curing was demonstrated by thiamphenicol sensitivity. After 7 passages, there was no indication that the pMTL-ME6 plasmid had been cured, which was unexpected for a plasmid maintained by the unstable pIM13 replicon. A PCR fragment of the *pyrE* locus was sent for sequencing and the results showed that at the *pyrE* locus the truncated/ disrupted *pyrE* genotype was found which would normally result in uracil auxotrophy. This was suspicious as the strains were demonstrating uracil prototrophy. At this point it was important to check that the initial in-frame knock out plasmid pMTL8-150Δ*bcd* had been properly cured from the strain, as it carried the *C. sporogenes pyrE* gene used as a counter selection marker, which would confer uracil prototrophy, should it still be present. A -80˚C cryogenic culture of the *C. acetobutylicum* Δ*pyrE* Δ*bcd*, stored by Dr. Gizynski was plated directly onto CBM^{Tm}, and the strain exhibited uracil prototrophy and thiamphenicol resistance (see figure 11).

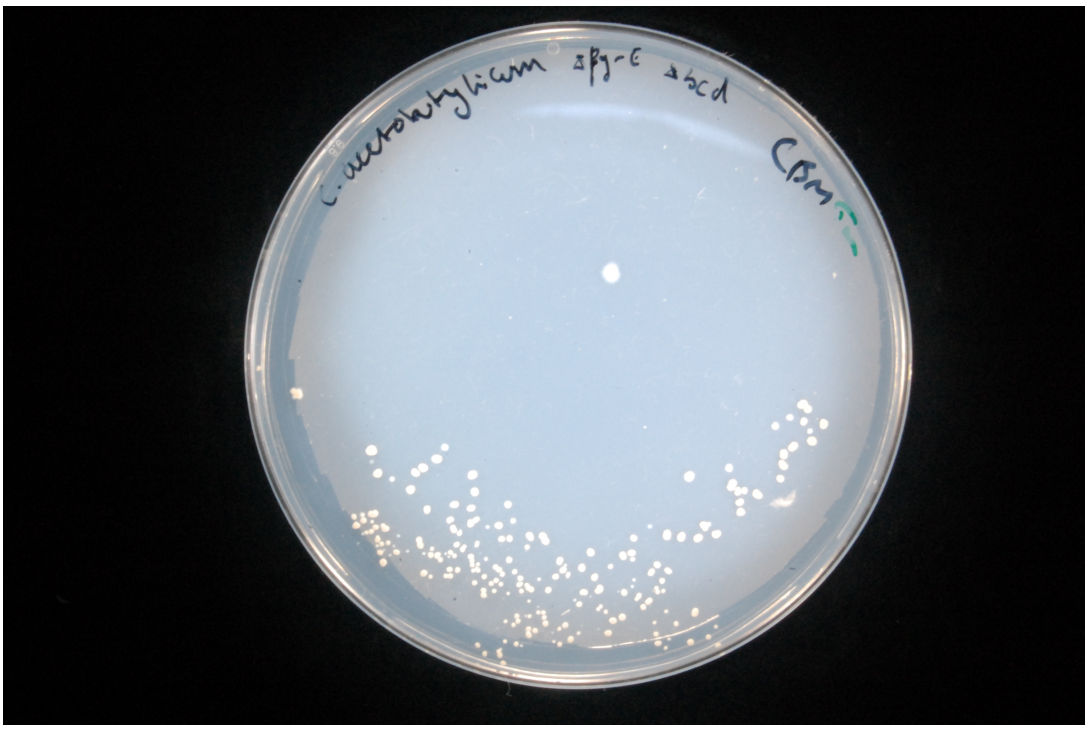


Figure 11: Streaking the *C. acetobutylicum* Δ*pyrE* Δ*bcd* strain from the culture collection onto minimal CBM media + 15 μg mL^{-1} thiamphenicol (lacking a uracil source) demonstrated uracil prototrophy and thiamphenicol resistance suggesting that the original in-frame deletion construct had not been effectively cured from the strain.

3.3.2 Plasmid curing of the Δ*bcd* mutant strain

Given the evidence suggesting that the original pMTL8-150Δ*bcd* knockout plasmid was still present in the mutant strain, PCR was used to confirm this on a screen of 18 colonies. The culture collection at the SBRC (this lab) consists of three vials of each strain and a colour coded system. Green represents the working stock, yellow is the stock for generating new working stocks, and red is a back up emergency stock. A streak plate was prepared from each stock and six colonies were picked from each - the basis being that for this strain the yellow and red stocks should be untouched and therefore the possibility that the vials had become contaminated can be ruled out.

Genomic DNA was prepared from overnight cultures grown in 2×YTG and three PCRs were used to genotype the strain. Firstly, the Δ*bcd* in-frame deletion was confirmed for all of the 18 colonies (figure 12). Primers targeting the *C. sporogenes pyrE* gene, present only on the plasmid DNA, successfully generated a correctly sized PCR product for the gene (figure 13), which was then confirmed by sequencing. Once it was evident that the pMTL8-150Δ*bcd* plasmid was still present, an attempt to cure the plasmid was made by passaging on CBM + 400 μg mL^{-1} 5-fluoroorotic acid (CBM^{FOA}), and plasmid loss was monitored using separate CGM^{Tm}. After 6 passages, 8 of the strains demonstrated thiamphenicol sensitivity, the remaining 10 still showed resistance. Genomic DNA was prepared again from each of the passaged strains and the PCRs targeting the *bcd* genomic locus and the *C. sporogenes pyrE* gene were repeated (see figures 14 & 15). An interesting pattern had emerged as a result of the passaging on counter selective media. In the strains which had shown thiamphenicol sensitivity, indicative of plasmid loss, the *bcd* locus PCRs using 1 kb flanking primers produced a 3.1 kb fragment suggesting the strains had reverted back to wild type and indeed sequencing confirmed that the *bcd* CDS had returned to its original position - suggesting it had always been present elsewhere in the strain's chromosome or pSOL megaplasmid. *C. sporogenes pyrE* PCRs also produced a negative result as would be expected in a strain that had lost the plasmid. Conversely, strains that produced a positive result for *C. sporogenes pyrE* PCRs continued to produce a 2 kb fragment in PCRs which flanked the *bcd* locus.

A third PCR was employed using internally binding primers targeted to *bcd*, the results showed that in all of the passaged strains *bcd* was present (figure 16).

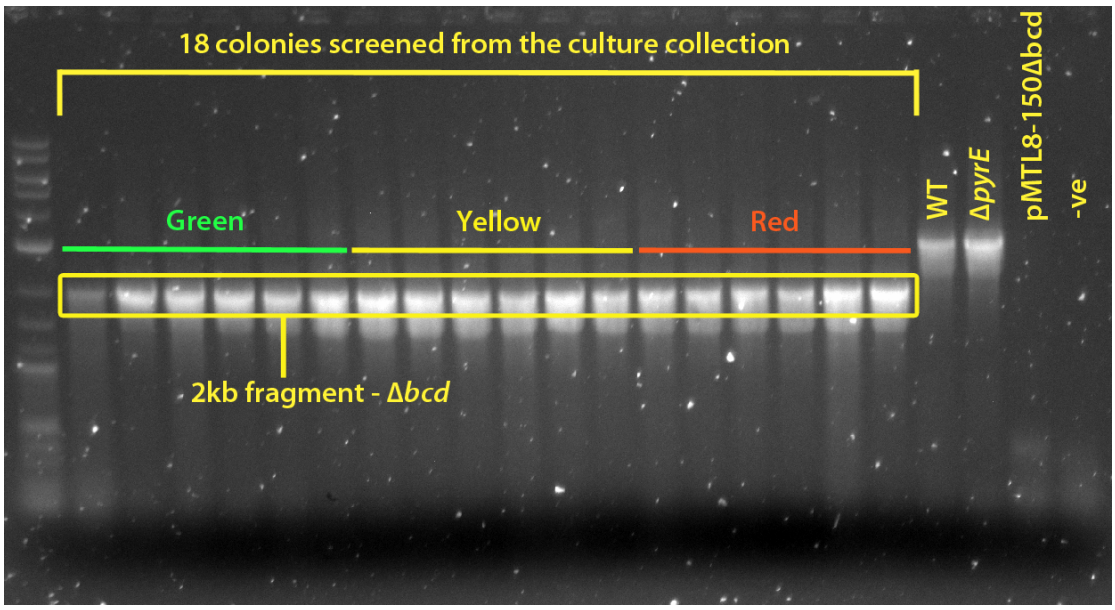


Figure 12: Genomic DNA from 18 colonies were analysed by PCR using primers flanking 1 kb upstream and downstream of the *bcd* locus. The 2 kb PCR product generated from all of the colonies suggests that the cultures stored are not a mixture of two strains containing the mutant and wild type.

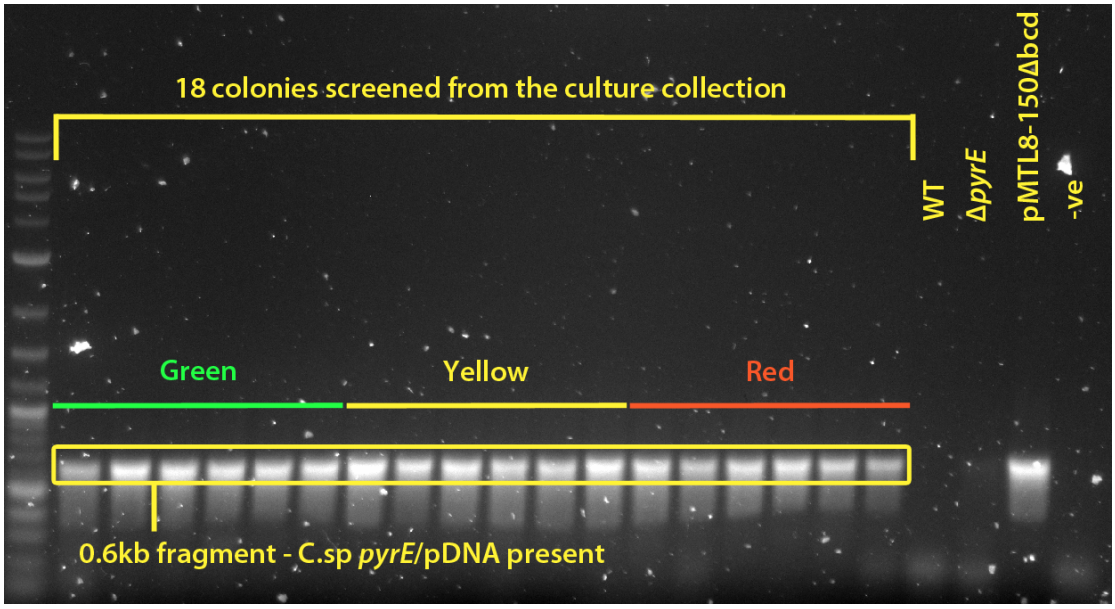


Figure 13: Each of the genomic DNA extracts also tested positive for the *Clostridium sporogenes pyrE* gene suggesting that the plasmid DNA of pMTL8-150 was present in the strain stored in the culture collection.

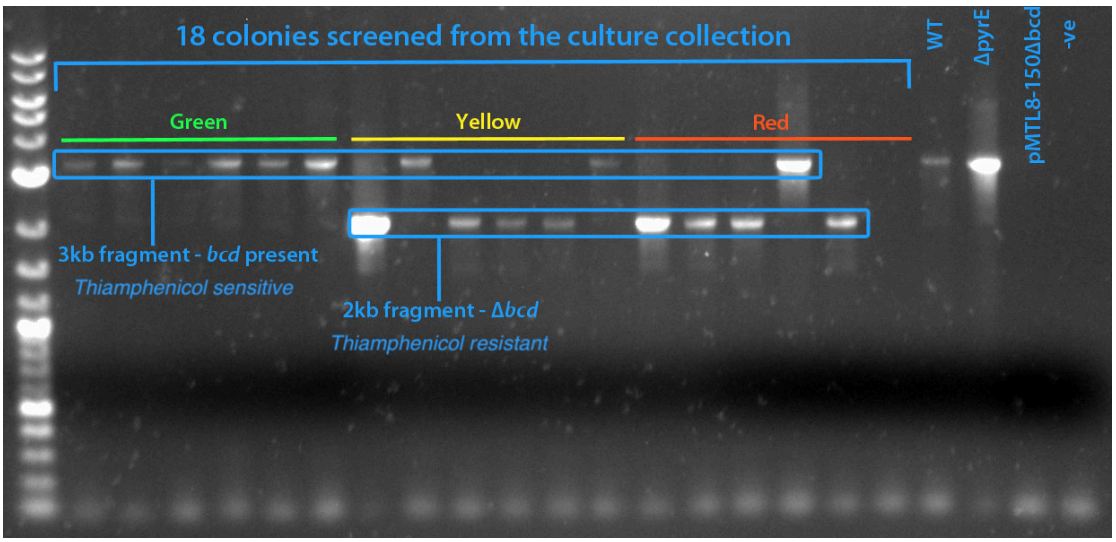


Figure 14: After passaging the strains on CBM^{FOA} to select for Δ*pyrE* strains. For 10 of the strains PCR products of the *bcd* chromosomal locus matched the size of the fragments produced by the wild type, the remaining 8 produced a 2 kb fragment.

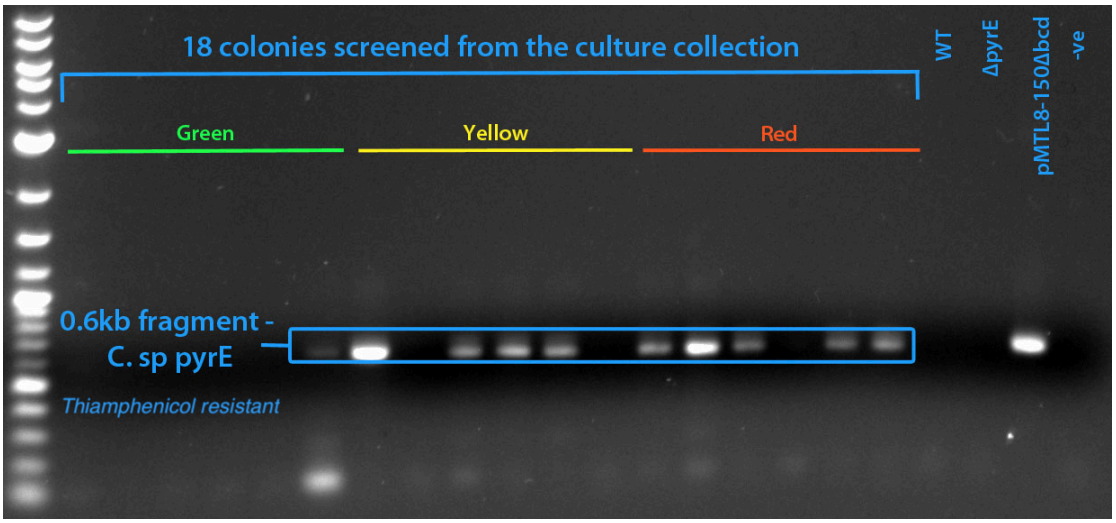


Figure 15: The 8 strains producing the 2 kb PCR fragment corresponding to the persistent deletion of *bcd* at its original locus also demonstrated the presence of the pMTL8-150 plasmid in the genomic DNA extraction. Strains that demonstrated a reversion to wild type at the *bcd* locus were cured of the *pyrE* containing plasmid.

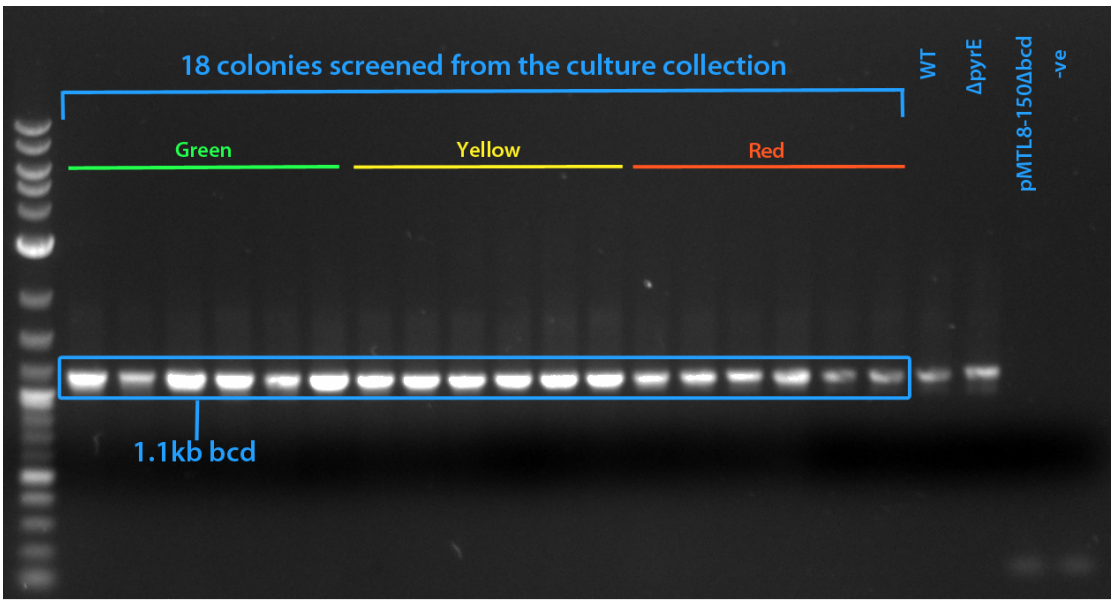


Figure 16: Importantly, after several passages on CBM^{FOA}, all of the strains tested positive when PCR reactions were primed to *bcd* using internally binding primers.

3.3.3 Discovery of plasmid sequence in the *C. acetobutylicum* Δ*pyrE* Δ*bcd* genome using next generation sequencing

Having demonstrated that the plasmid construct pMTL8-150Δ*bcd* remained present in the genome of the *C. acetobutylicum* Δ*pyrE* Δ*bcd* strain it was important to understand how this was possible since the construct lacked a gram positive replicon and only contained the ColE1 replicon for amplification in *E. coli*. The working hypothesis for the investigation was that maintenance of the plasmid was the result of a random integration event with either the chromosome or pSOL megaplasmid. Genomic DNA of the mutant strain was prepared by phenol chloroform extraction and sequenced using an Illumina MISEQ sequencer by The University of Nottingham's sequencing service. The reads were trimmed and then subjected to a *de novo* assembly using CLC Bio with the default settings. This produced 121 contigs. The contigs were searched using 100-200 bp sequences copied from the pMTL8-150Δ*bcd* plasmid sequence or *bcd*, 4 contigs were identified containing the entire sequence for both the plasmid DNA and *bcd.* Interestingly, the contigs all contained terminal complementary regions allowing the manual assembly of a complete plasmid sequence containing *bcd* (figure 17). The regions of complementarity were all 20 bp long with the exception of the region of complementarity joining contig 64 and contig 60 containing *bcd* and the LHA respectively, which was 5 bp in length.

The fact that the contigs generated for the plasmid DNA did not contain any chromosomal DNA may suggest that the genome contained a number of copies of the plasmid and the sequencing process generated a number of

reads which contained matching plasmid DNA but mismatching flanking chromosomal DNA, therefore the *de novo* assembly algorithm was only able to generate contigs for the plasmid DNA.

The alternative explanation would be that the plasmid existed independently as a circular structure and not as an integrant, however, given absence of a Gram positive origin present on the vector, this explanation would be inappropriate and in direct contradiction of the fundamental biological properties of the vector.

Another possible explanation is as follows: In a mixed culture containing both knock out and native genotypes at the target locus, PCRs performed using primers flanking the insertion site are likely to generate a signal for a knock out due to processivity effects of polymerase. Early rounds of replication will result in the amplification bias towards shorter templates which may then compete for polymerase and primers resulting in their domination of the amplification signal.

Recombination events may well be very “fluid” in genomes with low GC owing the weaker binding of thymine to adenine. It might be possible that in cases where genes are either essential or provide a significant fitness advantage knock out plasmids may exist in a constant state of in/out flux with the target site, where cells are strongly selected for should they manage to persist in maintaining the target gene.

In 2016, Wang *et al.,* were the first to report on the phenomenon of vector integration events (VIE) occurring as a result of homologous recombination plasmid borne editing. In their paper they describe using CRISPR-Cas9 nuclease technology to select against VIE strains in *C. beijerinckii*. When the symptoms of the VIE problem became apparent and results reported in this

chapter indicated a failure to lose the pMTL8-150Δ*bcd* plasmid, initially this was regarded as a rare event. However, Wang *et al.,* (2016) report that VIE is probably a very common problem, with one of their initial mutant screens of 17 colonies returning 9 strains showing signs indicative of VIE.

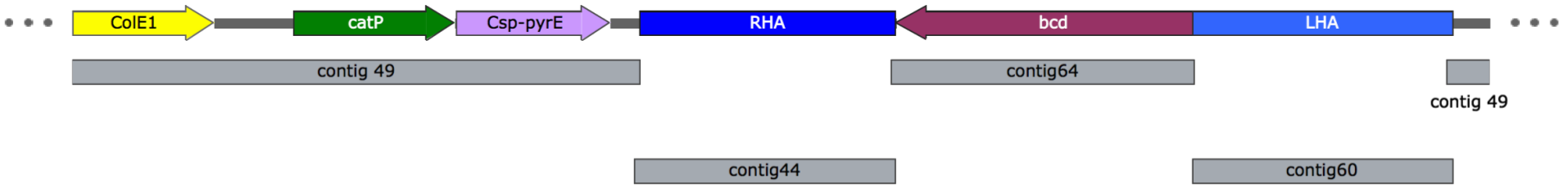


Figure 17: Contig assembly of four contigs from a *de novo* assembly of NGS sequence data produced from the *C. acetobutylicum* Δ*pyrE* Δ*bcd* genomic DNA. The contigs were assembled using short 20-5 bp terminal regions of complementarity.

3.3.4 ClosTron based mutagenesis of *bcd*

Lehmann and Lütke-Eversloh (2011) used ClosTron to generate an intron insertion mutant targeting the *hbd* CDS, inactivating the enzyme 3-hydroxybutyryl-CoA dehydrogenase and altering the product profile by eliminating butyrate and butanol production. Essentially, we can compare this work to a hypothetical strain with an intron insertion targeting *bcd* since they together form part of the C4 processing pathway which is expressed as an operon (figure 18). Although a *hbd* insertion was achievable in Lehmann and Lütke-Eversloh's work (2011) an attempt to achieve a similar strain targeting *bcd* was made. A ClosTron plasmid pMTL007C-E2 containing a group II intron sequence was re-targeted to the *C. acetobutylicum* ATCC 824 *bcd* sequence using the ClosTron targeting algorithm (http://ClosTron.com) and plasmid DNA was synthesised by DNA 2.0. Transformations into *C. acetobutylicum* ATCC 824 were plated onto CGM + 15 μg mL^{-1} thiamphenicol. A total of 6 colonies were picked and streaked on to CGM + 500 μg mL^{-1} erythromycin before transferring to 5 mL cultures of 2×YTG + 500 μg mL^{-1} erythromycin for genomic preparation of DNA and PCR screening for intron insertion at the *bcd* locus. All strains screened returned negative results (see figure 19) indicating off target insertion given that erythromycin resistance is only obtained once the retroactive marker is inserted downstream of a native promotor of a host gene, with the intron.

Several possible explanations are available for this discrepancy: A common reason ClosTron is thought to fail is due to significant target region similarity to another sequence in the genome resulting in a false positive signal during the selection process (Professor Nigel Minton: personal

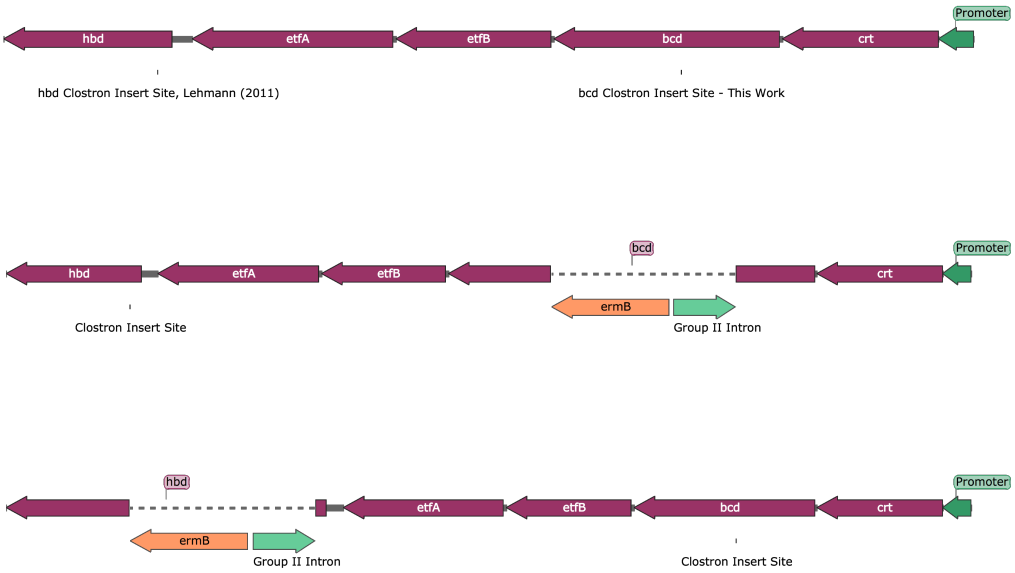


Figure 18: Comparison between the targeting of ClosTron insertional activation mutants of *bcd* (this study) and *hbd* (Lehmann and Lütke-Eversloh [2011]). The RAM-E2 insert length is 1167 bp so downstream genes remain in-frame.

communication). To expand on this, there may be a number of selective forces casting a spectrum of statistical likelihoods between the detection of either a true positive insertion signal or a false positive insertion signal. At the extreme true positive end of the spectrum, the target sequence will be highly novel compared the background genome sequence and loss of target gene function will be completely innocuous to the strains' viability and physiology. On the extreme false positive end of the spectrum the target sequence will have high sequence homology for many off-target sites in the genome sequence, and the target sequence itself will be essential for viability.

In reflection a positive control such as a ClosTron targeting *ldh* or *adhE1* would have provided further confidence in these results, however, the fact that both thiamphenicol followed by erythromycin resistant strains were

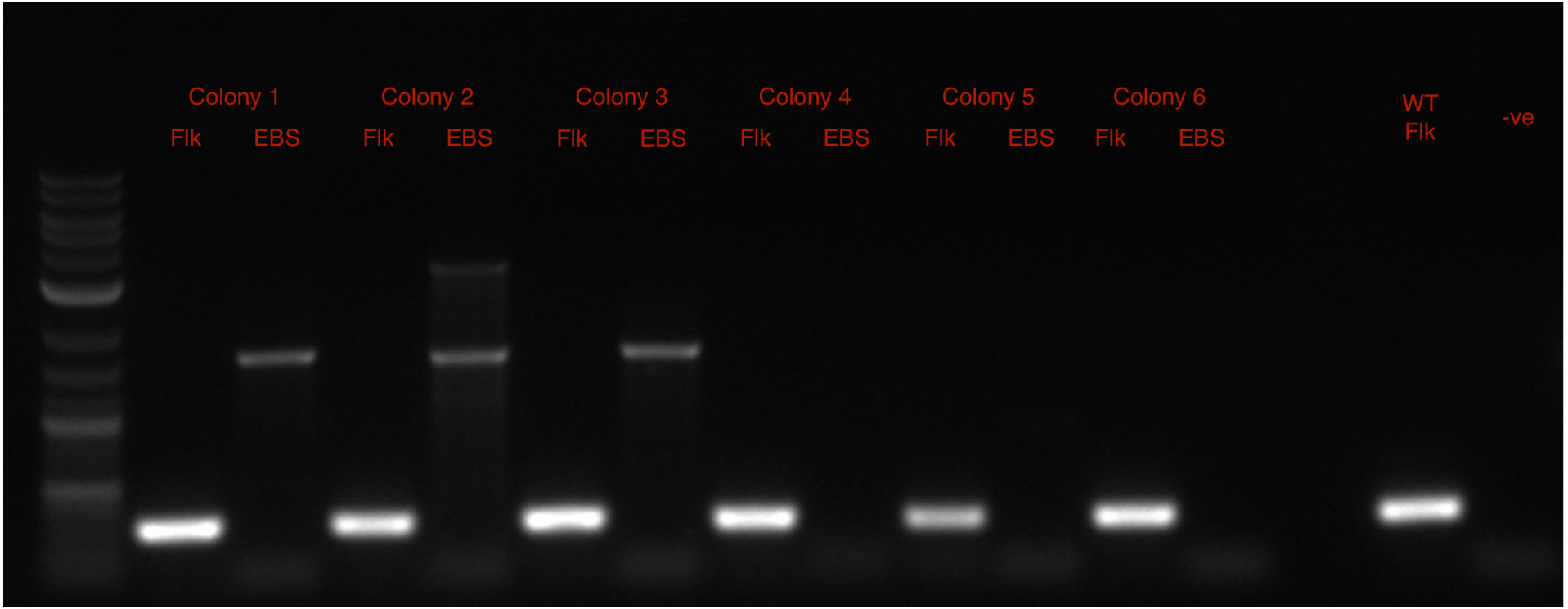


Figure 19: PCR analysis of clostron mutagenesis. Flanking primers produce a 300 bp PCR product in the case of no insertion as demonstrated in the wild type control. An insertion would produce a 2.1 kb product using the flanking primers. A positive result using the EBS binding primers would produce a 500 bp product, and a negative result would not produce a PCR product, no control is available for the EBS PCR analysis. Lane 1 contains NEB log 2 ladder. None of the strains indicate an insertion of the intron based on the criteria outlined. Strains 1, 2 and 3 produced a non-specific PCR product using the EBS binding primers likely due to mis-priming.

obtained should provide enough assurance that the negative result was not a result of technical error and that plasmid transformation had occurred, followed by off-target intron insertion.

A comparison between target sequence accuracy of Lehmann & Lütke-Eversloh's (2011) *hbd* targeted intron and the intron target sequence used in this study (*bcd*) showed similarity in accuracy according to the position specific score equation (Perutka *et al.,* 2004). Lehmann and Lütke-Eversloh's (2011) insertion at 69l70 bp produced a position specific score of 7.898 with two other targets scoring ≥7 and the lowest scoring target of 25 was 3.242. The *bcd* target sequence used in this study targeted insertion at 496l497 bp and produced a position specific score of 7.352 with two other targets scoring ≥7 and the lowest scoring target of 25 was 3.417. So the position specific scores were similar in the context of the range of scores produced by the targeting algorithm.

Given the similar targeting scores we must then consider the possible differences between a permitted Δ*hbd* and a null Δ*bcd*. Considering the impact on the physiology and viability, the resulting metabolism requires an examination of the modified ABE pathways for each mutant strain. The results published by Lehmann and Lütke-Eversloh (2011) showed that a successful intron insertion resulted in increased ethanol production of over 700 mM, acetone production was reduced to around 40 mM, and butanol production was abolished. Regarding a hypothetical intron mutated *bcd* pathway we must consider the impact of the accumulation of the intermediate crotonyl-CoA. In the cases of either butanol and butyrate pathway activity we can speculate the possibility of four β-unsaturated metabolites to subsequently emerge; Crotonic acid, crotonyl phosphate, crotonaldehyde and crotyl alcohol

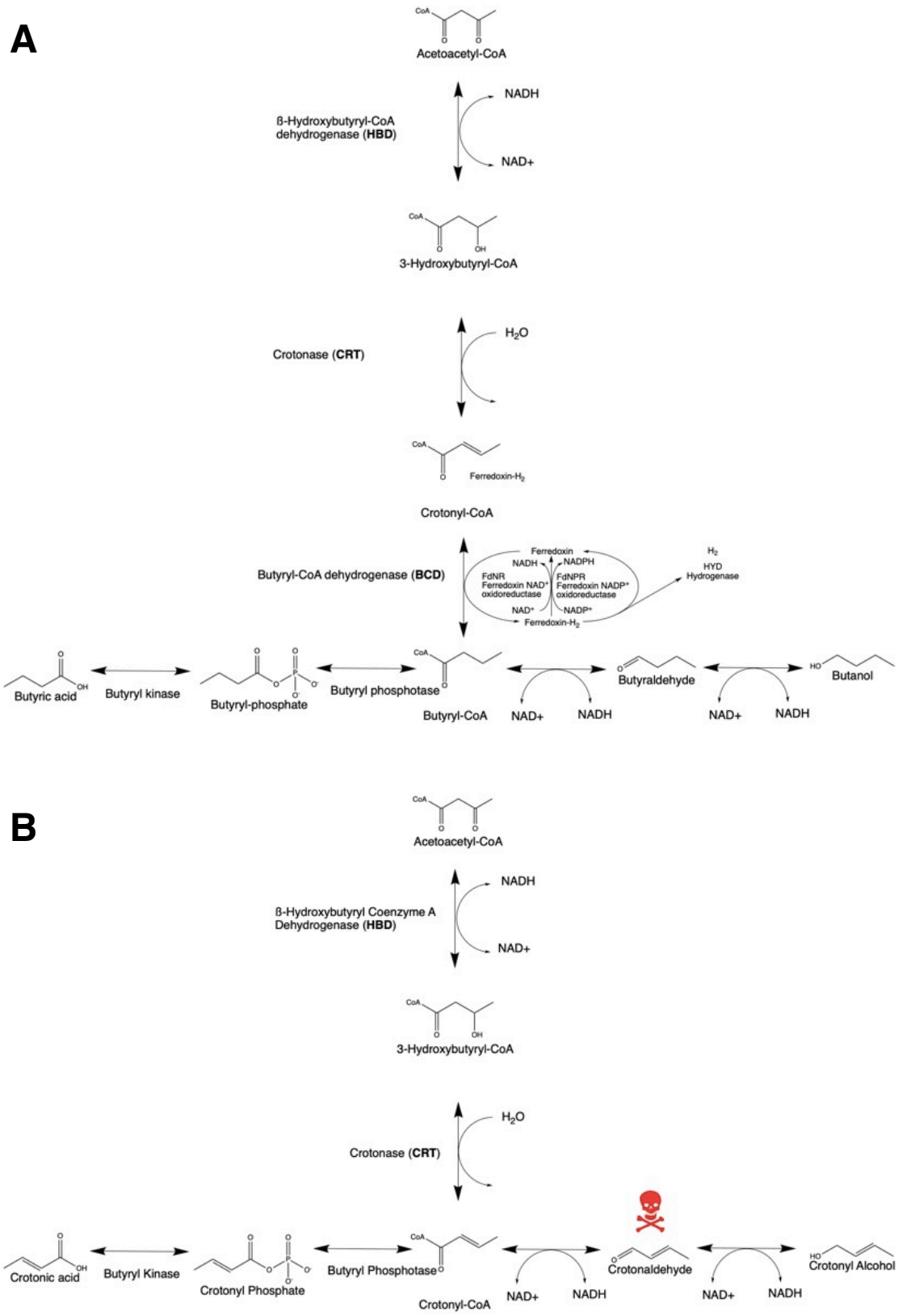


Figure 20: A comparison between (A) the native C4 intermediate processing pathway and (B) a hypothetical Δ*bcd* pathway where the toxic intermediate crotonaldehyde is generated.

(see figure 20).

The CoA-acylating aldehyde dehydrogenase from *Thermus thermophilus* has shown to be highly active at converting crotonyl-CoA to crotonaldehyde producing 1.04 µmol min^{-1} mg^{-1} total protein, the same enzyme converted butyryl-CoA to butylaldehyde + CoA 4.90 µmol min^{-1} mg^{-1} total protein demonstrating that aldehyde dehydrogenases have the potential to generate crotonaldehyde (Krutsakorn *et al.*, 2013).

When expressing the genes *hbd* and *ptb-buk* in *E. coli*, Fischer *et al.*, (2010) observed that crotonic acid was produced at 76 µM and hypothesised that in the absence of competing pathways (*tesB* and *phaB*) over 100 *µ*M might be expected.

Crotonic acid is a short chain β-unsaturated carboxylic acid, butyric acid being the equivalent product in the native pathway. Crotonic acid's biological interactions have been investigated in the literature and the chemical a known allelopathic and autotoxic herbicide (Jasicka-Misiak *et al.,* 2005) however, it has had no observed negative effect on the growth of *Saccharomyces cerevisiae* (Andreason and Stier, 1954) and in a comprehensive analysis of mutagenic properties of unsaturated hydrocarbons on, *Salmonella typhimurium,* crotonic acid showed no sign of possessing mutagenic properties (Lijinsky and Andrews, 1980).

Crotonaldehyde is a β-unsaturated aldehyde and - like many other known β-unsaturated aldehydes - is a highly reactive toxin owing to its two electrophilic sites (Yarbrough & Schultz, 2007). A mutagenesis assay using *supF* (a common mutagenesis reporter) as a reporter determined that shuttle plasmids incubated in 1mM crotonaldehyde at 37˚C for 2 hours demonstrated a mutation frequency reported by *supF* in 1% of *E. coli* colonies transformed,

and 100 mM crotonaldehyde produced a mutation frequency of 4%. Plasmid survival, indicative of mutations in the replicon, showed a survival rate of 75% for 1 mM incubations and 9% for 100 mM incubations (Czerny *et al.,* 1998).

Crotyl alcohol is a β-unsaturated four carbon alcohol that has also demonstrated mutagenic properties, albeit, its effects were less pronounced than those of crotonaldehyde when a side by side comparison was made (Lijinsky and Andrews, 1980).

The accumulation of crotonaldehyde contributing to the inability to obtain a mutant strain by ClosTron is dependant on whether the duel aldehyde/ alcohol dehydrogenases have activity for crotonyl-CoA, however, given that there are multiple duel aldehyde/alcohol dehydrogenases capable of catalysing both butanol and ethanol indicates there may be a possibility, having already demonstrated significant catalytic promiscuity (Yoo *et al*., 2015).

An alternative scenario explaining why Lehmann and Lütke-Eversloh (2011) were successful in obtaining an intron insertion into *hbd* might be the presence of a complimentary single nucleotide variation (SNV) permitting the survival of a strain. A closer look at the methods and results described by Lehmann and Lütke-Eversloh (2011) showed that 26 colonies were picked which contained only 3 clones, the equivalent of a 11.5% success rate. What is obvious from this observation, is that over four times as many more colonies were screened than were for this study, where only six colonies were analysed for insertion, for a reasonable chance of success at least 12 colonies should have been picked. Although, this experiment was repeated several times and in total >20 colonies were screened, all showing a negative result, suggesting that perhaps a single SNV or set of complimentary SNVs

occurred early in the genesis of electro-competent cells used to produce the Lehmann *et al*., (2011) ::*hbd* strain.

3.3.5 Genome sequencing of the ClosTron *ptb* mutant provided by Lehmann & Lütke-Eversloh (2012)

In response to the apparent inability to obtain mutations in the central carbon pathway in both this study and the work by Cooksley *et al.,* (2012) a sample of the *ptb* ClosTron mutant strain: *C. acetobutylicum ptb*::int was obtained from Lehmann and Lütke-Eversloh (2012b) at the request of Professor Nigel Minton, as ::*ptb* mutants are also difficult to obtain, like the C4 intermediate targets. A genomic DNA sample was prepared for genome sequencing and a subsequent SNV analysis was carried out by Dr. Wouter Kuit using a custom workflow in CLC Bio (Table 15). The analysis revealed 16 mutations in the chromosome and none in the pSOL megaplasmid when compared to the reference sequence (Genbank accession NC_003030). Of these, 6 were found in CDS annotated regions, 4 of which were SNVs and 2 were frame-shifts consisting of an insertion and a deletion. After investigation of the CDS 6 SNVs, it seems that two could be potentially complementary to the *ptb* ClosTron mutant.

Firstly an insertion into the *fliY* gene creating a frameshift (figure 20) was considered. In the firmicute *B. subtilis, fliY* has been shown to code for a flagellar chemotaxis mediating protein, where certain mutations result in loss of tumbling phenotype, producing smooth swimming and other mutations in *fliY* resulted in non-flagellate strains (Ward *et al.,* 2019). In *B. subtilis* , flagella is powered either by a Na^+ gradient or a H^+ gradient. This is interesting because in the case where motility in *C. acetobutylicum* is powered by the H^+ gradient generated in acidogenesis, then it could be argued that HydA activity is partly important for establishing and maintaining that gradient by converting intracellular H^+ to H_2 gas, indeed this mechanism has been demonstrated by

Reference Position	Type	Length	Reference	Allele	CDS Annotation	Coding region change	Amino acid change
109376	SNV	1	C	A	hemB	74C>A	Pro25His
162936	SNV	1	A	G	N/A - Intergenic	-	-
344361	SNV	1	T	C	CA_Cr026 (rRNA)	-	-
345993	SNV	1	C	T	CA_Cr027 (rRNA)	-	-
346083	SNV	1	C	G	CA_Cr027 (rRNA)	-	-
1105623	SNV	1	G	A	CA_Cr028 (rRNA)	-	-
1678691	SNV	1	C	A	CA_C1534	907G>T	Ala303Ser
1865678	Deletion	1	A	-	CA_C1715	909delA	Ala303fs
1911061	SNV	1	C	T	N/A - Intergenic	-	-
2102286	SNV	1	T	G	CA_C1989	332A>C	Gln111Pro
2306387	Insertion	1	-	C	fliY	849_850insG	Gln284fs
2772548	Deletion	1	T	-	N/A - Intergenic	-	-
2831844	SNV	1	T	C	hbd	711A>G	Ile237Met
3101780	Insertion	5	-	TTTTA	N/A - Intergenic	-	-
3241092	SNV	1	G	T	N/A - Intergenic	-	-
3337436	Deletion	1	T	-	N/A - Intergenic	-	-

Table 15: SNV analysis by Professor Nigel Minton and Dr. Wouter Kuit on genome sequence data from the phosphate butyryltransferase- negative mutant *C. acetobutylicum ptb*::int . All of the SNPs were discovered in the chromosome sequence (Genbank accession NC_003030).

Girbal *et al.,* (1994). Given that the *hbd* ClosTron mutant produced by Lehmann and Lütke-Eversloh (2011) reportedly produced <100 mL hydrogen from a 100 mL fermentation compared to the 300 mL produced by the wild type control, a non-flagellate strain arising from the *fliY* frameshift could plausibly compensate for the loss of hydrogenase activity supplied by the electron flux from BCD. This could have a significant impact on the rate of dissipation of the proton gradient, otherwise required for membrane transport and possibly F_1 F_0 ATPase activity under the correct stoichiometric conditions. To further investigate the possibility that the frameshift had rendered the strain immotile it would be useful to conduct a motility assay using soft, low density agar plates, and additional microscopy observations for qualitative changes in motility.

Had the authors complemented their ::*ptb* strain, and repeated the physiology assays they might have noticed a change in C4 product formation hinting that there were more genomic influences than the insertion mutation, leading to a better understanding of the roles of the C4 product pathway and hydrogenase activity at the BCD node.

The other mutations worth considering are the SNVs to *cac1989*, and *cac1534* annotated as an iron (III) ABC transporter ATPase and FtsX like ABC transporter ATP dependant permease, respectively. Each of these SNVs result in an amino acid substitution possibly causing a reduction or complete loss of activity. In either scenario the changes to these transporters would conserve ATP, which could result in a higher rate of proton extrusion from F_1F_0 ATPase, further compensating for the loss of hydrogenase activity.

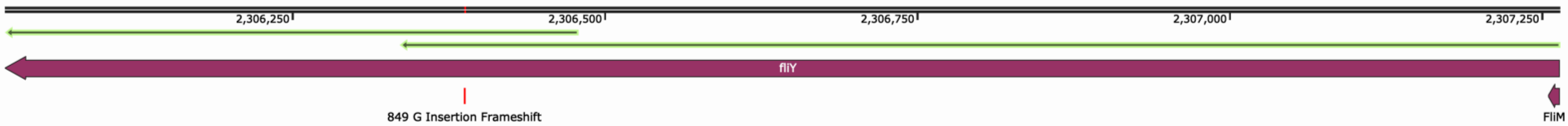


Figure 20: Frame shift mutation in the CDS *fliY* with the G insertion at position 849 indicated in red. Green arrows show open reading frames with the native reading frame truncating prematurely.

3.3.6 Forward engineering of *bcd* employing the use of growth media containing subverting components and a novel double counter selection in-frame deletion vector

A new approach to in-frame deletion using a double counter selection method was employed in an attempt to yield a *bcd* mutant. The double counter selection method makes use of two selection markers, namely *codA* and *pyrE,* separating the two homology arms. Using two counter selection markers in this arrangement meant that, in principle, double crossovers could be selected for immediately after transformation. Transforming into *C. acetobutylicum* Δ*pyrE* and then plating onto CBM^{FC+Tm} selected against the *codA* containing vector therefore favouring excision by double crossover and by positioning *pyrE* and *catP* between the homology arms, double crossovers that had resulted in an in-frame deletion could be positively selected for using thiamphenicol resistance. Double crossovers could then be plated onto $CBM^{FOA+URA}$ allowing for the counter selection of the *pyrE* marker positioned adjacently to *catP*, and between the homology arms, resulting in the removal by recombination occurring between two homologous 29 bp direct repeats, flanking the *pyrE* and *catP* markers.

To generate this vector a fragment containing a *C. sporogenes pyrE,* and *catP* was inserted between a left and right homology arm sequence before assembly using overhang extension PCR (described in section 2.2.2). Overlapping fragments were generated using primers described in table 3 and the resulting fragment was digested with MreI and MauBI then subsequently ligated to a fragment with complementary digests containing *codA* and ColE1 to produce the plasmid pMTL8-146 (figure 21). This plasmid was methylated using the pAN2 method (described in section 2.2.9) and a

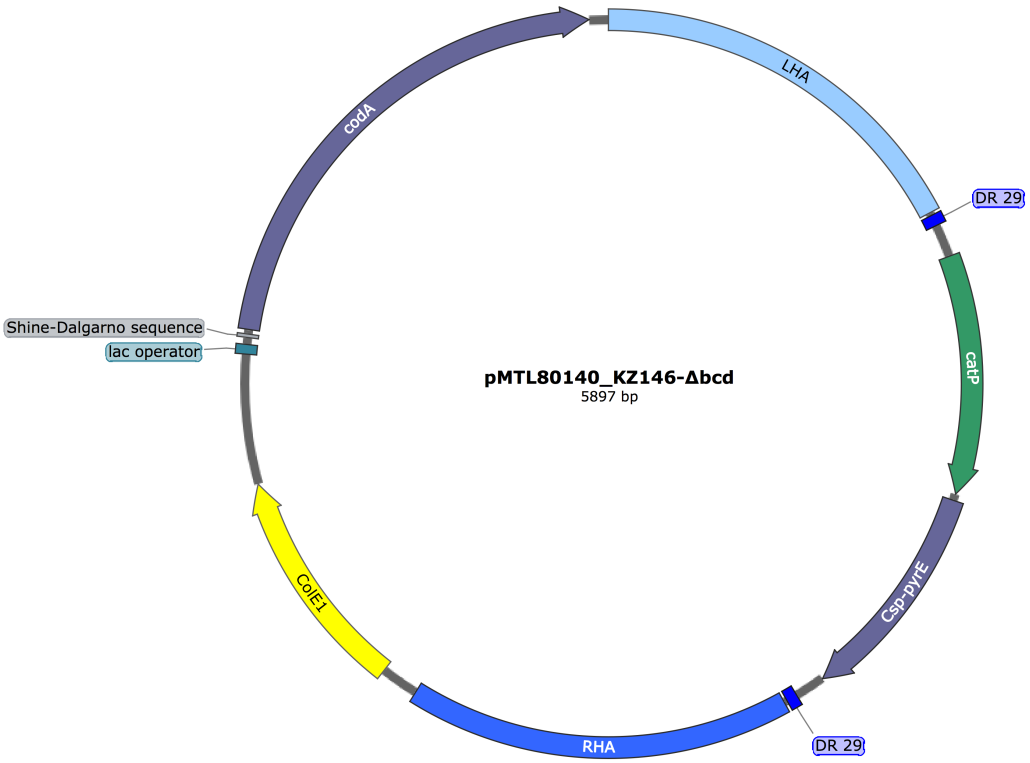


Figure 21: Plasmid map to illustrate the final configuration of the assembled pMTL8-KZ146 double counter selection plasmid targeting the in frame deletion of *bcd* in *C. acetobutylicum*. The plasmid is suicide for *C. acetobutylicum*, hence it is lacking an origin of replication for plasmid maintenance in *Clostridium*. This is a novel approach to achieving in frame deletions via homology arm recombination. ColE1: gram -ve origin; *codA*: cytosine deaminase CDS - counter selection marker; LHA/RHA: Left/Right homology arms; *catP* chloramphenicol acetyltransferase CDS - resistance selection marker; Csp-*pyrE*: *C. sporogenes* Δ*pyrE*; DR29: 29 bp direct repeat homology arms for removal of Csp-*pyrE* and *catP*.

plasmid preparation was made using the Qiagen midi prep kit for transformation according to the method described in section 2.3.1 using 20 μg of plasmid DNA. Transformants were plated out onto CGM^{Tm} to select for single crossovers. After 3 days anaerobic incubation at 37˚C plates were inspected for colony growth and all colonies if ≤ 10 were available were

streaked onto CBM 100 μg mL^{-1} 5-fluorocytosine (CBM^{FC}) + Tm to promote double crossover. After 24 hours incubation strains were then streaked onto CBM + 20 μg mL-1 uracil + 400 μg mL 5-fluoroorotic acid, and thiamphenicol sensitivity was monitored on CGM + Tm. This step was repeated until all of the strains exhibited thiamphenicol sensitivity at which point overnight cultures were prepared in 10 mL 2 × YTG for genomic DNA extraction and subsequent PCR analysis for double crossovers.

Although the design principle was theoretically sound, in practice the available transformation protocols are apparently not efficient enough to overcome the relatively unlikely event of obtaining a double crossover deletion mutant in a single selection step and after several failed attempts to yield a transformation colonies on CBM^{FC+Tm} the method was amended to select for single crossovers using CGM^{Tm}. Transformations plated onto this media would routinely produce <10 colonies, none of which demonstrated evidence of a double crossover event after passaging on CBM^{FC+Tm}, again suggesting essentiality of *bcd*. The plasmid pMTL83140 (containing the pCB102 Gram +ve replicon, *catP* marker and ColE1 Gram -ve replicon, plus a spacer) serving as a positive control for this transformation method comfortably produces >100 colonies.

In order to circumvent this essentiality and to select for double crossovers all of the available colonies were picked and streaked onto CBM^{FC+Tm} and modified CBM^{FC+Tm}. The modifications included (1) addition of 40 mM butyrate, (2) an iron limited version of CBM containing 1 mg mL^{-1} instead of 10 mg mL^{-1} in the original recipe in attempt to impair electron flux via ferredoxin and thus reducing the selective pressure against a Δ*bcd* strain. (3) Media was also supplemented with the physiologically active electron carrier -

methyl viologen - in a similar attempt to impair selection against Δ*bcd* by reducing electron flux to HydA, a redox balancing pathway fed by *bcd* via ferredoxin. Methyl viologen was supplemented in concentrations of 1 mM and (4) 2 mM and (5) a combination of iron limited and 1 mM methyl viologen was also experimented with. After re-streaking onto CBM^{FC+Tm} and the various modified media, the strains were transferred to CBM^{FOA} + 20 μg mL^{-1} uracil and thiamphenicol sensitivity was monitored using CGM^{Tm}. When strains demonstrated thiamphenicol sensitivity, genomic preps were analysed for double crossovers comparing to WT for evidence of a 1.1 kb in frame deletion. As shown in the gel image in figure 22, none of the conditions successfully selected for the isolation of a Δ*bcd* mutant.

Comparing the homology arms on the pMTL8-150Δ*bcd* knock out plasmid the GC content for the LHA is 32% and for the RHA GC content is 35% this discrepancy could contribute to strains reverting to the wild type genotype as the recombination events may be more favourable in the left homology arm, since base pairing is weaker, having significantly less hydrogen bonds.

The hypothesis that a *bcd* deletion could be achieved under the selective conditions was based on the work of Rao and Mutharasan (1986), Peguin, *et al*., (1994) and Peguin and Soucaille (1996) who experimented with iron deficiency and methyl viologen supplementation. Methyl viologen is an electron carrier which diverts electron flow away from hydrogen production and towards NADPH formation by subverting hydrogenase due to its slightly lower reduction potential (E_0' = -460 mv). Experiments by Peguin, *et al* (1994) and Peguin and Soucaille (1996) showed that media supplemented with methyl viologen produced less hydrogen and more alcohols as a result of higher intracellular concentration of NADPH.

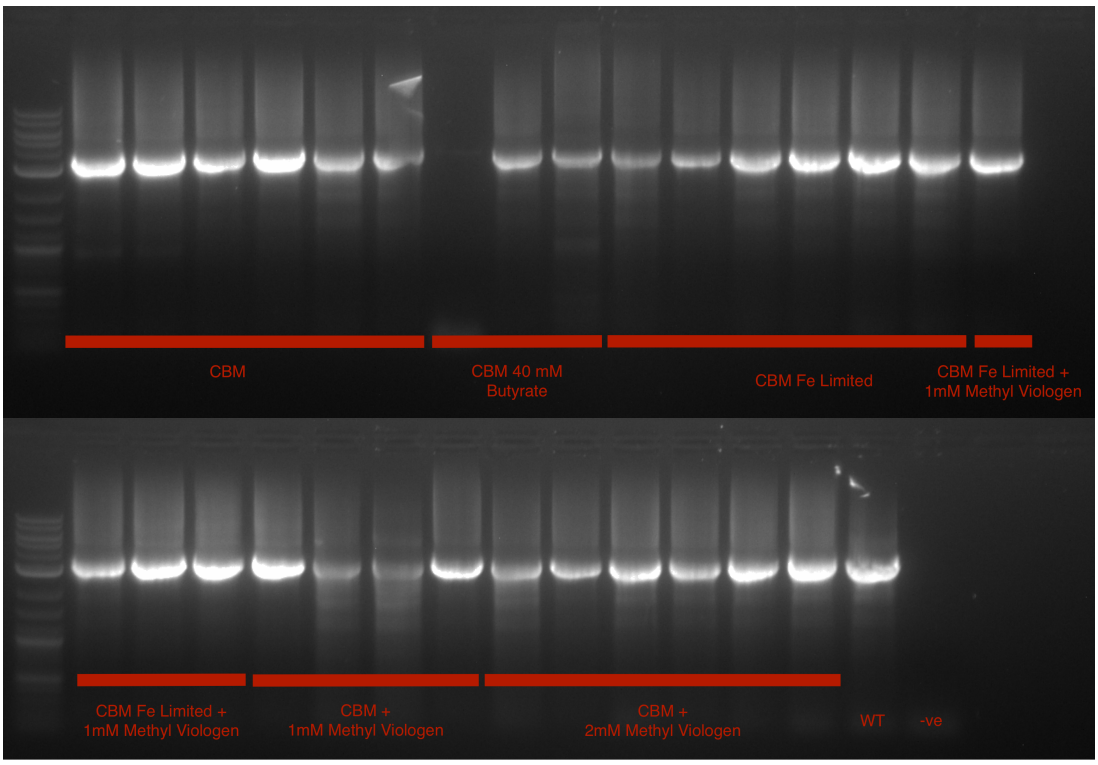


Figure 22: Analytical PCR results screening for the 1.1 kb deletion of the *bcd* gene from the *C. acetobutylicum* chromosome. Primers flanked the target 1 kb upstream and downstream on the chromosome. A 1.1 kb deletion of the target would produce a 2 kb PCR product and a wild type strain would produce a 3.1 kb, as demonstrated using a wild type control. This result indicates that the approach was unsuccessful as all of the experimental conditions produced 3.1 kb PCR products.

Iron is a necessary co-enzyme in Fe-hydrogenases and ferredoxin, and therefore hydrogenase and ferredoxin activity is reduced in an iron-limiting environment. Again, more alcohols are produced as an effect of this (Peguin, *et al.*, 1994). Since *bcd* is involved with this pathway due to its electron bifurcating property it was thought that the conditions causing the selection for a wild type strain might be subverted by provision of a source of redox stress via either iron limitation or methyl viologen supplementation, or a combination since their effects were shown to be additive (Peguin, *et al.*, 1994). This proved not to be the case. Given that in experiments where wild type *C.*

acetobutylicum is subjected to these redox stress inducing factors, cells respond by shifting carbon flux down NADPH consuming pathways in order to rebalance the oxidised co-factor availability (Peguin and Soucaille, 1996), resulting in an increased yield in butanol. This would not be possible in a Δ*bcd* strain as the pathway to butanol production is interrupted. However, the results obtained during Lehmann and Lütke-Eversloh's, (2011) *hbd* ClosTron work suggests that it is possible to knock out the C4 pathway and hydrogen data shows that far less hydrogen is produced. Instead the cells produce excessive amounts of ethanol.

Additionally, 40 mM sodium butyrate supplementation was also tested, the hypothesis being that butyrate can be re-assimilated to form butanol after two reductive steps. This approach also failed - likely due to the fact that hydrogenase activity maintains redox co-factor homeostasis during acidogenesis, however, butyrate is only re-assimilated during solventogenesis.

3.4 Conclusions

A Δ*bcd* mutant was unobtainable in this study. A number of molecular methods were employed to generate a *bcd* deletion or disruption mutant, coupled with attempts to subvert the activity of BCD dependant HydA activity using methyl viologen and iron limitation.

The approach of using counter-selective forces in order to force double crossovers led to vector integration events or simply the excision of the vector alone during the second recombination event.

Analysis of the discrepancies between Lehmann and Lütke-Eversloh's, (2011) ClosTron derived ::*ptb* intron insertion mutant and this work generated a number of interesting leads. Firstly the possibility of crotonyl-CoA to a number of β-unsaturated carbonyl compounds was examined, where in the case of crotonaldehyde production the isolation of a Δ*bcd* strain would be prevented due to the chemical's highly mutagenic properties.

Perhaps what was most interesting was the discovery of a frameshift mutation in the gene *fliY* identified by whole genome sequence SNV analysis of Lehmann and Lütke-Eversloh's, (2012b) ::*ptb* ClosTron strain. There is a possibility that this mutation resulted in improper function/formation of the flagella aiding the isolation of a mutant due to the effect of limiting the rate of proton gradient dissipation that would otherwise contribute to the proton motive force driving motility. This would provide evidence for the role of *C. acetobutylicum* hydrogenase in the establishment and maintenance of the pH gradient powering the proton motive force and that its role is of significant importance.

Given all of these points it is not clear whether native TER activity of *C. acetobutylicum* is able to compensate for the butyryl-CoA dehydrogenase activity, or whether it's the necessity of the bifurcating activity aiding the maintenance of redox co-factor homeostasis and/or proton motive force, or indeed whether the production of toxic crotonaldehyde is playing a role. These questions can be explored using a reverse genetics approach to over-express *ter* and then attempting to knock out *bcd*.

Chapter 4: Manipulation of the redox cofactor balance for enhanced production of butanol

4.1 Introduction

4.1.1 Trans-2-enoyl-CoA reductase is an alternative to butyryl-CoA dehydrogenase for the generation of butyryl-CoA

Trans-2-enoyl-CoA (TER) catalyses the reduction of crotonyl CoA to butyryl-CoA using 1 mole of NADH per mole of substrate consumed and was first described by Hoffmeister *et al.,* (2005) after purifying the enzyme from *Euglena gracilis* mitochondria. In this context the enzyme contributes to the synthesis of wax esters, which serve as an electron sink in anaerobic glycolytic ATP synthesis. The enzyme has specificity of C4 (crotonyl-CoA) and after purification the NADH specific activity was found to be 1.6 ± 0.02 μM mg^{-1} min^{-1} with crotonyl-CoA as a substrate. Hoffmeister *et al.,* (2005) also demonstrated that this enzyme has activity for C6 intermediates (trans-2-hexenoyl-CoA).

The first bacterial TER was described by Tucci and Martin (2007), and was isolated from the Gram negative spirochete *Treponema denticola* and is thought to belong to a novel anabolic fatty acid synthesis pathway, in contrast to the role in energy metabolism associated with *E. gracilis* TER. Tucci and Martin, (2007) found NADH specific enzyme activity to be 43 ± 4.8 μM mg^{-1} min^{-1}, however, a repeat of this assay conducted by Hu *et al.,* (2013) measured a ten fold higher activity at 455.8 ± 9.6 μM mg^{-1} min^{-1}. This discrepancy was attributed to the difference in purification method, where Tucci and Martin, (2007) used a one step affinity chromatography, however, Hu *et al.,* (2013) used two step process including a secondary a gel filtration.

In contrast to TER, BCD, requires 2 moles of NADH in order to reduce crotonyl-CoA, accounting for the bifurcating activity to ferredoxin. The ferredoxin independent reaction of TER has been used previously to replace the activity of BCD/EtfAB in a synthetic heterologous butanol pathway in *E. coli* (Shen *et al.,* 2011; Bond-Watts *et*

al., 2011) and for synthesis of fatty alcohols (Dekishima *et al.,* 2011; Machado *et al.,* 2012).

Shen *et al.,* (2011) produced an interesting useful hypothesis for the justification of employing TER in a synthetic butanol pathway. Given that TER does not bifurcate to ferredoxin as with BCD, NADH is conserved for use in the butanol pathway, and a shift in the NAD^+/NADH ratio towards the reduced cofactor represents a driving force guiding carbon flux from acetyl-CoA towards reduced butanol in order to reach thermodynamic equilibrium. In principle this works and Shen *et al.,* (2011) reported a 10 fold increase (30 g L^{-1}) in yields when compared to a pathway utilising BCD. Although, it seems the authors had neglected to consider that catalysis of crotonyl-CoA to butyryl-CoA by the enzyme BCD is ferredoxin dependant and unless this is coupled with an oxidative reaction to regenerate ferredoxin proceeding at a higher rate than the conversion of crotonyl-CoA the reaction will be limited. The study made no reference to the problem. Shen *et al.,* (2011) did state, however, that a previous study (Buelter *et al.,* 2009) using TER had only produced 250 mg L^{-1}, and attributed this to the use of *C. acetobutylicum* thiolase, rather than AtoB from *E. coli*, although here, the authors describe the use of *E. gracilis* TER rather than the *T. denticola* TER, making the comparison unreasonable.

Nevertheless, the principle of manipulating the redox co-factor ratio towards the reduced form stands as a good method for the creation of a driving force steering carbon flux towards reduced product formation i.e. butanol, and replacing BCD with TER in *C. acetobutylicum* should produce interesting results. Indeed, alcohologenesis, consistent with improvements in butanol yield, have been achieved already using various biochemical approaches to redox manipulation either by glycerol provision, iron limitation, CO gassing, or methyl viologen supplementation (Peguin and Soucaille, 1994; Vasconcelos *et al.,* 2003; Kim *et al.,* 1984). Subverting hydrogenase activity is plausible based on the principle that the role of hydrogenase is to maintain redox homeostasis while producing acetate (Gottschalk, 1986).

Furthermore, manipulating NADH/NAD^+ ratio is not uncommon in the field of metabolic engineering and many studies have demonstrated that it can be a successful strategy for yield improvements to reduced products (Qiao *et al.,* 2017; Lim *et al.,* 2013; San *et al.,* 2002).

4.1.2 Genetic strategies for metabolic engineering of carbon flux via the manipulation of the redox cofactor ratio in *Clostridium acetobutylicum*

The allele coupled exchange method (Heap *et al.,* 2012; Minton *et al.,* 2016) for stable chromosomal integration of genes in *Clostridium sp.* is a reliable and robust method for genetic engineering for synthetic biology purposes or for basic science purposes such as reverse genetics gene characterisation. The original paper (Heap *et al.,* 2012) provided two methods for integration into *C. acetobutylicum*. A markerless integration approach relied on a uracil auxotroph resulting from a partially deleted (although not entirely accurate, the partially deleted *pyrE* strain will be referred to simply as Δ*pyrE*) *C. acetobutylicum* strain which allowed for selection of integrants dependant on uracil prototrophy after successful integration of the remaining fragment of the *pyrE* gene the desired gene or operon downstream of *pyrE*. Later, the pMTL-ME6X integration plasmid was constructed to allow for over-expression of the desired genes or operon, by inclusion of the *C. sporogenes* ferredoxin promotor upstream of the multiple cloning site (Minton *et al.,* 2016).

The second method, although not as elegant, provided protection against toxicity of genes of interest throughout the cloning stages due to the absence of an upstream promotor on the vector. By allowing for insertion of a gene of interest downstream of the *thlA* locus, this ensures transcription can only occur once the gene has been integrated into the host chromosome, under the control of the strong thiolase promotor. The downside, however, was that this system relies on an erythromycin resistance marker,

ermB, in order for selection of a positive integrant, which is left behind after the integration procedure is complete.

Developments in in-frame deletion technology were also available for use for this work. Ehsaan *et al.,* (2016) have produced an in-frame deletion method which utilises *codA* as a counter selection marker which works similarly to the method describe by Ng *et al.,* (2013), however, this does not need to take place in a Δ*pyrE* background strain, negating the requirement to work with auxotrophs and the subsequent repair of Δ*pyrE.*

The most recent advancements in gene deletion methods includes the employment of CRISPR/Cas9 selection, three methods were trialled, two of which are now published (Ingle *et al.,* 2019; Cañadas *et al.,* 2019).

4.2 Results and Discussion

4.2.1 Generation of a heterologous trans-2-enoyl-CoA reductase expressing strain

The generation of a chromosome integrated gene for the expression of the *ter* originating from *Treponema denticola* was achieved using the method described by Heap *et al.,* (2012) (figure 23). *T. denticola ter* was selected over the *C. acetobutylicum ter* as Hu *et al*., (2013) demonstrated that this enzyme had almost 10 fold higher activity. The JH16 thiolase integration vector was used to insert the genes downstream from the *thlA* locus together with the antibiotic resistance selection marker *ermB*, providing high, polycistronic expression under the control of the strong constitutive *thlA* promotor while avoiding the apparent toxic effects of cloning the gene under a strong constitutive plasmid based promotor (figure 24). Altogether, three strains were created, a native, codon optimised *ter* expressing strain, a FLAG tagged *ter* expressing strain and a pMTL8-JH16 control which expressed the *ermB* marker alone. After transformation with each of the constructs all strains were successfully cured of the plasmid DNA demonstrating thiamphenicol sensitivity indicating loss of *catP* activity, whilst maintaining erythromycin resistance as a result of double crossover recombination of the insertion plasmid. Insertions were confirmed using PCR (see figure 25). The JH16ACEInsert_Fw/Rv PCR primers flanking the insertion site produced a 400 bp fragment in the wild type and a 2.5 kb band in the *ter* insertion strains, the JH16 *ermB* controls produced a 1.1 kb band.

Figure 2: The pMTL8-WOK40 plasmid construct comprising of *ter*_FLAG insert into the pMTL8-JH16 chromosomal insertion vector. The resulting chromosomal configuration of the gene inserts at the *thlA* locus. The *ermB* selection marker and Td_ter (with or without FLAG tag) are expressed polycistronically with *thlA* driven by the P_{thl} promotor. Td_ter has an additional Shine Dalgarno RBS 9 bp upstream of the start codon.

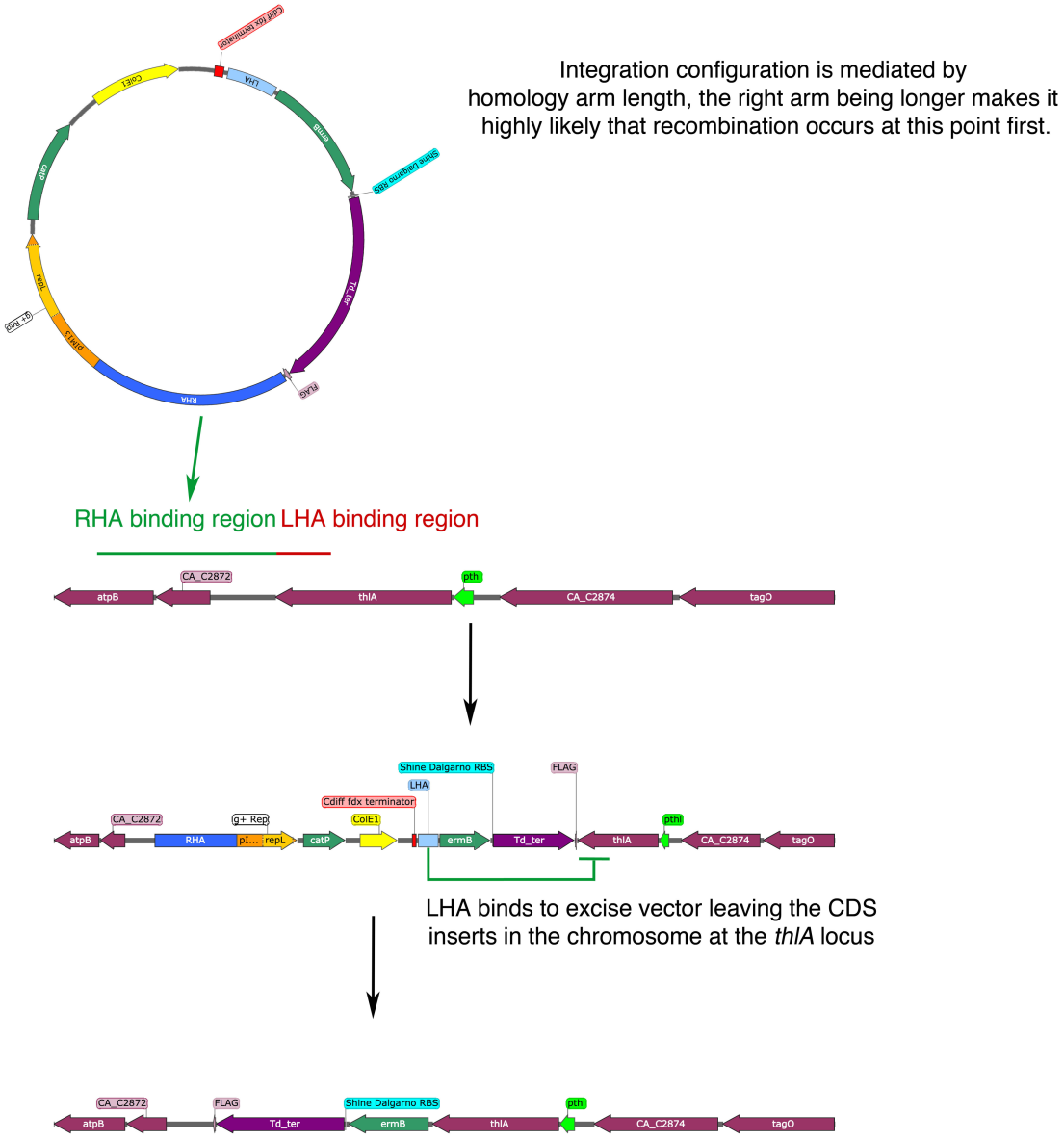


Figure 23: Insertion of heterologous genes downstream of *thlA* for constitutive expression using homologous recombination method described by *Heap et al*, (2012). The longer right homology arm of the insertion vector recombines first resulting in insertion of the vector. The second recombination event is less likely to occur than the repeated recombination of the right homology region, returning to the wild type genotype, however, the desired left homology recombination event can be selected for using the marker *ermB* as it is not correctly orientated until the left homology recombination event has taken place, and the vector backbone has been excised.

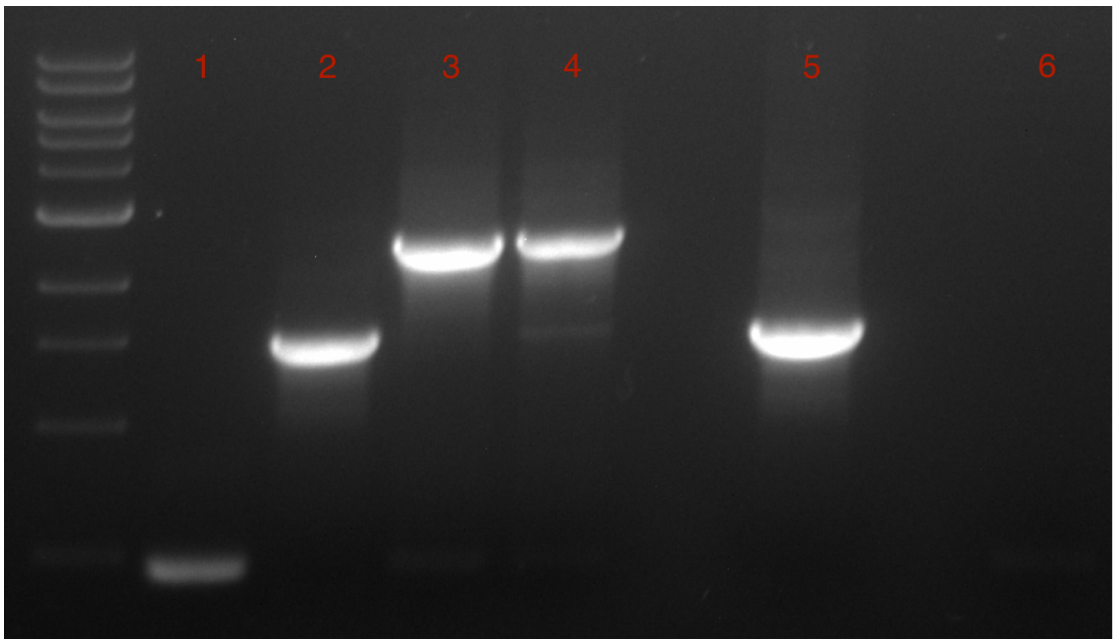


Figure 24: PCR on genomic DNA to demonstrate insertion of *ter* 1: WT; 2: pMTL8-JH16; 3: pMTL8-WOK37; 4: pMTL8-WOK40; 5: pMTL8-JH16 plasmid DNA control 6: PCR -ve control. NEB 1 kb ladder in the first lane.

4.2.2 Flask based physiological characterisation of the *ter* expressing strains

All three strains were characterised in 50 mL CBMS flask fermentations together with a *C. acetobutylicum* ATCC824 wild type control. *C. acetobutylicum* + JH16 -ve served as a control for expression of the *ermB* marker at the *thlA* locus. Fermentations were carried out in biological triplicate (n = 3).

The *ter* expressing strains demonstrated a profound difference in their physiological characteristics, with rapid acid re-assimilation and improved solvent productivity rates in the first 26 hours before TER strains showed productivities similar to the controls (figure 25).

The growth rates of the *ter* expressing strains were marginally affected, with *ter* doubling every 1.2 hours and the *ter*_FLAG doubling every 1.3 hours, compared with 1.1 for the wild type. The pH profiles of the *ter* expressing strains reflected the increased solvent productivity, with pH rising rapidly after initiation of solvent formation.

Butyrate re-assimilation was particularly pronounced in both strains expressing the FLAG tagged version of *ter* and the native, whereas acetate profiles displayed a relative similarity but with a marked degree of variation across all replicates at 50 hours. Interestingly, in the JH16 -ve control strain a diversion from the wild type control on the butyrate profile briefly occurred, where an increased rate of assimilation was observed at 50 hours. Wild type also produced double the amount of butyrate as acidogenesis continued until 26 - hours longer than the *C. acetobutylicum* + *ter*_FLAG strain, where acidogenesis had ended prior to this time point and acids had largely been re-assimilated - fuelling solventogenesis.

The solvent profile was particularly interesting, where the *C. acetobutylicum* + *ter* and *C. acetobutylicum* + *ter*_FLAG expressing strains reached the maximum final yield of 112 mM (± 8.6 SD) and 130 mM (± 17.1 SD), respectively, for butanol measured at the 78 hours time point, the wild type reached similar concentrations of

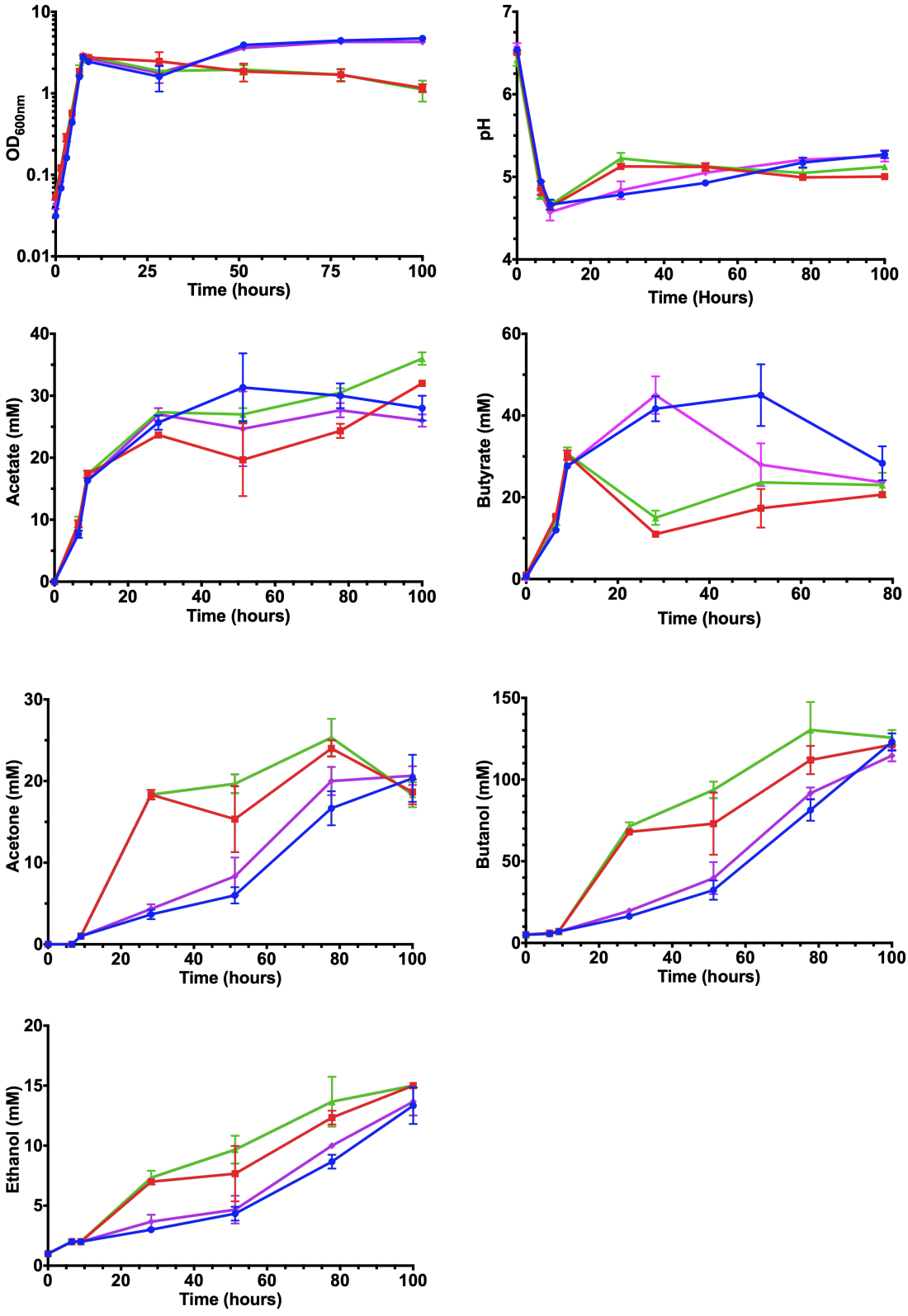


Figure 25: Fermentation analysis was conducted in 50 mL CBMS producing physiological profiles of *C. acetobutylicum* (wild type control) ●; *C. acetobutylicum* + *ter* ■; *C. acetobutylicum* + *ter*_FLAG ▲; *C. acetobutylicum* + JH16 control: ◆. Error bars show standard deviation where *n* = 3.

123 mM at 100 hours.

These gains were, however, mostly made during the interval between the 9 hour time point and the 28 hour time point where the *ter* strains both demonstrated a drastically improved productivity and cell specific productivity (table 16) before displaying solvent productivity rates comparable to the *C. acetobutylicum* ATCC 824 wild type control.

This might be explained if we consider the mechanism of the redox response regulator, Rex. Rex is a negatively regulating transcription factor/repressor which binds the rex operator when the NAD^+/NADH ratio is sufficiently high. In the reverse scenario, NADH binds to Rex and allows for the transcription of the rex regulon, which includes genes for butanol and butyric acid production including the dual aldehyde alcohol dehydrogenase *adhE2*, *thlA*, *crt-bcd-etfAB-hbd,* and *ptb-buk* (butyrate pathway) (Wietzke *et al.,* 2012; Zhang *et al.,* 2014). Given that AdhE2 has activity for both ethanol and butanol production this might help us to understand the increased alcohol productivity, although the rate of butyrate production only appeared to occur at a marginally increased rate, and differences are not significant.

Strain	Mean Butanol ± SD (mM)	Butanol Productivity (mM h^{-1})	Butanol Cell Specific Productivity (mM h^{-1} OD_{600nm}^{-1})
C. acetobutylicum	9.3 ±1.2	0.5	0.3
***C. acetobutylicum* + JH16 (control)**	13.7 ± 0.7	0.7	0.4
C. acetobutylicum* + *ter	61 ± 1.7	3.2	1.3
***C. acetobutylicum* + *ter*_FLAG**	64.3 ± 2.5	3.4	1.8

Table 16: Butanol productivities of the flask fermentations calculated in early solventogenesis - 9-28 hours. Mean butanol compared at 28 hrs.

Interestingly, and unexpectedly, acetone appeared to increase in yield for the *ter* strains with the *C. acetobutylicum ter* reaching 24 mM (SD ± 1) *C. acetobutylicum ter*_FLAG reaching 25 mM (SD ± 2.3) compared to the wild type *C. acetobutylicum* which reached 20.3 mM (SD ± 2.9) - although the SD overlaps by 0.2 mM and has a p-value = 0.1 for *C. acetobutylicum* vs *C. acetobutylicum ter* and a p-value = 0.07 for *C. acetobutylicum* vs *C. acetobutylicum ter*_FLAG meaning theres a significant possibly the difference is due to random sampling bias in the n = 3 experiment. Although it should also be noted that acetone production is not apparently regulated by Rex as neither *adc* or the *sol* operon (which includes CoA transferase - which constitutes as the first stage of acetone production).

The *C. acetobutylicum* + JH16 control fermentation did not deviate in any significant way from the *C. acetobutylicum* wild type except during butyrate re-assimilation measured at the 51 hours time point (figure 26). This may indicate that placing genes downstream of *thlA* has a possible impact on *thlA* expression resulting altered physiological characteristics.

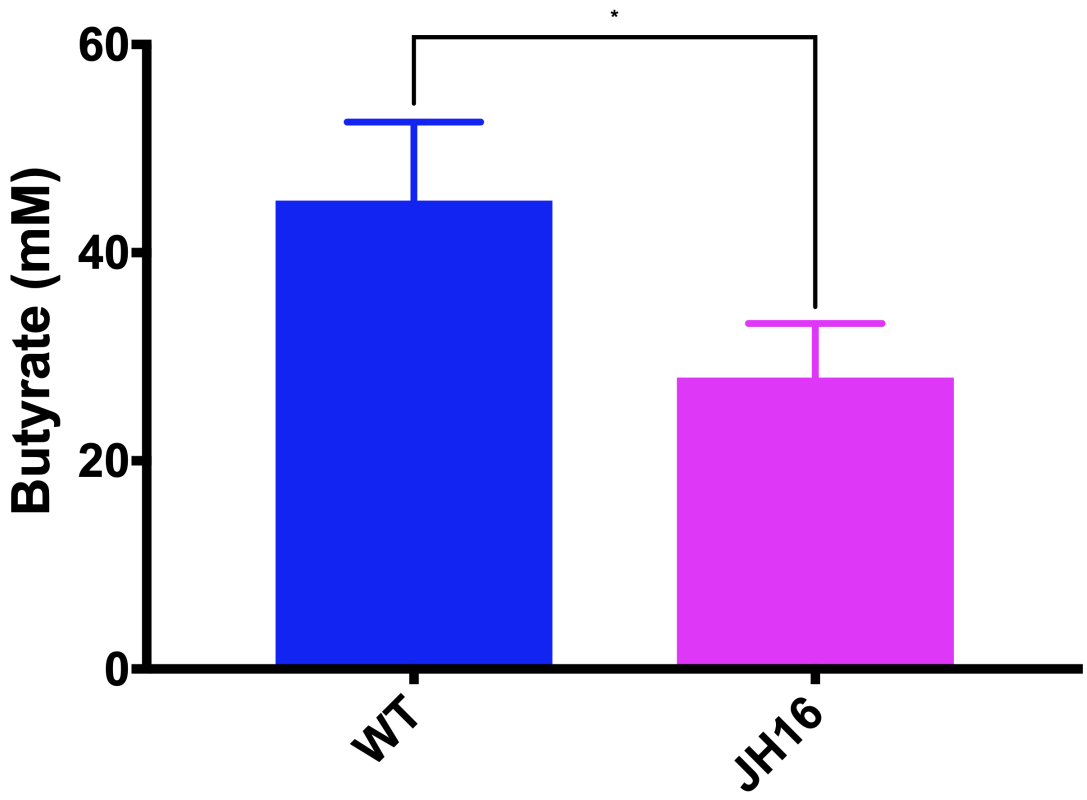


Figure 26: Unpaired, parametric t-tests (*n*=3) comparing the re-assimilation of butyrate between the *C. acetobutylicum* control and the *C. acetobutylicum* + JH16 control measured at 51 hours produced a p-value of 0.0325 indicating a significant difference between the wild type control and the JH16 control for butyrate at a single time point.

4.2.3 Bioreactor physiological characterisation of the *ter* expressing strains

With initial positive results obtained from the flask experiments, a scaled up comprehensive analysis was performed in 300 mL stirred tank, pH controlled bioreactors at 4.9 ± 0.05, with temperature maintained at 35˚C as a previous fermentation at 37˚C resulted in acid crash (Maddox *et al.,* 2000) in all of the cultures. Due to the relative similarity observed between the *C. acetobutylicum* control and the *C. acetobutylicum* + JH16 control the *C. acetobutylicum* ATCC 824 wild type was taken forward as the primary control and compared to the *C. acetobutylicum* + *ter*_FLAG which was used so that a western blot analysis could be performed to assess the drop in productivity noticed in the flasks after 28 hours. Fermentations were carried out in biological duplicate (n = 2), although measuring standard deviation used as an indication of biological variation between cultures is of limited value when using only two biological duplicates, these limitations are often unavoidable due to the expense and demand for fermentation equipment. It should be noted, however, that the bioreactor experiments were highly reproducible, although, this data was chosen due to the comprehensive sampling regime in the first 30 hours.

Strain analysis using bioreactors displayed similar differences in physiology, albeit with notable adjustments in profile dynamics (figures 27 & 28). Doubling times during exponential growth phase were greater in the bioreactor, probably on account of the reduced temperature and the elimination of concentration gradients, with wild type doubling every 1.6 hours and *ter* doubling every 1.9 hours. Butyrate analysis, again revealed a more rapid rate of re-assimilation by the *C. acetobutylicum* + *ter*_FLAG strain, however, a more comprehensive sampling regimen actually revealed that acid production occurred at a delayed rate lengthening the acidogenic period by 8.5 hours before butyrate was then rapidly re-assimilated in 7 hours, the wild type required 37.5 hours to achieve equivalent levels of butyrate re-assimilation.

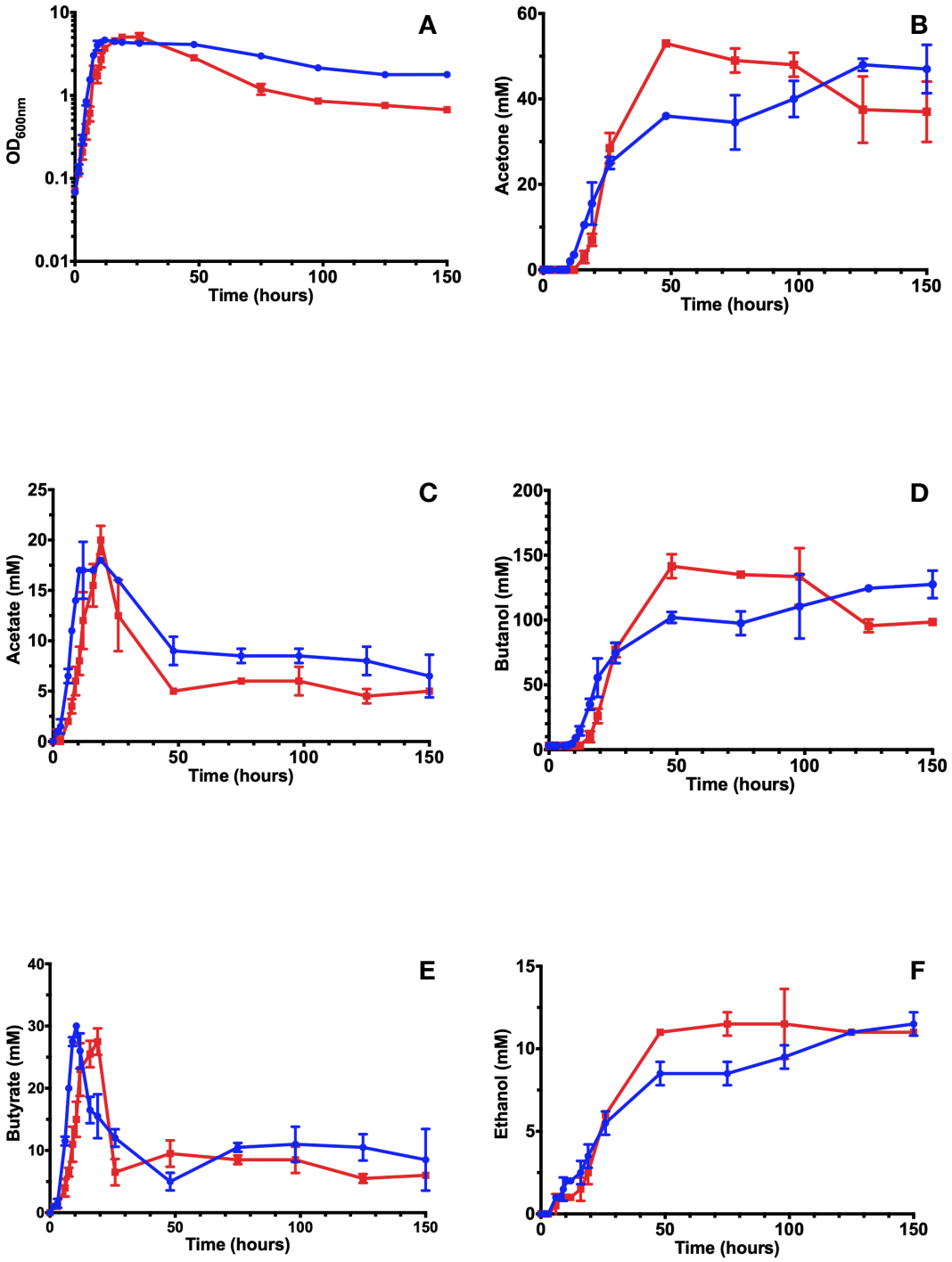


Figure 27: Fermentation analysis was conducted in bioreactors containing a working volume of 300 mL CBMS in order to produce physiological profiles of *C. acetobutylicum* ATCC 824 (wild type control) ●; *C. acetobutylicum* + *ter*_FLAG ■. Fermentations were carried out as two independent cultures (*n* = 2) of the same strain representing biological duplicates. Error bars show standard deviation. Growth (A) is monitored spectrophotometrically (OD 600nm) and solvents (B, D, F) and acids (C, E) in the culture supernatant were measured using GC.

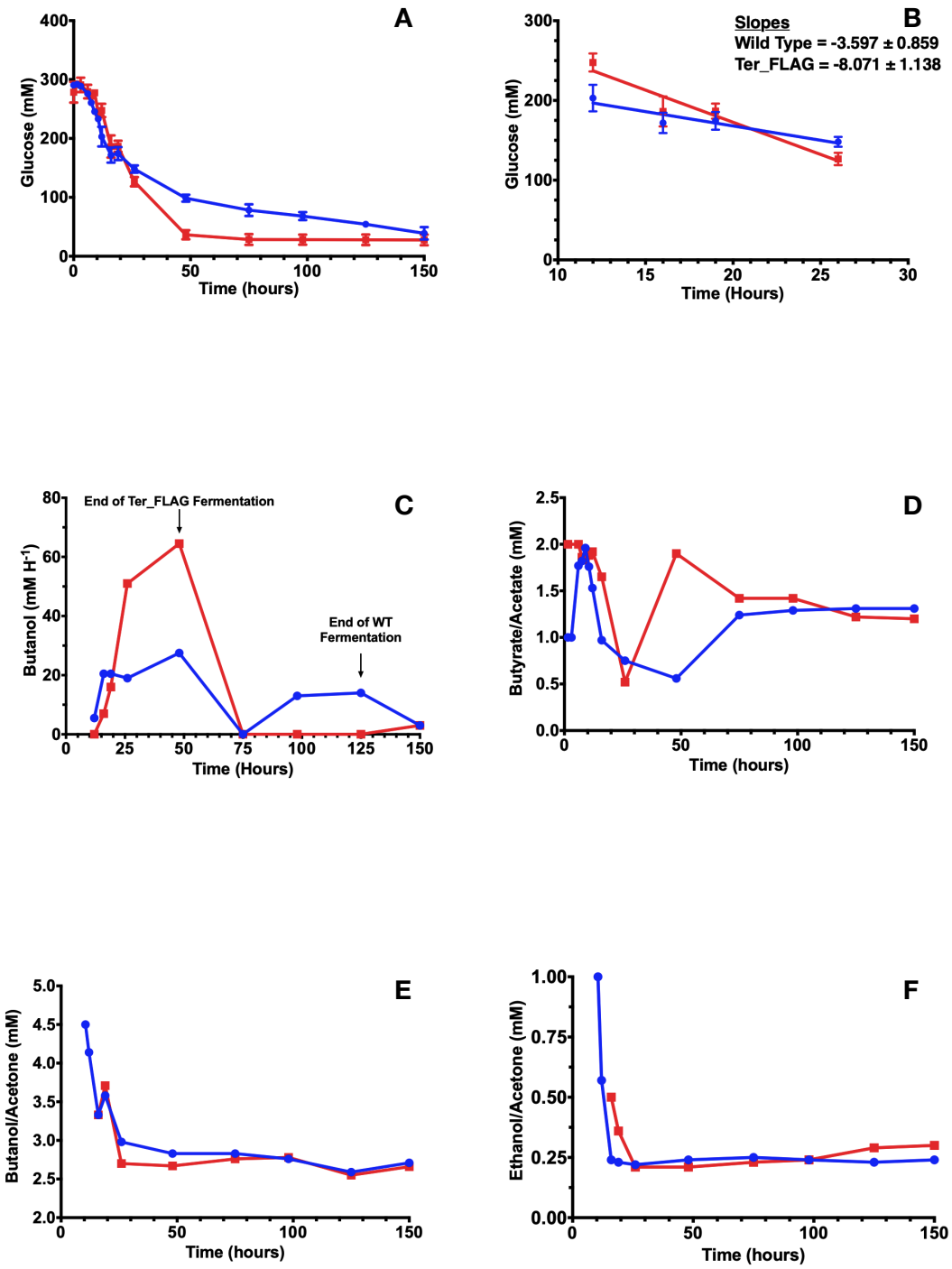


Figure 28: Glucose concentration was measured using HPLC (A) and a linear regression analysis was performed (B) on four data points selected between 12 and 24 hours to show a significant difference in glucose consumption (p=0.0086). Butanol productivity (mM H^{-1}) was calculated (C). Moles of reduced product i.e. products produced via NADH consuming pathways per mole of non reduced product were calculated to investigate whether Ter activity resulted in a shift towards alcohologenesis. (D, E, F). Colouring as in figure 27.

Acetate profiles also differed from the flask, where both strains demonstrated more complete re-assimilation dynamics, although the *C. acetobutylicum ter*_FLAG strain displayed a delay in the commencement of re-assimilation and an increase in the rate of assimilation, furthermore, re-assimilation was not as rapid for acetate as it was for butyrate - taking until 48 hours.

All of the solvent profiles showed a similar improvement in productivity during the 16-48 hour period, however, in the bioreactors, the benefit was not observed until 48 hour timepoint, as the wild type control performed much better in these conditions and during early solventogenesis, productivity was similar between the two strains (table 17). The *C. acetobutylicum* + *ter*_FLAG strain also performed better in solventogenesis with all three solvents reaching the maximum yield 100 hours before the wild type control peaking at 48 hours at which point the butanol and acetone concentrations were at significantly higher concentrations than the wild type, differences in ethanol produced a p-value of 0.0505 (figure 29). When comparing to the flask data, the *ter*_FLAG strain performed better in the bioreactors as the flasks did not reach the final yield for butanol until the 80 hour timepoint, and the mean maximum final yield was approximately 20 mM lower. Between 96 hours and 124 hours approximately 40 mM butanol was lost from the culture broth, as was the case for 15 mM of the acetone yield, this might be explained by the fact that nitrogen was sparged though the vessel at 0.05 vvm providing a positive pressure to maintain

Strain	Mean Butanol ± SD (mM)	Butanol Productivity (mM h^{-1})	Butanol Cell Specific Productivity (mM h^{-1} OD_{600nm}^{-1})
C. acetobutylicum	65.5 ± 7.1	3.9	0.9
***C. acetobutylicum* + *ter*_FLAG**	74 ± 5.7	4.3	0.8

Table 17: Summary of butanol productivities of the bioreactor fermentations calculated in early solventogenesis - 9-26 hours.

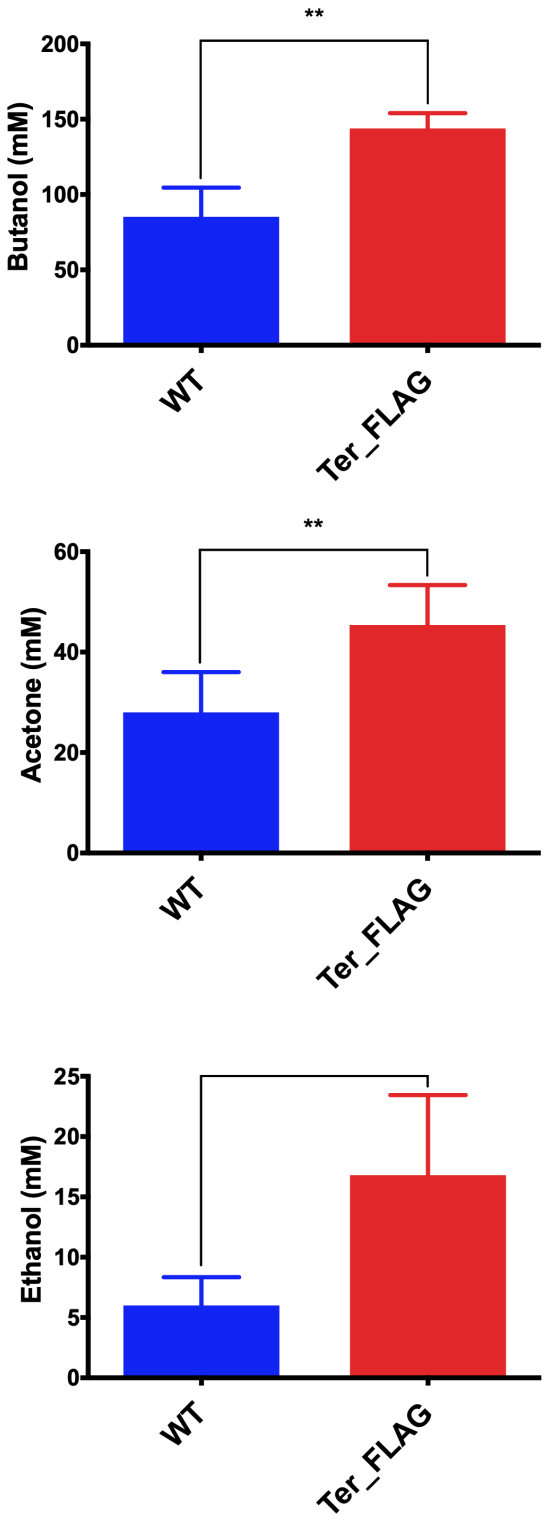


Figure 29: Unpaired parametric t-tests were used to compare difference in solvent concentrations at 48 hours. Data from a repeat of the same fermentation performed in triplicate (*n* = 3) were pooled with the *n* = 2 data described in this section. *n* = 5 Butanol p-value = 0.0040; Acetone p-value = 0.0058; Ethanol p-value = 0.0505. Error bars indicate standard deviation.

axenic conditions. This will certainly have had some gas stripping effects on the acetone which is a volatile solvent but it is also possible to remove some butanol from culture broths by gas stripping despite its high boiling point (Qureshi & Blaschek, 2001).

The glucose, in reflection of the solvent production profiles was almost completely consumed at 48 hours sparing 27.7 mM (SD ± 9.1), compared to the control, which took 150 hours to reach 38.9 mM (SD ± 10.5). This raises the question of whether improved yield could be achieved by addition of a glucose pulse prior to 48 hours, or whether the remaining glucose was a result of the loss of TER activity or indeed this was simply a result of the initiation of the sporulation process with the fermentation having reached its natural end with cells exposed to toxic concentrations of solvent. A linear regression on glucose consumption during exponential growth showed that the consumption rates were 3.6 mM H^{-1} for the wild type control and 8.1 mM H^{-1} for the *ter*_FLAG strain showing the removal of glucose occurring at over twice the rate of the wild type - regression analysis produced a p-value of 0.0086.

Butanol productivity in the bioreactor compared to the flasks over the same period of early solventogenesis (9-26 hours) was for both the *C. acetobutylicum* control and the *C. acetobutylicum* + *ter*_FLAG comparable to the performance of the *C. acetobutylicum* + *ter*_FLAG strain in the 50 mL static flasks. Again, the wild type is much more productive in the bioreactors, however, after the 26 hours timepoint solvent production and acid re-assimilation stalls before resuming again at 75 hours, *C. acetobutylicum* + *ter*_FLAG, however, continues producing solvents and consuming acids at a high rate until the fermentation ends at 48 hours. This observation might again be understood though the context of the redox response regulator, Rex. However, the adjusted production dynamics in the pH controlled, stirred tank bioreactors suggests some environmental sensing system is playing a role, where in the flasks, pH gradients will negatively affect the progression of

solventogenesis, as it is well understood that solvent production is not only a mechanism for the maintenance of redox cofactor homeostasis, but also environmental pH regulation, and that solventogenesis is pH induced (Millat *et al.,* 2013). Given that TER activity will only disrupt redox co-factor homeostasis during acidogenesis by competing with BCD/EtfAB activity - limiting the electron flux towards hydrogen production, which is only active during acidogenesis - the fact that solvent production in the wild type equilibrates at around 48 hours - entering a second short acidogenic period before returning to solvent production - probably represents the normal fluctuating levels of NAD^+/NADH and redox co-factor homeostasis. However, the fact that *C. acetobutylicum* + *ter*_FLAG continues producing solvents through to the end of the fermentation at 48 hours probably represents the imbalance NAD^+/NADH caused by TER.

Product ratios were compared between reduced and non-reduced products to investigate if TER activity was disrupting the NAD^+/NADH ratio resulting in a shift towards reduced product. Interestingly this was not seen in any particularly pronounced way, except perhaps with the ethanol/acetone profile in early solventogenesis, and the butyrate/acetate at 48 hours, which is actually more representative of re-assimilation rates. This was unexpected, considering that acetone production is abolished in batch cultures supplemented with methyl viologen or iron limitation (Peguin and Soucaille, 1994). Because CTFA/B activity is required for re-assimilation utilising acetoacetyl-CoA from the central pathway and producing acetone as a by-product, a shift towards reduced product formation will not occur without a sufficiently low NAD^+/NADH ratio, which demands more carbon from the central pathway for regeneration of NAD^+ in order to move closer to thermodynamic equilibrium. Arguably the competitive activity of TER in the presence of BCD/EtfAB is not capable of generating a sufficient shift in NAD^+/NADH ratio to affect product ratios towards alcohologenesis, where otherwise a failure to fully re-assimilate acids and a reduced acetone yield would have resulted. For both wild type and TER_FLAG

fermentations the molar acetone yield is the approximate sum of the molar butyrate and acetate yield, re-assimilation by moles are comparable between the two strains, though they differ in their dynamic rates.

4.2.4 Confirmation of genomic stability, expression and enzymatic activity

Further to initial PCR results and antibiotic resistance phenotypes demonstrating gene insertion at the *thlA* locus and subsequent plasmid loss, a study was carried out to investigate whether or not the insertion was stable during fermentation, in order to ensure that the fermentation had not ended due to genomic instability and loss of activity. Samples were taken from a triplicate repeat bioreactor fermentation at 72 hours which were subjected to three analyses: 1) a simple antibiotic resistance assay, to determine genotype stability, 2) a western blot analysis to confirm expression of FLAG tagged *ter* and, 3) an enzyme activity assay to confirm that the expressed gene is producing active enzyme.

4.2.5 Genotyping stability assay

Cells from triplicate fermenters for wild type control and the *ter*_FLAG expressing strain were sampled at 48 hours and serially diluted 1 mL in 10 mL down to 10^{-6} in PBS, 100 μL of each dilution was plated onto CGM, CGM^{Tm} and CGM^{Em}. Plates producing colonies in the range of 30-300 were selected for counting, all of which were produced by the 10^{-3} dilution. CGM only was used as a positive control, and as expected both cultures produced a comparable number of colonies on this media. The CGM^{Em} plates were used to test genotype stability for the *ter*_FLAG knock in which was inserted along with the *ermB* gene as a selection marker. This resulted in no growth for the wild type strain and the *ter*_FLAG strain demonstrated similar growth to on the CGM positive control media, indicating that the knock in locus was stable throughout the population. Finally, CGM^{Tm} was used as a negative control, to ensure the experimental results were not due to contamination with a different plasmid harbouring strain with *catP* selection, or due to an error in plasmid curing at an earlier stage, a negative result showed that this was not the case for either scenario (figure 30).

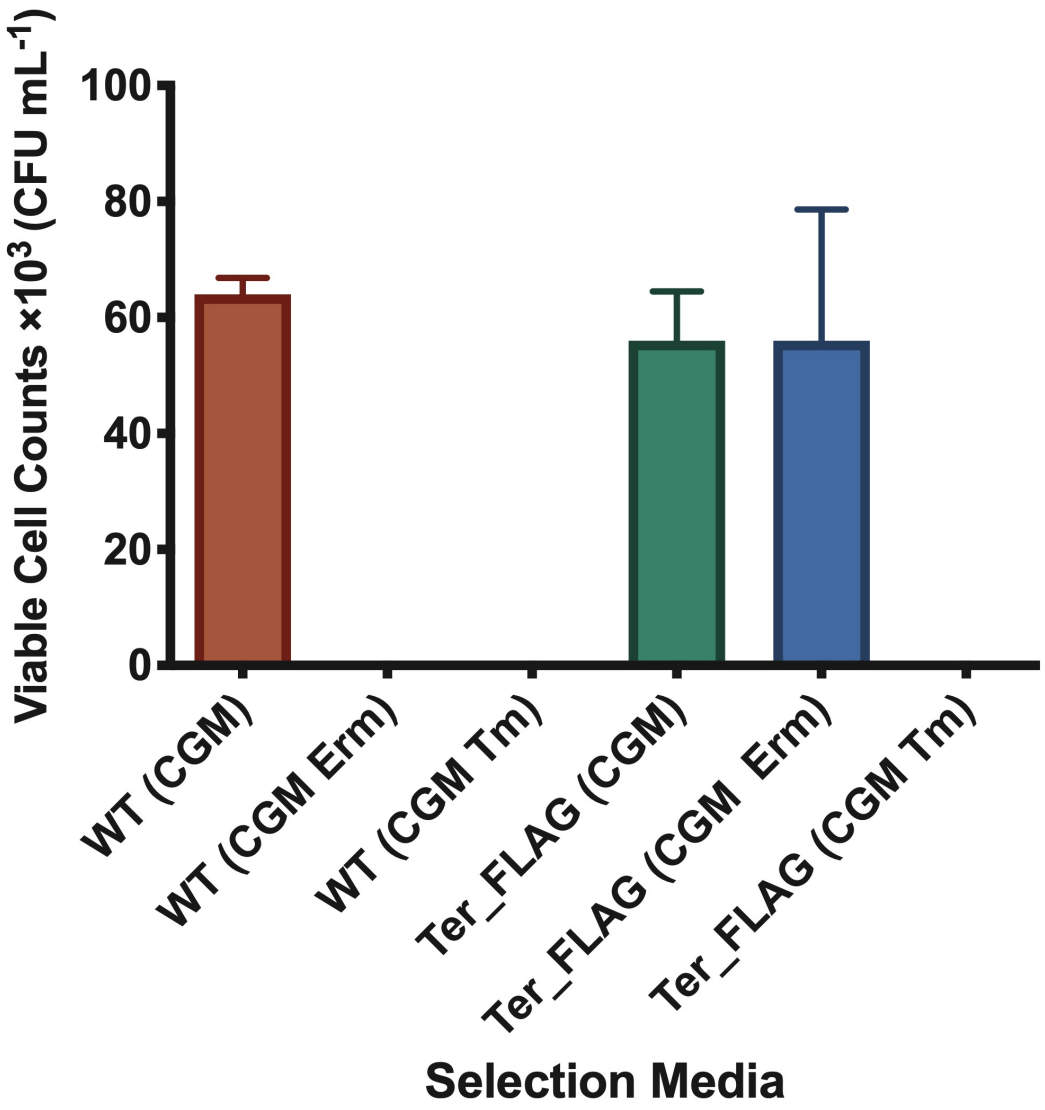


Figure 30: A stability study from 72 hour fermentation samples to test the stability of the insertion at the *thl* locus. Results show cell counts from serially diluted samples normalised by OD_{600nm} plated on selective media. Resistance to erythromycin (Erm) demonstrates *ermB* expression driven by $_{p}$*thl* indicating insertion stability at 72 hours (Ter_FLAG). Sensitivity to thiamphenicol (Tm) confirms that initial plasmid curing was successful. *C. acetobutylicum* ATCC (WT) was used as a negative control strain. Non selective media was used as a positive media control.

4.2.6 Trans-2-Enoyl-CoA reductase enzyme assays

Enzyme assays were initially carried out as a simple qualitative check for activity, in order to confirm that the *ter*_FLAG was expressing and the enzyme was active at the time that the strain ceased production of butanol in the *C. acetobutylicum* + *ter*_FLAG strain, *C. acetobutylicum* ATCC 824 fermentations served as the negative control. 10 mL cultures were sampled from triplicate fermenters at 48 hours for both strains and the cells were resuspended in 1 mL 2 × reaction buffer and lysed stirring on ice by sonication before being diluted 1:1 with NADH and crotonyl-CoA and assayed at 340 nm. The enzyme was assayed for crotonyl-dehydrogenase activity and cell lysates showed good activity (figure 31 & table 18) with the lysates from the *C. acetobutylicum* ATCC 824 producing undetectable levels of signal, indicating the native TER activity is substantially lower than the engineered strain. This is in congruence with the TER assays performed by Yoo *et al*., (2015) which also demonstrated undetectable levels of activity in wild type *C. acetobutylicum* ATCC 824 lysates. This rules out the possibility that strain degeneration was responsible for the end of butanol production observed at 48 hours in the bioreactor fermentations.

The results from the enzyme assay were compared to a metabolic flux analysis on *C. acetobutylicum* ATCC 824 in solventogenesis conducted by to Yoo *et al.,* (2015) and calculated according to dry cell weight (DCU). A conversion of OD_{600nm} data obtained in this study was made using conversion measurements obtained from Glazyrina *et al.,* (2010) in order to evaluate if the expression *ter* and subsequent enzyme activity should to be sufficient to replace the activity of BCD in the production of butyryl-CoA. Yoo *et al.,* (2015) calculated that 55.25 mM H^{-1} g DCW^{-1} intermediate traversed the C4 processing pathway (HBD, CRT, BCD). This is the equivalent of 0.921 mM min^{-1} $gDCW^{-1}$. TER_FLAG enzyme assay results were converted to from OD to DCW to produce a comparable figure: 1.092 mM min^{-1} $gDCW^{-1}$. This suggests that the expression and activity of *T. denticola ter*_FLAG polycistronically at the *thlA* locus provides adequate activity to serve as a replacement of BCD. Nevertheless,

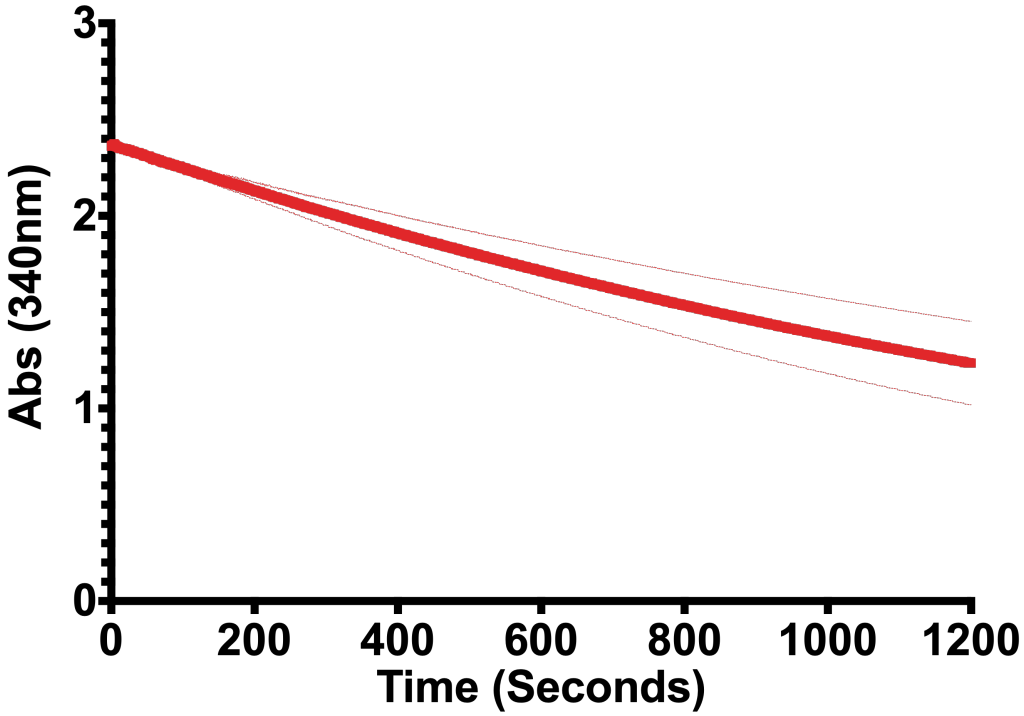


Figure 31: Ter_FLAG enzyme activity assay. Crotonyl-CoA conversion to butryl-CoA measured by consumption of NADH absorbed spectrophotometrically at 340nm. Assays performed in biological triplicate and the mean is represented by the central red line. The starting absorbance was normalised to the highest so that the line is representative of the rate of change. Standard deviation is represented by the flanking outer red lines.

Unit	WT	Ter_FLAG
Activity (mM min^{-1})	ND	0.048 ± 0.009
Cell Specific Activity (mM OD_{600nm}^{-1} mL^{-1} min^{-1})	ND	0.014
Dry Cell Weight Conversion (mM min^{-1} gDCW)	ND	1.092

Table 18: Enzyme activity results according to consumption of NADH measured at 340 nm.

these comparative figures were based on conversion values obtained from *E. coli* by Glazyrina *et al.,* (2010) and therefore some error should be expected. Ideally, a direct comparison of BCD and TER ought to be made, although this assay requires recombinant ferredoxin to be present in excess which was not readily available. Furthermore, comparing on these figures (based on activity alone) it appears that TER is competing with BCD, however, it cannot be considered dominant and that each enzyme will catalyse an approximately equal share of the substrate *in vivo*.

4.2.7 Western Blot Analysis

To confirm that the activity was related to the expression of the FLAG tagged *ter* gene fermentor samples were also subjected to western blot analysis (figure 32) using anti FLAG hrp conjugate monoclonal antibodies. A sample during the switch from acid to solvent production was taken at 10 hours and a second sample was taken at 48 hours to allowing for a semi quantitative visual comparison. Samples were normalised according to total protein concentration and the results of the western revealed a visually comparable level of protein expression for TER at the two time points. The wild type control lane did not produce a visible band, eliminating the possibility of a false positive, and a positive control lane produced a strong band at the expected Mw, plus two other unexpected bands. From this we can assume that the activity is consistent throughout the fermentation when expressed polycistronically at the *thlA* locus. The 50 kDa positive protein control from a commercial provider (Alpha Diagnostic Cat # FLAG16-R) demonstrated that the anti-FLAG western blot analysis was working as intended, however, there seems to be some evidence of dimerisation at 100 kDa and degradation at 45 kDa.

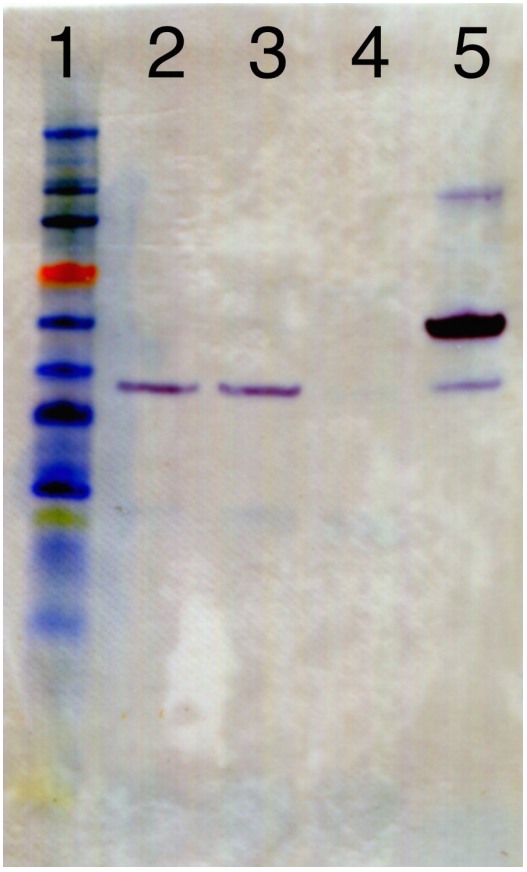


Figure 32: Mw marker is shown in lane 1: NEB Colour Pre-stained Protein Standard (broad range 11-245 kDa). Western blots on *C. acetobutylicum* ATCC824 +*ter*_FLAG fermentation broths taken at 8 and 48 hours (lanes 2 and 3 respectively) positive results conforming to the 44.8 kDa theoretical molecular mass and a wild type *C. acetobutylicum* ATCC824 negative control taken at 48 hours (lane 4) a positive FLAG tagged 50 kDa protein control (Alpha Diagnostic) is shown in lane 5, which appears to have formed a dimer at 100 kDa and a degradation product appears to be present at 45 kDa.

4.2.8 Production of a Δ*bcd* mutant in the *ter* expressing strain

The expression of the non-electron bifurcating, butyryl-CoA producing TER could in principle allow for the deletion of the *bcd* gene presuming the cell is able to compensate for the complete absence of election flux to molecular hydrogen at the BCD node in the ABE pathway. This can help to further characterise the role of BCD and also serve as a metabolic engineering strategy forcing an increase in reduced product resulting in alcohologenesis and an improvement of butanol yield.

4.2.9 Mutant generation by allelic exchange and *codA* counter selection

To build on previous attempts of obtaining a knockout mutant for *bcd* a new approach was taken using a suicide version of the *codA* counter selection method described by Ehsaan *et al*., (2016). The pMTL80150_RJH21Δ*bcd*/*etfAB* knock-out construct was produced using the Gibson assembly method (Gibson *et al.,* 2009) from a modified version of the vector described by Ehsaan *et al.,* (2016) with the pIM13 gram +ve replicon removed (supplied by Muhammed Ehsaan), this would ensure transformants harbouring single crossovers were selected for on the thiamphenicol selection media. The homology arms were also changed to flank *etfA, etfB* and *bcd* since the expression of all three genes leads to the formation of the electron transfer/dehydrogenase enzyme complex, it is rational to remove all of the associated genetic material required for this function. The plasmid construct was methylated using the usual pAN2 methylation system before transforming into the *C. acetobutylicum thlA*+*ter*_FLAG strain and wild type *C. acetobutylicum* ATCC 824 as a control. Transformations produced an unusually high number of thiamphenicol resistant colonies for a suicide system - >50 per plate from a 300 μL post transformation culture. Considering the previous method using the suicide double counter selection strategy (pMTL8-146) described in chapter 3 would only produce <10 colonies for the same transformation protocol, it appears that this method is

more efficient at producing single crossover colonies. From each strain, 48 colonies were picked into a 96 deep well plates containing 1 mL CBM + 100 μg mL FC in order to counter select for double crossovers. Cultures were incubated for 48 hours until all of the colonies had grown in the liquid media - many of which had grown in 24 hours. The cells were isolated in the 96 deep well plate by centrifugation, the genomic DNA was then extracted for analysis. The PCR results revealed an interesting pattern, where 18 colonies from the *C. acetobutylicum thlA+ter*_FLAG strain produced single bands of length consistent with a deletion of the *etfA/B bcd* genome section, a further 9 colonies produced a mixed banding pattern of consistent with both wild type and a deletion of the *bcd*/*etfAB* genes (figure 33). The wild type strain, however, only produced one colony with a single knockout band, and three others with a mixed knockout/wild-type banding pattern. Comparing the banding pattern between the wild-type control and the *C. acetobutylicum thlA+ter*_FLAG strain suggests that the knock out is more feasible in the *ter* expressing strain, due to the presence of the butyryl-CoA producing enzyme.

Due to previous issues with plasmid curing all strains were patch plated onto both CGM and CGM^{Tm} agar plates in order to confirm plasmid loss by thiamphenicol sensitivity. All of the mutant strains generated from the *C. acetobutylicum thlA+ter*_FLAG parent strains showing a 1.5 kb PCR product demonstrated thiamphenicol resistance. Similarly, strains from the wild-type background showing a 1.5 kb PCR product demonstrated thiamphenicol resistance. In an attempt to cure the mutant strains and isolate a true Δ*bcd/etfAB* mutant, strains were passaged on CBM + 100 μg mL FC agar plates and plasmid curing was monitored on CGM^{Tm}. Plasmid curing required 5-6 passages, after which only 16 of the original 29 strains showing a 1.5 kb deletion band had survived from the *C. acetobutylicum* + *ter*_FLAG parent strain, while only one of the strains showing a deletion band in the wild-type parent strain had survived. The remaining 16 strains were grown in 1 mL 2xYTG overnight and gDNA was prepared for PCR analysis (figure 33). This analysis

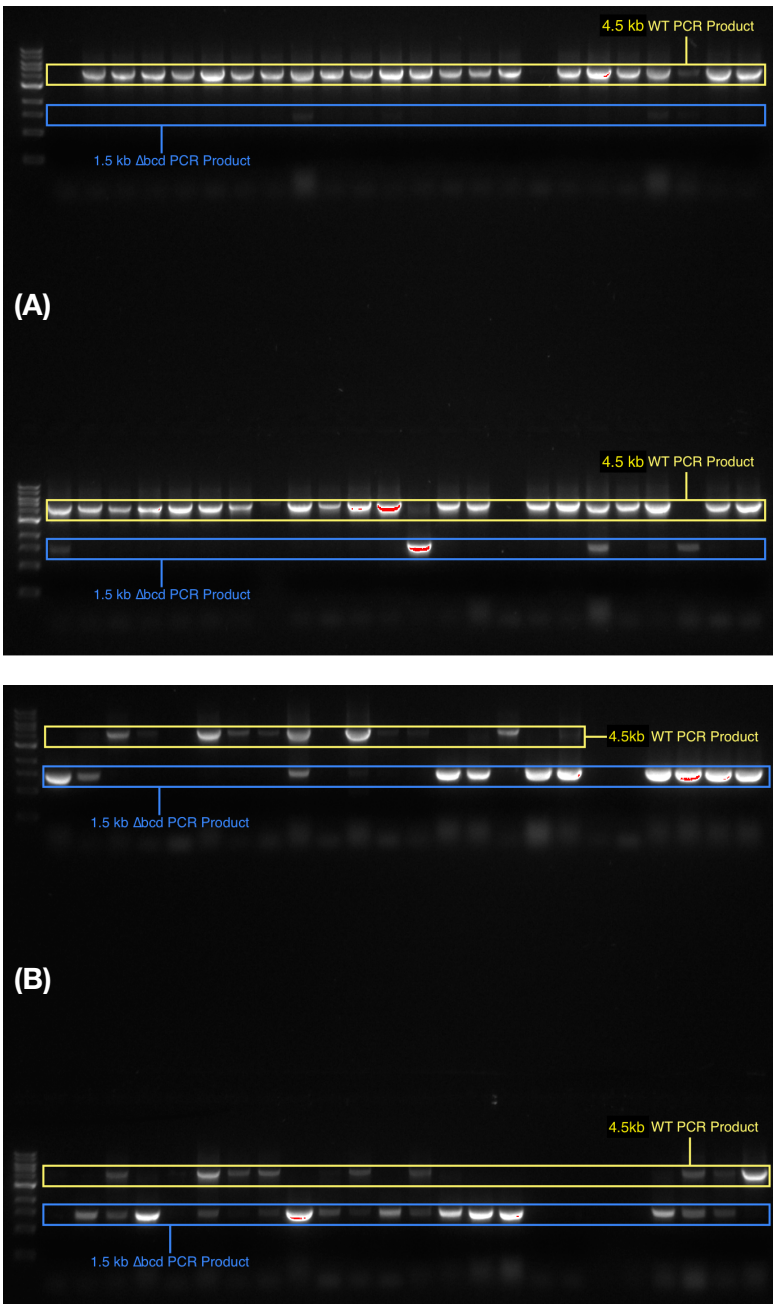


Figure 33: PCR screening analysis of 96 cultures after transformation with the pMTL80150-RJH21Δ*bcd/etfAB* plasmid using primers flanking 1 kb upstream and downstream of the target genes. Results from transformation into control strain *C. acetobutylicum* ATCC 824 after overnight growth in CBM + FOC liquid media shown on the (A). Results from transformation into *C. acetobutylicum* ATCC 824 + *ter*_FLAG shown on the (B). The 5 kb PCR product is indicative of a wild type genotype and a 1.5 kb PCR product is representative of a in-frame deletion of the target *bcd/etfAB* sequence.

showed that all of the remaining strains had reverted back to wild type with the exception of one. Again PCRs using primers binding sequence internal to the *bcd* CDS were used to investigate the possibility of VIE and again these produced a positive result for all reverted strains as well as the strain which maintained the 3 kb deletion at the *bcd*/*etfAB* locus had reverted back to a wild-type genotype with the exception of one strain, still showing a 1.5 kb band at the *bcd* locus. A second PCR was used to detect the presence of *bcd* using internally binding primers, and all of the remaining strains produced a 1.1 kb band indicating the presence of *bcd* in the genomic DNA.

Given the fact that 16 of the 48 strains showed clear evidence that a in-frame deletion had occurred at the *bcd/etfAB* locus, and given the clear positive distinction between the *C. acetobutylicum thlA*+*ter*_FLAG and the wild-type, the subsequent failure to isolate a mutant was an unexpected and disappointing result. On analysis, these results suggest that there is strong selective pressure for maintenance of electron flux to hydrogen production at the *bcd* node in the pathway, raising the question of whether or not it is actually possible to engineer the pathway in the desired manner to achieve alcohologenesis.

4.2.10 Mutant generation by CRISPR Cas9 selection for in-frame deletions targeted at the *bcd/etfAB* locus

Given that every attempt to knock out *bcd* using counter selection in-frame deletion methods had resulted in a VIE, an alternative approach using CRISPR Cas9 selection was taken. This approach should reduce the chances of a VIE occurring due to the targeted nuclease activity serving as the selection method, any cell containing the target gene, whether at the original loci or not, ought to be selected against using Cas9 nuclease. Development of a CRISPR/Cas9 platform vector for use in *Clostridium* species was in the prototype stages among several researchers at the SBRC at the time this work was conducted and several approaches to CRISPR Cas9 in-frame deletion were explored.

4.2.11 A frame shifted Cas9 putative nickase

The first vector used was constructed by Dr. Peter Rowe. This vector, originally harbouring native *S. pyogenes cas9*, was understood to contain a frameshift mutation resulting from an adenine insertion at base 133 generating a downstream stop codon at 159-162 bp, causing premature truncation of *cas9* translation, and also creating an alternative reading frame with a start codon at base 67, allowing for the expression of the complete Cas9 nuclease with the exception of the RuvCI nuclease domain lobe (Jinek *et al.,* 2014). This excision of the first nuclease domain results in a Cas9 possessing a nickase activity (Cas9n) - classically, nickase activity is the result of a D10A mutation (Ran *et al.,* 2013). A nickase is incapable of producing double strand breaks when only a single guide is present, conferring reduced toxicity.

A cloning protocol was followed according to a method developed by Dr. Daphne Groothuis. Homology arms were amplified from the pMTL-80150_RJH21Δ*bcd/etfAB* plasmid using primers to introduce the AsiSI and AscI restriction sites at the 5’ and 3’ ends, respectively. An algorithm provided by Benchling was used to generate gRNA

cassette sequences (table 18), cassettes were synthesised using primer dimers and included SalI and AsiSI restriction sites at the 5' and 3' ends, subsequent PCR amplification of these primer dimers produced the cloning fragments. Three variations of gRNA targets were selected from the top ranked list of results produced by the Benchling algorithm. The construct pMTL83140-RJH40 (P_{thl}_cas9n-pJ23119_gRNA) targeting *bcd* (figure 34) was generated after digesting the vector backbone with AscI and SalI, the homology arms with AscI and AsisI, and the guide RNA template with AsisI and SalI followed by a triple ligation reaction. To isolate correctly assembled constructs, colonies were screened by sequencing of PCR products generated using primers bridging the restriction sites.

This vector differs from the one initially used by Dr. Peter Rowe as it included the strong synthetic constitutive promoter P_{J23119} driving the transcription of the gRNA template instead of the xylose inducible promoter P_{araE}, as the efficacy of this promotor had been demonstrated by Dr. Daphne Groothuis in *C. botulinum.*

A second plasmid, pMTL83140-RJH41(pthl_cas9n-pJ23119_gRNA) targeting *pta* was generated as a control. This gene was selected as *pta* has been a consistently obtainable mutant target in *C. acetobutylicum* (Jang *et al.,* 2012; Lehmann *et al.,* 2012). Homology arms were generated in the usual fashion by amplifying 1 kb target flanking regions together with an overlapping complementary region for splicing overhang extension PCR (SOE), and AsiSI and AscI were introduced at the 5' and 3' ends. Template gRNA was generated as described for the *bcd* construct.

Plasmid DNA was prepared for transformation by methylation in the typical manner and then transformed into *C. acetobutylicum* + *ter*_FLAG, and transformants were plated onto CGM^{Tm} and incubated for 48 hours at 37˚C. A total of 24 colonies of each strain were picked into 1 mL 2×YTG and incubated overnight for preparation of genomic DNA. PCR analysis of the genomic DNA was performed in order to screen for mutants. The PCR results screening for mutants returned PCR products

Target	Position	Strand	Sequence	PAM	Specificity Score	Efficiency Score
bcd/etfAB	2175	-1	cgaagcaaga gcttacatga	agg	49.97	69.37
bcd/etfAB	224	-1	gttgataagg accttcaagt	agg	50.00	68.77
bcd/etfAB	1112	-1	actaaagaag ttaaaggaca	ggg	49.92	67.46
pta	990	1	actgttgttc aggctcaaag	agg	50.00	69.86
pta	944	1	atcaagaggc tgtagctcag	agg	50.00	65.18
pta	270	1	tatgaactta gaaaacacaa	agg	49.97	64.97

Table 18: Templates for Cas9 guide RNA were designed using the Benchling (benchling.com) gRNA design tool. Guides were ranked according to their efficiency and the top 3 variants were selected for use. Specificity score is a score indicating the on-target specificity - scores over 50 are considered to be good guides. Efficiency score indicates on-target efficiency where 60 or over is considered to be a good guide. Since all scores for specificity were 50 or below guide sequences were sorted by efficiency and the top 3 were selected.

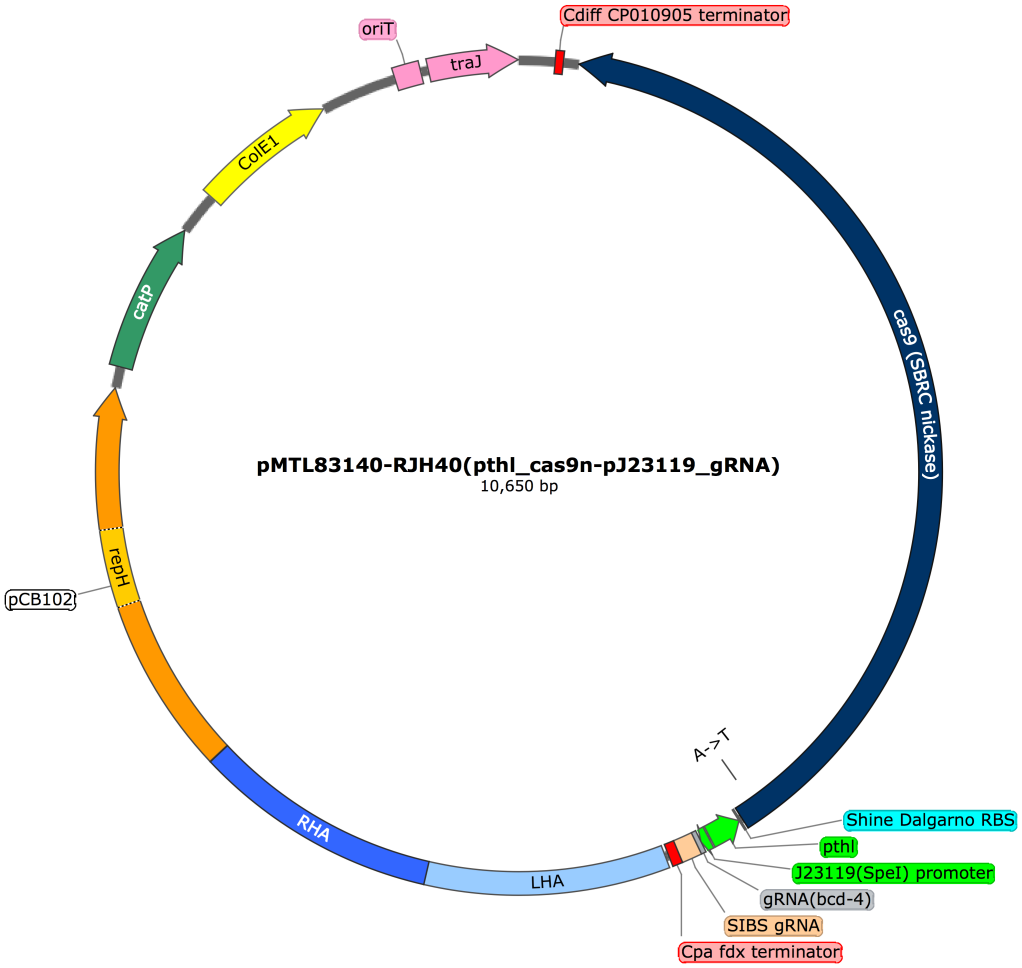


Figure 34: The first iteration of a CRISPR-Cas9 in-frame deletion vector targeting *bcd/etfAB*. This vector utilised a gram +ve origin pCD102 for plasmid maintenance in *Clostridium sp.* and a high copy number gram -ve origin for efficient plasmid propagation in *E. coli*. The antibiotic selection marker *catP*. For transformation via conjugation *oriT* and *traJ* is present. The *cas9* (SBRC nickase) is the nickase isolated at the Nottingham SBRC containing an insertion at base 67. This is driven by the strong constitutive P_{thl} promotor from *C. acetobutylicum* and is flanked by a strong terminator from *C. difficile*. Translation is initiated by the classical shine dalgarno ribosome binding site. SIBS gRNA refers to the scaffold component and the gRNA (bcd-4) refers to the guide RNA template. The gRNA components were transcribed under the control of the strong synthetic constitutive promotor P_{J23119} and transcription is terminated by the *C. pastarianum fdx* terminator.

consistent with the wild-type genotype for both the strain harbouring the pMTL83140-RJH40(pthl_cas9n-pJ23119_gRNA) plasmid and the pMTL83140-RJH41(pthl_cas9n-pJ23119_gRNA) plasmid - meaning neither attempt to produce an in-frame deletion mutant for *pta* or *bcd*/*etfAB* was successful (figure 35).

Given that this vector configuration has been successful in targeting several other genes in other species of *Clostridium* the failure to produce a Δ*pta* mutant was

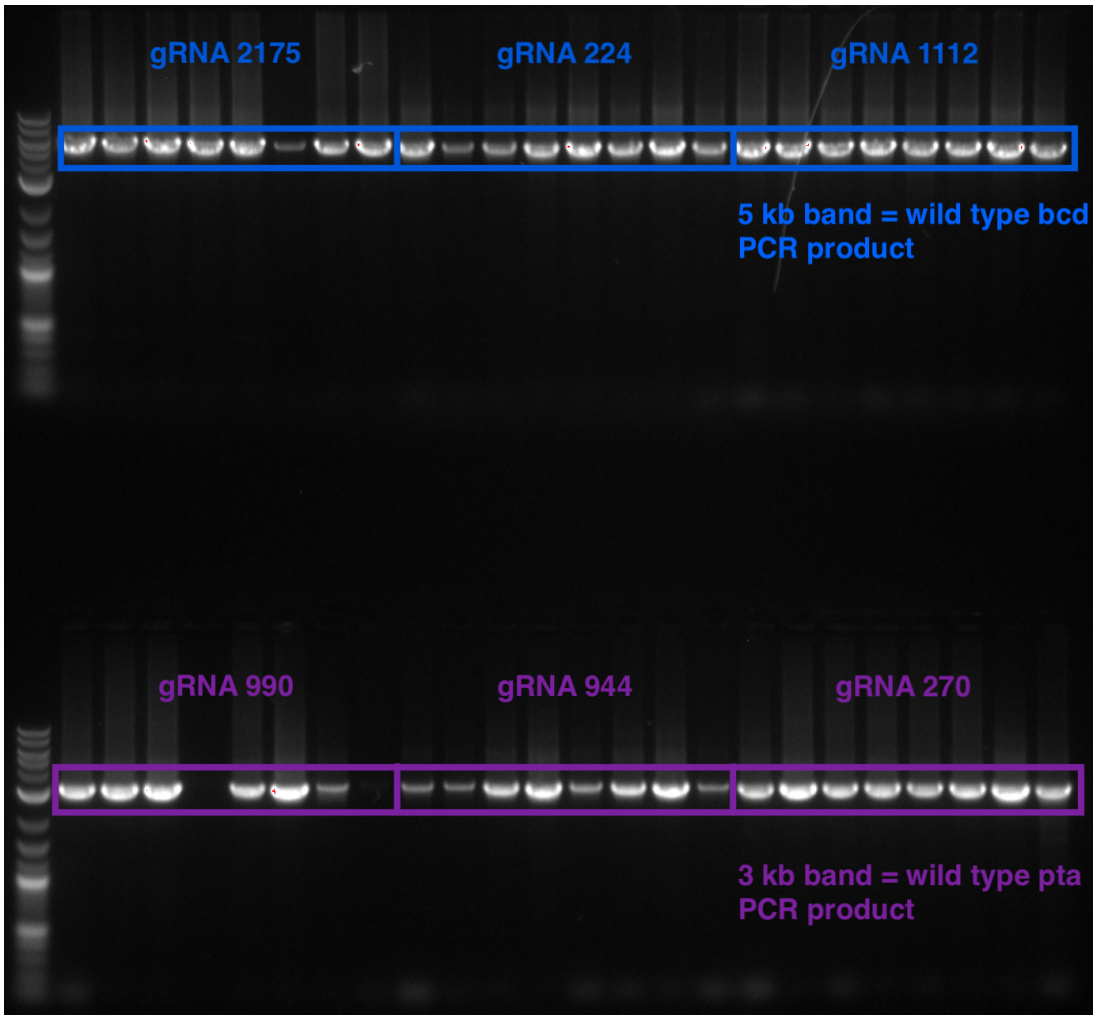


Figure 35: PCR screening for in-frame deletions at the *bcd*/*etfAB* loci after transformation with the pMTL83140-RJH40(Pthl_cas9n-pJ23119_gRNA) *bcd*/*etfAB*) (figure 14) and the pMTL83140-RJH41(Pthl_cas9n-pJ23119_gRNA) *pta* targeting in-frame deletion plasmids. All of the results show wild type sized bands indicating a failure to excise the target genes. Three different gRNAs were used - see table 18.

unexpected. It is worth noting that previous mutations to *pta* by Jang *et al.,* (2012); and Lehmann *et al.*, 2012 and were a result of group II intron insertion methods either ClosTron (Heap *et al.,* 2010) or Targetron (Liu *et al.,* 2015). It also worth nothing that *pta* is not an ideal control target, and although previous mutants have been obtained, an ideal control target would be one more readily obtainable than one coding for an enzyme in an ATP producing pathway. Such an example might be *ldh* coding for lactate dehydrogenase in the lactate producing pathway, or *alsS* coding for acetolactate synthase, in the acetoin producing pathway, as both pathways are responsible for producing small amounts of product and are relatively insignificant to the metabolism of *C. acetobutylicum*.

The ClosTron and Targetron methods both utilise mobile group II introns to inactivate target genes using a variable target region. Inserts are selected for using a retroactive antibiotic marker (RAM). Here a knockout was attempted using a nickase version of *cas9*, however, nickase is normally utilised with two complimentary guides one targeting the sense strand and one targeting the anti-sense strand producing a double strand break with some spacing region - where DNA is only held together by hydrogen bonding of the bases - ultimately resulting in a lethal break.

With this method, only one gRNA template was provided per vector, therefore double strand breaks could not take place. Furthermore, ligase activity from *ligA* will repair this single strand break. Although LigA is an NAD^+ dependant ligase (Uniprot - Q97JS8), and therefore this ligase activity is likely provide some selective power through stress to the maintenance of redox co-factor homeostasis - therefore the excision of the gene fragment by homologous recombination would be favoured in the case of benign mutations.

Additionally, it's reasonable to assume that single strand nicks will inhibit DNA replication and therefore slow the growth rate of cells with chromosomes containing the target DNA - serving as another selective force promoting excision of the target genes(s).

The fact that the frameshift in *cas9* resulted in a start codon which was 76 bp downstream from the RBS, will almost certainly result in very poor expression of the protein and consequentially, very poor activity (Chen *et al.,* 1994). In conclusion it seems that an alternative strategy should be taken to mitigate the toxic effects of Cas9 so that the native gene was available to provide the strong selective advantages associated with this targeted nuclease.

4.2.12 Development of a LacI repressed *cas9* CRISPR vector

To address the problems with the first generation CRISPR vector used in the previous experiment, a number of changes were made. After consultations with Dr. Peter Rowe and Professor Nigel Minton it was understood that the general problem with *cas9* cloning using an upstream strong constitutive promotor was that it is too toxic due to background non-specific nuclease activity. This results in disabling mutations occurring *cas9* to during the *E. coli* plasmid cloning amplification steps. Dr. Peter Rowe and Dr. Christopher Humphries expressed that this result had occurred after numerous attempts to obtain a mutation-free CRISPR vector. In light of these issues an alternative CRISPR vector was pursued.

Initially, the original facOID system developed by Willson *et al.,* (2016) was used as a foundation to generate a LacI repressed *cas9* CRISPR vector (see figure 36). The system employs the *lacI* repressor under the control of P_{laclq}, this promotor is based on the native $P_{lacI,}$ however, with the exception of a point mutation which results in 10 fold increase in transcription for tightly controlled repression in *E. coli* (Glascock and Weickert *et al.,* 1998). This promotor does not have affinity for *Clostridium sp.* sigma factors and therefore allows for constitutive expression of downstream genes when transferred to these hosts.

The P_{facOID} promotor system is built on the strong constitutive clostridial P_{fac} promotor, which has been subject to insertion of two symmetrical *lac* operators flanking the -35 and -10 SigA binding sequences. This results in repression by LacI in the absence of lactose or IPTG. A *C. acetobutylicum* codon optimised version of *cas9* was obtained from the culture collection placed by Dr. Wouter Kuit. The gene had been stored in a vector without an upstream promotor, minimising the risk of mutations occurring during *E. coli* amplification. This version of *cas9* was spliced using SOE PCR downstream of P_{facOID} with a Shine-Dalgarno RBS 9 bp upstream of the start codon, introducing NotI and XbaI restriction sites to replace the previous

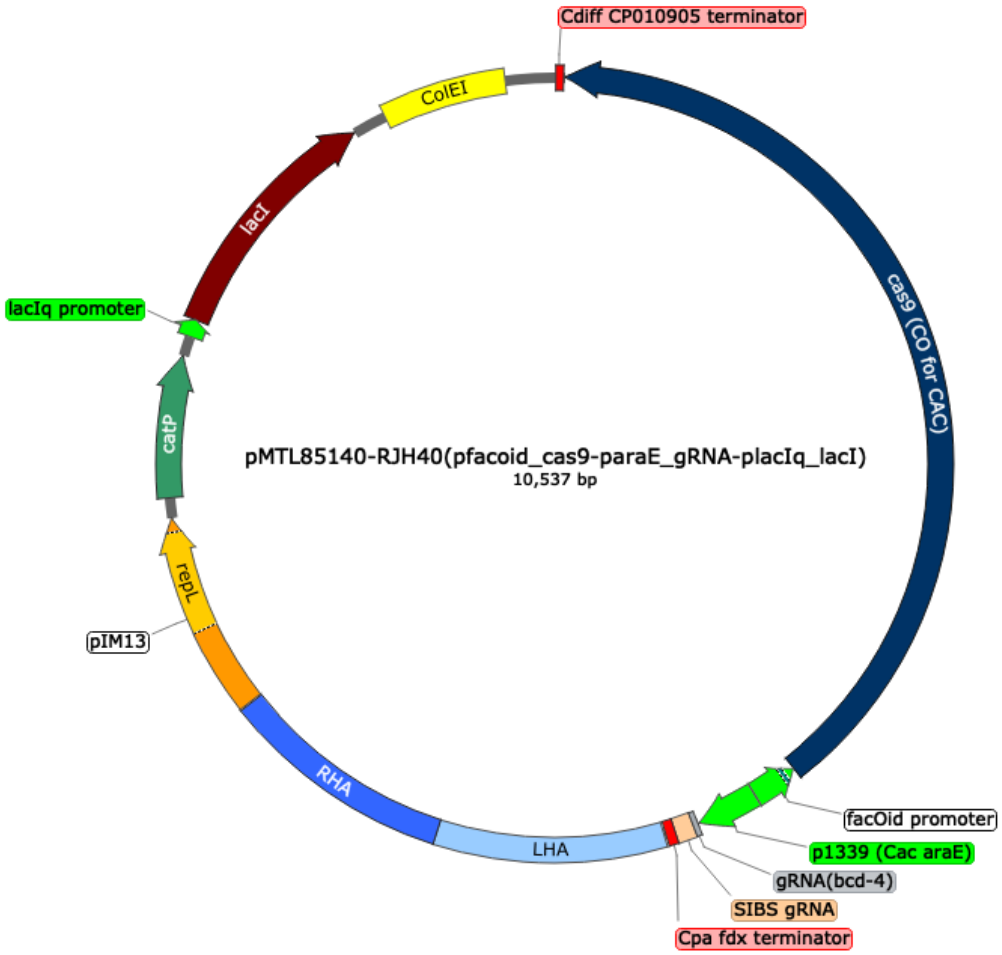


Figure 36: The second iteration of a CRISPR-Cas9 in-frame deletion vector targeting *bcd/etfAB*. This vector utilised a gram +ve origin pIM13 for low copy plasmid maintenance in *C. acetobutylicum* ensuring efficient plasmid curing. Selection marker *catP*, gram-negative origin, and homology arms remained the same as previous. The frame shifted *cas9n* has been replaced with the native *cas9* codon optimised for *C. acetobutylicum* which is expressed under the control of P_{facOID}. P_{facOID} is regulated by the transcriptional repressor gene *lacI* which itself is expressed under the control of P_{laclq}, which lacks activity for sigma factors of *C. acetobutylicum*. The gRNA promotor was changed to P_{araE} in order to allow for selection to take place after establishment of transformed colonies by transferring to inducing media containing xylose.

Cas9n and P_{thl} promotor of RJH40(Pthl_cas9n-pJ23119_gRNA). The repressor system, (*lacI* driven by P_{laclq}) was cloned in from pMTL-JH14_lacIQ_facOID described by Willson *et al.,* (2016) using the restriction sites PmeI and FseI. This also exchanged the pCB102 gram +ve replicon to the less stable pIM13, to ensure efficient plasmid curing.

The guide RNA template was placed under control of the xylose inducible P_{araE} promotor so that after transformation, transfer to plate media using xylose as a carbon source would allow for targeted nuclease activity after induction of the gRNA template. This was exchanged using the SalI and XbaI restriction sites. Again, two versions of the vector were constructed, one with appropriate 1 kb flanking homology arms and guide RNA template for *bcd*/*etfAB*: pMTL85140-RJH40 (pfacoid_cas9-paraE_gRNA-placIq_lacI) and a second construct serving as a control targeting *pta* pMTL85140-RJH41(pfacoid_cas9-paraE_gRNA-placIq_lacI) (see figure 16 for map of completed vector targeting *bcd*/*etfAB*). Sequencing of the completed vectors demonstrated an absence of mutations in the *cas9* gene.

After transforming into *C. acetobutylicum* + *ter*_FLAG plating onto CGM™ this method resulted in the failure to isolate transformants. Given that this was not the case when transforming into *E. coli* 5α suggests the un-mitigated activity of Cas9 in *C. acetobutylicum* + *ter*_FLAG under the strong constitutive P_{facOID} promotor was a lethal combination.

To address this problem, the promotor for *lacI* was exchanged from P_{laclq} to P_{thl} so that expression would remain strong and constitutive but in both *E. coli* and *C. acetobutylicum* - but with the requirement of lactose (or IPTG) as an additional inducer for expression of *cas9* in *C. acetobutylicum*.

The targeting components remained the same, including the same homology arms used previously, and the same gRNA components. However, the promotor for the gRNA was returned to P_{J23119} in order to simplify the induction protocol.

Transformation of the finished construct (figure 37) and the *pta* targeting control into *C. acetobutylicum* + *ter*_FLAG produced >50 colonies on CGM™, which were suspended in CBM then subcultured onto CBM agar with a variety of inducer concentrations: IPTG at 0.01, 0.1, 1, 10mM and lactose at 5, 50, 100mM. Interestingly, sub-cultured colonies took around 3 days before visible growth appeared on the plate and a further 5-7 days before they were sufficiently mature to enable inoculation into liquid media for gDNA prep - *C. acetobutylicum* typically forms healthy sized colonies within 18-24 hours after streaking onto rich media. Again, neither construct targeting *bcd/etfAB* or *pta* were successful in excising the target genes (figure 38).

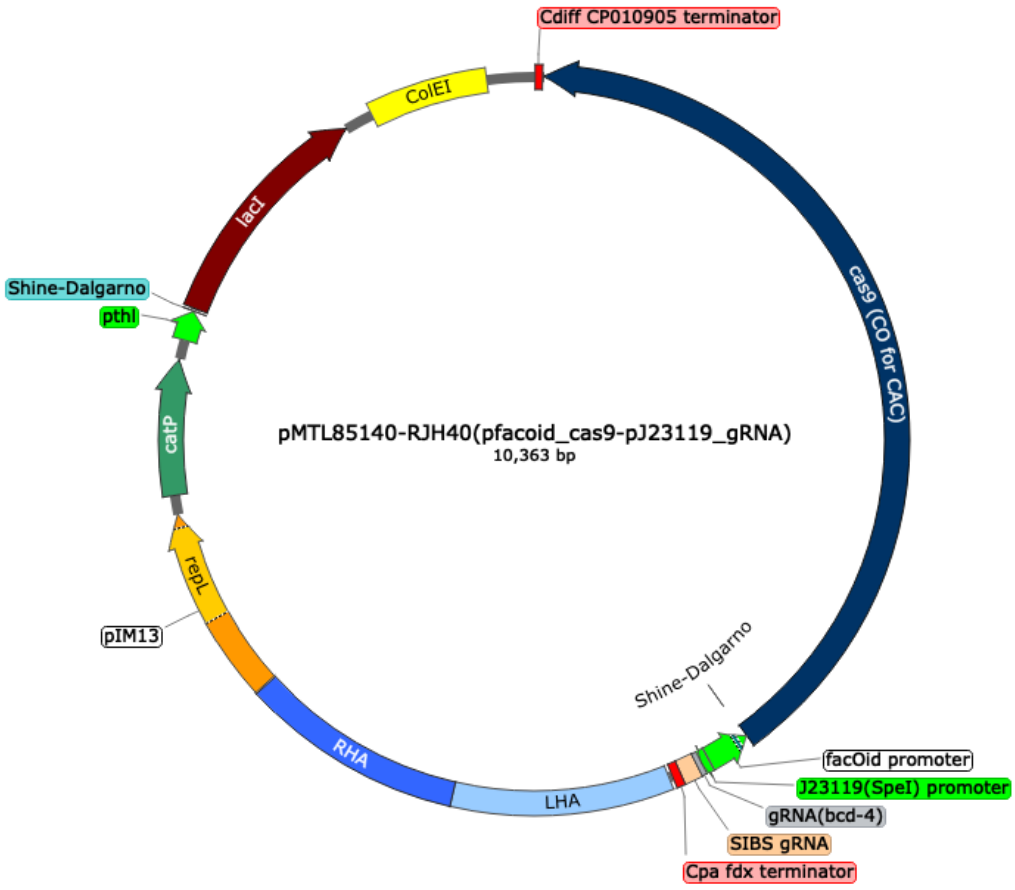


Figure 37: The third iteration of the CRISPR vector changed the promotor of the *lacI* to P_{thl} in order to ensure expression was occurring in both *E. coli* and *C. acetobutylicum*. The gRNA template promotor was changed to the strong constitutive synthetic promotor P_{J23119}.

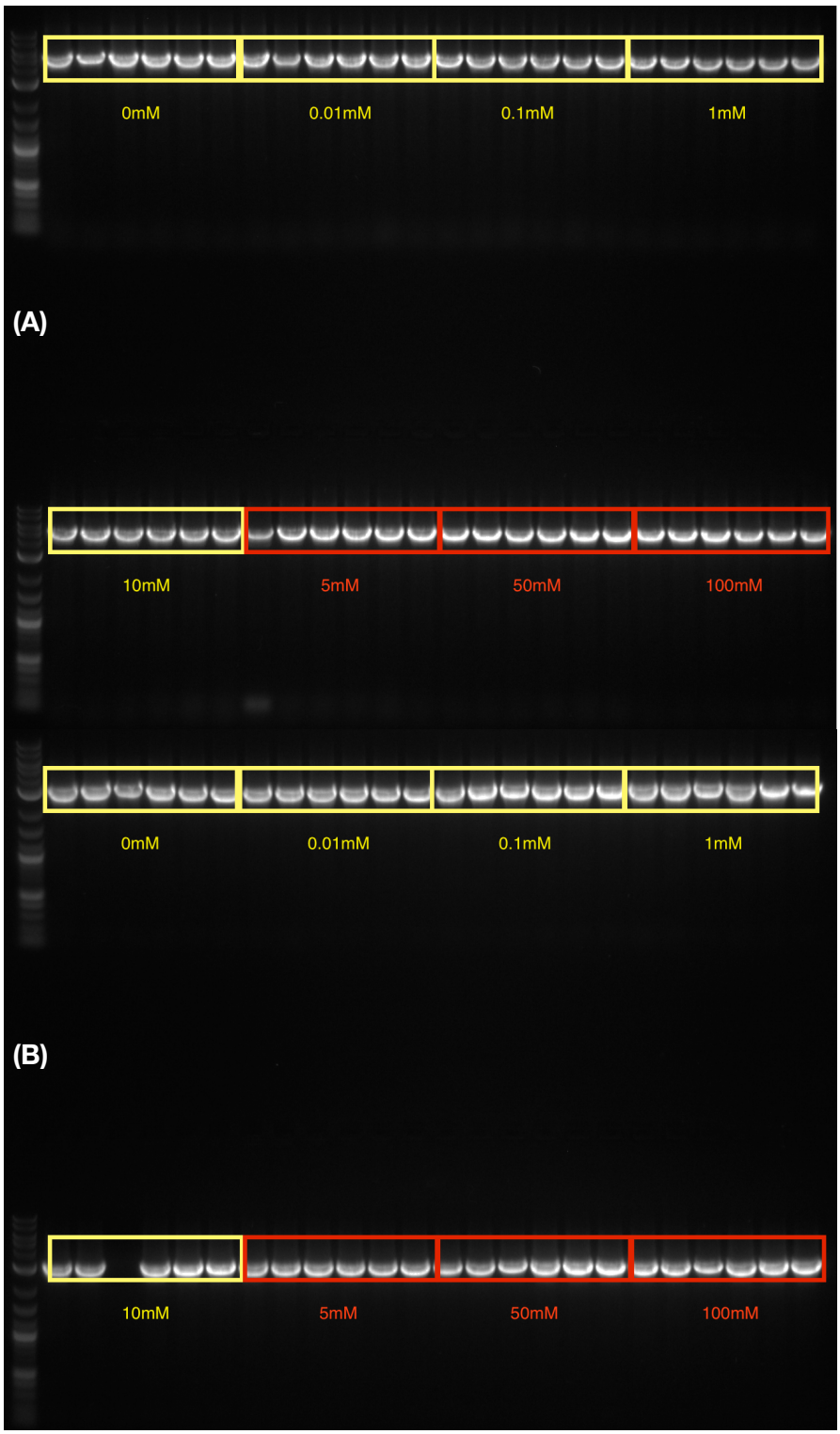


Figure 38: PCR mutant screening of the lactose inducible P_{facOID} *cas9* in-frame deletion construct targeting *bcd* (A) and *pta* (B). Two inducers were used; yellow bands indicate 0.01, 0.1, 1, 10mM IPTG. Red outlined bands indicate 5, 50, 100mM lactose. Bands all indicate WT genotype (cross ref with primers and anticipated fragment length)

4.2.13 Experimentation with a riboswitch vector in development

A colleague (Dr. Inés Cañadas) at the SBRC was also working on the problem of mitigating *cas9* toxicity. The approach undertaken utilised a riboswitch system which offered tight control of *cas9* expression (figure 39). In order to determine whether it was best to proceed with the LacI system or the riboswitch system an experiment was designed to test the background toxicity of each CRISPR vector used in this study compared to the riboswitch controlled *cas9*. A negative control plasmid pMTL8-5141 was used to compare which contained the pIM13 gram +ve replicon, the ColE1 gram -ve replicon, the catP resistance marker and a multiple cloning site. All of the targeting components (homology arms and gRNA template, scaffold, and promotor) were removed by restriction digest before overhangs were filled in using T4 polymerase and blunt ended plasmids were ligated together to generate the plasmid .

Plasmids are usually dosed by mass, however, in order to restrict variables which may directly impact transformation efficiency, plasmids were dosed according to molarity with each transformation receiving 2.5 mM of plasmid DNA. Transformations were carried out in triplicate and each transformation replicate was generated from the same competent cell prep for each vector tested. The test strain for this study was *C. acetobutylicum* ATCC 824. Transformations were plated onto CGM™ without inducers respective to the control system on the vector.

The results of this experiment (figure 40) demonstrated a clear benefit from using the riboswitch over every other variant used in this study. As expected the unmitigated expression of P_{facOID} *cas9* regulated by *lacI* under the control of P_{lacIQ} produced zero colonies, consistent with the previous experiment utilising this control method. The P_{facOID} *cas9* regulated by *lacI* under the control of P_{thl} produced a similar result to the *cas9n* vector. This was unexpected, and is likely due to impaired function of LacI in the *C. acetobutylicum* cytoplasm, which is typically maintained at

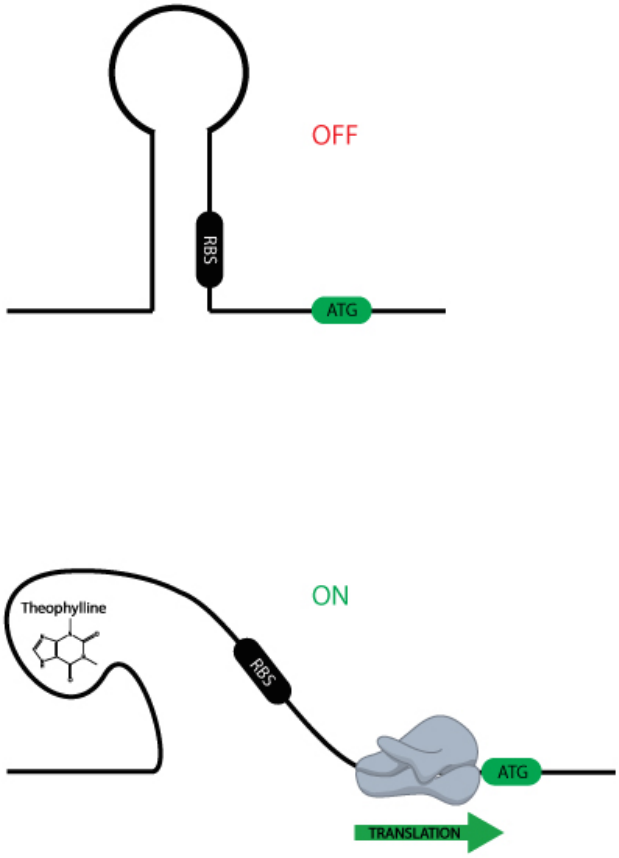


Figure 39: A schematic diagram to illustrate the mechanism of action of the synthetic theophylline activated riboswitch. The riboswitch is an RNA aptamer which in the absence of theophylline forms a secondary structure which sterically hinders the binding of ribosomes to the ribosome binding site. Addition of theophylline inhibits the formation of the RNA secondary structure, allowing access to the ribosome binding site and subsequent transcription of the downstream gene.

around pH 6.0 when external pH is between 4.5 and 6 (Huang *et al.*, 1985), whereas *E. coli* cytoplasm typically maintains an intracellular pH of 7.4-7.7 (Padan & Schuldiner, 1986). It is also worth considering the fact that CGM contains 2 g L^{-1} tryptone - a tryptic digest of the milk protein casein and it is therefore possible that this may result in the presence of trace amounts of lactose in the growth media, inadvertently inducing the expression of *cas9*. The riboswitch method, however, likely offers comparable transformation efficiency to the control plasmid.

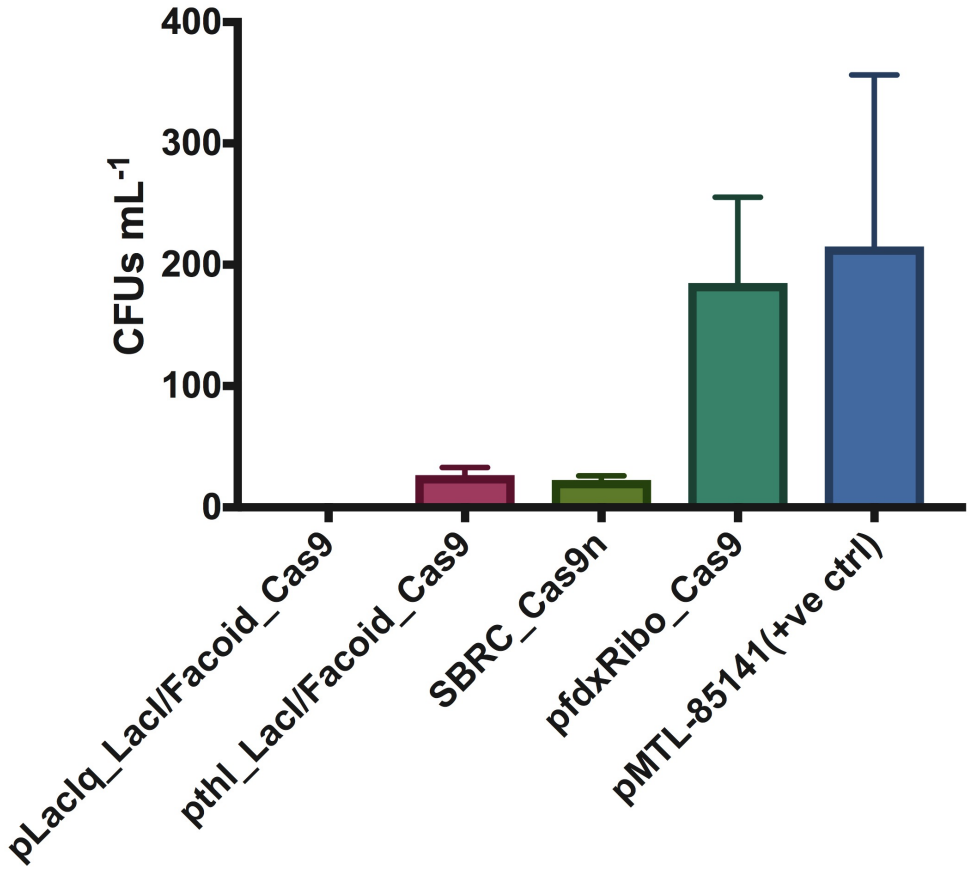


Figure 40: Transformation efficiencies of the four CRISPR/Cas9 vectors used in this study. Homology arms and gRNA were removed so that the relative toxicity of the *cas9* gene under the various mitigating mechanisms could be measured. pMTL-85141 serves as a positive control and contains the replicons pIM13, ColIEI, the selection marker *catP* and a multiple cloning site.

4.4 Conclusions

High expression of *T. denticola* TER seems to contribute to improved solventogenesis productivities but not final yield as the ratio of solvent products (acetone: butanol: ethanol) remains the same for the *ter* expressing strains as the wild type.

All analytical methods indicated that integration of *ter* at the P_{thlA} locus was stable and activity at 48 hours was comparable to the to the figures provided for the C4 intermediate synthesis pathway produced in a metabolic flux analysis performed by Yoo *et al.,* (2015).

Given that *ter* expression is deemed to be high enough to replace the conversion of crotonyl-CoA to butyryl-CoA normally performed by BCD, attempts were made using homologous recombination with a counter selection marker to generate an in-frame deletion mutant: Δ*bcd*.

Knockout attempts were to some degree successful in the *ter* expressing strain, where 17 double crossover daughter strains were obtained without a signal for the wild type genotype, whereas the same method only produced a single clean double crossover strain, indicating TER was favouring double crossover mutants of *bcd*. Nevertheless, as before curing the plasmid was difficult and re-streaking on CBM^{FOC} counter selective media resulted in either reversion to wild type or death of the strain with the exception of one. When internal primers were used to PCR genotype the apparent mutant this produced a positive result indicating that *bcd* had integrated at another locus on the chromosome - a so called vector integration event (Wang *et al.*, 2016).

Failure to cure the plasmid with the excised target fragment could either be down to the fact that hydrogenase activity at this position of the pathway is crucial for either redox co-factor homeostasis, preservation of the proton gradient, or both, or that perhaps activity of TER is actually not high enough and that errors in the calculations

to compare to the flux analysis were significant enough to provide a misleading comparison.

In order to prevent VIEs interfering with the ability to obtain a mutant CRISPR/Cas9 approach was explored, however, all prototype methods failed to reproduce the same results as the classical counter selection suicide plasmid method described by Ehsaan *et al.,* (2016). A final method has been produced by Cañadas *et al.,* (2019) and a Dr. Bunmi Omorotionmwan reports that this has been effective at knocking out *pyrE* (personal communication).

A new approach should include first knocking in *T. denticola ter* at the Δ*pyrE* locus however using the P_{facOID} integration vector, as prior cloning of *ter* indicated that unmitigated expression is toxic in *E. coli*. Furthermore, P_{facOID} is stronger than P_{thlA} (Willson *et al.,* 2016), which should ensure activities are significantly higher than flux requirements outlined by Yoo *et al*., (2015). Knocking in at the *pyrE* locus using the Willson *et al*., (2016) facOID method removes doubts that insertion at the P_{thlA} locus interferes with *thlA* expression.

Secondly, an attempt generate a non-flagellate strain in both the *ter* strain and the wild type should be made by deleting a key flagella gene component in order to limit dissipation of the proton gradient in the case where hydrogenase activity has a crucial role in maintain ΔpH. Following this, a Δ*bcd* strain should be attempted in the a *C. acetobutylicum* ATCC 824 wild type, a *C. acetobutylicum* Δ*flgX* a *C. acetobutylicum + ter*, and a *C. acetobutylicum +* P_{facOID} *ter*; Δ*flgX* strain (where Δ*flgX* refers to appropriate flagella gene deletion). The anticipation is that *C. acetobutylicum +* P_{facOID} *ter*; Δ*flgX* strain will be an alcohologenic strain with acetone production abolished and butanol yields and productivities improved.

Chapter 5: The application of an orthogonal inducible control system for enhanced butanol production

5.1 Introduction

In this chapter the application of a prototypic genetic expression circuit is investigated in *C. acetobutylicum*, where a lactose inducible transcriptional activator BgaR taken from *Clostridium perfringens* is applied to up-regulate the expression of a highly specific sigma factor (TcdR) the natural transcriptional regulator of the toxin A and toxin B genes from *C. difficile*. Heterologous genes downstream of P_{tcdB} in this expression system allows for their controlled expression, the utility of the application is explored and an increase in butanol yield and productivity is observed.

A control study expressing TcdR driven by the BgaR transcriptional activator produced a change in the solvent phenotype prompting a bioinformatics study into the background effects TcdR might be having in the *C. acetobutylicum* host - resulting in the proposition of a novel regulator of the *sol* operon.

5.1.1 Overview of inducible systems in synthetic biology

The most well-known inducible system to molecular biologists is the classic orthogonal, IPTG inducible, negatively regulated Lac repressor system employed in *E. coli*. The pET vectors originated in the 1980s with the publication of the seminal papers by Studier and colleagues. The system relies on the expression of a plasmid based gene of interest by the bacteriophage T7 RNA polymerase, expressing genes cloned downstream of a T7 promotor (Studier and Moffatt, 1986). Further work by Moffatt and Studier (1987) demonstrated the ability of T7 lysozyme to inhibit the activity of T7 RNA polymerase, which serves as a useful component in a synthetic expression system to inhibit the impacts of basal expression, expressed from the plasmid pLysS (Studier, 1991). Finally, the lactose repressor was engineered into the system allowing for easily inducible expression of the target gene. A lac operator placed in the *lac*UV5 promotor driving the expression of T7 RNA polymerase on the chromosome of *E. coli* BL21 DE3, and a second lac operator placed in the plasmid

based T7 promotor driving the expression of the target gene resulted in tight control of expression, which could be induced by IPTG - a non-metabolisable analogue of lactose. The most up to date version of this system is the now patent-expired, commercially available pET28 system originally developed by Novagen.

Clostridial expression systems are presently lacking in terms of choice and sophistication. An example of a lactose inducible positively-regulated system was developed for use in *Clostridium perfringens* (Hartman, 2011). The system itself was relatively simple in comparison to the *lacI*/T7 based system developed by Studier and colleagues, it was not orthogonal and lacked any protection against basal expression. The system utilised the regulatory, intergenic region for *bgaL* - which contained RNA polymerase binding regions for bi-directional translation. The left flanking *bgaR* gene was constitutively expressed and coded for the positive transcriptional regulator upstream of the target gene. The promotor driving the target gene required the lactose inducible binding of BgaR to positively regulate translation, and target genes were placed downstream of this promotor - flanking to the right of the intergenic region. The system works in the same way as the native system and is not the product of any real novel engineering besides toggling the gene of interest downstream of the BgaR regulated promotor. Furthermore, the BgaR activator is not responsive to IPTG, and therefore the metabolisable sugar lactose must be used instead. Nevertheless, the group did report a good dose related response to inducer concentrations, and a relatively low level of basal expression - barely quantifiable in the figures presented.

Another popular mechanism for heterologous control of gene expression which emerged from the field of mammalian cell biology was the deployment of the Tet repressor - of which there are many homologues. The bacterial TetR repressor based system was first demonstrated in mammalian HeLa cells by Gossen & Bujard (1992). The TetR repressor usually works as a negative regulator, where in the presence of the inducer TetR binds to the operator, inhibiting transcription. Gossen

and Bujard (1992) fused transcriptional activator VP16 from the herpes simplex virus to create the tetracycline controlled trans-activator (tTA), so that when bound to a tandem sequence of tet operators (*tetO*) placed in front of a minimal promotor, the human RNA polymerase II would initiate transcription. The tTA worked according to the TetR phenotype so that in the absence of the effector tetracycline (or doxycycline) the tTA could bind to the operator, addition of tetracycline would result in the release of the tTA protein and therefore this system was termed the Tet-Off system.

Conversely the Tet-On system worked in the opposite fashion where the tTA would bind to *tetO* and activate transcription in the presence of the effector. This system was developed by Gossen *et al* (1995) after random mutagenesis was used to identify which amino acid residues were necessary for binding to the operator. Specific mutations resulted in a *tetR* domain that would bind in the presence of the effector - this new transactivator was named the reverse tetracycline controlled transactivator (rtTA).

Tetracycline based heterologous inducible control systems have remained popular and have been adapted for a number of bacterial systems for metabolic engineering. Dong *et al*., (2011) utilised to TetR to engineer the first inducible heterologous expression system for *C. acetobutylicum.* In bacterial systems anhydrotetracycline (aTc) is the preferred effector due to its reduced toxicity and improved stability (Gossen & Bujard, 1992). Dong *et al*., (2011) demonstrated highly stringent control of basal expression by the placement of two *tetO* operators at the promotor site, one upstream, directly adjacent to the -10 binding site, and one in between the -35 and -10 binding sites. The plasmid based system placed the *tetR* gene under control of the strong constitutive P_{thl} and produced good dose related response to aTc induction, although not without impaired growth.

Soma *et al*., (2014) produced a more sophisticated system utilising both *tetR* and *lacI* to produce a metabolic toggle switch in *E. coli*. This system promoted the

flux of carbon from the TCA cycle and into the product pathway with the addition of a single inducer: IPTG. To prevent flux toward the citric acid cycle *gltA* was deleted from the chromosome, and replaced with a *tetO* mediated plasmid based version. The *tetR* gene was placed under control of LacI, as were the pathway genes for isopropanol synthesis. In minimal M9 media Δ*gltA* strains were unable to grow. The optimum induction point resulted in an increase in isopropanol yield from 13.7 mM for the control to 50.9 mM in the toggle switch flux controlled system.

5.1.2 The heterologous lactose inducible alternative sigma factor expression system for *Clostridium* spp.

Sigma factors in eubacteria are required for promotor recognition while bound to RNA polymerase. Cells typically have a sigma factor for general housekeeping purposes and a number of purpose specific sigma factors which recognise specific promotor sequences. Cells can efficiently respond to environmental changes - particularly stress inducing changes in a global manner by expressing a situation specific sigma factor (Gruber & Gross, 2003).

The σ^{70} family is the largest family of sigma factors and contains 4 diverse groups (Feklîstov *et al.,* 2014). Group I are known as house keeping sigma factors or RpoD type. Group II are similar in structure to group I, however, they are non-essential to the cells survival. Group III and IV are known as alternative sigma factors recruited by the cell when certain conditions require a specific, specialised response. Group IV are also known as Extra Cytoplasmic Function (ECF) sigma factors and regulate the cells ability to respond to changing environmental conditions - usually stress responses (Feklîstov *et al.,* 2014). Additionally, ECFs are known to auto-regulate, after initiation of transcription, a positive feedback loop between expression of the σ^{70} and binding and transcription initiation regulated by an upstream promotor site allows the cell to rapidly respond to environmental stimuli (Feklîstov *et al.,* 2014).

Sigma factors in solventogenic *Clostridium* spp. govern the progression of cells through the cell growth phase, into stationary and finally mediates sporulation. Alsaker & Papoutsakis (2005) used RNA micro array analysis to characterise the sigma factors responsible for mediating the transition through the physiological stages of batch growth in *C. acetobutylicum*. As is universal in Eubacteria, *C. acetobutylicum* possesses a σ^{70} - SigA for general housekeeping transcription. The sporulation program for *C. acetobutylicum* is similar to the model organism for sporulation, *Bacillus subtilis*, where the canonical σ^{F}, σ^{E}, σ^{G}, and σ^{K} mediate sporulation, following up-regulation by σ^{H}, after Spo0A phosphorylation (Al-Hinai, 2015).

Pathogenic strains of *C. difficile* possess a 19.6 kb chromosomal region known as the pathogenicity locus (PaLoc). The PaLoc region codes for two toxin producing genes *tcdA* and *tcdB* as well as three accessory genes *tcdR*, *tcdC* and *tcdE*. The expression of *tcdA* and *tcdB* is dependant on the expression of the group IV clostridial sigma factor TcdR, which is an alternative σ^{70}, which functions as the regulatory transcription factor for toxigenic expression of *tcdA* and *tcdB*, (Martin-Verstraete, 2016).

5.1.3 Aim of the study

This work sought to test the use of a prototype lactose inducible transcription factor to drive the transcription of *tcdR*, which would then activate the expression of genes placed downstream of the *tcdB* promotor for metabolic engineering applications. Since the TcdR/P_{tcdB} couple is the strongest promotor system available to synthetic biologists working in *Clostridium* spp. it was hypothesised that improved yield and productivity of butanol could be achieved under the principle that greater expression of pathway enzymes might lead to greater flux of carbon though this pathway.

Ideally, a dose dependant response would be observed and the regulatory control components would show to be inert in negative control studies. The system was originally designed and developed by Professor Nigel Minton, Dr. Ying Zhang, Dr. John Heap and Dr. Katrin Schwarz which was patented (Minton *et al.,* 2012), further development was carried out by Dr. Hengzheng Wang. The system has been used in *C. acetobutylicum* to produce CatP assay results suggesting it is the strongest Clostridial promotor system to date.

A fermentation of a control strain was used to test the phenotypic impact of the BgaR/P_{bgaL}; TcdR/P_{tcdB} control system and a follow-up genomic study was conducted to analyse the results.

5.2 Results and discussion

5.2.1 Construction of a butanol synthesis operon for enhanced butanol production in *C. acetobutylicum*

The genes chosen for expression under *bgaR/tcdR* orthogonal system were the native genes responsible for the metabolism of butanol in *C. acetobutylicum*: *thlA*, *hbd*, *crt*, and *adhE2*. These genes were obtained by PCR (with appropriate overhangs for Gibson assembly) from *C. acetobutylicum* genomic DNA and the fragments were assembled into a single operon driven by the P_{tcdB} promoter - with translation initiated by the RBS normally associated with the P_{tcdB} promotor (see figure 1). The genes were arranged as an operon to produce a 5.3 Kb transcript expressed in a polycistronic fashion. The insertion vector, contained the *bgaR* and *tcdR* regulatory components as well as homology arms for insertion into the *pyrE* locus, *catP* as a selection marker and the Gram-negative and Gram-positive origins ColEI and pIM13 were also included for plasmid maintenance in *E. coli and C. acetobutylicum*. The insertion vector for the orthogonal expression system was designed by Dr. Ying Zhang and assembled by Dr. Henzheng Wang at the University of Nottingham SBRC. The expression system containing the recombinant operon insert was cloned into the *C. acetobutylicum* Δ*pyrE* knock in vector described by Heap *et al.,* (2012) so that completed operons driven by the inducible promotor system may be expressed stably from the chromosome following insertion via homologous recombination. This system was inserted into a Δ*pyrE C. acetobutylicum thlA*+*ter*_FLAG strain at the *pyrE* locus to generate the *C. acetobutylicum* + *ter*_FLAG + *bgaR_tcdR* + C4 strain - expressing *thlA*, *hbd*, *crt*, and *adhE2* under the inducible control of BgaR_TcdR (figure 41).

The purpose of this strain was to test whether this inducible system was effective at enhancing pathway activity in a controlled and tuneable manner and to make maximum benefit of the NADH conserving activity uncoupled from the control of the

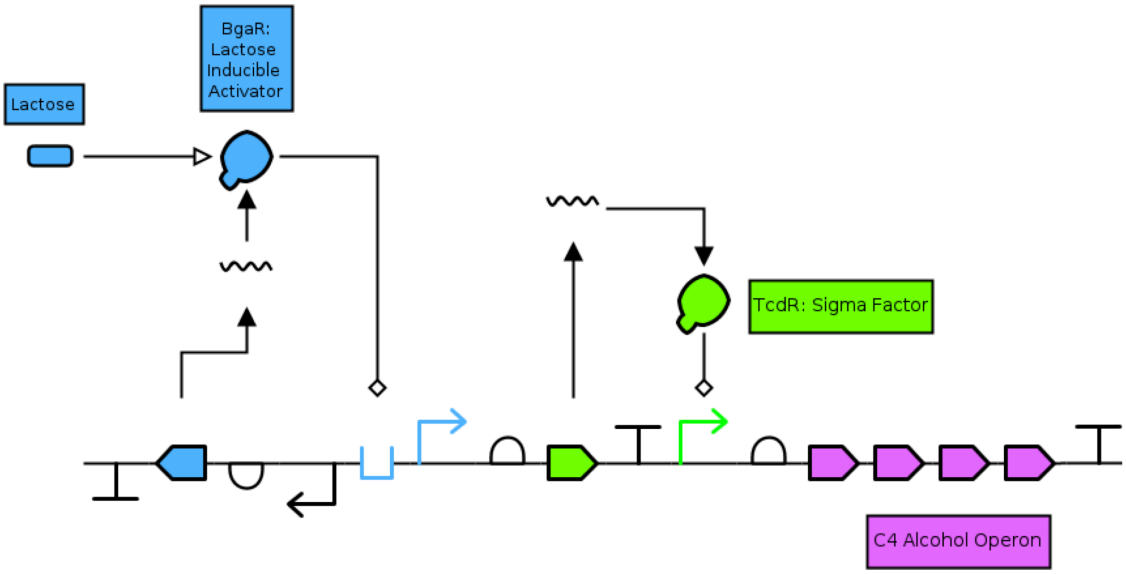


Figure 41: Schematic diagram of the synthetic, lactose inducible *tcdR* controlled gene expression system. P_{bgaL} contains two promotors (bent arrows), a left facing promotor constitutively expresses the lactose inducible positive regulator *bgaR* (blue CDS). In the presence of lactose, ligand binding to *bgaR* results in the binding to the left facing promotor P_{bgaL} and subsequent positive regulation. The left facing promotor drives the expression of the σ^{70} *tcdR* (green CDS), which complexes with RNA polymerase to bind P*tcdB* resulting in expression of downstream C4 alcohol genes. Half circles represent RBS, the RBS to right of P_{bgaL} is the native P_{bgaL} RBS from *C. perfringens*. The RBS to the right of *tcdR* is the native RBS found downstream of *tcdR* taken from *C. difficile*. The RBS to the right of P_{tcdB} is the native RBS found downstream of P_{tcdB} taken from *C. difficile.* The T symbols represent terminators. The terminator downstream of *bgaR* is a strong terminator taken from *C. botulinum.*The terminator downstream of *tcdR* is the strong *fdx* terminator taken from *C. pastarianum.* The terminator adjacent to *hydA* is the strong native *hydA C. acetobutylicum* terminator.

redox response regulator, Rex, where a bottle neck in normal expression levels may have resulted in an inability to make full use of the additional NADH as a result of TER activity. The expression of *thlA*, *hbd*, *crt*, and *adhE2* in native *C. acetobutylicum* ATCC 824 is understood to be under control of Rex, since they all contain Rex operators in their upstream intergenic promotor regions (Wietzke & Bahl, 2012). Although measuring expression by micro-array, *thlA*, *hbd*, and *crt* appear to be expressed at a constant rate during acidogenesis through to solventogenesis (Alsaker & Papoutsakis, 2005). Nevertheless, the thiolase expressed from the *thlA* gene has been reported to be a bottle-neck in the pathway due to the inhibitory effects of free CoA (Mann & Lütke-Eversloh, 2013), and the redox responsive disulphide bond formation regulating enzyme activity (Kim *et al.,* 2015) so it was considered particularly important to include this gene. In addition to the presence of *rex* operators in the promotor region and their potential detriment to butanol formation, the decision to include *hbd* and *crt* was also based on the reasoning that after up-regulating the expression of *thlA*, β-hydroxybutyryl-CoA dehydrogenase (HBD) and crotonase (CRT) may then become a bottle neck in the pathway to butanol. To encourage constant flux to butanol, the NADH dependant dual aldehyde/ alcohol dehydrogenase (ADHE2) was also included, which is otherwise also known to be regulated by Rex. Transcriptome analysis of this gene has demonstrated that regulation is growth phase dependant (Alsaker & Papoutsakis, 2005; Yoo *et al.,* 2015).

Following assembly, vectors were transformed into *E. coli* and screened by colony PCR for correct insert size. PCR fragments were purified using gel extraction and correct assembly was confirmed by sequencing. Once confirmed, pAN2 methylated plasmid DNA was transformed into *C. acetobutylicum* Δ*pyrE* + *ter*_FLAG and positive transformants were selected for using CBM^{Tm}. To cure the plasmid from the knock-in strain, eight colonies were re-streaked on CBM to select for uracil prototrophy and thiamphenicol sensitivity was monitored on CGM^{Tm}. Re-streaking

was repeated 4-5 times until a thiamphenicol sensitive, uracil prototroph was isolated. Insertion was confirmed using PCR with primers flanking the *pyrE* insertion site.

5.2.2 Phased induction fermentation of the butanol synthesis operon

In order to initially observe the capability of the lactose inducible operon to respond in a dose related manner, five concentrations of lactose were tested and the metabolic end products were subsequently measured. IPTG is not an active inducer for the BgaR transcriptional activator and so could not be used, therefore lactose was used as the inducer. Lactose was dosed at 0, 0.1, 0.25, 1 and 10 mM in 50 mL CBMS flask fermentations and was present from the inoculation. Solvent analysis by GC showed there was no clear, detectable, dose related response in butanol productivity or yield associated with variation in lactose concentration, although the upper limit of 10 mM seemed to have a negative effect on solvent synthesis, probably due to the metabolic burden of expression at this level of induction (figure 42).

The lack of dose dependant response for the lower inducer concentrations may not necessarily be a reflection of the capacity for the expression system to regulate protein expression dose dependently, however, it is very possible that the system is sufficiently leaky due to the promotors high strength, that in an un-induced state, basal expression is still sufficiently high to result in metabolic activity, resulting in an altered phenotype. Since none of the genes were tagged, protein expression could not be monitored directly by either ELISA or western blot. To identify if the basal expression of the pathway was comparable to native expression during acidogenesis, both the native gene and recombinant version might be FLAG tagged in both the wild type strain and the engineered strain. ELISA assays would provide a quantitative measure of how comparable the expression of the recombinant pathway.

Furthermore it would be necessary to obtain Δ*thl*, Δ*hbd*, Δ*bcd etfAB,* Δ*crt*, Δ*adhE2*, mutants in the background chromosome in order to obtain a more meaningful indication of the dose related response, with the background activity of this pathway subtracted from the tuneable recombinant pathway, however, this could

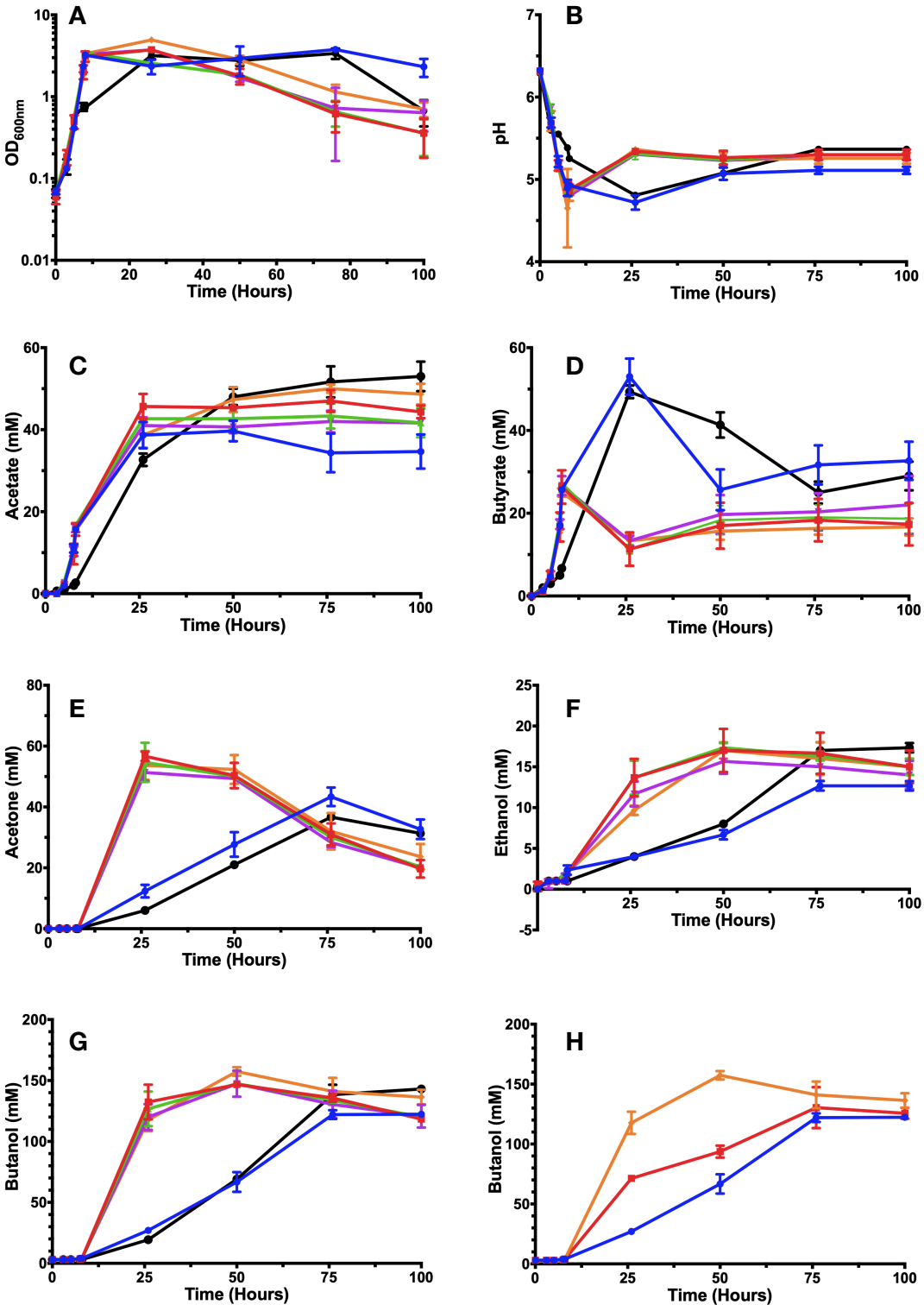


Figure 42: Inducer response of *C. acetobutylicum* + *ter*_FLAG + *bgaR_tcdR* + ButOH: ▪, 0mM; ▲ 0.1 mM; ▼ 0.25 mM; ◆ 1 mM; ● 10 mM; ●, *C. acetobutylicum* ATCC 824 (0 mM). Butanol from the *ter* strain is compared in H:●, *C. acetobutylicum* ATCC 824; ▪, *C. acetobutylicum* + *ter*_FLAG; ▲, *C. acetobutylicum* + *ter*_FLAG + *bgaR_tcdR* + ButOH.

not be considered until the issue of deleting *bcd* and *etfAB* has been resolved (see chapter 3 and 4).

In lieu of these results, previous CatP assay data was obtained from Dr. Henzheng Wang in order to gauge the level of basal expression in the un-induced state in order to gain insight into the contribution of basal expression. Data from the CatP assay, where the expression of the gene *catP* was placed under the BgaR/TcdR expression system showed there was a detectable level of activity in the un-induced samples (figure 43).

Despite the inability of this strain to produce a dose related response, the butanol yield and productivity in all of the batch fermentations was impressive. Comparing to the wild-type control, a 35 mM increase in butanol yield was observed, corresponding to a 29% improvement. Furthermore, the maximum final yield was reached at 50 hours, half the time taken for the wild type to reach the final yield, which represents a significant improvement in productivity. The improvement in yield and productivity was also superior to the *C. acetobutylicum* + *ter*_FLAG strain - which produces 27 mM less butanol, and required 25 hours more fermentation time to do so. The expression of the C4 pathway genes also resulted in a significant improvement when compared to the *C. acetobutylicum* + *ter*_FLAG strain, where the achieved maximum final yield was higher and was reached in half the time - as the improvements in productivity for the *C. acetobutylicum* + *ter*_FLAG strain were short lived and did not extend beyond 25 hours, and did not result in an improvement of maximum final yield.

It's interesting to note, also, that despite the presence of inducer in the fermentation medium from the start of the fermentation, inducing the duel aldehyde/alcohol dehydrogenase from inoculation, solventogenesis did not commence until the end of exponential growth, as is typical with batch growth for wild type strains. The explanation for this is not obvious. In the presence of TER, hypothetically, there should be enough NADH available to drive flux towards solvent production from the

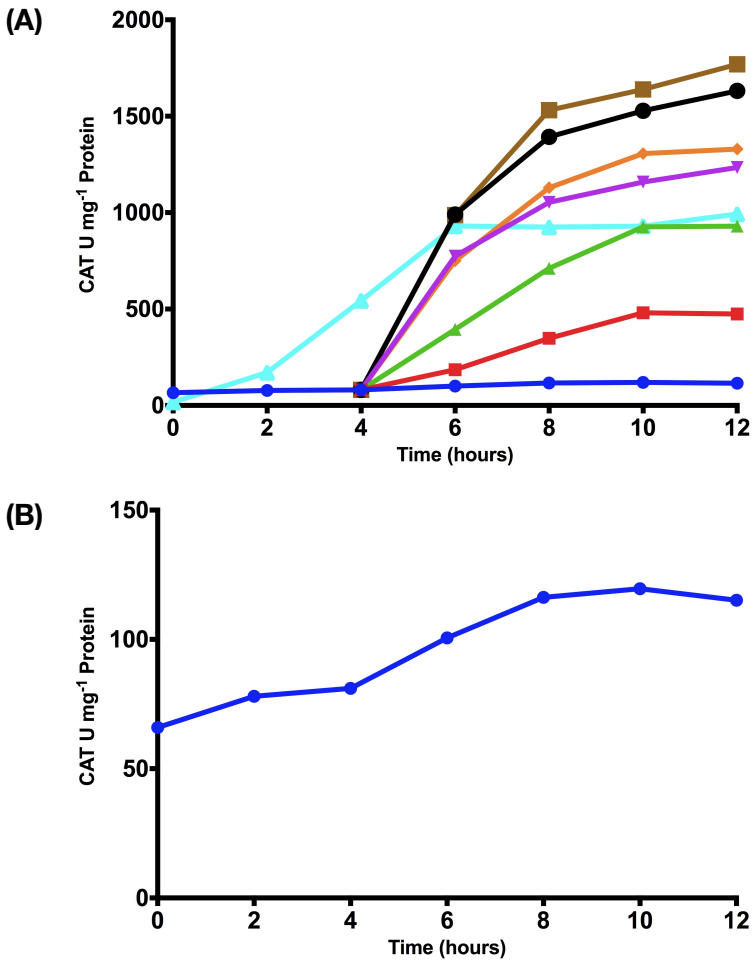


Figure 43: CatP assay results where the TcdR system was assayed using *catP* as a reporter in *C. acetobutylicum* grown on CGM. Various concentrations of the inducer (lactose) were used to obtain a dose dependant response profile (A). Leakiness of the system was examined by plotting the 0mM lactose inducer effect alone (B). ●, 0 mM lactose; ▪ 0.5 mM lactose; ▲, 1 mM lactose; ▼, 5 mM lactose; ◆, 10 mM lactose; ●, 50 mM lactose; ▪, 100 mM lactose; ▲, *C. pastarianum* P_{fdx} control.

Assay data produced and provided by Dr. Hengzheng Wang.

start of the fermentation.

The explanation as to why enhanced acetone production was observed in the *C. acetobutylicum* + *ter*_FLAG + *bgaR_tcdR* + ButOH strain might be provided by the fact that coenzyme A transferase is required for the re-assimilation of butyrate and acetate producing butyryl-CoA and acetyl-CoA via the transfer of the CoA group from acetoacetyl-CoA to produce acetoacetate - the direct precursor to acetone (Wiesenborn *et al.,* 1989). The increased availability of NADH and the enhanced expression of *adhE2* resulting in increased butanol synthesis rate would rapidly deplete butyryl-CoA leading to the rapid re-assimilation of butyrate (as observed) and the subsequent enhancement of acetone production.

The failure to produce a dose dependant response suggests that the un-induced fermentations were expressing significantly high amounts of pathway enzymes to produce an enhanced solventogenesis phenotype and further induction did not enhance flux through this pathway. This would suggest that the system is too leaky to serve any meaningful purpose as a metabolic engineering tool, at least when used in this context. It may be the case that a dose dependant response may be observable when measuring the metabolic products where use of a low activity enzyme is required.

To explore the effects of higher inducer concentration - ranging between 1 mM and 10 mM the experiment was repeated with 0, 1, 2, 4, and 6 mM, with *C. acetobutylicum* ATCC824 and the *C. acetobutylicum* + *ter*_FLAG strain as a control. All of the strains possessing lactose inducible *tcdR* system expressing the butanol operon displayed the “acid crash” phenotype (Maddox *et al.,* 2000) at every concentration of inducer, producing excessive amounts of butyrate and failing to achieve high concentrations of solvents, except for the un-induced which displayed a similar product profile as all of the fermentations for previous inducer optimisation experiment. The over production of acids was seen to a lesser extent in the 1 mM lactose fermentations, which were less productive during solventogenesis, producing

about 50% of the total solvents produced by the wild-type - this was in contradiction with the initial phased induction screen.

5.2.3 Repeat of the *C. acetobutylicum* + *ter*_FLAG + *bgaR_tcdR* + butOH flask fermentation with additional control studies

To test the reproducibility of the butanol overproducing *C. acetobutylicum* + *ter*_FLAG + *bgaR_tcdR* + ButOH strain, a repeat of the fermentation was conducted. The first major observation was that induction of the pathway resulted in poorer solvent production, likely a result of acid crash, where acetate yields reached the highest concentration before displaying a stunted shift to solventogenesis, similar to the 10 mM induction in the inducer optimisation experiment reported in the previous section. The un-induced strain produced the familiar enhanced phenotype, however, the strain did not reach as high a final yield as measured in the first fermentation; a maximum yield of 136 mM was measured at 48 hours rather than 157 mM as measured previously. Nevertheless, the performance relative to the WT control and the *C. acetobutylicum* + *ter*_FLAG control was still drastically improved. The discrepancy in performance might be due to variations in fermentation dynamics, given that sampling rate was approximately every 24 hours - a higher yield might have been observed slightly earlier or later in the fermentation (figure 44).

A further series of controls were performed to test the impact of lactose on the wild type *C. acetobutylicum* ATCC 824 and *C. acetobutylicum* + *ter*_FLAG control, and two additional control strains were also tested. Addition of 1 mM lactose to the wild-type and *C. acetobutylicum* + *ter*_FLAG fermentation media made no observable differences to the growth or acid and solvent profile, each strain performing as previously characterised. This provides assurance that lactose has no background regulatory activity acting on the acidogenic or solventogenic pathways under these fermentation conditions (figure 45).

Of the two control strains tested, the first was created using the lactose inducible TcdR driven expression system driving *bcd etfA/B* together with the rest of the native operon including *hbd* and *crt* with the addition of *thlA* and *adhE2* knocked into the

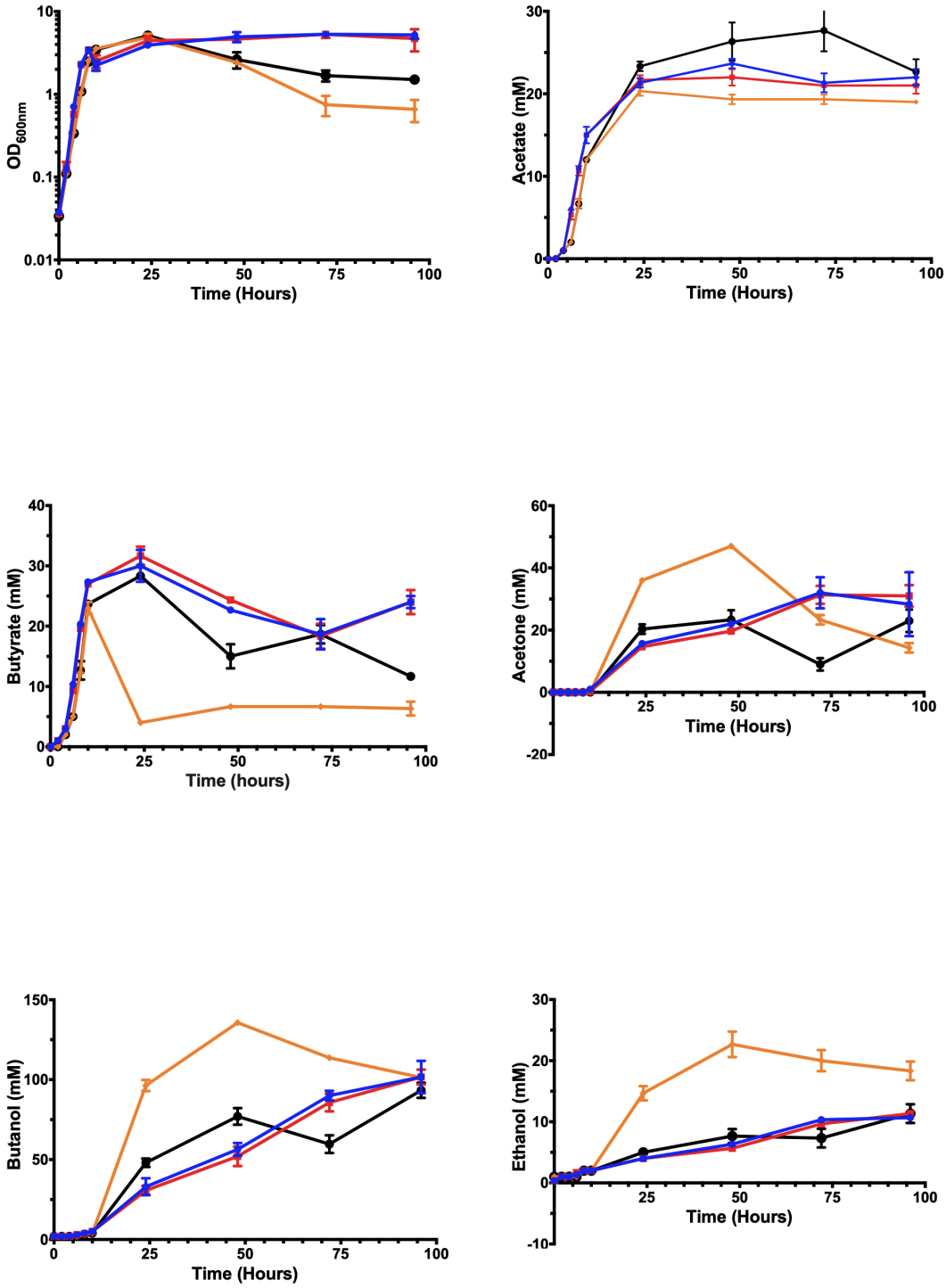


Figure 44: Fermentations to control for inducer effect on the wild type control. ●, Un-induced *C. acetobutylicum* ATCC 824; ▪ 1 mM lactose *C. acetobutylicum* ATCC 824; ◆, Un-induced *C. acetobutylicum* +*ter*_FLAG + *bgaR_tcdR* + ButOH; ●, 1 mM lactose *C. acetobutylicum* +*ter*_FLAG + *bgaR_tcdR* + ButOH. Error bars represent standard deviation. *n* = 3.

C. acetobutylicum Δ*pyrE* strain (without *ter*)- this strain shall be referred to as the *C. acetobutylicum bgaR_tcdR+bcd/etfAB*_ctrl strain. The purpose of this strain was to check that the effects observed in fermentations using the *C. acetobutylicum* + *ter*_FLAG + *bgaR_tcdR* + ButOH strain were partially owed to the NADH conserving properties of TER.

As before strains were either grown in the presence of 1 mM lactose inducer or absence (figure 45). The *C. acetobutylicum bgaR_tcdR + bcd/etfAB*_ctrl control strain produced two distinct phenotypes for the induced and un-induced state (figure 46). The un-induced strain produced an enhanced solvent phenotype, similar to the *C. acetobutylicum* + *ter*_FLAG strain showing a rapid re-assimilation of butyrate but no re-assimilation of acetate and enhanced solvent production with similar productivity dynamics and final yields. This suggests that *C. acetobutylicum* + *ter*_FLAG + *bgaR_tcdR* + ButOH strain is displaying an additive benefit from both the additional co-factor availability and the enhanced, expression of the C4 pathway enzymes with *adhE2*. Inducing the pathway resulted in an acid crash phenotype, where acids did not re-assimilate and the strain failed to switch to solventogenesis, as were some induced *C. acetobutylicum* + *ter*_FLAG *bgaR_tcdR* + ButOH fermentations.

The second control strain was tested in order to investigate the effects of the expression of the non-native clostridial *tcdR* gene - a strain created by Dr. Hengzheng Wang was used which expressed the lactose inducible *tcdR* under the control of the previously described system, inserted at the *pyrE* locus. This strain did not possess a P_{tcdB} promotor - therefore the effects of expressing the *tcdR* gene under the control of BgaR alone could be observed, while also mitigating against the effects of the expression of downstream chromosomal genes at the *pyrE* locus this strain will be referred to as *C. acetobutylicum bgaR_tcdR*_ctrl control. This strain was streaked onto CGM™ as a negative control before use, in order to ensure absence of contaminating plasmids.

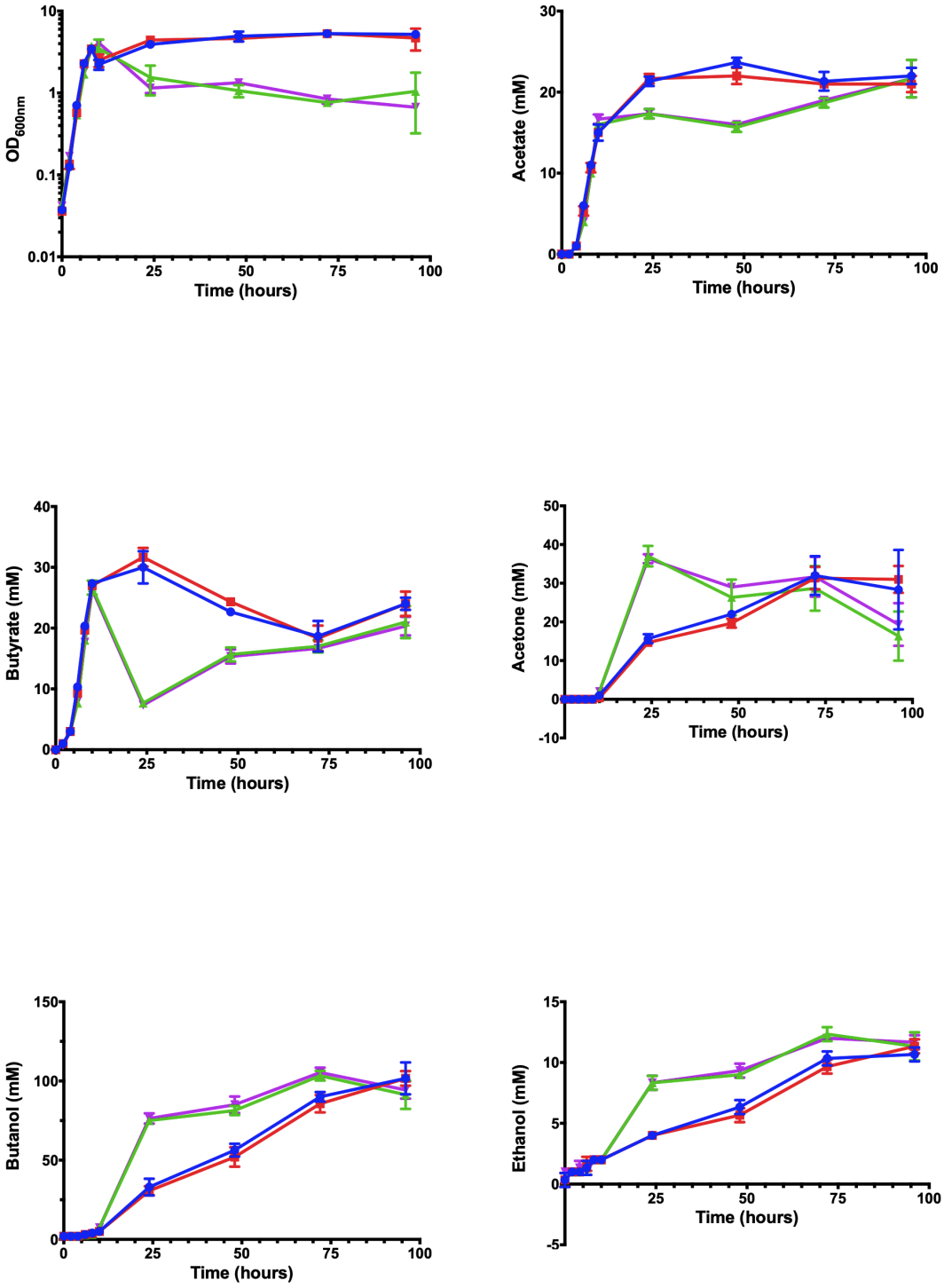


Figure 45: Fermentations to control for inducer effect on the Ter control ●, Un-induced *C. acetobutylicum* ATCC 824; ■ 1 mM lactose *C. acetobutylicum* ATCC 824; ▲, Un-induced *C. acetobutylicum* + *ter*_FLAG; ▼, 1 mM lactose *C. acetobutylicum* + *ter*_FLAG. Error bars represent standard deviation. *n* = 3.

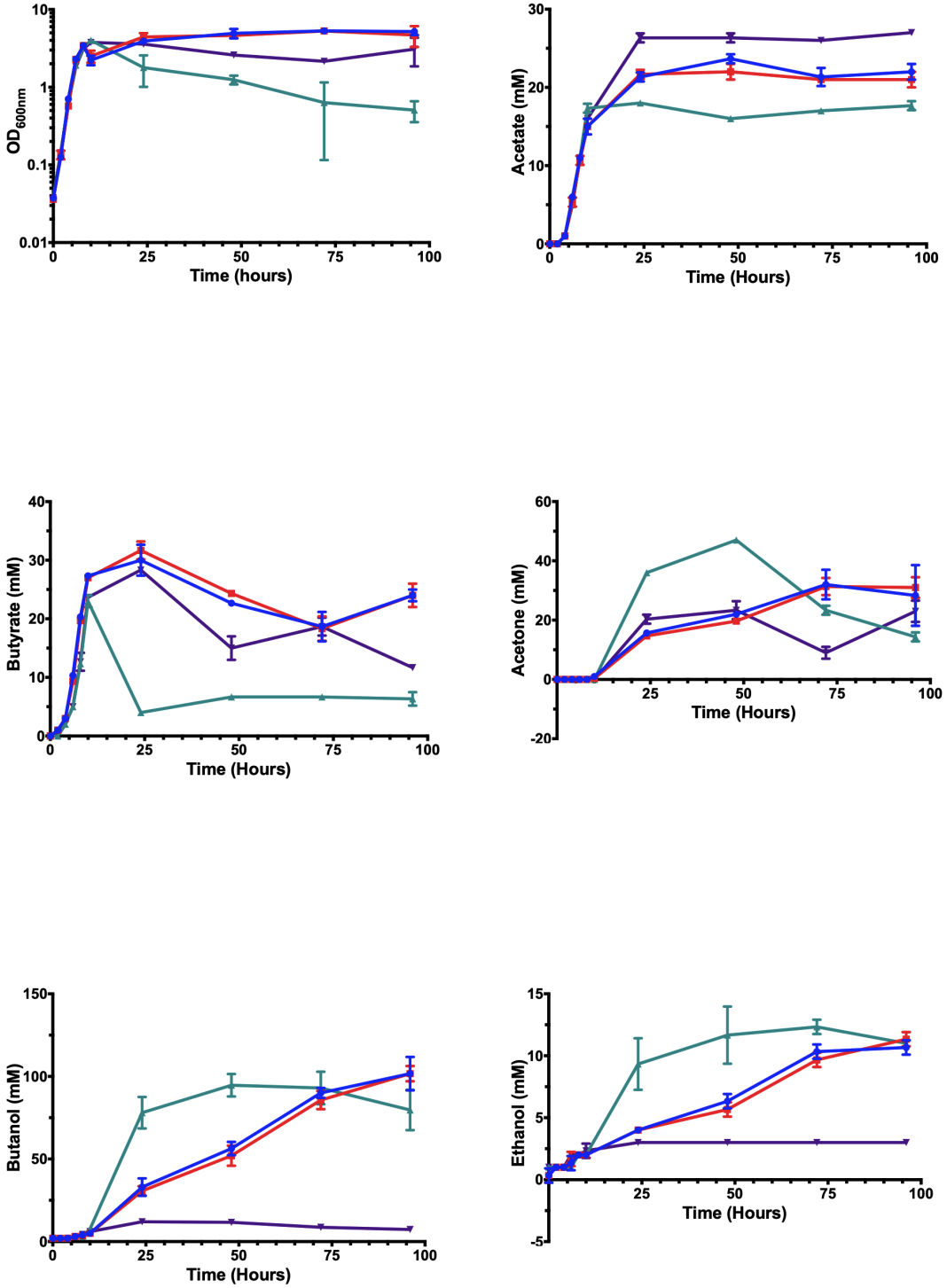


Figure 46: Physiology of the *C. acetobutylicum bgaR_tcdR + bcd/etfAB*_ctrl strain vs wild type control in fermentation. Lactose at 0 mM (un-induced) and 1 mM (induced) were compared: ●, Un-induced *C. acetobutylicum* ATCC 824; ■, Induced *C. acetobutylicum* ATCC 824; ▲, Un-induced *C. acetobutylicum bgaR_tcdR+bcd/etfAB*_ctrl; ▼, Induced *C. acetobutylicum bgaR_tcdR+bcd/etfAB*_ctrl. Error bars represent standard deviation. *n* = 3.

The *C. acetobutylicum bgaR_tcdR_ctrl* strain was tested in an un-induced and induced fermentation, using 0 and 1 mM lactose, as before. Expression of *tcdR* alone in both the induced and un-induced fermentations produced similar phenotypes, however, the induced *C. acetobutylicum bgaR_tcdR*_ctrl strain produced approximately 50% more ethanol. In both fermentations an enhanced solvent production profile may be seen, as well as a more rapid rate of re-assimilation of acids (see figure 47), overall, the performance was comparable to the *C. acetobutylicum* P_{thlA}+*ter*_FLAG strain.

This result would suggest that the TcdR σ^{70} factor taken from *C. difficile* has activity impacting the transcription of one or more genes in *C. acetobutylicum,* affecting the rate of progression of the metabolism though acidogenesis and into solventogenesis. Given that this σ^{70} is thought to be highly selective in its binding to P_{tcdB}/P_{tcdA} in *C. difficile* it is interesting that this result should occur.

A protein BLASTX search was conducted for the *tcdR* nucleotide sequence and a number of solventogenic species of *Clostridium* produced results with a sequence identity ranging from 30-37 %, including *Clostridium pasteurianum, Clostridium beijerinckii,* and *Clostridium saccharoperbutylacetonicum* (see table 1), however, this did not include *C. acetobutylicum.* Searching for the P_{tcdB} promotor region produced a number of hits when selecting *C. acetobutylicum* ATCC 824 with BLASTN (optimised for somewhat similar sequences) a number of sequence similarities are found, however, this tool ought to be deemed as too imprecise as the 17 bp gap region between the -35 and -10 binding sites will spuriously contribute to the generation of sequence matches. In order to identify important matches a bespoke script was written using BioPython.

A summary of the butanol production dynamics generated by the strains used in this study were compiled into a single graph (figure 48) to compare the performance of *C. acetobutylicum bgaR_tcdR + bcd/etfAB*_ctrl and *C. acetobutylicum bgaR_tcdR*_ctrl to previous strains used given that they both overproduced solvents.

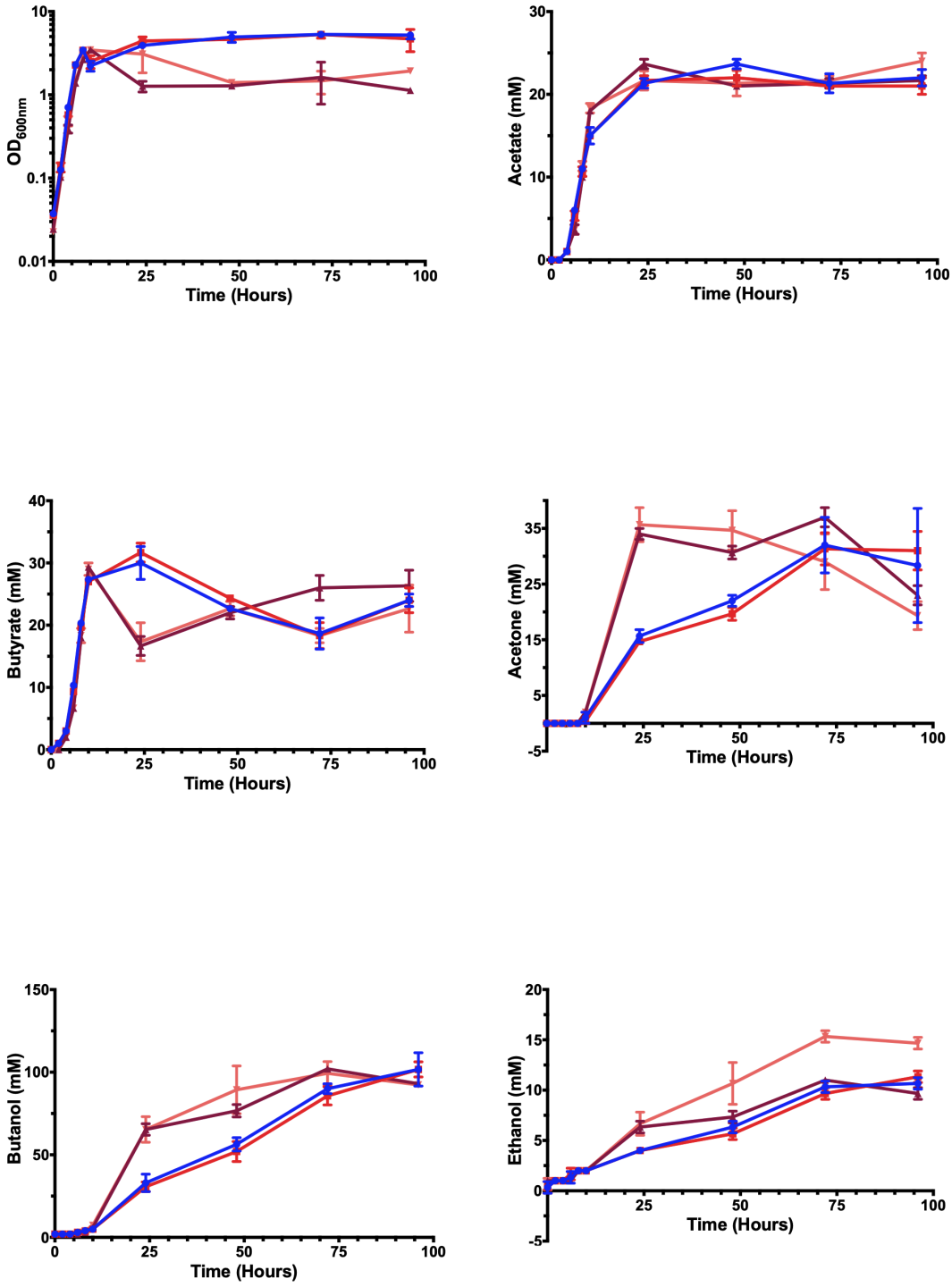


Figure 47: Physiology of the *C. acetobutylicum bgaR_tcdR*_ctrl strain for the effect of lactose, during flask fermentation. 0 mM (un-induced) and 1 mM (induced) were compared: ●, Un-induced *C. acetobutylicum* ATCC 824; ▪, Induced *C. acetobutylicum* ATCC 824; ▲, Un-induced *C. acetobutylicum bgaR_tcdR*_ctrl; ▼, Induced *C. acetobutylicum bgaR_tcdR*_ctrl. Error bars represent standard deviation. *n* = 3.

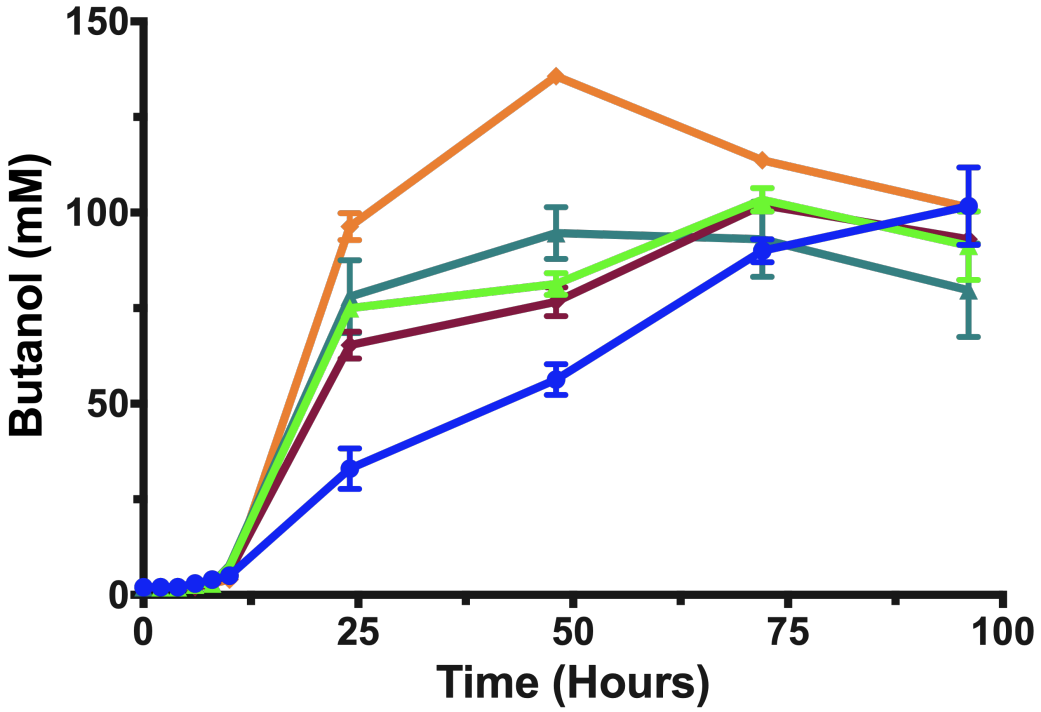


Figure 48: A summary comparison of butanol profiles produced by the un-induced strains used in this chapter: ●, *C. acetobutylicum* ATCC 824; ▲, *C. acetobutylicum* + *ter*_FLAG: ◆, *C. acetobutylicum* +*ter*_FLAG + *bgaR_tcdR* + ButOH; ▲, *C. acetobutylicum bgaR_tcdR* + *bcd/etfAB*_ctrl; ▼, *C. acetobutylicum bgaR_tcdR*_ctrl. Error bars represent standard deviation. *n* = 3.

Description	Identity (%)
hypothetical protein [Clostridium sulfidigenes]	48.276
RNA polymerase subunit sigma-24 [Clostridium beijerinckii]	37.58
sigma-70 family RNA polymerase sigma factor [Clostridium saccharoperbutylacetonicum]	37.195
sigma-70 family RNA polymerase sigma factor [Clostridium pasteurianum]	32.704

Table 1: Selected results from a BLASTX *tcdR* of *Clostridium* genus taxid: 1485 with the exclusion of *Clostridium difficile* [taxid:1496].

5.2.4 Bioinformatics analysis to investigate the potential activity of the TcdR sigma factor in the *C. acetobutylicum* genome

Given the change in phenotype observed from the expression of *C. acetobutylicum bgaR_tcdR*_ctrl strain, in the apparent absence of appropriate published web tools, an attempt to identify possible promotor binding sites was made using a Python algorithm utilising the Biopython package of libraries and tools (Cock *et al.,* 2009). The algorithm contained variables which allowed the -35 and -10 regions of the promotor to be defined, as well as a gap region between the two binding sites. It also allows for ambiguity to be defined in the -35 and -10 in the case of non-conserved bases. A full copy of the search algorithm is provided in the materials and methods section 2.4.2. A FASTA file directory is then referenced to search for all of the coding sequences that match the specifications defined in the aforementioned variables. The FASTA files referenced for *C. acetobutylicum* ATCC 824 were assigned with Genbank accession number NC_003030.1 for the chromosome, and NC_001988.2 for the pSOL megaplasmid.

TcdR has been assigned to a new group of sigma 70s: group V, however, they share significant similarity to group IV ECFs (Martin-Verstraete *et al.,* 2016). Dupuy *et al.,* (2006) characterised this new, additional class of sigma factors by conducting substitution experiments and found that BotR, TetR, TcdR and UviA were all capable of substituting one another *in vitro*. Although, in cells, BotR and TetR could substitute one another and UviA and TcdR could also substitute one another but neither BotR nor TetR could substitute for either UviA or TcdR - this discrepancy is possibly due to the fact that sigma factors compete for RNAP and without sufficiently high expression, sigma factors will not be able to titrate out the competitor - likely, in most cases to be the group I RpoD sigma 70.

Dupuy *et al.,* (2006) showed that the specificity of the TcdR sigma factors were dependant on the differences between the -10 regions. It might well be the case that

this specificity is lost when sigma factor concentrations become higher - which is particularly worth considering when designing artificial systems - and a weak binding affinity might be able to manifest gene expression where it otherwise would not in a native system.

Further to the four aforementioned members of the TcdR family of sigma factors, Dupuy *et al.,* (2006) identified a fifth sigma factor bearing significant sequence identity to the other TcdR family sigmas, which was designated "σ^{Y}" from *C. acetobutylicum*, this sigma factor could not be substituted for any of the other so called group V σ^{70} factors *in vitro,* and so was not considered a part of the TcdR family.

In a 2016 review, Martin-Verstraete *et al.,* (2016) compiled the 10 known promotors regulated by the group V σ^{70} TcdR family sourced from 8 species of *Clostridium*. The -35 sequence was conserved across the genus, however, the -10 sequence showed a high degree of variability, 4 of the 6 bases were not conserved across the 8 species, and among the -10 TcdR binding sites for P_{tcdA} and P_{tcdB} of *C. difficile,* 2 of the 6 bases were non-conserved (NTCNTT) (table 19), despite both binding the same TcdR sigma factor.

The *in vitro* interchangeability of TcdR with BotR, TetR, and UviA demonstrated by Dupuy *et al.,* (2006) suggests using the ambiguous -10 promotor sequence (NTNNTN) in the algorithm sequence parameters is acceptable for identifying matching viable TcdR binding sites. It is very likely, given the observations made by Dupuy *et al.,* (2006), that TcdR will bind at the promotor sites identified using these sequence parameters so long as the *tcdR* expression levels are sufficiently high, any downstream CDS with an upstream RBS and without interruption by a terminator is likely to be expressed.

Having established an acceptable range of parameters, two sets of variables were subsequently investigated. The initial sequence parameters were derived from the -35 and -10 conserved sequence between P_{tcdA} and P_{tcdB} from *C. difficile*,

Promotor	Sequence	Species	Loci	Gap (bp)	Distance between end of -10 and start of ATG
ntnh-bont/A	ATTTTAGGTTTACAAAAAATAGTGTGGCTATGTTATATATAAATG	*C. botulinum*	Chromosome	17	108 (96 ref)
ha34	ATTTTAGGTTTACAAAAAATAACTTGATTATGTTATATGTTATAT	*C. botulinum*	Chromosome	17	35 (34 ref)
tetX	ATTTTCAGTTTACAAAAAATAACCTGATTATGTTATATGTAATTG	*C. tetani*	Plasmid (pE88)	17	132 (124 ref)
tcdA	AAGATATGTTTACAAATTACTATCAGACAATCTCCTTATCTAATA	*C. difficile*	Chromosome	17	242 (234 ref)
tcdB	GAACAAAGTTTACATATTTATTTCAGACAACGTCTTTATTCAATC	*C. difficile*	Chromosome	17	172 (164 ref)
P1*bcn*	AAATTTAGTTTACAAAATTGAAGTCAAATTACTTTTTATATTATG	*C. perfringens*	Plasmid (pLLY_N11_1)	17	96 (90 ref)
P3*bcn*	TTTTTTAGGTTTACATTTTTAAAACTAAACTCTTTTTATTTATTA	No blast results	No blast results	16	(103 ref)
P*tpeL*	AATTTTTATTTACAAAATTTAAAATTAGAA-CTCTTTATATAATG	*C. perfringens*	Plasmid (pJIR4150)	17	148 (115 ref)
P*tcsL*	AAAGTATGTTTACATATTCATATAAAAATAATTCTTACTTTATGT	*C. sordellii*	Plasmid (pCS1)	17	186 (125 ref)
P*tcsH*	TTTTTTAGTTTACATTTTGCTTAAAATTTATTTTTTAATATAGGG	No blast results	No blast results	17	(176 ref)
tcdA/B conserved	NNNNNNNNTTTACANNNNNNNNNNNNNNNNNNTNCNTTNNNNNNNN	— —			
All conserved	NNNNNNNNTTTACANNNNNNNNNNNNNNNNNNTNNNTNNNNNNNNN				

Table 19: Promotors under regulatory control by TcdR family factors. Green denotes -35 binding region, teal denotes the -10. Conserved bases are identified in the bottom two rows. The -35 region is conserved across all species known to regulate under a TcdR family system. The -10 region is highly variable with 2 out of 6 bases conserved across the genus *Clostridium* and 4 out of 6 conserved between P_{tcdA} and P_{tcdB} of *C. difficile*. Conserved sequences were used to determine TcdR binding sites in *C. acetobutylicum*. Table adapted from Martin-Verstraete *et al* (2016), downstream start codons referenced from the NCBI genome database and compared to measurements recorded by Martin-Verstraete *et al* (2016) indicated in brackets.

allowing for ambiguity of non-conserved bases with a gap region of 17 bp. The genome of interest, *C. acetobutylicum* ATCC 824 was compared to a positive control *C. difficile* 630*,* the origin of TcdR and the two corresponding toxin regulation promotors, and a negative control, *Campylobacter jejuni* ATCC 700819, which was chosen as this organism might be considered sufficiently evolutionarily distinct, belonging to the phylum Proteobacteria. Furthermore, *C. jejuni* has a GC content comparable to the *Clostridium* strains at 30.5%, and therefore this sequence should generate similar match rates resulting from randomness.

These parameters returned a limited number of results among the *Clostridium* species, perhaps expectedly, owing to the notably specific nature of TcdR. Of the 61 matches returned for *C. acetobutylicum*, only two occurred intergenically, and only one of which was positioned so that downstream expression of a CDS was likely to occur. This gene was *dnaX,* and was found 234 bp downstream of the -10 binding site. The gene *dnaX* codes for is DNA polymerase III subunits τ and γ; this is a single strand DNA polymerase which is ATP dependent and requires the δ subunit for ATP activity (Walker *et al.,* 2000). There is no obvious reason why the expression of this gene would contribute to the observed phenotype in *C. acetobutylicum* *bgaR_tcdR*_ctrl fermentations.

The positive control in *C. difficile* returned 96 total matches, 9 of which were located in intergenic regions, and as is consistent with the literature, only two of these binding sites contained CDS in the correct orientation downstream from the binding site, these were the *C. difficile* toxin genes, *tcdA* and *tcdB* concurring with the established theory that TcdR specifically controls toxin gene expression in *C. difficile* (Mani & Dupuy, 2001; Dupuy & Sonenshein, 1998). It is interesting that although the TcdR conserved promotor sequences were found in 96 positions, only 2 of these were functional, resulting in the specific control of the toxin genes, and none of the apparently randomly occurring sequences resulted in unintended gene expression. Given that these promotors operate at a relatively high transcription rate it is likely

that any unintended gene expression was strongly selected against, driving out any randomly occurring sequences that would negatively impact fitness. Obtaining these results also provided assurance that the search algorithm was working as intended since the results were in agreement with the literature.

The negative control, *C. jejuni* produced 34 matches for the P_{tcdA}/P_{tcdB} conserved sequence, with one sequence found intergenically, which was not in the correct orientation for gene expression. This is a much lower number of matches comparing to the two *Clostridium* species, however, considering the fact that the size of the genome is also much smaller, it is to be expected that a lower number of matches would be found. With this in mind, a comparison of matches Mb^{-1} was made which produces a more comparable figure within the range of the two *Clostridium* species (table 20). It should also be noted that after visual inspection of the annotated GenBank file using the SnapGene GUI, the chromosome for *C. jejuni* appears to be much more condensed, containing far less and much shorter intergenic sequence regions than both *Clostridium* species.

A second set of searches were performed on the three sets of genomic data in order to investigate the effects of increased ambiguity in the -10 sequence to reflect the non-conservation among promotors recognised by all four members of the σ^{70} TcdR family (NTNNTN), given that they have been shown to work interchangeably when investigated *in vivo* (Dupuy *et al.,* 2006). A second investigation was made using the less conserved -10 sequence and although this risks generating more noise in the data, it is important to search for consensus sequences across the family, and not just restrict to the consensus between TcdR promotors, as TcdR has shown to be interchangeable with other members of this family of transcription factors. This produced a far greater number of matching sequences - returning 728 matches from the *C. acetobutylicum* ATCC 824 chromosome and plasmid reference sequences. Some interesting patterns were noticed within this data, with 85 sequence matches occurring intergenically in the correct orientation for expression of

		RefSeq Reference Files	Total Number of Matches	Exogenic Loci	Intergenic Loci	Percent Intergenic (%)	Percent Genome Non-Coding (%)	Correct orientation (CDS Downstream)	Percent Correct orientation (%)	Genome Size (Mb)	GC Content	Matches Mb^{-1}	Total Genome CDSs
-10 Conserved Across the TcdR Family (NTNNTN)	***C. acetobutylicum* ATCC 824**	NC_003030.1, NC_001988.2	728	580	148	20.33	13.04	86	58	4.13	30.9	176.3	3871
	***C. difficile* 630**	NC_009089.1, NC_009089.1	1047	835	212	20.25	-	75	35	4.29	29.1	244.1	3766
	***C. Jejuni* ATCC 700819**	NC_002163.1	344	330	14	4.07	-	8	57	1.64	30.5	209.8	4725
-10 Conserved Between P_{tcdA} and P_{tcdB} (NTCNTT)	***C. acetobutylicum* ATCC 824**	NC_003030.1, NC_001988.2	61	59	2	3.28	13.04	1	50	4.13	30.9	14.8	3871
	***C. difficile* 630**	NC_009089.1, NC_009089.1	96	87	9	9.38	-	2	22	4.29	29.1	22.4	3766
	***C. Jejuni* ATCC 700819**	NC_002163.1	34	33	1	2.94	-	0	-	1.64	30.5	20.7	4725

Table 20: Results summary compiling and comparing sequence matches across test sequence data and the controls. Percent non-coding data obtained from Poehlein *et al.,* (2017).

a downstream CDS. Since the data generated by the algorithm was analysed manually, terminator detection could not be feasibly performed, so therefore all genes downstream of a binding site were initially counted, it is understood that some genes may not actually be expressed due to this omission. Nevertheless, the issue was borne in mind when reviewing the impact of expression of qualifying genes.

When reviewing the importance of the location of the putative promotor sequences, promotors found inside CDS annotations were dismissed for simplicity. The question of overlapping genes and their contribution to any putative regulon would be potentially important, however, it is outside the scope of this initial study.

The positive control, *C. difficile* 630, produced the largest number of matches at 1047 across the genome including 75 sequences which had a downstream CDS in the correct orientation for transcription and translation. This indicates that off target binding is concentration dependant, given that TcdR binding is indiscriminate *in vitro* but not *in vivo*, it may be inferred that expression of TcdR must be titration dependent in order to ensure its precision.

The negative control was used to ideally provide an indication of the noise level in the data, where all matches may be presumed to have occurred as a matter of random chance providing an indication of the noise in the data produced by the variable test genome (*C. acetobutylicum* ATCC 824). The sample produced a comparable number of matches indicating a high level of noise. Nevertheless, TcdR ought to bind to the sequence defined in the parameters, but the binding would occur as a matter of random chance and not as an evolved function. A comparison between the number of matches per megabase was made between the negative control and the test strain. For this to be reliable the GC content of the test genome ought to be comparable to the negative control, indeed, this was the basis for the selection of the *C. jejuni* genome data, having a GC content of 30.5% compared to a GC content of 30.9% in *C. acetobutylicum*. When the algorithm was applied to the negative control using the TcdR family non-conserved -10 sequence the total number

of matches produced was 344, much lower than the matches produced for *C. acetobutylicum* and *C. difficile,* which produced 728 and 1047 matches, respectively. However, comparing matches per megabase produced much more comparable figures, where *C. jejuni* produced 209.8 matches Mb^{-1}, *C. acetobutylicum* produced a lower number of matches Mb^{-1}, at 176.3, and *C. difficile* produced the greatest at 244.1 matches Mb^{-1}. The negative control demonstrated that there were significantly high levels of random binding occurring, providing evidence to suggest that a large portion of the sequence matches in the *C. acetobutylicum* and *C. difficile* genomes can be considered random background noise and not necessarily the result of related function. When the more specific -10 sequence with bases conserved from P_{tcdA} P_{tcdB} only were compared on a matches Mb^{-1} basis, a similar ranking in binding frequency was observed, however, at a rate that was around 1 order of magnitude lower than the search using the more ambiguous -10 sequence.

Binding sites may well occur randomly without having any selective forces during the course of evolution, the shorter and more ambiguous the sequence the more likely they are to occur due to random chance. Furthermore, they may also exist as a vestigial relic from a common ancestor which may still have activity in the event of the presence of a heterologous, highly expressed, and compatible σ^{70}. Although, such relics should be expected to deteriorate rapidly without having any specific function due to the absence of selective forces preserving them. More importantly, binding sites might occur due to the presence of a related sigma factor which is still functional.

It is worth noting that previous nucleotide BLAST and BLASTX searches of various stringencies did not identify a *C. acetobutylicum* homologue in this study. A Clustal W analysis, however, showed significant similarity between TcdR and σ^{Y} which was initially conducted by Dupuy *et al.,* (2006) and repeated in this study (not shown). Domain 4.1 and 4.2 bearing particular similarity, which is known to confer -35 binding (Feklîstov *et al.,* 2014).

5.2.4 Promotor regions flanking downstream genes identified as sigma factors derived from the 728 sequence hits produced by the -10 conserved sequences from the TcdR family

The most interesting genes identified which were qualified as likely to be expressed, coded for three other putative sigma factors *cac0167*, *cac1766*, and *cap0167* (table 21). Clearly, given the regulatory nature of sigma factors, there is considerable potential for an adjusted phenotype to result from up-regulation of any of these genes. A fourth sigma factor *cac1770* - annotated on NCBI as *sigL* was also identified, however, this was largely discounted from the analysis due to the fact that the promotor was 1148 bp upstream of the gene in a 1231 bp intergenic region, the assumption being that it is possible that a terminator may interrupt this promotor.

The gene *cac0167* with a 48 bp upstream promotor site was identified as a sigF/sigE/sigG family sigma factor according to annotation on the Genbank file, annotations were performed according to the comparative genome analysis of *C. acetobutylicum* conducted by Nölling *et al.,* (2001). To date, this putative sigma factor remains uncharacterised. The sigF/sigE/sigG family sigma factors in *Clostridium* and *Bacillus* are generally associated with sporulation, and although this family of sigma factors are largely conserved across the Firmicutes, the order of activation and associated regulatory mechanisms prove to be diverse (Fimlaid and Shen, 2015). In *C. acetobutylicum*, knock out mutations, Tracy *et al*., (2011) demonstrated that *sigE* deletions had the potential to affect solvent production, although this was noted to be dependant on the physiological state of the starter culture used to inoculate the fermentations. Deletion of *sigG* resulted stalled sporulation during endospore formation - however, normal solvent production did occur. These results suggest that over-expression of a sigF/sigE/sigG family sigma factor has the potential to impact solventogenesis, and therefore up-regulation of this gene by TcdR might explain the phenotype observed in the TcdR -ve control fermentation.

Locus/ Gene Identity	Genbank Annotation	Comparative ID by PBLAST	Number of bases upstream of start codon	Promotor Sequence
CA_C0167	sigF/sigE/ sigG family sigma factor	Histidine Phosphatase	58	TTTACATTAAAATTGACTGGGTTTTAATA
CA_C1766	RNA polymerase factor sigma-70	RNA polymerase factor sigma-70	48	TTTACAAAGAATTAAGATGATGCATTATG
CA_C1770	RNA polymerase sigma factor SigI	Spore sigI superfamily	1148	TTTACATAATTATAACATATTTGCTACTA
CA_P0167	Specialized sigma factor (SigF/SigE family)	sigma70-ECF superfamily	289	TTTACATTATAACACATATTGTTTTTTTA
CA_P0167	Specialized sigma factor (SigF/SigE family)	sigma70-ECF superfamily	252	TTTACACCATATATTAATTTTTTCTACTA

Table 21: Summary of putative sigma factor CDS sites potentially transcribed by TcdR. Promotors identified according to -35 (green) and -10 consensus sequence (teal with conserved "t" bases underlined).

The gene *cac1766* has been annotated in the NCBI genbank sequence with accession number NC_003030 as a "RNA polymerase factor sigma-70" and the binding site occurred 48 bp upstream of the start codon, and so is well positioned for a promotor. A BLASTP query confirmed this annotation with the first 100 results ranging from 75 - 37 % sequence identity with other sigma 70s, all within the Firmicutes phylum. This produced 8 results from *C. acetobutylicum*, 2 of which were labelled as "ECF RNA polymerase sigma factor SigL." Furthermore, a sigma factor sharing 75% identity from *Clostridium roseum* was annotated "ECF RNA polymerase sigma factor SigL" (GenBank: OOL91654.1).

Nie *et al.,* (2016), characterised a gene *sigL* which was responsible for regulation of cellobiose metabolism, however, this was identified as a σ^{54} by comparative genome analysis so comparisons a group IV σ^{70} cannot be used. One interesting reference in the literature "*sigL*" was for a σ^{54} type sigma factors which are generally involved in regulation of carbon metabolism via the carbon catabolite repressor (CcpA) signalling during vegetative growth of *B. subtilis* (Choi *et al.,* 2005). This is interesting as CcpA has been shown to negatively modulate TcdR activity in *C. difficile.* In *C. acetobutylicum* CcpA is known to positively regulate the *sol* operon and negatively regulate alternative carbon catabolism genes in the presence of glucose (Ren *et al.,* 2012). Hahn, *et al.,* (2005) did discover an ECF type σ^{70} in *Mycobacterium tuberculosis* that they named SigL, which was shown to be involved in polyketide synthesis and pathogenesis after marked attenuation was demonstrated in Δ*sigL* mutants.

The third sigma factor was described as a "hypothetical protein" in the Genbank feature notes for accession number AE001438 and is located on the pSOL megaplasmid at *cap0167*. A BLASTP query placed the CDS translation into the sigma70-ECF superfamily - a sigma factor of extracytoplasmic function (ECF), making it a member of σ^{70} group IV.

cap0167 appears to be expressed polycistronically with *cap0166* which lies immediately upstream and a BLASTP query places this hypothetical protein in the YvrJ superfamily, which UNIPROT (accession: Q97TE4) identifies the CDS as coding for a membrane protein with a single transmembrane helix domain and a coiled coil domain.

Yoo *et al.,* (2015) produced a comprehensive whole genome microarray transcriptomics reference for *C. acetobutylicum* in three metabolic steady states, acidogenesis, solventogenesis and alcohologenesis (GEO accession number: GSE69973). All cultures were continuous phosphate limited cultures with maintained with a constant total carbon amount of 995 mM. Acidogenic cultures were maintained

at pH 6.3 and fed on a 995 mM glucose feed. Solventogenic cultures were maintained at pH 4.4 and also fed with a 995 mM glucose feed. Alcohologenic cultures were maintained at pH 6.3 and were fed with 498 mM glucose and 498 mM glycerol, glycerol being a more reduced substrate - its catabolism results in a higher $NADH/NAD^+$ ratio forcing acetone out from the product profile, the supplementary transcriptomics data provided with the paper allowed for comparison of transcription rates of the three putative sigma factors identified by the promotor search algorithm.

Comparing the relative transcription rates of the three sigma factors identified by the search algorithm (figure 49) provides some insight into their potential regulatory roles. Firstly *cac0167* shows similar transcription patterns to an RpoD group I σ^{70}. Transcript concentration is consistent under all three fermentation conditions,

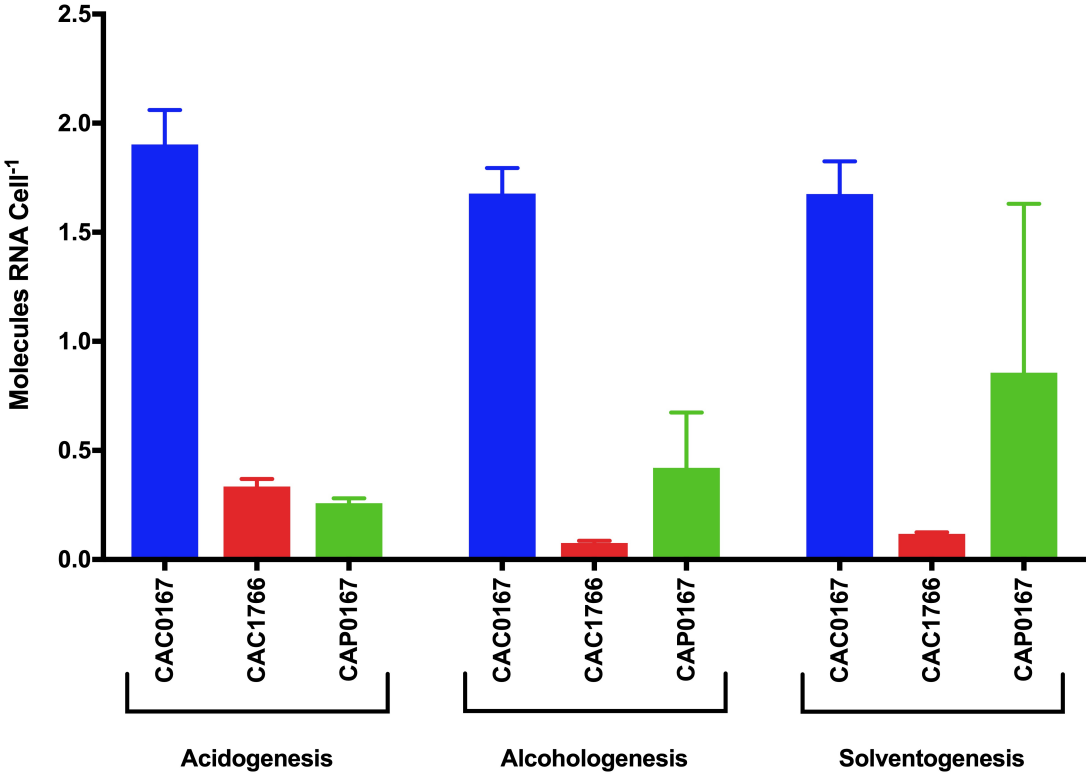


Figure 49: Selected data from the whole genome micro-array provided by Yoo *et al.,* (2015) allowing for comparison of transcription rates of putative sigma factors during three different metabolic states in a steady state constant phosphate limited fermentation system. Error bars represent standard deviation where *n* = 3.

although is marginally higher in acidogenesis. *cac0167* is more highly transcribed than *cac1766* and *cap0167* in all metabolic states. *cac1766* has a relatively low rate of transcription under all three conditions, although measurements are the highest during acid fermentations. *cap0167* shows markedly increased transcription rates under conditions required for solventogenesis, however, standard deviation is high. *cap0166*, which is co-expressed polycistronically with *cap0167* displayed a similar transcription pattern (not shown).

The function of the putative sigma factor encoded at the *cap0167* CDS has been investigated previously and results have been reported by several groups. In 2000, Behrens *et al.,* phylogenetically classified this as an ECF σ^{70} and designated it as *sigY*, along with another previously uncharacterised but related sigma factor coding sequence - *sigX*, situated on the chromosome at *cac1505*. Amino acid sequence alignment demonstrated significant similarity of σ^{Y} to UviA of *C. perfringens* and the early sporulation regulator σ^{H} belonging to *B. subtilis*. The strongest resemblance was found to be in domain 4 which is responsible for -35 binding. This corroborates with observations of a highly conserved -35 sequence summarised in the 2016 review by Martin-Verstraete *et al.* It also suggests that it's expectable to find a promotor for σ^{Y} bearing a similar -35 sequence to the TcdR family of sigma factors. The study also cited unpublished work by Dürre and Schaffer mentioning that σ^{Y} was responsible for expression of *amyA*, at the adjacent downstream locus, coding for alpha amylase.

In 2002 Dürre *et al.,* discussed *sigY* regulation and noted that the target operons of σ^{Y} were yet unknown. The authors made reference to the work of Nakotte *et al.,* (1998) who had observed the binding of a protein to the promotor region of *sigY* during late solventogenesis. This information was then used to speculate that negative supercoiling in the regions 250 bp upstream of genes being transcribed in the reverse configuration (i.e. *sigY*), as is consistent with the "twin transcriptional loop" model (Liu and Wang, 1987), could be responsible for down-regulation of *adc.*

By 2006 Dupuy *et al.,* had began work on classification of the TcdR family and had discovered that σ^Y had significant protein sequence similarity to TcdR and had designated TcdR into a new group of sigma 70 factors: group V, along with TetR from *C. tetani*, BotR from *C. botulinum*, and UviA from *C. perfringens*. However, due to the apparent lack of *in vivo* interchangeability of σ^Y with other members of the group, this sigma factor was not included in the proposed group V TcdR family by the authors. Presently, both TcdR and σ^Y are classified as ECF type group IV sigma 70 factors according to BLASTP results. The study did confirm earlier suggestions that σ^Y was responsible for regulating the expression of *amyA*, *in vitro*; by adding a DNA fragment containing *amyA* and its putative promotor region to a reaction mix containing recombinant σ^Y and *E. coli* core RNAP the reaction produced a 150 bp transcript. This was not the case when TcdR or UviA were added to the mixture in place of σ^Y.

In 2008, a transcription analysis using RNA microarrays was reported by Jones *et al*. This study highlighted *cap0167*, noting it as a potentially important regulator of sporulation and/or solventogenesis. Furthermore, the study identified *cac1766* as a novel sigma factor, which was also identified in this study as having a putative TcdR-like promotor, as discussed previously. *cap0166* was also shown to be expressed as an operon with *cap0167* and their function, along with *cac1766* was investigated using asRNA knockdowns.

Although, Jones *et al.,* (2008) used a batch fermentation process, the microarrays showed similar results to those published by Yoo *et al.,* (2015) for *cap0166* and *cap0167* transcription rates peaking at the commencement of solventogenesis. Although, they differed in that *cac1766* transcripts reached the highest concentration towards mid stationary solventogenic phase and this possibly represents a role in sporulation, where Yoo *et al.,* (2015) reported higher transcription rates during the acidogenic continuous fermentation.

The asRNA knock down results reported by Jones *et al.,* (2008) for *cac1766*, *cap0166* and *cap0167* indicated a possible delayed switch from acidogenesis to solventogenesis at 96 hours. Although this effect was also seen with the pSOS95del plasmid control (asRNA delivery plasmid absent of an asRNA coding region), although to a less pronounced degree meaning any effect seen in the knock downs was only subtle. At 144 hours it appeared as though the asRNA *cap0167* strain was exhibiting a high degree of metabolic activity as the glucose concentration was markedly lower than both controls and all other test strains - while the butanol concentration was the highest of all the strains. It is not known whether this is a result of inhibition of sporulation and continued fermentation, or a result of the strain sporulating early and then germinating in the culture vessel (see figure 8 for substrate and product measurements reported in Jones *et al.,* 2008). It should be noted that these experiments were not performed in biological replicates, so the results presented are not at all statistically significant. Nevertheless, the study serves as the only attempt reported in the published literature to characterise these genes in *C. acetobutylicum*.

Since ECF type sigma factors are auto-regulating, in that they bind to and transcribe from their own promotors to rapidly up-regulate their own expression in a positive feedback loop, it is probably the case that the two TcdR-like binding promotors found upstream of *sigY* bind σ^Y itself in this auto-regulating fashion and it is known that the σ^Y -35 binding domain shares significant sequence homology with TcdR (Behrens *et al.,* 2000). Furthermore, we can check that the -35 binding region of TcdR and σ^Y is present upstream of the two amylase genes (CA_P0098 and CA_P0168), indeed this turns out to be the case - with both genes having a -35 TcdR/σ^Y binding region - (see table 22) and Dupuy *et al.,* (2006) suggesting that σ^Y regulates CA_P0168.

With consideration to the results of this study and the results of Jones *et al.,* (2008) it seems plausible to suggest solventogenesis might well be regulated by σ^Y,

and should this be the case a TcdR like σ^Y binding promotor region should also be found upstream of important solvent regulating genes, particularly the solventogenic (*sol*) operon, given the close proximity to σ^Y - as is often the case with ECF sigma factors they can be found nearby the genes they are known to regulate (Davis *et al.,* 2016; Kuehne *et al.,* 2010; Llamas *et al.,* 2009). Again, this turns out to be the case, in the intergenic region upstream of the *sol* operon there are two TcdR/putative σ^Y -35 binding sites. To provide additional assurance that these do in fact represent promotors it is interesting to compare putative promotor regions to existing annotations. Ralston and Papousakis, (2018) provide access to an RNAseq transcriptome assisted automated annotated genomic assembly for *C. acetobutylicum* (with access instructions described in their supplementary materials). Referencing this database shows that two of these TcdR/putative σ^Y promotor regions have a relationship with annotations on the database, one of which perfectly aligns with a putative σ^Y promotor, the second is overlapping (table 22). It would be interesting to repeat the experiments of Dupuy, *et al*., (2006) and take a DNA fragment containing the *sol* operon and the upstream intergenic region to serve as a transcription template in an *in vitro* transcription reaction using recombinant *E. coli* RNAP and σ^Y in order to confirm that these are indeed regulatory promotors and that σ^Y is a regulator of solventogenesis.

It ought to be considered, however, that the algorithm used to identify promotors by Ralston and Papoutsakis (2018) relied on the consensus sequence for *B. subtilis* SigA with distance from transcription start sites factored in, and that promotor annotations are estimations and not the result of direct experimental evidence. Nevertheless, upstream of the *sol* operon in the intergenic region, there are two binding sequences which match the -35 consensus sequence for TcdR. One of these perfectly aligns with the SigA type binding sequence identified by the Ralston and Papoutsakis (2018) annotation (see table 22).

Upstream CDS	-10 to start codon (bp)	Putative SigY Binding Promotor Sequence Location (bp)	Promotor Location Predicted By Ralston and Papoutsakis (2018) (bp)
amyA **(CA_P0168)**	93	183 362 .. 183 990	183 956 .. 183 990
	tttacattatatcaataaaatacaataaa attgtattttattgatataatgtaaa		
sol **operon (CA_P0162-CA_P0164)**	281	175 500 .. 175 528	175 500 .. 175 528
	tttacaatatctatctccaaatctgcttt tttacaatatctatctccaaatctgcttt		
sol **operon (CA_P0162-CA_P0164)**	80	175 701 .. 175 729	175 685 .. 175 717
	tttacaatatctatctccaaatctgcttt ttgttttgttttgcagtttacaatatctatctc		
adc **(CA_P0165)**	197	180 679 .. 180 708	180 678 .. 180 707
	tttacttaaaaaaacaatatgtgttataat ttacttaaaaaaacaatatgtgttataatg		
agrB **(CA_C0078)**	143	86 556 .. 86 584	86 556 .. 86 590
	tttacaaacaacaaatttccagagaaaaaa tttacaaacaacaaatttccagagaaaaatacagt		
sigY **locus (CA_P0166-CA_P0167)**	112	180 674 .. 180 702	180 636 .. 180 675
	tttacatttaaatacacagctgttatatt ttcacattataaatcgcctctagtcctatttatttatatt		
sigY **locus (CA_P0166-CA_P0167)**	175	180 711 .. 180 739	n/a
	tttacattataacacatattgttttttta No match		

Table 22: Putative promotors based on TcdR Family -35 consensus sequence and the common 17 bp gap. The final 6 bp were selected as the a tentative -10 sequence based on the review by Martin-Verstraete *et al.,* (2016). Database of annotations accessible through the genome browser provided by Ralston & Papoutsakis 2018. DNA sequences show in the top row: TcdR parameters including the -35 consensus (green) sequence followed by a 17 bp gap, the following 6 bp were then defined as the -10 sequence. The -10 sequence has poor consensus however "t's" in the 2nd and 5th position are conserved for TcdR and are underlined in putative σ^Y -10 sequences. In the bottom row: Promotor sequence identified by Ralston and Papoutsakis (2018).

5.2.5 Protein modelling of Cap0166

If indeed σ^Y is an ECF σ^{70} regulator of solventogenesis then this raises the question of the role of the upstream CDS *cap0166*. Generally speaking, genes co-expressed with ECF sigma factors are usually environment sensing regulatory systems such as two component systems or anti-sigma factors. Given that there is only one CDS upstream of σ^Y, this rules out the possibility of a two component regulatory system. As previously mentioned BLASTP and UNIPROT results showed the 53 residue protein to be a member of the currently uncharacterised YvrJ superfamily of membrane proteins, with an additional coiled coil domain. To confirm, TMHMM 2.0 (Krogh et al., 2001; Sonnhammer *et al.,* 1998) (see figure 50) and TMPred (K. Hofmann & W. Stoffel, 1993) (results not shown) topology modelling algorithms were used to check the predicted topology of *cap0166*. Both of these tools predicted a high probability single transmembrane region. Both tools predict

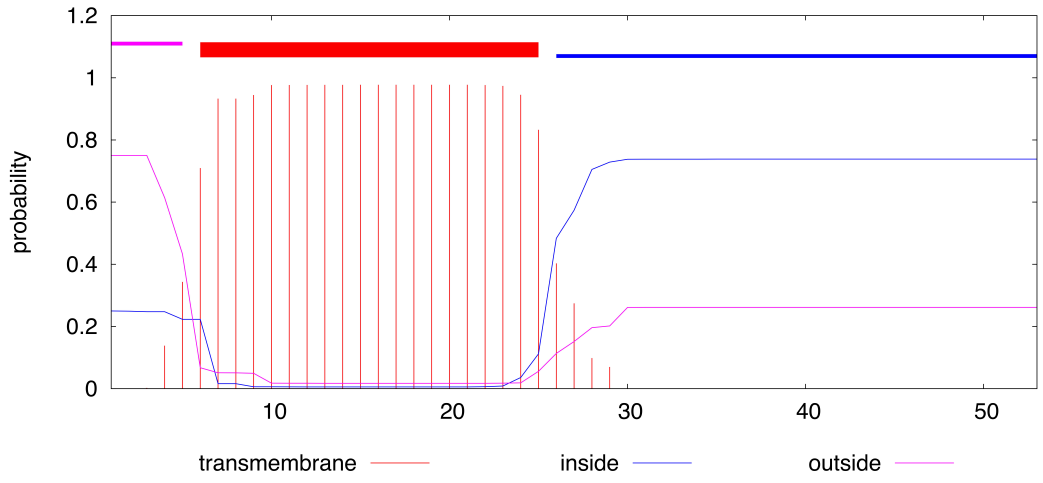


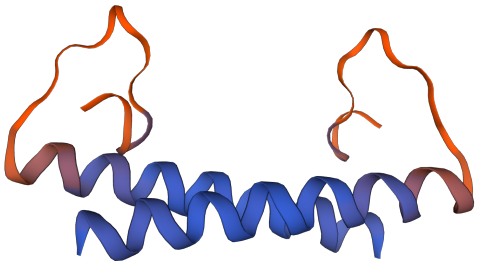
Figure 50: TMHMM algorithm which uses hidden markov models to make predictions of membrane protein helix topology. The output indicates probability of inside, outside or transmembrane residues on the y-axis, the x-axis indicates amino acid sequence. The results from the prediction tool indicate a high probability 20 residue transmembrane helix running from residue 6-25.

that the transmembrane region spans from residue 6 to residue 25 with the first 5 residues protruding into extracellular environment and the final n-terminal 27 residues extending into the cytoplasm - suggesting that the coiled coil domain listed on UNIPROT is intracellular.

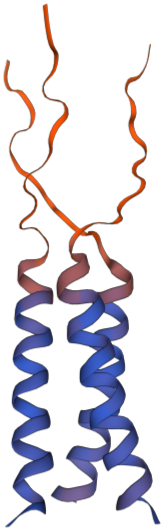
Further modelling of this putative membrane protein was carried out using ExPASY Swiss Model (Waterhouse *et al.,* 2018). This generated three putative structures: a monomer, a homodimer and, a homotrimer. The monomer produced a poor QMEAN (Benkert *et al.,* 2011) score, both globally and locally throughout the structure and did not corroborate the previous models in any way. The models depicting a homodimer and a homotrimer predicted coiled coil helical structures for residues residing in the portion identified as intracellular by the membrane protein prediction algorithms - residues 26-53 producing a high QMEAN locally rated score structures for residues within the coiled coil region. The homodimeric helices spanned from the predicted intracellular residues 24-49 and the homotrimeric helices spanned from residues 25-47 (figure 51).

The dimer has been arranged as an anti-parallel coiled coil and was based on a template for the *B. subtilis* transcriptional activator multi-drug activation, Mta. The structure was solved for the DNA binding apo-protein N terminal fragment - MtaN in 2001 by Godsey, *et al*. The Mta protein is part of the MerR family of winged helix-turn-helix (HTH) multi-drug efflux transporter transcriptional regulators. Novic *et al.,* (2003) notes that the SarA transcriptional regulator involved in the regulation of the *agr* quorum sensing system of *Staphylococcus aureus* is also a member of this family. It was interesting to discover this connection as *C. acetobutylicum* is known to have an *agr* quorum sensing system (Steiner *et al.,* 2011; Jabbari *et al.,* 2013). SarA is now known to be a late exponential/early stationary RNA binding accessory regulator (Morrison *et al.,* 2012) that is involved in acid stress response regulation (Weinrick *et al.,* 2004).

1



2



1

```
Model_01:A MNSGDIVSLIGNVGFPIAVSI--------YLLVRIEGKMEALTSSINNLTNTINNSNTKNK 53
Model_01:B MNSGDIVSLIGNVGFPIAVSI--------YLLVRIEGKMEALTSSINNLTNTINNSNTKNK 53
  1jbg.1.A ---------FKEIGFRLDEIKEMLDHPNFDRKAALQSQKEILMKKKQRMDEMIQTID---- 102
```

2

```
Model_02:A MNSGDIVSLIGNVGFPIAVSIYLLVRIEGKMEALTSSINNLTNTINNSNTKNK 53
Model_02:B MNSGDIVSLIGNVGFPIAVSIYLLVRIEGKMEALTSSINNLTNTINNSNTKNK 53
Model_02:C MNSGDIVSLIGNVGFPIAVSIYLLVRIEGKMEALTSSINNLTNTINNSNTKNK 53
  1slq.1.A ------------YQTPIDNSVTVRQDLERQLGELREEFNALSQEIAM------ 267
```

Figure 51: Protein models produced by ExPASy SWISS-MODEL 1a shows a coiled coil homodimer modelled from the intracellular N-terminal portion of the protein sequence. 2a shows a coiled coil homotrimer modelled from the intracellular N-terminal portion of the protein sequence. 1b and 2b show the model alignment of the amino-acid sequence with the templates. 1b shows the homodimer based on the template MtaN (*B. subtilis* Multidrug Transporter Activator, N-terminus) with PDB accession number: 1JBG. 2b shows the homotrimer based on the template VP4 (rhesus rotavirus membrane interaction domain) with PDB accession number: 1SLQ. Local QMEAN scores indicated by red to blue fade. Red indicating low probability and blue indicating high probability

Referring to back to the *C. acetobutylicum* (Genbank: NC_003030) annotated genome assembly, a TcdR/σ^Y like -35 consensus sequence was found 143 base pairs upstream of the *agrB* start codon, again this was cross-referenced against the Ralston & Papoutsakis (2018) annotated genome and indeed there was a promotor sharing the same -35 region, however, the -10 was recorded a further 6 bp upstream. AgrB is the transmembrane quorum sensing peptide receptor, it is known to regulate quorum sensing for sporulation regulation in *C. acetobutylicum* and it is co-transcribed as an operon with the quorum sensing signalling peptide AgrD (Steiner *et al.,* 2011).

The trimer was arranged as a parallel helix coiled coil based on a template for a section of the structure of the VP4 membrane interaction domain of the rhesus rotavirus (Dormitzer *et al.,* 2004). This is a hemagglutinin type membrane protein and as is characteristic of hemagglutinin type membrane proteins, it undergoes a conformational change in response to pH conferring the ability to puncture the lipid membrane.

UNIPROT and BLASTP classifies *cap0166* as a member of the YvrJ superfamily based on its transmembrane region, rather than the intracellular coiled coils. The original *yvrJ* was annotated in the *B. subtilis* genome and remains uncharacterised. Referring to the *B. subtilis* Genbank record (accession: NC_000964) *yvrJ* can be found adjacent to the CDS *sigO* coding for an ECF σ^{70}. The rest of this locus has been characterised in some detail. In 2009, MacLellan *et al.,* reported that the gene *oxdC*, which lies immediately downstream of *yvrJ* codes for oxalate decarboxylase, which is regulated by an ECF sigma factor YvrI/SigO under the control of three then unknown accessory proteins,YvrJ, YvrL and YvrHa. The expression of oxalate decarboxylase is a stress response to low pH where oxalate and H^+ is converted to formate and CO_2, thereby raising pH (MacLellan *et al.,* 2009). The gene *yvrHa* has been subsequently shown to function as a two component signalling system along with *yvrG* (Serizawa *et al.,* 2005). Interestingly, YvrL has been characterised as an

acid sensing, pH sensitive, membrane bound, anti-sigma factor and renamed to RsiO (Davis *et al.,* 2016). This leaves the question of the function of YvrJ of *B. subtilis* and *cap0166* of *C. acetobutylicum*.

It would be tempting to hypothesise that the function of *cap0166* is an acid sensing, membrane bound anti-sigma factor, however, with the discovery of RsiO in *B. subtilis* and its characterisation as an anti-sigma factor regulator of *sigO* (Xue *et al.,* 2016) it seems to be an unlikely scenario, given that *yvrJ* is located within the same genomic locus and is also thought to be a regulator of *sigO*. It would seem unlikely that both are anti-sigma factors.

An alternative hypothesis should be considered. It seems likely, given previously mentioned associations with transcription factors that *cap0166* is an acid sensing, membrane bound, one component transcriptional activator which regulates the expression of σ^Y and possibly other genes in a putative σ^Y regulon as an acid stress response regulator. One component systems are the dominant form of signal transduction for regulation of responses to the prokaryotic cellular environment (Ulrich *et al.,* 2005). A possible mechanism for signal transduction and regulatory activity is as follows: The membrane anchoring region of the homodimer is labile yet stable under neutral pH conditions through electrostatic interactions between the transmembrane helices. A drop in pH results in interference of these electrostatic interactions as H^+ ions diffuse between the labile helices and displace positively charged residues allowing for a separation of these two domains. The disassociation also results in sufficient solubility of the transmembrane domain and allows for the diffusion into the cytoplasm before making its way to the chromosome where it binds at the putative P_{sigY} sequence upstream of the *cap0166/7* operon, up-regulating the expression of *cap0167*/*sigY* in a positive feedback loop. Then σ^Y up-regulates the σ^Y regulon which possibly includes the *sol* operon and other genes listed in table 5. The expression of the *sol* operon results in the re-assimilation of the free carboxylic acids through CTFA/B activity (CTFA/B does not have activity for the conjugate bases

acetate or butyrate) before conversion to ethanol and butanol by ADHE1, resulting in a drop in environmental pH.

Further evidence that *cap0166* serves as a transcriptional activator can be taken from Jones *et al.,* (2008) where knock down asRNA targeted at *cap0166* and *cap0167* both resulted in a drop in solvent concentrations observed at 96 hours (see figure 8). Given that the plasmid negative control produced a significant effect to the strain phenotype it would be important to make use of more advanced homologous recombination in-frame deletion technology (Heap *et al.,* 2010; Ehsaan *et al.,* 2015) to independently investigate the roles of *cap0166* and *cap0167*. Further experiments could include analytical centrifugation of reconstituted phospholipid liposomes containing recombinant protein product of *cap0166* at pH conditions ranging from 7-5 as demonstrated in the characterisation of the influenza ion channel pH dependant tetramer (Salom *et al.,* 2000).

In order to confirm the role of *cap0166* as a DNA binding transcriptional activator, the protein product might be tested in DNA binding experiments using the intergenic upstream promotor region of *cap0166* and *cap0167* - or perhaps the *sol* operon, and expressed and purified *cap0166*.

5.2.6 Protein modelling using Iterative Threading Assembly Alignment (I-TASSER)

In an effort to provide further validation for the hypothetical function of *cap0166* the I-TASSER predictive algorithm was employed to produce a 3D model of the structure based on the protein sequence (Roy *et al.,* 2010; Yang and Zhang, 2015; Yang *et al.,* 2015). Threading methods allow for the proposition of a hypothetical structure by matching sequences to previously solved protein domain structures.

The amino acid sequence was input into the I-TASSER algorithm which was ran under default conditions. The results included 5 protein structure models with the top scoring model producing a confidence score (C) of -1.65 where C scores tend to range between -5 and 2, higher scores indicating higher confidence. C scores are calculated based on significance of threading template alignments and the convergence parameters of the structure assembly simulations. The highest ranking model (model 1, see figure 52) also produced a TM score of 0.51 ± 0.15 which is based on accuracy of threading of the model to known structures available on PDB, TM scores greater than 0.5 are generally accepted as reliable (Yang *et al.,* 2015). Upon inspection of model 1 it also appeared to be the best fit in accordance with previous modelling results; containing a helical structure at the N-terminus consistent with the coiled coil models in figure 51. The C-terminus which had previously modelled as a transmembrane helix was presented as an interrupted HTH rather than a single 20 residue helix. This might be expected, as per the hypothetical role of CAP0166 as a pH sensitive membrane anchored transcription factor, the model may be representative of the protein in its soluble conformation. The PDB database, which I-TASSER sources to predict structure, is biased towards soluble proteins, since so comparatively few membrane proteins have been submitted due to their insolvent nature resulting in an inability to coalesce as protein crystals during screens - to date there are 938 membrane proteins of known structure (White, 2019) compared to the total of 151228 available protein structures on PDB. The lack of available structural data also explains why SwissModel failed to predict secondary structure for the c-terminal transmembrane helix.

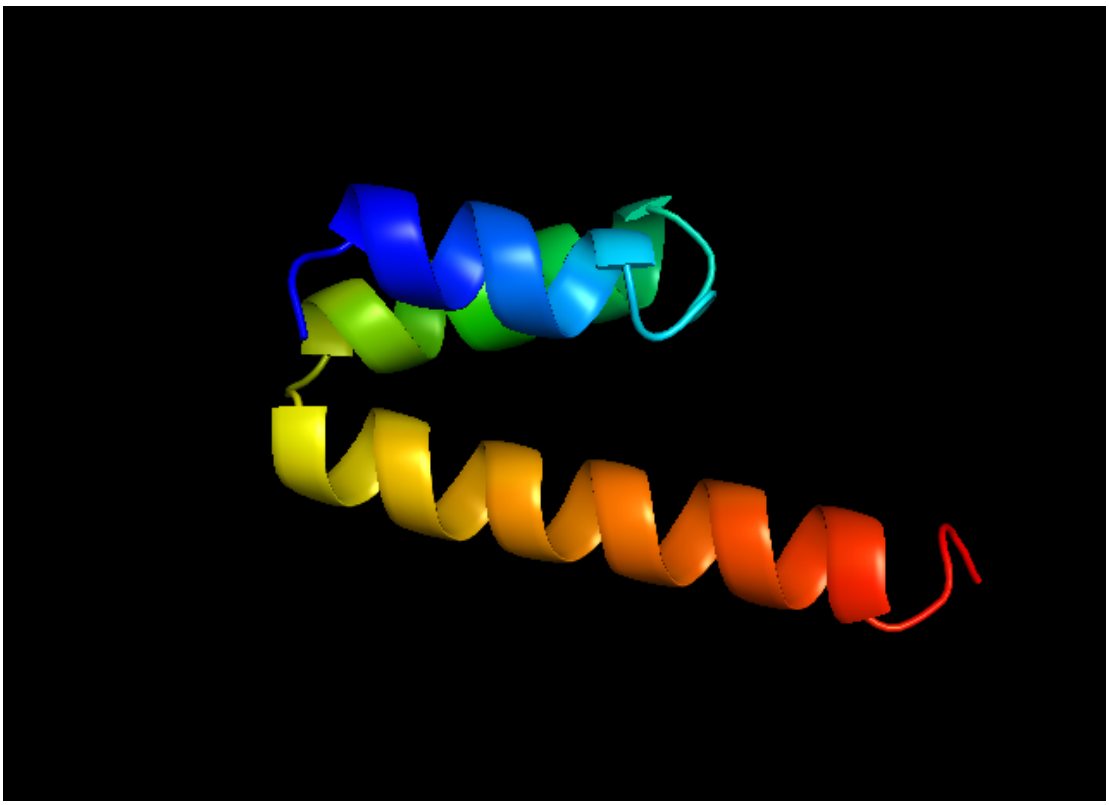


Figure 52: Model 1 result produced by I-TASSER protein modelling algorithm displayed as a cartoon ribbon diagram with the rainbow colouration applied in PYMOL. The C-terminus is coloured blue and the N-terminus is coloured red. The N terminal helix previously identified as a coiled coil by Swiss Model and UNIPROT is clearly present. The previously identified 20 amino acid C-terminal transmembrane helix identified by TMHMM and TMPred, also logged on UNIPROT was not modelled this way, instead a HTH motif was generated.

Membrane anchored proteins which are sensitive to pH, undergoing conformational changes are an established biological phenomenon, as previously discussed the VP4 membrane penetration spike of rhesus rotavirus, undergoes conformational changes in response to pH in order to penetrate the host membrane. The diphtheria toxin becomes insoluble in response to lowered pH at which point a transmembrane region emerges facilitating the transfer of the c-domain into the cytoplasm. The toxin gene is in fact coded by a prophage (Blewitt *et al.,* 1985; Bell & Eisenberg, 1996).

It ought to be considered that the putative transcription factor is possibly a leucine zipper, with the zipper region actually spanning the putative TMH, which contains 3 leucines and 4 isoleucines. Leucine zippers are quite small and a 53 residue protein is not unusual for a leucine zipper (Krylov & Vinson, 2001).

A protein dimer was generated using SymmDock (Schneidman-Duhovny *et al.,* 2005; Schneidman-Duhovny *et al.,* 2005) using model 1 from I-TASSER from as the input and the algorithm parameters for symmetry order was set to 2 in order to generate dimers, as informed by the previous model results generated by SwissModel. SymmDock generated geometry based models and results are not dependent on predicted residue interactions, therefore models are not ranked by probability. The result which contained a region resembling an N-terminal coiled coil dimer similar to the model generated by SwissModel (figure 51) was selected (figure 53).

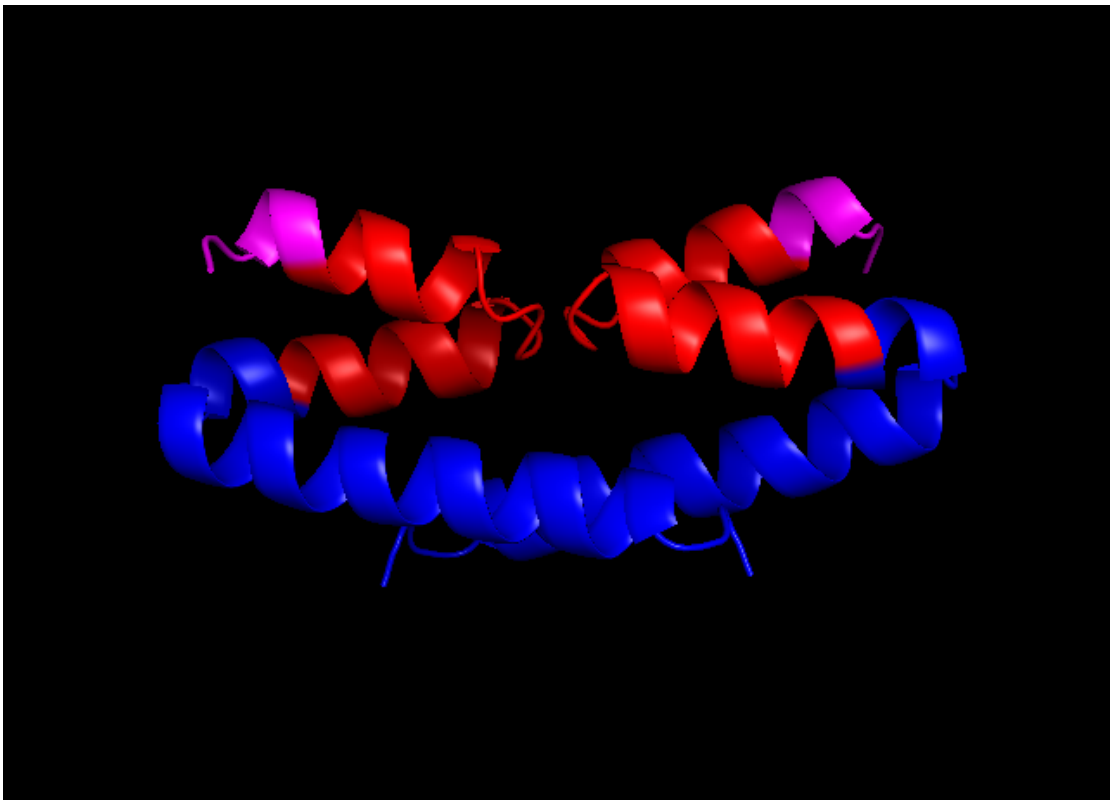


Figure 53: Model 1 result from I-TASSER displayed as a dimer using SymmDock. Protein sequence has been coloured according to the scheme used by TMHMM (see figure 10). Extracellular residues: Magenta; transmembrane residues: red; intracellular residues: blue.

Finally, to explore the hypothesis that CAP0166 is a transcriptional activator, the amino acid sequence was processed using DP-Bind which a DNA binding prediction algorithm utilising 3 machine learning methods for identification of key DNA binding residues. The machine learning algorithms are supervised learning algorithms and are trained on a set of 62 solved non-redundant crystal structures (a very small training set, relatively speaking) Hwang, *et al.,* (2007). DP-Bind identified 9 residues as majority consensus - residues in binding positions for the majority of the machine learning methods used i.e. 2 out of 3, and 6 of these residues were found to be strictly conserved i.e. all three of the machine learning methods identified these residues as DNA binding. All of these residues with one exception appeared in the N terminus intracellular region which was modelled as a coiled coil based on the DNA binding coiled coil of MtaN (figure 51). These residues (table 22) were colour labeled the CAP0166 dimer in PyMOL (figure 55). When the model is rotated 90˚ upwards the “bottom” of the protein dimer is displayed where all of the DNA binding residues are visible suggesting that this area would be where the protein recognises and binds to the corresponding DNA bases (see table 6 for positive DP-Bind results).

This raises the next question: does the CAP0166 dimer model have the correct dimensions to bind DNA with the “bottom” facing area orientated towards the bases? This question was explored using the DNA PDB model structure 6GN4 of *Homo sapiens* A-form DNA solved by NMR. The major groove was measured using the PyMOL Wizard>Measure tool producing a measurement of 17.9 Å, this was compared to the width of the CA_P0166 dimer model which produced measurements of 13.4 Å and 13.9 Å. A direct visual comparison was also made where an image of both the DNA (6GN4) model was superimposed with an image of the CAP0166 dimer placed directly over the major groove (figure 56). All indications suggest that the putative transcription factor would be able to fit within the major groove without steric hinderance and with DNA binding residues correctly orientated to recognise exposed bases.

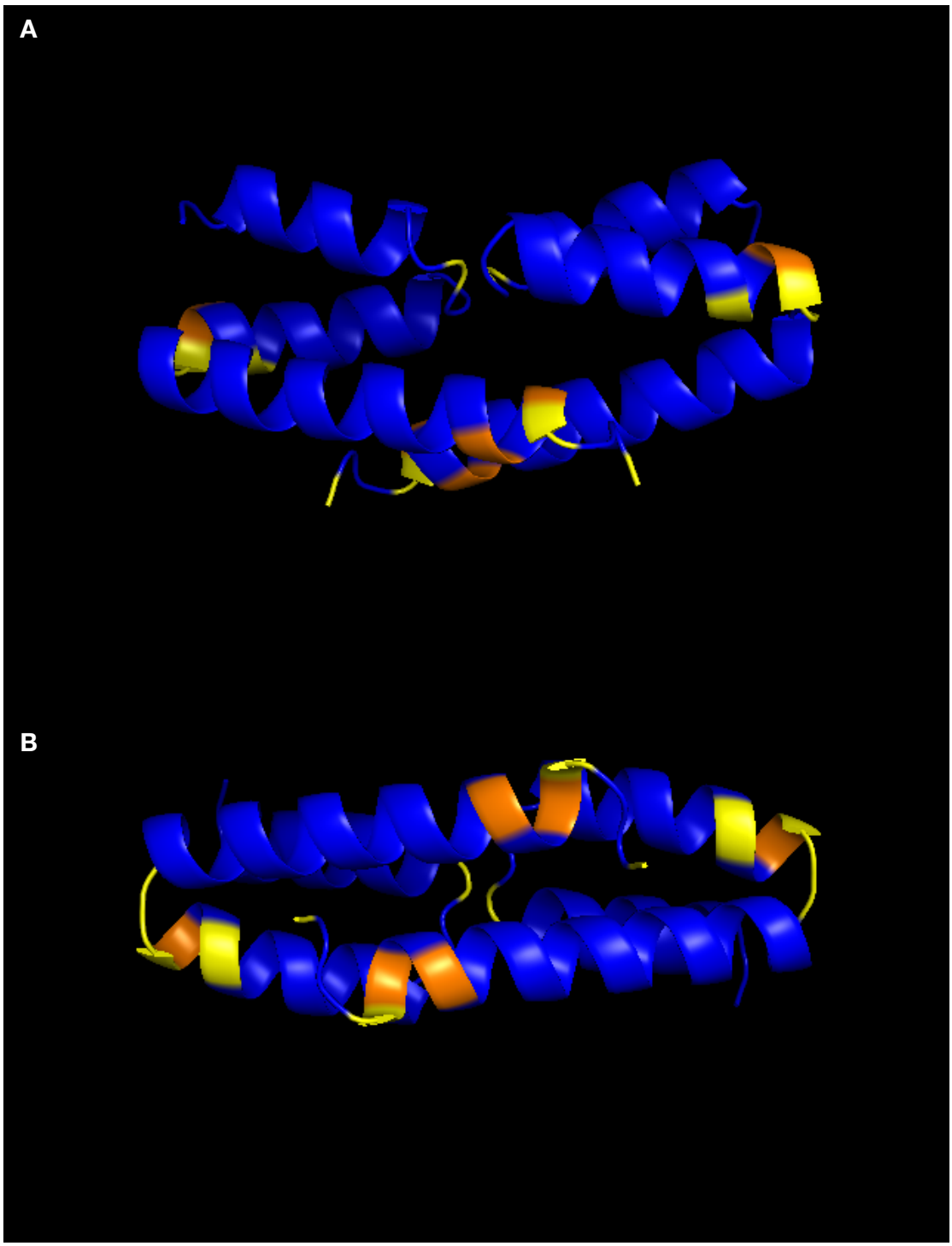


Figure 55: Positive scoring residues and their respective positions from the DP-Bind results were used to label the CA_P0166 structural model (blue default). Majority consensus residues are displayed in orange and strict consensus residues are displayed in yellow. Image A shown from typical lengthways side view with the C-termini at the top and the N-termini at the bottom. Image B has been rotated 90° to show the N-terminal intracellular coiled coils where all DNA binding residues can be clearly seen.

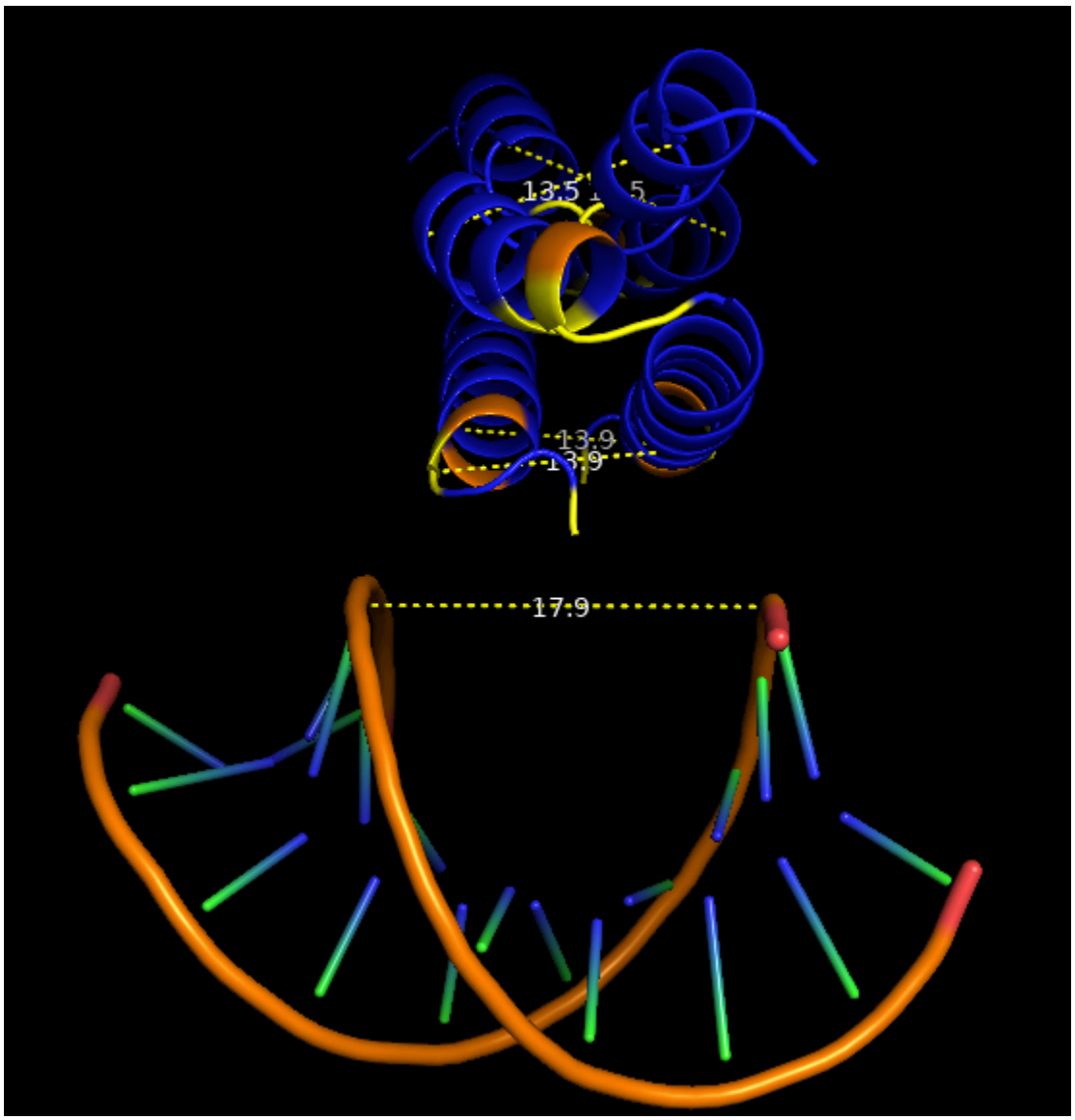


Figure 56: The putative structure of CA_P0166 orientated complementarily to the major groove of DNA to demonstrate that it would be capable of binding without steric hinderance and with predicted DNA binding residues correctly distributed in order to recognise DNA bases. DNA structure is A-form *Homo sapiens* obtained by NMR (PDB: 6GN4). Measurement annotations are in angstroms (Å).

Pos	Resi	SVM	SVM Prob	KLR	KLR Prob	PLR	PLR Prob	Maj Con	Str Con
14	G	1	0.6652	1	0.6021	1	0.6517	1	1
26	R	1	0.6485	1	0.6396	1	0.6996	1	1
28	E	1	0.5111	1	0.5534	0	0.7136	1	NA
29	G	1	0.5574	1	0.5774	1	0.5515	1	1
30	K	1	0.5492	1	0.6276	1	0.7088	1	1
47	N	1	0.5695	0	0.5101	1	0.6946	1	NA
49	N	1	0.5680	0	0.5107	1	0.5940	1	NA
50	T	1	0.6708	1	0.6422	1	0.7498	1	1
53	K	1	0.6565	1	0.5778	1	0.6041	1	1

Table 22: DP-Bind results showing CA_P0166 positions (pos) of residues (resi) which received either positive majority consensus scores or strict consensus scores. Support vector machine (SVM), kernel logistic regression (KLR), and penalised logistic regression (PLR) represent the three machine learning methods used by DP-Bind and their relative probability scores (Prob). Residues achieving a positive score for 2 out of 3 of the results are classified as majority consensus (Maj Con). Residues achieving a positive score for all three methods are classified as strict consensus (Str Con).

5.2.7 Further observations drawn from the results TcdR family promotor search algorithm in light of the characterisation of σ^Y

Given the insight provided from the investigation into the role and activity of *cap0166* (putative transcription factor) and *cap0167* (*sigY*), and given the suspected shared consensus between the TcdR -35 site and the σ^Y -35 it was interesting to review the remainder of the 85 genes identified using the TcdR family promotor identification algorithm (i.e. the genes identified as downstream of putative promotors found using the more ambiguous -10 sequence) as potential members of a σ^Y regulon implicated in regulation of acid stress response. Again, the transcriptomic data generated by Yoo *et al.,* (2015) was utilised to compare regulation of all of the genes downstream of a putative σ^Y binding promotor during acidogenesis and solventogenesis. A calculation dividing mRNA molecules per cell during solventogenesis by mRNA molecules per cell during acidogenesis was made in order to highlight genes up-regulated during solventogenesis i.e solventogenesis mRNA molecules cell^{-1}/acidogenesis mRNA molecules cell^{-1} = up-regulation factor (figure 57). This showed that of the 4 highest increases in gene transcription (all greater than 3 fold increase) 3 of those genes had multiple (2 or 3) σ^Y binding promotor sequences with the *cap0166*/*cap0167* (*sigY*) operon having 2 promotors, *galK* also having 2 promotors, *nrpE* having 3 promotors and *cac0717* having just 1 σ^Y promotor. This strongly suggests there is significance to the TcdR/σ^Y promotor sequence and that it may be regulated by pH conditions.

The most interesting of these genes is annotated as *nrpE* which putatively codes for a neutral extracellular metalloprotease. This gene has 3 σ^Y binding promotors in the intergenic region upstream of its start codon and sees a 4.4 fold increase in transcription when comparing acidogenic (pH 6.3) conditions to solventogenic conditions (pH 4.4). NprR is structurally related to NprE (Rocha-Estrada *et al.,* 2010) and is the quorum sensing transcriptional regulator and a key component RNPP-type family (Rap, NprR, PlcR, and PrgX) of quorum sensing systems. NprR binds mature quorum sensing peptides resulting in transcriptional

regulation in *Bacillus cereus* (Zouhir *et al.,* 2013). NprB is known to be responsible for the maturation of quorum sensing peptide in *B. cereus* (Pomerantsev *et al.,* 2009). The fact that *nprE* is highly up-regulated during solventogenic fermentations (conducted at pH 4.4) and has three upstream σ^Y binding promotors is interesting in terms of understanding the putative role of σ^Y and the conditions it operates under, however, the story becomes yet more interesting when attention is turned to the gene *oppA.* This gene, which has one putative σ^Y promotor, is up-regulated 2.67 fold under solventogenic conditions compared to fermentations carried out during acidogenic conditions. This gene codes for the oligopeptide ABC-importer and is essential for re-uptake of the mature quorum sensing peptide (Slamti and Lereclus, 2019). Recently Kotte *et al.,* (2017) described an RNPP-type quorum sensing system in *C. acetobutylicum* which appears to be responsible for triggering solventogenesis. If the σ^Y system and the RNPP quorum sensing are linked, it may be in accordance with the following regime: concentration gradients are the norm in nature, unlike the shaken flasks or stirred bioreactors used in the lab, certain subpopulations will respond to acid stress before others further up the hydrogen ion concentration gradient. Upon reception of a sufficiently strong pH signal, up-regulation of the quorum sensing peptidase (possibly *nprE*) and subsequent secretion into the extracellular environment propagates a distress signal as immature peptide is cleaved and mature peptide diffuses amongst the bacterial population triggering solventogenesis in a bid to collectively mitigate the pH stress. The gene *oppA* has high expression in acidogenesis (4.083 SD ± 0.509 mRNA molecules $cell^{-1}$) increasing during solventogenesis (10.902 SD ± 4.983 mRNA molecules $cell^{-1}$), which would allow for initial reception of the peptide signal, followed by increased sensitivity to a higher quorum sensing signal during solventogenesis.

The other two genes (besides the *sigY* operon) which were highly up-regulated during solventogenesis were *galK* and *cac0717*. *cac0717* has one putative σ^Y binding promotor, is listed as a hypothetical protein, and the nearest BLASTP identity is a “chitinase.” The reason for the up-regulation of the gene *galK* is not

clear, however, it may be related to exopolysaccharide breakdown, a phenomenon which occurs after transitioning to solventogenesis.

There was one other gene which had two promotors - the serene protein kinase *prkA* which is responsible for modulation of sporulation via regulation of σ^K in *B subtilis* (Yan *et al.,* 2015). The transcripts were undetectable in acidogenesis, and measured 0.066 (SD ± 0.012) mRNA molecules cell^{-1} during solventogenesis. Any role implicated with initiation of sporulation via a σ^Y, PrkA, σ^K pathway would be in keeping with the observations of Jones *et al*., (2008) where asRNA approaches implicated *cap0166* and *cap0167* in the sporulation differentiation process.

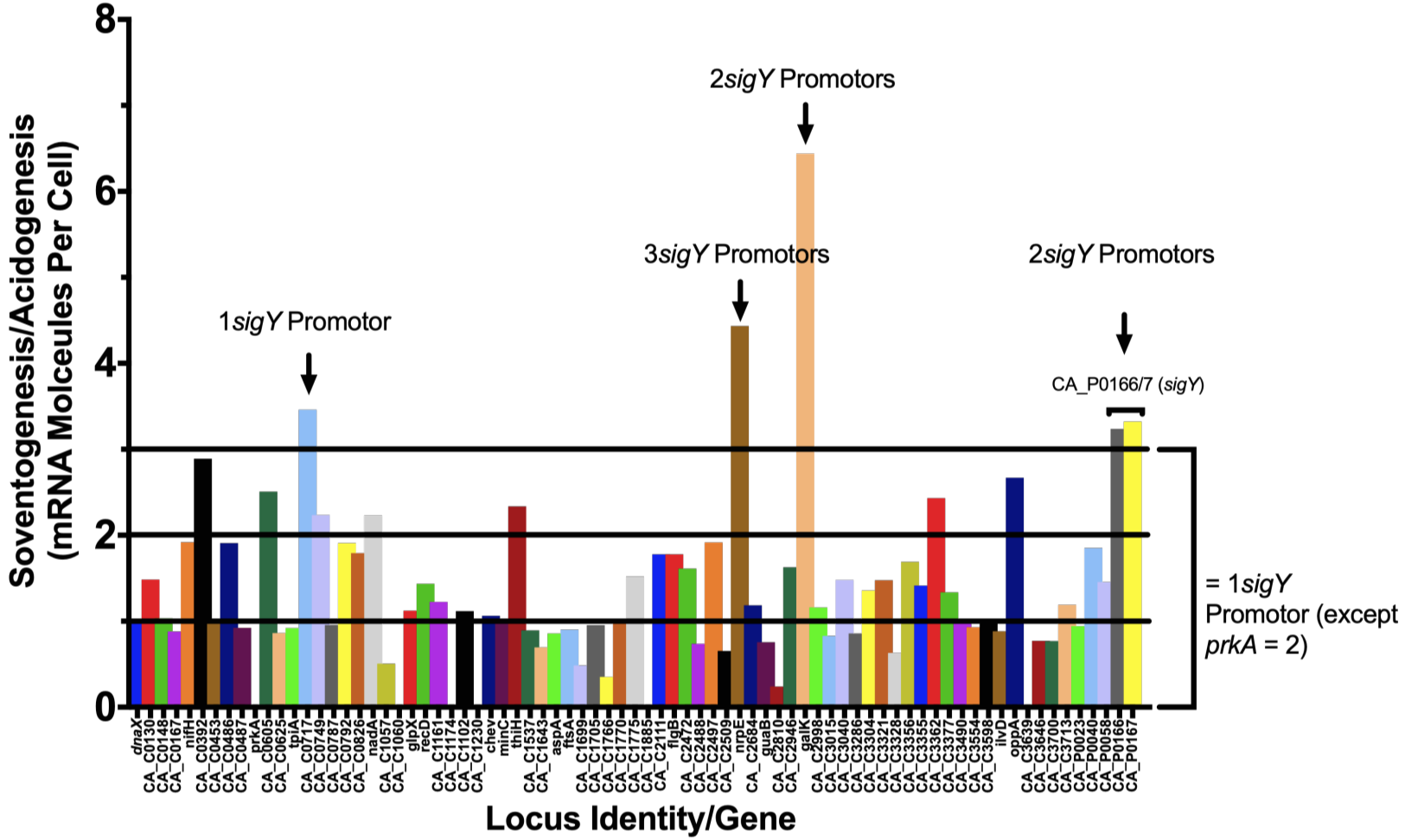


Figure 57: Transcriptome data (Yoo *et al.,* 2015) encompassing a putative SigY regulon - mRNA transcripts per cell based on the promotor sequence parameters derived from Martin-Verstraete *et al.*, (2016). Transcript concentration measured during solventogenesis and expressed per molecule mRNA per cell in acidogenesis.

Conclusions

The aim of this study was to test the applicability of a lactose inducible sigma factor for specific control of recombinant genes for metabolic engineering purposes. The system was not sufficiently able to prevent expression of pathway genes in the absence of inducer as a clear difference in phenotype was shown. Furthermore the response to addition of inducer was poor and either showed no increase in pathway product yield, or a phenotype which was detrimental to productivity. Ideally the system ought to tuned by exchanging the transcription factor/promotor system from P_{bgaL}/BgaR to something more sensitive to bring the P_{tcdB}/TcdR system into the dynamic inducer range. Alternatively random mutagenesis techniques might be used to create a promotor library of P_{tcdB} to lower the strength, dampen leakiness and bring the system into the dynamic range. Furthermore, native copies of the genes under the control of the BgaR/TcdR system, ought to be deleted from the chromosome in order to provide a more accurate indication of the response to inducer.

The negative control for the BgaR/TcdR system produced some interesting results, where solvent production occurred at a higher rate than in the wild type. This is not an ideal quality for such a system, and any genetic components used in a synthetic control system ought to be completely inert in the host organism. In order to eliminate any background activity, whether it's a result of random matching promotor sequence binding sites, the presence of a related, yet undiscovered sigma factor regulon, or the presence of vestigial relics from a common ancestor, a similar orthogonal system could employ a sigma factor from an phylogenetically distinct organism, this ought to mitigate against unwanted activity. Further assurance might be obtained by first screening genomes using the algorithm designed for this study and checking that there are no binding sites in problematic positions before selecting an appropriate sigma factor.

Regardless of the fact that the inducible promotor system failed to produce a dose related response, the results comparing the strains *C. acetobutylicum* + *ter*_FLAG, *C. acetobutylicum* + *ter*_FLAG + *bgaR_tcdR* + ButOH, and *C.*

*acetobutylicum bgaR_tcdR*_ctrl suggests that generating an excess of NADH has a separate advantage to unregulated expression of the genes responsible for converting acetyl-CoA to butanol and when both systems are active in the same strain they have an additive effect, resulting in a highly productive strain. Although it may be the case that this benefit might be achieved as a result of expressing one or a combination of the four genes in the synthetic operon of *C. acetobutylicum* + *ter*_FLAG + *bgaR_tcdR* + ButOH. Appropriate controls would be required in order to rule this out.

A follow-up bioinformatics study sought to identify one or more genes responsible for the altered phenotype observed in the negative control fermentation of the strain expressing the BgaR/TcdR system alone. This highlighted TcdR family conserved promotor sequences upstream of *C. acetobutylicum* ECF sigma factors which may be implicated in solventogenesis regulation, namely *cap0167* also known as *sigY*. Protein modelling of a co-expressed putative regulator - Cap0166 provided the basis for a hypothesis for its role as a membrane anchored one-component pH sensing transcriptional activator responsible for regulating solvent production via *sigY* activation in the response to acid stress from the environment.

Sequences matching the TcdR/σ^Y -35 consensus sequence were discovered upstream of the *sol* operon, which were corroborated by other annotation studies. In order to confirm the significance of these putative σ^Y binding sequences a chromatin immunoprecipitation coupled to high-throughput sequencing (ChIP-seq) (Grainger *et al*., 2009) study might be undertaken in order to reveal the genome wide associations to help characterise the regulon.

Interestingly the *sol* operon and *adc* possess upstream putative σ^Y promotors, however, they do not possess Rex operators. On the other hand *adhE2* doesn't possess a putative σ^Y promotor but does possess a Rex operator. When comparing the expression of the alcohol dehydrogenases during alcohologenesis and solventogenesis, Yoo *et al.,* (2015) showed that during alcohologenesis *adhE2* was largely responsible for butanol production, and during solventogenesis *adhE1* (part of the *sol* operon) expressed the dominant aldehyde dehydrogenase and *bdhB* was

the responsible for conversion to butanol. This might to suggest that σ^Y and the *sol* operon is responsible for regulation of solvent production in response to acid stress and Rex regulates the expression of genes producing solvents in order to achieve redox homeostasis.

Bibliography

ABOULNAGA, E. H., PINKENBURG, O., SCHIFFELS, J., EL-REFAI, A., BUCKEL, W. & SELMER, T. 2013. Effect of an Oxygen-Tolerant Bifurcating Butyryl Coenzyme A Dehydrogenase/Electron-Transferring Flavoprotein Complex from Clostridium difficile on Butyrate Production in *Escherichia coli*. Journal of Bacteriology, 195, 3704-3713.

AL-HINAI, M. A., JONES, S. W. & PAPOUTSAKIS, E. T. 2015. The Clostridium Sporulation Programs: Diversity and Preservation of Endospore Differentiation. Microbiology and Molecular Biology Reviews, 79, 19-37.

ALPANDA, S. & PERALTA-ALVA, A. 2010. Oil crisis, energy-saving technological change and the stock market crash of 1973-74. Review of Economic Dynamics, 13, 824-842.

ALSAKER, K. V., PAREDES, C. J. & PAPOUTSAKIS, E. T. 2005. Design, optimization and validation of genomic DNA microarrays for examining the *Clostridium acetobutylicum* transcriptome. Biotechnology and Bioprocess Engineering, 10, 432-443.

ALTIZER, S., OSTFELD, R., JOHNSON, P., KUTZ, S. & HARVELL, C. 2013. Climate Change and Infectious Diseases: From Evidence to a Predictive Framework. Science, 341, 514-519.

AMADOR-NOGUEZ, D., BRASG, I. A., FENG, X. J., ROQUET, N. & RABINOWITZ, J. D. 2011. Metabolome Remodeling during the Acidogenic-Solventogenic Transition in *Clostridium acetobutylicum*. Applied and Environmental Microbiology, 77, 7984-7997.

ANDERSCH, W., BAHL, H. & GOTTSCHALK, G. 1983. Level of enzymes involved in acetate, butyrate, acetone and butanol formation by *Clostridium acetobutylicum* . European Journal of Applied Microbiology and Biotechnology, 18, 327-332.

ANDREASEN, A. A. & STIER, T. J. B. 1954. Anaerobic nutrition of *Saccharomyces cerevisiae*. II. Unsaturated fatty and requirement for growth in a defined medium. Journal of Cellular and Comparative Physiology, 43, 271-281.

ANTONI, D., ZVERLOV, V. V. & SCHWARZ, W. H. 2007. Biofuels from microbes. Applied Microbiology and Biotechnology, 77, 23-35.

BAHL, H., ANDERSCH, W., BRAUN, K. & GOTTSCHALK, G. 1982. Effect of pH and butyrate concentration on the production of acetone and butanol by *Clostridium acetobutylicum* grown in continuous culture. European Journal of Applied Microbiology and Biotechnology, 14, 17-20.

BAHL, H., GOTTWALD, M., KUHN, A., RALE, V., ANDERSCH, W. & GOTTSCHALK, G. 1986. Nutritional factors affecting the ratio of solvents produced by *Clostridium acetobutylicum*. Applied and Environmental Microbiology, 52, 169-172.

BALASSA, B. 1985. Exports, policy choices, and economic-growth in developing-countries after the 1973 oil shock. Journal of Development Economics, 18, 23-35.

BARRANGOU, R., FREMAUX, C., DEVEAU, H., RICHARDS, M., BOYAVAL, P., MOINEAU, S., ROMERO, D. A. & HORVATH, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science, 315, 1709-1712.

BAUMEISTER, C. & KILIAN, L. 2016. Forty Years of Oil Price Fluctuations: Why the Price of Oil May Still Surprise Us. Journal of Economic Perspectives, 30, 139-160.

BEHRENS, S., MEYER, U., SCHANKIN, H., LONETTO, M. A., FISCHER, R. J. & BAHL, H. 2000. Identification of two genes encoding putative new members of the ECF subfamily of eubacterial RNA polymerase sigma factors in Clostridium acetobutylicum. Journal of Molecular Microbiology and Biotechnology, 2, 265-269.

BENKERT, P., BIASINI, M. & SCHWEDE, T. 2011. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics, 27, 343-350.

BERTSCH, J., PARTHASARATHY, A., BUCKEL, W. & MULLER, V. 2013. An Electron-bifurcating Caffeyl-CoA Reductase. Journal of Biological Chemistry, 288, 11304-11311.

BLANCHARD, O. J. & RIGGI, M. 2013. Why are the 2000s so different from the 1970s? A structural interpretation of changes in the macroeconomic effects of oil prices. Journal of the European Economic Association, 11, 1032-1052.

BLEWITT, M. G., CHUNG, L. A. & LONDON, E. 1985. Effect of pH on the conformation of diphtheria-toxin and its implications for membrane penetration. Biochemistry, 24, 5458-5464.

BOND-WATTS, B. B., BELLEROSE, R. J. & CHANG, M. C. Y. 2011. Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways.

Nature Chemical Biology, 7, 222-227.

BRUDER, M. R., PYNE, M. E., MOO-YOUNG, M., CHUNG, D. A. & CHOU, C. P. 2016. Extending CRISPR-Cas9 Technology from Genome Editing to Transcriptional Engineering in the Genus *Clostridium*. Applied and Environmental Microbiology, 82, 6109-6119.

BUCKEL, W. & THAUER, R. K. 2013. Energy conservation via electron bifurcating ferredoxin reduction and proton/Na+ translocating ferredoxin oxidation. Biochimica Et Biophysica Acta-Bioenergetics, 1827, 94-113.

BUELTER, T., HAWKINS, A., KERSH, K., MEINHOLD, P., PETERS, M. & SUBBIAN, E. 2009. Engineered microorganisms for producing n-butanol and related methods. France patent application.

CALERO, P. & NIKEL, P. I. 2019. Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms. Microbial Biotechnology, 12, 98-124.

CANADAS, I. C., GROOTHUIS, D., ZYGOUROPOULOU, M., RODRIGUES, R. & MINTON, N. P. 2019. RiboCas: A Universal CRISPR-Based Editing Tool for Clostridium. Acs Synthetic Biology, 8, 1379-1390.

CATO, E. P. & STACKEBRANDT, E. 1989. Clostridia. Biotechnology Handbooks. 1st ed. Springer Science & Business Media New York: Springer US.

CHEN, H. Y., BJERKNES, M., KUMAR, R. & JAY, E. 1994. Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* messenger-RNAs. Nucleic Acids Research, 22, 4953-4957.

CHOI, S. K. & SAIER, M. H. 2005. Regulation of sigL expression by the catabolite control protein CcpA involves a roadblock mechanism in *Bacillus subtilis*: Potential connection between carbon and nitrogen metabolism. Journal of Bacteriology, 187, 6856-6861.

CHOWDHURY, N. P., KAHNT, J. & BUCKEL, W. 2015. Reduction of ferredoxin or oxygen by flavin-based electron bifurcation in *Megasphaeraelsdenii*. Febs Journal, 282, 3149-3160.

CHOWDHURY, N. P., KLOMANN, K., SEUBERT, A. & BUCKEL, W. 2016. Reduction

of Flavodoxin by Electron Bifurcation and Sodium Ion-dependent Reoxidation by NAD(+) Catalyzed by Ferredoxin-NAD(+) Reductase (Rnf). Journal of Biological Chemistry, 291, 11993-12002.

CHOWDHURY, N. P., MOWAFY, A. M., DEMMER, J. K., UPADHYAY, V., KOELZER, S., JAYAMANI, E., KAHNT, J., HORNUNG, M., DEMMER, U., ERMLER, U. & BUCKEL, W. 2014. Studies on the Mechanism of Electron Bifurcation Catalyzed by Electron Transferring Flavoprotein (Etf) and Butyryl- CoA Dehydrogenase (Bcd) of *Acidaminococcus fermentans**. Journal of Biological Chemistry, 289, 5145-5157.

COOKSLEY, C. M., ZHANG, Y., WANG, H. Z., REDL, S., WINZER, K. & MINTON, N. P. 2012. Targeted mutagenesis of the *Clostridium acetobutylicum* acetone-butanol-ethanol fermentation pathway. Metabolic Engineering, 14, 630-641.

COUNTRIES, O. O. T. P. E. 2019. The OPEC Annual Statistical Bulletin. In: MACNEILL, M. & QUINN, M. (eds.) 54 ed. Vienna, Austria.

CZERNY, C., EDER, E. & RUNGER, T. M. 1998. Genotoxicity and mutagenicity of the alpha,beta-unsaturated carbonyl compound crotonaldehyde (butenal) on a plasmid shuttle vector. Mutation Research-DNA Repair, 407, 125-134.

DAVIS, M. C., SMITH, L. K. & MACLELLAN, S. R. 2016. The atypical two-subunit sigma factor from *Bacillus subtilis* is regulated by an integral membrane protein and acid stress. Microbiology-Sgm, 162, 398-407.

DEKISHIMA, Y., LAN, E., SHEN, C., CHO, K. & LIAO, J. 2011. Extending Carbon Chain Length of 1-Butanol Pathway for 1-Hexanol Synthesis from Glucose by Engineered Escherichia coli. Journal of the American Chemical Society, 133, 11399-11401.

DEMMER, J. K., BERTSCH, J., OPPINGER, C., WOHLERS, H., KAYASTHA, K., DEMMER, U., ERMLER, U. & MULLER, V. 2018. Molecular basis of the flavin-based electron-bifurcating caffeyl-CoA reductase reaction. Febs Letters, 592, 332-342.

DEMMER, J. K., CHOWDHURY, N. P., SELMER, T., ERMLER, U. & BUCKEL, W. 2017. The semiquinone swing in the bifurcating electron transferring flavoprotein/ butyryl-CoA dehydrogenase complex from *Clostridium difficile*. Nature Communications, 8.

DONG, H., TAO, W., ZHU, L., ZHANG, Y. & LI, Y. 2011. CAC2634-disrupted mutant

of *Clostridium acetobutylicum* can be electrotransformed in air. Letters in Applied Microbiology, 53, 379-382.

DORMITZER, P. R., NASON, E. B., PRASAD, B. V. V. & HARRISON, S. C. 2004. Structural rearrangements in the membrane penetration protein of a non-enveloped virus. Nature, 430, 1053-1058.

DUPUY, B., RAFFESTIN, S., MATAMOUROS, S., MANI, N., POPOFF, M. R. & SONENSHEIN, A. L. 2006. Regulation of toxin and bacteriocin gene expression in *Clostridium* by interchangeable RNA polymerase sigma factors. Molecular Microbiology, 60, 1044-1057.

DUPUY, B. & SONENSHEIN, A. L. 1998. Regulated transcription of *Clostridium difficile* toxin genes. Molecular Microbiology, 27, 107-120.

DURRE, P. 2011. Fermentative production of butanol - the academic perspective. Current Opinion in Biotechnology, 22, 331-336.

DURRE, P., BOHRINGER, M., NAKOTTE, S., SCHAFFER, S., THORMANN, K. & ZICKNER, B. 2002. Transcriptional regulation of solventogenesis in Clostridium acetobutylicum. Journal of Molecular Microbiology and Biotechnology, 4, 295-300.

DUSSEAUX, S., CROUX, C., SOUCAILLE, P. & MEYNIAL-SALLES, I. 2013. Metabolic engineering of *Clostridium acetobutylicum* ATCC 824 for the high-yield production of a biofuel composed of an isopropanol/butanol/ethanol mixture. Metabolic Engineering, 18, 1-8.

EL-MANSI, E. M. T., BRYCE, C. F. A., HARTLEY, S. & DEMAIN, A. D. 2007. Fermentation Microbiology and Biotechnology: A Historical Perspective. In: EL-MANSI, E. M. T. (ed.) Fermentation Microbiology and Biotechnology. 2nd ed. Boca Raton, FL, USA: CRC Press.

FARAH, D. & BABINEAU, K. 2019. Extra-regional Actors in Latin America The United States is not the Only Game in Town. PRISM, 8, 96-113.

FAWCETT, P., EICHENBERGER, P., LOSICK, R. & YOUNGMAN, P. 2000. The transcriptional profile of early to middle sporulation in Bacillus subtilis. Proceedings of the National Academy of Sciences of the United States of America, 97, 8063-8068.

FEKLISTOV, A., SHARON, B. D., DARST, S. A. & GROSS, C. A. 2014. Bacterial Sigma Factors: A Historical, Structural, and Genomic Perspective. Annual Review of

Microbiology, Vol 68, 68, 357-376.
FIMLAID, K. A. & SHEN, A. 2015. Diverse mechanisms regulate sporulation sigma factor activity in the Firmicutes. Current Opinion in Microbiology, 24, 88-95.
FISCHER, C. R., TSENG, H. C., TAI, M., PRATHER, K. L. J. & STEPHANOPOULOS, G. 2010. Assessment of heterologous butyrate and butanol pathway activity by measurement of intracellular pathway intermediates in recombinant *Escherichia coli*. Applied Microbiology and Biotechnology, 88, 265-275.
GEA 2012. Global Energy Assessment - Toward a Sustainable Future, Cambridge University Press, Cambridge, UK and New York, NY, USA and the International Institute for Applied Systems Analysis, Laxenburg, Austria.
GIBSON, D. G., YOUNG, L., CHUANG, R. Y., VENTER, J. C., HUTCHISON, C. A. & SMITH, H. O. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods, 6, 343-U41.
GIRBAL, L. & SOUCAILLE, P. 1994. Regulation of *Clostridium acetobutylicum* metabolism as revealed by mixed-substrate steady-state continuous cultures - role of NADH/NAD ratio and ATP pool. Journal of Bacteriology, 176, 6433-6438.
GLASCOCK, C. B. & WEICKERT, M. J. 1998. Using chromosomal lacI(Q1) to control expression of genes on high-copy-number plasmids in *Escherichia coli*. Gene, 223, 221-231.
GLAZYRINA, J., MATERNE, E. M., DREHER, T., STORM, D., JUNNE, S., ADAMS, T., GRELLER, G. & NEUBAUER, P. 2010. High cell density cultivation and recombinant protein production with *Escherichia coli* in a rocking- motion- type bioreactor. Microbial Cell Factories, 9.
GODSEY, M. H., BARANOVA, N. N., NEYFAKH, A. A. & BRENNAN, R. G. 2001. Crystal structure of MtaN, a global multidrug transporter gene activator. Journal of Biological Chemistry, 276, 47178-47184.
GOSSEN, M. & BUJARD, H. 1992. TIGHT CONTROL OF GENE-EXPRESSION IN MAMMALIAN-CELLS BY TETRACYCLINE-RESPONSIVE PROMOTERS. Proceedings of the National Academy of Sciences of the United States of America, 89, 5547-5551.
GOSSEN, M., FREUNDLIEB, S., BENDER, G., MULLER, G., HILLEN, W. &

BUJARD, H. 1995. TRANSCRIPTIONAL ACTIVATION BY TETRACYCLINES IN MAMMALIAN-CELLS. Science, 268, 1766-1769.

GOTTSCHALK, G. 1986. Bacterial Metabolism, New York, Springer-Verlag.

GRAINGER, D. C., LEE, D. J. & BUSBY, S. J. W. 2009. Direct methods for studying transcription regulatory proteins and RNA polymerase in bacteria. Current Opinion in Microbiology, 12, 531-535.

GREEN, E. M. 2011. Fermentative production of butanol - the industrial perspective. Current Opinion in Biotechnology, 22, 337-343.

GREEN, E. M. & BENNETT, G. N. 1996b. Inactivation of an aldehyde/alcohol dehydrogenase gene from Clostridium acetobutylicum ATCC 824. Applied Biochemistry and Biotechnology, 57-8, 213-221.

GREEN, E. M., BOYNTON, Z. L., HARRIS, L. M., RUDOLPH, F. B., PAPOUTSAKIS, E. T. & BENNETT, G. N. 1996. Genetic manipulation of acid formation pathways by gene inactivation in *Clostridium acetobutylicum* ATCC 824. Microbiology-SGM, 142, 2079-2086.

GRUBER, T. M. & GROSS, C. A. 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. Annual Review of Microbiology, 57, 441-466.

GYULEV, I. S., WILLSON, B. J., HENNESSY, R. C., KRABBEN, P., JENKINSON, E. R. & THOMAS, G. H. 2018. Part by Part: Synthetic Biology Parts Used in Solventogenic Clostridia. Acs Synthetic Biology, 7, 311-327.

HAHN, M. Y., RAMAN, S., ANAYA, M. & HUSSON, R. N. 2005. The mycobacterium tuberculosis extracytoplasmic-function sigma factor SigL regulates polyketide synthases and secreted or membrane proteins and is required for virulence. Journal of Bacteriology, 187, 7062-7071.

HALLEGRAEFF, G. M. 2010. Ocean climate change, phytoplankton community responses, and harmful algal blooms: a formidable predictive challenge. Journal of Phycology, 46, 220-235.

HAMILTON & JAMES 1983. Oil and the macroeconomy since World-War-II. Journal of Political Economy, 91, 228-248.

HAMILTON & JAMES 1996. This is what happened to the oil price macroeconomy relationship. Journal of Monetary Economics, 38, 215-220.

HAMILTON & JAMES 2003. What is an Oil Shock? 113.

HARRIS, L. M., WELKER, N. E. & PAPOUTSAKIS, E. T. 2002. Northern, morphological, and fermentation analysis of spo0A inactivation and overexpression in *Clostridium acetobutylicum* ATCC 824. Journal of Bacteriology, 184, 3586-3597.

HARTMAN, A. H., LIU, H. L. & MELVILLE, S. B. 2011. Construction and Characterization of a Lactose-Inducible Promoter System for Controlled Gene Expression in *Clostridium perfringens*. Applied and Environmental Microbiology, 77, 471-478.

HEAP, J. T., EHSAAN, M., COOKSLEY, C. M., NG, Y. K., CARTMAN, S. T., WINZER, K. & MINTON, N. P. 2012. Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. Nucleic Acids Research, 40, 10.

HEAP, J. T., KUEHNE, S. A., EHSAAN, M., CARTMAN, S. T., COOKSLEY, C. M., SCOTT, J. C. & MINTON, N. P. 2010. The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. Journal of Microbiological Methods, 80, 49-55.

HEAP, J. T., PENNINGTON, O. J., CARTMAN, S. T., CARTER, G. P. & MINTON, N. P. 2007. The ClosTron: A universal gene knock-out system for the genus *Clostridium*. Journal of Microbiological Methods, 70, 452-464.

HEAP, J. T., PENNINGTON, O. J., CARTMAN, S. T. & MINTON, N. P. 2009. A modular system for *Clostridium* shuttle plasmids. Journal of Microbiological Methods, 78, 79-85.

HEAP, J. T., THEYS, J., EHSAAN, M., KUBIAK, A. M., DUBOIS, L., PAESMANS, K., VAN MELLAERT, L., KNOX, R., KUEHNE, S. A., LAMBIN, P. & MINTON, N. P. 2014. Spores of *Clostridium* engineered for clinical efficacy and safety cause regression and cure of tumors in vivo. Oncotarget, 5, 1761-1769.

HERRMANN, G., JAYAMANI, E., MAI, G. & BUCKEL, W. 2008. Energy conservation via electron-transferring flavoprotein in anaerobic bacteria. Journal of Bacteriology, 190, 784-791.

HIRSCH, A. & GRINSTED, E. 1954. Methods for the growth and enumeration of anaerobic spore-formers from cheese, with observations on the effect of nisin. Journal of Dairy Research, 21, 101-110.

HOCH, J. A. 2002. Control of signal transduction in the sporulation phosphorelay.

Signals, Switches, Regulons and Cascades: Control of Bacterial Gene Expression, 61, 199-212.

HOFMANN, K. & STOFFEL, W. 1993. TMBASE— A database of membrane spanning protein segments. Biological Chemistry, 347, 166.

HONICKE, D., LUTKE-EVERSLOH, T., LIU, Z. Y., LEHMANN, D., LIEBL, W. & EHRENREICH, A. 2014. Chemostat cultivation and transcriptional analyses of *Clostridium acetobutylicum* mutants with defects in the acid and acetone biosynthetic pathways. Applied Microbiology and Biotechnology, 98, 9777-9794.

HU, K., ZHAO, M., ZHANG, T. L., ZHA, M. W., ZHONG, C., JIANG, Y. & DING, J. P. 2013. Structures of trans-2-enoyl-00A reductases from Clostridium acelobutylicum and Treponema denticola: insights into the substrate specificity and the catalytic mechanism. Biochemical Journal, 449, 79-89.

HUANG, H. Y., WANG, S. N., MOLL, J. & THAUER, R. K. 2012. Electron Bifurcation Involved in the Energy Metabolism of the Acetogenic Bacterium Moorella thermoacetica Growing on Glucose or H-2 plus CO2. Journal of Bacteriology, 194, 3689-3699.

HUANG, L., GIBBINS, L. N. & FORSBERG, C. W. 1985a. Transmembrane pH gradient and membrane-potential in *Clostridium acetobutylicum* during growth under acetogenic and solventogenic conditions. Applied and Environmental Microbiology, 50, 1043-1047.

HUANG, L., GIBBINS, L. N. & FORSBERG, C. W. 1985b. Transmembrane pH gradient and membrane-potential in *Clostridium acetobutylicum* during growth under acetogenic and solventogenic conditions. Applied and Environmental Microbiology, 50, 1043-1047.

HWANG, S., GOU, Z. K. & KUZNETSOV, I. B. 2007. DP-Bind: a Web server for sequence-based prediction of DNA-binding residues in DNA-binding proteins. Bioinformatics, 23, 634-636.

IGBP, I., SCOR 2013. Ocean Acidification Summary for Policymakers – Third Symposium on the Ocean in a High-CO2 World. International Geosphere-Biosphere Programme. Stockholm, Sweden.

INGLE, P., GROOTHUIS, D., ROWE, P., HUANG, H., COCKAYNE, A., KUEHNE, S.

A., JIANG, W. H., GU, Y., HUMPHREYS, C. M. & MINTON, N. P. 2019. Generation of a fully erythromycin-sensitive strain of *Clostridioides difficile* using a novel CRISPR-Cas9 genome editing system. Scientific Reports, 9.

IPCC 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In: PACHAURI, R. K. & MEYER, L. A. (eds.). Geneva, Switzerland: IPCC.

JABBARI, S., STEINER, E., HEAP, J. T., WINZER, K., MINTON, N. P. & KING, J. R. 2013. The putative influence of the agr operon upon survival mechanisms used by *Clostridium acetobutylicum*. Mathematical Biosciences, 243, 223-239.

JANG, Y. S., LEE, J. Y., LEE, J., PARK, J. H., IM, J. A., EOM, M. H., LEE, S. H., SONG, H., CHO, J. H., SEUNG, D. Y. & LEE, S. Y. 2012. Enhanced Butanol Production Obtained by Reinforcing the Direct Butanol-Forming Route in *Clostridium acetobutylicum*. Mbio, 3.

JASICKA-MISIAK, I., WIECZOREK, P. P. & KAFARSKI, P. 2005. Crotonic acid as a bioactive factor in carrot seeds (Daucus carota L.). Phytochemistry, 66, 1485-1491.

JIA, K. Z., ZHANG, Y. P. & LI, Y. 2012. Identification and Characterization of Two Functionally Unknown Genes Involved in Butanol Tolerance of *Clostridium acetobutylicum*. Plos One, 7.

JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science, 337, 816-821.

JINEK, M., JIANG, F. G., TAYLOR, D. W., STERNBERG, S. H., KAYA, E., MA, E. B., ANDERS, C., HAUER, M., ZHOU, K. H., LIN, S., KAPLAN, M., IAVARONE, A. T., CHARPENTIER, E., NOGALES, E. & DOUDNA, J. A. 2014. Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. Science, 343, 1215-+.

JOHNSON, E., SARCHAMI, T., KIESSLICH, S., MUNCH, G. & REHMANN, L. 2016. Consolidating biofuel platforms through the fermentative bioconversion of crude glycerol to butanol. World Journal of Microbiology & Biotechnology, 32, 14.

JONES, A. J., FAST, A. G., CLUPPER, M. & PAPOUTSAKIS, E. T. 2018. Small and

Low but Potent: the Complex Regulatory Role of the Small RNA SolB in Solventogenesis in *Clostridium acetobutylicum*. Applied and Environmental Microbiology, 84, 19.

JONES, D. T. & WOODS, D. R. 1986. Acetone-butanol fermentation revisited. Microbiological Reviews, 50, 484-524.

JONES, S. W., PAREDES, C. J., TRACY, B., CHENG, N., SILLERS, R., SENGER, R. S. & PAPOUTSAKIS, E. T. 2008. The transcriptional program underlying the physiology of clostridial sporulation. Genome Biology, 9.

JONES, T. C. 2012. America, Oil, and War in the Middle East. Journal of American History, 99, 208-218.

KARBERG, M., GUO, H. T., ZHONG, J., COON, R., PERUTKA, J. & LAMBOWITZ, A. M. 2001. Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. Nature Biotechnology, 19, 1162-1167.

KASTER, A. K., MOLL, J., PAREY, K. & THAUER, R. K. 2011. Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. Proceedings of the National Academy of Sciences of the United States of America, 108, 2981-2986.

KELL, D. B., PECK, M. W., RODGER, G. & MORRIS, J. G. 1981. On the permeability to weak acids and bases of the cytoplasmic membrane of *Clostridium pasteurianum*. Biochemical and Biophysical Research Communications, 99, 81-88.

KELLY, M. L., NG, Y. K., GARTMAN, S. T., COLLERY, M. M., COCKAYNE, A. & MINTON, N. P. 2016. Improving the reproducibility of the NAP1/B1/027 epidemic strain R20291 in the hamster model of infection. Anaerobe, 39, 51-53.

KIM, B. H., BELLOWS, P., DATTA, R. & ZEIKUS, J. G. 1984. Control of carbon and electron flow in *Clostridium acetobutylicum* fermentations - utilization of carbon-monoxide to inhibit hydrogen-production and to enhance butanol yields. Applied and Environmental Microbiology, 48, 764-770.

KIM, S., JANG, Y. S., HA, S. C., AHN, J. W., KIM, E. J., LIM, J. H., CHO, C., RYU, Y. S., LEE, S. K., LEE, S. Y. & KIM, K. J. 2015. Redox-switch regulatory mechanism of thiolase from *Clostridium acetobutylicum*. Nature Communications, 6.

KINZER, S. 2003. All The Shah's Men: An American Coup and the Roots of Middle

East Terror, John Wiley & Sons.

KNOTT, G. J. & DOUDNA, J. A. 2018. CRISPR-Cas guides the future of genetic engineering. Science, 361, 866-869.

KOTTE, A. K., SEVERN, O., BEAN, Z., SCHWARZ, K., MINTON, N. P. & WINZER, K. 2020. RRNPP-type quorum sensing affects solvent formation and sporulation in *Clostridium acetobutylicum*. Microbiology-Sgm, 166, 579-592.

KOVACS, K., WILLSON, B. J., SCHWARZ, K., HEAP, J. T., JACKSON, A., BOLAM, D. N., WINZER, K. & MINTON, N. P. 2013. Secretion and assembly of functional mini-cellulosomes from synthetic chromosomal operons in *Clostridium acetobutylicum* ATCC 824. Biotechnology for Biofuels, 6, 14.

KROGH, A., LARSSON, B., VON HEIJNE, G. & SONNHAMMER, E. L. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. Journal of Molecular Biology, 305, 567-580.

KRUTSAKORN, B., HONDA, K., YE, X. T., IMAGAWA, T., BEI, X. Y., OKANO, K. & OHTAKE, H. 2013. In vitro production of n-butanol from glucose. Metabolic Engineering, 20, 84-91.

KRYLOV, D. & VINSEN, C., R. 2001. Leucine Zipper, Chichester, John Wiley & Sons Ltd.

KUEHNE, S. A., CARTMAN, S. T., HEAP, J. T., KELLY, M. L., COCKAYNE, A. & MINTON, N. P. 2010. The role of toxin A and toxin B in Clostridium difficile infection. Nature, 467, 711-U97.

KUIT, W., MINTON, N. P., LOPEZ-CONTRERAS, A. M. & EGGINK, G. 2012. Disruption of the acetate kinase (*ack*) gene of *Clostridium acetobutylicum* results in delayed acetate production. Applied Microbiology and Biotechnology, 94, 729-741.

LEE, K. & NI, S. 2002. On the dynamic effects of oil price shocks: a study using industry level data. Journal of Monetary Economics, 49, 823-852.

LEHMANN, D., HONICKE, D., EHRENREICH, A., SCHMIDT, M., WEUSTER-BOTZ, D., BAHL, H. & LUTKE-EVERSLOH, T. 2012. Modifying the product pattern of *Clostridium acetobutylicum*. Applied Microbiology and Biotechnology, 94, 743-754.

LEHMANN, D. & LUTKE-EVERSLOH, T. 2011. Switching *Clostridium acetobutylicum* to an ethanol producer by disruption of the butyrate/butanol fermentative pathway.

Metabolic Engineering, 13, 464-473.
LEHMANN, D., RADOMSKI, N. & LUTKE-EVERSLOH, T. 2012b. New insights into the butyric acid metabolism of Clostridium acetobutylicum. Applied Microbiology and Biotechnology, 96, 1325-1339.
LI, F., HINDERBERGER, J., SEEDORF, H., ZHANG, J., BUCKEL, W. & THAUER, R. K. 2008. Coupled ferredoxin and crotonyl coenzyme a (CoA) reduction with NADH catalyzed by the butyryl-CoA dehydrogenase/Etf complex from *Clostridium kluyveri*. Journal of Bacteriology, 190, 843-850.
LI, J., CHEN, X. R., QI, B. K., LUO, J. Q., ZHANG, Y. M., SU, Y. & WAN, Y. H. 2014. Efficient production of acetone-butanol-ethanol (ABE) from cassava by a fermentation-pervaporation coupled process. Bioresource Technology, 169, 251-257.
LIJINSKY, W. & ANDREWS, A. W. 1980. MUTAGENICITY AND CARCINOGENICITY OF NITROSAMIDES. Environmental Mutagenesis, 2, 256-256.
LIM, J. H., SEO, S. W., KIM, S. Y. & JUNG, G. Y. 2013. Refactoring redox cofactor regeneration for high-yield biocatalysis of glucose to butyric acid in *Escherichia coli*. Bioresource Technology, 135, 568-573.
LIU, L. F. & WANG, J. C. 1987. SUPERCOILING OF THE DNA-TEMPLATE DURING TRANSCRIPTION. Proceedings of the National Academy of Sciences of the United States of America, 84, 7024-7027.
LIU, Y. J., ZHANG, J., CUI, G. Z. & CUI, Q. 2015a. Current progress of targetron technology: Development, improvement and application in metabolic engineering. Biotechnology Journal, 10, 855-865.
LIU, Z., QIAO, K., TIAN, L., ZHANG, Q., LIU, Z. Y. & LI, F. L. 2015b. Spontaneous large-scale autolysis in *Clostridium acetobutylicum* contributes to generation of more spores. Frontiers in Microbiology, 6.
LLAMAS, M. A., VAN DER SAR, A., CHU, B. C. H., SPARRIUS, M., VOGEL, H. J. & BITTER, W. 2009. A Novel Extracytoplasmic Function (ECF) Sigma Factor Regulates Virulence in *Pseudomonas aeruginosa*. PLOS Pathogens, 5.
LOPES, M. L., PAULILLO, S. C. D., GODOY, A., CHERUBIN, R. A., LORENZI, M. S., GIOMETTI, F. H. C., BERNARDINO, C. D., NETO, H. B. A. & DE AMORIM, H. V. 2016. Ethanol production in Brazil: a bridge between science and industry. Brazilian

Journal of Microbiology, 47, 64-76.
LUND, H. 2007. Renewable energy strategies for sustainable development. Energy, 32, 912-919.
MACHADO, H. B., DEKISHIMA, Y., LUO, H., LAN, E. I. & LIAO, J. C. 2012. A selection platform for carbon chain elongation using the CoA-dependent pathway to produce linear higher alcohols. Metabolic Engineering, 14, 504-511.
MACLELLAN, S. R., HELMANN, J. D. & ANTELMANN, H. 2009. The YvrI Alternative sigma Factor Is Essential for Acid Stress Induction of Oxalate Decarboxylase in *Bacillus subtilis*. Journal of Bacteriology, 191, 931-939.
MADDOX, I. S., STEINER, E., HIRSCH, S., WESSNER, S., GUTIERREZ, N. A., GAPES, J. R. & SCHUSTER, K. C. 2000. The cause of "acid crash" and "acidogenic fermentations" duping the batch acetone-butanol-ethanol (ABE) fermentation process. Journal of Molecular Microbiology and Biotechnology, 2, 95-100.
MANI, N. & DUPUY, B. 2001. Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. Proceedings of the National Academy of Sciences of the United States of America, 98, 5844-5849.
MANN, M. S. & LUTKE-EVERSLOH, T. 2013. Thiolase engineering for enhanced butanol production in *Clostridium acetobutylicum*. Biotechnology and Bioengineering, 110, 887-897.
MARTIN-VERSTRAETE, I., PELTIER, J. & DUPUY, B. 2016. The Regulatory Networks That Control *Clostridium difficile* Toxin Synthesis. Toxins, 8.
MERMELSTEIN, L. D. & PAPOUTSAKIS, E. T. 1993. Invivo methylation in *Escherichia coli* by the *Bacillus subtilis* phage-Φ3TI methyltransferase to protect plasmids from restriction upon transformation of *Clostridium acetobutylicum* ATCC-824. Applied and Environmental Microbiology, 59, 1077-1081.
MEYER, C. L., ROOS, J. W. & PAPOUTSAKIS, E. T. 1986. Carbon-monoxide gassing leads to alcohol production and butyrate uptake without acetone formation in continuous cultures of *Clostridium-acetobutylicum*. Applied Microbiology and Biotechnology, 24, 159-167.
MILLAT, T., JANSSEN, H., BAHL, H., FISCHER, R. J. & WOLKENHAUER, O. 2013a. Integrative modelling of pH-dependent enzyme activity and transcriptomic

regulation of the acetone-butanol-ethanol fermentation of *Clostridium acetobutylicum* in continuous culture. Microbial Biotechnology, 6, 526-539.
MILLAT, T., JANSSEN, H., THORN, G. J., KING, J. R., BAHL, H., FISCHER, R. J. & WOLKENHAUER, O. 2013b. A shift in the dominant phenotype governs the pH-induced metabolic switch of *Clostridium acetobutylicum* phosphate-limited continuous cultures. Applied Microbiology and Biotechnology, 97, 6451-6466.
MILLAT, T., JANSSEN, H., THORN, G. J., KING, J. R., BAHL, H., FISCHER, R. J. & WOLKENHAUER, O. 2013c. A shift in the dominant phenotype governs the pH-induced metabolic switch of *Clostridium acetobutylicum* phosphate-limited continuous cultures. Applied Microbiology and Biotechnology, 97, 6451-6466.
MILLAT, T. & WINZER, K. 2017. Mathematical modelling of clostridial acetone-butanol-ethanol fermentation. Applied Microbiology and Biotechnology, 101, 2251-2271.
MINTON, N. P., EHSAAN, M., HUMPHREYS, C. M., LITTLE, G. T., BAKER, J., HENSTRA, A. M., LIEW, F., KELLY, M. L., SHENG, L., SCHWARZ, K. & ZHANG, Y. 2016. A roadmap for gene system development in *Clostridium*. Anaerobe, 41, 104-112.
MITCHELL, W. J. 2016. Sugar uptake by the solventogenic clostridia. World Journal of Microbiology & Biotechnology, 32, 10.
MOFFATT, B. A. & STUDIER, F. W. 1987.T7 lysozyme inhibits transcription by T7 RNA-polymerase. Cell, 49, 221-227.
MOLLE, V., NAKAURA, Y., SHIVERS, R. P., YAMAGUCHI, H., LOSICK, R., FUJITA, Y. & SONENSHEIN, A. L. 2003. Additional targets of the *Bacillus subtilis* global regulator CodY identified by chromatin immunoprecipitation and genome-wide transcript analysis. Journal of Bacteriology, 185, 1911-1922.
MOORE, S. J., LAI, H. E., KELWICK, R. J. R., CHEE, S. M., BELL, D. J., POLIZZI, K. M. & FREEMONT, P. S. 2020. EcoFlex: A Multifunctional MoClo Kit for E. coli Synthetic Biology (vol 5, pg 1059, 2016). Acs Synthetic Biology, 9, 1225-1225.
MORDAKA, P. M. & HEAP, J. T. 2018. Stringency of Synthetic Promoter Sequences in *Clostridium* Revealed and Circumvented by Tuning Promoter Library Mutation Rates. Acs Synthetic Biology, 7, 672-+.

MORRISON, J. M., ANDERSON, K. L., BEENKEN, K. E., SMELTZER, M. S. & DUNMAN, P. M. 2012. The staphylococcal accessory regulator, SarA, is an RNA-binding protein that modulates the mRNA turnover properties of late-exponential and stationary phase *Staphylococcus aureus* cells. Frontiers in Cellular and Infection Microbiology, 2.

NASA 1991. Kuwait oil fire. In: 669828MAIN_KUWAIT-TIGHT.JPG (ed.) NASA's Earth Observator. http://www.nasa.gov/images/content/669828main_kuwait-tight.jpg: NASA.

NGUYEN, N. P. T., RAYNAUD, C., MEYNIAL-SALLES, I. & SOUCAILLE, P. 2018. Reviving the Weizmann process for commercial n-butanol production. Nature Communications, 9.

NGUYEN, P. T., LINDER, S., FLITSCH, S. K., SCHIEL-BENGELSDORF, B., DURRE, P. & SOUCAILLE, P. 2016. Cap0037, a Novel Global Regulator of *Clostridium acetobutylicum* Metabolism. Mbio, 7.

NIELSEN, D. R. & PRATHER, K. J. 2009. In Situ Product Recovery of n-Butanol Using Polymeric Resins. Biotechnology and Bioengineering, 102, 811-821.

NITSCHKE, W. & RUSSELL, M. J. 2012. Redox bifurcations: Mechanisms and importance to life now, and at its origin. Bioessays, 34, 106-109.

NOLLING, J., BRETON, G., OMELCHENKO, M. V., MAKAROVA, K. S., ZENG, Q. D., GIBSON, R., LEE, H. M., DUBOIS, J., QIU, D. Y., HITTI, J., WOLF, Y. I., TATUSOV, R. L., SABATHE, F., DOUCETTE-STAMM, L., SOUCAILLE, P., DALY, M. J., BENNETT, G. N., KOONIN, E. V., SMITH, D. R. & FINISHING, G. T. C. S. C. P. 2001. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. Journal of Bacteriology, 183, 4823-4838.

NOVAGEN 1999. pET System Manual. 8th Edition ed.

NOVICK, R. P. 2003. Autoinduction and signal transduction in the regulation of staphylococcal virulence. Molecular Microbiology, 48, 1429-1449.

O'BRIEN, R. W. & MORRIS, J. G. 1971. Oxygen and the Growth and Metabolism of *Clostridium acetobutylicum*. Microbiology, 68, 307-318.

PADAN, E. & SCHULDINER, S. 1986. Intracellular pH regulation in bacterial cells. Methods in Enzymology, 125, 337-352.

PADDON, C. J., WESTFALL, P. J., PITERA, D. J., BENJAMIN, K., FISHER, K., MCPHEE, D., LEAVELL, M. D., TAI, A., MAIN, A., ENG, D., POLICHUK, D. R., TEOH, K. H., REED, D. W., TREYNOR, T., LENIHAN, J., FLECK, M., BAJAD, S., DANG, G., DENGROVE, D., DIOLA, D., DORIN, G., ELLENS, K. W., FICKES, S., GALAZZO, J., GAUCHER, S. P., GEISTLINGER, T., HENRY, R., HEPP, M., HORNING, T., IQBAL, T., JIANG, H., KIZER, L., LIEU, B., MELIS, D., MOSS, N., REGENTIN, R., SECREST, S., TSURUTA, H., VAZQUEZ, R., WESTBLADE, L. F., XU, L., YU, M., ZHANG, Y., ZHAO, L., LIEVENSE, J., COVELLO, P. S., KEASLING, J. D., REILING, K. K., RENNINGER, N. S. & NEWMAN, J. D. 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature, 496, 528-+.

PEGUIN, S., GOMA, G., DELORME, P. & SOUCAILLE, P. 1994. METABOLIC FLEXIBILITY OF CLOSTRIDIUM-ACETOBUTYLICUM IN RESPONSE TO METHYL VIOLOGEN ADDITION. Applied Microbiology and Biotechnology, 42, 611-616.

PEREGO, M. 1998. Kinase-phosphatase competition regulates Bacillus subtilis development. Trends in Microbiology, 6, 366-370.

PERUTKA, J., WANG, W. J., GOERLITZ, D. & LAMBOWITZ, A. M. 2004. Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. Journal of Molecular Biology, 336, 421-439.

PIGGOT, P. J. & HILBERT, D. W. 2004. Sporulation of *Bacillus subtilis*. Current Opinion in Microbiology, 7, 579-586.

POEHLEIN, A., SOLANO, J. D. M., FLITSCH, S. K., KRABBEN, P., WINZER, K., REID, S. J., JONES, D. T., GREEN, E., MINTON, N. P., DANIEL, R. & DURRE, P. 2017. Microbial solvent formation revisited by comparative genome analysis. Biotechnology for Biofuels, 10.

POMERANTSEV, A. P., POMERANTSEVA, O. M., CAMP, A. S., MUKKAMALA, R., GOLDMAN, S. & LEPPLA, S. H. 2009. PapR peptide maturation: role of the NprB protease in *Bacillus cereus* 569 PlcR/PapR global gene regulation. Fems Immunology and Medical Microbiology, 55, 361-377.

PYNE, M. E., BRUDER, M. R., MOO-YOUNG, M., CHUNG, D. A. & CHOU, C. P. 2016. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*. Scientific Reports, 6.

QIAO, K. J., WASYLENKO, T. M., ZHOU, K., XU, P. & STEPHANOPOULOS, G. 2017. Lipid production in *Yarrowia lipolytica* is maximized by engineering cytosolic redox metabolism. Nature Biotechnology, 35, 173-177.

QURESHI, N. & BLASCHEK, H. P. 2001. Recovery of butanol from fermentation broth by gas stripping. Renewable Energy, 22, 557-564.

RAGSDALE, S. W. 2003. Pyruvate ferredoxin oxidoreductase and its radical intermediate. Chemical Reviews, 103, 2333-2346.

RALSTON, M. T. & PAPOUTSAKIS, E. T. 2018. RNAseq-based transcriptome assembly of *Clostridium acetobutylicum* for functional genome annotation and discovery. Aiche Journal, 64, 4271-4280.

RAN, F. A., HSU, P. D., LIN, C. Y., GOOTENBERG, J. S., KONERMANN, S., TREVINO, A. E., SCOTT, D. A., INOUE, A., MATOBA, S., ZHANG, Y. & ZHANG, F. 2013. Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Cell, 154, 1380-1389.

RAO, G. & MUTHARASAN, R. 1986. Alcohol production by *Clostridium acetobutylicum* induced by methyl viologen. Biotechnology Letters, 8, 893-896.

RAVAGNANI, A., JENNERT, K. C. B., STEINER, E., GRUNBERG, R., JEFFERIES, J. R., WILKINSON, S. R., YOUNG, D. I., TIDSWELL, E. C., BROWN, D. P., YOUNGMAN, P., MORRIS, J. G. & YOUNG, M. 2000. Spo0A directly controls the switch from acid to solvent production in solvent-forming clostridia. Molecular Microbiology, 37, 1172-1185.

REN, C., GU, Y., HU, S. Y., WU, Y., WANG, P., YANG, Y. L., YANG, C., YANG, S. & JIANG, W. H. 2010. Identification and inactivation of pleiotropic regulator CcpA to eliminate glucose repression of xylose utilization in *Clostridium acetobutylicum*. Metabolic Engineering, 12, 446-454.

REN, C., GU, Y., WU, Y., ZHANG, W. W., YANG, C., YANG, S. & JIANG, W. H. 2012. Pleiotropic functions of catabolite control protein CcpA in Butanol-producing *Clostridium acetobutylicum*. BMC Genomics, 13, 20.

RITCHIE, H. & ROSER, M. 2019. CO_2 and Greenhouse Gas Emissions. Retrieved from: 'https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions' [Online Resource]: Published online at OurWorldInData.org.

ROCHA-ESTRADA, J., ACEVES-DIEZ, A. E., GUARNEROS, G. & DE LA TORRE, M. 2010. The RNPP family of quorum-sensing proteins in Gram-positive bacteria. Applied Microbiology and Biotechnology, 87, 913-923.

ROOS, J. W., MCLAUGHLIN, J. K. & PAPOUTSAKIS, E. T. 1985. The effect of pH on nitrogen supply, cell-lysis, and solvent production in fermentations of *Clostridium acetobutylicum*. Biotechnology and Bioengineering, 27, 681-694.

ROY, A., KUCUKURAL, A. & ZHANG, Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols, 5, 725-738.

SALOM, D., HILL, B. R., LEAR, J. D. & DEGRADO, W. F. 2000. pH-dependent tetramerization and amantadine binding of the transmembrane helix of M2 from the influenza A virus. Biochemistry, 39, 14160-14170.

SAN, K. Y., BENNETT, G. N., BERRIOS-RIVERA, S. J., VADALI, R. V., YANG, Y. T., HORTON, E., RUDOLPH, F. B., SARIYAR, B. & BLACKWOOD, K. 2002. Metabolic engineering through cofactor manipulation and its effects on metabolic flux redistribution in *Escherichia coli*. Metabolic Engineering, 4, 182-192.

SCHNEIDMAN-DUHOVNY, D., INBAR, Y., NUSSINOV, R. & WOLFSON, H. J. 2005. Geometry-based flexible and symmetric protein docking. Proteins-Structure Function and Bioinformatics, 60, 224-231.

SCHUCHMANN, K. & MULLER, V. 2012. A Bacterial Electron-bifurcating Hydrogenase. Journal of Biological Chemistry, 287, 31165-31171.

SCHUCHMANN, K. & MULLER, V. 2014. Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. Nature Reviews Microbiology, 12, 809-821.

SENGER, R. S. & PAPOUTSAKIS, E. T. 2008. Genome-Scale Model for Clostridium acetobutylicum: Part I. Metabolic Network Resolution and Analysis. Biotechnology and Bioengineering, 101, 1036-1052.

SERIZAWA, M., KODAMA, K., YAMAMOTO, H., KOBAYASHI, K., OGASAWARA, N. & SEKIGUCHI, J. 2005. Functional analysis of the YvrGHb two-component system of Bacillus subtilis: Identification of the regulated genes by DNA microarray and northern blot analyses. Bioscience Biotechnology and Biochemistry, 69, 2155-2169.

SERVINSKY, M. D., KIEL, J. T., DUPUY, N. F. & SUND, C. J. 2010. Transcriptional

analysis of differential carbohydrate utilization by *Clostridium acetobutylicum*. Microbiology-Sgm, 156, 3478-3491.

SHEN, B., ZHANG, W. S., ZHANG, J., ZHOU, J. K., WANG, J. Y., CHEN, L., WANG, L., HODGKINS, A., IYER, V., HUANG, X. X. & SKARNES, W. C. 2014. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. Nature Methods, 11, 399-+.

SHEN, C. R., LAN, E. I., DEKISHIMA, Y., BAEZ, A., CHO, K. M. & LIAO, J. C. 2011. Driving Forces Enable High-Titer Anaerobic 1-Butanol Synthesis in Escherichia coli. Applied and Environmental Microbiology, 77, 2905-2915.

SHIVERS, R. P. & SONENSHEIN, A. L. 2004. Activation of the *Bacillus subtilis* global regulator CodY by direct interaction with branched-chain amino acids. Molecular Microbiology, 53, 599-611.

SLAMTI, L. & LERECLUS, D. 2019. The oligopeptide ABC-importers are essential communication channels in Gram-positive bacteria. Research in Microbiology.

SOMA, Y., TSURUNO, K., WADA, M., YOKOTA, A. & HANAI, T. 2014. Metabolic flux redirection from a central metabolic pathway toward a synthetic pathway using a metabolic toggle switch. Metabolic Engineering, 23, 175-184.

SONNHAMMER, E. L. L., EDDY, S. R., BIRNEY, E., BATEMAN, A. & DURBIN, R. 1998. Pfam: multiple sequence alignments and HMM-profiles of protein domains. Nucleic Acids Research, 26, 320-322.

STEINER, E., DAGO, A. E., YOUNG, D. I., HEAP, J. T., MINTON, N. P., HOCH, J. A. & YOUNG, M. 2011. Multiple orphan histidine kinases interact directly with Spo0A to control the initiation of endospore formation in *Clostridium acetobutylicum*. Molecular Microbiology, 80, 641-654.

STRAUCH, M., WEBB, V., SPIEGELMAN, G. & HOCH, J. A. 1990. The Spo0A protein of *Bacillus subtilis* is a repressor of the *abrB* gene. Proceedings of the National Academy of Sciences of the United States of America, 87, 1801-1805.

STUDIER, F. W. 1991. Use of bacteriophage-T7 lysozyme to improve an inducible T7 expression system. Journal of Molecular Biology, 219, 37-44.

STUDIER, F. W. & MOFFATT, B. A. 1986. Use of bacteriophage-T7 RNA-polymerase to direct selective high-level expression of cloned genes. Journal of Molecular

Biology, 189, 113-130.
THAMER, W., CIRPUS, I., HANS, M., PIERIK, A. J., SELMER, T., BILL, E., LINDER, D. & BUCKEL, W. 2003. A two 4Fe-4S -cluster-containing ferredoxin as an alternative electron donor for 2-hydroxyglutaryl-CoA dehydratase from *Acidaminococcus fermentans*. Archives of Microbiology, 179, 197-204.
TOMAS, C. A., WELKER, N. E. & PAPOUTSAKIS, E. T. 2003. Overexpression of groESL in Clostridium acetobutylicum results in increased solvent production and tolerance, prolonged metabolism, and changes in the cell's transcriptional program. Applied and Environmental Microbiology, 69, 4951-4965.
TRACY, B. P., JONES, S. W. & PAPOUTSAKIS, E. T. 2011. Inactivation of sigma(E) and sigma(G) in Clostridium acetobutylicum Illuminates Their Roles in Clostridial-Cell-Form Biogenesis, Granulose Synthesis, Solventogenesis, and Spore Morphogenesis. Journal of Bacteriology, 193, 1414-1426.
TUCCI, S. & MARTIN, W. 2007. A novel prokaryotic trans-2-enoyl-CoA reductase from the spirochete *Treponema denticola*. Febs Letters, 581, 1561-1566.
ULRICH, L. E., KOONIN, E. V. & ZHULIN, I. B. 2005. One-component systems dominate signal transduction in prokaryotes. Trends in Microbiology, 13, 52-56.
UYEDA, K. & RABINOWITZ, J. C. 1971. Pyruvate-ferredoxin oxidoreductase .3. Purification and properties of the enzyme. Journal of Biological Chemistry, 246, 3111-+.
VASCONCELOS, I., GIRBAL, L. & SOUCAILLE, P. 1994. Regulation of carbon and electron flow in *Clostridium acetobutylicum* grown in chemostat culture at neutral pH on mixtures of glucose and glycerol. Journal of Bacteriology, 176, 1443-1450.
WALKER, J. R., HERVAS, C., ROSS, J. D., BLINKOVA, A., WALBRIDGE, M. J., PUMAREGA, E. J., PARK, M. O. & NEELY, H. R. 2000. *Escherichia coli* DNA polymerase III tau- and gamma-subunit conserved residues required for activity in vivo and in vitro. Journal of Bacteriology, 182, 6106-6113.
WANG, S. N., HUANG, H. Y., MOLL, J. & THAUER, R. K. 2010. NADP(+) Reduction with Reduced Ferredoxin and NADP(+) Reduction with NADH Are Coupled via an Electron-Bifurcating Enzyme Complex in *Clostridium kluyveri*. Journal of Bacteriology, 192, 5115-5123.

WANG, Y., ZHANG, Z. T., SEO, S. O., LYNN, P., LU, T., JIN, Y. S. & BLASCHEK, H. P. 2016. Bacterial Genome Editing with CRISPR-Cas9: Deletion, Integration, Single Nucleotide Modification, and Desirable "Clean" Mutant Selection in *Clostridium beijerinckii* as an Example. Acs Synthetic Biology, 5, 721-732.

WARD, E., KIM, E. A., PANUSHKA, J., BOTELHO, T., MEYER, T., KEARNS, D. B., ORDAL, G. & BLAIR, D. F. 2019. Organization of the Flagellar Switch Complex of *Bacillus subtilis*. Journal of Bacteriology, 201.

WASELS, F., JEAN-MARIE, J., COLLAS, F., LOPEZ-CONTRERAS, A. M. & FERREIRA, N. L. 2017. A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. Journal of Microbiological Methods, 140, 5-11.

WATERHOUSE, A., BERTONI, M., BIENERT, S., STUDER, G., TAURIELLO, G., GUMIENNY, R., HEER, F. T., DE BEER, T. A. P., REMPFER, C., BORDOLI, L., LEPORE, R. & SCHWEDE, T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Research, 46, W296-W303.

WEGHOFF, M. C., BERTSCH, J. & MULLER, V. 2015. A novel mode of lactate metabolism in strictly anaerobic bacteria. Environmental Microbiology, 17, 670-677.

WEINRICK, B., DUNMAN, P. M., MCALEESE, F., MURPHY, E., PROJAN, S. J., FANG, Y. & NOVICK, R. P. 2004. Effect of mild acid on gene expression in Staphylococcus aureus. Journal of Bacteriology, 186, 8407-8423.

WEIZMANN, C. & ROSENFELD, B. 1937. The activation of the butanol-acetone fermentation of carbohydrates by *Clostridium acetobutylicum* (Weizmann). Biochemical Journal, 31, 619-639.

WHEELER, T. & VON BRAUN, J. 2013. Climate Change Impacts on Global Food Security. Science, 341, 508-513.

WHITE, S. Membrane Proteins of Known Structure [Online]. https://blanco.biomol.uci.edu/mpstruc/. [Accessed 2019].

WIEDENHEFT, B., STERNBERG, S. H. & DOUDNA, J. A. 2012. RNA-guided genetic silencing systems in bacteria and archaea. Nature, 482, 331-338.

WIESENBORN, D. P., RUDOLPH, F. B. & PAPOUTSAKIS, E. T. 1989. COENZYME-A TRANSFERASE FROM CLOSTRIDIUM-ACETOBUTYLICUM ATCC-824 AND ITS ROLE IN THE UPTAKE OF ACIDS. Applied and Environmental Microbiology, 55,

323-329.

WIETZKE, M. & BAHL, H. 2012. The redox-sensing protein Rex, a transcriptional regulator of solventogenesis in *Clostridium acetobutylicum*. Applied Microbiology and Biotechnology, 96, 749-761.

WILLSON, B. J., KOVACS, K., WILDING-STEELE, T., MARKUS, R., WINZER, K. & MINTON, N. P. 2016. Production of a functional cell wall-anchored minicellulosome by recombinant *Clostridium acetobutylicum* ATCC 824. Biotechnology for Biofuels, 9.

XU, M. M., ZHAO, J. B., YU, L., TANG, I. C., XUE, C. & YANG, S. T. 2015a. Engineering *Clostridium acetobutylicum* with a histidine kinase knockout for enhanced n-butanol tolerance and production. Applied Microbiology and Biotechnology, 99, 1011-1022.

XU, T., LI, Y. C., SHI, Z., HEMME, C. L., LI, Y., ZHU, Y. H., VAN NOSTRAND, J. D., HE, Z. L. & ZHOU, J. Z. 2015b. Efficient Genome Editing in *Clostridium cellulolyticum* via CRISPR-Cas9 Nickase. Applied and Environmental Microbiology, 81, 4423-4431.

XUE, X. W., DAVIS, M. C., STEEVES, T., BISHOP, A., BREEN, J., MACEACHERON, A., KESTHELY, C. A., HSU, F. S. & MACLELLAN, S. R. 2016. Characterization of a protein-protein interaction within the SigO-RsoA two-subunit sigma factor: the sigma(70) region 2.3-like segment of RsoA mediates interaction with SigO. Microbiology-Sgm, 162, 1857-+.

YANG, J. Y., YAN, R. X., ROY, A., XU, D., POISSON, J. & ZHANG, Y. 2015. The I-TASSER Suite: protein structure and function prediction. Nature Methods, 12, 7-8.

YANG, J. Y. & ZHANG, Y. 2015. I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Research, 43, W174-W181.

YARBROUGH, J. W. & W., S. T. 2007. Abiotic Sulfhydryl Reactivity: A Predictor of Aquatic Toxicity for Carbonyl-Containing α,β-Unsaturated Compounds. Chemical Research in Toxicology, 20, 558-562.

YETIV & STEPHEN 2008. The Absence of Grand Strategy: The United States in the Persian Gulf, 1972-2005, Johns Hopkins University Press.

YOO, M., BESTEL-CORRE, G., CROUX, C., RIVIERE, A., MEYNIAL-SALLES, I. & SOUCAILLE, P. 2015. A Quantitative System-Scale Characterization of the

Metabolism of Clostridium acetobutylicum. Mbio, 6, 12.

ZHANG, J., LIU, Y. J., CUI, G. Z. & CUI, Q. 2015a. A novel arabinose-inducible genetic operation system developed for *Clostridium cellulolyticum*. Biotechnology for Biofuels, 8, 13.

ZHANG, L., NIE, X. Q., RAVCHEEV, D. A., RODIONOV, D. A., SHENG, J., GU, Y., YANG, S., JIANG, W. H. & YANG, C. 2014. Redox-Responsive Repressor Rex Modulates Alcohol Production and Oxidative Stress Tolerance in *Clostridium acetobutylicum*. Journal of Bacteriology, 196, 3949-3963.

ZHANG, Y., GROSSE-HONEBRINK, A. & MINTON, N. P. 2015b. A Universal Mariner Transposon System for Forward Genetic Studies in the Genus Clostridium. PLOS One, 10, 21.

ZHOU, Z. M. & SWENSON, R. P. 1996. The cumulative electrostatic effect of aromatic stacking interactions and the negative electrostatic environment of the flavin mononucleotide binding site is a major determinant of the reduction potential for the flavodoxin from *Desulfovibrio vulgaris* Hildenborough. Biochemistry, 35, 15980-15988.

ZIMMERMAN, T. 2013. Untersuchungen zur Butanolbildung von Hyperthermus butylicus und Clostridium acetobutylicum PhD thesis, University of Ulm.

ZOUHIR, S., PERCHAT, S., NICAISE, M., PEREZ, J., GUIMARAES, B., LERECLUS, D. & NESSLER, S. 2013. Peptide-binding dependent conformational changes regulate the transcriptional activity of the quorum-sensor NprR. Nucleic Acids Research, 41, 7920-7933.