Engineering Bacteria for Biofuel Production

Heather-Rose Victoria Macklyne

Thesis submitted for the degree of

Doctor of Philosophy

University of Sussex

September 2016

Summary

This thesis addresses the need for environmentally and socially responsible sources of energy. Biofuels, made from organic matter, have recently become a viable alternative to petroleum-based fossil fuel. Sugar and starch make up the majority of feedstock used in biofuel production as it is easily digested. However, the use of these feedstocks is problematic as they consume resources with negative implications. By using a bacterium able to utilise five and six carbon sugars, such as the thermophile *Geobacillus thermoglucosidans*, organic lignocellulosic waste material can be used as a feedstock.

The aim of this project was to investigate and utilise key genetic regulators of fermentation in *G. thermoglucosidans* and to construct genetic engineering tools that enable strain development for second generation biofuel production. We have focused on the redox-sensing transcriptional regulator Rex, widespread in Gram-positive bacteria, which controls the major fermentation pathways in response to changes in cellular NAD+/NADH ratio. Following the identification of several members of the Rex regulon via bioinformatics analysis, ChIP-seq and qRT-PCR experiments were performed to locate genome-wide binding sites and controlled genes in *G. thermoglucosidans*. Initial electromobility shift assay experiments were performed to demonstrate the potential for use of Rex from *Clostridium thermocellum* as an orthogonal regulator. To further this research, novel in vivo synthetic regulatory switches were designed and tested with the aim of controlling gene expression in response to changes in cellular redox state. In addition, new tools for the efficient genetic engineering of *G. thermoglucosidans* were produced and optimised, including an *E. coli-G. thermoglucosidans* conjugation method for plasmid transfer and gene disruption.

Acknowledgments

Firstly I would like to thank Dr Mark Paget for the opportunity to work in his lab, and acknowledge the huge amount of time and energy he has dedicated to helping me during my PhD. To all members of the Paget research group, past and present, including Heena Jagatia, Laurence Humphrey, Aline Tabib-Salazar: thank you for your daily support and encouragement. I would also like to extend my thanks to the Morley research group, including Ella Lineham, Alice Copsey and Simon Cpoper for your friendship. I wish to thank my co-supervisor Dr Neil Crickmore for his help and guidance throughout this project. I will think back on my time at Sussex fondly and miss all the friends, fellow students and staff members who made it such an enjoyable place to work and study.

The Student Support team has been an incredibly important part of my time at Sussex and I want to acknowledge that their work has enabled me to succeed. I am particularly grateful to my dyslexia tutors Moira Wilson and Alison Firbank for believing in me, even when I didn't believe in myself. I could never have achieved this without you.

For the funding of this research and my PhD studentship I am thankful to the BBSRC. I am also thankful for initial sponsorship and support from TMO Renewables Ltd, particularly to Dr Alex Pudney, Gareth Cooper, Kirstin Eley. I would also like to acknowledge Joanne Neary, Kevin Charman and Zaneta Walanus for their help whilst visiting Rebio Technologies Ltd.

Ursula Gooding deserves my gratitude for making sure I didn't fall through the cracks when I left home at a young age; her support was vital and always offered graciously. I would also like to thank my oldest friends James Corbett, Jess Chitty and Christina Griffin who have cheered me on through the good times, and the bad.

Lastly I would like to thank my fiancé Oliver Hill-Andrews who I met at the beginning of these four years. His love and intellect are a continual inspiration to me and have spurred me to complete this project.

Table of Contents

1. Chapter I: Introduction

Summary

Biofuels and other bio-based products are renewable, viable alternatives to petroleum-based fuels and chemicals. The bacterium *Geobacillus thermoglucosidans* has been utilised as a converter of lignocellulosic biomass to bioethanol. However, to diversify products and optimise yields, innovation in genetic engineering tools is badly needed. The metabolism of *G. thermoglucosidans* must be optimised and redirected towards production of the required chemical. For this to be realised, not only are improved genetic tools required, but an in-depth knowledge of the regulators of gene expression is needed.

The sections that follow make the case for biofuels as an alternative to petroleum and highlight the use of synthetic biology in aiding this transition process. The *Geobacillus* genus, particularly *Geobacillus thermoglucosidans*, is then introduced as a biofuel producer, including the advantages of the organism that have led to its industrial application. An overview of how bacteria regulate respiration and fermentation in response to environmental conditions follows, including the mechanism by which the main regulators operate. A description of Rex, a redox sensing regulator that is the focus of this thesis, is then given, followed by an account of the aims of this project.

1.1. The need for renewable energy

Energy is humanity's most important resource, its consumption continually growing, yet the method of its production poses the greatest challenge of our time. At present fossil fuels remain the main resource used to meet energy demands. The impact of fossil fuel utilisation, however, is clearly environmentally damaging and unsustainable (Armaroli & Balzani, 2011; WEF, 2016). A 2°C global temperature increase limit has been set, a commitment intended to avoid the most serious effect of climate change – such as the melting of the Greenland ice sheet and sea level rises of 7 meters (Centre & Office, 2013). The international community has recognised the need for change and governments and industry have made

investments in an array of alternative technologies, an investment increase of 80% since 2000 (IEA, 2015). These are being developed and implemented, resulting in diversification of the energy market, including wind power, solar energy, geothermal heat, hydroelectricity and bioenergy. A global low-carbon energy target has been set of 21% of the total energy supply by 2025 (IEA, 2015). Each energy source has advantages and disadvantages associated with it, and each will be most efficient when applied to particular applications. The niche biofuels occupy is as an alternative to crude oil. Crude oil is a resource which is rapidly declining and increasing in price as it becomes more difficult to extract (despite this the market supply has steadily increased to drive economic growth) (Miller & Sorrell, 2014). Biofuels can not only be renewable and potentially carbon neutral, but they also represent a fresh commerce opportunity to replace the production and sale of petroleum-based fuel and chemicals, often monopolised and linked to the geographical location and natural resources of a country.

1.1.1. Biofuels as a solution, for better or worse?

Crude oil is the predominant resource used for production of both liquid transportation fuels and a vast array of chemical products. The development of biofuels, using microorganisms to convert renewable feedstocks, offers an alternative, potentially superior in terms of both environmental impact and renewability (Hill, Nelson, Tilman, Polasky, & Tiffany, 2006). When combusted, biofuels also emit polluting gases, however they can be considered carbon neutral as the feedstock is based upon renewable plant material. Plants absorb carbon dioxide whilst growing and so the quantity released back into the atmosphere is equal to that absorbed by the process of plant growth (Bogner et al., 2008). Although biofuel feedstock is renewable and carbon neutral, several issues concerning the source and exact type of feedstock exist, as this too can generate environmental and social problems. Lignocellulosic biomass is often discarded as a waste material by the agricultural and forestry industries, and makes a convincing candidate for use as a second generation biofuel feedstock. Lignocellulosic biomass, however, is not widely utilised by the current model microorganisms such as *E. coli*.

1.1.2. First generation biofuel

First generation biofuels, although utilising a renewable feedstock, have detrimental social and environmental impact factors associated with production. Feedstocks such as sugar and starches have been utilised for biofuel production due to the wide range of microorganisms able to process them and the minimal pre-treatment requirements. Negative implications of first generation biofuel production have become evident (Sims, Taylor, Jack, & Mabee, 2008). Feedstocks such as sugar cane or corn have become diverted from consumption as nutrition, for either cattle or humans, towards biofuel production. This has repercussions for availability and pricing of food, especially in the developing world. The resources needed to grow the feedstock — such as fertilisers, land and water — are at risk of over consumption, with potentially harmful environmental implications (Lopes, 2015). The move towards alternative, second generation biofuels aims to alleviate many of these concerns and their utilisation has recently become a focus of research (Simmons, Loque, & Blanch, 2008).

1.1.3. Lignocellulosic biomass and advanced second generation biofuel production

Second generation biofuels differ from first generation by the feedstock utilised in their production. The alternative feedstock attempts to minimise any negative aspects of first generation biofuels, but they tend to be organic waste materials that are harder to break down. Substantial pre-treatment is usually needed which includes multiple stages, requiring specialised equipment and energy use, impacting overall energy yield and efficiency (Kumar, Barrett, Delwiche, & Stroeve, 2009; Williams, Inman, Aden, & Heath, 2009). Lignocellulosic biomass, an abundant potential feedstock, has been identified as being an alternative to first generation feedstocks.

Lignocellulose biomass is composed of cellulose, hemicellulose (such as non-cellulosic polysaccharides including xylans) and lignin (Abdel-Rahman, Tashiro, & Sonomoto, 2010). The biomass can be found in several forms, such as crop residues, municipal waste, wood and paper waste (John, Nampoothiri, & Pandey, 2007). The largest proportion of lignocellulose biomass is composed of cellulose, made up of glycosidic bond D-glucose. Lignin surrounds cellulose fibres and is

composed of a complex matrix of aromatic alcohol bond linkages which make it extremely hard to degrade (Abdel-Rahman et al., 2010). However, hemicellulose — which is also sheathed around cellulose — is more easily hydrolysed. Hemicellulose is composed of pentose (xylose) and hexose (glucose) sugars which can be used as fermentation feedstock (Nieves, Panyon, & Wang, 2015). The proportion of cellulose, hemicellulose and lignin differs between biomass source and pre-treatment must be optimised accordingly.

The pre-treatment of lignocellulosic biomass involves initial mechanical breakdown. This is followed by acidic chemical treatment used to degrade hemicellulose polymers (Alvira, Tomás-Pejó, Ballesteros, & Negro, 2010). A costly enzymatic step then follows to act upon cellulose and lignin, breaking it down and making it accessible to microorganisms. Next, fermentation and the conversion of the sugar monomers and dimers to fermentation products occurs (Saha, 2003). Extraction of the biofuel and downstream processing follows. Each stage of pre-treatment is optimised to maximise yield, but the process is demanding of both time and energy, reducing the overall yield and therefore margin of profit. Industrially relevant microorganisms must be capable of the co-utilisation of both pentose (xylose) and hexose (glucose) sugars to be efficient lignocellulosic biomass to biofuel converters. Although a variety of organisms exist that are able to utilise lignocellulosic biomass such as fungi, they are not commercially viable for a number of reasons. Not only must the organism possess the ability to utilise the feedstock; it must also have the ability to produce and tolerate high concentrations of a biofuel such as ethanol or butanol. Furthermore, the organism must produce minimal amounts of unwanted products. A fast metabolism for quick product turnaround and robustness for changes in temperature and pressure are also essential attributes for a commercially viable organism.

1.2. The industrial biofuel producer Geobacillus thermoglucosidans

The *Geobacillus* genus has emerged as an important industrial strain, commercially adopted for biofuel production (Taylor et al., 2009). It is a gram positive, rod shaped bacterium characterised as facultative anaerobic and thermophilic, able to ferment a variety of feedstocks, including sugars derived from lignocellulose and

municipal waste materials to products such as ethanol. The range of genetic manipulation tools for use with this organism is underdeveloped, however. The application of traditional genome editing techniques alongside synthetic biology methods should quicken strain development, increasing yields and enabling diversification of product production from ethanol to butanol, and indeed to other high value commodities.

1.2.1. *Geobacillus* classification, environmental isolation and history

The *Geobacillus* genus is thermophilic and isolated from heated environments such as compost heaps, hot springs, oil and gas wells. *Geobacillus thermoglucosidans,* originally known as *Bacillus thermoglucosidasius* (Coorevits et al., 2012; Nazina, Tourova, & Poltaraus, 2001), is the organism of focus throughout this thesis. Therme, meaning heat, refers to the thermophilic nature of the bacterium, the optimal growth temperature being between 55-65°C. The glucosidans component of its name indicates its starch-hydrolyzing glucosidase activity, a key advantage of the species as a biofuel producer. *G. thermoglucosidans* was first isolated from Japanese soil and has been found to tolerate growth temperatures of between 42–69°C (Suzuki et al., 1983). It was phenotypically analysed and originally categorised as *Bacillus thermoglucosidasius* but, following the discovery of large numbers of other thermophilic strains, and subsequent 16S rRNA gene sequence analysis, it was transferred to a new genus, that of *Geobacillus* (Nazina, Tourova, Poltaraus, et al., 2001). Further phenotypic analysis of multiple *Geobacillus thermoglucosidasius* strains revealed that four were in fact of the species *G. toebii*. The remaining eight *G. thermoglucosidasius* strain descriptions were amended and the name revised to *G. thermoglucosidans* (Coorevits et al., 2012).

Prokaryotic organisms are divided into two domains, bacteria and archaea. The bacterial domain is further subdivided into phyla including the Firmicutes which are characterised as having a low C+G content (Bergey, 2009). Bacillales is an order within this phylum, and the genus *Geobacillus* is a further subcategory to which the species *G. thermoglucosidans* belongs, based on 16S rRNA gene sequence analysis. The genus *Geobacillus* is part of a vast assortment of predominantly soil dwelling organisms that contribute to the degradation of biomass and thereby play an important role in the carbon cycle, maintaining a soil's ability to support plant

life (Gougoulias, Clark, & Shaw, 2014). Each of these organisms, including *G. thermoglucosidans,* occupies a specialist niche. Many soil characteristics such as nutrient levels, temperature and oxygen concentration change dramatically throughout a soil profile, giving rise to unique micro-ecosystems. The optimal environment for *G. thermoglucosidans* may only arise infrequently; often the bacterium will have to cope with more stressful environments.

During times of intolerable environmental stress, organisms can lay dormant until circumstances improve. The genus *Geobacillus* achieves this by forming tough endospores capable of resisting harsh conditions (Zeigler, 2014). When the environment becomes more favourable, endospore reactivation occurs in stages. Firstly, activation arises (triggered commonly by heating), germination shortly follows with an increase in metabolic activity, finally leading to outgrowth and a fully functional replicating cell.

G. thermoglucosidans are facultative anaerobic bacteria, able to respire in the absence of oxygen. This is an important adaptation, as after the predominantly obligate aerobic fungi have broken down complex organic matter at the surface of soil, facultative anaerobes are integral for the continued decomposition within the depth of soil, where low oxygen concentrations arise (Juo, 2003). For short periods temperatures can reach upwards of 55°C due to the exothermic reactions and decomposition of plant matter (such as in compost heaps), a temperature range *G. thermoglucosidans* thrives at (Zeigler, 2014).

1.2.2. Genes and Genome

Many species of the *Geobacillus* genus have been sequenced and annotated, and are stored on the NCBI genome database. These genomes include six *G. thermoglucosidans* species strains (see Table 1.1). The median GC content of the *G. thermoglucosidans* genome is 43.8%, and of a median length of 3.87 Mb in total. The chromosome of *G. thermoglucosidans* is circular: in one strain the genome has been found to contain two additional plasmids (Brumm, Land, & Mead, 2015).

The genome of C56-YS93 strain is unique in that it includes two circular plasmids approximately 81 Kb and 20 Kb in size. This strain was also found to have a xylan degradation cluster, complete with enzymes and transporters, not found in other

strains (Brumm et al., 2015). The C56-YS93 strain also contains nitrate respiration clusters for reduction of nitrate to nitrite.

Strain	Size (Mb)	GC%	Genes	Proteins	Isolation source	Date
C56-YS93	3.99379	43.98	3992	3717	Hot Spring, USA	2011/06/10
DSM 2542	390714	43.70	3895	3667	Soil, Japan	2015/03/30
TNO-09.020	3.74024	43.90	3750	3521	Casein fouling sample from a dairy factory, geographical location unknown	2012/04/05
NBRC 107763	3.87116	43.70	3849	3644	Japan, specifics unknown	2014/04/12
W-2	3.89456	43.20	3919	3717	Oil reservoir, China	2016/06/03
GT23	3.69646	43.80	3666	3508	Casein pipeline, Netherlands	2016/05/25

Table 1.1 Genomes of G. thermoglucosidans species strains stored on NCBI database (http://www.ncbi.nlm.nih.gov/genome/genomes/2405).

The *G. thermoglucosidans* strain 11955 was isolated and sequenced by NCIMB, a microbiology and chemical company which collects industrially valuable microorganisms. A wild type strain was selected by the second generation biofuel producing company TMO Renewables Ltd/ReBio Technologies Ltd to undergo strain development. It was chosen as a suitable strain for bioethanol production as it can utilize C_5 and C_6 sugars.

1.2.3. Industrial use of *Geobacillus thermoglucosidans*

G. thermoglucosidans has been industrially adopted as a producer of bioethanol by the British company TMO Renewables Ltd, and more recently ReBio Technologies Ltd. It has the advantage over first generation bioethanol producers of being able to convert cellobiose, xylitol and other lignocelluloses-based feed stocks into

fermentable sugars (Cripps et al., 2009). Naturally, the organism would convert these into a range of products including acetate, lactate, formate and ethanol. Strain development, however, has enabled flux to be diverted into ethanol production. The ability of *G. thermoglucosidans* to utilise second generation feedstock, combined with thermophilic associated properties, led to this strain being considered as a strong candidate for industrial biofuel production. *G. thermoglucosidans* was named as a most promising candidate for biofuel production, owing to its comparatively high product tolerance and broad substrate utilisation range (Tang et al., 2009).

Although *G. thermoglucosidans* demonstrates many advantageous characteristics for bioethanol production, the application of metabolic engineering has further improved the strain. The main areas of research importance within industry have been to increase ethanol yield and tolerance (Jingyu Chen, Zhang, Zhang, & Yu, 2015). Superior ethanol yield has been achieved by repressing alternative fermentation pathways. The commercial strain TM242 has multiple-gene knockouts, designed to divert metabolism towards ethanol production (Cripps et al., 2009).

1.2.4. Advantages of *Geobacillus thermoglucosidans* as a biofuel producer

The thermophilic characteristics of *G. thermoglucosidans* lend themselves to the industrial environment. There are a number of reasons for this. Firstly, large scale mesophilic fermentations require cooling, however thermophiles limit the need for this energy intensive practice (Zeldes et al., 2015). Secondly, an increased rate of metabolism and substrate conversion leads to a quicker product output and better efficiency (Bhalla, Bansal, Kumar, Bischoff, & Sani, 2013). Finally, the risk from contamination is minimised, as few microorganisms can tolerate such high temperatures. In addition, as ethanol has a low boiling temperature, its removal is facilitated by higher temperatures (Cripps et al., 2009; Lynd, van Zyl, McBride, & Laser, 2005; Shaw et al., 2008).

TMO Renewables Ltd, a former company utilising *G. thermoglucosidans* as a biofuel producer, used genetic manipulation tools to modify the metabolism of *Geobacillus thermoglucosidans* to enhance ethanol production (Cripps et al., 2009). Naturally, *G. thermoglucosidans* can produce ethanol, lactate, formate and acetate during anaerobic

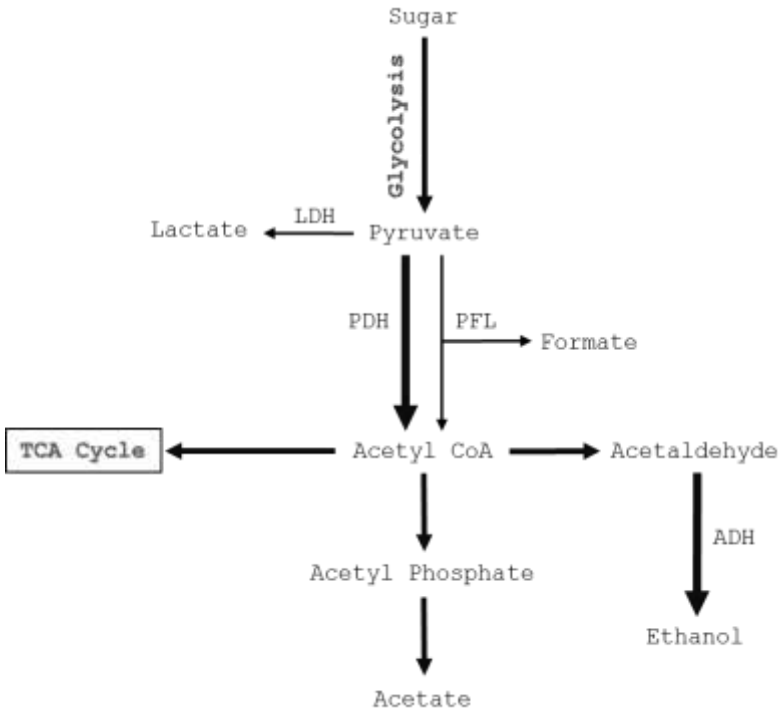


Figure 1.1 Overview of major relevent fermentation metabolic pathways in *G. thermoglucosidans*. Industrial strain TM242 is an *ldh* and *pfl* mutant with an enhanced *pdh* promoter to redirect the pathway to enhance ethanol yields. Abbreviations: LDH lactate dehydrogenase, PFL pyruvate formatelyase; PDH pyruvate dehydrogenase; ADH alcohol dehydrogenase; (Cripps et al., 2009).

respiration and fermentation. It does this to regenerate NAD^+ from NADH. By repressing alternative pathways not involved in the direct production of ethanol, yields were enhanced. The commercial strain TM242 has three main mutations designed to increase flux to ethanol production and decrease flux of alternative fermentation products. TM242 has gene deletions in lactate dehydrogenase (LDH) and pyruvate formate lyase (PFL) which led to a reduction in unwanted lactate and formate production and an increase in ethanol yield (see Figure 1.1) (Cripps et al., 2009). Gene knockout was achieved by use of homologous recombination. Gene flanking regions were cloned into temperature sensitive vectors. Vectors were transformed into *G. thermoglucosidans* via electroporation and integration was induced by raising the temperature while applying kanamycin selection. The second crossover event was encouraged by removing the kanamycin selection during several rounds of sub culturing. The gene pyruvate dehydrogenase (PDH) was

upregulated by using a strong promoter (taken from the lactate dehydrogenase gene which contains a Rex binding site). Upregulation of PDH was done to redirect metabolic flux towards ethanol production. Recently the genome of the wild type and the industrial strain have been sequenced, and gene annotation is underway.

Despite advancements in strain development for biofuel production, the techniques used (although effective) are time consuming and limited. Genetic manipulation tools are still hindering the rate of progress and a superior understanding of the regulatory mechanisms controlling fermentation genes will help to enable the precise control over metabolism and optimisation of industrial *G. thermoglucosidans* strains (Kananaviciute & Citavicius, 2015).

1.3. Control of respiration in bacteria

Bacteria have evolved complex pathways to adapt to environmental fluctuations, including to oxygen deprivation. To adapt, bacteria have developed environmental sensors and genetic regulators for flexible respiration and energy generation (Green & Paget, 2004). Oxygen is the most prevalent electron accepter during respiration, however some bacteria can utilise others. Depending upon the environmental conditions, bacteria will respire aerobically, anaerobically or in a fermentative way. Some bacteria are obligate aerobes, unable to respire and grow in anoxic environments. Others are facultative anaerobes, capable of utilising alternative electron acceptors such as nitrate instead of oxygen. Obligate anaerobes also exist, to which trace levels of oxygen in the environment are toxic. Bacteria such as these are thought of as extremophiles now, but before biological oxygen production during the Archean Eon they were the predominant life form on the planet (Hamilton, Bryant, & Macalady, 2016). Many environmental niches exist where facultative and obligate anaerobes thrive, breaking down organic compounds to alcohols and organic acids to produce energy. Facultative anaerobes in particular face environments where oxygen concentrations fluctuate regularly. These bacteria must be capable of rapidly shifting from utilising oxygen as an electron acceptor to another energy generating pathway.

1.3.1. Overview of the respiratory chain

Respiration is an essential process for the production of adenosine triphosphate (ATP), the energy currency of the living cell. Aerobic respiration can be split into three main stages: carbohydrate catabolism, the citric acid (TCA) cycle and electron transport. During glycolysis a sugar molecule is broken down to form ATP and pyruvate which then enters the TCA cycle. Acetyl-CoA is subject to a series of oxidation reactions that reduce NAD^+ to NADH, which is utilised as an electron donor for electron transport chain. Electron transport involves a series of oxidation and reduction reactions that transfer electrons between membrane-associated donors and acceptors, the function of which is to create an electrochemical gradient created by hydrogen ions to produce a transmembrane proton motive force (PMF) that is used for ATP synthesis (Marreiros et al., 2016).

Respiration can occur under two main conditions. Aerobic respiration occurs in oxygen-abundant environments, whilst anaerobic respiration occurs in oxygen-limited or anoxic conditions. During aerobic respiration oxygen is the ultimate electron acceptor, resulting in the production of water (Borisov & Verkhovsky, 2015). Facultative anaerobes are able to transition from aerobic to anaerobic respiration by utilising a complex regulatory network that senses and responds to changes in environment (Richardson, 2000). By utilising alternative electron acceptors, they are able to generate energy most efficiency depending upon the abundancy of electron acceptors in the environment.

1.3.1.1. Glycolysis

Glycolysis is an oxygen independent process that occurs in the cytoplasm of bacteria and is the first stage of ATP production. It is the metabolic pathway responsible for converting carbohydrates, such as glucose, to pyruvate through a series of reactions, releasing energy. During the process a six carbon glucose molecule, over the course of nine reactions, is broken up into two molecules of three carbon pyruvate. Pyruvate is utilised in subsequent metabolic pathways to generate additional energy, including during either fermentation or the TCA cycle. Along with the pyruvate molecules and ATP, two molecules of NADH are also produced (see Figure 1.2). NADH molecules act as electron carriers and are re-oxidised by entering oxidative phosphorylation, where ATP synthesis occurs.

$$\text{Glucose} + 2\,\text{ATP} + 2\,\text{NAD}^+ + 2\,p_i \dashrightarrow 2\,\text{Pyruvate} + 2\,\text{ATP} + 2\,\text{NADH} + 2\,\text{H}^+$$

Figure 1.2 Glycolysis reaction. One glucose molecule produces two ATP and two NADH molecules plus two hydrogen ions during glycolysis. The pyruvate molecules are used in the TCA cycle and NADH feeds electrons into oxidative phosphorylation to generate a PMF.

1.3.1.2. TCA cycle

The TCA cycle (commonly referred to as the Krebs cycle after the 1953 Nobel Prize winning scientist Hans Krebs [1900–1981], or alternatively known as the citric acid cycle) follows on from glycolysis and is made up of a series of eight reactions. Once pyruvate is converted into acetyl-CoA, it feeds into the TCA cycle and undergoes reactions that yield reduced electron carriers NADH and FADH (Spaans, Weusthuis, van der Oost, & Kengen, 2015). ATP and GTP are also produced along with CO_2. NADH is the most abundant product of the TCA cycle, which is used to help form the PMF, which is subsequently used to generate ATP in the oxidative phosphorylation. The TCA cycle also produces many intermediates which are important molecules for other metabolic pathways and the production of structural components. Each of the eight reactions is catalysed by a specific enzyme that is precisely regulated by the cell, ensuring energy is metabolised in the most efficient way possible.

1.3.1.3. Oxidative phosphorylation

The majority of energy released during respiration in glycolysis and the Krebs cycle is locked in the coenzyme/electron acceptor NADH. During oxidative phosphorylation this energy is utilised to produce ATP, and is the most efficient and abundant method of production. Many different mechanisms of generating a PMF for ATP synthesis exist, but all pass electrons from a donor, down the electron transport chain, releasing energy in manageable increments to produce a

proton motive force coupled to ATP synthesis. Electrons are carried from proteins with low reduction potential, to those with higher reduction potential, and ultimately to a terminal electron acceptor, which is oxygen during aerobic respiration (Borisov & Verkhovsky, 2009).

Oxidative phosphorylation is a universal metabolic pathway; within prokaryotes this process is located in the inner membrane, whilst in eukaryotes it occurs in mitochondria. An overview of the electron transport chain in *E. coli* will facilitate a good understanding of the mechanisms by which oxidative phosphorylation leads to energy production.

Oxidative phosphorylation is conducted by the electron transport chain, made up of a series of membrane bound dehydrogenases that reduce the quinone pool, and terminal reductases that can oxidise it. The chain components are arranged so that energy is released by a series of oxidative/reductive reactions to create a proton motive force (PMF) coupled to the production of ATP. ATP synthases (known as complex V) use the energy generated by the PMF to form the bonds needed for ATP synthesis. Complexes can be comprised of numerous types of prosthetic groups (hemes, flavins, iron-sulfur clusters and copper clusters) depending upon the most energetically efficient method of electron transport given the circumstances (Gottfried Unden, Steinmetz, & Degreif-Dünnwald, 2014). Gene regulators control the expression of proteins involved in the electron transport chain, in response to environmental conditions and respiration type (aerobic or anaerobic).

1.3.1.3.1. Oxidative phosphorylation in *E. coli*

E. coli has a branched respiration chain, with multiple dehydrogenases and terminal reductases having been described, which enable multiple electron donors and acceptors to be utilised by the organism (see Table 1.2). Electron acceptors include oxygen during aerobic respiration, which is the most energetically efficient, or alternatively nitrate or fumarate can be used (G Unden & Bongaerts, 1997). Each enzyme has an associated redox potential, which increases along the transport chain as electrons are transferred.

Electron donors	dehydrogenases	quinone pool	reductases	Electron acceptors
NADH	NADH dehydrogenase I	Menaquinone	Cytochrome *bo*	Oxygen
Succinate	NADH dehydrogenase II	Ubiquinone	Cytochrome *bd*	Nitrate
Glycerophosphate	Cytochrome c	Dimethylmenaquinone		Fumarate
Formate	Succinate dehydrogenase			
Hydrogen	Glycerophosphate			
Pyruvate	Dehydrogenase			
Lactate				

Table 1.2 Lists of the most prevalent enzymes and electron donors/acceptors in *E.coli* *(G Unden & Bongaerts, 1997).*

A dehydrogenase is the first protein in the electron transport chain. Many alternative dehydrogenases exist that are able to accept electrons from a variety of donors such as lactate and succinate. However, usually an NADH dehydrogenase accepts electrons from NADH and passes them to the quinone pool (see Figure 1.3). In *E. coli* there are two types of NADH dehydrogenases (I and II), which are expressed under different environmental conditions (Yagi, 1993). NADH dehydrogenase I (prosthetic groups FMN, FeS) is a large protein complex that spans the membrane, and is able to pass hydrogen ions into the intermembrane space, contributing to the proton motive force (PMF). It also catalyses the reduction of ubiquinone to ubiquinol and is expressed during times of oxygen starvation. NADH dehydrogenases II (prosthetic group FAD) is a single subunit enzyme and does not span the membrane, and therefore is unable directly to contribute to the PMF. It also catalyses the reaction of ubiquinone to ubiquinol, but it is expressed when oxygen is abundant. During conditions when NADH dehydrogenase II is most highly expressed, the terminal reductase cytochrome *bo* is also upregulated, a membrane spanning protein capable of generating a PMF. This compensates for the

$$NADH \rightarrow NAD^+ + 2H^+ + 2e^- + \tfrac{1}{2} O_2 \rightarrow H_2O$$

Figure 1.3 The overall reaction of the electron transport chain. One NADH molecule contributes two electrons, whose combined energy can transport ten protons to create a PMF. The electrons are eventually transferred to oxygen and combine with two hydrogen ions to form water.

downregulation of NADH dehydrogenases I and allows the cell to recycle NADH to NAD^+ whilst maintaining a PMF during both aerobic and anaerobic conditions. In addition to NADH, intermediates of the TCA cycle such as formate and succinate are also able to act as electron donors. Succinate dehydrogenase catalyses the reaction of succinate to fumarate and ubiquinone to ubiquinol, but does not contribute to the PMF directly. Glycerophosphate dehydrogenase is also an alternative dehydrogenase utilised during anaerobic respiration (X. Zhang, Shanmugam, & Ingram, 2010).

Once electrons have been donated to a dehydrogenase they are passed to the quinone pool. Ubiquinone, ubiquinol, menaquinol and menaquinone make up the quinone pool in *E. coli* and link the transfer of electrons from the dehydrogenases to terminal reductases. Each of the quinones are most abundant during particular growth conditions and are able to carry electrons or hydrogen ions and are small mobile molecules. Ubiquinone is most abundant during aerobic respiration, whereas menaquinone and dimethylmenoquinone dominate during anaerobic conditions (G Unden & Bongaerts, 1997). Due to the midpoint potentials of the quinone, each is only able to function with particular electron donors and acceptors. Ubiquinone is used when oxygen is the electron acceptor, whereas menaquinol are optimum for use with a nitrate terminal electron acceptor (Gottfried Unden et al., 2014). Menaquinone is most optimal for DMSO and fumarate electron donors (G Unden & Bongaerts, 1997).

In *E. coli* there are two main terminal oxidases: cytochrome *bo*, present at high levels during oxygen abundance in aerobic conditions, and cytochrome *bd*, expressed most

during periods of oxygen starvation (Borisov, Gennis, Hemp, & Verkhovsky, 2011). Cytochrome *bd* oxidises menaquinol but is unable to pump protons to contribute to the PMF (Borisov, Gennis, Hemp, & Verkhovsky, 2011). It does however have a higher affinity for oxygen then cytochrome *bo* and can transfer electrons to produce water during microaerobic respiration (Borisov & Verkhovsky, 2015; Jasaitis et al., 2000). Cytochrome *bo*, however, does contribute to the PMF as it contains prosthetic groups Heme O_3 and Cu_B (also Heme b, although this does not help generate a PMF) and oxidises ubiquinone in aerobic conditions, utilising oxygen to produce water (Borisov et al., 2011).

1.3.1.4. ATP Synthase

The final stage of respiration is ATP synthesis which uses the electrochemical gradient, created by the electron transport chain, to drive ATP synthesis. The F_0F_1 ATP synthase enzyme spans the membrane and allows hydrogen ions to flow down a concentration gradient, thereby capturing energy and using it to form the bonds needed for ATP synthesis. The F_1 subunit comprises the catalytic site where ATP synthesis occurs (using ADP and phosphate) (Trchounian, 2004; Weber & Senior, 2003). The F_0 subunit makes up the section which spans the membrane, containing a channel running though it for hydrogen ions to pass through. Variations of ATP synthases are also used to catalyse the reverse reaction to enable other processes such as movement by flagellum (Trchounian, 2004; Weber & Senior, 2003).

1.3.1.5. Oxygen independent respiration

There are two forms of oxygen independent energy production that can be engaged depending upon environmental conditions: anaerobic respiration and fermentation. In contrast to aerobic respiration, where oxygen serves as the final electron acceptor, anaerobic respiration uses an alternative electron acceptor such as nitrate. In this case the electrochemical gradients are generated using a specialised membrane-associated electron transport chain. Fermentation, however, occurs without the use of external electron acceptors: rather, an internally generated metabolite such as pyruvate is used as an electron acceptor and is reduced to energy-rich products such as lactate, butanol and ethanol (Härtig & Jahn, 2012). The main purpose of this fermentative process is to create redox balance by the

process of recycling NADH back into NAD^+, allowing continued ATP production by substrate-level phosphorylation. Many fermentation products are toxic at high concentrations and are excreted by the cell.

Bacteria can be categorised by their ability to adapt to changes in oxygen abundance and their method of energy production. Obligate aerobes require oxygen for oxidative phosphorylation and cannot grow in its absence (although they can often adapt their aerobic respiratory pathways to account for variation in oxygen levels). Facultative anaerobic bacteria are able to respire in the presence of oxygen, and adapt to the absence of oxygen by switching to anaerobic respiratory and/or fermentative pathways. Obligate anaerobes can only grow anaerobically and often perish in oxygen-rich environments.

It is the concentration of oxygen and the substrates available that determine the state of respiration within the cell of a facultative anaerobic bacterium. Facultative anaerobes are therefore able to quickly adapt to changing conditions by directly sensing the amount of oxygen or indirectly, by sensing for example redox signals. Thus, they can control gene expression in response to a variety of environmental conditions, allowing them to inhabit a diverse range of territories and adapt to changing conditions (Morris & Schmidt, 2013).

The *Geobacillus* genus is made up of a combination of both obligate and facultative anaerobes. Soil inhabiting microorganisms such as these face constantly fluctuating oxygen concentrations. During the decomposition of organic matter within the depths of soil, oxygen can become scarce. For this reason, facultative anaerobic bacteria are integral to the continued recycling of carbon. After obligate aerobes at the surface of soil have secreted enzymes that break down the complex organic matter of wood and leaves (Juo, 2003), bacteria such as *G. thermoglucosidans* are able to continue the degradation of biomass beneath soil by respiring not only aerobically, but also surviving oxygen limitation by switching to fermentation or anaerobic respiration. The sensing of oxygen levels and the switch between respiration types involves a complex regulatory network.

1.3.2. Regulators and sensors

Control of metabolic flux and the transition between aerobic, anaerobic respiration and fermentation is regulated by a number of sensors and regulators that influence expression of important genes. The sensors and regulators involved form interconnected networks that allow for a high degree of control over respiration in response to a variety of environmental factors. This ensures that the cell generates energy in the most efficient way possible, in relation to growth conditions, switching off pathways that are unnecessary. There are many sensors and regulators associated with changes in respiration including ArcAB in *E. coli*, ResDE in *B. subillis*, Fnr in both organisms, and Rex in Gram positive bacteria. It is important to understand these regulators, especially when using an organism for practical applications such as biofuel production, so that control of key genes can be engineered for optimal yield.

1.3.2.1. ResDE

The two component sensor kinase and response regulator ResD-ResE system in *B. subtilis* is a key regulator of anaerobic respiration (Durand et al., 2015; Geng, Zuber, & Nakano, 2007). Fnr expression is controlled by the ResD-ResE system in *B. subtilis*, in response to oxygen and nitric oxide levels. ResE has the ability to sense, via a PAS domain, both oxygen and nitric oxide. ResE then modulates ResD activity via phosphorylation, by either phosphorylating or dephosphorylating ResD (Baruah et al., 2004). The ResD-ResE system controls the switch to alternative terminal electron acceptors, from oxygen to nitrate. It does this by regulating genes such as cytochrome *bf* complex, ctaA haem biosynthesis, and the regulator Fnr (Durand et al., 2015; Esbelin, Armengaud, Zigha, & Duport, 2009).

1.3.2.2. Fnr

The Fnr (Fumarate and Nitrate reductase) regulator senses oxygen in the environment directly (Mazoch & Kucera, 2002). The Fnr protein is expressed by both *E. coli* and *B. subtilis* and is activated under anaerobic conditions. Fnr has combined sensory and regulatory activities, unlike the ArcAB two component system in which they are separate, and the crystal structure of Fnr from *Aliivibrio fisheri* has been solved recently (Volbeda, Darnault, Renoux, Nicolet, & Fontecilla-

Camps, 2015). Fnr contains an oxygen-sensitive iron-sulphur cluster (4Fe-4S) that is required to allow it to bind DNA target sequences and thereby control gene expression (Lazazzera, Beinert, Khoroshilova, Kennedy, & Kiley, 1996; Mettert & Kiley, 2005). In the absence of oxygen, the *E. coli* 4Fe4S cluster binds via cysteine residues to promote the formation of a protein dimer which can bind to DNA (Tolla & Savageau, 2010). However in the presence of oxygen the dimerization bonds are broken, preventing DNA binding (Lazazzera et al., 1996). In *B. subtilis*, however, the Fnr protein forms a constant dimer independent of oxygen and evidence suggests it binds to ResD, indicating that the mechanism of activation is different to that in *E. coli* (Härtig & Jahn, 2012). In *B. subtilis,* Fnr positively regulates genes involved in the transition between aerobic and anaerobic respiration and fermentation, including: *narGHJI* operon that encodes nitrate reductase*; arfM*, an anaerobic modulator; and *narK*, a nitrite extrusion transport protein (Nakano, Dailly, & Zuber, 1997; Reents et al., 2006). Fnr has also been found to auto-regulate as well as regulating the expression of other regulators such as ArcA in *E. coli* (Reents et al., 2006).

1.3.2.3. ArcAB

E. coli is a facultative bacterium that can use nitrate as an alternative electron acceptor instead of oxygen. The ArcAB two component system in *E. coli* regulates the transition from aerobic to anaerobic respiration in *E. coli*, allowing it to use nitrate as an alternative electron acceptor, and the transition involves initially the use of alternative oxygen-dependent terminal oxidases (Alexeeva, Hellingwerf, & Teixeira de Mattos, 2003). The ArcAB system is comprised of the response regulator ArcA, along with ArcB, a membrane associated sensor kinase/phosphatase (X. Liu & De Wulf, 2004; Loui, Chang, & Lu, 2009). ArcB is a multi-domain protein with a signal sensor PAS domain that is thought to sense the redox state of the quinone pool in the cell membrane through the reversible formation of an intermolecular disulphide bond that occurs between two subunits (Bekker et al., 2010). During aerobic respiration the majority of the quinone pool is in the oxidised form, and it has been proposed to accept electrons from two cysteine residues in the PAS domain of ArcB, leading to the formation of a disulphide bond that inhibits its activity (Malpica, Franco, Rodriguez, Kwon, & Georgellis, 2004).

During oxygen limitation (microaerobic growth), the accumulation of the quinone pool in a reduced state prevents the formation of a disulphide bridge, giving rise to large structural changes which activate the auto kinase activity of ArcB. This leads to the phosphorylation of ArcA and the binding to DNA to repress genes involved in aerobic respiration and activate those important for anaerobic respiration, such as pflA (X. Liu & De Wulf, 2004).

1.4. Rex is a key regulator of respiratory and fermentation genes in Gram-positive bacteria

Rex is a redox sensor that controls the expression of numerous genes involved in energy generation, such as fermentation genes, in many Gram-positive organisms including strict anaerobes. The general model for the action of Rex is that it acts as a repressor during growth that is not limited by terminal electron acceptor availability, conditions in which the cellular NADH to NAD+ ratio is low on account of the rapid recycling of NADH. During times of oxygen limitation, which causes NADH levels to increase, Rex binds directly to NADH with high affinity and disassociates from DNA, thereby allowing the transcription of target genes (Brekasis & Paget, 2003). Rex is therefore instrumental in allowing the cell to sense and respond to redox stress caused by changes in the rate of respiration and fermentation caused, for example, by oxygen limitation.

1.4.1. Discovery of Rex in *Streptomyces*

Originally Rex was discovered in *Streptomyces coelicolor* and identified as a repressor of cytochrome *bd* terminal oxidase operon (cydABDC; Brekasis & Paget, 2003). The cydABDC operon is controlled by two promoters: cydp2 is constitutive and cydp1 is induced during hypoxia. Mutagenesis of the cydp1 region revealed an operator site (sequence 5'-TGTGAACGCGTTCACA-3'), which was designated ROP (Rex operator) (see Figure 1.4). This was used as a template to identify other ROP sites within the genome. The nuoA-N and SCO3320-hemACD operons were found to contain upstream ROP sites. Interestingly, the first gene in the SCO3320-hemACD transcriptional unit encoded a potential DNA-binding protein which when overexpressed produced a protein that bound to the cydp1 ROP site (Brekasis & Paget, 2003) and acted as a repressor. Subsequent deletion of the SCO3320 gene

resulted in the constitutive expression of the cytochrome cydA gene, verifying its control of expression at this promoter site.

Figure 1.4 Map of cydABC operon in *S. coelicolor* (Brekasis & Paget, 2003; Sickmier et al., 2005).

The amino acid sequence of SCO3320 was predicted to include a Rossmann fold, which is associated with the binding of pyridine nucleotides such as NADH. Strikingly, the ability of SCO3320 protein to bind to DNA was inhibited by NADH, whereas the addition of NAD^+ to the assay counteracted the NADH effects, allowing SCO3320 protein to bind DNA and repress gene transcription again. The conclusion was therefore that the SCO3320 protein is able to detect the ratio of NADH:NAD^+ in the cell and control gene expression accordingly; it was consequently named Rex (for redox regulator) due to its ability to detect redox poise.

1.4.2. Rex is a redox regulator in bacteria

Following its initial discovery in *S. coelicolor*, many Rex orthologues have been recognised in Gram-positive bacteria, including facultative (*Geobacillus*) and obligate (*Clostridia*) anaerobes (Clostridia; Wietzke & Bahl, 2012) and facultative anaerobes (Härtig & Jahn, 2012). The Rex binding site (ROP; TTGTGAnnnnnnTCACAA) is remarkably well conserved across a number of taxonomic groups (Zheng et al.,

2014) and Rex has been identified as a key redox sensor and regulator of fermentation genes in Gram-positive organisms (Härtig & Jahn, 2012; Sickmier et al., 2005b; Wietzke & Bahl, 2012). A *rex* mutant of an industrial biofuel producer, *Clostridium acetobutylicum,* has been studied and it was demonstrated that it produced lesser amounts of hydrogen and acetone, but higher amounts of ethanol and butanol compared to wild type (Wietzke & Bahl, 2012). It was discovered that the *adhE* gene was over expressed in the *C. acetobutylicum* Δ *rex* and therefore that Rex was of key importance to the solventogenic shift. The Rex gene was found to be located just upstream of an operon, including genes vital for conversion of acetyl-CoA to butyryl-CoA (see Figure 1.5), and many *Clostridium* species also have similar gene arrangements, where *rex* is found in close proximity to carbon metabolism and fermentative genes (Wietzke & Bahl, 2012).

The further investigation of and understanding of Rex is key to optimisation and engineering of industrial biofuel-producing bacterial strains. The study of Rex will provide a better understanding of the optimum bioreactor conditions to induce the genes needed for biofuel production, and may also lead to the discovery that *rex* regulates several genes important for biofuel production, which were previously unknown.

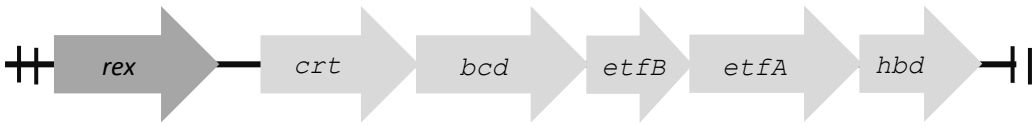


Figure 1.5 The Rex gene and the crt-bcd-etfA/B-hbd (bcs) operon.

1.4.3. The structure and function of Rex

Although the structure of *G. thermoglucosidans* Rex has not been solved, many others have been. X-ray crystal structures are available for Rex proteins from *Thermus*

aquaticus (T-Rex) the identical *Thermus thermophiles* (T-Rex), *B. subtilis* (B-Rex), *Streptococcus agalactiae* (S-Rex), and *Thermoanaerobacter ethanolicus* (RSP) (Protein Data Bank [PDB] accession number 1XCB, 2DT5, 2VT2, 2VT3, 3IKT, 3IKV, 3IL2, 3KEO, 3KEQ and 3KET) (McLaughlin et al., 2010a; Nakamura, Sosa, Komori, Kita, & Miki, 2007; Sickmier et al., 2005a; Wang, Ikonen, Knaapila, Svergun, Logan, & von Wachenfeldt, 2011; Zheng et al., 2014). Of particular use is the thermophilic *Thermus aquaticus* Rex x-ray structures, including it bound to NADH (2.9 A resolution), NAD^+, and DNA together with NAD^+ (Sickmier et al., 2005b; Wang, Ikonen, Knaapila, Svergun, Logan, & Wachenfeldt, 2011). These structures have helped to establish which protein residues are important for DNA and NADH interactions and the mechanism of redox sensing. G-Rex has a 40% identity to T-Rex (see Figure 1.6) and binds to the same consensus DNA rex operator (ROP) site. T-Rex, like all known Rex proteins, consists of three main domains and naturally forms a homodimer. The C-terminus of Rex comprises of a Rossmann fold that can bind NADH (Sickmier et al., 2005b; Wang, Bauer, Rogstam, Linse, Logan, & von Wachenfeldt, 2008). In the absence of NADH, Rex binds to DNA via an N- terminal winged helix-turn-helix motif, consisting of four α-helices and two β-strands (Sickmier et al., 2005). The NADH binding protein residues of T-Rex are located within the centre of the entire protein sequence. As Figure 1.7 illustrates, the C-terminal domain of Rex is comprised of four α-helices and seven β-strands, and forms a Rossmann fold. The final α-helix is domain-swapped, interacting with the α-helix on the opposing subunit, and inserting between the two domains to aid dimer formation (Sickmier et al., 2005). Rex has been found to not only bind NADH, but also NAD^+, and is sensitive to the ratio of $NADH/NAD^+$ within the cell (Brekasis & Paget, 2003; Sickmier et al., 2005). Although Rex can bind NADH, it does not re-oxidise it: Rex merely senses redox poise. Any effect on cellular redox state is indirect and via the regulation of gene expression.

The binding of Rex to NADH causes a change in its conformation, inducing dissociation and preventing interaction with DNA ROP sites. NADH/NAD+ molecules bind to Rex at the dimer interface and induce the rotation of the Rex homodimer (Figure 1.8) (Mclaughlin et al., 2010). The NADH-bound Rex conformation is “closed” and the positions of key residues central in DNA

interactions result in the inability of Rex to fit into the major groove of DNA, and critically to form hydrogen bonds (McLaughlin et al., 2010). However, when NADH dissociates from the T-Rex triggers 43° rigid body rotation of entire subunit and 22Å movement in WH domain allowing DNA binding at ROP sites.

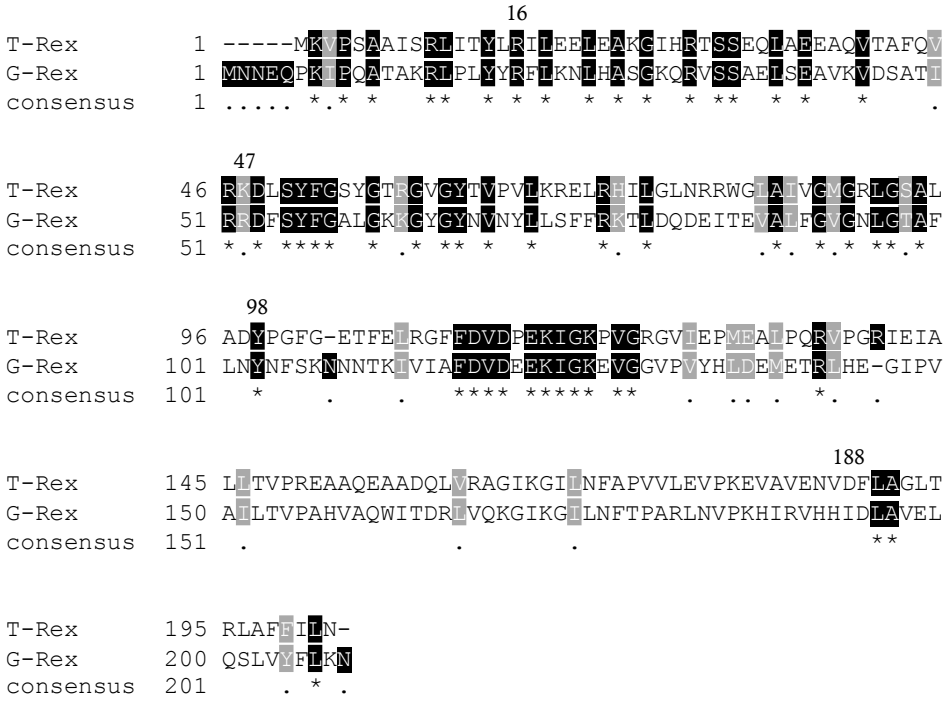


Figure 1.6 Protein sequence alignment of *Thermus aquaticus* (T-Rex) *Geobacillus thermoglucosidans* (G-Rex). Sequences share a 40% identity. The N-terminus of Rex comprises of a winged helix-turn-helix motif. The C-terminus includes a Rossman fold. The NADH binding triad includes R16, Y98, D188. R16 and D188 also form the salt bridge that stabilizes Rex-ROP complex. Residue 47 is integral for ROP binding specificity.

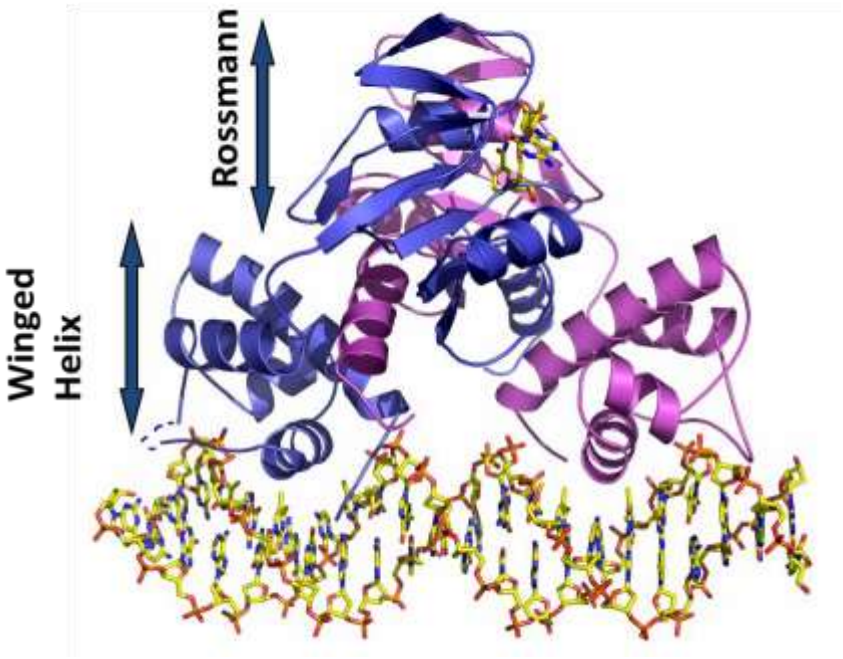


Figure 1.7 T-Rex structure of dimer. The Rossmann fold is a NADH binding domain and the wiggled helix is a DNA binding domain (Mclaughlin et al., 2010; Sickmier et al., 2005).

The ROP site is a highly conserved 18 bp palindromic sequence present at promoter regions controlled by Rex. Rex interacts with both the major and minor groove through its winged helix-turn-helix domains, and the spacing and orientation of these is therefore of major importance. Interestingly, while ROP sites are generally highly conserved across bacteria, Clostridia appear to deviate, having a binding consensus of: TTGTTAANNNNTTAACAA, compared to Actiobacteria and Bacilli; TTGTGAANNNNTTCACAA (D. a. Ravcheev et al., 2012). This difference in binding motif corresponds to substitutions in protein residues found in the DNA recognition α -helices. Clostridia has been the only Rex to have a substituted Gln47, an important hydrogen-bond-forming residue that interacts with guanine7 ROP base. Actionobacteria have a Lys47, and Bacilli a Arg47, both of which are positively charged amino acids, essential for hydrogen bond formation (Zhang et al., 2014). The Rex regulon varies across taxonomic groups from

including genes involved in energy metabolism, NAD(P) biosynthesis, fermentation, nitrogen and sulphur reduction and central carbon metabolism (D. a. Ravcheev et al., 2012). Harnessing and controlling Rex and the role it plays in controlling such genes would be of use in the industrial production of biofuels and other biocommodities.

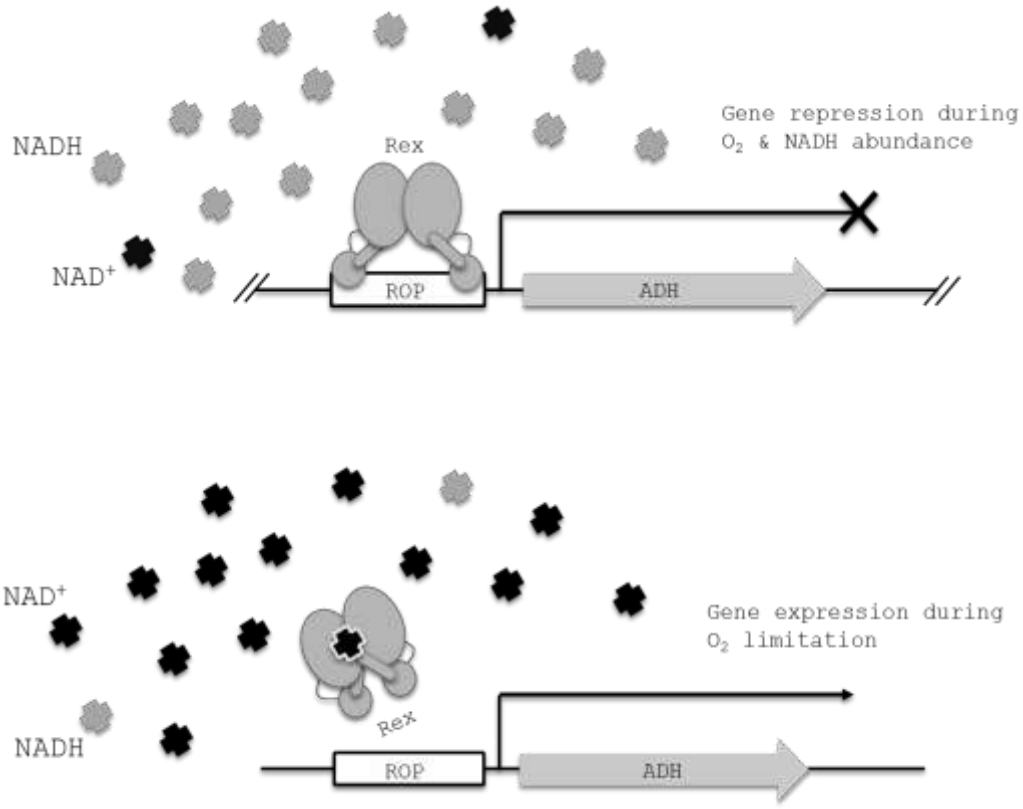


Figure 1.8 Rex binding to *adh* ROP site. During periods of oxygen limitation NAD+ builds up and interacts with Rex, making it disassociate from the ROP site and allowing expression of *adh.*

1.5. Project aims

The aims of this study center on understanding the role of Rex, a regulator of fermentation. To investigate Rex, it was aimed to construct two strains using two alleles: a strain in which Rex was tagged with a FLAG epitope, and a strain where *rex* was deleted. The *rex*-FLAG allele would be integrated into the native site, replacing the original *rex* gene. The tag would allow the investigation of the *G. thermoglucosidans* Rex regulon using chromatin immunoprecipitation techniques, combined with high-throughput DNA sequencing. The *rex* deletion mutant allele would also be integrated into the native locus and the resulting strain analysed for changes in the transcriptome using qRT-PCR and changes in the metabolome (e.g. fermentation products such as ethanol, lactate and formate) using HPLC.

This study also aimed to develop new genetic manipulation tools for the improved transformation of *G. thermoglucosidans.* This was to include the development of vectors and *E. coli* strains for intergeneric conjugation as well as the optimisation of protocols.

Finally, it was aimed to apply the basic understanding of the Rex system towards the construction of synthetic genetic regulatory systems to manipulate the metabolism of *G. thermoglucosidans.* The work will be proof of concept, demonstrating the ability to control gene expression by using a potentially orthoganol Rex.

Ultimately the study aims to contribute to the understanding of fermentation regulation in these industrial organisms and to develop genetic engineering tools that can be implemented to demonstrate the potential for metabolism engineering for the efficient production of biofuels and other biocommodities.

2. Chapter II: Materials and methods

2.1. Chemicals, reagents, enzymes

2.1.1. Chemicals and reagents

- Acrylamide solutions (Severn Biotech Ltd)
- Agarose powder (Melford Laboratory)
- Ammonium persulphate (Sigma)
- Ampicillin (Melford)
- Apramycin (Duchefa Biochemie)
- Bradford Reagent (Sigma)
- Bromophenol blue (Amersham Biosciences)
- Casamino acids (Difco)
- Chloramphenicol (Melford)
- Chloroform (Fisher)
- Dimethylsulphoxide (DMSO) (BDH)
- Dithiothreitol (DTT) (Melford)
- dNTPs (New England Biolabs)
- Glycogen (Fisher)
- Hepes (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) (Fisher)
- Isopropyl-β-D-thiogalactopyranoside (IPTG) (Melford)
- Isoamyl alcohol (Sigma)
- Kanamycin (Melford)
- N, N-dimethyl-formamide (Fisher)
- N-Tris(hydroxymethyl)methyl-2-aminoethane sulfonic acid (TES) (Fisher)

- Nutrient agar (Difco)
- Phenol (Fisher Scientific)
- PEG 1000 (Polyethylene glycol) (BDH)
- Phenol (Fisher Scientific)
- Phenylmethylsulfonyl Fluoride (PMSF) (Sigma)
- Radionuclides ([γ-^{32}P]-ATP) (Perkin Elmer)
- Sodium dodecyl sulphate (SDS) (Fisher Scientific)
- Spectinomycin
- Tetramethyl-ethylenediamine (TEMED) (Fisher Scientific)
- N-Tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid (TES) (Fisher Scientific)
- Trichloroacetic acid (TCA) (Sigma)
- Tris (2-Amino-2-hydroxymethylpropane- 1,3-diol) (Fisher Scientific)
- Tryptone soya broth (TSB) (Oxoid)
- Uracil (Sigma)
- 2-Amino-2-hydroxymethyl-propane-1,3-diol (Tris) (Fisher Scientific)
- 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) (Melford Laboratories Ltd)
- 5-Fluoroorotic Acid (5-FOA) (Fisher Scientific)

2.1.2. Enzymes

- Phusion™ (New England Biolabs)
- QuantiTect SYBR green PCR Kit (QIAGEN)
- Reverse transcriptase iScript (Bio-Rad)
- Taq DNA polymerase (New England Biolabs)

- Shrimp alkaline phosphatase (Promega)
- T4 DNA Ligase (New England Biolabs)
- T4 polynucleotide kinase (New England Biolabs)
- Restriction endonucleases (New England Biolabs)
- RQ RNAse-free DNAse (Promega)
- RiboShredder (Cambio)
- RNase A (Sigma)

2.1.3. Antibodies

- Anti-FLAG M2 monoclonal antibody, from mouse, Sigma F18041.

2.1.4. Solutions and buffers

10 x TBE buffer: 107.8 g Tris base, 55 g boric acid and 7.44 g EDTA, made up to 1 L with H_2O.

Kirby buffer: 1% (w/v) sodium-triisopropylnapthalene sulphonate (TPNS), 6% (w/v) sodium 4-amino salycilate, 6% (v/v) phenol, made up in 50 mM Tris-HCl, pH 8.3.

Phenol/chloroform: 25 ml phenol, 24 ml chloroform, 1 ml isoamyl alcohol.

TE Buffer: 10 mM Tris/HCl, pH 8, 1 mM EDTA in 1 L.

Solution I: 50 mM Tris-HCl, pH 8; 10 mM EDTA. Stored at RT.

Solution II: 200 mM NaOH; 1% (w/v) SDS. Stored at RT

Solution III: 3 M potassium acetate, pH 5.5. Stored at 4ºC.

10 x Primer annealing buffer: 100 mM Tris-HCl, pH 8.0, 150 mM $NaCl_2$, 10 mM EDTA.

STE buffer: 10.3% sucrose, 25 mM Tris HCL pH 8.0, 25 mM EDTA, 2 µg/ µl lysozyme.

Blocking solution: 5ml TBS tween, 2.5 g milk powder, made up to 50 ml with ultra-pure water.

20 x TBS tween: 500mM Tris, 60mM KCl, 2.8M NaCl, 1.0% Tween 20, pH 7.

IP buffer: 50mM Hepes-KOH at pH7.5, 150mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS.

IP Salt buffer: 50mM Hepes-KOH at pH7.5, 500mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS.

5X MOPS buffer: 0.2 M MOPS pH 7.0, 0.05 M sodium acetate, 5 mM EDTA. For 1 L of buffer, dissolve in 0.8 L of double autoclaved ultra-pure H_2O with stirring, adjust the pH to 7.0 with 10N NaOH. Bring the final volume to 1L with double autoclaved ultra-pure H_2O water. Dispense into 200ml aliquots and autoclave.

RNA Sample buffer: 10ml deionized formamide, 3.5 ml 37% formaldehyde, 2 ml 5X MOPS Buffer. Dispense into single-use aliquots and store at –20°C in tightly sealed, screw-cap tubes. These can be stored for up to 6 months.

RNA loading buffer: 50% (v/v) glycerol, 1 mM EDTA, 0.4% (w/v) bromophenol blue. Prepare in double autoclaved ultra-pure H_2O, dispense into single-use aliquots and store at –20°C.

6 x Orange DNA loading dye: 25 mg Orange G and 30% (v/v) glycerol in 10 ml and filter-sterilise.

GelRed Stain: 100 ml ultra-pure H_2O and 2 µl GelRed (Cambridge Bioscience). Stored in the dark.

Binding buffer: 20mM Tris-HCl pH7.9, 500mM NaCl, 0.1mM PMSF.

Charge buffer: 50mM $NiCl_2$

Wash buffer: 20mM Tris-HCl pH 7.9, 0.5M NaCl, 60mM imidazole.

Elution buffer: 20mM Tris-HCl pH 7.9, 0.5M NaCl, 0.5M or 1M imidazole.

Strip buffer: 20mM Tris-HCl pH7.9, 0.5M NaCl, 100mM EDTA.

Semi-Dry buffer: 5.82g Tris Base plus 2.92g Glycine plus 0.0375 % SDS and 20 % Ethanol made in 1 L.

2x protein loading dye: 250 mM Tris-HCl pH 6.8, 2% (v/v) SDS, 10% (v/v) glycerol, 0.01% (w/v) bromophenol Blue and 20 mM DTT.

1x SDS running buffer: 100 ml 10% (v/v) SDS, 30.9 g Tris-HCl, pH 8.3, 144.1 g glycine in 1 L.

Coomassie Brilliant Blue stain: 0.25% (w/v) Coomassie Brilliant Blue (Sigma), 50% (v/v) methanol, 10% (v/v) acetic acid.

Destainer solution: 20% (v/v) methanol, 10% (v/v) acetic acid.

2.2. Strains, vectors and oligos

2.2.1. Plasmids and expression vectors

Plasmids	Features	Source/Reference
pN109	*E. coli - G. thermoglucosidans* conjugative vector: *oriR*, *oriT, traJ*, Amp^R, Kan^R	TMO Renewables
pSX700	*E. coli - G. thermoglucosidans* conjugative vector: *oriT, traJ*, Amp^R , Kan^R	This study
pUCG18	*E. coli, G. thermoglucosidans* bifunctional vector: Amp^R , Kan^R	(Taylor, Esteban, & Leak, 2008)
pSET152	*Streptomyces* conjugative/integrative vector: *lacZα, φC31 int/attP,* Apr^R	(Bierman et al., 1992)
pBlueScript II SK (+)	*E. coli* cloning vector: *bla, lacZα,* Amp^R	Stratagene (Alting-Mees & Short, 1989)
pET15b	*E. coli* T7 based His-Tag expression vector: Amp^R	Novagen
pTMO31	*G. thermoglucosidans* temperature sensitive suicide vector: *pUB110 oriR,* Kan^R	(Cripps et al., 2009)
pG1AK-sfGFP	*G. thermoglucosidans* reporter plasmid: *sfGFP,* Amp^R , Kan^R	(Reeve, Martinez-Klimova, de Jonghe, Leak, & Ellis, 2016)
pG1AK-sfGFP_OriT	*E. coli - G. thermoglucosidans* conjugative reporter plasmid: *sfGFP,* AmpR , Kan^R , *oriT, traJ*,	This study

Table 2.1 list of the plasmids and vectors used during this project.

2.2.2. Bacterial strains

2.2.2.1. *E. coli* strains

E. coli Strain	Description	**Source/Reference**
S17-1	*TpRS mR, recA, thi pro hsdR-M+RP4: 2-Tc:Mu: Km, Tn7 λpir*	(Simon, Priefer, & Pühler, 1983)
DH5α	F- Φ80 *dlacZ, ΔM15* Δ*(lac*ZYA-*argF)* U169, *recA1, endA1, hsdR17* (r^- m^-), *supE44 medλ- thi-1 gyrA, relA1*	Invitrogen
BL21 (λDE3) (pLysS)	*F' ompT hsdSB (rB- mB-) dcm, gal, λ(DE3), pLysS,* Cm^R	Novagen
S17-1 *Δdcm*	*Δdcm,* Apr^R	This study

Table 2.2 List of the *E. coli* strains used during this project. S17-1 strains were used as donors during conjugation. DH5α strain was used during routine laboratory cloning and is suitable for blue/white screening. BL21 (λDE3) (pLysS) strain was used for protein over expression.

2.2.2.2. *G. thermoglucosidans* strains

Geobacillus Strain	**Description**	**Source/Reference**
TM444	A derivative of the industrial TM242 Strain	(Cripps et al., 2009)
11955	WT isolate	David Leak Lab, University of Bath
TM242	Industrial Strain	(Cripps et al., 2009)
11955 Δ*rex*	Used for analysis of rex	This study
TM444 3xFLAG tagged *rex*	Used for analysis of rex	This study

Table 2.3 List of the *G. thermoglucosidans* strains used during this project. Unless otherwise stated work in this study was carried out using the 11955 strain.

2.2.1. Oligonucleotides

Name/target	**Sequence (5'->3')**
F_DCMApr	ATGCAGGAAAATATATCAGTAACCGATTCATACAGCACCATGTGCAGCTCCATCAGCA
R_DCMApr	TTATCGTGAACGTCGGCCATGTTGTGCCTCTTGCTGACGTCCAACGTCATCTCGTTCT
Col_S17-1Dcm_F	GCTACCGCAAACCATGCAAA
Col_S17-1Dcm_R	TCTCTATCGCCGTATCACGC
UPP-LFlank_Fwd	CCTGCAGGTCGACTCTAGAGGCACGGCAGCCCAGTCAA
UPP-LFlank_Rev	CCACCCAGGGTCCTCTTTGTTACGGGTTACACTTCC

attP_pSET1_Fwd	ACAAAGAGGACCCTGGGTGGGTTACAC
attP_pSET1_Rev	CTACTTCTTCCGTTGGCGCTACGCTGT
UPP-RFlank_Fwd	AGCGCCAACGGAAGAAGTAGAAATTGAAACGCCG
UPP-RFlank_Rev	ATTCGAGCTCGGTACCCGGGTCGTTCGCAACATCGCAG
LF-attP_F	GAAGATTTGGCGCGCGAA
LF-attP_R	CTTGCTGCGCTCGAATTCTC
attP-RF_F	GTTACACGACGCCCCTCTAT
attP-RF_R	TGTTCCGAAAAGCAAGCTCA
Int_F	CCCTCGAGATGACACAAGGGGTTGTG
Int_R	GGACTAGTAAAAAAACGCCCCCTTTCGGGGCGCGATTCTACGCCGCTACGTCTTC

Table 2.4 List of the oligonucleotides used to generate and test conjugation in Chapter 3.

Name	**Sequence (5'->3')**	**Use/description**
F1MuantRex_6.12	ATACGACGAGAATAGCGCCG	Δrex generation
R1MuantRex_6.12	ACGTTTTGCGGTTGCTTGTGG	Δrex generation
F2MuantRex_6.12	CACCATATTGATTTGGCGGTC	Δrex generation
R2MuantRex_6.12	GGAATGTTTTTGCGCCGTGA	Δrex generation
Cas2_L	TTCAGACAGCTTTGGTCGTG	qPCR
Cas2_R	CAATCCCTTCAGCGGTAAAA	qPCR
Cbd1_L	GCGATTTTCCTTGGCATTTA	qPCR
Cbd1_R	CCCCGTCTTTCAACTCAAAA	qPCR
adhE2_L	ACAATGCGTACATTCCGTCA	qPCR
adhE2_R	TATCCGAGCTTGACCATTCC	qPCR
Pgi2_F	TGGCGGACACACTGTAGAAA	qPCR
Pgi2_R	ATAAATGGCAAGGCATTGGA	qPCR
L_ldh	AGCCGGGAGAAACAAGACTG	qPCR
R_ldh	TGGGTTCGTTGCCACAAGAA	qPCR
L_ndh	ACGGAAGAAGAAAAGCGGGA	qPCR
R_ndh	CGGCATAATTGTGGAAGGCG	qPCR
L_L-ao	TTTTGCATGCGGGAGGAGAT	qPCR
R_L-ao	TGACGCATTCTCCCTCGATG	qPCR
L_hcbd	CCGGCGCTAAGGAAGAGAAA	qPCR
R_hcbd	CTCGGAGTGCGGATCAAGTT	qPCR
L_mhk	CTGCTGGCAAGAGGGAACAG	qPCR
R_mhk	CTCCGCTTCAGACTCCATCA	qPCR
L_Fnr	GGATGGACGCAGCAGAACTA	qPCR
R_Fnr	ATTTCTCCAGGGCCGCAAAT	qPCR
L_mABC	GCCGCAAGTGTCGTTTCTTT	qPCR
R_mABC	TTCATTTGCTCCATCGACGC	qPCR

Table 2.5 List of the oligonucleotides used to generate and test a *rex* mutant.

Name	Sequence (5'->3')
GRex_reg_F	CGGATCCTATGATATGTGTAAAGCGCTGG
GRec_reg_R	CGGATCCACATCACAACAACCGGCAGCTG
GRex_Flag_F	CACGATATCGACTACAAAGACGATGACGATAAGTAAAGGAGAGAACCAAAAGTG
GRex_Flag_R	GTCCTTATAATCGCCGTCATGATCTTTGTAGTCTTCGGCTGGATAATTTTTCAA
L_cplpny_Re-FLG	ATATTGATTTGGCGGTCGAG
R_cplpny_Re-FLG	GCCCTTTTGTCGCATCTTTA

Table 2.6 List of the oligonucleotides used to generate and test a 3xFLAG tagged *rex* strain.

Name	Sequence (5'->3')
CT1T_F	CCTAACTGTTGTTAAATTGTTAACATCCTAACTGC
CT1T_R	GCAGTTAGGATGTTAACAATTTAACAACAGTTAGG
CT1G_F	CCTAACTGTTGTGAAATTGTTCACATCCTAACTGC
CT1G_R	GCAGTTAGGATGTGAACAATTTCACAACAGTTAGG
GTLD_F	CCTAACTGTTGTGAAATAATGCACAACCTAACTGC
GTLD_R	GCAGTTAGGTTGTGCATTATTTCACAACAGTTAGG
C04422_F	ATTTATTTTTTGTGAAAACTATATAAAACTTTCAT
C04422_R	AATGAAAGTTTTATATAGTTTTCACAAAAATAAAT

Table 2.7 List of the oligonucleotides used during EMSA experiments.

Name	Sequence (5'->3')
autrex/RBSgfp_F	GCACTCTCCTCTTTGTTACGG
autrex/RBSgfp_R	CCAGAATAGGGACGACACCA
GRex_S_top	TCGAATTGAATGATACCGATGTTTGTGATAGAATTGCTTGTGAAATAATGCACAAT
GRex_S_bottom	CTAGATTGTGCATTATTTCACAAGCAATTCTATCACAAACATCGGTATCATTCAAT
GRex_D_top	TCGAATTGAATGATACCGATGTTTGTGATAGAATTGCTTGTGAATGTGAACACATTCACAAT
GRex_D_bottom	CTAGATTGTGAATGTGTTCACATTCACAAGCAATTCTATCACAAACATCGGTATCATTCAAT
CRex_S_top	TCGAATTGAATGATACCGATGTTTGTGATAGAATTGCTTGTTAAATTGTTAACATT
CRex_S_bottom	CTAGAATGTTAACAATTTAACAAGCAATTCTATCACAAACATCGGTATCATTCAAT
CRex_D_top	TCGAATTGAATGATACCGATGTTTGTGATAGAATTGCTTGTTAATGTTAAAACATTAACAAT
CRex_D_bottom	CTAGATTGTTAATGTTTTAACATTAACAAGCAATTCTATCACAAACATCGGTATCATTCAAT
qPCR_1Rex_Fwd	TTCTTATTTCGGTGCGCTTGG
qPCR_1Rex_Rev	CCAAGTTTCCGACGCCAAAT
CRexqPCR_F	TATTCGTCGTCTGCCTCGTT
CRexqPCR_R	GCTTGCGGTAATGCCCATAC
GRexqPCR_F	CAATGAGCAGCCAAAGATCCC
GRexqPCR_R	CACGCTGTTTTCCTGAAGCA

Table 2.8 List of the oligonucleotides used to construct and test G/C-Rex promoter/operator expression circuit.

2.3. Growth media, storage and selection

2.3.1. Media

E.coli Media	Use and preparation	Ingredients per litre of media
Lennox broth (LB)	Liquid media used for routine growth of E. coli strains. 50 ml aliquots in universal bottles were sterilised by autoclaving. LB media was prepared with no glucose for BACTH assays.	10 g Difco Bacto-tryptone, 5 g Difco yeast extract, 5 g NaCl and 1 g glucose.
Lennox agar (LA)	Prepared as LB, but with the addition of 1.5 g of agar per 100 ml. Flask was stoppered with foam bung, covered with foil and autoclaved.	15g agar, 10 g Difco Bacto-tryptone, 5 g Difco yeast extract, 5 g NaCl and 1 g glucose.
Low salt LB/LA	Similar to LA but used when selecting for Kan^R	15g agar, 10 g Difco Bacto-tryptone, 5 g Difco yeast extract, 1 g glucose.
2xYT	A rich liquid medium used for high efficiency transformation. 10 ml aliquots in universal bottles were sterilised by autoclaving.	16 g Difco Bacto-tryptone, 10 g Difco Yeast Extract, 5 g NaCl.
SOC	Rich media for use after high efficiency transformation. Adjust pH to 7.0 with 10N NaOH. Made up to 1 L with distilled water. Autoclave to sterilize.	20 g tryptone, 5 g yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose.

Table 2.9 List of the media used for *E. coli* growth.

G. thermoglucosidans Media	Use and preparation	Ingredients per litre of media
Tryptone soya broth (TSB)	General growth media. Made up to 1 L in distilled water. Autoclave to sterilize.	30g tryptone soya broth powder
Tryptone soya agar (TSA)	General growth media. Made up to 1 L in distilled water, 100ml aliquoted into conical flasks and then 1.5 g agar is added before autoclaving.	30g tryptone soya broth powder, 15g agar
USM	Fermentation tube culture evaluations of strains uses the USM media base. When making USM seed media, the ingredients except for biotin are made up to ~500ml then pH adjusted to 6.0. The biotin, yeast extract and C-source are then added before pH adjusting to 6.7 and the final volume is made up to 1 L. In standard seed medium, yeast extract is added at a concentration of 20 g/L and glycerol at 30 g/L Store out of direct light or in an opaque bottle.	12.4 ml 2M NaH_2PO_4, 12.5 ml 4M Urea, 50 ml 0.5M K_2SO_4, 5 ml 1M Citric acid, 12.5 ml 0.25M $MgSO_4$, 2.5 ml 0.02M $CaCl_2$, 12.5 ml Sulphate TE solution, 94 µl 0.8 mg/ml Biotin, 0.25 ml 10 mM Na_2MoO_4
Sulphate TE solution	Made up to 1 L in distilled water. Store out of direct light or in an opaque bottle in fridge.	Conc. H_2SO_4 5 ml, $ZnSO_4.7H_2O$ 1.44g, $Fe_SO_4.7H_2O$ 5.56g, $MnSO_4.H_2O$ 1.69g, $CuSO_4.5H2O$ 0.25g, $CoSO_4.7H2O$ 0.562g, H_3BO_3 0.06g, $NiSO_4.6H2O$ 0.886g

Minimal media (MM)	Solutions were combined after each had been autoclaved separately. Made up to 1 L in distilled water.	20% (v/v) M9 salt (5x) solution, 80% (v/v) casamino acids + D-glucose solution
Casamino acids and D-glucose solution	Made up to 1 L with RO water and the solution was autoclaved.	1.25g casamino acids and 12.5g D-glucose
M9 salt (5x) solution	The pH was adjusted to pH 7.5 and the solution autoclaved.	1.5g K_2SO_4, 1.03g Na_2HPO_4, 5g NH_4Cl, 3.7g $MgSO_4.7H_2O$, 0.015g $MnCl_2.4H_20$, 0.025g $CaCl_2.H_2O$, 0.021g $FeCl_3$, 10mM Tris-HCl
R2YE trace elements	Made up to 1 L with RO water. Store out of direct light or in an opaque bottle in fridge.	40mg $ZnCl_2$, 200mg $FeCl_3.6H_2O$ 10mg $CuCl_2.2H_2O$, 10mg $MnCl_4.4H_2O$, 10mg $Na_2B_4O_7.10H2O$, 10mg (NH4)6Mo7O_2.4H2O

Table 2.10 List of the media used for *G. thermoglucosidans* growth.

2.3.2. Antibiotics and additives

Name	**Stock concentration**	**Stock solvent**	**Working concentration (liquid media)**	**Working concentration (solid media)**
Ampicillin	100 mg/ml	60% ethanol	100 µg/ml	100 µg/ml
Apramycin	50 mg/ml	dH2O (Filter-sterilised)	20-50 µg/ml	20 µg/ml
Choramphenicol	34 mg/ml	100% ethanol	34 µg/ml	25 µg/ml
IPTG	1 M	dH2O (Filter-sterilised)	1 mM	0.5 mM
Kanamycin	50 mg/ml	dH2O (Filter-sterilized)	12.5 µg/ml	12.5 µg/ml
X-gal	40 mg/ml	N, N-dimethyl-formamide	N/A	40 µg/ml
5-FOA	100 µg/ml	DMSO (Filter-sterilised)	1-200 µg/ml	1-200 µg/ml
Uracil	10 µg/ml	100% ethanol	1-10 µg/ml	1-10 µg/ml

Table 2.11 List of antibiotics and additives used in growth media for selection.

2.3.3. Growth and storage of bacterial strains

2.3.3.1. Growth of *E. coli* cultures and storage

E. coli strains were grown at 37°C on solid media or in liquid media shaking at 250 rpm. Colonies formed on solid media plates were kept for up to 3 weeks at 4°C. To achieve long-term storage, overnight LB cultures were suspended in a final concentration of 20% (v/v) glycerol, and stored at -80°C.

2.3.3.2. Preparation of chemically competent *E. coli* by $CaCl_2$ method

Chemically competent *E. coli* cells were prepared by inoculating 5 ml of LB media from glycerol stocks and grown overnight at 37 ºC with shaking at 250 rpm. The overnight culture was used to inoculate (by a ratio of 1:100) 50 ml of LB media in a 250 ml conical flask and incubated at 37 ºC with shaking at 250 rpm until it reached an OD^{600} of 0.4. The culture was harvested by centrifugation at 4ºC, 2000 x g for 8 min. The cell pellet was resuspended in 20 ml of chilled 50 mM $CaCl_2$, incubated in an ice bath for 20 min, and centrifuged again at 4ºC, 2000 x g for 8 min. The pellet was gently resuspended in 2.5 ml of chilled 0.1 M $CaCl_2$, 10% glycerol. 100 µl aliquots were frozen in a dry ice ethanol bath and stored at -80°C.

2.3.3.3. Preparation of electrocompetent E. coli

Electrocompetent *E. coli* cells were prepared by inoculating 5 ml of LB media from glycerol stocks and grown overnight at 37 ºC with shaking at 250 rpm. The overnight culture was used to inoculate (by a ratio of 1:100) 50 ml of LB media in a 250 ml conical flask and incubated at 37 ºC with shaking at 250 rpm until it reached an OD^{600} of 0.35. The culture was harvested by centrifugation at 4ºC, 2000 x g for 10 min. The cell pellet was resuspended in 50 ml of chilled 10% v/v glycerol. This was centrifuged again at 4ºC, 2000 x g for 5 min and resuspended in 25 ml of chilled 10% v/v glycerol. A further centrifugation at 4ºC, 2000 x g for 10 min followed and the pellet resuspended in 10 ml of chilled 10% v/v glycerol. A final centrifugation followed at 4ºC, 2000 x g for 10 min and the pellet resuspended in the residual volume ~1.5ml. 50 µl aliquots were frozen in a dry ice ethanol bath and stored at -80°C.

2.3.3.4. Growth of *G. thermoglucosidans* cultures and storage

G. thermoglucosidans strains were grown at 60°C on solid media or in liquid media shaking at 250 rpm, all media and equipment was pre-warmed to reduce temperature stress. To grow *G. thermoglucosidans*, frozen stocks were used to inoculate solid media (e.g. TSA) and incubated overnight. Plates were inoculated by spreading 150 µl of glycerol stock or a sterile loopful of biomass onto a plate. Distinct colonies or confluent growth was harvested and the biomass was dispersed in a small amount of liquid media before being used to inoculate further pre-warmed media.

Aerobic growth of a 50 ml culture was achieved in a 250 ml flask. Oxygen limited growth was achieved by completely filling a 15 ml tube or a 1.5 ml microfuge tube with culture and tightly sealing it. When grown on solid media, agar plates were placed in plastic bags including a damp cloth to prevent drying out. Colonies that formed on solid media plates were kept for 1 week at 4°C before being discarded. To achieve long-term storage, overnight cultures were suspended in a final concentration of 33% (v/v) glycerol, and stored at -80°C.

2.3.3.5. Preparation of electrocompetent *G. thermoglucosidans* cells

Electrocompetent *G. thermoglucosidans* cells were prepared by inoculating 50 ml of TSB media from glycerol stocks and grown overnight at 52ºC with shaking at 250 rpm overnight. 5 ml of the overnight culture was used to inoculate 50 ml of TSB media in a 250 ml conical flask and incubated at 55ºC with shaking at 250 rpm until it reached an OD^{600} of between 1.5 - 2.0. The culture was placed on an ice bath for 10 min and cells harvested by centrifugation at 4ºC, 2000 x g for 15 min. The cell pellet was resuspended in 50 ml of chilled electroporation buffer (0.5 M sorbitol, 0.5 M mannitol and 10% v/v Glycerol). The cell pellet was sequentially washed in 30 ml, 20 ml, 10 ml of chilled electroporation buffer. A final centrifugation at 4ºC, 2000 x g for 15 min followed and the pellet resuspended in the residual volume of ~2 ml. 60 µl aliquots were frozen in a dry ice ethanol bath and stored at -80°C.

2.4. DNA and RNA manipulation and molecular cloning

2.4.1. DNA gel electrophoresis

To analyse DNA size and quality agarose gel electrophoresis was used. DNA samples were prepared by adding 2µl Orange G loading dye, DNA sample 1-10 µl, H_2O make up to a total of 12 µl before loading into the gel wells. Gels of between 0.8-1% agarose, in TBE buffer were made and run at 2 V/cm. Gels were stained after electrophoresis using GelRed Nucleic Acid Gel Stain and visualised using a UV transilluminator.

2.4.2. RNA gel electrophoresis

To analyse RNA size and quality it must be separated on a denaturing gel to disrupt secondary structure. RNA samples were prepared by adding RNA Sample buffer in a ratio of 1:2 to make a total volume of 10 µl. Samples were then heated to 60°C for 5 min and allowed to cool for 2 min before 2 µl of RNA Loading buffer was added. Gels consisted of 36.5 mM MOPS, 9.1 mM sodium acetate, 0.9 mM EDTA, 2 M formaldehyde, and 1.2% agarose and were run at 4-5 V/cm. Gels were stained using GelRed Nucleic Acid Gel Stain and visualised using a UV transilluminator.

2.4.3. DNA gel purification

For extraction and purification DNA from agarose gel QIAquick™ or MinElute™ Gel Extraction Kits were used according to manufacturer's instructions.

2.4.4. General polymerase chain reaction (PCR)

Every reaction consisted of 1 µl 50 ng/µl template DNA, 5 µl 10 pmol/µl of each primer, 5 µl 10x Polymerase buffer, 1.5 µl 10 mM dNTPs, 1 µl polymerase. Each reaction was made up to 50 µl with ultra-pure H_20. Reactions were also supplemented with either 2.5µl DMSO if appropriate. The polymerase used was determined by the intended purpose of the product: Phusion if product is to be used downstream, Taq if diagnostic.

Thermocycler conditions:

1. Initial denaturation: 2 min at 95 °C
2. Denaturation: 30 sec at 95 °C
3. Annealing: 30 sec at 50-68 °C
4. Extension: ~30 sec/kb of product at 72 °C
5. *Repeat steps 2-4 for 30 cycles*
6. Final extension: 5 min at 72°C
7. Hold at 4 °C

The conditions of the annealing and extension phases were optimised for each reaction according to the primer and template used.

2.4.5. Inverse PCR

The reaction was prepared as described in Section 1.4.4, however the primers were phosphorylated and for larger products such as vectors (>3kb), the following PCR cycle was used:

1. Initial denaturation: 2 min at 95 °C
2. Denaturation at 95 °C for 1 min
3. Annealing at 55-65 °C for 45 s
4. Extension at 72 °C for 2.5 min per 1 kb
5. *Repeat steps 2-3 for 10 cycles*
6. Denaturation at 96 °C for 1 min
7. Annealing at 55-65 °C for 45 s
8. Extension at 72 °C for 3.5 min per 1 kb
9. *Repeat steps 6-8 for 10 cycles*
10. Final Extension at 72 °C for 15 min
11. Hold at 4 °C

2.4.6. Colony PCR

Colony PCR was used to identify mutants or genetically altered strains. A single colony was picked from a plate and biomass suspended in sterile ultra-pure H_2O

and heated to 100 °C for 10 min. Cell debris was removed by centrifugation and 5 µl of the supernatant used as template DNA in a standard Taq PCR.

2.4.7. DNA ligation

Ligation reactions consisted of 1 µl 10X T4 DNA Ligase buffer, 1 µl T4 DNA ligase, 100 ng vector DNA, the appropriate quantity of insert DNA (usually a ratio of 1:3 see equation). This was made up to a total volume of 10 µl. The reaction proceeded at either room temperature for 2 h, 4 °C for 16 h, or 16 °C for 4 h.

$$\text{Insert (ng)} = \frac{\text{100 ng X size of insert (kb)}}{\text{Size of vector (kb)}} \text{ X ratio (e.g 1/3)}$$

Figure 2.1 Equation used to calculate amount of DNA required for an optimal ligation reaction.

2.4.8. Oligonucleotide annealing for construction of dsDNA fragments

Equal molar amounts of each primer mixed with 10 x Annealing Primer buffer (diluted to 1 x in reaction). This was made up to volume with ultra-pure H_2O, heated to >95 ºC for 2 min, then cooled to 50 ºC over at least 30 min. Placed on ice until use or stored at -20 ºC.

2.4.9. Restriction digest

Restriction digests contained ~1 µg DNA, ~10 U of each enzyme, 2 µl 10X Restriction buffer and were made up to a total volume of 20µl using ultra-pure H_2O. Digestion occurs over 2-4 hs at the temperature recommended by the manufacturer of the enzyme. Samples were analysed via agarose gel electrophoresis for the

production of the desired fragment sizes. Fragments, if necessary for downstream use, were isolated from the agarose gel and purified.

2.4.10. Dephosphorylation of DNA

To inhibit self-ligation 5' ends of digested vectors were dephosphorylated. 1 µl shrimp alkaline phosphatase, ~1 µg restriction digested vector, 3.5 µl 10 x Phosphatase buffer were mixed in a final volume of 35 µl with ultra-pure H_2O and incubated at 37 °C for 30 min, then heat inactivated at 65 °C for 20 min.

2.4.11. Phosphorylation of DNA primers

4 µl 200 pmol/µl primers were phosphorylated using 1 µl T4 polynucleotide kinase, 2 µl 10X Kinase buffer, 2 µl 10 mM ATP. Made up to 20µl with ultra-pure H_2O and incubated at 37 °C for 20 min, and then heated at 90 °C for 2 min to inactivate the enzyme before using for PCR.

2.4.12. Determination of DNA/RNA concentration

Determination of RNA or DNA concentrations was achieved using a NanoDrop 1000 (Thermo Fisher Scientific). Between 0.5-2 µl of sample was loaded and measured to give an approximation of the concentration. For a more accurate measurement the Qubit High Sensitivity DNA assay kit was used along with the Qubit 2.0 Fluorometer.

2.4.13. Gibson assembly

PCR amplified fragments with overlapping regions (20 bp) were Gibson assembled in one reaction. 5 µl of the DNA fragments (0.5 pmol of each fragment) were added to 15 µl of Gibson Assembly master mix (NEB) and incubated at 50 °C for 1 h. 1-2 µl was then used in a transformation with extra competent *E. coli* (High Efficiency, NEB #C2987) cells.

2.4.14. DNA sequencing

Sequencing of DNA was carried out externally by Eurofins Genomics using the value read service. 15 µl of 50-100 ng/µl purified plasmid samples were sent for each strand to be sequenced in 1.5 ml Eppendorf tubes.

2.4.15. Generating an in-frame disruption strain using λ RED

This technique can be used for gene targeted deletion and uses a λ Red recombinase under the control of an inducible promoter on a low copy plasmid (Gust, Rourke, Bird, Kieser, & Chater, 2004). It was applied in this project to generate a S17-1 Δ*dcm* strain of *E. coli.* To target the gene of interest (F_DCMApr and R_DCMApr) primers were used that amplify an apramycin antibiotic resistance cassette. These primers had homology of 39 nt in length to the S17-1 *dcm* gene that was to be inactivated. The resulting PCR product therefore contained an apramycin resistant gene flanked by regions of *dcm* homology. S17-1 *E. coli* cells were transformed with the λ-Red recombination plasmid pIJ970 and Electocompetent S17-1 (pIJ970) cells were then transformed with 100 ng of the *dcm::apr* allele cassette (PCR product). Transformants were grown over night at 30 °C on LA plates containing chloramphenicol to select for pIJ970, 1 mM L-arabinose to induce λ-Red mediated hyper-recombination, and apramycin to select for recombination of cassette. Ex-transformants were then grown at a raised temperature of 37 °C to induce the loss of the pIJ970 plasmid. The strain was screened for the loss of Cm^R and gain of Apr^R, then analysed by colony PCR and methylation specific restriction digest.

2.4.16. DNA purification

2.4.16.1. Small scale miniprep plasmid purification

A cell-pellet resulting from a ~5 ml culture was resuspended in 200 µl of solution I. Immediately 400 µl of solution II was added and mixed by inverting. 300 µl of solution III was then added and the contaminating RNA was degraded by the addition of 1 µl 10 mg/ml RNAse A and incubated for 10 min at room temperature. The samples were centrifuged at 16,100 x g for 5 min and the supernatant extracted with 150 µl phenol/chloroform. The mixtures were vortexed

for 2 min, centrifuged as before and precipitated with 600 µl isopropanol. The samples were incubated on ice for 10 min and centrifuged as above. The pellet was washed with 70% ethanol, allowed to dry and finally resuspended in 30 µl TE buffer or ultra-pure H2O.

2.4.16.2. Large scale midiprep plasmid purification

Each preparation required a cell pellet from 50 ml of culture and QIAGEN Plasmid Midiprep kit was utilised for plasmid purification and the manufacturer's procedure followed.

2.4.16.3. Chromosomal DNA extraction

For *G. thermoglucosidans* chromosomal DNA extraction and purification 5 ml of culture was harvested by centrifugation at 13,000 x g for 1 min and the supernatant discarded. The pellets were washed with 1 ml 10.3% sucrose and then resuspended in 250 µl of STE buffer and incubated at 37 °C for 30 min. 330 µl of Kirby mix was added and vortexed for 30 sec. 670 µl of phenol/ chloroform was added and vortexed for another 30 sec before centrifuging again at 13,000 x g for 5 min. The upper phase was transferred to a new tube and 250 µl of phenol/ chloroform was added, vortexed, and centrifuged at 13,000 x g 4 °C for 5 min. The upper phase was transferred to a new tube and an equal volume of isopropanol and a 1/10 volume of 3 M sodium acetate was added and mixed by inversion. The mixture was stored at -20 °C for 1 h. This was again centrifuged at 13,000 x g 4 °C for 5 min and the supernatant carefully discarded. The pellet was resuspended in TE buffer and incubated for 30 min at 37 °C with 10 µl/ml of RNase A. Chromosomal DNA was extracted with 100 µl of phenol/chloroform and precipitated with equal volume of isopropanol and a 1/10 volume of 3 M sodium acetate as before. The pellet was washed with 150 µl of ethanol and left to air dry for 5 min before being resuspended in 100 µl of TE buffer.

2.4.17. RNA Isolation and purification

2.4.17.1. Cryogenic grinding method

To isolate RNA from *G. thermoglucosidans,* 15 ml of cell culture was added to 400 µl of ice-cold 95% ethanol, 5% phenol/chloroform/isoamyl alcohol (25:24:1) mix per ml of culture to protect RNA, and then vortex mixed. Samples were then centrifuged at 6,000 x g 4 °C for 10 min and the supernatant carefully discarded. The pellet was vortexed in 1ml of ice cold TE buffer to dislodge it and small droplets flash frozen in liquid nitrogen. The cells were then placed into cryogenic grinding apparatus and beaten for 1.5 min at 30 Hz for three cycles, cooling in liquid nitrogen between cycles. After removing the lysed cells from the cryogenic grinding apparatus a Qiagen RNeasy midi kit was utilised to purify RNA according to the manufacturer's instructions.

2.4.17.2. Rapid high throughput sonication method

To rapidly isolate RNA from *G. thermoglucosidans* 1.5ml of culture was harvested by centrifugation at top speed at 4 °C for 30 sec. The pellet was immediately resuspended in chilled 400 µl Kirby mix. Lysed cells were sonicated using a Diagenode Bioruptor sonicating waterbath for 10 cycles of 30s on/30s off pulses. To isolate RNA from the cell debris 300 µl of phenol/chloroform was added, vortexed for 1 min, and centrifuged at 13,200 x g 4 °C for 5 min and the upper phase transferred to a new tube. 200 µl of phenol/ chloroform was added, vortexed, and centrifuged at 13,200 x g 4 °C for 5 min. The upper phase was transferred to a new tube and an equal volume of isopropanol and a 1/10 volume of 3 M sodium acetate was added and mixed by inversion. The mixture was stored at -20 °C for 1 h. This was again centrifuged at 13,200 x g 4 °C for 5 min and the supernatant carefully discarded. The pellet was washed with 150 µl of ethanol. The pellet was resuspended in 200 µl of 1x DNAse buffer and 0.5 µl of RNAse-free DNAse added and incubated at 37 °C for 30 min. After incubation 200 µl of RNAse-free H_2O was added and 200 µl of phenol/chloroform, then vortex for 1 min. Centrifuged at 13,200 x g 4 °C for 5 min and the upper phase was transferred to a new tube and an equal volume of isopropanol and a 1/10 volume of 3 M sodium acetate was added and mixed by inversion. The mixture was stored at -20 °C for 1 h and centrifuged at

13,200 x g 4 °C for 5 min and the supernatant carefully discarded. The pellet was washed with 150 µl of ethanol and left to air dry for 5 min before the pellet was resuspended in 50 µl of RNAse-free H_2O. Samples were analysed for quality by gel electrophoresis (28S RNA and 18S RNA band intensity) and NanoDrop and stored at -80 °C.

2.4.18. DNA introduction into bacteria

2.4.18.1. *E. coli* transformation

2.4.18.1.1. Heat shock

A 150 µl aliquot of chemical competent cells was thawed on ice, mixed with 1-2 µl of DNA (10-300ng) and gently mixed by pipetting, then incubated on ice for 10-30 mins. The vial was then heat shocked at 42 °C for exactly 30 sec and immediately placed back on ice for 5 min. 950 µl of pre warmed SOC or LB was added and the vial incubated at 37 °C shaking at 250 rpm for 1 h. 50-100 µl of the transformation mix was then plated onto pre warmed LA plates containing the appropriate antibiotic. For low efficiency transformations the transformation mix was centrifuged at 13,000 x g for 30 sec, the pellet resuspended in a residual ~50 µl and then plated out as outlined before.

2.4.18.1.2. Electroporation

A 60 µl aliquot of electrocompetent cells was thawed on ice, mixed with 1-2 µl of plasmid DNA (~100 ng) and gently mixed by pipetting, then incubated on ice for 10 min. The mixture was then transferred to an ice-cold (1 mm gap) electroporation cuvette and electroporated at 2500 V (capacitance 10 µF, and resistance 600 Ω, time constant 4-6). Immediately after pulsing, 500 µl of pre warmed SOC or LB media was added to the cuvette and the suspension transferred to an Eppendorf tube containing an additional 500 µl of pre warmed media. The cells were incubated at 37 °C for 1 h with shaking at 250 rpm and plated on pre-warmed LA plates containing appropriate selective antibiotic.

2.4.18.2. *G. thermoglucosidans* transformation

2.4.18.2.1. Electroporation

A 60 µl aliquot of electrocompetent cells was thawed on ice, mixed with 1-2 µl of plasmid DNA (~100ng) and gently mixed by pipetting, then incubated on ice for 10 mins. The mixture was then transferred to an ice-cold (1 mm gap) electroporation cuvette and electroporated at 2500 V (capacitance 10 µF, and resistance 600Ω, time constant 4-6). Immediately after pulsing, 500 µl of pre warmed TSB media was added to the cuvette and the suspension transferred to an Eppendorf tube containing an additional 500 µl of pre warmed TSB. The cells were incubated at 55 °C for 2 h with shaking at 250 rpm and plated on pre-warmed TSA plates containing appropriate selective antibiotic. Colonies usually became visible after 1 day incubation, but could take longer to grow.

2.4.18.2.2. Conjugation

For conjugation into *G. thermoglucosidans, E. coli* strain S17-1 Δ*dcm* was used. 100 µl of *E. coli* overnight culture was used to inoculate 5 ml LB containing 100 µg/ml ampicillin to select for the conjugative vector. This was grown at 37 °C with shaking at 250 rpm to an OD_{600} 0.4-0.6. The cells were harvested by centrifugation at 6,000 x g at 37 °C for 6 min and washed twice in pre warmed LB (37 °C). *G. thermoglucosidans* was grown by using biomass to inoculate pre warmed (60 °C) TSB and grown till OD ^{600} 0.4 at 60 °C with 250 rpm shaking. This was then cooled to 37 °C before being added in a 1:9 ratio to washed and pelleted *E. coli* cells and vortex mixed to resuspend all cells. The mixed *G. thermoglucosidans* and *E. coli* cells were then centrifuged at 6,000 x g at 37 °C for 10 min and resuspended in residual ~100 µl of pre warmed (37 °C) TSB per 10 ml of mixed cells. 100 µl of cells were plated onto pre-warmed (37 °C) 20 ml TSA + 10 mM $MgCl_2$ plates, allowed to dry in 37 °C for 5 min before placing lids. Plates were incubated at 37 °C for between 1-3 hs to allow conjugation to occur. Plates were then incubated at 60 °C for 1-3 h to allow *G. thermoglucosidans* to recover. To select for uptake of the conjugative vector, an antibiotic overlay was applied evenly. Each plate was covered in a dilution of kanamycin (5 µl of Kan 50 mg/ml, in 500 µl of TSB) and allowed to dry in 60 °C incubator for 10 min before replacing lids. The final concentration of selection was

12.5 µg/ml kanamycin. Plates were incubated overnight to produce visible ex-conjugate colonies.

2.4.19. Analysis of nucleic acids

2.4.19.1. qRT PCR

RNA samples were reverse transcribed using High Capacity RNA-to-cDNA Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. RT-qPCR experiments were performed using a 96-well plate and a StopOnePlus Real-Time PCR system (Applied Biosystems). Each reaction contained 12.5 µl maxima SYBR Green qPCR master mix, 2 µM of a specific set of up and downstream primers, 1 µl of template cDNA (cDNA was diluted 1/50) and made up to a total of 25 µl with nuclease free H_2O. Data was analysed by StopOnePlus Applied Biosystems Real-Time PCR Software.

2.5. Protein expression and purification

2.5.1. Protein expression

BL21 (pLysS) transformants were inoculated into LB with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol and incubated overnight at 37°C with 250 rpm shaking. The overnight culture was used to inoculate 20-250 ml cultures grown to an OD^{600} of 0.4-0.7. The cultures were cold-shocked in an ice bath for 15 min to induce cold-shock proteins. Protein expression was then induced by adding 1 mM IPTG and grown at 30-37°C for 3 h with 250 rpm shaking.

2.5.2. Protein purification

Cell pellets were resuspended in Binding buffer, 25 µg/ml PMSF and a protease inhibitor tablet (Roche) was added. This was sonicated 6 x 10 s with 50 s intervals at 30% amplitude. The cell suspension was centrifuged at 20,000 x g for 10 min at 4 °C. The cleared supernatant was then loaded on a Ni-affinity column for purification.

2.5.3. Nickel sepharose hand-made column

A Ni-NTA sepharose column was prepared by filling the tip of a syringe with glass wool and then topping with ~4ml of iminodiacetic acid (IDA) sepharose fast-flow resin. The column was washed with 3 column volumes (CV) of dH_2O and Ni-charged with 5 CV of Charge buffer. The column was washed again with Binding buffer and the cleared supernatant loaded onto the column. The column was washed with 10 CV of Binding buffer 5 CV of Wash buffer. 1-2 CV of Elution buffer was applied and the samples collected from the flow-through and analysed by SDS polyacrylamide gel. Before storage at 4°C the column was washed with 6 CV of Strip buffer and 3 CV of dH_2O then stored in 20% ethanol.

2.5.4. Concentrating Protein

Protein samples were concentrated using VIVASPIN-6 membrane 5,000 MWCO PES (Sartorius stedim biotech) and VIVASPIN-500 membrane 3,000 MWCO PES (Sartorius stedim biotech) by following the manufacturer's instructions. To prepare the membrane dH_2O was added to the column and centrifuged at 2,000 x g for 5 min at 4 °C. The protein sample was centrifuged at 10,000 x g at 4 °C until the required concentration or volume was achieved.

2.5.5. Determining a protein concentration

2.5.5.1. Bradford assay

Protein quantification using the Bradford assay is based on an absorption shift of Brilliant Blue G dye in the Bradford Reagent (Sigma) from 465 to 595 nm when the protein binds. To measure the concentration of an unknown protein concentration 1-200 µl of protein was added to 300-499 µl dH2O and 500 µl Bradford reagent. The absorbance was measured and compared to a standard curve previously generated using a variety of known concentrations of Bovine Serum Albumin (BSA).

2.5.5.2. Qubit

Protein samples were accurately determined using a Qubit protein assay and Qubit 2.0 Fluorometer according to manufacturer's instructions.

2.6. SDS polyacrylamide gel

SDS polyacrylamide gels were prepared using 15% resolving gel and a 5% stacking gel. The equipment used was a Hoefer Mighty Small dual gel caster (Hoefer miniSE 260 unit– Amersham Biosciences) and was assembled as per manufacturer's recommendation. The resolving gel consisted of 5 ml 30% (v/v) acrylamide/bis-acrylamide mix solution (Severn Biotech), 2.5 ml 1.5 M Tris pH 8.8, 0.1 ml 10% (v/v) SDS, 50 µl 10% (w/v) Ammonium Persulphate (APS), 20 µl TEMED and 2.33 ml ultra-pure H_2O. This was poured to fill three quarters and allowed to set before the stacking gel was added. 70% ethanol was used to cover the gel as it set and was then rinsed away with ultra-pure H_2O. The stacking gel consisted of 0.75 ml 30% (v/v) acrylamide/bis-acrylamide mix solution (Severn Biotech), 1.25 ml 0.5 M Tris pH 6.8, 50 µl 10% (v/v) SDS, 50 µl 10% APS, 8 µl TEMED and 2.89 ml ultra-pure H_2O. A comb was inserted carefully ensuring no air bubble was introduced immediately after the stacking gel was poured. The gel was left to set for 30 min before use.

2.6.1. Sample preparation

Protein samples were prepared to load on the gel by adding 2 x protein loading dye and DTT. The samples were denatured at 100 °C for 5 min and then run on the SDS polyacrylamide gel on Hoefer SE260, a mini-vertical gel electrophoresis unit (Amersham Biosciences) at 200 V for 70 min using 1x SDS Running buffer. The protein bands were visualized by staining with Coomassie Brilliant Blue and the background removed by a destainer solution.

2.7. Western blot

2.7.1. Semi-dry transfer

Samples of 3xFLAG tagged protein were analysed via western blot. Samples were normalised (using a Bradford assay or Qubit protein assay) and run on an SDS gel. Six sheets of blotting paper were cut to the desired size and soaked in Semi-dry buffer. The SDS gel was placed down on three sheets of soaked blotting paper then a single sheet of nitrocellulose membrane (Thermo Fisher) was then placed on top

of the gel followed by a further three sheets of soaked blotting paper. The transfer was run at 20 V, 150 mA for 1 h in a Hoefer semiPhor tank before undergoing washes and antibody attachment.

2.7.2. Washes and antibodies

The nitrocellulose membrane with the transferred proteins was incubated in blocking solution for 30 min at room temperature with slow agitation. The membrane was washed, incubating with 2x TBS tween for 10 min at room temperature with slow agitation. It was then incubated overnight at 4 °C in 1/1000 dilution of primary antibody, which was diluted in freshly made blocking solution, with slow agitation. The membrane was again washed in 2x TBS tween twice more at 4 °C. The membrane was then incubated in an appropriate dilution of secondary antibody HRP conjugate, made in fresh blocking solution at 4 °C for 1 h. The membrane was again washed 3 times with 2x TBS tween for 10 min and then subjected to ECL detection.

2.7.3. ECL detection and development

The Amersham™ ECL™ Prime Western Blotting Detection Reagent (GE Healthcare) was used according to the manufacturer's instructions. 1 ml detection reagent and 1 ml of enhancer solution (GE Healthcare) were mixed together, added to the membrane and the detection reagent added. The mixture was carefully pipetted over the membrane for 1 min. The membrane was then removed with forceps and gently touched on the filter paper to drain excess solution. The soaked membrane was then wrapped in Saran wrap and exposed to film for between 20 sec – 5 min depending on the intensity of the bands.

2.8. Chromatin immunoprecipitation (Chip –seq)

The DNA was crosslinked with the protein by adding 37% formaldehyde to the culture making a final concentration of 1%. Incubation was continued as before for 20 min, then 10.5 ml 3 M glycine added and incubated for a further 5 min. Cells were harvested by centrifugation at 6,000 x g for 2 min at 4 °C. The pellet was washed with 25 ml of ice cold PBS and flash frozen by using liquid nitrogen. Using

the cryogenic grinder, cell samples were lysed by running at 30 Hz for 1 min 30 sec for three cycles, cooling in liquid nitrogen between runs. The samples were transferred to a 15 ml tube and 2.2 ml of IP buffer added containing 0.1 mg/ml RiboShredder and a protease inhibitor tablet. Samples were divided into 400 µl aliquots and Bioruptor for 30 sec on 30 sec off for 35 cycles. 12 µl samples were separated by agarose gel electrophoresis to check that fragments ranged in size from 200-500 bp.

Once the size of the fragments was confirmed, samples were aliquoted into 2 ml microfuge tubes and 25 µl of magnetic protein G beads (NEB) added, followed by incubation for 90 min at 4 °C, rotating at 20 rpm. Beads were then immobilised using a magnetic rack and the supernatant removed and discarded. The beads were resuspended in 750 µl of chilled IP buffer and transferred to a new 1.5 ml tube and incubated for 10 min at 4 °C rotating at 20 rpm. After incubating the beads were immobilised again and the supernatant removed. The beads were resuspended in 1 ml of chilled IP salt buffer and incubated for 10 min at 4 °C rotating at 20 rpm. The beads were immobilised again and the supernatant removed. The beads were resuspended in 1 ml of chilled IP buffer with additional salt and incubated for 10 min at 4 °C rotating at 20 rpm. The beads were immobilised again and the supernatant removed. The beads were resuspended in 1 ml chilled TE buffer pH 8 and incubated for 10 min at 4 °C rotating at 20 rpm. The beads were immobilised again and the supernatant removed. The beads were resuspended again in 1 ml chilled TE buffer pH 8 and incubated for 10 min at 4 °C rotating 20 rpm. The beads were resuspended again in 100 µl Elution buffer and 5 µl of Riboshreader added, then incubated for 30 min at room temperature. Samples were incubated at 65 °C overnight to de-crosslink and the beads immobilised and the supernatant containing the immunoprecipitated DNA was transferred to a new tube. This was incubated with 5 µl of 10mg/ml proteinase K at 55 °C for 2 h. Samples were used for qPCR and sent to The Genome Analysis Centre (Norwich) for sequencing, 50 bp single end on the HiSeq 2500.

2.9. Electromobility shift assay (EMSA)

EMSA was used to analyse and detect DNA/protein complexes. DNA was radiolabelled and mixed with protein before being loaded and run on a gel. The detection of complex formation is known by comparing the bands of the individually-run components with the combined components.

2.9.1. Generating radiolabelled DNA probe

Oligos were annealed to generate a probe and then purified from single stranded DNA using the QIAgen gel extraction kit according to the manufacturer's instructions. Probes were radiolabelled, 100 ng probe, 1.11 MBq of γ[^{32}P]-ATP, 1 µl T4 polynucleotide kinase, 2 µl 10 x Kinase buffer, made up to 20 µl with nuclease free H_2O was incubated at 37 °C for 30 min. The probe was purified again by QIAgen PCR purification kit according to the manufacturer's instructions.

2.9.2. Running EMSA gel

Samples were prepared by adding 1 ng of γ-labelled DNA, 2 µl of 5x Binding buffer and 2 µl of 6x loading dye and a specific amount of NAD^+/H as well as the Rex protein or 1µg herring-sperm DNA. This was incubated at room temperature for 20 min and then loaded onto a 6% TBE-polyacrylamide gel and run at 120 V for 1 h 20 min. The gel was then vacuum dried, exposed to a storage phosphor screen, then analysed using a Typhoon phosphorimager (GE Healthcare).

3. Chapter III: Development of tools

Overview

Previously few genetic tools were available to assist in *G. thermoglucosidans* genetic manipulation; an important aim was therefore to develop genetic tools for the engineering of the organism. Until recently the introduction of DNA into *G. thermoglucosidans* was achieved solely by electroporation for which the size of the plasmid limited transformation frequency (Cripps et al., 2009). It is important that vectors with the ability to tolerate large synthetic gene clusters are developed, as has been the case for antibiotic biosynthesis cluster engineering in *Streptomyces* (Fong et al., 2007). Electroporation is also costly due to the need for specialised equipment, whereas conjugation does not (although it is slightly more time consuming). We therefore set out to develop and optimise a conjugative system with improved transformation frequency, especially for low copy, large or non-replicating plasmids.

Conjugation is a method of direct horizontal gene transfer where an origin of transfer (*oriT*) is nicked by a relaxase enzyme and one of the vector DNA strands is transferred into the recipient cell (Lee & Grossman, 2007). Prior to this study, an *oriT* fragment from the plasmid pSET152 was cloned into the MCS of the bifunctional replicating plasmid pUCG18 (M. Jury & M. Paget, personal communication) and conjugation between *E. coli* and *G. thermoglucosidans* TM444 was demonstrated but not properly quantified. The donor strain was *E. coli* S17-1, a strain in which the conjugation apparatus of plasmid RP4 is integrated in the genome. However, this strain methylates DNA and it was reported that *G. kaustophilus* had a methyl-directed restriction system (Suzuki & Yoshida, 2012). Simultaneously, work at TMO Renewables led to the construction of an alternative conjugative plasmid pN109, in which the RK2 *oriT* fragment along with *traJ* from pRK2013 (Kreuzer, Gärtner, Allmansberger, & Hillen, 1989) was inserted into pUCG18 at a position that did not disrupt the MCS, which was therefore more favourable for cloning. pN109 was subsequently used and adapted further to generate a non-replicating plasmid for gene disruptions and fusions. Initial conjugation experiments at TMO Renewables used tri-parental mating protocols to

bypass the methylation system. Tri-parental mating uses three bacterial strains including a helper, a donor and a recipient strain. The helper strain caries the genes required for DNA transfer and conjugation, the donor carries the plasmid for transfer, and the recipient is the bacteria where the plasmid will be introduced into. However this method, although successful, was overcomplicated and time consuming. This chapter describes the development of a single donor strain along with a simplified conjugation protocol that reduces the time involved from three days to two. We also investigated the possible importance of a methyl-directed restriction system in *G. thermoglucosidans* by developing a S17-1 Δ*dcm* donor strain and testing its conjugation frequency against wt S17-1.

This chapter describes the development of strains, vectors and methodology for efficient *E. coli* - *Geobacillus* conjugation. The expansion of the genetic tool kit by the development of a conjugative method will aid academic and industrial research and ultimately will be vital for the commercial success of *G. thermoglucosidans* as a chassis for the production of biocommodities.

This work builds upon an early example of conjugation into a *Geobacillus* strain, demonstrating the efficient conjugation from an *E. coli* donor strain to *G. kaustophilus* (Suzuki & Yoshida, 2012). Since completing this research, efficient conjugation between *E. coli* and *G. thermoglucosidans* was reported (Tominaga, Ohshiro, & Suzuki, 2016). However, our research optimises the protocol, making it far less time consuming. We also present a conjugative module for a range of recently created shuttle vectors.

3.1. Preliminary evidence for *E. coli* to *G. thermoglucosidans* intergeneric conjugation.

3.1.1. *E. coli* strains

The conjugative protocols first produced by TMO Renewables tried using BL21 or JM109 as a donor strain, with HB101 (pRK2013) as a helper strain to conjugate into *G. thermoglucosidans* recipient. JM109 lacks the *E. co*li K restriction system, and BL21 has a non-functional *dcm* due to mutation; the lack of such methylation systems were thought to enhance the success rate of conjugation as DNA was not digested by the recipient. The plasmid pUCG16T in *E. coli* BL21 (*dcm* background) was used as the conjugal donor and *E. coli* HB101 (pRK2013) as the conjugal helper

strain. In this system pRK2013 aided the transfer of pUCG16T from *E. coli* BL21 into *G. thermoglucosidans*. The purpose of this triparental system was to ensure that conjugating DNA was not dcm methylated, thereby bypassing a suspected methyl-directed restriction system. To reduce the complexity of the protocol and to hopefully increase conjugation frequency it was an aim to develop and adapt *E. coli* S17-1 in a biparental method (genotype: Tp^r Sm^r *recA thi pro hsd*$^-$ *hsdM*$^+$ RP4:2-Tc::Mu::Km^r::Tn7 *λpir*). The conjugal transfer functions (RP4) of S17-1 are stably integrated in the chromosome; this ensures that the helper plasmid (pRK2013) is not co-transferred, which might otherwise complicate strain construction. However, S17-1 methylates DNA and so initial studies were performed to see if, despite this, the strain was suitable for conjugation into *G. thermoglucosidans*. Since the growth stage of donor strains has an important effect on conjugation efficiencies, a key benefit of using a single donor strain is that it removes the requirement to coordinate growth of two donor strains and thereby provides more consistent conjugation rates.

3.1.2. Initial biparental mating protocol

A biparental mating protocol was designed where *G. thermoglucosidans* and *E. coli* were grow up as liquid cultures to an OD^{600} of 0.5, and then placed on a nitrocellulose disk via vacuum pump in a ratio of 1:9. The nitrocellulose disk was incubated on plates at 37°C over night to allow for conjugation to occur. The disk was then washed to collect the cells, and the cells plated onto TSA + 12.5 µg/ml Kanamycin and incubated at 60°C overnight to select for ex-conjugates. The described protocol has undergone many alterations and optimisations throughout this project. Our next aim was to streamline and optimise the bi-parental mating protocol, refining the incubation periods, growth media, selection process, and reducing the need for specialist equipment, thereby shortening the time requirements and making the protocol more accessible. The optimised protocol can be found in section 2.4.18.2. The main improvements were at stage 4 and 5 of Figure 3.1 where the use of a nitrocellulose disc was discontinued and its function of allowing transfer of cells from non-selective to selective media was replaced by an antibiotic overlay. The conjugation period was shortened from overnight to ~2

hours and a period prior to antibiotic selection at 60 °C was found to lead to far enhanced ex-conjugate recovery.

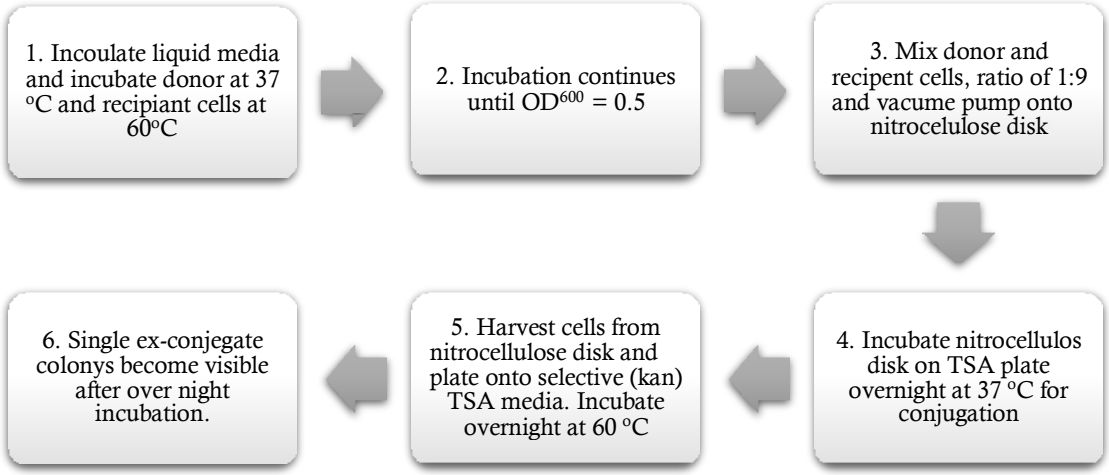


Figure 3.1 Flow diagram of original biparental mating conjugation protocol.

3.2. Growth media

Initially, to optimise conditions for *E. coli* LA growth medium was used during the conjugation. However, no ex-conjugates were recovered, possibly because *G. thermoglucosidans* was unable to survive the combined stress of sub-optimal medium and temperature (data not shown). We therefore used TSA plates for further experiments, which did not have any noticeable effect on the growth of *E. coli* and is an optimal growth medium for *G. thermoglucosidans.* Divalent cations such as magnesium have been shown to improve conjugation in some systems (Yu & Tao, 2010). Experiments were therefore performed to test whether magnesium improved the conjugation. A concentration of 10mM $MgCl_2$ more than doubled transfer frequency and was therefore used in future conjugation experiments (Table 3.1).

Donor Strain	methylase Gene	pN109 transfer frequency (recipient-1)*	
		$MgCl_2$ 0 mM	$MgCl_2$ 10 mM
S17-1	*dam*$^+$, *dcm*$^+$	$(7.1 \pm 9.9) \times 10^{-4}$	$(1.9 \pm 1.8) \times 10^{-3}$
S17-1 Δ*dcm*	*dam*$^+$, *dcm*$^-$	$(1.2 \pm 2.0) \times 10^{-3}$	$(3.0 \pm 3.9) \times 10^{-3}$

Table 3.1 Effect of magnesium supplementation on transfer frequency. *Transfer efficiencies are expressed as the number of kanamycin-resistant transformants per total number of recipients. Results are expressed as the mean ± SD ($MgCl_2$ 0mM n=10, $MgCl_2$ 10mM n=5)

3.3. Investigating the importance of DNA methylation on conjugation frequency

Published work using *Geobacillus kaustophilus* (Suzuki & Yoshida, 2012) suggested that DNA which is methylated by the dam dcm methylation system of a donor strain such as *E. coli*, may be restricted. More specifically it appeared that dcm methylated was particularly restricted, while dam-methylated DNA actually transferred more efficiently (Suzuki & Yoshida, 2012). The donor strain S17-1 has both dam and dcm methylases, therefore our initial aim was to construct an S17-1 dcm mutant; this was achieved using a λ red recombineering approach described in Section 2.4.15 (Gust et al., 2004).

3.3.1. Construction of S17-1 Δ*dcm::apr* donor strain using a REDIRECT approach

REDIRECT is a method of gene disruption using homologous recombination to replace a target gene with an antibiotic resistance cassette (Gust et al., 2004). The *dcm* gene in *E. coli* S17-1 was disrupted by replacing the *dcm* open reading frame with an apramycin resistance gene (*apr*). A disruption cassette consisting of the apramycin resistance gene (apr; aac(3)-IV) flanked by *dcm*-specific coding DNA was amplified from the vector pSET152. Primers were designed to include the entire

aac(3)-IV gene including an upstream region of 130 bp, which was likely to include the promoter. The 58nt primers included the DNA equivalent to the N-terminal and C-terminal 13 amino acids (39 bp) of dcm at their 5' ends and *apr*-specific DNA at their 3' ends. pIJ790 (λred inducible by arabinose, Cm^R) was introduced into S17-1 and the λred system induced. The linear mutant allele was electroporated into S17-1 (pIJ790) selecting for apramycin resistance. Several apramycin resistant (Apr^R) colonies were isolated and screened for the mutant allele by colony PCR (See section 2.2.3 col_S17-1_Dcm_F and col_S17-1Dcm_F '). An *E. coli* S17-1 *dcm* mutant strain was thereby successfully isolated. Colony PCR was performed to distinguish mutants from non-mutants (see Figure 3.2 A); all isolates were found to be mutants. To ensure that the *dcm* methylation system has been disrupted, a ScrF1 restriction enzyme digest profile of pN109 from each of the strains was compared (see Figure 3.2 B). ScrF1 cleavage at restriction sites becomes blocked if DNA is subject to either dam or dcm methylation. The loss of dcm methylation due to the creation of a *E. coli* S17-1 *dcm* mutant strain is indicated, therefore, by an increase/change in the restriction pattern created when plasmid pN109 is subject to ScrF1 digestion.

Testing the effect of dcm methylation on conjugation frequency

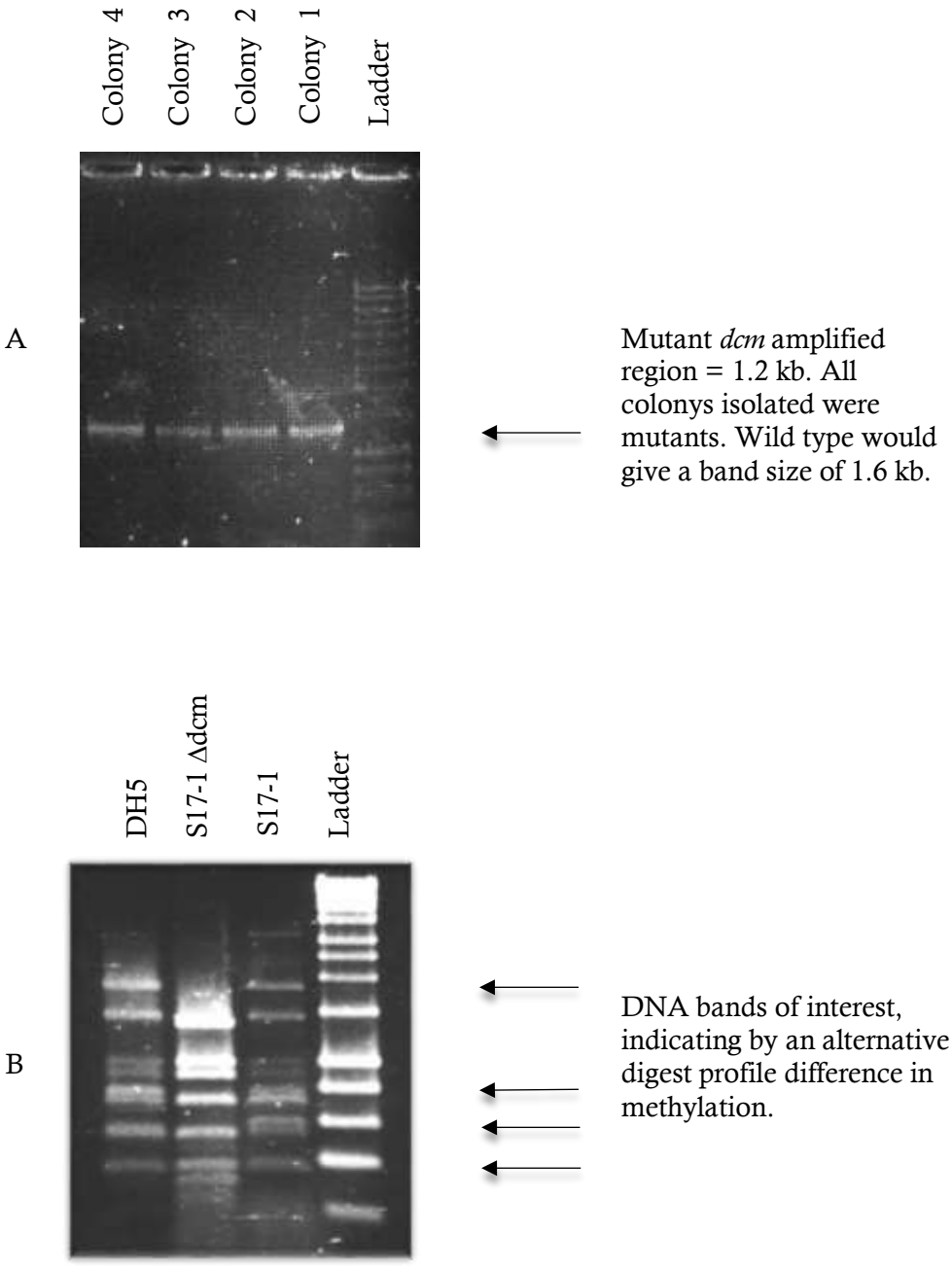


Figure 3.2 Agarose gel electrophoresis showing ScrFI restriction enzyme profile of pN109 isolated from different E. coli strains. A Lane 1-4 isolated dcm mutant colonies that underwent PCR, lane 5: DNA ladder. B Lane 1, DH5α; Lane 2, S17-1 Δ*dcm::apr*; Lane 3: S17-1 dcm+. Lane 4: DNA ladder. Note the appearance of smaller fragments in the *dcm* mutant indicating that dcm methylation was not preventing ScrFI digestion.

To test whether dcm methylation affects conjugation, experiments were performed using *E. coli* S17-1 (pN109) or *E. coli* S17-1 Δ*dcm* (pN109) and *G. thermoglucosidans*. In addition, a non-replicating plasmid that integrates by homologous recombination was tested: pSX700::Δ*rex* is described in Section 4.3 and contains 1.3 kb of homologous DNA either side of a Δ*rex* mutant allele. The transfer frequency of plasmid pN109 from S17-1 Δ*dcm* strain was ~1.7 fold higher than that from S17-1 (see Table 3.2), suggesting that *dcm* methylation might contribute to the degradation of incoming plasmid DNA. However, it should be noted that if *G. thermoglucosidans* possesses a methyl-directed restriction system, but that sites were not present on pN109, we would not expect a difference in frequencies; i.e. other/larger recombinant plasmids might benefit from the use of the Δ*dcm* strain to a higher degree. The non-replicating plasmid pSX700 showed a similar ~3 fold increase in frequency (see Table 3.2), therefore the *dcm* mutant was routinely used in further experiments.

donor Strain	**methylase gene**	**transfer frequency (recipient-1)***	
		pN109	pSX700::Δ*rex*
S17-1 wild type	dam^+, dcm^+	$(7.1 \pm 9.9) \times 10^{-4}$	$(4.9 \pm 2.9) \times 10^{-6}$
S17-1 Δ*dcm*	dam^+, dcm^-	$(1.2 \pm 2.0) \times 10^{-3}$	$(1.5 \pm 1.5) \times 10^{-5}$

Table 3.2 Transfer frequency of both replicating and non-replicating plasmids. *Transfer efficiencies are expressed as the number of kanamycin-resistant transformants per total number of recipients. Results are expressed as the mean ± SD (pN109 n=10, pSX700 n=4)

3.4. Construction of a non-replicating conjugative plasmid

It was anticipated that optimisation of the conjugation protocol would lead to sufficiently high transfer rates to allow the use of a non-replicating plasmid for gene disruption via homologous recombination. The non-replicating plasmid was constructed by deleting a substantial fragment of DBNA from the repBST1 origin of replication. repBST1 is an origin or replication derived from pBST22 and shares a high similarity with the origin of replication from pBM300 (H. H. Liao & Kanikula, 1990). Alongside pN109, repBST1 is also utilised by the *G. thermoglucosidans*

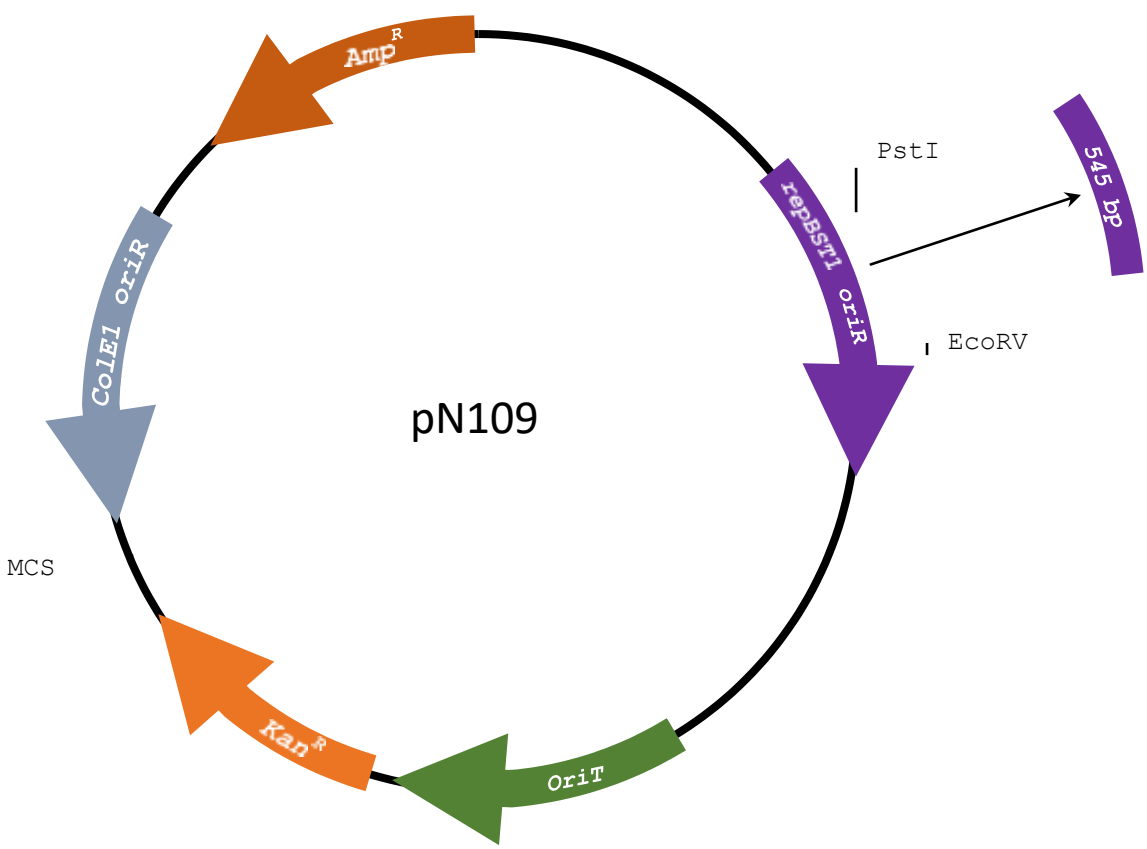


Figure 3.3 The structure of pN109_oriT. The EcoRVand PstI restriction site locations on the repBST1 origin of replication are indicated, used to render the plasmid non-replicating.

Figure 3.4 Non-replicating conjugative plasmid pSX700.

plasmid pUCG18 (Taylor, Esteban, & Leak, 2008). To disrupt the functioning of repBST1 a 545 bp region of the origin of replication was deleted from pN109 by digestion with PstI and EcoRV (see Figure 3.3). The vector was then re-circularised by blunt end ligation and the resulting vector named pSX700 (Figure 3.4).Vector pSX700 was tested to ensure it was unable to replicate in *G. thermoglucosidans*. The non-replicating conjugative pSX700 plasmid was then successfully used to promote integration of a cassette into the genome of *Geobacillus thermoglucosidans* via homologous recombination.

3.5. Development of and optimising a rapid conjugation protocol

3.5.1. Conjugation temperature and time optimisation

The donor and recipient organisms have very different optimal growth temperatures. The *G. thermoglucosidans* optimal growth temperature is 60°C with a tolerance down to 37 °C, whereas the *E. coli* optimal growth temperature is 37°C with a tolerance up to 45 °C. It was therefore no surprise that the duration and temperature of conjugations were key variables that had a dramatic effect on transformation frequency.

Initial experiments indicated that conjugations should be performed at 37°C to ensure *E. coli* can express all necessary transfer functions; no ex-conjugates were recovered when a conjugation temperature of 45 °C was trialled (data not shown). Although at first this incompatibility in temperature was considered a disadvantage it does prove a useful additional selection to eradicate *E. coli* growth later in the protocol, and prevents *G. thermoglucosidans* from growing during the early stages of conjugation, thereby generating siblings.

The length of time for which conjugation was allowed to proceed was also found to be significant (see Table 3.3); optimum frequencies were obtained when conjugation was performed overnight (17 h) at 37°C (Table 3.3). However, shortening this period to 2 h only reduced frequency by a factor of ten. A further

Conjugation duration (hours)	pN109 Transfer frequency (recipient^{-1})*
1	$(2.4 \pm 1.1) \times 10^{-6}$
2	$(1.2 \pm 2.6) \times 10^{-5}$
17	$(3.8 \pm 3.2) \times 10^{-4}$

Table 3.3 Conjugation duration optimisation. *Transfer efficiencies are expressed as the number of kanamycin-resistant transformants per total number of recipients. Results are expressed as the mean ± SD, n=3. All recovery times before kanamycin selection were a duration of 1 h at 60ºC.

reduction of conjugation incubation time to 1 h diminished frequency again by a factor of ten.

The 1 h conjugation period achieves sufficient frequency of 2.4 x 10^{-6} if a replicating plasmid is used, and shortens the protocol time by the largest amount, however a conjugation period 2 h is almost 10 times more effective and is still achievable in one day and so was the preferred choice. Non-replicating plasmids that integrate via homologous recombination require a high conjugation frequency and so 3 h is routinely used (see Table 3.4), and in the case of pSX700::Δrex was found to have an ample frequency of 2.2 x 10^{-6}. When conducting the protocol, it can be expected that ~300 colonies will be found to grow on a single plate after transforming the replicating plasmid, and ~30 for the non-replicating plasmid.

Conjugation duration (hours)	pSX700::Δrex Transfer frequency (recipient-1)*
3	(2.2 ± 8.5) x 10^{-6}
17	(1.4 ± 1.9) x 10^{-5}

Table 3.4 Conjugation duration optimisation of non-replicating plasmid. *Transfer efficiencies are expressed as the number of Kanamycin-resistant transformants per total number of recipients. Results are expressed as the mean ± SD n=3. All recovery times before kanamycin selection were a duration of 1.5 hours at 60°C.

3.5.2. Recovery Period

An important step in any transformation is the recovery period during which the recipient strain is given time to express the antibiotic resistance gene used for selection. The plasmids used in this work contained a kanamycin resistance gene, and selection for ex-conjugates took place at a concentration of 12.5 μg/ml kanamycin. In early protocols before optimisation, cells were plated out and immediately incubated after a period of conjugation on TSA + 12.5 μg/ml kanamycin plates. However, this resulted in a near zero conjugation success rate. A period of recovery was introduced into the protocol to allow the ex-conjugate *G. thermoglucosidans* cells to recover from temperature stress.

In typical bacterial transformation experiments, 1 h at the recipients' optimal growth temperature is usually sufficient before selection is applied. However, in this

case the recipient *G. thermoglucosidans* must not only begin to express the resistance gene, it also must first recover from a stressful incubation period at 37°C. For this reason, we explored and optimised the period of recovery needed.

If the 60 °C recovery duration was increased from 30 mins to 90 mins a 30-40 fold increase in frequency was achieved (see Figure 3.5). To ensure this increase was not due to the overlay redistributing *G. thermoglucosidans* cells after they had undergone division, a control was set up. For the control, mixed donor and recipient cells were plated onto a non-selective medium (note 10^{-3}, 10^{-4} and 10^{-5} cell dilutions were used) and grown at 37 °C for 2 h, then allowed to recover for 30 min at 60 °C. Half the plates were then overlaid with 500 µl of liquid media only (representing a kanamycin overlay), and all plates then incubated at 60 °C overnight. The difference in cell count was measured between those plates which underwent the overlay and those that did not. It was found that the overlay did not have the effect of redistributing cells so as to superficially increase colony numbers. The opposite effect was in fact seen, whereby the overlay disturbed cell growth and reduced cell numbers slightly. We therefore believe the overlay is not contributing to exaggerated ex-conjugate number, but instead reduces overall frequency of the ex-conjugates recovery.

The optimal time for the transfer of pN109 from the donor to the recipient was found to be 2 h, and the recovery time 90 min, giving a transformation frequency of 2.4 x 10^{-5}. These time parameters can be adjusted depending upon the plasmid for transfer and the number of ex-conjugates needed. For example, when using a non-replicating plasmid, or for generating a mutant library, a higher frequency will be required.

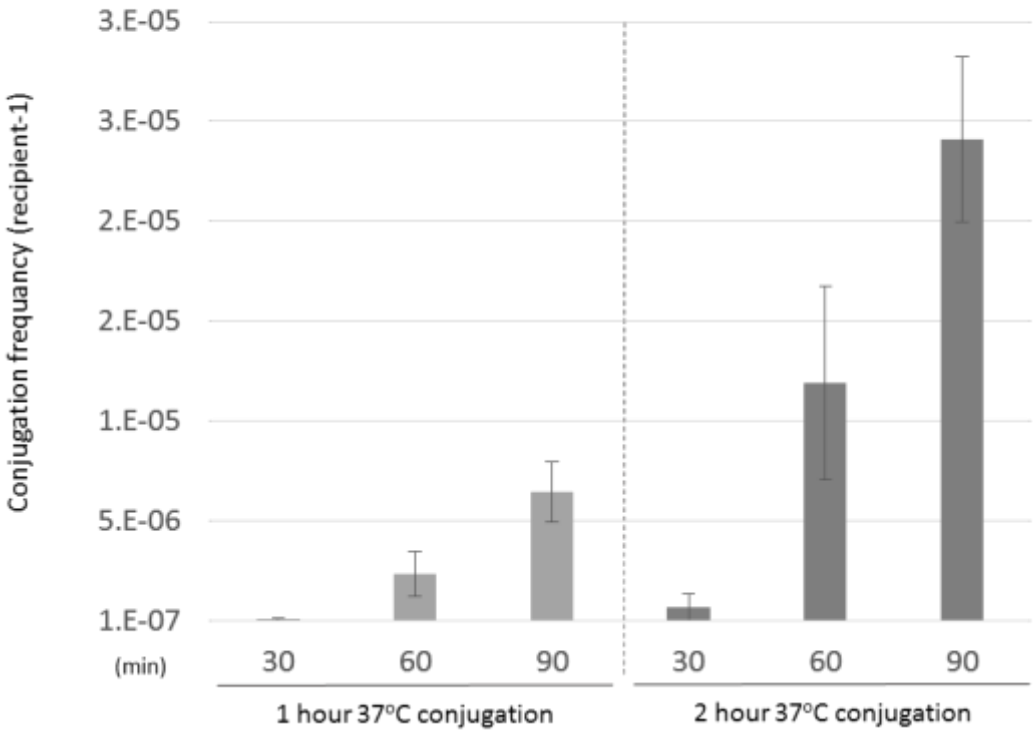


Figure 3.5 Transformation efficiency rates of pN109 with variation in time. Comparison of transformation frequency rates of pN109 S17-1 Δ*dcm* for varying conjugation durations (1 and 2 hours) and 19955 ex-conjugate 60 °C recovery durations (30, 60, 90 minuets).

3.5.3. Description and Comparison of Protocols

The biparental mating protocol was optimised to reduce the time involved and the specialist equipment needed. The recovery of ex-conjugates was optimised and a S17-1 Δ*dcm* strain constructed to circumvent methylation restriction. A non-replicating plasmid (pSX700) was also produced to integrate by homologous

recombination, and was use to create a *Δrex* strain. Figure 3.6 is of an overview of the different protocol stages. It differs from the original conjugation protocol at stages 4, 5 and 6. The introduction of an overlay step is the main addition, which replaced the need for a nitrocellulose disk which was used to transfer cells from non-selective, to selective media.

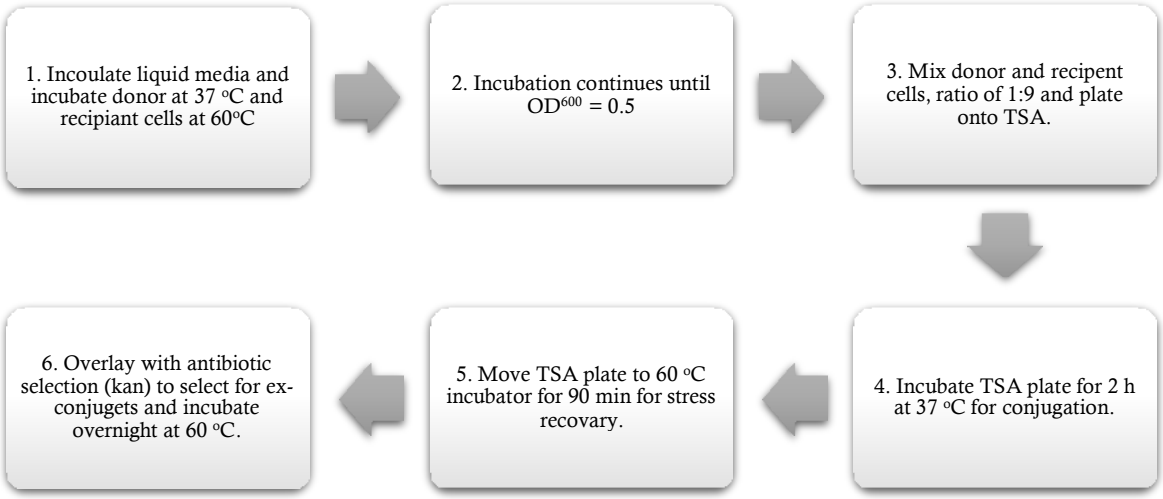


Figure 3.6 Overview of optimised conjugation protocol.

3.6. Adapting SEVA modular shuttle vectors

Recently, a range of *G. thermoglucosidans* - *E.coli* shuttle vectors were created as part of a modular toolkit by the Ellis group (Reeve et al., 2016). The plasmids were based on the 'Standard European Vector Architecture' (SEVA) model of interchangeable parts, allowing deconstruction and reconstruction of plasmids, giving the flexibility to exchange functional parts according to the needs of the user. These vectors contain a variety of reporter genes, including a *gfp,* which is applied in Chapter 6 to assess gene expression. However, the plasmids lack *oriT* and so require electroporation to transform *Geobacillus*. It was therefore decided to incorporate *oriT* into these plasmids thereby permitting high frequency conjugation.

Our aim was to make a modular *oriT* component, designed to be compatible with the existing *G.thermoglucosidans* / *E.coli* shuttle vectors and the current range of parts. The conjugative module was designed to conform, where possible, to the SEVA format (see Figure 3.7). This requires the absence of particular internal restriction enzyme sites (*HindIII, PstI, XbaI, BamHI, SmaI, KpnI, SacI, SalI, EcoRI, SfiI, SphI, AatII, AvrII, PshAI, SwaI, AscI, FseI, PacI, SpeI, SanDI* and *NotI*) along with flanking sites that meet the SEVA format and ensure modular compatibility.

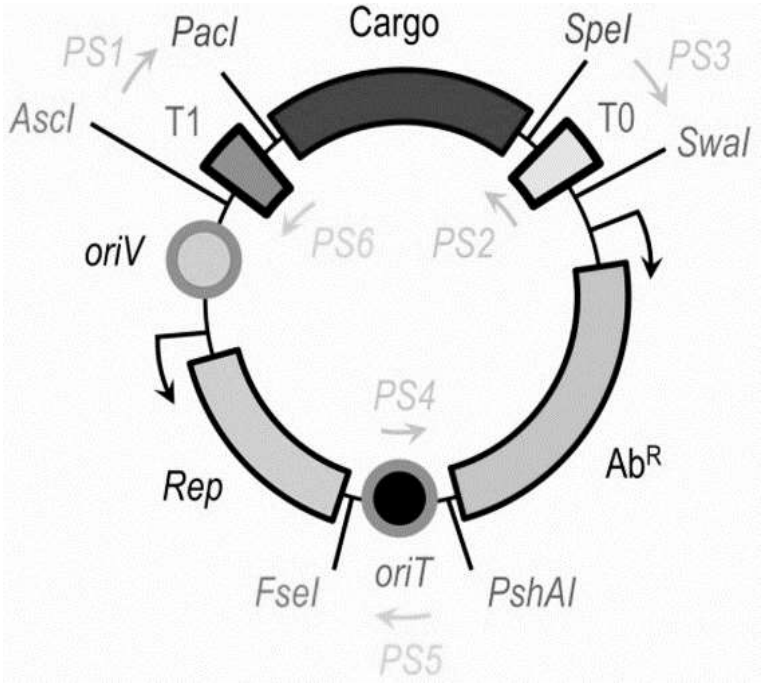


Figure 3.7 The elementary design of a SEVA plasmid. Each functional part is divided by rare restriction sites AscI, PacI, SpeI, SwaI, PshAI and FseI. Terminators T1 and T0 flank the cargo region and aid in plasmid stability. An oriT permits the conjugative transfer of the plasmid between the donor strain E.coli and a recipient (Silva-Rocha et al., 2013).

The conjugative module was based on the broad host rage RK2 family oriT and its sequence synthesised by Life Technologies as a double stranded GeneArt string of 792 bp in length. *EagI* and *NotI* sites were placed at the ends of the fragment to allow it to be subclonned at the *NotI* site of the shuttle vectors, specifically pG1AK-sfGFP (as this vector was to be used for further experiments due to it containing gfp). The oriT double stranded DNA fragment was first cloned into the EcoRV site

of pBlueScript II SK+, then subcloned as a EagI-NotI fragment into the NotI site of pG1AK-sfGFP (which had been digested with *EagI* to give compatible overhangs, see Figure 3.8). When ligated, the *NotI* site would be recreated at only at one end, and a *EagI* would form at the other so to conform to the SEVA format (see Figure 3.9).

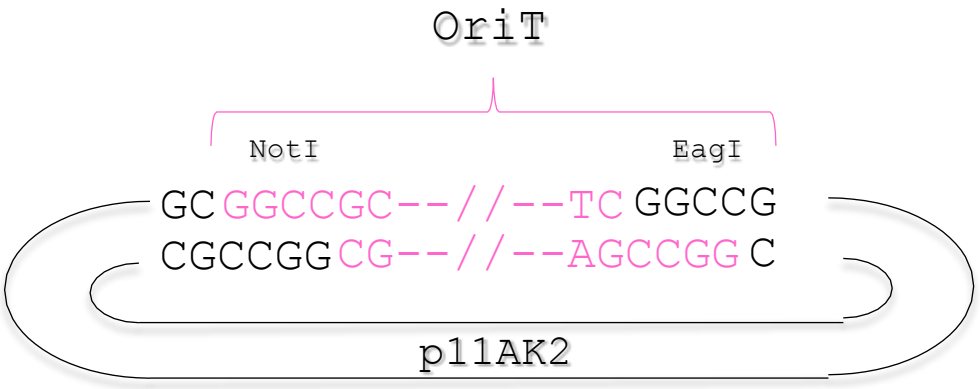


Figure 3.8 Ligation of OriT into p11AK2. The sticky end overhangs after EagI digestion to exercise the oriT fragment, once ligated to the NotI cut vector backbone, resulted in the recreation of NotI and EagI sites. This was designed so to conform to SEVA.

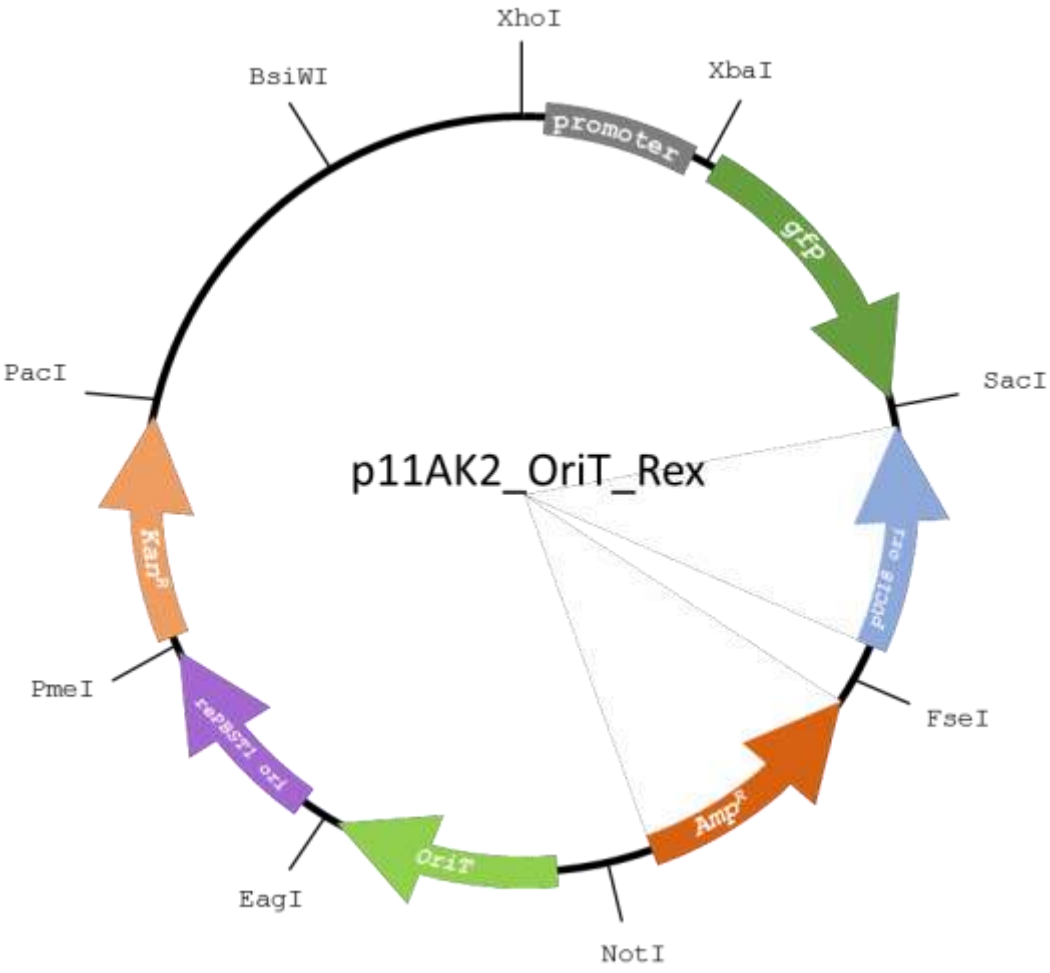


Figure 3.9 Vector map of p11KA2_oriT_Rex. An oriT has been ligated into the p11AK2 NotI site (which had been cut with EagI).

3.6.1. Testing conjugation frequency of pG1AK-sfGFP_oriT

The conjugation frequency of pG1AK-sfGFP_oriT was compared with that of pN109 using a range of incubation times (see Table 3.5). This revealed that the pG1AK-sfGFP_oriT transfer frequency was ~10-fold lower than that of pN109. The reason for this is unclear but is unlikely to be due to differences in size (pN109 is only 298 bp a smaller plasmid than pG1AK-sfGFP_oriT). Other possible explanations are differences in copy number and in host restriction.

Plasmid	**Transfer frequency (recipient-1)***	
	60 mins (recovery duration)	**90 mins** (recovery duration)
pN109	$(2.4 \pm 1.1) \times 10^{-6}$	$(6.5 \pm 1.5) \times 10^{-6}$
pG1AK-sfGFP_oriT	$(2.9 \pm 0.9) \times 10^{-7}$	$(5.1 \pm 1.5) \times 10^{-7}$

Table 3.5 Transfer efficiencies with varying recovery times. Transfer efficiencies are expressed as the number of Kanamycin-resistant transformants per total number of recipients. Results are expressed as the mean ± SD n=3. All conjugation times were a duration of 1 hour at 60°C. pG1AK-sfGFP_oriT = 6435 bp, pN109 = 6137 bp. Donor strain used was S17-1 Δ***dcm***.

3.7. Attempts to develop an integrative conjugative system

The use of chromosomal integration of biosynthetic modules avoids issues associated with multi-copy vectors such as plasmid segregation and structural instability. It also removes the need for maintaining antibiotic selection, which is particularly important for industrial applications. Currently, only homologous recombination has been used to integrate new genes into the *G. thermoglucosidans* genome. This is a laborious two-step process not suited to the screening of multiple constructs. Flanking regions are generated which are homologous to the site of integration. Recombination of this type is an event reliant on probability, its

frequency of occurrence proportional to the length of flanking regions and the recombinogenic tendency of the specific genome region (Marshall Stark et al., 1992). Therefore a site-specific DNA integration system to rapidly engineer *G. thermoglucosidans* is of particular interest and was partly developed.

Site-specific recombination relies upon DNA-binding recombinases, and is far more efficient than homologous recombination (Lyznik et al., 2003). It evolved as a mechanism by which mobile elements including viral DNA could integrate into host genomes (Groth and Calos, 2004). The bacteriophage φC31 is a natural virus of *Streptomyces coelicolor*; upon infection the circular DNA integrates into a specific small site in the genome via site-specific integration. The natural φC31 site-specific recombination system has been adapted for the generation of several integrative vectors in *Streptomyces* (Bierman et al., 1992). Furthermore, it has been adapted for use in higher organisms including fly and mammalian cells (Bateman *et al.*, 2006, Marshall Stark et al., 1992, Sadowski, 1986). The system consists of three main components: integrase enzyme, attachment phage (*attP*) and attachment bacterial (*attB*) sites. The integrase enzyme cuts DNA at specific *att* sites and initiates recombination: one attachment site on the host genome, and the other on the plasmid. These are brought into close proximity by protein-protein-interactions mediated by the integrase enzyme, cleavage occurs of the plasmid backbone and genome after which strands undergo a strand exchange reaction and ligate together (Coates et al., 2005, Fogg et al., 2014).

In this work, it was aimed to generate a conjugative-integrative system based on φC31 for use with *G. thermoglucosidans.* Bioinformatic searches showed no natural attB attachment site present in the *G. thermoglucosidans* genome, which therefore necessitated the artificial engineering of a site. An integrative vector containing the φC31integrase (*int*) and *attB* sites was constructed, the vector backbone based upon the non-replicating conjugative plasmid pSX700, allowing manipulation in *E. coli,* and transfer by conjugation. It was aimed to integrate the attP site in the host genome rather than attB, thereby preventing the (unlikely) infection of the strain with a related bacteriophage.

3.7.1. Integrative Vector

The integrative plasmid includes φC31 system parts cloned into pSX700 for *E. coli* to *G. thermoglucosidans* conjugation. The necessary parts were cloned into pSX700 in two stages. Firstly the minimal attB integration site (70bp) was determined using the literature (Bateman & Wu, 2008) and synthesised as a single fragment including the chosen promoter for the integrase gene (pup3 – a constitutive promoter based on pUP constitutively expressed uracil phosphoribosyl transferase gene of G. thermoglucosidans and has been found to be functional in both *E. coli* and *G. thermoglucosidans* (Reeve et al 2016)) and homologus flanking regions for Gibson assembly into pSX700 (see Figure 3.10). XhoI and SpeI restriction sites were placed in between the promoter and the attB site so that the integrase gene could be ligated in at a later stage. The fragment was produced by Life Technologies, as a double stranded GeneArt string and was directly Gibson assembled into NotI cut pSX700 to create pSX700_attB. The second stage was to PCR amplify the φC31 integrase gene (int) from pSET152 using primers (see section 2.2.3 Int_F and Int_R) with restriction sites XhoI on the forward, and SpeI on the reverse for later ligation. The reverse primer was designed to also contain a terminator sequence. Although the φC31 integrase gene is unlikely to be thermostable at 60°C this is unimportant since conjugation with *E. coli* is carried out at 37°C. After amplification the fragment was purified and cloned into EcoRV site of pBlueScript, sequenced to check for errors (one silent mutation was found), and cut out as an XhoI and SpeI fragment. It was then ligated into pSX700_attB to create pSX700_attB_*int* (Figure 3.11). Conjugation experiments set up with S17-1 *dcm* (pSX700_attB_*int*) and *G. thermoglucosidans* 11955 resulted in no ex-conjugants, as expected since a genomic attP site is not present in the genome.

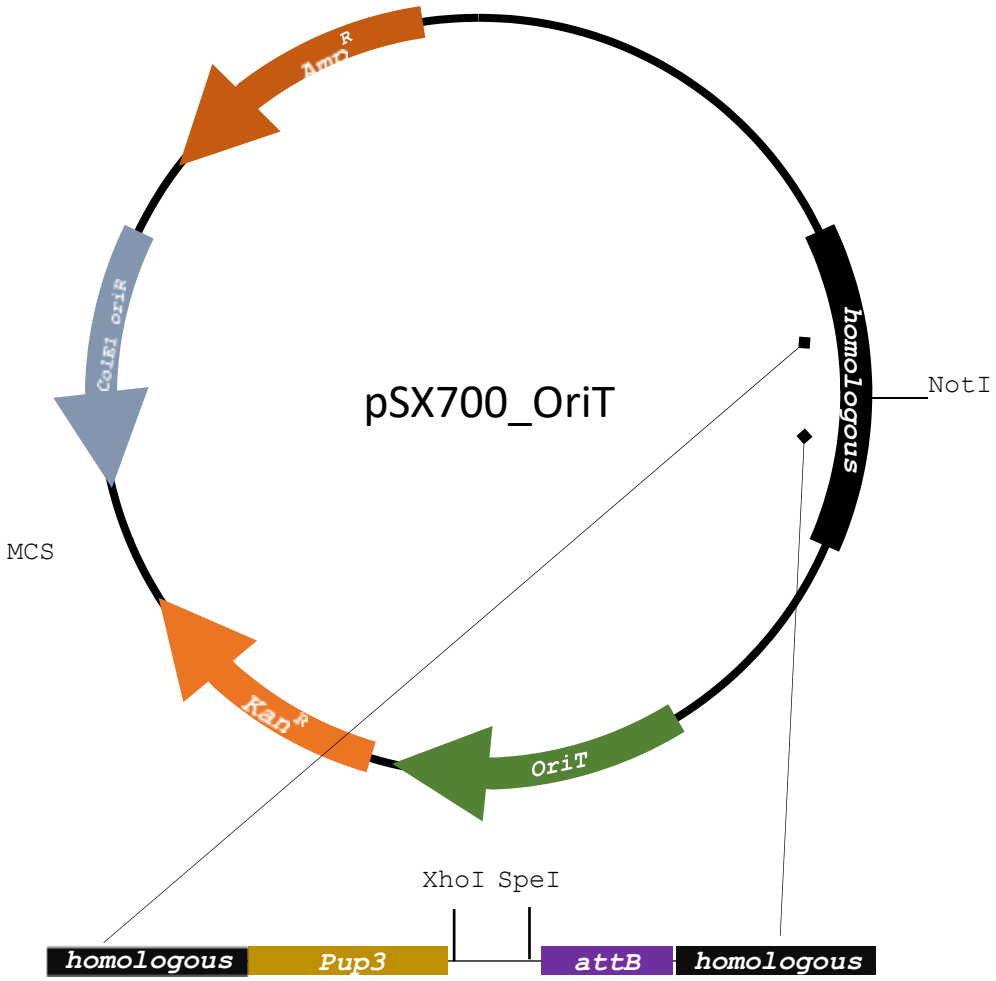


Figure 3.10 Vector map of the pSX700 and linier synthesised attB fragment. The NotI site where the pSX700 vector was linearized for a synthesised fragment to be Gibson assembled in, is indicated. The synthesised fragment contains: 40 bp pSX700 homologues regions at either end; a pup3 constitutive promoter for integrase expression; XhoI and SpeI sites for ligation of the PCR amplified integrase gene; a 70 bp attB site.

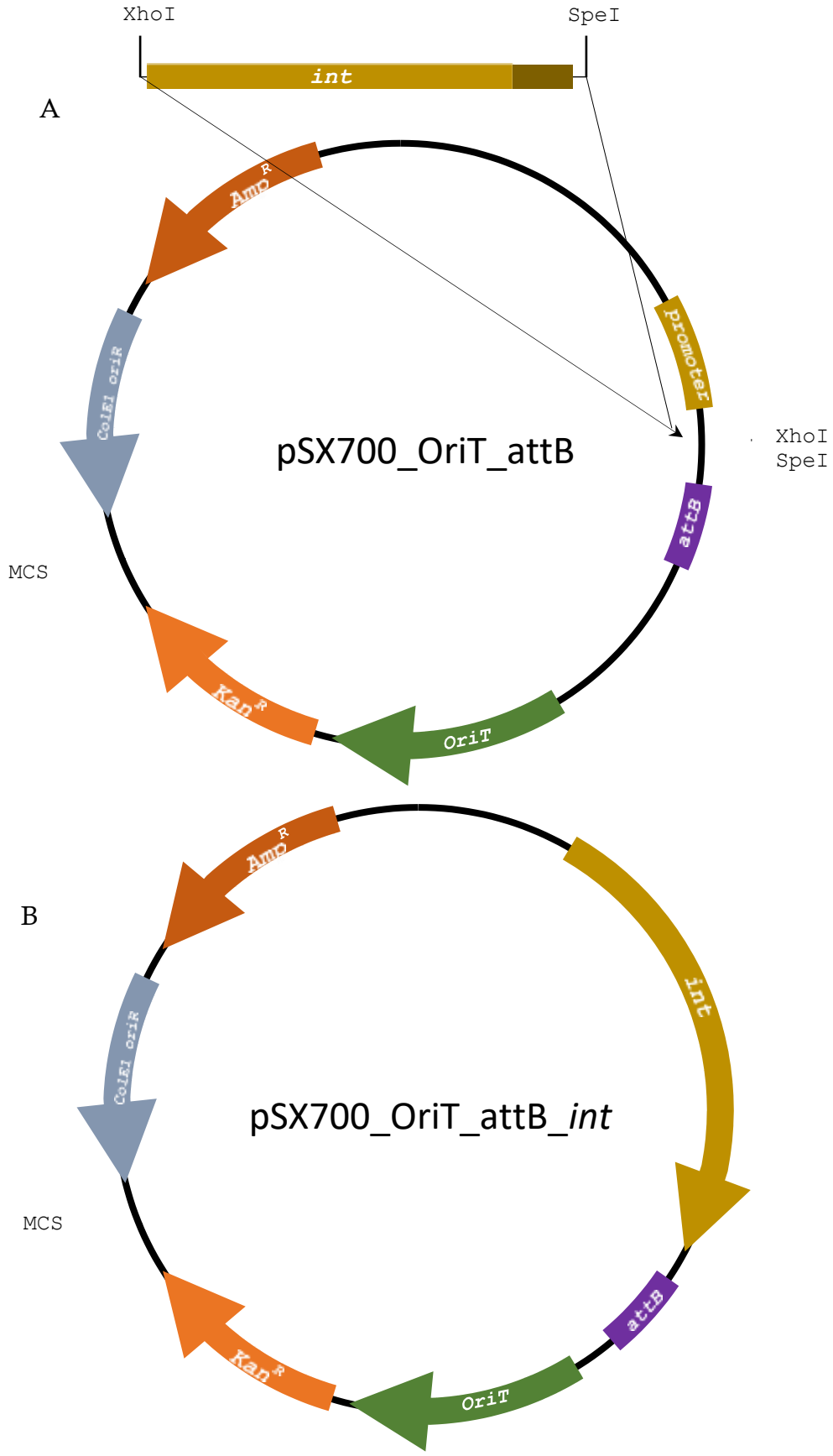


Figure 3.11 Integrative vector construction. A) Vector map of pSX700 and the int fragment (dull yellow, the terminator a darker yellow) to be ligated into the XhoI and SpeI restriction sites. B) Final map of the integrative vector pSX700_attB_int.

3.7.2. Genomic attachment site

During this study several alternative genomic locations for the attP attachment site were chosen, and attempts made in each case to generate a stable recombinant strain; however, none were successful. The genomic sites selected were a mutant *ldh* gene, and two genes that provide the possibility of efficient counter selection: *pyrE* and *upp*.

3.7.2.1. The orotate phosphoribosyltransferase (pyrE) gene as an attachment site

In order to rapidly generate strains that have an integrated *attP* site, it would be beneficial to place the attachment site in a gene, the disruption of which could be selected for (counter selection). The orotate phosphoribosyltransferase (*pyrE*) gene has been used as counter selectable marker in many organisms including in yeast, *Bacillus* and *Clostridium* (Boeke, La Croute, & Fink, 1984; Haas, Cregg, & Gleeson, 1990; Suzuki, Murakami, & Yoshida, 2012; Tripathi et al., 2010). The *pyrE* gene is part of an operon integral for the pyrimidine biosynthetic pathway, and it catalyses one of the last steps in this process and disruption of this gene prevents the conversion of orotic acid to UMP and produces uracil auxotrophs (Dong & Zhang, 2014). Disruption of *pyrE* expression also prevents the conversion of 5-fluoroorotic acid (5-FOA) into toxic 5-fluoro-UMP. Therefore the replacement of *pyrE* with *attP* by homologous recombination should result in 5-FOA resistance. However, as *pyrE* is essential for growth on pyrimidine-deficient medium, uracil is incorporated into 5-FOA selection media.

3.7.2.1.1. pyrE gene disruption

The *pyrE* is located within an operon with orotidine 5'-phosphate decarboxylase (*pyrF*) and dihydroorotate dehydrogenase (*pyrD*) (see Figure 3.12). A 2.02 kb fragment of DNA including the centrally located *pyrE* gene was amplified by PCR using *G. thermoglucosidans* chromosomal DNA as template and primers that incorporated flanking BamHI sites. Following cloning into pBlueScript, inverse PCR was performed to delete 444 bp of *pyrE* coding region, replacing it with the 60 bp *attP* site. The pyrE_att construct was subcloned into pSX700 and transformed into S17-1 Δdcm. pSX700_ pyrE_att (see figure 3.13) was then conjugated into *G. thermoglucosidans* and ex-conjugants were selected using kanamycin, thereby forcing genomic integration by homologous recombination. Colonies were picked, purified as kanamycin resistant clones, grown in TSB to allow time for the second crossover event to occur, then plated to minimal media containing 50 mg/l 5-FOA, supplemented with 10 mg/l uracil. This was predicted to select for the second crossover event forcing the loss of the wild-type *pyrE* gene and the generation of a stable mutant allele that includes a central *attP* site. Cell dilutions were also plated onto TSA only, and TSA with kanamycin; this was done to determine the phenotypic makeup of the liquid culture. The results indicated (see Table 3.6) that few second crossovers arose during the non-selective liquid incubation period. The 7 5-FOA resistant colonies were streaked onto TSA with kanamycin selection, MM with 5-FOA and uracil, and TSA with 5-FOA. Of the seven potential pyre mutant colonies only two were found to be both kanamycin sensitive and 5-FOA resistant. As a control measure WT *G. thermoglucosidans* was also streaked alongside the potential *pyrE* mutants.

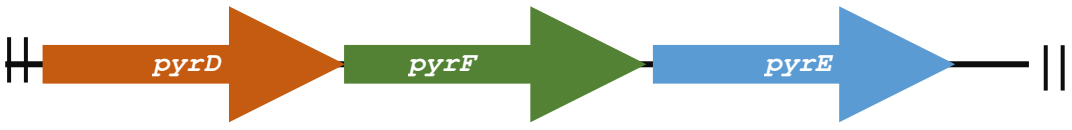


Figure 3.12 The genes surrounding pyre. The *pryE* gene encodes an orotate phosphoribosyltransferase, *pyrF* encodes orotidine 5'-phosphate decarboxylase and *pyrD* encodes a dihydroorotate dehydrogenase. Genes overlap one another and exist in the same operon.

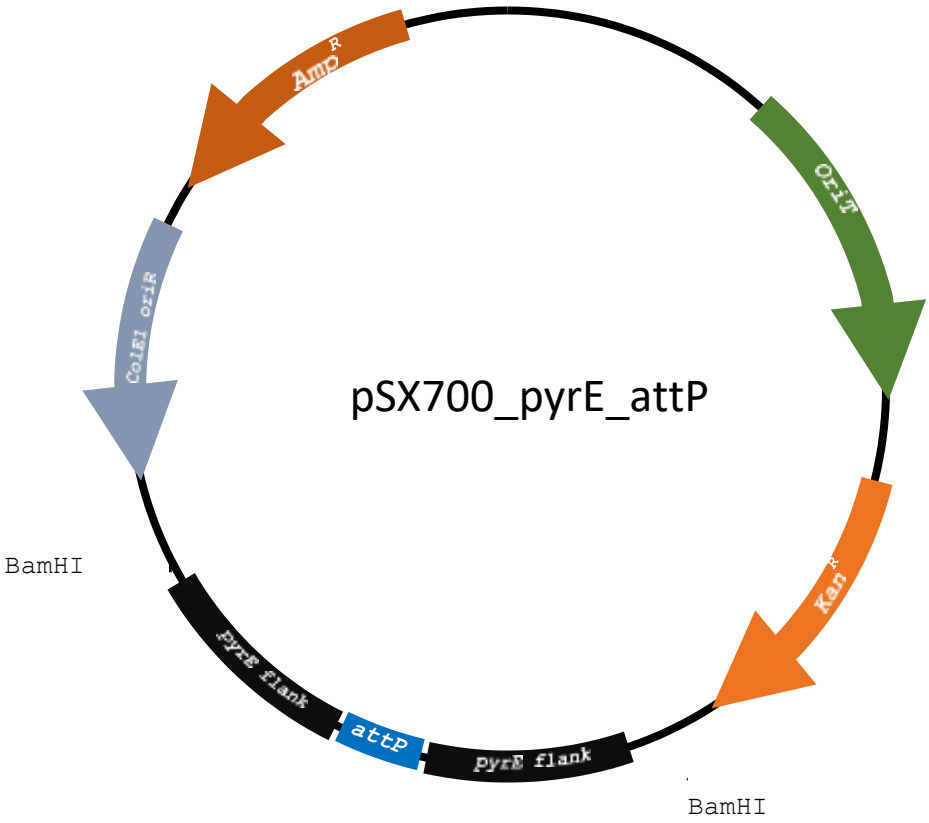


Figure 3.13 Vector map pSX700_pyrE_attP. pyrE flanks were ligated into the BamHI site of pSX700. The attP site was constructed by inverse PCR and is situated in-between the *pyrE* flanks so to be inserted into the genome of G. *thermoglucosidans* via homologous recombination.

Dilutions of cells grown in TSB 60°C with no selection & no Uracil addition.	**TSA only**	**TSA + Kan**	**MM + 5-FOA + Uracil**
10^{-2} dilution	10,000 +	10,000+	7
10^{-4} dilution	1384	1104	0
10^{-5} dilution	352	83	0

Table 3.6 Serial dilution of TSB cell coulters grown from first crossovers for phenotypic analysis.

***G. thermoglucosidasius* strain type**	**TSA**	**MM + FOA**	**MM + U**	**TSA + Kan**	**MM + FOA + U**
Δ*pyrE*::attP	++	-	+	-	+
11955	++	-	+	-	++
11955 first crossover	++	-	+	+	++

Table 3.7 Comparison of G. thermoglucosidasius strain phenotypes. Vigorous growth is indicated as ++, moderate +, and no growth as - .

Unexpectedly the WT strain was able to grow on 5-FOA selective media (see Table 3.7). The results indicated the phenotypes of both the potential *pyrE* mutants and WT strains was indistinguishable. The potential *pyrE* mutants were tested via colony PCR, the results indicated that no mutants were isolated (results not shown). Several attempts were made to isolate a potential *pyrE* mutant using variations of the method previously outlined. A range of 5-FOA concentration were tested (1-200 μg/ml) and uracil supplementation (1-10 μg/ml), despite this we were unable to differentiate between WT and potential *pyrE* mutants using phenotypic analysis. Literature suggests this may be due to uracil supplementation and its inhibitory effect on the expression of the operon containing *pyrE*. A regulator of the pyrimidine biosynthetic genes, pyrR was found, able to sense UMP and UTP and inhibit gene expression in *Bacillus* (Dong & Zhang, 2014). UMP is produced from uracil by the protein encoded in *upp* gene. PyrR works as a negative regulator and is a mRNA-binding attenuator (Turner, Bonner, Grabner, & Switzer, 1998). A bioinformatics search revealed a *pyrR* homologue in *G. thermoglucosidans* with a 73% identity.

Our results indicated that counter selection was not possible as the necessary uracil enriched media caused the *pyr* operon to be inhibited and WT strains were therefore no longer susceptible to 5-FOA selection. Compounding this issue was that any *pyrE* mutant strain was likely slower growing than the WT. For these reasons the

counter selection system failed and in fact *pyrE* mutants/ genomic *attP* site integration was more difficult to isolate than they would have been otherwise by mutant screening and colony PCR.

3.7.2.2. The uracil phosphoribosyl-transferase gene (*upp*) as an attachment site

Based on successful work in *Bacillus subtilis* (Fabret, Ehrlich, & Noirot, 2002) it was decided that the attP site would be placed into the uracil phosphoribosyl-transferase gene (*upp*) of *G. thermoglucosidans*. The uracil-phosphoribosyl-transferase (UPRTase) enzyme converts uracil into UMP, allowing the cell to use exogenous uracil (Neuhard, 1983; Nygaard, 1993). It also converts the pyrimidine analogue 5-fluorouracil (5-FU) to a toxic product and can therefore be used as a counter selectable marker. The *upp* gene is the only gene present within the operon.

3.7.2.2.1. *upp* gene disruption

Two 980 bp flanks were designed based on the *upp* gene and PCR amplified from *G. thermoglucosidans* using primers of 38 nucleotides in length, designed with a 20 nucleotide overlap with the next fragment in the Gibson assembly (See Section 2.2.3 UPP-LFlank_Fwd, UPP-LFlank_Rev , UPP-RFlank_Fwd , UPP-RFlank_Rev). The attP site was PCR amplified as a 300bp fragment, using pSET152 as a template using primers designed for downstream Gibson assembly (See Section 2.2.3, attP_pSET1_Fwd, attP_pSET1_Rev). In total 4 fragments were assembled in a single reaction including the pSX700 (BamHI cut) backbone, both the left and right upp flanks and the attP site amplified from pSET152 (see Figure 3.14).

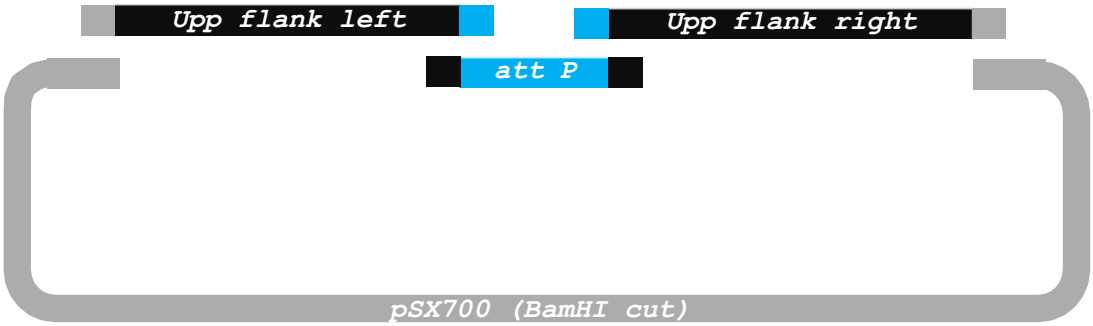


Figure 3.14 representation of the overlap of fragments to be assembled via a signal Gibson reaction. Overlaps were a minimum of 20 nt in length.

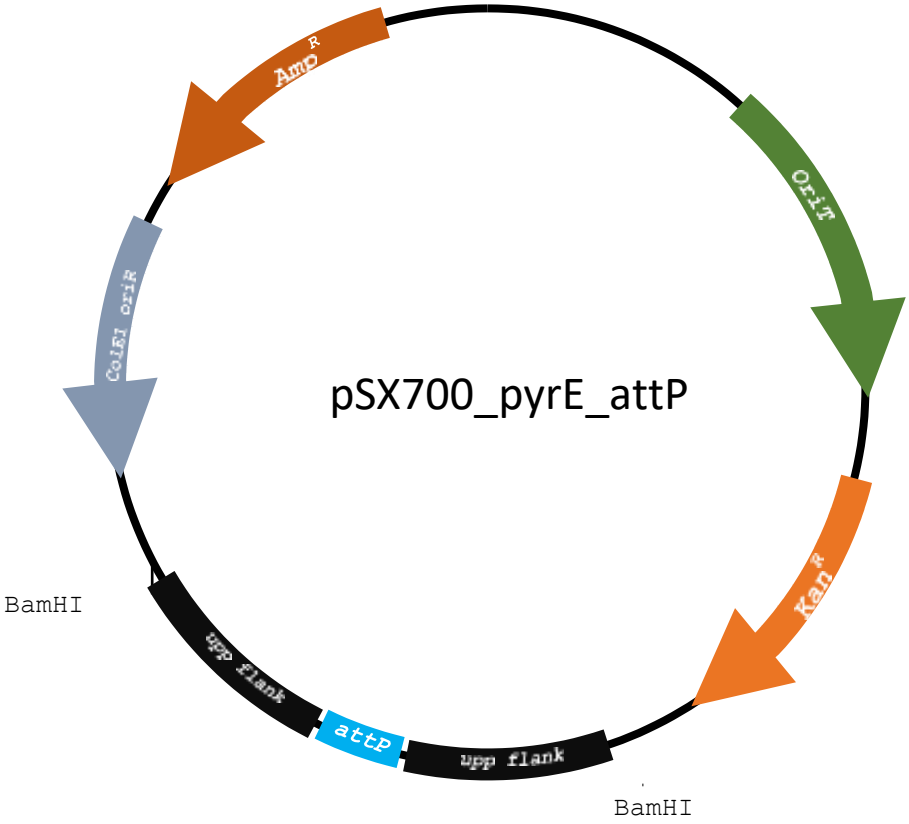


Figure 3.15 Vector map pSX700_upp_attP. The attP site, along with upp flanks were Gibson assembled into the BamHI site of pSX700.

A *upp::attP* allele was generated by using Gibson assembly cloning the fragment into the BamHI site of pSX700 (Figure 3.15). The assembly was checked via diagnostic digest before transforming into S17-1 Δ*dcm* strain and conjugating into *G. thermoglucosidans*. First crossover recombinants were selected for by kanamycin resistance and then subcultured in non-selective TSB media, giving time for the second crossover event to occur. The culture was plated to TSA containing 100 µg/ml 5-FU to select for the second crossover. The time taken for colonies to grow to a visible size was 48 h compared to the usual ~6 h for a WT strain. Three Kan^S 5-FU^R colonies were isolated after screening for kanamyacin sensitivity. The Kan^S 5-FU^R colonies were identified as being possible mutants, and one was confirmed by colony PCR. Despite attempts made to conjugate into the *G. thermoglucosidans* Δ*upp*/attP strain using the vector pSX700_attB to test if the integrative system was functional, no ex-conjugates were isolated. This may be due to the weakness of the *G. thermoglucosidans* Δ*upp*/attP strain which takes 48 h to grow to a visible size and may not cope with the temperature (37 °C) at which conjugation occurs. The negative result could also be due to the integrative system being non-functional.

3.8. Discussion and future work

The genetic engineering tools available for use with *G. thermoglucosidans* have, during this project, been expanded. The development of a conjugative system for use in *G. thermoglucosidans* represents a major addition to the genetic tool kit, enabling reliable and high frequency transformation of replicating and non-replicating vectors. The optimisation of the rapid conjugation protocol has also made this form of transformation advantageous when compared to electroporation in terms of ease and time duration. It also eliminates the need for specialist equipment. A number of vectors are now available for use with this protocol, including a non-replicating vector and an adapted modular shuttle vector component, compatible with a range of plasmids adopted by several research groups (Reeve et al., 2016).

The production and use of a S17-1 Δ*dcm*::*apr* donor strain did improve conjugation frequency. Previously published work using *G. kaustophilus* demonstrated a 10 fold improvement in transformation efficiency when using an *E. coli Δdcm* donor strain

(Suzuki & Yoshida, 2012). In comparison a lesser improvement of between 1.7-3 fold was recorded for *G. thermoglucosidans* using a S17-1 Δ*dcm*::*apr* donor strain (depending on the DNA conjugated). Whilst work here demonstrates the importance of DNA methylation on transformation frequency, investigations into the exact nature of this could be carried out. Perhaps an alternative methylation system to the *dcm* also plays a role and investigation of this could help to enhance transformation frequency further.

Temperature during the conjugation protocol was found to play a key role in a successful outcome. *E. coli* and *G. thermoglucosidans* optimal growth temperatures differ substantially. By regulating growth temperature carefully, switching between 37 °C and 60 °C, optimising each step of the protocol, what at first seemed a major hindrance became an advantage. A balance was struck between allowing for a long enough incubation period at 37 °C for conjugation to occur, but not too long as to negatively affect *G. thermoglucosidans.* A recovery period for *G. thermoglucosidans* at 60 °C was introduced, before the antibiotic selection was applied. By performing the conjugation at 37 °C, optimal for *E. coli,* the overgrowth of *G. thermoglucosidans* was halted, allowing for an accurate measurement of transformation frequency. The switch to *G. thermoglucosidans's* optimal temperature of 60 °C after conjugation killed the *E. coli* cells and acted as a selection method. It was found that pre-warming of all equipment and media was key to successful conjugation, as *G. thermoglucosidans* was especially sensitive to cold shock, even at room temperature.

One of the many optimisations made to the conjugation protocol was the addition of an antibiotic overlay, which replaced the need for nitrocellulose disks. Nitrocellulose disks were used in early protocols to aid the transfer of cells from TSA plates to TAS + kanamycin selection plates. Disks were removed from the TSA plate and cells washed off, then resuspended and plated onto selective media. During this wash step the cells were disturbed which led to a reduction in ex-conjugate recovery. Furthermore, this method required specialised equipment to place the cells on the nitrocellulose disk. It also exposed *G. thermoglucosidans* to room temperature (known to be harmful, resulting in cell death) whilst the culture was vacuum pulled through the discs, trapping the cells on the nitrocellulose matrix. To eliminate the need for nitrocellulose disks, an overlay step was added to the protocol. After conjugation, instead of transferring cells by the method outlined

above, the TSA plates containing ex-conjugates (without a nitrocellulose disk) were overlaid with a kanamycin-TSB dilution. This gave good results, was quicker to perform and was less expensive.

The work we have carried out to develop an integrative conjugative system will hopefully be of use to guide future research. It is as yet unclear if the integrative system is viable in *G. thermoglucosidans.* The integrative vector remains to be fully tested for its capability to express the integrase machinery at the correct time point. The greatest task remains the engineering of an attachment site into the genome of *G. thermoglucosidans.* This could be achieved by simple homologous recombination and screening methods for the loss of antibiotic resistance.

It is likely that the creation of a *G. thermoglucosidans* Δ*pyrE* strain and its use as a counter selectable marker may still be possible despite the limitations of this project. If the *pyrR* gene was to be inactivated by deletion it would no longer be able to act as a negative regulator of *pyrE* expression (by sensing UMP and UTP levels which were elevated by the enrichment of media with uracil) and it might become possible to differentiate the phenotype of WT and Δ*pyrE G. thermoglucosidans* strains.

4. Chapter IV: Analysing the Rex regulon

Overview

Rex is a transcriptional repressor that senses cellular NAD^+/NADH levels and regulates specific genes, often key to fermentation and anaerobic respiration, so to maintain redox balance. During periods of environmental oxygen abundance, aerobic respiration functions as the main process to produce energy, and NADH is constantly recycled to NAD^+, maintaining redox balance. During oxygen limited growth conditions however, NADH builds up, as the terminal electron acceptor, oxygen, is scarce. Just as the lack of oxygen limits the electron transport chain and ATP synthesis, the lack of NAD^+ becomes a rate limiting step for the glycolysis pathway. To cope with these conditions bacteria often switch to nitrate as an alternative terminal electron acceptor. Soil-dwelling bacteria such as *Geobacillus thermoglucosidans* often encounter oxygen limited environments and are able to respire anaerobically, or switch to fermentation if no terminal electron acceptor is available (Bueno, Mesa, Bedmar, Richardson, & Delgado, 2012). The Rex regulon has previously been described in a wide variety of Gram positive bacterium and found to be a regulator of genes involved in fermentation, glycolysis, NAD biosynthesis, and general energy metabolism, as well as some miscellaneous and uncharacterised genes (D. A. Ravcheev et al., 2012).

Rex is of particular interest during this project due to its key role in regulating fermentative genes. Understanding and utilising Rex to optimise the industrial potential of *G. thermoglucosidans* is an important aim. To gain insight into the Rex regulon a 3xFLAG tagged Rex strain was constructed for use with ChIP-sequencing, a technique able to map all Rex binding sites within the *G. thermoglucosidans* genome. In addition to this, a *rex* mutant strain was constructed and analysed for effects of deregulation. The Δ*rex* strain was predicted to experience major loss in the control of fermentation genes. Results from the ChIP-seq were used to validate potential members of the Rex regulon using qRT-PCR experiments. This information has allowed us to gain a sophisticated knowledge of the Rex regulon. qRT-PCR data suggested that other regulators might also act on Rex targets, producing unexpected variations in transcript abundance, in response

to changes in oxygen availability. Fermentative genes such as *adhE* (alcohol dehydrogenase) were shown to be deregulated in a *Δrex* strain, although there was surprisingly small changes in overall fermentation products, as shown by HPLC.

4.1. The genetic context of *rex*

G. thermoglucosidans rex is a 639 bp gene located within a predicted operon consisting of four genes (see Figure 4.1). *rex* is the second gene of the operon. Upstream of *rex* is the first gene of the operon (ATO13_04535; gene numbering for *G. thermoglucosidans* DSM2542 is used throughout this thesis) which encodes a molybdenum cofactor biosynthesis protein C. Molybdenum cofactor is utilised by many enzymes to catalyze redox reactions and is found universally in bacteria, archaea and the eukaryotic domains of life (Bittner & Mendel, 2010). The genes downstream of *rex* encode preprotein translocase subunits (AOT13_04545 and AOT13_04550), part of a sec-independent selective twin-arginine translocation (Tat) system (Berks, Sargent, & Palmer, 2000). The Tat system has been identified as commonly transporting redox enzymes (Weiner et al., 1998). The *rex* gene in *B. subtilis* is also arranged within a similar genomic context, with TatAy and TatCy positioned downstream of *rex*. The TatAy – TatCy pathway is dissimilar from other systems, such as the TatAd–TatCd pathway in *E. coli* (Goosens, Monteferrante, & van Dijl, 2014). The TatAd–TatCd pathway was capable of translocating the TatAy–TatCy-dependent substrate EfeB, however the TatAy – TatCy pathway was not capable of translocating the TatAd–TatCd dependent PhoD (Eijlander, Jongbloed, & Kuipers, 2009). Differences were also found in expression profiles: TatAd and TatCd are expressed only in low phosphate conditions whereas TatAy and TatCy were constitutively expressed in low phosphate conditions (Goosens et al., 2014).

Figure 4.1 Organisation of the predicted *G. thermoglucosidans* DSM2542 *rex* (AOT13_04540) operon. *rex* is the second gene in the four-gene operon. The start codon of rex overlaps with the last three codons of the first gene, molybdenum cofactor biosynthesis protein C (*moaC*) suggesting translational coupling. The gene downstream, preprotein translocase subunit TatA (*tatA*) is separated by 15 bp. Preprotein translocase subunit TatC (*tatC*) is 196 bp downstream of *tatA*.

4.2. Identification of Rex targets by ChIP-seq analysis

To fully understand the biological role of Rex we used chromatin immunoprecipitation assay in combination with a DNA sequencing (ChIP-Sequencing), an approach that could globally map all Rex target binding sites within the genome of *G. thermoglucosidans*. This revealed the potential regulon and also helped identify Rex consensus binding motif using ChIP-Seq enriched DNA regions.

The ChIP-Sequencing technique involves chemically crosslinking DNA binding proteins to chromosomal DNA in vivo, and then selectively immunoprecipitating the protein of interest — along with the bound DNA — using a specific antibody. The co-immunoprecipitated DNA is then sequenced and mapped to the genome revealing binding sites and a qualitative view of relative occupancy. An epitope tag can be used to modify the DNA binding protein if no poly- and mono-clonal antibodies against the protein are available. In this instance the addition of a 3xFLAG tag to the C-terminus of Rex was used to allow immunoprecipitation with an anti-FLAG antibody.

4.2.1. Construction of a 3xFLAG tagged rex strain

The addition of a triple FLAG tag to the terminus of Rex was achieved in a number of stages that included cloning steps for assembly and then subsequent integration

into the genome to replace the WT *rex* gene. Firstly, the *rex* gene and surrounding region was amplified from chromosomal DNA using (Grex_reg_F and Grex_reg_R) primers containing BamHI restriction sites (Figure 4.2). The resulting fragment contained the entire coding region of rex, including a left flank of 636bp and a right flank of 714bp. This fragment was then blunt end ligated into pBlueScript. Inverse PCR (primers Grex_FLAG_F and Grex_FLAG_R) was used to insert a 3XFLAG tag to the C-terminus of the rex gene immediately upstream of the stop codon. The triple FLAG tag cassette, including the flanks, was isolated as a BamHI fragment and cloned into the non-replicating conjugative vector pSX700. pSX700::*rex*-3FLG was conjugated into into *G. thermoglucosidans* TM444, using kanamycin resistance to select for first crossover homologous recombination events. Screening for the loss of kanamycin resistance resulted in the isolation of a second crossover 3xFLAG tagged *rex* allele which was confirmed via colony PCR (data not shown). Initially experiments were performed to ensure 3xFLAG tagged Rex was expressed. The FLAG tagged strain and WT cultures were grown in a variety of aeration conditions and western blots performed. Oxygen limitation was created by filling cultures to the top of tubes and closing the lid to restrict airflow. Western blot results showed that 3xFLAG tagged Rex was expressed constitutively (Figure 4.3). This expression pattern is consistent with *B. subtilis tatAy* and *tatCy* constitutive expression, adding to the evidence that Rex in *G. thermoglucosidans* is part of the Tat operon (Goosens et al., 2014).

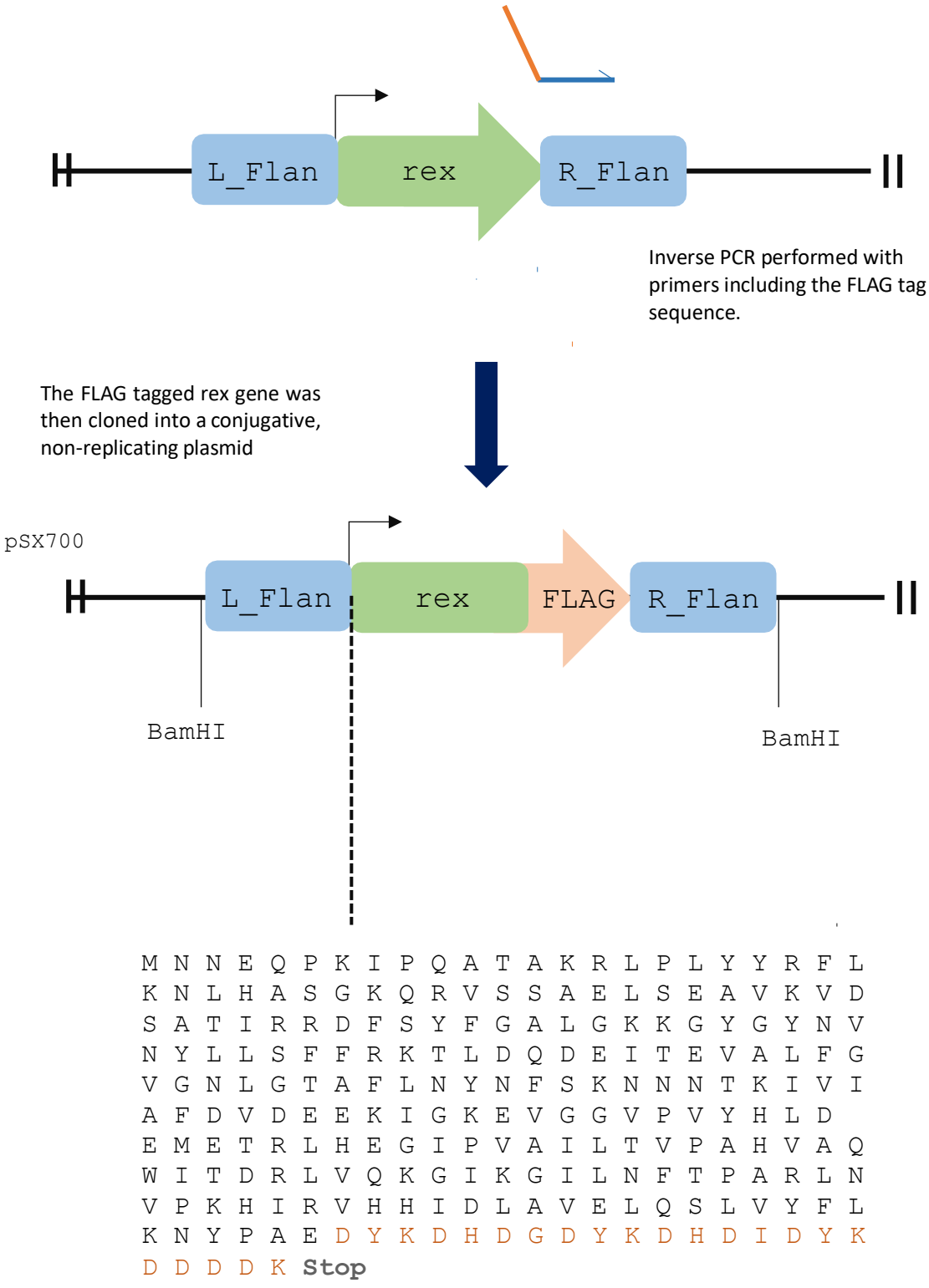


Figure 4.2 Diagram of the construction of a 3xFLAG tagged Rex allele. The in-frame FLAG tag sequence is depicted.

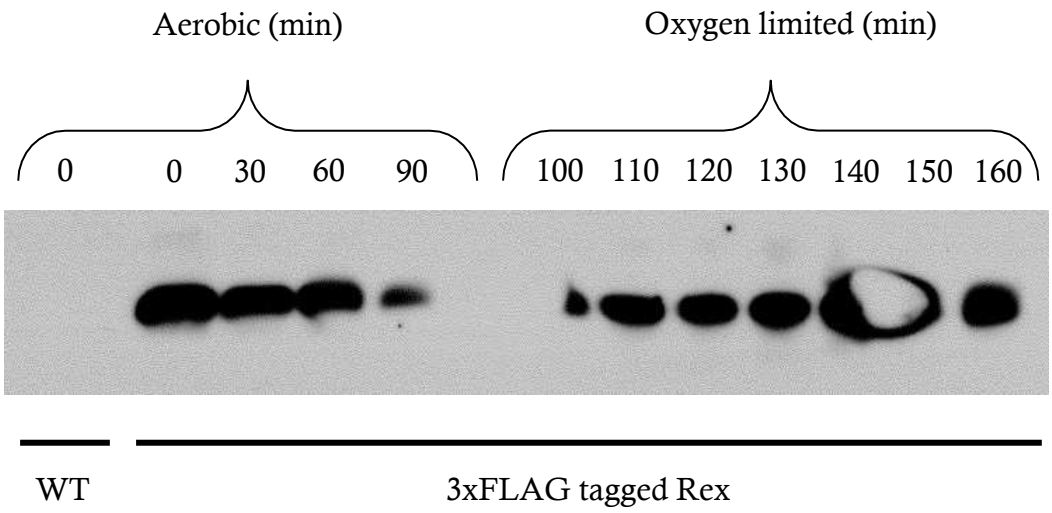


Figure 4.3 Western blot of 3xFLAG tagged Rex being expressed under different oxygen abundant conditions over a time course.

4.2.2. Optimisation of chromatin immunoprecipitation

The Chromatin immunoprecipitation method described by Grainger et al. (Efromovich et al., 2008) was optimised for *G. thermoglucosidans* (section 2.8). Specifically, the cell disruption and sonication steps required improvement as DNA fragments were not initially of the correct size and consistency for use in ChIP-seq.

The cell disruption step was first attempted with sonication (Bioruptor), then optimised using cryogenic grinding, and the purified DNA fragments from each method were analysed by gel electrophoresis (data not shown) for the amount of desirable DNA of a consistent 250 bp size range. Cell disruption by sonication was initially measured by a fall in OD^{600} measurement. Results indicated that cell disruption was successful but fragment analysis revealed that although a strong signal was detected, size was inconsistent. Cell disruption by cryogenic grinding was far more successful and resulted in cleaner, more consistent fragment sizes. It was reasoned that cryogenic grinding only disputed cells and had no impact on DNA fragments, unlike sonication, so more consistent sizes resulted.

Shearing of DNA was optimised by altering the number of biorupter cycles with the aim of generating fragment sizes of between 100-500bp for maximum resolution. A range of sonication cycles were performed on cryogenically lysed aliquots, and the fragment sizes analysed on agarose gel. The optimal number of cycles using a biorupter was found to be 35 x 30sec on/ 30 sec off which produced sizes of between 100-600bp; this was considered acceptable.

4.2.3. Chromatin immunoprecipitation of *G. thermoglucosidans* grown in liquid cultures

Biological replica cultures of strains TM444 *rex*-3xFLAG along the WT parental strain (TM444) were grown in 50 ml TSB, 60°C with 250 rpm shaking. The WT acted as a no antigen control, which provided an indicator of the amount of DNA that is non-specifically purified during the assay. Strains were grown to late-exponential phase (OD600 1.0-1.5) and then crosslinked by adding formaldehyde to a final concentration of 1% to covalently stabilise the protein-DNA complex. Cultures were incubated for a further 20 min and then any formaldehyde remaining was quenched with glycine. The immunoprecipitation with anti-FLAG antibody was performed resulting in an enriched DNA-protein sample. As a control, a sample was taken prior to immunoprecipitation and referred to as total input DNA control. These samples then underwent a series of wash steps to reduce non-specific binding, followed by de-crosslinking and a DNA purification as described in section 2.8. The samples were purified using a Qiagen MinElute PCR purification kit, and eluted in 22 µl of ultra-pure water.

4.2.4. ChIP-qPCR of predicted Rex target sites

The samples were analysed using quantitative real-time PCR (qPCR) to test efficiency of crosslinking and immunoprecipitation specificity. Enrichment at previously identified putative Rex target sites was measured using primers that bind to the upstream regions of *adhE* and *cydA* genes, as well as a negative control *GPI* primer set. Each set of data represents one biological repeat and a technical repeat (Figure 4.4). The *adhE* was the greatest enriched by an average factor of 812, whilst *cydA* was enriched by 202 fold, compared to the control *GPI* gene. The results

validated that enrichment occurred at putative Rex targets (*adhE* and *cydA*) and samples were therefore analysed by deep sequencing.

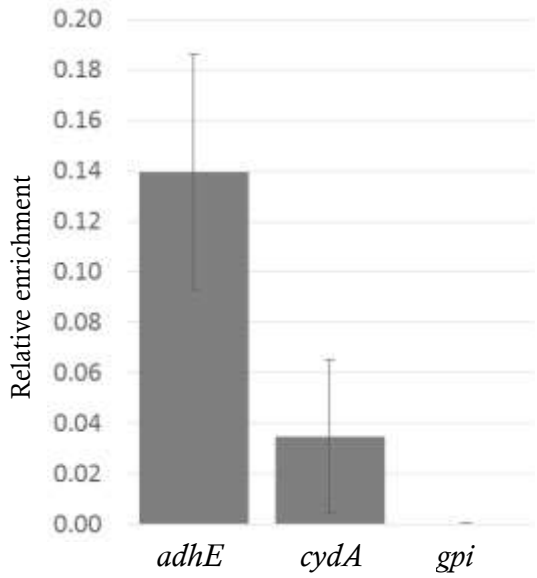


Figure 4.4 Relative enrichment as determined by qPCR using immunoprecipitated DNA using *adhE, cydA* and *gpi* primers. Values were then made relative for enrichment by subtracting results of the no antigen control strain value from the 3xFLAG tagged Rex strain.

4.2.5. Chip-Seq and identifying the Rex regulon

Samples followed the steps outlined in Figure 4.5 and were sent to The Genome Analysis Centre (Norwich, UK) for library preparation and sequencing using the Illumina HiSeq 2500 platform with a 50bp single-end read metric. Size selection was 200-300bp and two biological repeats were run. Results were received as FASTQ files and uploaded to Galaxy (http://usegalaxy.org) and converted to Sanger, from Illumina format (Galaxy Tool version 1.0.4). Quality control was performed using FASTQC (Galaxy Tool Version 0.63) and sequence data from the two lanes on the HiSeq 2500 platform were concatenated. At the time of analysis the genome most closely related to TM444 was that of *Geobacillus thermoglucosidans* DSM-2542 and so the FASTQ files were then aligned with this genome using Bowtie for Illumina (Galaxy Tool version 1.1.2) (Langmead, Trapnell, Pop, & Salzberg, 2009). To avoid artefacts, the first 5 nt and last 5 nt of each read were trimmed from each read prior to alignment and two mismatches were permitted.

The output SAM files were compressed to BAM files using Sam to Bam (Galaxy Tool version 2.0), and then visualised efficiently by creating bigWig files (Galaxy Tool version 1.5.9.1.0) (Ramírez, Dündar, Diehl, Grüning, & Manke, 2014). BigWig files were visualised using Integrated Genome Browser (IGB) to align the histogram against the annotated *Geobacillus thermoglucosidans* genome. Rex enrichment was broadly very specific, giving rise to some intense peaks across the genome (Figure 4.6).

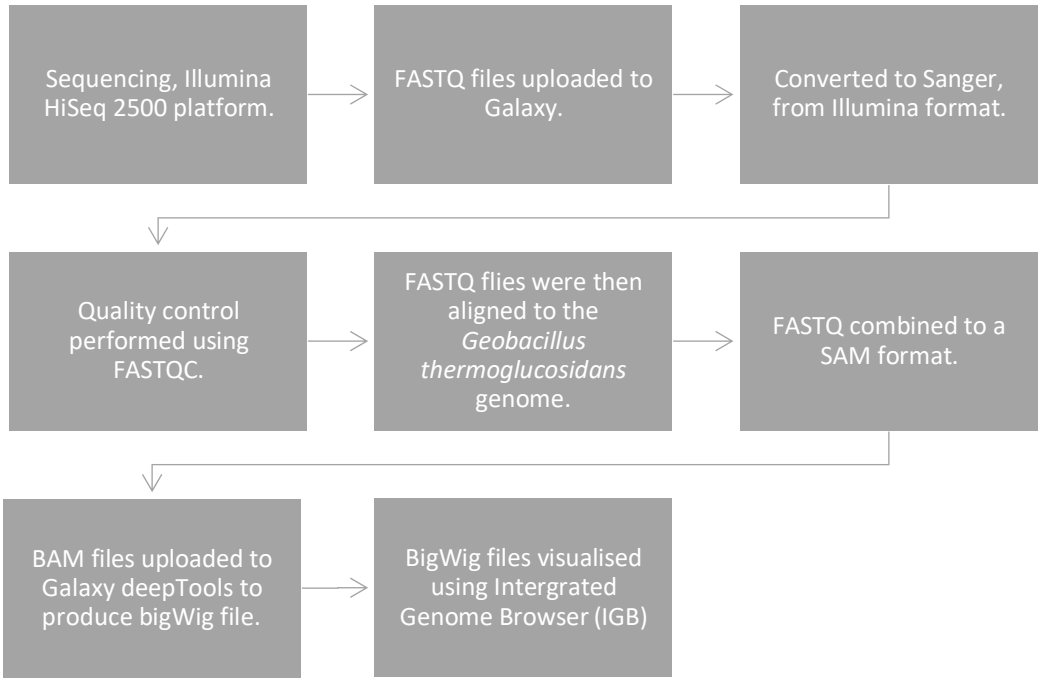


Figure 4.5 Flow chart of the stages involved in ChIP-seq data analysis.

Peaks were identified using MACS2 on Galaxy using a q value cut-off of <0.05. This resulted in the overestimation of peaks (>4000 peaks) caused by the absence of a control sample to calculate local background. Peaks with −log10 P-values of >20 were chosen for further analysis as listed in Table 4.2; peaks with lower −log10 P-values were considered less likely to be biologically meaningful since many were within open reading frames rather than regulatory regions. The final list consisted of 22 genes and the corresponding *Bacillus* orthologues and reference have been given alongside for comparison (Table 4.2). The peaks are illustrated by histograms

in Figure 4.7 which illustrate the position of enrichment. These regions were then investigated for Rex binding motifs and a *G. thermoglucosidans* consensus reached. Many of the genes highlighted as part of the Rex regulon were previously identified in other bacteria, however some new and unexpected discoveries were made.

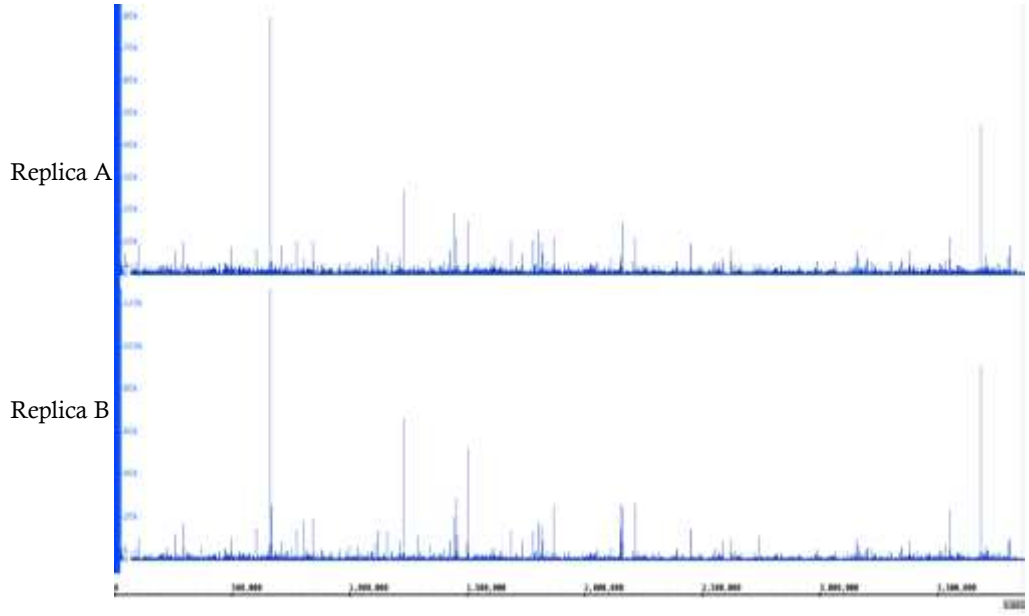


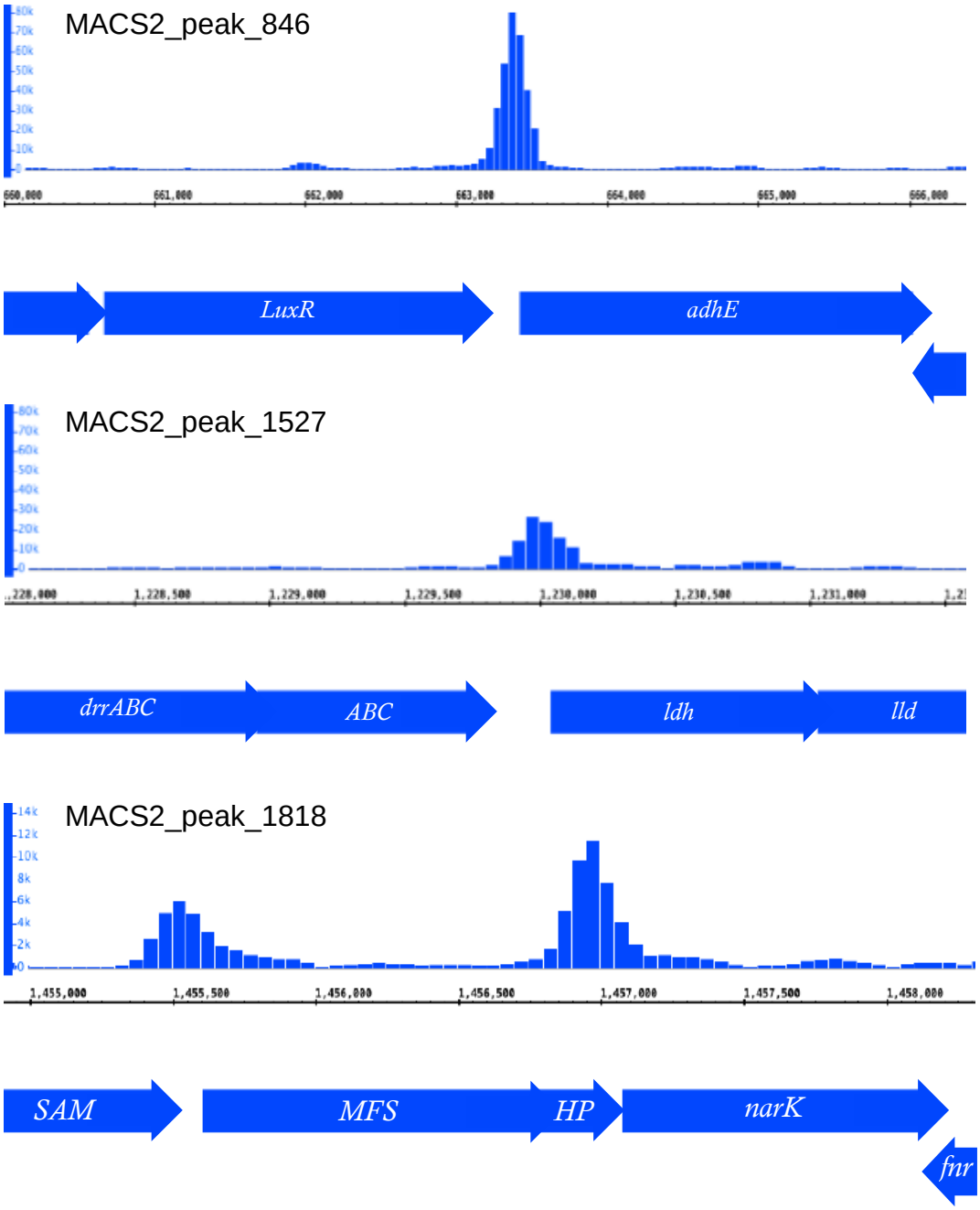
Figure 4.6 Visualisation of BigWig histogram files with integrated Genome Browser. ChIP-seq analysis was performed on biological replicas of *G. thermoglucosidans* and the entire genome represented.

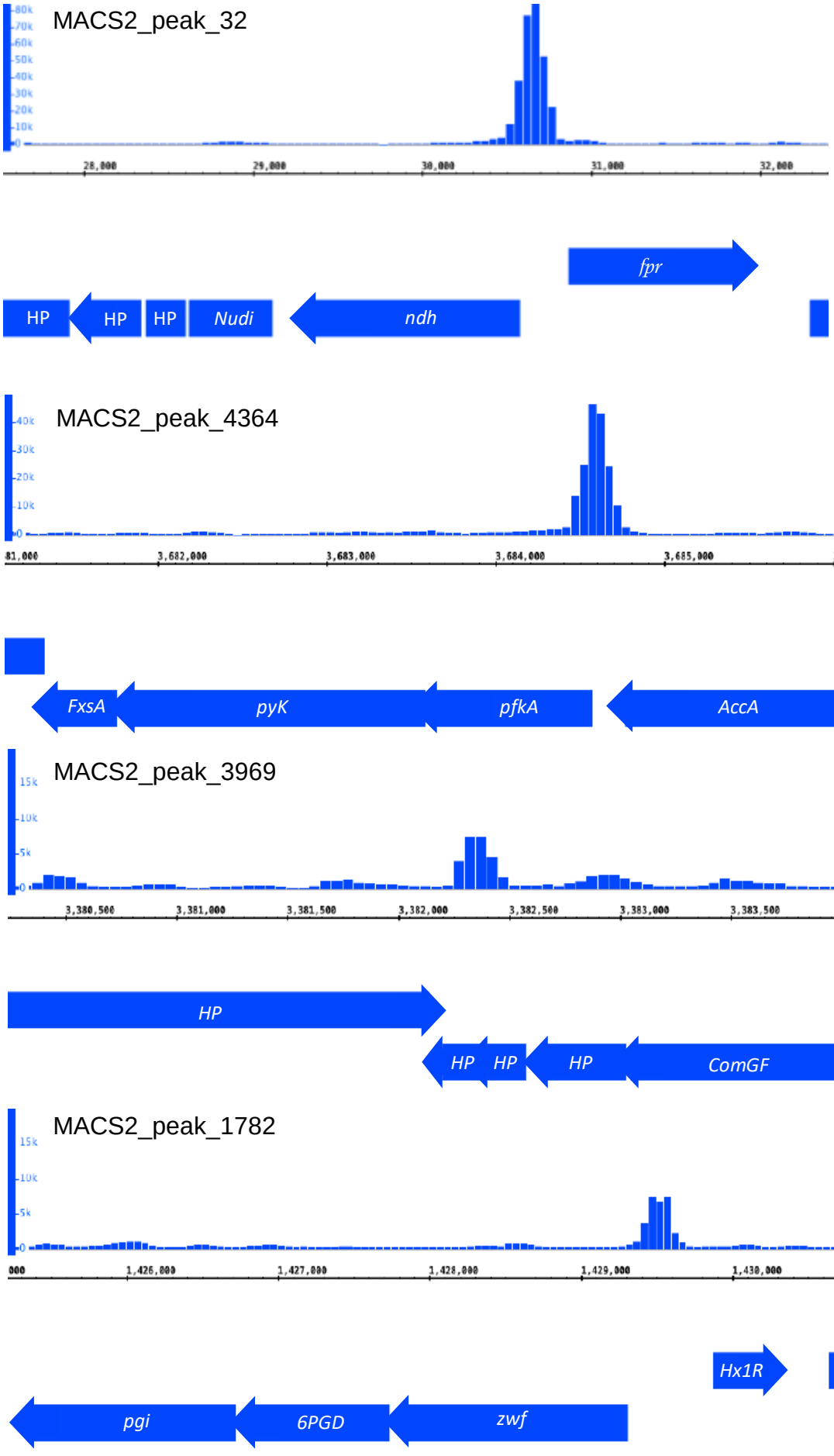
Peak name	**Locus tag**	**Function**	**−log10 P-value**	**Distance to start codon**	**Location consistent with regulation?**	***B. subtilis* orthologue**	***Reference to B. subtilis***
MACS2_peak_32	AOT13_00160 Geoth_0611	NADH dehydrogenase	20.7602	-50	Yes	*ndh (YjlD)*	(Marino, Hoffmann, Schmid, Möbitz, & Jahn, 2000)
MACS2_peak_846	AOT13_03315 Geoth_3897	Bifunctional acetaldehyde-CoA/ alcohol dehydrogenase	20.7602	-53	Yes	*adhA adhB*	(Romero, Merino, Bolívar, Gosset, & Martinez, 2007)
MACS2_peak_1527	AOT13_05975 Geoth_3351	L-lactate dehydrogenase	20.7602	-51	Yes	*ldh*	(Cruz Ramos et al., 2000; Marino et al., 2000)
MACS2_peak_1818	AOT13_07200 Geoth_3084	nitrite reductase	20.7602	-120	Yes	*narK*	(Cruz Ramos et al., 1995)
MACS2_peak_2340	AOT13_09330 Geoth_2650	polynucleotide phosphorylase	20.7602	-175	Yes	*pnpA*	(Bechhofer & Wang, 1998; Farr, Oussenko, & Bechhofer, 1999)
MACS2_peak_2739	AOT13_10935 Geoth_2365	nitrite transporter	20.7602	0	Yes	*nirC*	(Reents et al., 2006)
MACS2_peak_2689	AOT13_10655	hypothetical protein	20.7602	259	No - internal	No homology	
MACS2_peak_4364	AOT13_18150 Geoth_0897	6-phosphofructo kinase	20.7602	-44	Yes	*pfkA*	(Tobisch, Zühlke, Bernhardt, Stülke, & Hecker, 1999)
MACS2_peak_2256	AOT13_08990 Geoth_2719	signal peptidase I	20.43195	-100	Yes	*sipS*	(Bolhuis et al., 1996)
MACS2_peak_631	AOT13_02460 Geoth_0118	50S ribosomal protein L1	20.43195	629	No - internal	*rplA*	
MACS2_peak_1800	AOT13_07120	fatty acid-binding protein DegV	20.43195	183	No - internal		
MACS2_peak_3969	AOT13_16625 Geoth_1249	hypothetical protein	20.43195	5	Yes, close to N-terminus	*YqzE*	Functionally uncharacterised
MACS2_peak_1079	AOT13_04165	alpha-amylase	20.1058	166	No - internal		
MACS2_peak_1816	AOT13_07190 Geoth_0646	Nitrate/nitrite transporter	20.1058	-68	Yes	*NarK*	(Miethke, Schmidt, & Marahiel, 2008)
MACS2_peak_2229	AOT13_08865 Geoth_2744	serine/threonine protein kinase	20.1058	1565	No - internal	*YabT*	(Bidnenko et al., 2013; Pompeo, Foulquier, & Galinier, 2016)
MACS2_peak_2269	AOT13_09045	recombinase XerC	20.1058	448	No - internal		
MACS2_peak_326	AOT13_01325	metallophosphoesterase	20.1058	1731	No - internal		

MACS 2_peak _1782	AOT13_ 07050 Geoth_3 115	HxlR family transcriptional regulator	20.1058	-358	Yes	*HxlR*/*yckH*	(Yurimoto et al., 2005)
MACS 2_peak _3135	AOT13_ 12775	transposase	20.1058	468	No - internal		
MACS 2_peak _3726	AOT13_ 15405	endonuclease III	20.1058	160	No - internal		
MACS 2_peak _4549	AOT13_ 18745	transposase	20.1058	499	No - internal		
MACS 2_peak _909	AOT13_ 03545	spermidine synthase	19.78181	631	No - internal		

Table 4.2 List of the enrichment sites identified by ChIP-seq. Included in the table is the gene identifier, locus tags, −log10 P-value, distance to the start codon, if enrichment is internal to a gene, and the homologous gene found in *B. subtitles* including references.

The Rex enriched regions in Table 4.2 are represented in histograms (Figures 4.7) and can be segregated into three categories based upon location in relation to the start codon, and whether or not the gene is predicted to be regulated by Rex based on this position.


MACS2_peak_846
660,000
661,000
662,000
663,000
664,000
665,000
666,000
LuxR
adhE
MACS2_peak_1527
1,228,000
1,228,500
1,229,000
1,229,500
1,230,000
1,230,500
1,231,000
drrABC
ABC
ldh
lld
MACS2_peak_1818
1,455,000
1,455,500
1,456,000
1,456,500
1,457,000
1,457,500
1,458,000
SAM
MFS
HP
narK
fnr


MACS2_peak_32
80k
70k
60k
50k
40k
30k
20k
10k
0
28,000
29,000
30,000
31,000
32,000
fpr
HP
HP
HP
Nudi
ndh
MACS2_peak_4364
40k
30k
20k
10k
0
81,000
3,682,000
3,683,000
3,684,000
3,685,000
FxsA
pyK
pfkA
AccA
MACS2_peak_3969
15k
10k
5k
0
3,380,500
3,381,000
3,381,500
3,382,000
3,382,500
3,383,000
3,383,500
HP
HP
HP
HP
ComGF
MACS2_peak_1782
15k
10k
5k
0
000
1,426,000
1,427,000
1,428,000
1,429,000
1,430,000
Hx1R
pgi
6PGD
zwf

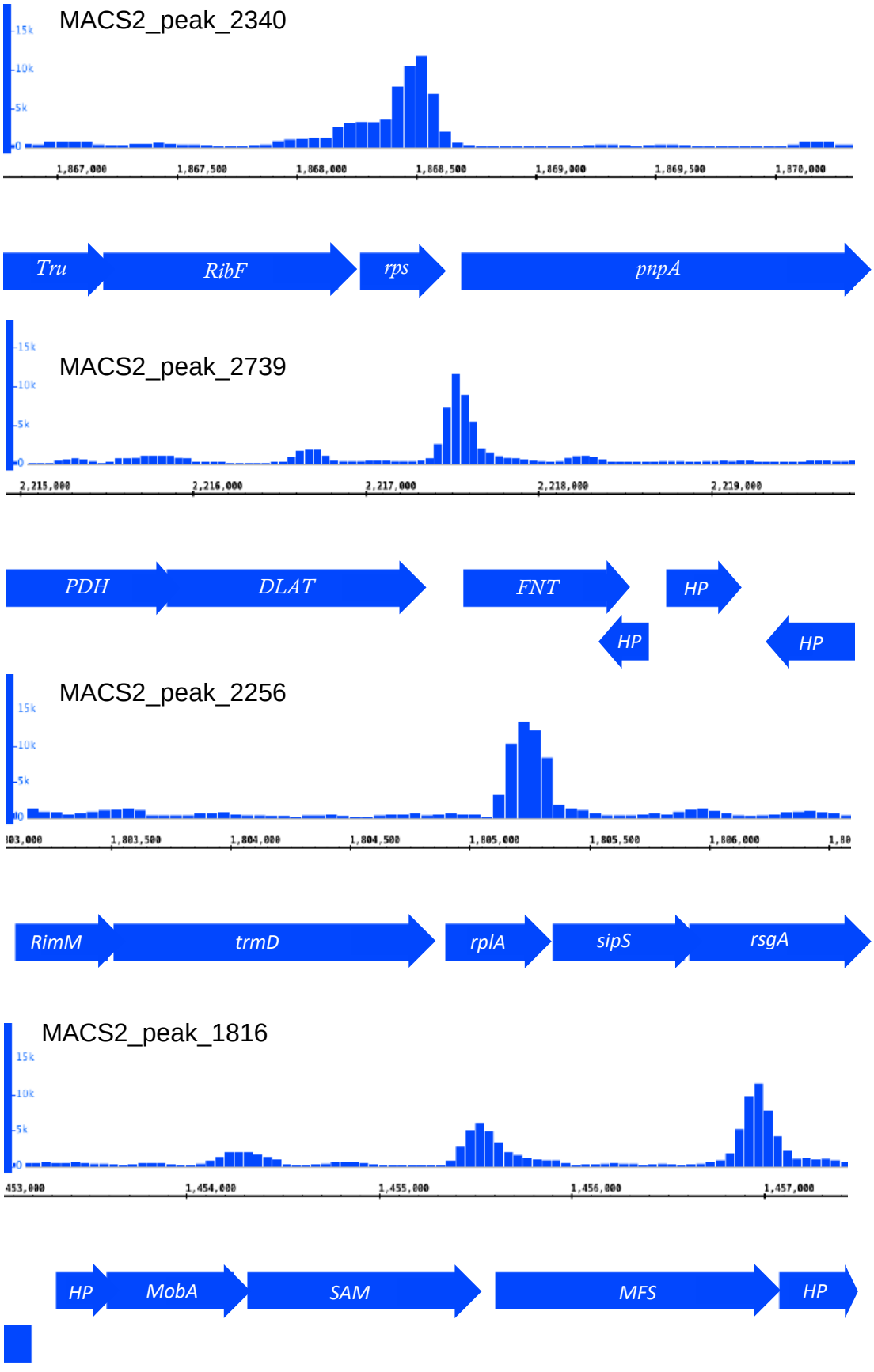


Figure 4.7 Histograms of Rex-3FLG enrichment sites and genes.

4.2.6. Putative Rex targets involved in fermentation and anaerobic respiration

The strong MACS2_peak_32 is upstream from an NADH dehydrogenase (*ndh*) gene, this enzyme is important for catalysing the transfer of electrons from NADH to the quinone pool. The regulatory relationship between Rex and Ndh has been referred to as a feedback loop that reduces oscillation of $NADH/NAD^+$ (Gyan, Shiohira, Sato, Takeuchi, & Sato, 2006). *ndh* is not likely to be part of a bigger operon, however a divergent gene may also be controlled by the ROP site indicated by the strong enrichment peak. The divergent gene is a ferredoxin-NADP reductase (*fpr*), also transcribed alone, which has not been well studied in bacteria but may play a role in adapting to oxidative stress and maintaining NADPH/NADP+ homeostasis (Jun Chen, Shen, Solem, & Jensen, 2015).

Enrichment at MACS2_peak_846 indicates that bifunctional acetaldehyde-CoA/ alcohol dehydrogenase (*adhE*) is Rex regulated. This gene is transcribed alone and is integral to responding to oxygen limitation and a lack of a terminal electron acceptor by fermenting to generate ATP.

L-lactate dehydrogenase (*ldh*) is a likely Rex target as enrichment upstream from the start codon was found at MACS2_peak_1527. Downstream of the *ldh* gene is an L-lactate permease encoding gene, involved in the transport of lactate across the cell membrane, which is likely to occupy the same operon and is therefore also likely to be regulated by Rex.

MACS2_peak_1816 and 1818 are located upstream of a predicted nitrate/nitrite extrusion gene (*narK*) and a copper containing nitrite reductase (*nirK*) gene. It is worth noting that the next gene downstream and convergent is an Fnr encoding gene. In *B. subtillis* and *E. coli* both *narK* and *Fnr* form an operon (Cruz Ramos et al., 1995). Analysis of this operon in *B. subtillis* demonstrated that a conserved Fnr binding site was located upstream and Fnr was needed for induction and expression of the operon (Reents et al., 2006).

Enrichment MACS2_peak_2739 indicated that nitrite transporter (*nirC*) is Rex regulated. This gene is not part of a larger operon and is implicated in the nitrite

export system during anaerobic respiration to minimise toxicity (Ramírez et al., 2000).

4.2.7. New Rex targets

Enrichment at MACS2_peak_4364 indicates that 6-phosphofructo kinase (*pfkA*), the first gene of an operon including pyruvate kinase (*pyK*) is a Rex target. The *pfkA* and *pyK* genes encode for enzymes that catalyse essentially irreversible reactions and are therefore critical in controlling flux through the glycolytic pathway (Figure 4.8). PfkA catalyses the transfer of phosphate from ATP to fructose-6-phosphate to form fructose-1,6-bisphosphate and ADP, whereas PyK catalyses the transfer of a phosphoryl group of phosphoenolpyruvate (PEP) to ADP to form pyruvate and ATP . Neither of these genes have previously been described as Rex targets, although several other genes involved in glycolysis including *PgK, Tpi, Fba* have been identified as part of the core Rex regulon (D. a Ravcheev et al., 2012). In *B. subtillis. pfkA* and *pyK* also constitute an operon that is weakly induced by glucose (Ludwig et al., 2001). The *pfkA pyK* operon has been studied in the bacterium *Corynebacterium glutamicum,* an industrial amino acid producer, and it was demonstrated that overexpression of glycolytic enzymes, including *pfkA* and *pyK,* enhanced glucose consumption, improved alanine product yield and reduced $NADH/NAD^+$ ratio (Yamamoto et al., 2012). It seems that *pfkA* and *pyK* upregulation during oxygen starvation or elevated NADH conditions may help to produce ATP and rebalance $NADH/NAD^+$ ratio by providing abundant substrates for fermentation. Protein blast search revealed that 6-phosphofructo kinase (*pfkA*) and pyruvate kinase (*pyK*) are the only copies of these genes within the *G. thermoglucosidans* genome. It is implausible, therefore, that the regulation of this operon is regulated simply by Rex, but it is expected that other regulators induce expression and that Rex acts as a modulator.

Enrichment at MACS2_peak_1782 indicates that a ROP site is located upstream of an HxlR family transcriptional regulator. In *B. subtillis* HxlR has been demonstrated to be an activator of the formaldehyde inducing enzymes encoded by the *hxlAB* operon involved in the ribulose monophosphate pathway (Yurimoto et al., 2005). This operon was located downstream of *HxlR* in *B. subtilis*, and is similarly positioned in *G. thermoglucosidans* genome, as just four genes separate HxlR and the

hxlAB operon. The *hxlAB* operon encodes 6-phospho 3-hexuloisomerase (*hps*) and 3-hexulose-6-phosphate synthase (*phi*), used in formaldehyde fixation, which were once thought to be confined to methylotrophs (Yasueda, Kawahara, & Sugimoto, 1999). An operon divergent to *HxlR* includes glucose-6-phosphate dehydrogenase (*Zwf*) 6-phosphogluconate dehydrogenase (*pgl*) and glucose-6-phosphate isomerase (*pgi*). However, our transcriptome analysis indicates (data not shown) this operon is very unlikely to be a Rex target and in *B. subtitles* is confirmed as being constitutively expressed (Ludwig et al., 2001).

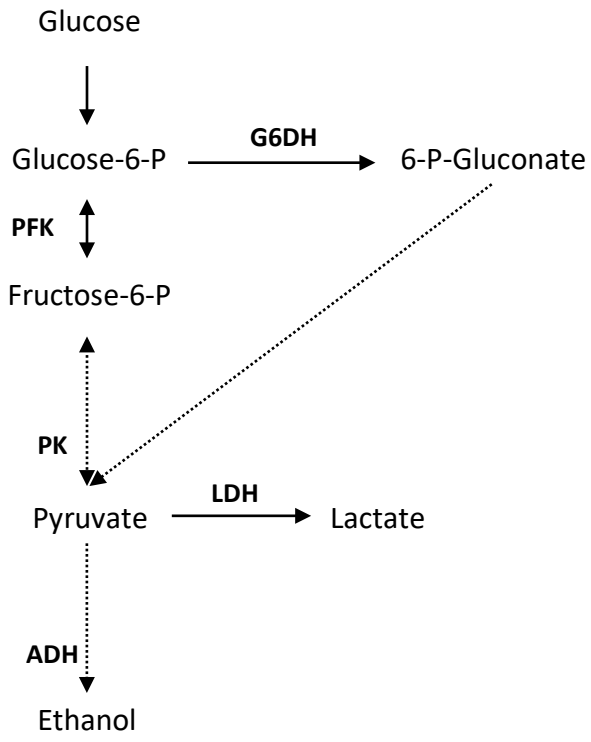


Figure 4.8 Brief outline diagram of the glycolytic and pentose phosphate pathway with important enzymes. Dotted arrows indicate that enzymatic reactions have been simplified and omitted.

Rex may control signal peptidase I (*sipS*) gene indicated by enrichment at MACS2_peak_2256. Proteins such as signal peptidase I remove signal peptides

from secreted pre-proteins (Tuteja, 2005) and in *B. subtillis sipS* expression is modulated to control protein secretion efficiency (Bolhuis et al., 1996). Another enriched region was present upstream of a polynucleotide phosphorylase. This is an enzyme found in eukaryotic, mitochondrial, chloroplasts and bacterial cells, and functions to decay mRNA (Bechhofer & Wang, 1998; Farr et al., 1999).

The enrichment at MACS2_peak_2340 is located upstream of an operon encoding a polynucleotide phosphorylase (*pnpA*) and a GTPase gene. PNPase is a 3' exonuclease that decays mRNA, and although not essential in *B. subtilis,* Δ*pnpA* strains demonstrate a clear phenotype including cells growing in chain formations (B. Liu, Kearns, & Bechhofer, 2016).

Closer inspection of the ChIP-seq data to look at genes beyond the cut-off point of −log10 P-value >20 revealed an enriched region upstream of *fnr* (Figure 4.9). *fnr* has not been identified as a Rex target before but is known to be controled by the ResDE system in many bacteria including *B. subtilis* (Geng, Zhu, Mullen, Zuber, & Nakano, 2007).

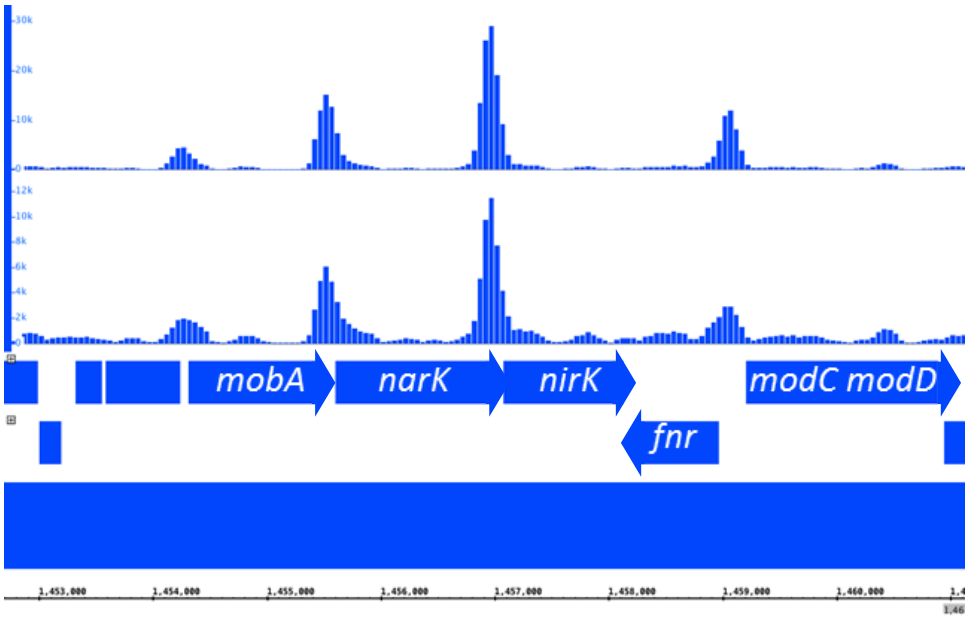


Figure 4.9 Histograms of enrichment at *fnr* site and the surrounding region.

4.2.8. Rex binding site analysis

The sequence regions of Rex enrichment were analysed for a binding site motif using http://meme-suite.org/, and the results listed in Table 4.3. Only peaks that correspond with the start of a gene were included. Relative to the centre of each peak, 50 bp upstream and downstream DNA was extracted for analysis. Using the sequences of predicted ROP sites in Table 4.3, a potential consensus Rex binding site was generated by compiling and aligning them and feeding them through a sequence logo generator (Figure 4.10). The logo in Figure 4.11 indicates that ROP sites may have a preference for GTG-n8-CAC, as do the majority of Rex ROP sites across species (D. Ravcheev et al., 2012). The nucleotides residing within the centre of the ROP site tend to favour adenine and thymine as do those eminently flanking the GTG-n8-CAC motif.

Peak number	Target gene	Position in 100 bp around peak	P-value	Binding site motif
MACS2_peak_32	NADH dehydrogenase	70	1.02e-10	AACGTGAATATTTTCACAAA
MACS2_peak_846	acetaldehyde dehydrogenase	67	5.53e-10	ATTGTGAAATGTTTCACAAA
MACS2_peak_1818	nitrite reductase	38	7.22e-9	ATTGTTCATTATTTCACAAA
MACS2_peak_1875	hypothetical protein	46	1.54e-8	ATCGTGAATAAAATCACAAA
MACS2_peak_4364	6-phosphofructokinase	21	1.80e-8	TTCGTTCATTATTTCACAAA
MACS2_peak_1527	L-lactate dehydrogenase	25	3.11e-8	ATTGTGCATTATTTCACAAT
MACS2_peak_2340	polynucleotide phosphorylase	34	5.27e-8	ATCGTGAATTGATTAACAAA
MACS2_peak_2739	nitrite transporter	4	7.51e-8	AATGTGAAAATCATCACAAA
MACS2_peak_2689	hypothetical protein	47	1.50e-7	AATGTGCATATGTTCGCGAA
MACS2_peak_1816	MFS transporter	34	4.97e-7	TTTGTGACAAACATCACAAA
MACS2_peak_991	stage II sporulation protein D	33	4.97e-7	ATCGCGCACTGTTTCACCAA
MACS2_peak_3726	endonuclease III	54	5.45e-7	TTCGTTACTTTGTTCACCAA
MACS2_peak_909	agmatinase	36	9.26e-7	AATCGGGAAAATTTCACGAA
MACS2_peak_631	50S ribosomal protein L1	40	1.29e-6	ACGGTGACGTTTTTCACATA
MACS2_peak_3969	YqzE family protein	22	1.40e-6	TACGTTAAGTTTTTCACGCA
MACS2_peak_4195	pilus assembly protein PilM	50	1.40e-6	AACGTGAATATGTTAACATT
MACS2_peak_2269	ATP-dependent protease subunit HslV	51	1.64e-6	AATGTGACAACATTCACTAA
MACS2_peak_4549	transposase	57	3.91e-6	TCCCGGCGAAGTTTCACGTA
MACS2_peak_3135	transposase	26	3.91e-6	TCCCGGCGAAGTTTCACGTA
MACS2_peak_1079	Hypothetical protein	38	5.11e-6	AATGTTTATTTAATCACGTA
MACS2_peak_2256	signal peptidase I	40	6.57e-6	AACGTTCAATAATGAACATA
MACS2_peak_2229	serine/threonine protein kinase	77	3.04e-5	ATTCCGACAAAGTTCCGGAA

Table 4.10 Predicted ROP sites from each enriched peak region analysed by using a motif search engine (http://motifsearch.com/index.php?page=motifseq). The start of the sequence in relation to the 100 bp sequence analysed is indicated.

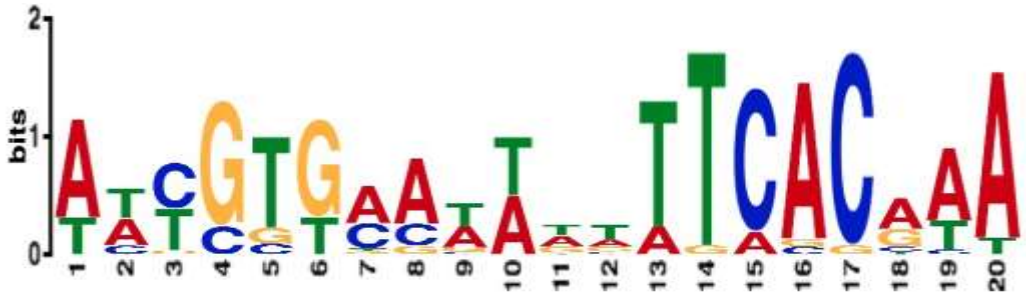


Figure 4.11 Sequence logo derived from the genomic enriched regions

4.3. Construction and analysis of a *rex* mutant in *G. thermoglucosidans*

To further understand the role of Rex and to validate potential members of the Rex regulon, a mutant strain was constructed and the expression of selected genes analysed. The creation of a *rex* mutant strain was achieved by homologous recombination, which involved transforming a rex mutant allele into *G. thermoglucosidans* and applying selection to force the first homologous recombination event, fusing the plasmid with the chromosome via integration at a flanking region. The second recombination event was then screened for and confirmed by colony PCR. Both transcriptome and metabolome data were collected and analysed.

4.3.1. Initial attempts to isolate a Δ*rex* mutant

An in-frame Δ*rex* mutant allele was constructed by M. Karigithu during his MSc, which consisted of two 0.5kb flanking regions, which would produce a Δ*rex* mutant protein with 175 central amino acids missing: MSNEQPKIPQATAKR/GS/HHIDLAIELQSLVYFLRNYPLPS*. The mutant allele was flanked by EcoRI sites and was initially cloned into the temperature-

sensitive plasmid pTMO31. However, *G. thermoglucosidans* transformants could not be isolated, possibly due to a problem with the kanamycin resistance gene on temperature sensitive plasmid pTMO31.

The Δ*rex* allele was instead subcloned into an alternative temperature sensitive plasmid pUB31 and used to transform *G. thermoglucosidans* by electroporation, selecting for kanamycin resistance at 52 °C. Single cross-over recombinants were then isolated at 68°C, purified via single colonies, then grown in absence of selection to isolate second cross-over events (Kan^S). Colony PCR screening revealed several potential mutants that were further tested for kanamycin sensitivity and further rounds of colony PCR to ensure clonal populations. A Δ*rex* mutant could not be isolated as only WT revertants and single crossovers could be identified.

An alternative approach was therefore taken by subcloning the Δ*rex* allele into a non-replicative conjugative vector pSX700. However, although initial first crossover recombinants were isolated, a second crossover Δ*rex* mutant was again not obtained, possibly due to the low rate of recombination caused by the relatively short flanking DNA. Furthermore, the difference in size of the flanks, 569 bp and 475 bp, would favour isolation of second crossover wild-type revertants. For this reason, a new mutant allele was designed with longer flanks of similar size.

4.3.2. Recombination of Δ*rex* allele into *G. thermoglucosidans*

New primers were designed (Section 1.2.3 F1MuantRex_6.12, R1MuantRex_6.12 and F2MuantRex_6.12, F2MuantRex_6.12) to amplify the Rex flanking region of 1336 bp and 1335 bp. The flanks were redesigned to delete 175 central amino acids from the *rex* gene. The larger size exponentially increases the likelihood of recombination events and the similarity in size, differing by only a single bp, results in a theoretically equal likelihood of the second crossover event, increasing the likelihood of isolating a *rex* mutant.

The flanking regions were amplified from chromosomal DNA, then independently "blunt end" cloned into pBlueScript and the orientation and DNA sequence confirmed. The upstream flank was then subcloned adjacent to the right flank using BamHI and PstI restriction sites and the construct confirmed via restriction digest. The resulting Δ*rex* allele was cloned into pSX700 via flanking EcoRI and PstI sites.

The vector pSX700-Δ*rex* was transformed via conjugation and recombined into the *G.thermoglucosidans* strain 11955 genome using kanamycin selection. Subsequent removal of selection and screening for kanamycin sensitives allowed second crossover strains to be isolated and the identification of a Δ*rex* strain. The Δ*rex* strain was confirmed using colony PCR, which indicated that the wild-type allele was replaced with a mutant allele with the loss of 175 amino acids.

4.4. Gene expression in a Δrex mutant

To confirm whether genes identified as putative Rex targets were indeed controlled by Rex, RT-qPCR was performed using RNA samples isolated from the Δ*rex* strain compared to those isolated from the WT parent. The impact of a variety of growth conditions, specifically oxygen abundance, was also investigated.

4.4.1. Optimisation of growth conditions and RNA isolation

As there were no RNA harvesting methods available for *G. thermoglucosidans,* we needed to design and optimise our own. There were two major steps to consider during RNA isolation. Firstly, the lysis of cells without the degradation of RNA. Secondly, the purification of RNA and degradation/removal of DNA. Initial attempts using Qiagen kits failed to isolate intact RNA of a high enough concentration. It was thought that this was due to poor cell breakage and therefore a new method was designed that involved cryogenic grinding to disrupt cells. Cell disruption was then followed by using an RNeasy midi kit (Qiagen) protocol. This method yielded RNA concentrations 10 fold higher than the previous method (data not shown).

Due to the large number of samples which were to be processed, the protocol was further optimised. Cell disruption by sonication using a Biorupter replaced the cryogenic grinding stage, thereby increasing maximum processing capacity from two, to twelve samples per run. To ensure this method of cell disruption was effective and did not degrade RNA, the samples were run on an RNA gel to check ribosomal RNA was intact and abundant (data not shown). This method proved to yield good quality RNA and was preferred.

For large-scale purification of RNA after cell disruption, the RNeasy midi kit was replaced by phenol/chloroform extraction and isopropanol precipitation steps described in Section 2.4.2.1. A DNase treatment kit (Ambion) was then used to remove DNA. A reverse transcription kit (BioRad) was then used to generate cDNA, and diluted 1/50 fold before use in qPCR. Primers for qRT-PCR were designed using the Primer 3 tool (Section 2.2.3).

4.4.2. Transcript analysis of WT and Δrex mutant strains

The WT and *Δrex G. thermoglucosidans* strains were grown aerobically in 250 rpm shaking flasks to OD600 1.0-1.5, at which point a 1.5 ml sample was taken (time point = 0 min) and processed to isolate RNA. The remaining culture was used to completely fill 1.5 ml microfuge tubes, and sealed such that no air gap remained. The microfuge tubes were then incubated without shaking to produce oxygen limited/fermentative growth conditions. At 30 min and 60 min intervals, the tubes were processed to isolate RNA, and qRT-PCR was performed. The normalised transcript abundance ratio was calculated by dividing *Δrex* by WT strain y-axis values, and indicates the relative level of expression caused by the loss of Rex.

The transcript profiles for *ldh, adhE, ndh* and *fnr* genes are given in Figure 4.12. Results here demonstrate that *ldh* is a likely Rex target as in the WT strain expression increases during oxygen starvation at time point 60 min and the expression profile in a rex mutant demonstrates deregulation at time point 0. However, surprisingly *ldh* expression decreases in the *Δrex* strain following further oxygen limitation, suggesting the involvement of additional effectors.

Unlike *ldh*, *adhE* expression remains supressed at T=0 in the *Δrex* strain and then, at time points 30 and 60 min, is overexpressed in the *Δrex* strain compared to WT. This again suggests that adhE is a direct target and that regulators other than Rex may also be controlling *adhE.*

Expression of *ndh* differs from that of *ldh* than *adhE.* In WT *ndh* expression is decreased by oxygen starvation. However, it is clearly a Rex target because the T=0 levels are ~4-fold increased in the *Δrex* strain. Nonetheless, the downregulaton of *ndh* during oxygen limitation remains in the *Δrex* strain again suggesting that other regulators may act to repress this gene.

The expression profile of *fnr* is of interest due to the role this regulator plays in controlling anaerobic respiration. Furthermore, the similarity of Fnr and Rex DNA binding site sequences (see Discussion), raises the possibility of interplay between the two regulators. As would be expected, in the WT strain oxygen limited growth conditions resulted in upregulation of *fnr*. In a *rex* mutant *fnr* was also induced during oxygen limited growth conditions, but to a higher level, suggesting the involvement of Rex in negatively regulating *fnr* expression.

Across all of the qRT-PCR experiments genes (*ldh, ndh, adhE* and *fnr*) in *Δrex* strains, higher maximal expression levels were obtained demonstrating a partial loss of control in responding to redox stress. However, no gene was constitutively expressed, indicating that other regulators are likely to act on these genes. The dissimilar nature of the *ldh, ndh* and *adhE* expression profiles indicates that the regulators that act upon these genes might be different or might act differently due to, for example, binding site location.

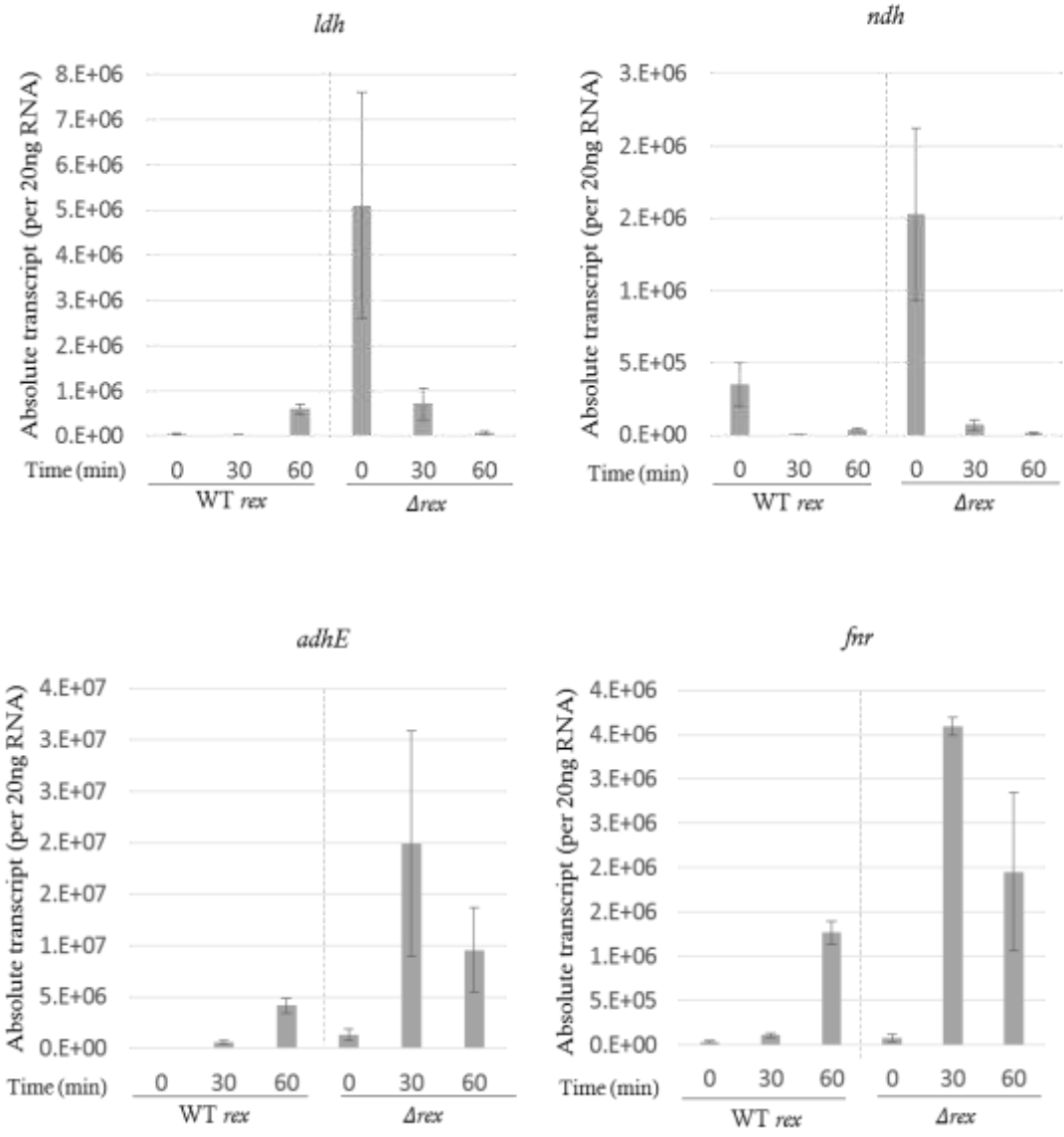


Figure 4.12 Absolute transcript of *ldh, ndh, adhE* and *fnr* measured by qRT-PCR. Error bars indicate the standard deviation take of three biological repeats.

4.4.3. General growth characteristics of the Δ*rex* mutant in liquid medium

Growth curves of both WT and Δ*rex G. thermoglucosidans* 11955 strains were conducted to discover if the mutant strain grew differently, and to inform the design of future fermentation experiments. Both WT and *Δrex* strains were grown first in pre inoculum medium 2TY, and then in rich, buffered USM medium. Results indicated that a loss of *rex* led to a reduced lag phase (Figure 4.13).

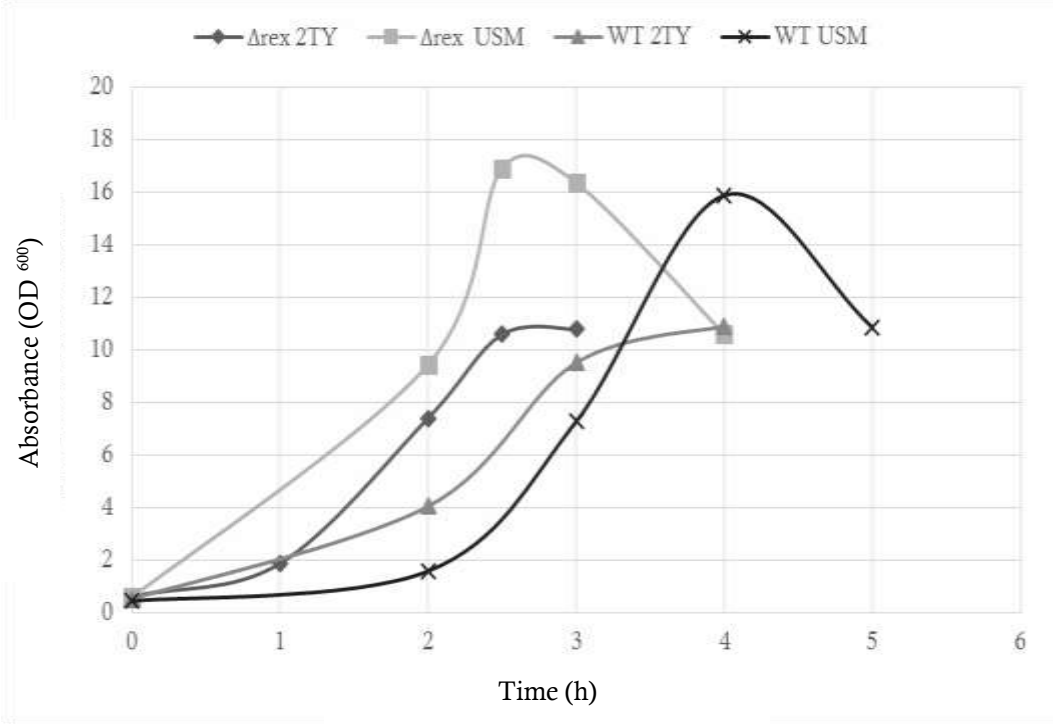


Figure 4.13 Wild type (TM960) and Δ rex (TM961) growth curves in two types ofmedia, 2TY and USM. Experiments were conducted in shaking flasks and samples were taken at near hourly intervals.

4.5. Fermentation analysis of *Δrex* mutant

To understand how the loss of *rex* might affect product formation, larger-scale fermentation analysis was carried out. Strains were grown at our industrial collaborators, ReBio Technologies, using specialist equipment to regulate oxygen

exchange and measure substrate usage and product formation through HPLC. A photograph of a fermentation tank used can be seen in Figure 4.14.

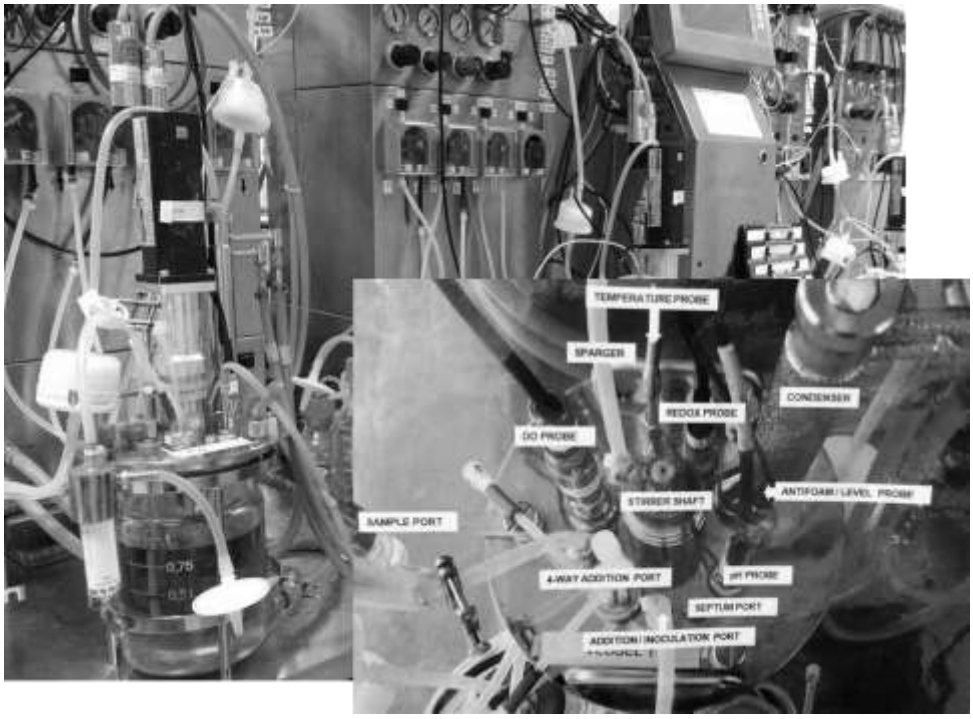


Figure 4.14 Photographs of a 1L fermentation tank at ReBio Technologys. An additional photograph of the tank from above with labels.

4.5.1. Growth and analysis of *Δrex* mutant in fermenters

To investigate the growth and product formation of *G. thermoglucosidans Δrex* and its WT parent, strains were grown in 1L fermentation tanks with an aerobic to anaerobic switch. This work was carried out collaboratively with ReBio Technologies and the University of Sussex (specifically thanks to J. Neary, K. Charman and Z. Walanus from ReBio Technologies). Samples were taken at specific time points to analyse product formation and substrate utilisation.

Confluent TAS plates were harvested and used to inoculate seed flasks of 200ml buffered 50 mM USM, 2% YE, 2% glucose at pH 7. These were grown to a target OD^{600} of 8-10 before being used to inoculate the fermentation tanks. The 1L

fermentation tanks were run in biological duplicate using media USM, 4% glucose, 2% YE, at pH 6.7. During the aerobic phase aeration was of the rate 1 Lpm until an OD of 5-7 was reached, at which point conditions were switched to oxygen limited conditions with an aeration rate of 0.2 Lpm. Samples were taken both before and after the switch for metabolome (HLPC) analysis.

The growth curves WT 11955 and *Δrex* replicate strains are shown in Figure 4.15. The replicate WT strains performed extremely similarly, however the *Δrex* strain did not. One of the strains did not exhibit a lag phase and stopped growing early, at a lower OD^{600} compared to the WT. The other strain grew similarly to the WT, but reached a lower final OD^{600}. Unfortunately we were unable to re-perform this experiment due to limited time and resources (ReBio went into administration shortly before the visit).

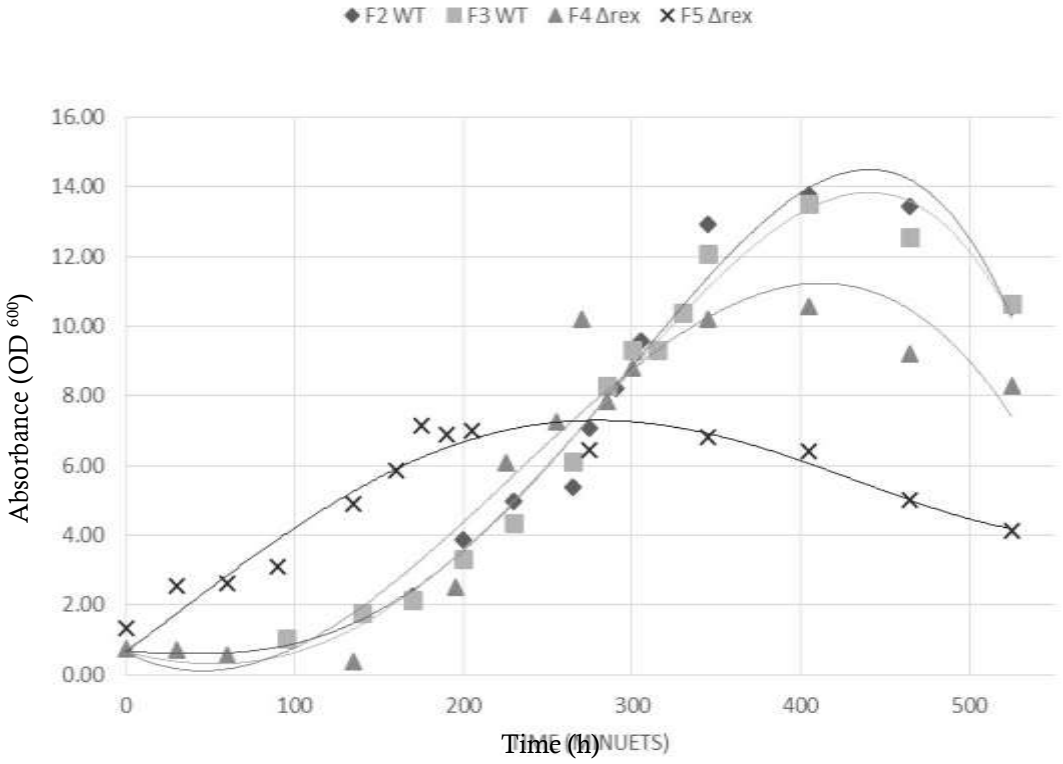


Figure 4.15 Growth curves of 11955 WT and Δ*rex* strains in fermentation tanks. Medium was USM, 4% glucose, 2% YE, at pH 6.7

4.5.2. HPLC analysis

HPLC was carried out on samples taken from the fermentation tanks. Both the products produced, and the substrate glucose used during growth, were analysed for changes in abundance (mM) periodically. HPLC analysis was performed by ReBio technologies and the data sent to the University of Sussex for analysis. The metabolic pathway in Figure 4.16 highlights the products being analysed, including ethanol, acetate, succinate, citrate, lactate and some of the genes Rex is predicted to regulate.

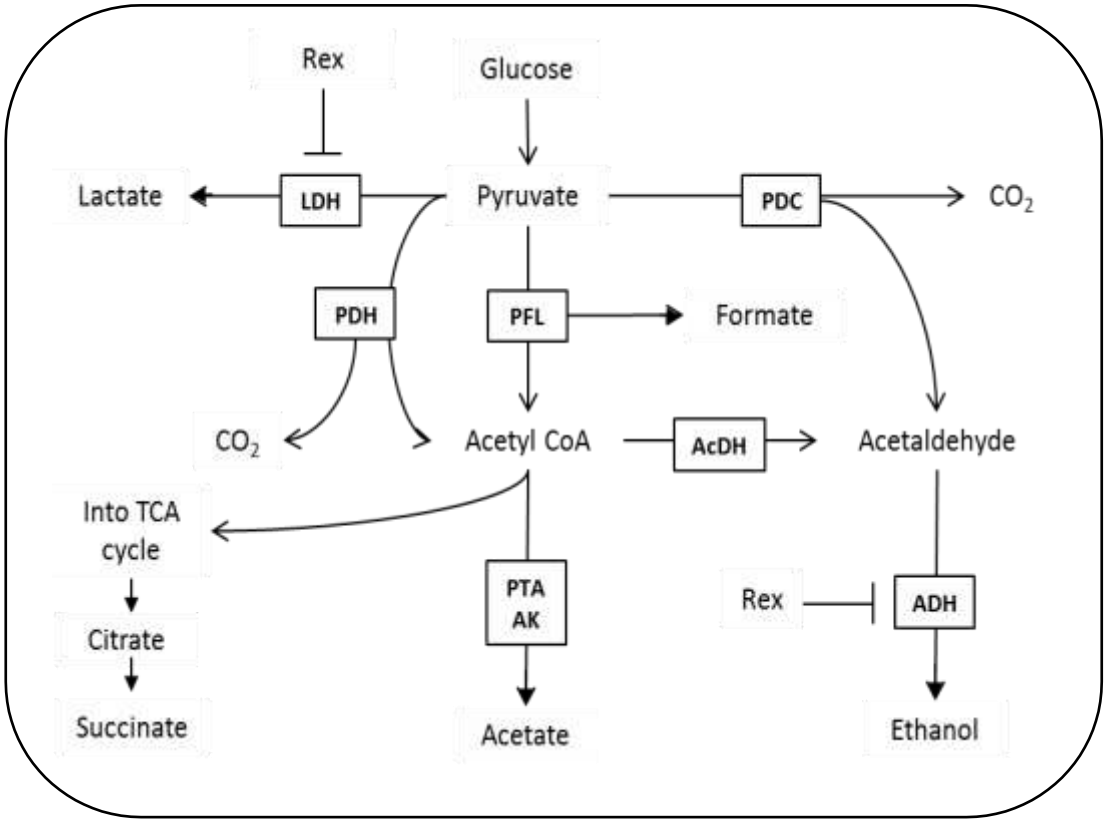


Figure 4.16 The relevant metabolic pathways and products analysed by HPLC during fermentation. LDH, lactate dehydrogenase; PFL, pyruvate formate lyase; PDH, pyruvate dehydrogenase; AcDH, acetaldehyde dehydrogenase; ADH, alcohol dehydrogenase; PTA, phosphotransace- tylase; AK, acetate kinase; TCA cycle, tricarboxylic acid cycle. Adapted from (Cripps et al., 2009).

4.5.2.1. Metabolite analysis of the Δrex strain

The production of a variety of metabolites is represented in Figure 4.17 and 4.18. The fermentative switch occurred at OD^{600} 5 for all tanks and resulted in inducing the production of formate, lactate and ethanol. The metabolites differ between the

WT and *Δrex* strain primarily in ethanol and lactate yield. The *Δrex* strain has a higher yield of lactate compared to the WT strain. This is consistent with the role of Rex as a repressor of lactate dehydrogenase (*ldh*). Counter intuitively, however, the *Δrex* strain produced less ethanol than the WT strain despite *adh* being a Rex target gene.

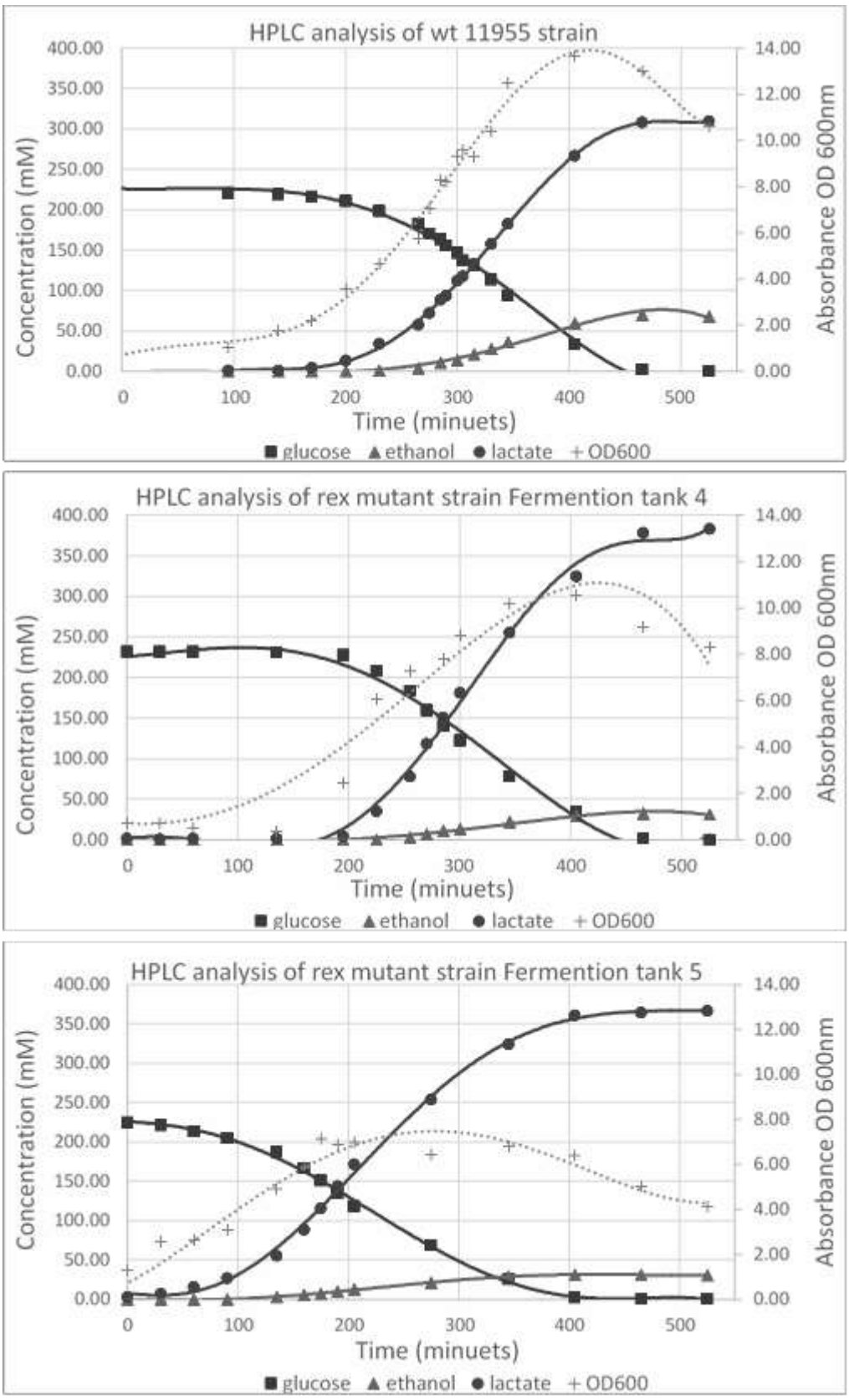


Figure 4.17 Production of metabolites by *G. thermoglucosidans* WT and *Δrex* strains. The x-axis represents the time post inoculation. Primary Y-axis represents glucose consumption (square) and the of products ethanol (square), lactate (circle). Secondary y-axis represents the OD^{600} of the culture (dotted line and cross). The oxygen limiting switch occurred when the culture reached OD^{600} 5 at which point aeration dropped from 1 Lpm and 600 RPM to 0.2 Lpm 300 RPM.

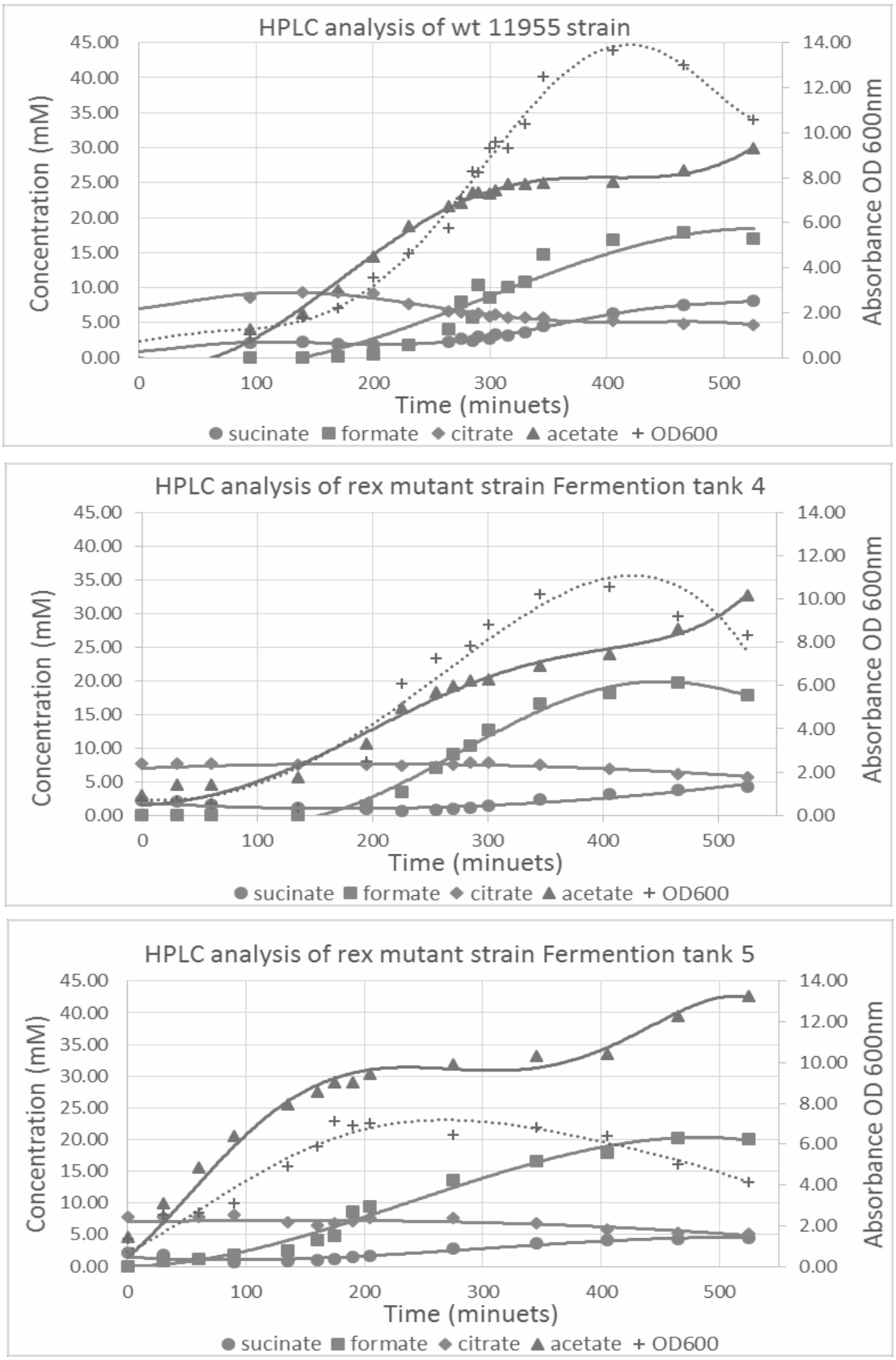


Figure 4.18 Production of metabolites by *G. thermoglucosidans* WT and *Δrex* strains. The x-axis represents the time post inoculation. Primary Y-axis represents the concentration of products sucunate (circle), formate (square), citrate (diamond) and acetate (triangle). Secondary y-axis represents the OD^{600} of the culture (dotted line and cross). The oxygen limiting switch occurred when the culture reached OD^{600} 5 at which point aeration dropped from 1 Lpm and 600 RPM to 0.2 Lpm 300 RPM.

The final yield of all metabolites was very similar in both strains, with the exception of ethanol and lactate (see Table 4.3), which decreased and increased respectively. Succinate yield was very slightly less in a *Δrex* strain, and acetate yield raised in one of the two *Δrex* strains.

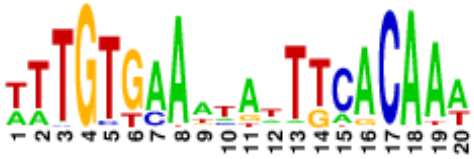
	Product yield (mM product/mM glucose substrate used)					
	Ethanol	Lactate	Succinate	Formate	Citrate	Acetate
WT tank 2	0.31	1.37	0.04	0.08	0.02	0.13
WT tank 3	0.29	1.37	0.03	0.08	0.02	0.14
Δ*rex* tank 4	0.13	1.63	0.02	0.08	0.02	0.14
Δ*rex* tank 5	0.13	1.56	0.02	0.09	0.02	0.18

Table 4.3 Metabolite yields at point of full glucose substrate utilisation.

4.6. Future work and discussion

By combing genome wide Rex ChIP-seq data with analysis of *Δrex* strain via RT-qPCR we confirmed putative Rex targets and uncovered potentially new ones. Confirmation of the *G. thermoglucosidans* consensus sequence by analysis of Rex enriched DNA regions was achieved. One example of a potential new target unrelated to central metabolism and fermentation is the polynucleotide phosphorylase (PNPase) encoding gene at peak 2340. Speculative answers as to why this gene might be controlled by Rex could include a need for rapid turnover of mRNA during times of changing redox state, so that previously expressed genes are no longer translated and newly expressed genes are better represented. This would, for example, allow the cell to more quickly adapt to oxygen limited conditions and focus resources. The majority of genes potentially regulated by Rex included those involved in fermentation, anaerobic respiration and, more generally, those involved in the recycling of NADH to NAD+ and ATP production. However, some enriched regions were internal to genes, suggesting that Rex plays a role in controlling sRNA expression.

Fnr consensus sequence in Bacillus subtilis



Rex consensus sequence in Bacillus subtilis

Figure 4. 19 Consensus DNA binding site sequences of Rex in (Wang, Bauer, Rogstam, Linse, Logan, & von Wachenfeldt, 2008) **and Fnr in *Bacillus subtilis* (Reents et al., 2006).**

4.6.1. Other regulators are likely to influence Rex target gene expression

Transcriptional analysis of specific genes (Table 4.5) in a *Δrex* strain suggested the involvement of additional regulators. One of these might be Fnr, which itself appears to be regulated by Rex. Interestingly, the expression profiles of *adh* and *ldh* differ, suggesting different mechanisms of control by Rex and/or additional regulators.

For example, *adh* may have an additional repressor, or require an activator such as Fnr, which lessens the effect of the *Δrex* mutation during oxygen abundant conditions. Fnr binds in *B. subtilis* to *ldh*, inducing expression, and it could well be true that Fnr regulates Rex targets in *G. thermoglucosidans* too (Reents et al., 2006). It has been shown that in *E. coli* Fnr works in combination with other regulators and that Fnr sites can be blocked by these regulators, preventing occupancy and gene expression induction (Myers et al., 2013). Rex and Fnr binding sites are strikingly similar (Figure 4.19) which could conceivably result in competition for binding. In a *Δrex* strain Fnr might no longer have to compete with Rex, and so may have increased access to its binding site to overexpress genes. The extent to which Rex and Fnr compete may be dependent not only on the binding site but also on oxygen and NAD^+/NADH levels, the signals for activation of these regulators. A system like this would help to differentiate between the need for expression of anaerobic or fermentative genes.

In the future, to investigate the possibility that Fnr and Rex are competing regulators, an EMSA experiment could be carried out using Fnr and Rex proteins with different binding sites and concentrations of NADH and O_2. Chip-Sequencing could also reveal the Fnr regulon and if Rex and Fnr sites overlap. Construction of an *fnr* and *rex* double mutant and its use to investigate deregulation by qRT-PCR on genes such as *ldh* and *adh,* would help to reveal if Fnr is necessary for expression of these genes. Analysis of Fnr expression in a *Δrex* strain results showed that over expression occurred during oxygen limited growth conditions only. Fnr expression is controlled in *B. subtilis* by the two-component ResDE system, induced in response to anaerobic conditions, but the exact signal is unknown (Geng, Zhu, et al., 2007; Härtig & Jahn, 2012).

Gene expression profile	**WT *rex* strain**		***Δrex* strain**	
	O_2 abundant	O_2 limited	O_2 abundant	O_2 limited
***Ldh* expression**	Low	Elevated	Extreme	Low
***Adh* expression**	Low	Elevated	Low	Extreme
***Ndh* expression**	Elevated	Low	Extreme	Low
***Fnr* expression**	Low	Elevated	Low	Extreme

Table 4.4 Expression profiles found of *ldh, adh, ndh, fnr* in WT *rex* and *Δrex* stains in both oxygen abundant and limited growth environments.

4.6.2. Fermentation and the metabolome

A *Δrex* strain was thought likely to increase the yield of fermentation products, especially ethanol. Results demonstrate that lactate yield did increase in a *Δrex* strain, but that crucially ethanol yield decreased. Although the reason behind this is unknown, a probable cause is that the ADH enzyme was not the limiting factor, and that LDH is a more efficient enzyme at pyruvate reduction and NAD^+ regeneration. The deletion of LDH may result in ethanol yield increasing, although such mutants in *B. subtilis* have a reduced growth rate. This may be because LDH is able to use both NADH and NADPH and could be integral for equilibrium of NADPH/$NADP^+$ under fermentative conditions (Romero et al., 2007). For these

reasons future research into directing metabolic flux towards ethanol production should consider the role of LDH in maintaining redox balance, substituting for another enzyme capable of utilising NADPH, preferably an alcohol dehydrogenase. An example of a strictly NADPH-dependent alcohol dehydrogenase was discovered in a thermophilic *Clostridium* and converts acetoin to 2,3-butanediol (Ismaiel, Zhu, Colby, & Chen, 1993; Korkhin et al., 1998). Although this particular alcohol dehydrogenase is thermostable it does not produce ethanol and would not be suitable, but it does indicate that potentially a NADPH-dependent one could be discovered or even engineered to replace LDH.

5. Chapter V: Development of orthogonal redox -responsive expression vectors

Overview

Bioinformatics analysis indicated that the Rex found in Clostridia might bind to an alternative motif sequence compared to that of most other bacteria including Firmicutes, Thermotogales, Actinobacteria, Chloroflexi, Deinococcis-Thermus and Proteobacteria phylum. This raised the prospect of an orthogonal Rex regulator/operator couple for *Geobacillus*, generating an alternative and independently controlled regulatory network. This might, for example, allow target genes to be induced at different redox potentials. The use of an orthogonal Rex, with a differing redox sensing potential, may offer a means of optimising gene expression for the production of ethanol or other fermentation products. To explore this, *Clostridium thermocellum* Rex (C-Rex) was chosen for experimentation as this bacterium grows at a similar temperature to *G. thermoglucosidans*. Electromobility shift assays (EMSA) were used to confirm that C-Rex and G-Rex differ in their DNA binding specificity at ROP sites. Expression vectors were then designed, expressing synthetic FLAG-epitope-tagged Rex proteins, within a specially engineered genetic circuit (several variations were constructed). To characterise the functioning of the different Rex proteins, and the ability of the circuit to express genes in a controlled manner, a fluorescent reporter gene was put under the control of a variety of ROP sites. qRT-PCR and western blot analysis were used to investigate RNA expression and protein levels, in response to changing environmental oxygen abundance, and to assess the success of the expression circuits.

5.1. Identification of potentially orthogonal and thermostable Rex regulators

Clostridia have been proposed to possess a different Rex DNA-binding motif and Rex protein compared to most other Gram-positive bacteria (see Figure 5.1) (D. A. Ravcheev et al., 2011; D. Ravcheev et al., 2012). Experiments have been conducted that determine if the different binding sites can be used as an orthogonal Rex

regulatory system with the ability to regulate fermentative genes independently of the host Rex. A search for Rex proteins in the thermophile *C. thermocellum* found that it possesses two rex paralogs, Cthe_0422 and Cthe_1798. Although it is rare for a bacterium to possess rex paralogs, *C. thermocellum* is not alone. *Clostridium bartlettii* and *Thermotogales maritima* also have paralogues and are some of the most divergent, with the largest number of Rex amino acid substitutions being found in *T. maritima* (D. a Ravcheev et al., 2012). It is not currently known why some bacteria have paralogues or even whether both paralogues sense NAD(H).

C .thermocellum – TTGTTAANNNNTTAACAA

G. thermoglucosidasius – TTGTGAANNNNTTCACAA

Figure 5.1 Proposed Rex binding site of *C .thermocellum* compared to that of *G. thermoglucosidasius.* Underlined are the key differing nucleotides.

Cthe_0422 appears to be cotranscribed with bi-functional acetaldehyde-CoA/alcohol dehydrogenase (*adhE*), whereas Cthe_1798 appears to be monocistronic, and surrounded by genes unrelated to fermentation or central metabolism. The *C .thermocellum* paralogues were aligned with the Rex protein sequences from *G. thermoglucosidans* and *B. subtilis* (Figure 5.2). Rex Cthe_1798 shared a 38% identity and Cthe_0422 shared a 35% identity to B-Rex. G-Rex shared an 83% identity with B-Rex. The protein alignment includes the Rossmann fold (NADH binding domain) and the winged helix-turn-helix (DNA binding domain). Of particular interest are the 10 residues, Pro4, Arg10, Ser31, Arg46, Lys47, Tyr51, Gly56, Gly59, Gly61, Tyr62, which are highly conserved throughout the Rex family and T-Rex solved structure indicates they contact DNA (McLaughlin et al., 2010). Phe43 and Arg58 are also important for Rex-DNA interaction, however these are far less well conserved. Cthe_0422 protein sequence has several substitutions of key residues including 4, 31, 43, 51, 58 and 59 and the protein sequence of Cthe_1798 also contains slightly fewer substitutions. Residue 47 is of particular importance as it contacts and forms a hydrogen bond with the 5th

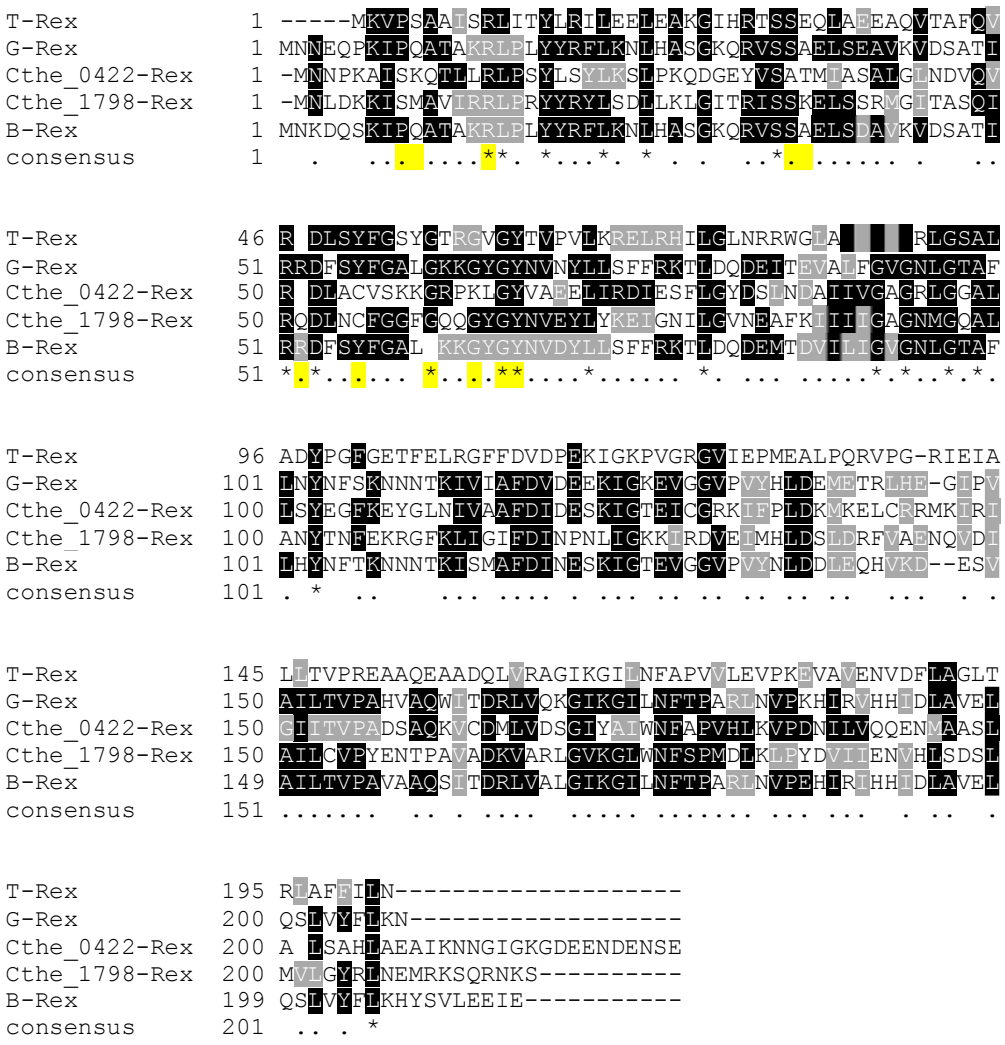


Figure 5.2 Sequence alignment of *T. aquatics, C .thermocellum, G. thermoglucosidans and B. subtilis* Rex homologues. Black highlights residues that are the same, grey those with similar properties and white which are markedly different.

G/T nucleotide of Rex operator sites and inserts into the major groove of DNA, nucleotide 5 is contacted by Lys47 in Rex proteins from Actinobacteria and Thermus, or Arg47 in Bacilli Rex, both of which are positively charged amino acids (L. Zhang et al., 2014). The Cthe_0422 protein also contains a Lys47 residue, however in Cthe_1798 protein it has been substituted for Gln (Luscombe, Laskowski, & Thornton, 2001). The Gln47 has only been found in *Clostridia* and correlates with the Rex DNA binding motif: TTGTTAANNNNTTAACAA as opposed to the binding site found both in Actinobacteria and Bacilli:

TTGTGAANNNNTTCACAA (D. A. Ravcheev et al., 2012). The key differences being at positions G5 and C14, changing to T5 and A14. This is likely to result in alteration of the way Rex can interact with DNA. To understand more about the activity of *C. thermocellum* Cthe_0422 and Cthe_1798 Rex and how they interact with ROP sites compared to *G. thermoglucosidans* Rex, the proteins were expressed, purified and used in EMSA experiments to assess their binding affinity for a variety of operator sites.

5.2. Cloning and overexpression of *Clostridium thermocellum* rex paralogues

Both *C. thermocellum* Rex paralogs, along with the *G. thermoglucosidasius* Rex were over expressed and purified in preparation for EMSA experiments to investigate relative affinities for the artificial "ideal" ROP sites: TTGTTAANNNNTTAACAA and TTGTGAANNNNTTCACAA were designed as ideal sites for clostridial Rex and Geobacillus Rex, respectively.

5.2.1. Design and overexpression of *rex*

C. thermocellum Rex (Cthe_0422 and Cthe_1798) were codon-optimised for *E. coli* expression and the sequences ordered from Invitrogen as double stranded linear DNA Strings™ Fragments with flanking NdeI and BamHI restriction sites. The two genes, along with a previously constructed G-rex NdeI-BamHI cassette (Michael Murage Karigithu, 2012) were cloned independently into NdeI / BamHI-digested pET15b which fused, then overexpressed using *E. coli* strain BL21λDE3(pLysS) (Section 2.5.1). Rex was purified by Ni-affinity chromatography (Section 2.5.2), then buffer exchanged using VIVASPIN-500 (section 2.5.3) to lower imidazole and salt concentrations and to concentrate the protein to 150-950 µg/ml. The proteins were stored in Storage buffer, on ice at 4 °C overnight. For longer storage periods proteins were suspended in 40% glycerol, snap frozen using liquid nitrogen and stored at -80 °C until needed.

While *C. thermocellum* Rex paralogues (Cthe_0422 and Cthe_1798) were successfully expressed and purified, G-Rex was found to be quite insoluble, and was only present at low levels in the soluble cell lysate. Multiple 50 ml small-scale

methods were trialled and optimised before scaling up to 500ml for a larger yield. A reduced rate of protein synthesis, by a combination of a lower growth temperature and a lesser concentration of IPTG (Table 5.1), was found to be successful for helping to increase the solubility of G-Rex. The slower rate of protein synthesis allows time for chaperones and foldases to be expressed at the proteins to fold correctly, promoting solubility (Sørensen & Mortensen, 2005). Figure 5.3 demonstrates the difference in solubility when reducing the rate of protein synthesis. G-Rex is more soluble when expressed under optimised conditions and became more concentrated in the cleared cell lysate (lane 4), compared to the whole cell lysate (lane 5).

Culture trial	OD^{600} prior to induction	Initial growth temp	IPTG concentration (mM)	After induction growth temp	Growth duration (hour)
Original	0.4 - 0.7	37 °C	1	30 °C	3.0
Final optimised	0.3 - 0.4	37 °C	0.4	16 °C	16

Table 5.1 Growth conditions of original methods of overexpressing G-Rex, and final optimised method. Changes in conditions aimed to slow the rate of expression and help promote folding, to produce more soluble protein.

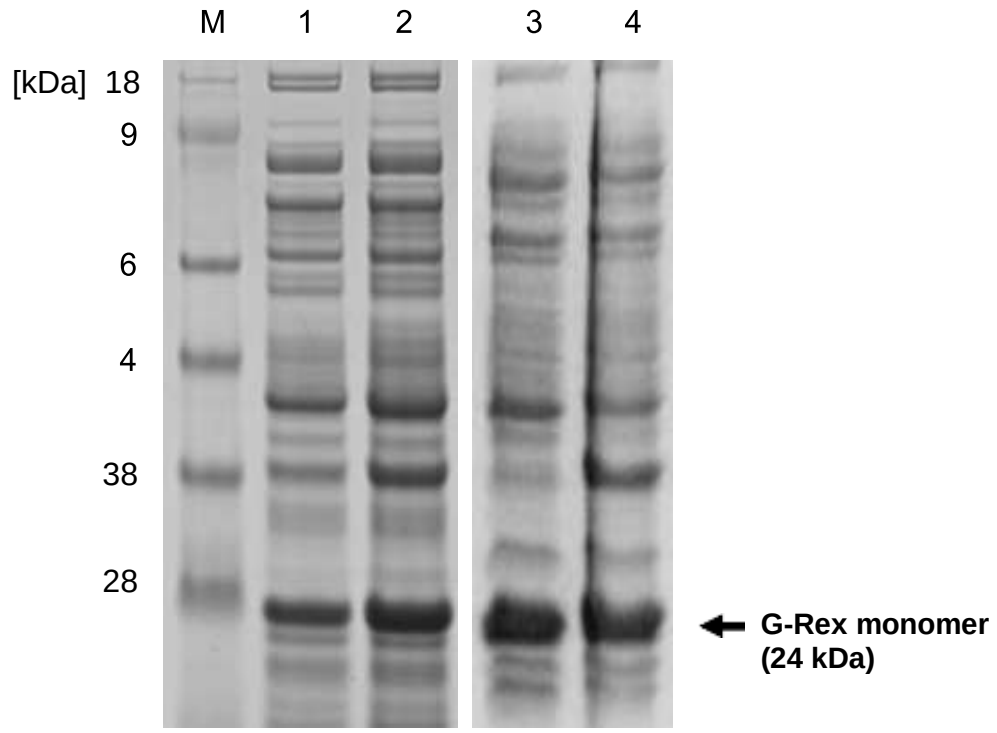


Figure 5.3 Optimisation of G-Rex expression. Samples were assessed by separating proteins using 12% SDS PAGE. Numbers to the left represent molecular weight maker (M) in kDa. Lanes 1-2 represent samples taken during the original method of G-Rex protein expression. Lane 1 is the cleared cell lysate, lane 2 is the whole cell lysate. A clear loss of G-Rex in lane 1 (compared to lane 2) was caused by the insolubility of the protein and it pelleting during centrifugation. Lanes 3-4 represent samples taken during the final optimised method. Lane 3 is the cleared cell lysate, lane 4 is the whole cell lysate. Instead of a loss of G-Rex, it becomes more concentrated in lane 4 (compared to lane 5), due to the removal of the cell debris, indicating it is far more soluble when expressed under the optimised conditions.

5.3. Design of EMSA probes

Short (~24 bp) DNA oligonucleotide probes were designed based upon predicted Rex targets, following RSAT (http://embnet.ccg.unam.mx/rsa-tools/) genome-wide searches using the consensus sequence inputs: TTGTTAANNNNTTAACAA for *C. thermocellum (DSM 1313)*; and TTGTGAANNNNTTCACAA for *G. thermoglucosidans (C56-YS93). C. thermocellum* results included the nuoA-N operon with sequence TTGTTAAATTGTTAACAT, pensioned -49 in relation to the start codon. An ROP site upstream of the nouA-N operon in *Streptomyces coelicolor* has also been identified and in vitro binding assays confirmed Rex bound (Brekasis &

Paget, 2003) and the *ndh* gene in *Bacillus subtilis* has been confirmed as a Rex target in vitro (Gyan et al., 2006). Due to this the nuoA-N operon was deemed to be a likely Rex target and a probe was designed (named CT1T, table 5.1). A second potential Rex binding site in *C. thermocellum* site was found manually. It was thought likely that the C-Rex paralogue, which is located upstream of an *adhE* gene located within an operon, is a Rex target. A site resembling a ROP was located — 211 in relation to the C-Rex paralogue start codon, a probe was also designed based upon this sequence (named C0422, Table 5.2). Two probes were also designed based upon *G. thermoglucosidasius* Rex consensus sequence. The predicted Rex binding site within the promoter region of *ldh* (lactate dehydrogenase) was used to design a probe GTLD1. The other *G. thermoglucosidasius* probe was based on the C-Rex probe CT1T, but the key nucleotides (at positions 5,14) were substituted to switch specificity to match the *G. thermoglucosidans* ROP consensus.

The sequences were ordered as single stranded DNA oligos from Eurofins and one of the oligos (per probe) was end-labelled with ϒ-ATP. The probes were then annealed using an excess of the unlabelled oligo. These were purified on a NAP-5 column and eluted in TE buffer ready for use during EMSA. During EMSA experiments probes were mixed with Rex proteins, incubated for 15 min at room temperature, and analysed non-denaturing 6% polyacrylamide gels.

Name	ROP short hand	ROP Sequence
CT1T	$ROP^{T5/A14}$	TTGT**T**AAATTGTT**A**ACAT
CT1G	$ROP^{G5/C14}$	TTGT**G**AAATTGTT**C**ACAT
GTLD	$ROP^{G5/C14}$	TTGT**G**AAATAATG**C**ACAA
C0422	$ROP^{G5/A14}$	TTGT**G**AAAACTAT**A**TAAA

Table 5.2 EMSA probe Rex binding sites. Comparison and names of the Rex binding sites chosen for EMSA experiments. Key nucleoties are highlighted in bold.

5.4. DNA binding analysis of Rex proteins using EMSA

Initially the interaction of G-Rex with CT1G and CT1T probes that differ by the presence of G or T at position 5 (and corresponding change of C to A at position 14) in the ROP sequence was investigated. As predicted, G-Rex bound more strongly to CT1G compared to CT1T (figure 5.4). This confirms the importance of G5/C14 for *G. thermoglucosidasius* Rex –DNA interactions and suggests that ROP sites that contain T5/A14 might be poorly recognised by G-Rex *in vivo*. In addition to Rex-DNA complexes that migrated just above the free probe, a smear of slower-migrating higher order complexes was also present.

EMSA studies with C-Rex Cthe_1798 and Cthe_0422 paralogues demonstrated that Cthe_1798 showed strongest binding to the CT1T probe, with only weak binding to CT1G or to GTLD1 (Figure 5.5). Surprisingly, the paralogue Cthe_0422 did not bind to any of the probes (CT1T, CT1G, GTLD1). A fourth probe named C0422 was designed (as previously described) based on the predicted ROP site upstream of Cthe_0422 and the nouA-N operon. However Cthe_0422 Rex also failed to bind to this probe (data not shown). This result was unexpected as Cthe_0422 shares its operon with bi-functional acetaldehyde-CoA/alcohol dehydrogenase, often a Rex target. In light of this, experiments using Cthe_0422

were discontinued and remaining studies focused on Cthe_1798, which is referred to as C-Rex.

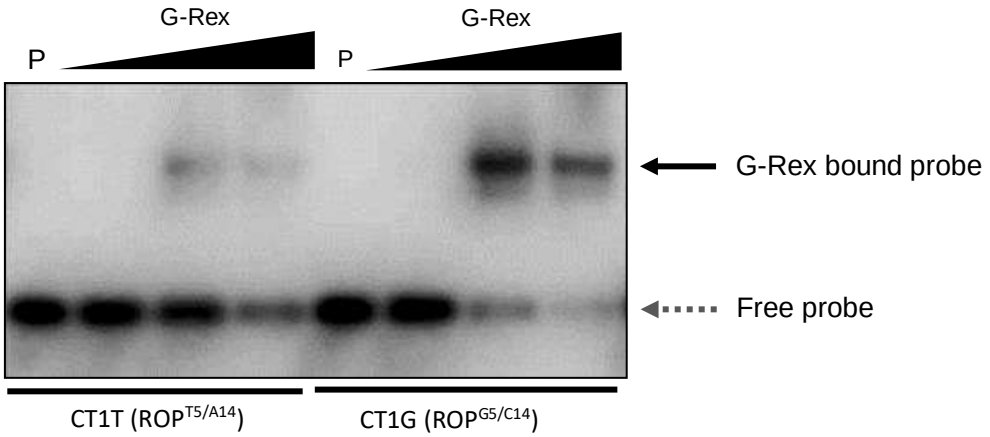


Figure 5.4 G-Rex binds to the G. thermoglucosidasius consensus sequence probe more strongly than to the probe comprising of a C. thermocellum consensus sequence. EMSA binding reactions contained 1ng of CT1T or CT1G 5' end labelled DNA probe, 0.2 µg of non-homologous herring sperm DNA, increasing concentrations of G-Rex protein. Black arrows indicate G-Rex DNA complexes. Grey dashed arrow indicates unbound probe.

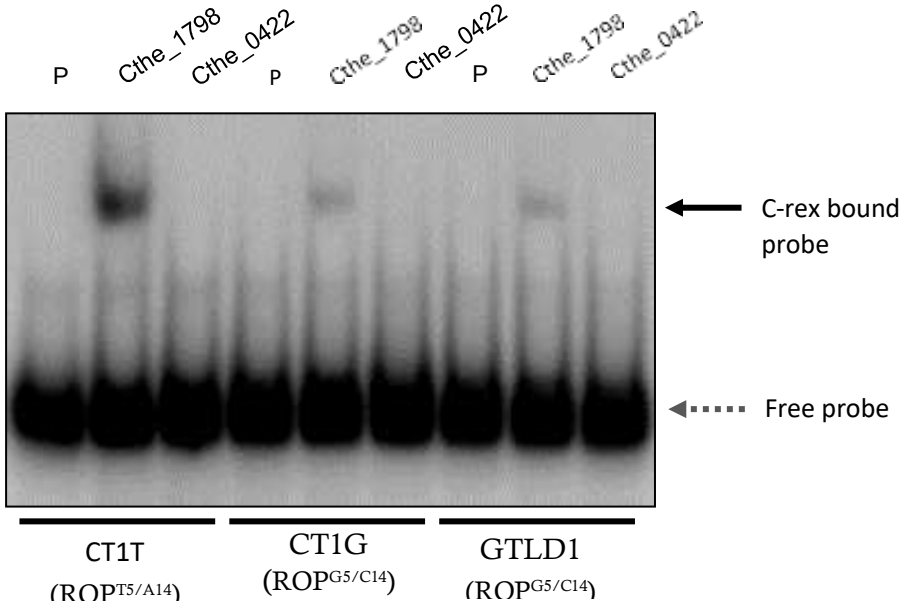


Figure 5.5 C-Rex Cthe_1798 binds to the C. thermocellum consensus sequence probe more strongly than to the probe comprising of a G. thermoglucosidasius consensus sequence. EMSA binding reactions contained 2 ng of CT1T, CT1G or GTLD1 5' end labelled DNA probe, 0.2μg of non-homologous herring sperm DNA, 0.1 μM of either Cthe_0422 or Cthe_1798 proteins. Black arrows indicate C-Rex DNA complexes. Grey dashed arrow indicates unbound probe.

5.4.1. C-Rex is responsive to NADH and NAD+

To investigate if the DNA binding activity of C-Rex is responsive to pyridine nucleotides, EMSA experiments were performed in the presence of NADH and NAD^+. Figure 5.6 demonstrates that as concentrations of NADH increase, DNA-binding decreases.

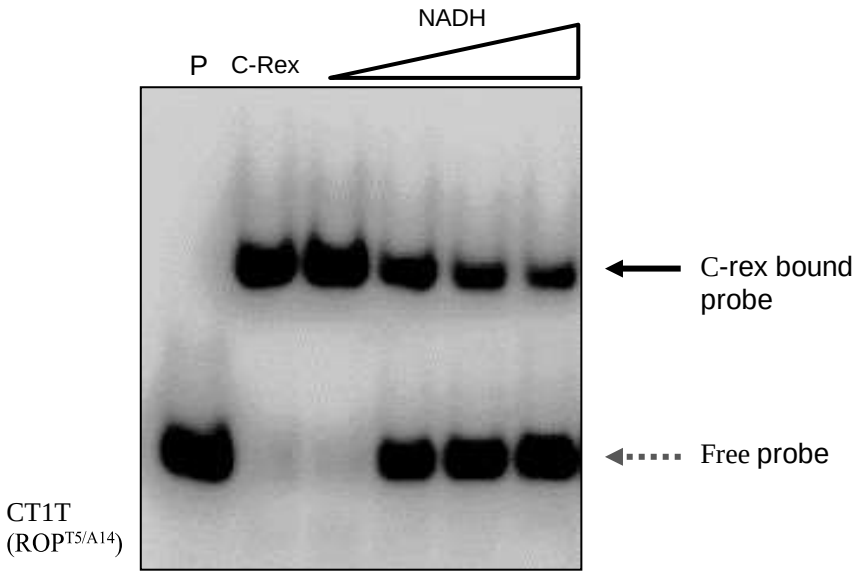


Figure 5.6 C-Rex DNA binding activity is inhibited by increasing concentrations of NADH. EMSA binding reactions contained 1 ng of CT1T 5' end labelled DNA probe, 0.2 µg of non-homologous herring sperm DNA, C-Rex protein, increasing concentrations of NADH from 0 mM, 0.001 mM, 0.01 mM, 0.1 mM, 1 mM. Black arrow indicates C-Rex DNA complex. Grey dashed arrow indicates unbound probe.

To determine whether C-Rex was also responsive to NAD^+, experiments were repeated in the presence of 1 mM NAD^+, with C-Rex of increasing concentrations (Figure 5.7). NAD+ appeared to improve DNA-binding. This is not the case throughout the Rex family. NAD^+ does not increase *S. coelicolor* Rex binding affinity, however it does in the case of the Rex orthologues from *S. aureus* and importantly *B. subtilis* (Brekasis & Paget, 2003; Gyan et al., 2006; Pagels et al., 2010).

The EMSA data indicate that the *C. thermocellum* Rex paralogue Cthe_1798 possess qualities such as an alternative consensus binding sequence (compared to G-Rex) and redox sensing, which makes it a suitable choice in the design and construction of Rex-based orthogonal expression vectors.

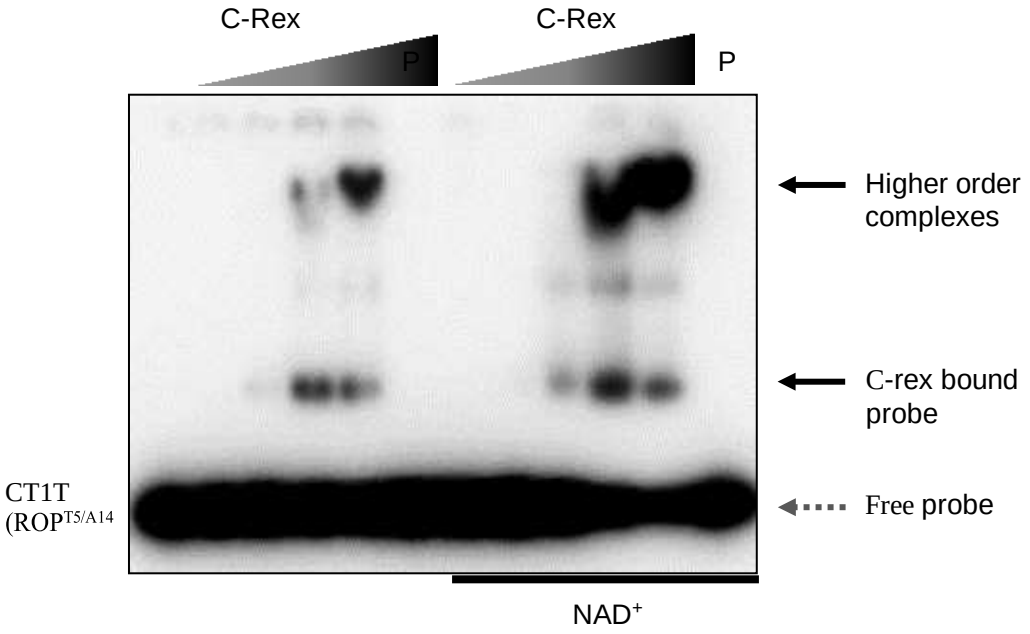


Figure 5.7 C-Rex DNA binding activity is enhanced by NAD^+. EMSA binding reactions contained 2 ng of CT1T 5' end labelled DNA probe, 0.2 µg of non-homologous herring sperm DNA, 1 mM NAD^+, increasing concentrations of C-Rex protein. Black arrows indicate Cthe_1798 C-Rex DNA complexes. Grey dashed arrow indicates unbound probe.

5.5. Design and construction of Rex-based genetic circuit expression vectors

To investigate the usefulness of a C-Rex-based expression system in vivo, a set of plasmids containing modular components was assembled. A variety of promoters were designed that included ROP sites based on either C-Rex ($ROP^{T5/A14}$) or G-Rex ($ROP^{G5/C14}$) consensus sequences (labelled in Figure 5.8 as Rex operator), in a single or double ROP site format. These promoters were located upstream of a reporter gene *gfp*. Plasmids also contained a copy of either C-Rex or G-Rex genes, the expression of which was auto-regulated through the inclusion of another ROP site. Expression studies were carried out in the *G. thermoglucosidasius* 11955 Δ*rex* strain to prevent interference from the host Rex. Initially Western blot analysis was performed to investigate whether the C-Rex and G-Rex proteins were expressed and controlled in a redox dependent manner. Their expression levels were compared to a *G. thermoglucosidasius* 11955 3xFLAG tagged Rex strain. Expression of the *gfp* reporter gene and that of selected genomic rex targets was also then investigated by qRT-PCR. Figure 5.6 is a schematic of the genetic circuit and

illustrates that Rex expression leads to negative regulation of *gfp* and autoregulation of *rex* itself.

5.5.1. Design of *rex* gene modular components and auto-regulating promoter

The design of the C-Rex gene modular fragment was based on the *C. thermocellum* cthe_1798 gene, which was found to be the only of the two paralogs which bound to probes during EMSA experiments. This gene was codon optimised for *Bacillus subtillus*, as was the G-Rex gene, to help ensure both would be expressed in a very similar way in *G. thermoglucosidans*. A 3xFLAG tag was placed just before the stop codon of *rex* (see in yellow Figure 5.9) for western blot analysis of protein levels. *Streptococcus mutans* and *S. coelicolor* have been found to express Rex via an auto-regulated feedback loop (Bitoun, Liao, Yao, Xie, & Wen, 2012; Brekasis & Paget, 2003), and bioinformatics has predicted that out of a total of 119 genomes analysed, 51 demonstrate Rex autoregulation (D. A. Ravcheev et al., 2012). An auto-regulatory mechanism may be preferable for this project so *rex* is not overexpressed on multicopy plasmids. Auto regulation was incorporated into the design as it would also prevent overexpression that could occur when using a constitutive promoter. Negative autoregulation has also been found to increase

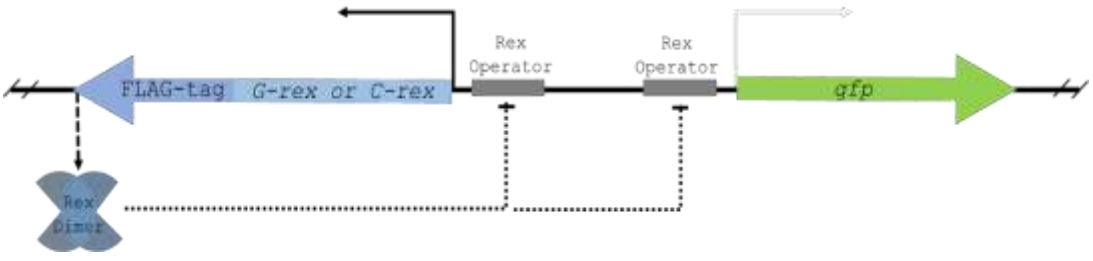


Figure 5.8 Schematic of the rex genetic circuit module.

```
  1 ATGCCTCGAG GTCAATGATT GTTAAATACT TAACAATGTG ATAGAATGGT GTTTAGGAAA AGCGTACGCC GTAACAAAGA
 81 GGAGAGTGCA AAATGAACCT GGACAAAAAA ATCAGCATGG CAGTTATTCG TCGTCTGCCT CGTTATTATC GTTATCTGAG
161 CGATCTGCTG AAACTGGGTA TTACCCGTAT TAGCAGCAAA GAACTGAGCA GCCGTATGGG CATTACCGCA AGCCAGATTC
241 GTCAGGATCT GAATTGTTTT GGTGGTTTTG GTCAGCAGGG TTATGGTTAT AATGTGGAAT ATCTGTACAA AGAGATCGGC
321 AATATTCTGG GTGTTAATGA GGCCTTCAAA ATCATTATTA TCGGTGCCGG TAATATGGGT CAGGCACTGG CAAATTATAC
401 CAACTTTGAA AAACGCGGTT TCAAACTGAT CGGCATCTTT GACATTAATC CGAACCTGAT CGGGAAAAAA ATCCGTGATG
481 TTGAAATCAT GCATCTGGAT AGCCTGGATC GTTTTGTTGC AGAAAATCAG GTTGATATTG CCATTCTGTG TGTGCCGTAT
561 GAAAATACAC CGGCAGTTGC AGATAAAGTT GCACGTCTGG GTGTGAAAGG TCTGTGGAAT TTTAGCCCGA TGGATCTGAA
641 ACTGCCTTAT GATGTGATTA TCGAAAATGT GCATCTGTCC GATAGCCTGA TGGTTCTGGG TTATCGTCTG AATGAAATGC
721 GTAAAAGCCA GCGCAATAAA AGCGACTACA AAGATCATGA CGGCGATTAT AAGGACCACG ATATCGACTA CAAAGACGAT
801 GACGATAAGT AATTAATTAA ATGC
```

Figure 5.9 Synthesised C-Rex fragment modular component. Sequence includes a specially designed auto regulating promoter, and a FLAG tag. Pink highlights are Rex binding sites. Green highlights represent the extended -10 region. Underlined font colour indicates -35 or -10 region. XhoI, BsiWI and PacI restriction sites are highlighted in blue. Yellow highlighted nucleotides are the 3xFLAG tag.

response time, leading to a more dynamic gene expression system (Rosenfeld, Elowitz, & Alon, 2002). Therefore to control Rex expression, modular fragments were designed to include a promoter sequence that included a ROP site. This promoter was based on the pUP, which derives from the constitutively expressed uracil phosphoribosyl transferase gene of *G. thermoglucosidans* and is functional in both *E. coli* and *G. thermoglucosidans* (Reeve et al., 2016). From the pUP promoter library, pUP3 was selected as it possessed one of the midrange expression profiles. The promoter was modified so that a Rex binding site was placed partially overlapping the -35 region (Figure 5.9). As this results in a suboptimal -35 region, an extended -10 region was also added as this element enhances *Bacillus* promoter strength, and was found to mitigate the loss of similarity within the -35 region in *E. coli* (Voskuil, Voepel, & Chambliss, 1995). A TRTG motif was found to be conserved in extended -10 regions of *Bacillus* sp. (G. Chen, Kumar, Wyman, & Moran, 2006). The sequence TGTG was chosen as crucially only a single nucleotide of the pUP3 sequence would need to be substituted to achieve it (figure 5.10).

The G-Rex and C-Rex promoters were identical, except for the two discriminatory nucleotides within the ROP site. An XhoI restriction site (CTCGAG) was placed

upstream the promoter and a BsiWI (CGTACG) site downstream, just before the ribosome binding site (RBS). These sites allow for exchange of components so that other promoters could be studied in the future. The C-Rex and G-Rex modular components, including the auto-regulating promoters, were synthesised by Invitrogen as gene Art double stranded strings, and cloned into pG1AK-sfGFP_OriT via XhoI and PacI sites (Figure 5.11).

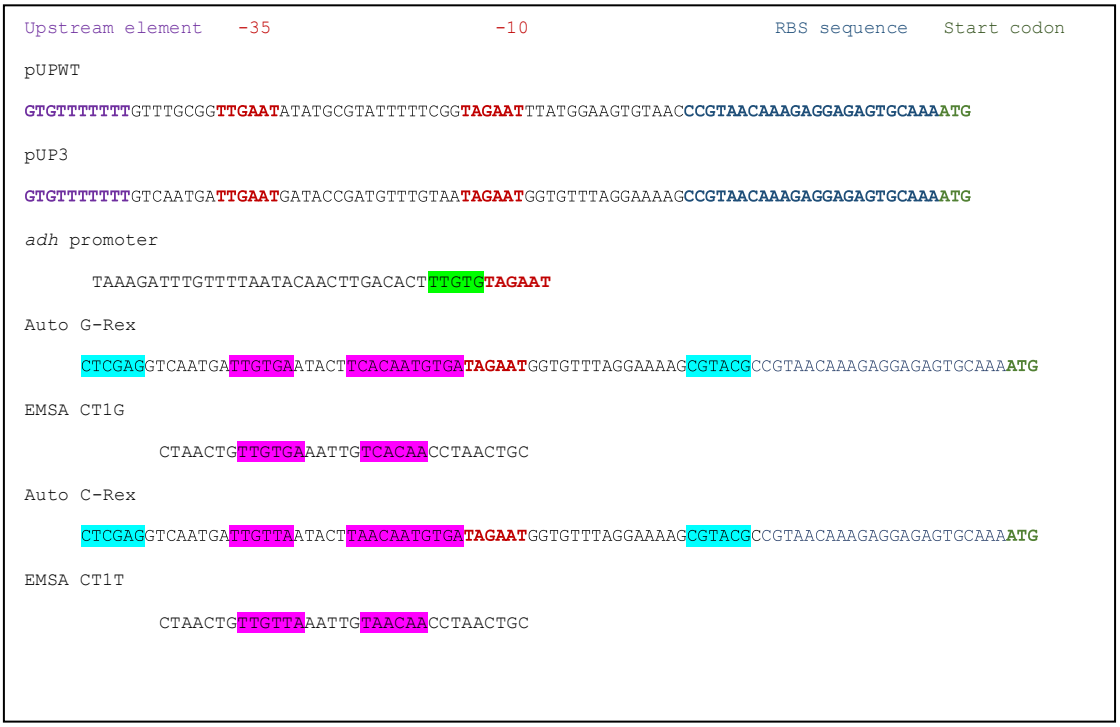


Figure 5.10 The design of Auto G-Rex promoter. Sequence of pUPWT a uracil phosphoribosyl transferase gene promoter in G. thermoglucosidans. pUP3 from a promoter library based on pUPWT. adh promoter region, including an extended -10 region. Auto G-Rex promoter designed based on pUP3 and adh extended -10 region, incorporating a G-rex binding site. The EMSA CT1G probe sequence including a Rex consensus binding site. Auto C-Rex promoter. EMSA CT1T probe sequence. Hilighted in blue are XhoI and BsiWI restriction sites. Pink highlights are Rex binding sites. Green highlights represent the extended -10 region. Red font colour indicates -35 or -10 region.

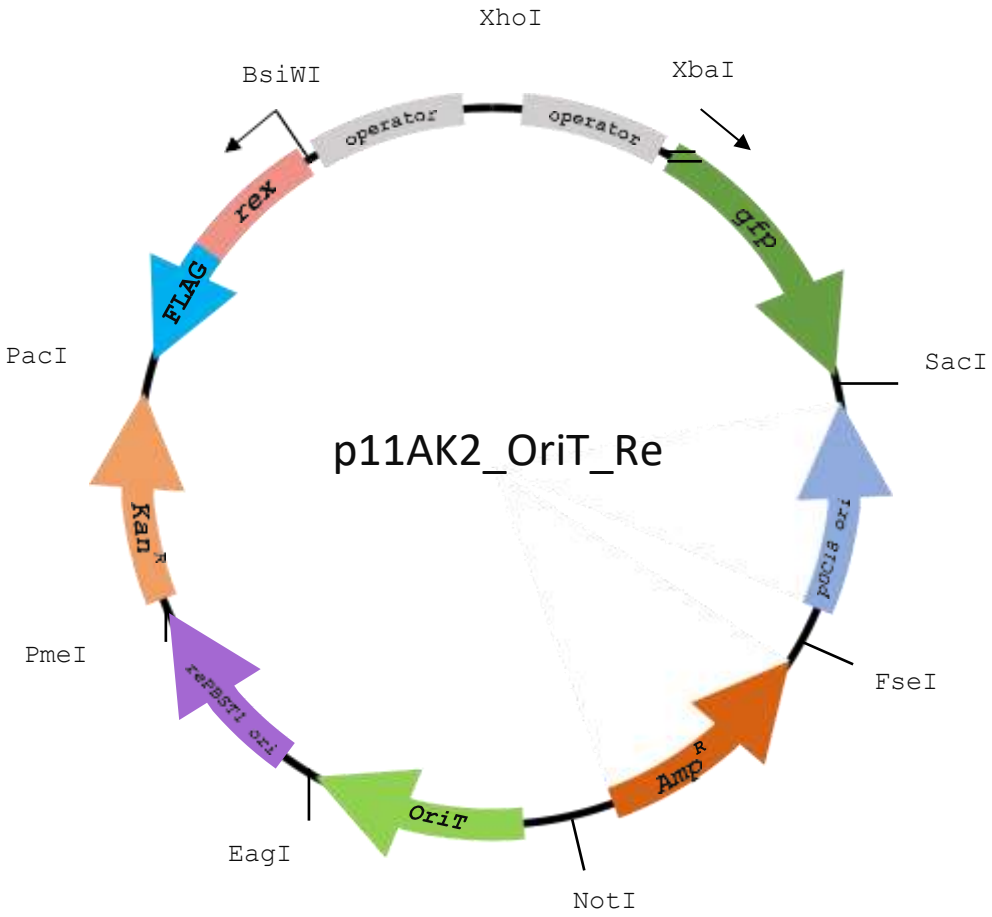


Figure 5.11 Plasmid Map of p11AK2_OriT_Rex.

5.5.2. Analysis of the auto regulating promoter controlling *rex* gene expression

It was expected that the plasmid-borne *rex* genes should be autoregulated such that a shift in the NADH/NAD+ redox poise would increase the cellular level of Rex protein. To test this, western blots were performed, and the data compared to that obtained with *G. thermoglucosidasius* 11955 3xFLAG tagged *rex* strain, in its natural locus. Overnight TSA plate cultures were used to inoculate 50 ml TSB, then grown aerobically with shaking to an OD_{600} of 1.5. Following harvesting of an aerobic sample, two 15 ml tubes were filled completely, sealed to eliminate air, then incubated without shaking for 30 or 60 mins, prior to harvesting. Following the

washing of pellets 0.9% NaCl (w/v) they were resuspended 1 ml of ice-cold Lysis buffer (0.9% NaCl, containing protease inhibitors). 250 µl of 50% (v/v) TCA was then added and, following incubation on ice for 15 min, samples were fully disrupted using a biorupter. The TCA insoluble material, including protein, was harvested by centrifugation, washed with 100% acetone, then finally resuspended in a denaturing SDS-PAGE gel loading dye prior to SDS PAGE and Western analysis.

The results revealed that Rex expressed from the native locus is constitutively expressed under these growth conditions in the WT strain. This concurs with the absence of an obvious Rex binding site upstream of the operon and the absence of a Rex peak in the Chip-seq experiments. It therefore appears that Rex is not autoregulated or even induced by oxygen limitation through a different mechanism in *Geobacillus*. However, as expected, when C-rex or G-rex were placed under the control of promoters that included ROP operators, protein levels increased during oxygen limitation (Figure 5.12). It therefore appears that each Rex protein is active for both DNA binding and NAD(H) sensing in vivo: C-Rex and G-Rex bind to their respective ROP sites to inhibit expression during high oxygen abundance and low NADH levels, then dissociate when aerobic respiration is inhibited and NADH levels rise. Presumably, basal levels of Rex ensure very low levels of expression under oxygen abundant conditions.

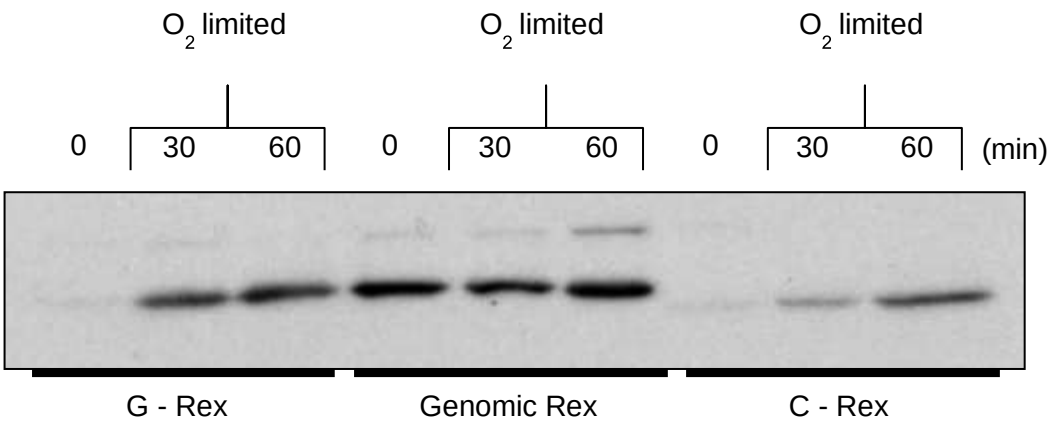


Figure 5.12 Oxygen limitation results in more abundant Rex protein levels when auto-regulated.

Importantly, plasmid-expressed C-Rex and G-Rex protein levels did not exceed that seen for the WT strain, which should minimize the possible pleiotropic effects that might result from excess regulator. These data indicate that the auto-regulating promoter worked as intended and that experiments should proceed.

5.5.3. Design and analysis of promoter controlling *gfp* gene expression

An expression vector was designed containing synthetic autoregulated G-Rex or C-Rex genes and a target promoter. To characterise the functioning of the circuit, a fluorescent reporter gene, *gfp*, was put under the control of the target promoter, allowing several promoter/operator variants to be tested. The constructed vectors were transformed into *G. thermoglucosidasius* 11955 Δ*rex* and RNA expression analysed by qRT-PCR (initial results revealed that GFP fluorescence could not be used to assess gene expression under oxygen limited conditions).

5.5.3.1. Design of Rex responsive promoters / operators

Controllable promoters were designed to include a Rex ROP site based on the EMSA probes previously tested. While most Rex-controlled promoters have single ROP sites, in *Streptomyces* double Rex sites have been found upstream of the *cyd*, *wblE* and SCO5207 genes (Brekasis & Paget, 2003; Strain-damerell, n.d.). These

double sites overlap one another by 5 bp and therefore occupy opposite sides of the DNA helix. EMSA experiments using *cydA* and *wblE* probes result in double shifts indicating two Rex dimers bound (Brekasis & Paget, 2003). Importantly, *cydA* is the most tightly controlled of the genes in *S. coelicolor*, perhaps on account of the high occupancy of the double ROPs site (Brekasis & Paget, 2003). On account of this, both single and double ROP sites were integrated into promoters. The arrangement of the G-Rex and C-Rex double ROP sites, with respect to the promoter elements, were designed based on *S. coelicolor cydA* and *wbIE* (Figure 5.13 and 5.14). The promoters included -35 and -10 regions based on pUP3 with an additional extended -10 region. Unlike the design of the auto-regulated promoter (section 5.5.1), the ROP site was positioned downstream of the -10 region similar to the arrangement in *S. coelicolor cydA*. Flanking XhoI and XbaI restriction sites were incorporated to allow cloning into the previously-constructed pG1AK- sfGFP_oriT_G/C-Rex plasmids (Figure 5.11). The plasmids were conjugated into *G. thermoglucosidasius* 11955 Δ*rex* strain.

Name	Sequence
	-35 extended -10 ---- Rex site ---- -----
G-rex single ROP	TTGAATGATACCGATGTTTGTGATAGAATTGCTTGTGAAATAATGCACAA
G-rex double ROP	TTGAATGATACCGATGTTTGTGATAGAATTGCTTGTGAATGTGAACACATTCACAA
C-rex single ROP	TTGAATGATACCGATGTTTGTGATAGAATTGCTTGTTAAATTGTTAACAT
C-rex double ROP	TTGAATGATACCGATGTTTGTGATAGAATTGCTTGTTAATGTTAAAACATTAACAA

Figure 5.13 Spacing of double ROP sites including cydA and wbIE found in Streptomyces. Double site sequences designed for G-Rex and C-Rex binding are based on cydA and wbIE. The main interaction points of Rex are indicated in yellow or pink.

Name	Sequence
	ROP1 ROP2 ROP1 ROP2
Streptomyces cydA Rex ROP site	TTGTGAATGTGAACGCGTTCACAA
Streptomyces wblE Rex ROP site	TTGTGAATGTGAACGCTTTCACGA
Streptomyces Rex consensus	TTGTGAACGCGTTCACAA TTGTGAACGCGTTCACAA
G-rex double binding site design	TTGTGAATGTGAACACATTCACAA
G-Rex consensus	TTGTGAAATAATGCACAA TTGTGAAATAATGCACAA
C-rex double binding site design	TTGTTAATGTTAAAACATTAACAA
C-Rex consensus	TTGTTAAATAATGAACAA TTGTTAAATAATGAACAA

Figure 5.14 Design of single and double C/G Rex operator sites. Red font indicates -35 or -10 regions. Blue highlight is extended -10 region. The main interaction points of Rex are indicated in yellow or pink.

5.5.3.2. Study of double and single ROP sites by expression analysis of gfp reporter

Following the cloning of the various promoter/operator fragments, activity could be qualitatively assessed using the downstream fluorescence reporter gene *gfp*. TAS plates were streaked with *G. thermoglucosidasius* 11955 Δ*rex* (pG1AK-sfGFP_oriT_G/C-Rex) strains, and grown at 60 °C overnight. A clear difference in *gfp* expression levels could be visually gauged and the fluorescence recorded (Table 5.3)

Name	*gfp* fluorescence strength
G-Rex Single ROP$^{G5/C14}$ site	++
G-Rex Single ROP$^{T5/A14}$ site	+++
G-Rex Double ROP$^{G5/C14}$ site	+
G-Rex Double ROP$^{T5/A14}$ site	++
G-Rex constitutive promoter	+++
C-Rex Single ROP$^{T5/A14}$ site	+
C-Rex Single ROP$^{G5/C14}$ site	+++
C-Rex Double ROP$^{T5/A14}$ site	-
C-Rex Double ROP$^{G5/C14}$ site	++
C-Rex constitutive promoter	+++

Table 5.3 Visual record of florescence caused by gfp expression. The variety of different gfp promoters are represented, including a strong constitutive promoter. Each promoter was cloned into both C-Rex and G-Rex expression vectors.

As expected the highest fluorescence was observed for the promoters that lacked ROP sites as well as those that included a non-cognate ROP site (e.g. G-Rex Single ROP$^{G5/C14}$ site). However, expression was reduced for both G-Rex and C-Rex constructs when the promoter contained single cognate ROP sites; ROP$^{G5/C14}$ and ROP$^{T5/A14}$, respectively. In agreement with the EMSA data, this suggests that *in vivo* G-Rex binds more strongly to ROP$^{G5/C14}$ than to ROP$^{T5/A14}$ and that G-Rex binds more strongly to ROP$^{T5/A14}$ than to ROP$^{G5/C14}$. Furthermore, fluorescence levels

were greatly diminished when cognate double ROP sites were used. This suggests tighter regulation at promoters with double ROP sites, which might be due to higher degree of Rex occupancy, thereby preventing RNA polymerase binding. A small decrease in fluorescence was observed between *G-Rex Signal* $ROP^{T5/A14}$ and *G-Rex Double* $ROP^{T5/A14}$ sites indicating that G-rex does bind to Crex sites, just less strongly than to the Grex consensus. Similar results were seen with C-Rex binding activity and $ROP^{G5/C14}$.

5.5.4. Response of Rex expression vectors to oxygen limitation

A key aim of this work was to generate a Rex-dependent expression vector that allowed increased gene expression in response to oxygen limitation. Experiments were therefore performed where *G. thermoglucosidasius* 11955 Δ*rex* (pG1AK-sfGFP_oriT_G/C-Rex) strains were grown in oxygen abundant, or limited conditions. A potential problem was that the gfp reporter requires molecular oxygen for the maturation of the fluorophore. Nonetheless gfp has been used for similar experiments in the past by allowing oxygen-dependent maturation following the inhibition of further protein synthesis (Roggiani & Goulian, 2015). In our hands, cultures grown under oxygen limitation had very low levels of fluorescence and attempts to allow chromophore maturation did not work. This might be because this gfp is a superfolder with increased resistance to denaturation, but becomes trapped in a non-fluorescent folding intermediate state (Dammeyer & Tinnefeld, 2012; Fisher & DeLisa, 2008). Therefore an alternative approach of using qRT-PCR to measure *gfp* transcript abundance was used; a benefit of this approach is that the expression of other known Rex targets within the genome could also be assessed.

To confirm previous visual qualitative assessment of *gfp* expression, strains were grown aerobically at 250 rpm in shaking flasks to OD_{600} 1.0-1.5, and a 1.5 ml sample was taken of each culture at time point 0 (oxygen abundant time point) and processed to isolate RNA and qRT-PCR performed. Results in Figures 5.15 and 5.16 indicated that *gfp* transcript abundance was reduced for nearly all G-Rex and C-Rex constructs, compared to the constitutive promoter (apart from C-Rex double ROPG5/C14 site, this was negligibly higher). A substantial decrease in *gfp* transcript abundance was measured between C-Rex single/double $ROP^{T5/A14}$ site

and C-Rex single/double $ROP^{G5/C1}$ indicating that C-Rex binds more strongly to $ROP^{T5/A14}$ than $ROP^{G5/C1}$ sites (Figure 5.15). This was not reciprocal in G-Rex constructs, and no significant difference in *gfp* transcript abundance was measured between $ROP^{T5/A14}$ and $ROP^{G5/C1}$ (Figure 5.16). *gfp* transcript abundance decreased in cognate double ROP $^{T5/A14}$ sites, but not double $ROP^{G5/C1}$ sites (compared to single ROPs) in both G-Rex and C-Rex constructs. The results suggest that the double ROP $^{T5/A14}$ site is more favourable to rex regulation, potentially permitting two Rex dimers to bind simultaneously, producing tighter control over gene expression. The viable qualitatively florescence results may differ slightly from qRT-PCR data potentially due to the fast turnover of mRNA, whereas gfp protein builds up over a far longer period of time.

To analyse if the activity of Rex is responsive to changing oxygen abundance in vivo, strains were grown as stated before but after time point 0 the remaining culture was used to fill 1.5 ml microfuge tubes to the top, so no air remained. Tubes were then incubated without shaking to produce oxygen limited growth conditions and harvested every 5 min for a total of 25 min, then processed to isolate RNA and qRT-PCR performed. *gfp* transcript abundance increased as time under oxygen limited conditions increased (Figure 5.17). This indicates both G-Rex and C-Rex sense a change in redox caused by limited oxygen and associate with NADH, causing disassociation with ROP sites. *gfp* expression levels were also overall higher in G-Rex constructs than C-Rex, suggesting that C-Rex binds more tightly to the ROP $^{T5/A14}$ site than G-Rex binds the $ROP^{G5/C1}$ site.

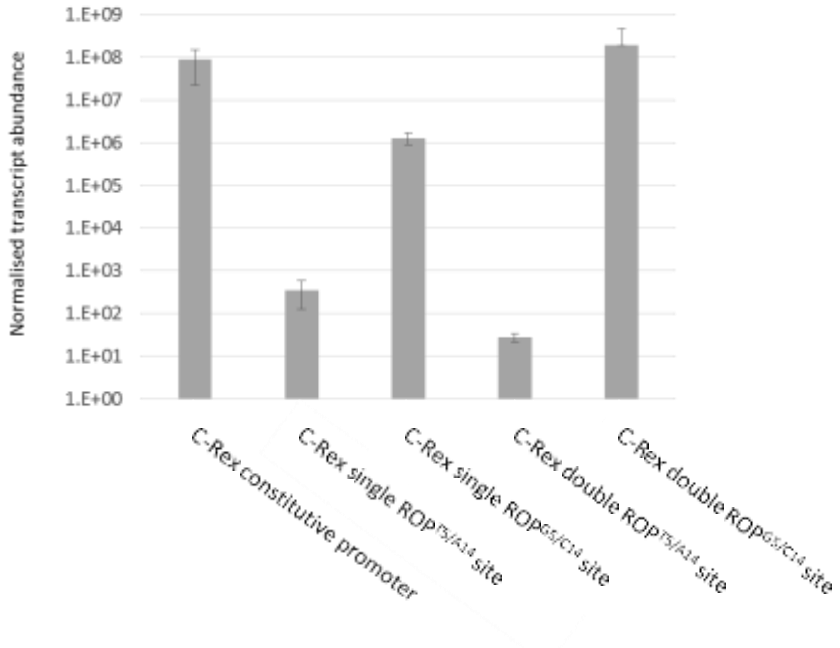


Figure 5.15 C-Rex binds to ROPT5/A14 site more strongly than to ROPG5/C14 site. qRT-PCR of *gfp* target with C-Rex, both double and single G/C rex ROP. As a control *gfp* was put under the control of a constitutive promoter. All samples are of time point 0, oxygen abundant growth conditions. Graph is a logarithmic scale. Data was normalised against *pgi* transcript abundance. Error bars are of standard deviation and data represents two biological repeats.

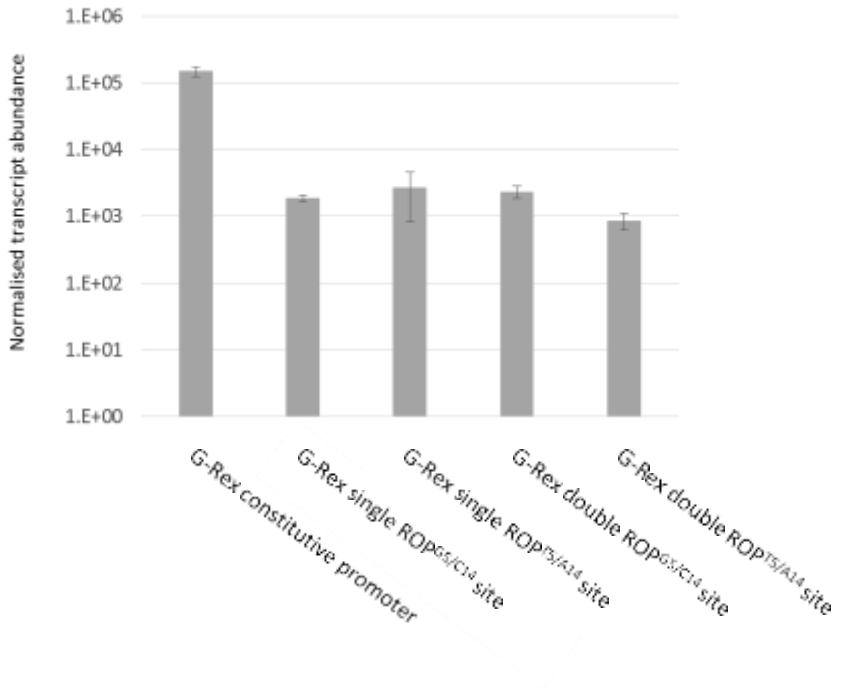


Figure 5.16 G-Rex binds to ROPG5/C14 site and ROPT5/A14 site equaly. qRT-PCR of gfp target with G-Rex, both double and single G/C rex ROP. As a control gfp was put under the control of a constitutive promoter. All samples are of time point 0, oxygen abundant growth conditions. Graph is a logarithmic scale. Data was normalised against pgi transcript abundance. Error bars are of standard deviation and data represents two biological repeats.

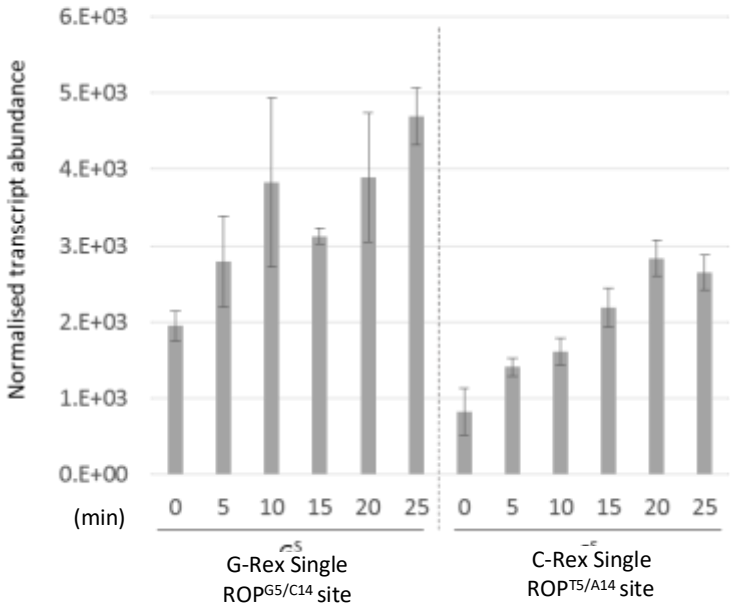


Figure 5.17 Oxygen limitation affects C-Rex and G-Rex binding. qRT-PCR of gfp with two strains, G. thermoglucosidasius 11955 Δrex (p11AK2_oriT_G/C-Rex) strains with over time points 0-25 min. At time point 0 samples were taken during oxygen abundant growth conditions. Samples 5-25 min were increasingly oxygen limited. Graph is a logarithmic scale. Data was normalised against pgi transcript abundance. Error bars are of standard deviation representative of two biological repeats.

5.5.5. Analysis of chromosomal Rex targets in response to oxygen limitation

These experiments were set up to establish if the *G. thermoglucosidasius* 11955 Δ*rex* (pG1AK-sfGFP_oriT_G/C-Rex) strains are capable of regulating known rex targets in vivo. Strains containing G-Rex Single ROPG5/C14 site and C-Rex Single ROPT5/A14 site were grown as stated before, in oxygen limited conditions. The transcript abundance of *adhE* and *ldh* were measured over a time course and the results presented in Figure 5.18 and 5.19. Results show that *adhE* becomes more highly expressed with increasing oxygen limitation in both strains, as expected (Figure 5.18). However, *adhE* becomes far overexpressed in the C-Rex strain, rising to a transcript abundance of over 1200, compared to G-Rex which reached a height of just over 200. This indicates that G-Rex binds more strongly to the genomic ROP

site than C-Rex, and is therefore able to inhibit expression to a greater degree. *Ldh* transcript abundance is more highly expressed in the C-Rex strain than G-Rex (Figure 5.19) indicating again that G-Rex is more able to bind *G. thermoglucosidasius* genomic ROP sites than C-Rex. These results demonstrate the orthogonal nature of C-Rex in *G. thermoglucosidasius* bacterium.

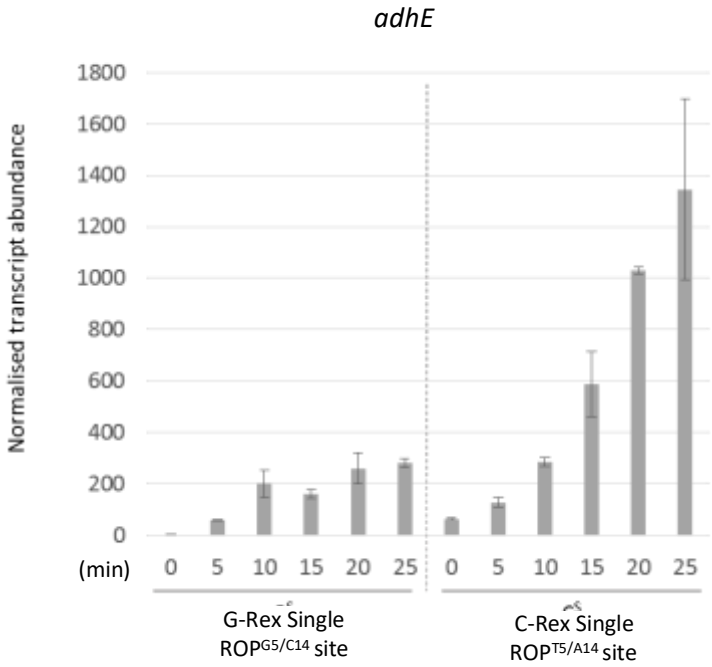


Figure 5.18 qRT-PCR of adh target with two strains, G/C-Rex_single ROP over time points 0-25 min. At time point 0 samples were taken during oxygen abundant growth conditions. Samples 5-25 min were increasingly oxygen limited. Data was normalised against pgi transcript abundance. Error bars are of standard deviation and data represents two biological repeats.

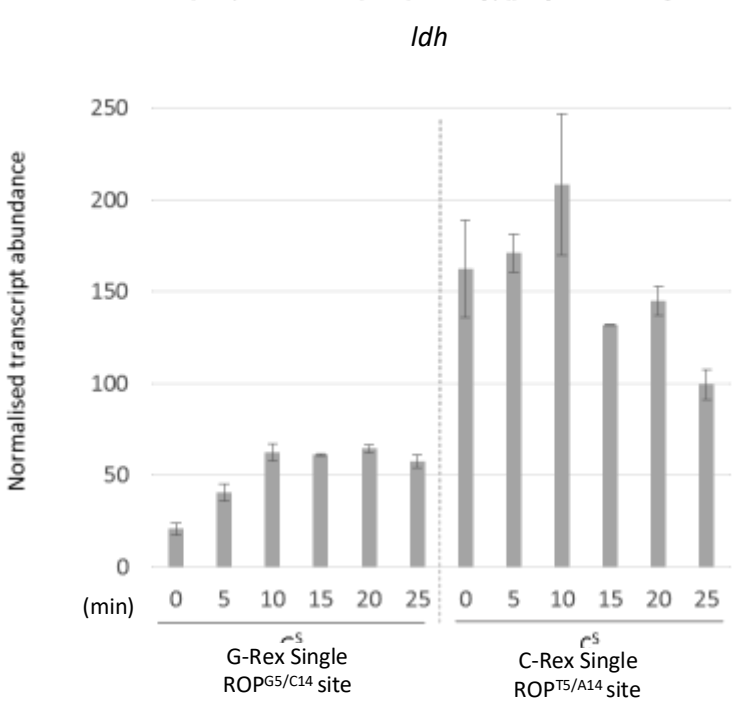


Figure 5.19 qRT-PCR of ldh target with two strains, G/C-Rex_single ROP over time points 0-25 min. At time point 0 samples were taken during oxygen abundant growth conditions. Samples 5-25 min were increasingly oxygen limited. Data was normalised against pgi transcript abundance. Error bars are of standard deviation and data represents two biological repeats.

5.5.6. Complementation of the rex mutant

Experiments were conducted to establish if the *G. thermoglucosidasius* 11955 Δ*rex* (pG1AK-sfGFP_oriT_G/C-Rex) strains are capable of Δ*rex* complementation. Cultures were grown of the 11955 WT, 11955 Δ*rex,* 11955 Δ*rex* (pG1AK-sfGFP_oriT_G-Rex Single ROP$^{G5/C14}$) and 11955 Δ*rex* (pG1AK-sfGFP_oriT_C-Rex Single ROP$^{T5/A14}$) in both oxygen abundant (T=0) and limited (T=30) conditions, processed to isolate RNA and qRT-PCR performed on known Rex targets. In both Figures 5.20 and 5.21 the expression profile of WT matches that of the G-Rex complement. Regulation of *adhE* and *ldh* was restored by the pG1AK-sfGFP_oriT_G-Rex Single ROPG5/C14 plasmid almost completely. A slight elevation of transcript abundance was present and this may be due to the difference in *rex* expression, as the WT was constituently expressed compared to the complement G-Rex being auto-regulated. These results indicate that the 11955 Δ*rex* (pG1AK-sfGFP_oriT_G-Rex Single ROP$^{G5/C14}$) Complement strain was capable of controlling *rex* target gene expression.

The transcript abundance of *adhE* and *ldh* Rex targets was extremely elevated in the 11955 Δ*rex* (pG1AK-sfGFP_oriT_C-Rex Single ROP$^{T5/A14}$) strain compared to WT, and matches closely to the Δ*rex* strain (Figure 5.20 and 5.21). These results indicate that C-Rex was not able to bind genomic ROP sites of *adhE* and *ldh* targets to regulate the expression of these genes. This demonstrates that C-Rex is orthogonal gene regulator in *G. thermoglucosidasius*.

adhE

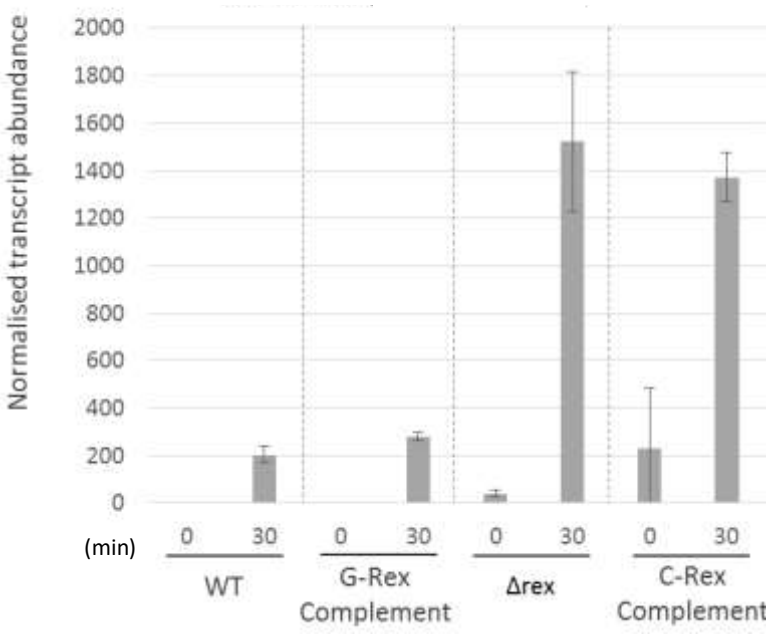


Figure 5.20 G-Rex, but not C-Rex complements a Δrex. qRT_PCR of adhE target using RNA RNA isolated from four strains Δrex, wild type 11955 and G. thermoglucosidasius 11955 Δrex (p11AK2_oriT_G/C-Rex) strains. Time point 0 was taken during arobic respiration and time pint 30 min was taken during oxygen limited growth conditions. Data was normalised against pgi transcript abundance. Error bars are of standard deviation and data represents two biological repeats

ldh

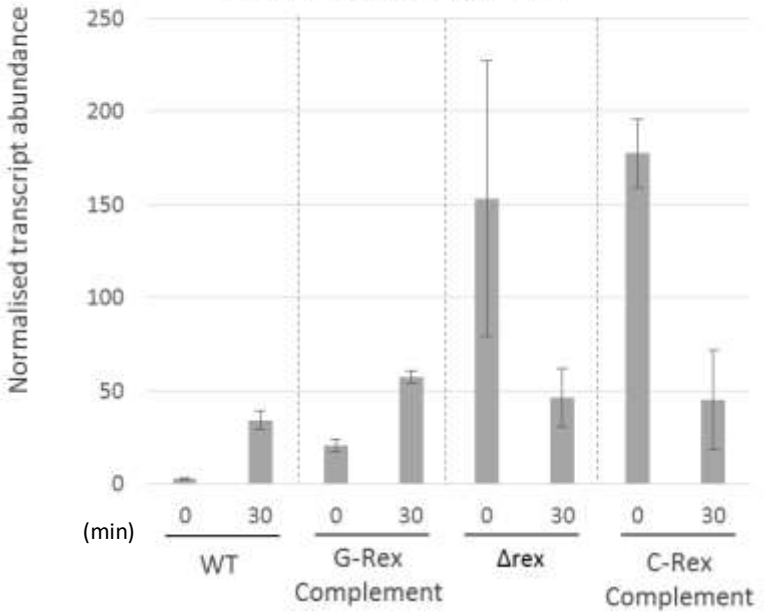


Figure 5.21 G-Rex, but not C-Rex complements a Δrex. qRT_PCR of ldh target using RNA isolated from four strains Δrex, wild type 11955 and G. thermoglucosidasius 11955 Δrex (p11AK2_oriT_G/C-Rex) strains. Time point 0 was taken during arobic respiration and time pint 30 min was taken during oxygen limited growth conditions. Data was normalised against pgi transcript abundance. Error bars are of standard deviation and data represents two biological repeats

fnr

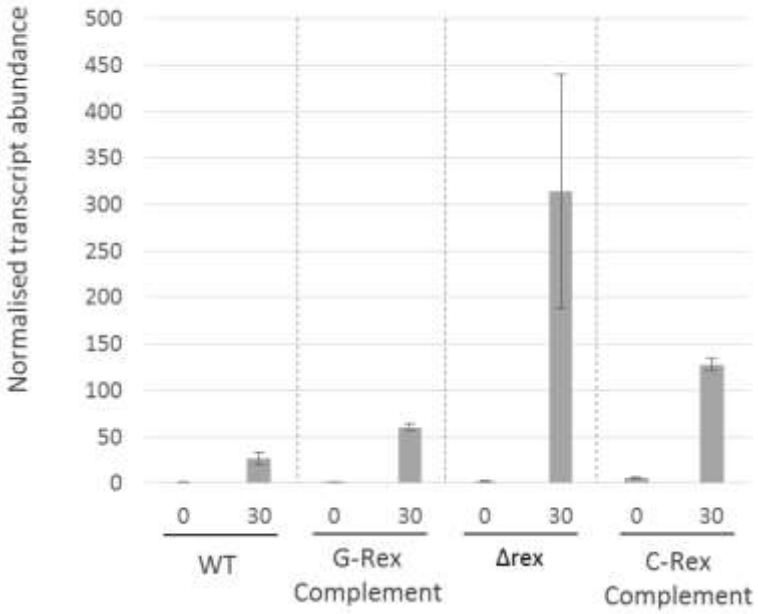


Figure 5.22 G-Rex, but not C-Rex complements a Δrex. qRT_PCR of *fnr* target using RNA isolated from four strains Δrex, wild type 11955 and G. thermoglucosidasius 11955 Δrex (p11AK2_oriT_G/C-Rex) strains. Time point 0 was taken during arobic respiration and time pint 30 min was taken during oxygen limited growth conditions. Data was normalised against pgi Error bars are of standard deviation and data represents two biological repeats.

5.6. Discussion and conclusions

Prior to this study *Clostridia spp.* had been predicted to have an alternative ROP sequence (TTGTTAANNNNTTAACAA) compared to that found in most other groups of bacteria (TTGTGAANNNNTTCACAA) (D. A. Ravcheev et al., 2012). This raised the possibility that a Rex orthologue would be able to discriminate between such ROP sites that differ only by two positions: $ROP^{G5/C14}$ and $ROP^{T5/A14}$, which might allow the development of orthologous expression vectors. However, since *Clostridium thermocellum* contains two paralogues, an initial aim was to determine whether the Cthe_0422 and Cthe_1798 paralogues were each active in DNA binding and redox sensing. Surprisingly, Cthe_0422 Rex protein did not bind to any of the ROP sites tested. This was particularly unusual as Cthe_0422 was predicted to be co-transcribed with bi-functional acetaldehyde-CoA/alcohol dehydrogenase, a gene that is part of the core Rex regulon across many taxonomic

groups. A similarly non-functioning Rex protein, TM1427. was previously found in *Thermotogales maritima*, which did not bind to any predicted ROP, and a functioning paralog TM0169 was also found (D. A. Ravcheev et al., 2012). The TM1427, TM0169, Cthe_0422 and Cthe_1798 protein sequences were aligned and it seems likely that TM1427 and Cthe_0422 are orthologues, more closely related to one another than to the paralogs they share a genome with.

Unlike Cthe_0422, Cthe_1798 did bind to all the EMSA probes, although there was a clear preference for $ROP^{T5/A14}$ compared to $ROP^{G5/C14}$. Furthermore it was found that G-Rex demonstrated more binding activity to $ROP^{G5/C14}$ than to $ROP^{T5/A14}$. These results agree with the predictions and suggested that the ROP of *Clostridia,* combined with the Rex paralog Cthe_1798 could be used as an alternative and possibly orthogonal redox responsive regulatory system in *G. thermoglucosidans.* Such a system would offer relatively similar and related regulation, yet independent. To test this in vivo C/G-Rex-based genetic circuit vectors were constructed and transformed into a *Δrex* background strain. A *gfp* reporter gene was put under the control of either signal/double $ROP^{T5/A14}$ or $ROP^{G5/C14}$ sites. Results indicated that C-Rex acted as an orthologue, inhibiting expression of *gfp* at $ROP^{T5/A14}$ far more than at $ROP^{G5/C14}$ sites. C-Rex also appeared to be capable of distinguishing single from double sites as *gfp* was more tightly inhibited at double ROP sites. Results indicated that G-Rex was not as able at distinguishing between single and double sites as *gfp* was expressed at near equal amounts. G-Rex did not demonstrate differential binding between $ROP^{T5/A14}$ and $ROP^{G5/C14}$ sites as *gfp* expression did not differ significantly. These results are promising as alterations such as the replacement of $ROP^{G5/C14}$ for $ROP^{T5/A14}$ sites, and the introduction of double sites may have no effect on G-Rex regulation, but C-Rex could differentiate between such sites readily. C-Rex and G-Rex were demonstrated to both be reactive to changing oxygen abundance, as *gfp* expression increased with longer duration of oxygen starvation, and are therefore likely to be able to sense cell redox.

Genomic Rex targets *adhE* and *ldh* were also analysed and it was found that expression of G-Rex by pG1AK-sfGFP_oriT_G-Rex plasmids was capable of complementing the *G. thermoglucosidasius* 11955 *rex* mutant strain. Figure 5.6.1 illustrates how regulators Rex, Fnr and ResDE are thought to interact after conducting this research. C-Rex expression in *G. thermoglucosidasius* 11955 *rex* mutant strain was not, however, capable of

complementing, and the expression of Rex targets *adhE* and *ldh* mimicked that of the mutant, thereby demonstrating again that C-Rex is a likely orthologue.

The genetic circuit vectors could be diversified further to experiment with the expression of Rex. During this project we demonstrated that *G. thermoglucosidasius* wild type Rex expression is likely constitutive. To optimise previous research and generate orthogonal Rex systems, a major focus of research should be of the control over the expression of the orthogonal Rex itself, especially once integrated into the genome. Inducible promoter systems offer robust control, however the cost of adding large quantities to a bioreactor quickly becomes expensive and is not cost effective. Autoregulation proved to be a successful mechanism to prevent the over expression of Rex when on multicopy plasmids and utilising autoregulation is likely to be the most robust, dynamic and cost effective mechanism of orthogonal C-Rex control. If the strength of the promoter and the tightness of Rex regulation can be balanced, as it was in the design of pG1AK-sfGFP_oriT_G/C-Rex plasmids, then this may offer the best solution. Perhaps research can be accelerated by studying the similarities between previously demonstrated Rex autoregulation (Bitoun et al., 2012; Brekasis & Paget, 2003) and also those predicted to be (D. A. Ravcheev et al., 2012).

The use of orthogonal systems to divide gene expression into independent modules can aid in achieving predictable outcomes and optimisation of product yield. Cells and biological systems are interconnected and any change in one pathway is likely to have unpredictable functional or temporal effects on another (Kavscek, Strazar, Curk, Natter, & Petrovc, 2015). For this reason orthogonality is not absolute, but instead a sliding scale. Many varieties of orthogonal systems have been developed, some centring on generating promoter libraries but others are far more ambitious, and some aim expand the genetic code (Blount, Weenink, Vasylechko, & Ellis, 2012; Gilman & Love, 2016; Maranhao & Ellington, 2016). The development of an orthogonal system makes it possible to engineer the cell metabolism. To gain more control over the expression of key enzymes, such as those specifically involving fermentation is a powerful way to divert metabolic flux towards the product of interest, ethanol. Here we have demonstrated the orthogonal potential of C-Rex and expanded the repertoire of ROP sites that could be used in such a system. Further work could explore how C-Rex and G-Rex might compete for $ROP^{T5/A14}$

sites. This would help to develop a more sophisticated understanding of how C-Rex could be used in an orthogonal system. It seems likely that C-Rex would 'override' G-Rex for ROP$^{T5/A14}$ sites but not ROP$^{G5/C14}$ sites and specific genes could be targeted for tighter regulation. An understanding of the exact redox sensing properties of both G-Rex and C-Rex would enable prediction as to what cellular and environmental conditions would result in association, or dissociation with ROP sites, giving greater insight into how the orthogonal system could be implemented in large bioreactors to trigger gene expression. The orthogonal C-Rex system analysed during this project has wider applications beyond *G. thermoglucosidasius*, and could be applied in any of the many bacteria conforming to the typical ROP$^{G5/C14}$ consensus sequence (D. A. Ravcheev et al., 2012).

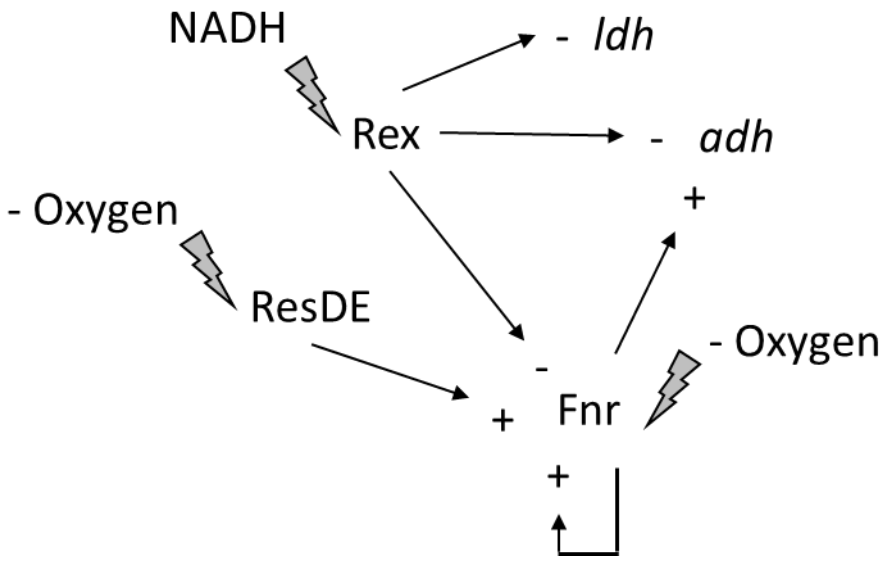


Figure 5.6.1 Gene regulatory network. Rex binding to ROP sites is induced by NADH, when bound it inhibits expression of *ldh*, *adh* and *fnr*. Fnr competes with Rex for ROP sites and positively regulates expression of *adh,* and auto regulates itself. ResDE is active under oxygen limited conditions and induces expression of *fnr*.

6. Chapter VI: Discussion

Overview

To produce biofuel in the most efficient way possible it is important to fully understand the regulators controlling fermentation and anaerobic metabolism, and the factors triggering their activation. Rex is a key regulator of fermentation genes, able to sense redox balance within the cell. By understanding and manipulating Rex it is anticipated that fermentation can be more precisely controlled and the expression of genes optimised for industrial bioreactors. Rex evolved as a mechanism to cope with redox stress caused, for example, by oxygen deprivation, often faced by soil-dwelling bacteria. Rex is found in many Gram-positive bacteria and predictions of regulons based on the conserved ROP operator have revealed that certain "core" genes are particularly prevalent. One example is the bifunctional acetaldehyde-CoA/alcohol dehydrogenase (ADH) (D. a Ravcheev et al., 2012). ADH converts acetyl-coenzyme A to ethanol via the intermediate acetaldehyde, and oxidises two NADH molecules to NAD+ in the process, thereby helping to maintain the redox balance. It is vital for the cell to recycle NADH in the absence of oxygen, as the glycolytic pathway depends upon NAD^+ to generate ATP. Only during a build-up of NADH (usually occurring during oxygen limited growth conditions) would a less efficient form of energy production, fermentation, be necessary. Rex is a direct sensor of NADH/NAD+ ratio (Brekasis & Paget, 2003) and is therefore uniquely placed to regulate fermentation and anaerobic respiration. Much work has been carried out in industrially important *Clostridium* to this effect (Wietzke & Bahl, 2012b; L. Zhang et al., 2014). In *C. acetobutylicum* Rex was shown to control *adhE* and *ldh* genes as well as the *bca* operon encoding genes important for the C_4 metabolic pathway (Wietzke & Bahl, 2012). Predicted Rex target genes in a range of *Clostridium* species include *adhA*, *adhE*, *crt-bcd-etfBA-hbd* and *ptb-buk* operon encoding enzymes important for alcohol synthesis (L. Zhang et al., 2014). Other predicted Rex controlled genes included those involved in sulphite and nitrate reduction, hydrogen production, fermentation, the TCA cycle, and NAD biosynthetic genes, and some targets were confirmed by EMSA (L. Zhang et al., 2014). This project focused on *G. thermoglucosidans,* which also has great industrial potential, but has an underdeveloped genetic manipulation toolkit. The aim was to

investigate the *G. thermoglucosidans* Rex regulon, explore the use of a clostridial Rex as an orthogonal regulator for use in *Geobacillus*, and to construct new tools for genetic manipulation.

6.1. General discussion and future directions

The Rex regulon of *G. thermoglucosidans* was investigated by Chip-seq and qRT-PCR and was found to include genes that encode enzymes involved in fermentation, anaerobic respiration, and carbon metabolism, as well as several genes with no known function. ChIP-seq also identified a number of binding sites that are internal to genes; while Rex is unlikely to regulate these genes it is conceivable that Rex regulates sRNAs that are encoded with the genes. Interestingly, Rex itself was not auto-regulated, unlike the case in *Streptomyces coelicolor* (Brekasis & Paget, 2003), but rather appears to be unaffected by oxygen limitation.

Much work has been focused on identifying members of the rex regulon, however these have focused on the use of bioinformatics and a consensus sequence. The study conducted here is the first to apply ChIP-sequencing to identify Rex binding locations at a gnomically global level. Rex was previously shown to be important for the control of fermentation genes in a variety of Gram positive bacteria including *Staphlococcus aureus,* strains of *Clostridia* and *Bacillus* (Pagels et al., 2010; D. a Ravcheev et al., 2012; Wang, Bauer, Rogstam, Linse, Logan, & von Wachenfeldt, 2008; L. Zhang et al., 2014). The predicted conserved Rex regulon across 11 taxonomic groups of bacteria included *adhE, adhA, adhB, ldh, pflAB, dhaT, butA* and *bcd-hbd* (D. a Ravcheev et al., 2012). Some of these fermentative Rex targets such as *ldh* and *adhE* have been confirmed experimentally in *G. thermoglucosidans* during this study. The *G. thermoglucosidans Δrex* strain constructed during this study was predicted to produce high amounts of ethanol, however this was not what our results showed. A *Δrex* strain of *C. acetobutylicum* had been demonstrated to increase yields of butanol and ethanol, and to start production much earlier than in a WT strain (Wietzke & Bahl, 2012). Deregulation in our *G. thermoglucosidans Δrex* strain did lead to an increases in yield and earlier production of one fermentation product, lactate. This outcome could possibly be improved by

the generation of a double mutant (*Δrex, Δldh*) to redirect flux towards the ADH pathway. A *G. thermoglucosidans Δldh* has been previously analysed for the production of metabolites and it was found that although ethanol yields increased, the remaining pathways were not able to utilise the increase in carbon flux and pyruvate built up (Cripps et al., 2009). To deal with this the PDH gene was upregulated which resulted in almost a doubling in ethanol yield (Cripps et al., 2009). Our results indicate that Rex deregulation results in increased carbon flux towards LDH, however if LDH was deleted this may have the effect of redirecting it again towards ADH. Since *adhE* is overexpressed in a *Δrex*, this strain might be able to cope with an in increased carbon flux well, especially if combined with an upregulated PDH.

Genes involved in carbon metabolism and glycolysis such as *gldA, gap, pgk, tpi* and *fba* have previously been described as Rex targets (D. a Ravcheev et al., 2012). We discovered that Rex binds upstream of an operon encoding 6-phosphofructo kinase (*pfkA*) and pyruvate kinase (*pyK*), key enzymes in the glycolytic pathway. PFK is particularly important as it controls the rate limiting step in glycolysis and therefore plays a central role in carbon flux. In *E. coli* PFK is controlled in response to levels of the product fructose 6-phosphate, by the central carbon pathways regulator Cra, and in Bacillus CcpA (Shimada, Yamamoto, & Ishihama, 2011; Tobisch et al., 1999). In addition the *adhE* gene in *E. coli* was found to also be controlled by Cra (Mikulskis, Aristarkhov, & Lin, 1997). PYK is controlled by Cra and CcpA, and in *Bacillus* PYK has been shown to be involved in acetate production and a Δpyk demonstrated significant reduced in organic acids (Fry et al., 2000). The overall role of CcpA includes the activation of glycolysis in *B. subtitlis* and response to changing conditions by directing carbon overflow metabolism. Rex may control *pfkA* and *pyK* expression in conjunction with CcpA to help to direct flux and address redox imbalance.

Redox imbalance is also adjusted by the expression of *ndh*, shown to be a Rex target in *G. thermoglucosidans* during this study. *ndh* is a NADH dehydrogenase and in *B. subtilis* is also regulated by Rex forming a regulatory loop that avoids large fluctuations in the ratio of $NADH/NAD^+$ within the cell (Gyan et al., 2006).

This study has also demonstrated that Rex binds upstream of nitrite reductase (*narK*) and a nitrite transporter (*nirC*) in *G. thermoglucosidans*; both genes were previously identified as putative Rex targets in a variety of bacteria (D. a Ravcheev et al., 2012). A number of other anaerobic, nitrite reduction genes such as *nark* and *nirK* were also identified as potentially controlled by Rex during ChIP-seq analysis. NirK is a copper containing nitrite reductase found in several denitrifying bacteria, and NarK a nitrite extrusion protein used in protecting cells from stress of nitrite within the cytoplasm, also regulated by Fnr in *B. subtitlis* (Bueno et al., 2012; Rodionov, Dubchak, Arkin, Alm, & Gelfand, 2005). The data suggests that Rex might control nitrite dissimilation, which could be a way of addressing redox balance. An alternative possibility, given the close location of *fnr* to *nark* and *nirK,* and the fact that these enzymes are controlled by Fnr in Bacillus, is that the binding seen is spurious, possibly caused by the similarity between binding sites. *B. subtilis* Fnr consensus binding site is very similar (TGTGA-n6-TCACT) to the Rex consensus (TTGTG-n8-CACAA) (Reents et al., 2006; Wang, Bauer, Rogstam, Linse, Logan, & von Wachenfeldt, 2008). Is it also possible that Rex and Fnr have overlapping specificity and control the some of the same genes. It could be that Rex modulates expression under anaerobic conditions, depending on redox state of the cell, and Fnr induces gene expression in an oxygen limited environment. Data to support this demonstrated that neither *adhE* nor *ldhA* are constitutively expressed in a rex mutant, indicating other regulators induce expression of these genes, regulators such as Fnr, or ResDE. A Rex binds upstream of other genes important for nitrate reductase including *fnr/ modCD* operon and a *mobA* gene, discovered during this study. Mob converts molybdopterin to molybdopterin guanine dinucleotide which is require for nitrate reductase and *rex* shares an operon with a gene involved in molybdopetrin production (*moaC*). Data indicates that Rex is a regulator of not only genes involved in anaerobic respiration and nitrate reduction, but also an important inducer of anaerobic respiration, the regulator Fnr. ChIP-seq revealed that *fnr* does appear to be directly controlled and that Rex deregulation did have a significant impact on *fnr* expression, but this was confined to oxygen limited growth conditions. Interestingly the expression pattern in a *Δrex* strain mimicked that of *adhE*, however other Rex targets such as *ldh* and *ndh* did not match this expression profile. The differing expression profiles of genes such as *ldh* and *adhE* in

a deregulated *Δrex* strain indicated that other regulators may also act upon Rex targets. Perhaps future research should focus on ascertaining if the regulators Rex and Fnr compete (or bind together at double ROP sites) to fine-tune gene expression in response to environmental oxygen, and cellular redox conditions. Experiments should focus on whether Rex may act to dampen Fnr gene expression induction, especially in relation to the *adhE* gene.

An alignment of G-Rex binding sites identified using Chip-seq data allowed the generation of a G-Rex ROP consensus sequence with the most conserved residues including the GTG-n8-CAC motif. The Rex binding site is extremely well conserved among Gram-positive bacteria and extends to the remaining residues in the 16 bp site. However, the prediction that clostridial Rex proteins bind with an altered specificity (GTT n8 AAC) was exploited here to develop a *Clostridium thermocellum* Rex (C-Rex)-based expression system that might be orthogonal in *Geobacillus*. The data presented indicated that C-Rex was expressed but did not restore regulation in a *Δrex* strain, which suggested that it might not be binding to host Rex binding sites and therefore might be orthogonal.

An orthogonal regulator opens up the possibility for control over fermentation regulation. ROP sites of specific genes could be modified and optimised for C-Rex binding, so that altering C-Rex levels would give control over these genes or any new genes which might be integrated. C-Rex potentially has a different affinity for NADH and genes could be switched on or off at slightly altered redox potentials. In addition to this a library of ROP sites, including single and double sites, could be produced that range in strength. Approaches in the past have generated promoter libraries to construct orthogonal genetic circuits (Stanton et al., 2014). This would allow a form of temporal control over the expression of genes in response to redox. We have shown that the design of an auto regulating Rex expression system functions efficiently and can dynamically respond to changing conditions. Other systems have utilised dynamic regulatory networks to optimise the production of fatty acid biosynthesis (Xu, Li, Zhang, Stephanopoulos, & Koffas, 2014). Inducible promoters are not cost effective in an industrial setting, but auto regulation does provide a real alternative.

C-Rex was proven to exhibit a greater affinity for an alternative ROP site than to the predicted *G. thermoglucosidans* consensus sequence. G-Rex did however bind to both its own, and to C-Rex ROP sites. These experiments were conducted in vitro and vivo and C-Rex was demonstrated to respond not only to NADH levels but also NAD^+ (which enhanced binding to ROP sites). Experiments also included gene regulatory networks that incorporated auto-regulated G-rex and C-Rex, which was successful in dynamically responding to changing environmental conditions (oxygen abundance). The orthogonal system could be used in the future to optimise the expression of particularly significant genes, by replacing the original ROP site for that of *C. thermocellum,* thereby creating an independent redox responsive regulatory system separate form G-Rex. In addition to the use of C-Rex as an expression control aid, double ROP sites were demonstrated to lead to tighter regulation than single sites. In combination, orthogonal C-Rex, autoregulation and double ROP sites could expand the repertoire of the rex regulatory system and lead to precise modulation of fermentation.

The cloning and transfer of large (~80kb) biosynthetic clusters remains challenging, even in model organisms (Nah, Woo, Choi, & Kim, 2015). Optimisation of DNA transfer will be important for the genetic engineering of *G. thermoglucosidans* genome. Before work began to characterise the Rex regulon or explore C-rex as an orthogonal regulator, genetic manipulation tools were constructed. Much work was done to generate an integrative vector system, including the construction of an integrative vector, however this was unsuccessful. The placement of an attachment site within the *G. thermoglucosidans* genome proved difficult due to the combining of this aim with another, for the production of a counter selection system. Future work would build upon the pyrE counter selection system to construct a ΔpyrE ΔpryR double mutant strain for counter selection to be successful. The counter selection system could then be used to insert an attachment site into the *G. thermoglucosidans* genome. An alternative integrative vector system could utilise the φBT1 phage attachment sites instead of φC31 (Gregory, Till, & Smith, 2003).

Functional selection markers are also limiting, as currently only two thermostable antibiotic markers are available for use in *G. thermoglucosidans* (Reeve et al., 2016). The development of additional antibiotic marks needs both a thermostable antibiotic with a long half-life, and also an antibiotic resistance gene active under

high growth temperatures of 60°C. This is difficult to achieve: Kanamycin is a commonly used antibiotic as it has good thermostability but standard mesophilic kanamycin resistance genes encode enzymes which are not thermostable. Therefore several thermostable resistant derivatives were developed by forced evolution (H. Liao, McKenzie, & Hageman, 1986). Gentamicin, a member of the aminoglycosides class of antibiotics, is effective against Gram positive bacteria and relatively thermostable and often used as a bacterial agent in cell culture (Fischer, 1975). A gentamicin resistance has been developed for use in cloning and may not need to undergo forced evolution to work in *G. thermoglucosidans* (Bian, Fenno, & Li, 2012; Poggi, Oliveira de Giuseppe, & Picardeau, 2010).

Prior to this project DNA transformation was inefficient and required specialist equipment, hampering progress. A conjugative method was developed (including strains and plasmids) which facilitated our research. Other research had also produced a working conjugative system for use with *Geobacillus kaustophilus*, which was then expended and tested in other *Bacillus* and *Geobacillus* species (Suzuki & Yoshida, 2012; Tominaga, Ohshiro, & Suzuki, 2016). Non-replicating conjugative vectors were constructed and used for gene disruption (to produce a *Δrex* strain) and gene replacement (3xFLAG tagged Rex Strain). Optimisation of the conjugation protocol led to the use of magnesium supplementation and *dcm-* donor *E. coli* strain. Results indicated that methylation did play a role in transformation frequency, but whether or not *dcm* is the only system to be important is yet to be discovered. Throughout experimentation with *G. thermoglucosidans* it become clear that it is sensitive to cold shock at room temperature and that minimising this leads to enhanced transformation. A modular origin of transfer component was also constructed for use with the modular plasmid toolkit set (Reeve et al., 2016), which was utilised during the analysis of the orthogonal C-Rex protein.

This thesis has made a novel contribution to the understanding of the Rex regulon, and significantly expanded the genetic engineering tools that enable strain development. The work presented has demonstrated the potential of Rex to be used in an othoganol system for second generation biofuel production. Many more questions have arisen during the course of our research that if pursued could potentially aid the progress of *G. thermoglucosidans,* or other gram negative bacteria, to be successful biocommodities producers.

7. Bibliography

Abdel-Rahman, M. A., Tashiro, Y., & Sonomoto, K. (2010). Lactic acid production from lignocellulose-derived sugars using lactic acid bacteria: overview and limits. *Journal of Biotechnology*, *156*(4), 286–301. doi:10.1016/j.jbiotec.2011.06.017

Alexeeva, S., Hellingwerf, K. J., & Teixeira de Mattos, M. J. (2003). Requirement of ArcA for Redox Regulation in Escherichia coli under Microaerobic but Not Anaerobic or Aerobic Conditions. *Journal of Bacteriology*, *185*(1), 204–209. doi:10.1128/JB.185.1.204-209.2003

Alting-Mees, M. A., & Short, J. M. (1989). pBluescript II: gene mapping vectors. *Nucleic Acids Research*, *17*(22), 9494. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=335171&tool=pmcentrez&rendertype=abstract

Alvira, P., Tomás-Pejó, E., Ballesteros, M., & Negro, M. J. (2010). Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. *Bioresource Technology*, *101*(13), 4851–61. doi:10.1016/j.biortech.2009.11.093

Armaroli, N., & Balzani, V. (2011). The legacy of fossil fuels. *Chemistry, an Asian Journal*, *6*(3), 768–84. doi:10.1002/asia.201000797

Baruah, A., Lindsey, B., Zhu, Y., Michiko, M., Baruah, A., Lindsey, B., … Nakano, M. M. (2004). Mutational Analysis of the Signal-Sensing Domain of ResE Histidine Kinase from Bacillus subtilis Mutational Analysis of the Signal-Sensing Domain of ResE Histidine Kinase from Bacillus subtilis. doi:10.1128/JB.186.6.1694

Bateman, J. R., & Wu, C. (2008). A simple polymerase chain reaction-based method for the construction of recombinase-mediated cassette exchange donor vectors. *Genetics*, *180*(3), 1763–6. doi:10.1534/genetics.108.094508

Bechhofer, D. H., & Wang, W. (1998). Decay of ermC mRNA in a polynucleotide phosphorylase mutant of Bacillus subtilis. *Journal of Bacteriology*, *180*(22), 5968–77. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=107672&tool=pmcentrez&rendertype=abstract

Bekker, M., Alexeeva, S., Laan, W., Sawers, G., Teixeira de Mattos, J., & Hellingwerf, K. (2010). The ArcBA two-component system of Escherichia coli is regulated by the redox state of both the ubiquinone and the menaquinone pool. *Journal of Bacteriology*, *192*(3), 746–54. doi:10.1128/JB.01156-09

Berks, B. C., Sargent, F., & Palmer, T. (2000). The Tat protein export pathway. *Molecular Microbiology*, *35*(2), 260–274. doi:10.1046/j.1365-2958.2000.01719.x

Bhalla, A., Bansal, N., Kumar, S., Bischoff, K. M., & Sani, R. K. (2013). Improved lignocellulose conversion to biofuels with thermophilic bacteria and thermostable enzymes. *Bioresource Technology*, *128*, 751–9. doi:10.1016/j.biortech.2012.10.145

Bian, J., Fenno, J. C., & Li, C. (2012). Development of a modified gentamicin resistance cassette for genetic manipulation of the oral spirochete Treponema denticola. *Applied and Environmental Microbiology*, *78*(6), 2059–62. doi:10.1128/AEM.07461-11

Bidnenko, V., Shi, L., Kobir, A., Ventroux, M., Pigeonneau, N., Henry, C., … Mijakovic, I. (2013). Bacillus subtilis serine/threonine protein kinase YabT is involved in spore development via phosphorylation of a bacterial recombinase. *Molecular Microbiology*, *88*(5), 921–35. doi:10.1111/mmi.12233

Bierman, M., Logan, R., O'Brien, K., Seno, E. T., Rao, R. N., & Schoner, B. E. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp. *Gene*, *116*(1), 43–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1628843

Bitoun, J. P., Liao, S., Yao, X., Xie, G. G., & Wen, Z. T. (2012). The Redox-Sensing Regulator Rex Modulates Central Carbon Metabolism, Stress Tolerance Response and Biofilm Formation by Streptococcus mutans. *PLoS ONE*, *7*(9). doi:10.1371/journal.pone.0044766

Bittner, F., & Mendel, R. R. (2010). Cell biology of molybdenum. *Plant Cell Monographs*, *17*, 119–143. doi:10.1007/978-3-642-10613-2_6

Blount, B. a, Weenink, T., Vasylechko, S., & Ellis, T. (2012). Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. *PloS One*, *7*(3), e33279. doi:10.1371/journal.pone.0033279

Boeke, J. D., La Croute, F., & Fink, G. R. (1984). A positive selection for mutants lacking orotidine-5′-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Molecular and General Genetics MGG*, *197*(2), 345–346. doi:10.1007/BF00330984

Bogner, J., Pipatti, R., Hashimoto, S., Diaz, C., Mareckova, K., Diaz, L., … Gregory, R. (2008). Mitigation of global greenhouse gas emissions from waste: conclusions and strategies from the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report. Working Group III (Mitigation). *Waste Management & Research*, *26*(1), 11–32. doi:10.1177/0734242X07088433

Bolhuis, A., Sorokin, A., Azevedo, V., Ehrlich, S. D., Braun, P. G., deJong, A., … Maarten van Dijl, J. (1996). Bacillus subtilis can modulate its capacity and specificity for protein secretion through temporally controlled expression of the

sipS gene for signal peptidase I. *Molecular Microbiology*, *22*(4), 605–618. doi:10.1046/j.1365-2958.1996.d01-4676.x

Borisov, V. B., Gennis, R. B., Hemp, J., & Verkhovsky, M. I. (2011). The cytochrome bd respiratory oxygen reductases. *Biochimica et Biophysica Acta - Bioenergetics*, *1807*(11), 1398–1413. doi:10.1016/j.bbabio.2011.06.016

Borisov, V. B., & Verkhovsky, M. I. (2009). Oxygen as Acceptor. *EcoSal Plus*, *3*(2). doi:10.1128/ecosalplus.3.2.7

Brekasis, D., & Paget, M. S. B. (2003). A novel sensor of NADH / NAD + redox poise in Streptomyces coelicolor A3 (2). *EMBO Journal*, *22*(18).

Brumm, P. J., Land, M. L., & Mead, D. A. (2015). Complete genome sequence of Geobacillus thermoglucosidasius C56-YS93, a novel biomass degrader isolated from obsidian hot spring in Yellowstone National Park. *Standards in Genomic Sciences*, *10*(1), 73. doi:10.1186/s40793-015-0031-z

Bueno, E., Mesa, S., Bedmar, E. J., Richardson, D. J., & Delgado, M. J. (2012). Bacterial adaptation of respiration from oxic to microoxic and anoxic conditions: redox control. *Antioxidants & Redox Signaling*, *16*(8), 819–52. doi:10.1089/ars.2011.4051

Centre, H., & Office, M. (2013). International symposium on the stabilisation of greenhouse gases, (February), 1–16. Retrieved from papers2://publication/uuid/F9A1E5A7-A145-4BBB-8A57-5E3BABF936D7

Chen, G., Kumar, A., Wyman, T. H., & Moran, C. P. (2006). Spo0A-dependent activation of an extended -10 region promoter in Bacillus subtilis. *Journal of Bacteriology*, *188*(4), 1411–8. doi:10.1128/JB.188.4.1411-1418.2006

Chen, J., Shen, J., Solem, C., & Jensen, P. R. (2015). A New Type of YumC-Like Ferredoxin (Flavodoxin) Reductase Is Involved in Ribonucleotide Reduction. *mBio*, *6*(6), e01132–15. doi:10.1128/mBio.01132-15

Chen, J., Zhang, Z., Zhang, C., & Yu, B. (2015). Genome sequence of Geobacillus thermoglucosidasius DSM2542, a platform hosts for biotechnological applications with industrial potential. *Journal of Biotechnology*, *216*, 98–9. doi:10.1016/j.jbiotec.2015.10.002

Cripps, R. E., Eley, K., Leak, D. J., Rudd, B., Taylor, M., Todd, M., … Atkinson, T. (2009). Metabolic engineering of Geobacillus thermoglucosidasius for high yield ethanol production, *11*, 398–408. doi:10.1016/j.ymben.2009.08.005

Cruz Ramos, H., Boursier, L., Moszer, I., Kunst, F., Danchin, A., & Glaser, P. (1995). Anaerobic transcription activation in Bacillus subtilis: identification of distinct FNR-dependent and -independent regulatory mechanisms. *The EMBO Journal*, *14*(23), 5984–94. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=394718&tool=pmcentrez&rendertype=abstract

Cruz Ramos, H., Hoffmann, T., Marino, M., Nedjari, H., Presecan-Siedel, E., Dreesen, O., … Jahn, D. (2000). Fermentative Metabolism of Bacillus subtilis: Physiology and Regulation of Gene Expression. *Journal of Bacteriology*, *182*(11), 3072–3080. doi:10.1128/JB.182.11.3072-3080.2000

Dammeyer, T., & Tinnefeld, P. (2012). Engineered fluorescent proteins illuminate the bacterial periplasm. *Computational and Structural Biotechnology Journal*, *3*, e201210013. doi:10.5936/csbj.201210013

Dong, H., & Zhang, D. (2014). Current development in genetic engineering strategies of Bacillus species. *Microbial Cell Factories*, *13*, 63. doi:10.1186/1475-2859-13-63

Durand, S., Braun, F., Lioliou, E., Romilly, C., Helfer, A.-C., Kuhn, L., … Condon, C. (2015). A nitric oxide regulated small RNA controls expression of genes involved in redox homeostasis in Bacillus subtilis. *PLoS Genetics*, *11*(2), e1004957. doi:10.1371/journal.pgen.1004957

Eijlander, R. T., Jongbloed, J. D. H., & Kuipers, O. P. (2009). Relaxed specificity of the Bacillus subtilis TatAdCd translocase in Tat-dependent protein secretion. *Journal of Bacteriology*, *191*(1), 196–202. doi:10.1128/JB.01264-08

Esbelin, J., Armengaud, J., Zigha, A., & Duport, C. (2009). ResDE-dependent regulation of enterotoxin gene expression in Bacillus cereus: evidence for multiple modes of binding for ResD and interaction with Fnr. *Journal of Bacteriology*, *191*(13), 4419–26. doi:10.1128/JB.00321-09

Fabret, C., Ehrlich, S. D., & Noirot, P. (2002). A new mutation delivery system for genome-scale approaches in Bacillus subtilis. *Molecular Microbiology*, *46*(1), 25–36. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12366828

Farr, G. A., Oussenko, I. A., & Bechhofer, D. H. (1999). Protection against 3'-to-5' RNA decay in Bacillus subtilis. *Journal of Bacteriology*, *181*(23), 7323–30. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=103696&tool=pmcentrez&rendertype=abstract

Fischer, A. B. (1975). Gentamicin as a bactericidal antibiotic in tissue culture. *Medical Microbiology and Immunology*, *161*(1), 23–39. doi:10.1007/BF02120767

Fisher, A. C., & DeLisa, M. P. (2008). Laboratory evolution of fast-folding green fluorescent protein using secretory pathway quality control. *PloS One*, *3*(6), e2351. doi:10.1371/journal.pone.0002351

Fong, R., Hu, Z., Hutchinson, C. R., Huang, J., Cohen, S., & Kao, C. (2007). Characterization of a large, stable, high-copy-number Streptomyces plasmid that requires stability and transfer functions for heterologous polyketide overproduction. *Applied and Environmental Microbiology*, *73*(4), 1296–1307. doi:10.1128/AEM.01888-06

Fry, B., Zhu, T., Domach, M. M., Koepsel, R. R., Phalakornkule, C., & Ataai, M. M. (2000). Characterization of growth and acid formation in a Bacillus subtilis pyruvate kinase mutant. *Applied and Environmental Microbiology*, *66*(9), 4045–9. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=92257&tool=pmcentrez&rendertype=abstract

Geng, H., Zhu, Y., Mullen, K., Zuber, C. S., & Nakano, M. M. (2007). Characterization of ResDE-dependent fnr transcription in Bacillus subtilis. *Journal of Bacteriology*, *189*(5), 1745–55. doi:10.1128/JB.01502-06

Geng, H., Zuber, P., & Nakano, M. M. (2007). Regulation of respiratory genes by ResD-ResE signal transduction system in Bacillus subtilis. *Methods in Enzymology*, *422*, 448–64. doi:10.1016/S0076-6879(06)22023-8

Gilman, J., & Love, J. (2016). Synthetic promoter design for new microbial chassis. *Biochemical Society Transactions*, *44*(3), 731–7. doi:10.1042/BST20160042

Goosens, V. J., Monteferrante, C. G., & van Dijl, J. M. (2014). The Tat system of Gram-positive bacteria. *Biochimica et Biophysica Acta*, *1843*(8), 1698–706. doi:10.1016/j.bbamcr.2013.10.008

Gougoulias, C., Clark, J. M., & Shaw, L. J. (2014). The role of soil microbes in the global carbon cycle: tracking the below-ground microbial processing of plant-derived carbon for manipulating carbon dynamics in agricultural systems. *Journal of the Science of Food and Agriculture*, *94*(12), 2362–71. doi:10.1002/jsfa.6577

Green, J., & Paget, M. S. (2004). Bacterial redox sensors. *Nature Reviews. Microbiology*, *2*(12), 954–66. doi:10.1038/nrmicro1022

Gregory, M. A., Till, R., & Smith, M. C. M. (2003). Integration Site for Streptomyces Phage BT1 and Development of Site-Specific Integrating Vectors, *185*(17), 5320–5323. doi:10.1128/JB.185.17.5320

Gust, B., Rourke, S. O., Bird, N., Kieser, T., & Chater, K. (2004). Recombineering in Streptomyces coelicolor, 1–22.

Gyan, S., Shiohira, Y., Sato, I., Takeuchi, M., & Sato, T. (2006). Regulatory loop between redox sensing of the NADH/NAD(+) ratio by Rex (YdiH) and oxidation of NADH by NADH dehydrogenase Ndh in Bacillus subtilis. *Journal of Bacteriology*, *188*(20), 7062–71. doi:10.1128/JB.00601-06

Haas, L. O., Cregg, J. M., & Gleeson, M. A. (1990). Development of an integrative DNA transformation system for the yeast Candida tropicalis. *J. Bacteriol.*, *172*(8), 4571–4577. Retrieved from http://jb.asm.org/content/172/8/4571.abstract?ijkey=cf9a47802b16cefc092f61c39b76b60f481c47eb&keytype2=tf_ipsecsha

Hamilton, T. L., Bryant, D. A., & Macalady, J. L. (2016). The role of biology in planetary evolution: cyanobacterial primary production in low-oxygen Proterozoic oceans. *Environmental Microbiology*, *18*(2), 325–40. doi:10.1111/1462-2920.13118

Härtig, E., & Jahn, D. (2012). Regulation of the Anaerobic Metabolism in Bacillus subtilis, *61*, 195–216. doi:10.1016/B978-0-12-394423-8.00005-6

Hill, J., Nelson, E., Tilman, D., Polasky, S., & Tiffany, D. (2006). Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels.

Proceedings of the National Academy of Sciences of the United States of America, *103*(30), 11206–10. doi:10.1073/pnas.0604600103

IEA. (2015). Energy and Climate Change. *World Energy Outlook Special Report*, 1–200. doi:10.1038/479267b

Ismaiel, A. A., Zhu, C. X., Colby, G. D., & Chen, J. S. (1993). Purification and characterization of a primary-secondary alcohol dehydrogenase from two strains of Clostridium beijerinckii. *J. Bacteriol.*, *175*(16), 5097–5105. Retrieved from http://jb.asm.org/content/175/16/5097.long

Jasaitis, A., Borisov, V. B., Belevich, N. P., Morgan, J. E., Konstantinov, A. A., & Verkhovsky, M. I. (2000). Electrogenic Reactions of Cytochrome bd †. *Biochemistry*, *39*(45), 13800–13809. doi:10.1021/bi001165n

John, R. P., Nampoothiri, K. M., & Pandey, A. (2007). Fermentative production of lactic acid from biomass: an overview on process developments and future perspectives. *Applied Microbiology and Biotechnology*, *74*(3), 524–34. doi:10.1007/s00253-006-0779-6

Juo, A. S. R. K. F. (2003). *Tropical Soils : Properties and Management for Sustainable Agriculture: Properties and Management for Sustainable Agriculture*. Oxford University Press, USA. Retrieved from https://books.google.com/books?id=dVGgQcuoibAC&pgis=1

Kananaviciute, R., & Citavicius, D. (2015). Genetic engineering of Geobacillus spp. *Journal of Microbiological Methods*, *111*, 31–9. doi:10.1016/j.mimet.2015.02.002

Kavscek, M., Strazar, M., Curk, T., Natter, K., & Petrovc, U. (2015). Yeast as a cell factory: current state and perspectives. *Microbial Cell Factories*, *14*(1), 94. doi:10.1186/s12934-015-0281-x

Korkhin, Y., Kalb(Gilboa), A. J., Peretz, M., Bogin, O., Burstein, Y., & Frolow, F. (1998). NADP-dependent bacterial alcohol dehydrogenases: crystal structure, cofactor-binding and cofactor specificity of the ADHs of Clostridium beijerinckii and Thermoanaerobacter brockii. *Journal of Molecular Biology*, *278*(5), 967–981. doi:10.1006/jmbi.1998.1750

Kreuzer, P., Gärtner, D., Allmansberger, R., & Hillen, W. (1989). Identification and sequence analysis of the Bacillus subtilis W23 xylR gene and xyl operator. *Journal of Bacteriology*, *171*(7), 3840–5. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=210133&tool=pmcentrez&rendertype=abstract

Kumar, P., Barrett, D. M., Delwiche, M. J., & Stroeve, P. (2009). Methods for Pretreatment of Lignocellulosic Biomass for Efficient Hydrolysis and Biofuel

Production. *Industrial & Engineering Chemistry Research*, *48*(8), 3713–3729. doi:10.1021/ie801542g

Lazazzera, B. a, Beinert, H., Khoroshilova, N., Kennedy, M. C., & Kiley, P. J. (1996). DNA binding and dimerization of the Fe-S-containing FNR protein from Escherichia coli are regulated by oxygen. *The Journal of Biological Chemistry*, *271*(5), 2762–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8576252

Lee, C. A., & Grossman, A. D. (2007). Identification of the origin of transfer (oriT) and DNA relaxase required for conjugation of the integrative and conjugative element ICEBs1 of Bacillus subtilis. *Journal of Bacteriology*, *189*(20), 7254–7261. doi:10.1128/JB.00932-07

Liao, H. H., & Kanikula, A. M. (1990). Increased efficiency of transformation ofBacillus stearothermophilus by a plasmid carrying a thermostable kanamycin resistance marker. *Current Microbiology*, *21*(5), 301–306. doi:10.1007/BF02092095

Liao, H., McKenzie, T., & Hageman, R. (1986). Isolation of a thermostable enzyme variant by cloning and selection in a thermophile. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(3), 576–80. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=322906&tool=pmcentrez&rendertype=abstract

Liu, B., Kearns, D. B., & Bechhofer, D. H. (2016). Expression of multiple Bacillus subtilis genes is controlled by decay of slrA mRNA from Rho-dependent 3??? ends. *Nucleic Acids Research*, *44*(7), 3364–3372. doi:10.1093/nar/gkw069

Liu, X., & De Wulf, P. (2004). Probing the ArcA-P modulon of Escherichia coli by whole genome transcriptional analysis and sequence recognition profiling. *The Journal of Biological Chemistry*, *279*(13), 12588–97. doi:10.1074/jbc.M313454200

Lopes, M. S. G. (2015). Engineering biological systems toward a sustainable bioeconomy. *Journal of Industrial Microbiology & Biotechnology*, *42*(6), 813–38. doi:10.1007/s10295-015-1606-9

Loui, C., Chang, A. C., & Lu, S. (2009). Role of the ArcAB two-component system in the resistance of Escherichia coli to reactive oxygen stress. *BMC Microbiology*, *9*, 183. doi:10.1186/1471-2180-9-183

Ludwig, H., Homuth, G., Schmalisch, M., Dyka, F. M., Hecker, M., & Stülke, J. (2001). Transcription of glycolytic genes and operons in Bacillus subtilis: evidence for the presence of multiple levels of control of the gapA operon. *Molecular Microbiology*, *41*(2), 409–422. doi:10.1046/j.1365-2958.2001.02523.x

Luscombe, N. M., Laskowski, R. a, & Thornton, J. M. (2001). Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Research*, *29*(13), 2860–2874. doi:10.1093/nar/29.13.2860

Lynd, L. R., van Zyl, W. H., McBride, J. E., & Laser, M. (2005). Consolidated bioprocessing of cellulosic biomass: an update. *Current Opinion in Biotechnology*, *16*(5), 577–83. doi:10.1016/j.copbio.2005.08.009

Malpica, R., Franco, B., Rodriguez, C., Kwon, O., & Georgellis, D. (2004). Identification of a quinone-sensitive redox switch in the ArcB sensor kinase. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(36), 13318–23. doi:10.1073/pnas.0403064101

Maranhao, A., & Ellington, A. D. (2016). Evolving orthogonal suppressor tRNAs to incorporate modified amino acids. *ACS Synthetic Biology*. doi:10.1021/acssynbio.6b00145

Marino, M., Hoffmann, T., Schmid, R., Möbitz, H., & Jahn, D. (2000). Changes in protein synthesis during the adaptation of Bacillus subtilis to anaerobic growth conditions. *Microbiology (Reading, England)*, *146 (Pt 1*(1), 97–105. doi:10.1099/00221287-146-1-97

Marreiros, B. C., Calisto, F., Castro, P. J., Duarte, A. M., Sena, F. V, Silva, A. F., … Pereira, M. M. (2016). Exploring membrane respiratory chains. *Biochimica et Biophysica Acta*, *1857*(8), 1039–67. doi:10.1016/j.bbabio.2016.03.028

Mazoch, J., & Kucera, I. (2002). Control of gene expression by FNR-like proteins in facultatively anaerobic bacteria. *Folia Microbiologica*, *47*(2), 95–103. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12058404

Mclaughlin, K. J., Strain-damerell, C. M., Xie, K., Brekasis, D., Soares, A. S., Paget, M. S. B., & Kielkopf, C. L. (2010). Article Structural Basis for NADH / NAD + Redox Sensing by a Rex Family Repressor. *MOLCEL*, *38*(4), 563–575. doi:10.1016/j.molcel.2010.05.006

McLaughlin, K. J., Strain-Damerell, C. M., Xie, K., Brekasis, D., Soares, A. S., Paget, M. S. B., & Kielkopf, C. L. (2010). Structural Basis for NADH/NAD+ Redox Sensing by a Rex Family Repressor. *Molecular Cell*, *38*(4), 563–575. doi:10.1016/j.molcel.2010.05.006

Mettert, E. L., & Kiley, P. J. (2005). ClpXP-dependent proteolysis of FNR upon loss of its O2-sensing [4Fe-4S] cluster. *Journal of Molecular Biology*, *354*(2), 220–32. doi:10.1016/j.jmb.2005.09.066

Miethke, M., Schmidt, S., & Marahiel, M. A. (2008). The major facilitator superfamily-type transporter YmfE and the multidrug-efflux activator Mta mediate bacillibactin secretion in Bacillus subtilis. *Journal of Bacteriology*, *190*(15), 5143–52. doi:10.1128/JB.00464-08

Mikulskis, A., Aristarkhov, A., & Lin, E. C. (1997). Regulation of expression of the ethanol dehydrogenase gene (adhE) in Escherichia coli by catabolite repressor activator protein Cra. *Journal of Bacteriology*, *179*(22), 7129–34. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=179656&tool=pmcentrez&rendertype=abstract

Miller, R. G., & Sorrell, S. R. (2014). The future of oil supply. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *372*(2006), 20130179. doi:10.1098/rsta.2013.0179

Morris, R. L., & Schmidt, T. M. (2013). Shallow breathing: bacterial life at low O(2). *Nature Reviews. Microbiology*, *11*(3), 205–12. doi:10.1038/nrmicro2970

Myers, K. S., Yan, H., Ong, I. M., Chung, D., Liang, K., Tran, F., … Kiley, P. J. (2013). Genome-scale analysis of escherichia coli FNR reveals complex features of transcription factor binding. *PLoS Genetics*, *9*(6), e1003565. doi:10.1371/journal.pgen.1003565

Nah, H.-J., Woo, M.-W., Choi, S.-S., & Kim, E.-S. (2015). Precise cloning and tandem integration of large polyketide biosynthetic gene cluster using Streptomyces artificial chromosome system. *Microbial Cell Factories*, *14*(1), 140. doi:10.1186/s12934-015-0325-2

Nakamura, A., Sosa, A., Komori, H., Kita, A., & Miki, K. (2007). Crystal structure of TTHA1657 (AT-rich DNA-binding protein; p25) from Thermus thermophilus HB8 at 2.16 A resolution. *Proteins*, *66*(3), 755–9. doi:10.1002/prot.21222

Nakano, M. M., Dailly, Y. P., & Zuber, P. (1997). Characterization of anaerobic fermentative growth of Bacillus subtilis : identification of fermentation end products and genes required for growth . Characterization of Anaerobic Fermentative Growth of Bacillus subtilis : Identification of Fermentation En.

Nieves, L. M., Panyon, L. A., & Wang, X. (2015). Engineering Sugar Utilization and Microbial Tolerance toward Lignocellulose Conversion. *Frontiers in Bioengineering and Biotechnology*, *3*, 17. doi:10.3389/fbioe.2015.00017

Pagels, M., Fuchs, S., Pané-Farré, J., Kohler, C., Menschner, L., Hecker, M., … Engelmann, S. (2010). Redox sensing by a Rex-family repressor is involved in

the regulation of anaerobic gene expression in Staphylococcus aureus. *Molecular Microbiology*, *76*(5), 1142–61. doi:10.1111/j.1365-2958.2010.07105.x

Poggi, D., Oliveira de Giuseppe, P., & Picardeau, M. (2010). Antibiotic resistance markers for genetic manipulations of Leptospira spp. *Applied and Environmental Microbiology*, *76*(14), 4882–5. doi:10.1128/AEM.00775-10

Pompeo, F., Foulquier, E., & Galinier, A. (2016). Impact of Serine/Threonine Protein Kinases on the Regulation of Sporulation in Bacillus subtilis. *Frontiers in Microbiology*, *7*, 568. doi:10.3389/fmicb.2016.00568

Ramírez, S., Moreno, R., Zafra, O., Castán, P., Vallés, C., & Berenguer, J. (2000). Two nitrate/nitrite transporters are encoded within the mobilizable plasmid for nitrate respiration of Thermus thermophilus HB8. *Journal of Bacteriology*, *182*(8), 2179–83. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=111266&tool=pmcentrez&rendertype=abstract

Ravcheev, D. a, Li, X., Latif, H., Zengler, K., Leyn, S. a, Korostelev, Y. D., … Rodionov, D. a. (2012). Transcriptional regulation of central carbon and energy metabolism in bacteria by redox-responsive repressor Rex. *Journal of Bacteriology*, *194*(5), 1145–57. doi:10.1128/JB.06412-11

Reents, H., Münch, R., Dammeyer, T., Härtig, E., Mu, R., Jahn, D., & Ha, E. (2006). The Fnr Regulon of Bacillus subtilis The Fnr Regulon of Bacillus subtilis †, *188*(3). doi:10.1128/JB.188.3.1103

Reeve, B., Martinez-Klimova, E., de Jonghe, J., Leak, D. J., & Ellis, T. (2016). The Geobacillus Plasmid Set: A Modular Toolkit for Thermophile Engineering. *ACS Synthetic Biology*. doi:10.1021/acssynbio.5b00298

Richardson, D. J. (2000). Bacterial respiration: a flexible process for a changing environment. *Microbiology (Reading, England)*, *146 (Pt 3*(3), 551–71. doi:10.1099/00221287-146-3-551

Rodionov, D. A., Dubchak, I. L., Arkin, A. P., Alm, E. J., & Gelfand, M. S. (2005). Dissimilatory metabolism of nitrogen oxides in bacteria: comparative reconstruction of transcriptional networks. *PLoS Computational Biology*, *1*(5), e55. doi:10.1371/journal.pcbi.0010055

Roggiani, M., & Goulian, M. (2015). Oxygen-Dependent Cell-to-Cell Variability in the Output of the Escherichia coli Tor Phosphorelay. *Journal of Bacteriology*, *197*(12), 1976–1987. doi:10.1128/JB.00074-15

Romero, S., Merino, E., Bolívar, F., Gosset, G., & Martinez, A. (2007). Metabolic engineering of Bacillus subtilis for ethanol production: lactate dehydrogenase plays a key role in fermentative metabolism. *Applied and Environmental Microbiology*, *73*(16), 5190–8. doi:10.1128/AEM.00625-07

Rosenfeld, N., Elowitz, M. B., & Alon, U. (2002). Negative Autoregulation Speeds the Response Times of Transcription Networks. *Journal of Molecular Biology*, *323*(5), 785–793. doi:10.1016/S0022-2836(02)00994-4

Saha, B. C. (2003). Hemicellulose bioconversion. *Journal of Industrial Microbiology & Biotechnology*, *30*(5), 279–91. doi:10.1007/s10295-003-0049-x

Shaw, a J., Podkaminer, K. K., Desai, S. G., Bardsley, J. S., Rogers, S. R., Thorne, P. G.Lynd, L. R. (2008). Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(37), 13769–74. doi:10.1073/pnas.0801266105

Shimada, T., Yamamoto, K., & Ishihama, A. (2011). Novel members of the Cra regulon involved in carbon metabolism in Escherichia coli. *Journal of Bacteriology*, *193*(3), 649–59. doi:10.1128/JB.01214-10

Sickmier, E. A., Brekasis, D., Paranawithana, S., Bonanno, J. B., Paget, M. S. B., Burley, S. K., & Kielkopf, C. L. (2005). X-Ray Structure of a Rex-Family Repressor / NADH Complex Insights into the Mechanism of Redox Sensing, *13*, 43–54. doi:10.1016/j.str.2004.10.012

Silva-Rocha, R., Martínez-García, E., Calles, B., Chavarría, M., Arce-Rodríguez, A., de Las Heras, A.de Lorenzo, V. (2013). The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Research*, *41*(Database issue), D666–75. doi:10.1093/nar/gks1119

Simmons, B. A., Loque, D., & Blanch, H. W. (2008). Next-generation biomass feedstocks for biofuel production. *Genome Biology*, *9*(12), 242. doi:10.1186/gb-2008-9-12-242

Simon, R., Priefer, U., & Pühler, A. (1983). A BROAD HOST RANGE MOBILIZATION SYSTEM FOR INVIVO GENETIC-ENGINEERING - TRANSPOSON MUTAGENESIS IN GRAM-NEGATIVE BACTERIA. *BIO-TECHNOLOGY*, *1*(9). Retrieved from https://pub.uni-bielefeld.de/publication/1659456

Sims, R., Taylor, M., Jack, S., & Mabee, W. (2008). From 1st to 2nd Generation Bio Fuel Technologies: An overview of current industry and RD&D activities. *IEA Bioenergy*, (November), 1–124.

Sørensen, H. P., & Mortensen, K. K. (2005). Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microbial Cell Factories*, *4*(1), 1. doi:10.1186/1475-2859-4-1

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

Stanton, B. C., Nielsen, A. A. K., Tamsir, A., Clancy, K., Peterson, T., & Voigt, C. A. (2014). Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nature Chemical Biology*, *10*(2), 99–105. doi:10.1038/nchembio.1411

Strain-damerell, C. M. (n.d.). Functional analysis of Rex , a sensor of the NADH / NAD + redox poise in Streptomyces coelicolor.

Suzuki, H., Murakami, A., & Yoshida, K. (2012). Counterselection system for Geobacillus kaustophilus HTA426 through disruption of pyrF and pyrR. *Applied and Environmental Microbiology*, *78*(20), 7376–83. doi:10.1128/AEM.01669-12

Suzuki, H., & Yoshida, K. (2012). Genetic transformation of Geobacillus kaustophilus HTA426 by conjugative transfer of host-mimicking plasmids. *Journal of Microbiology and Biotechnology*, *22*(9), 1279–87. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22814504

Tang, Y. J., Sapra, R., Joyner, D., Hazen, T. C., Myers, S., Reichmuth, D., … Keasling, J. D. (2009). Analysis of metabolic pathways and fluxes in a newly

discovered thermophilic and ethanol-tolerant Geobacillus strain. *Biotechnology and Bioengineering*, *102*(5), 1377–86. doi:10.1002/bit.22181

Taylor, M. P., Eley, K. L., Martin, S., Tuffin, M. I., Burton, S. G., & Cowan, D. A. (2009). Thermophilic ethanologenesis: future prospects for second-generation bioethanol production. *Trends in Biotechnology*, *27*(7), 398–405. doi:10.1016/j.tibtech.2009.03.006

Taylor, M. P., Esteban, C. D., & Leak, D. J. (2008). Development of a versatile shuttle vector for gene expression in Geobacillus spp. *Plasmid*, *60*(1), 45–52. doi:10.1016/j.plasmid.2008.04.001

Tobisch, S., Zühlke, D., Bernhardt, J., Stülke, J., & Hecker, M. (1999). Role of CcpA in regulation of the central pathways of carbon catabolism in Bacillus subtilis. *Journal of Bacteriology*, *181*(22), 6996–7004. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=94174&tool=pmcentrez&rendertype=abstract

Tolla, D. A., & Savageau, M. A. (2010). Regulation of aerobic-to-anaerobic transitions by the FNR cycle in Escherichia coli. *Journal of Molecular Biology*, *397*(4), 893–905. doi:10.1016/j.jmb.2010.02.015

Tominaga, Y., Ohshiro, T., & Suzuki, H. (2016). Conjugative plasmid transfer from Escherichia coli is a versatile approach for genetic transformation of thermophilic Bacillus and Geobacillus species. *Extremophiles : Life under Extreme Conditions*, *20*(3), 375–81. doi:10.1007/s00792-016-0819-9

Trchounian, A. (2004). Escherichia coli proton-translocating F0F1-ATP synthase and its association with solute secondary transporters and/or enzymes of anaerobic oxidation-reduction under fermentation. *Biochemical and Biophysical Research Communications*, *315*(4), 1051–7. doi:10.1016/j.bbrc.2004.02.005

Tripathi, S. A., Olson, D. G., Argyros, D. A., Miller, B. B., Barrett, T. F., Murphy, D. M.,Caiazza, N. C. (2010). Development of pyrF-based genetic system for targeted gene deletion in Clostridium thermocellum and creation of a pta mutant. *Applied and Environmental Microbiology*, *76*(19), 6591–9. doi:10.1128/AEM.01484-10

Turner, R. J., Bonner, E. R., Grabner, G. K., & Switzer, R. L. (1998). Purification and Characterization of Bacillus subtilis PyrR, a Bifunctional pyr mRNA-binding Attenuation Protein/Uracil Phosphoribosyltransferase. *Journal of Biological Chemistry*, *273*(10), 5932–5938. doi:10.1074/jbc.273.10.5932

Tuteja, R. (2005). Type I signal peptidase: an overview. *Archives of Biochemistry and Biophysics*, *441*(2), 107–11. doi:10.1016/j.abb.2005.07.013

Unden, G., & Bongaerts, J. (1997). Alternative respiratory pathways of Escherichia coli: energetics and transcriptional regulation in response to electron acceptors. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1320*(3), 217–234. doi:10.1016/S0005-2728(97)00034-0

Unden, G., Steinmetz, P. A., & Degreif-Dünnwald, P. (2014). The Aerobic and Anaerobic Respiratory Chain of Escherichia coli and Salmonella enterica: Enzymes and Energetics. *EcoSal Plus*, *6*(1). doi:10.1128/ecosalplus.ESP-0005-2013

Volbeda, A., Darnault, C., Renoux, O., Nicolet, Y., & Fontecilla-Camps, J. C. (2015). The crystal structure of the global anaerobic transcriptional regulator FNR explains its extremely fine-tuned monomer-dimer equilibrium. *Science Advances*, *1*(11), e1501086. doi:10.1126/sciadv.1501086

Voskuil, M. I., Voepel, K., & Chambliss, G. H. (1995). The ? 16 region, a vital sequence for the utilization of a promoter in Bacillus subtilis and Escherichia coli. *Molecular Microbiology*, *17*(2), 271–279. doi:10.1111/j.1365-2958.1995.mmi_17020271.x

Wang, E., Bauer, M. C., Rogstam, A., Linse, S., Logan, D. T., & von Wachenfeldt, C. (2008). Structure and functional properties of the Bacillus subtilis transcriptional repressor Rex. *Molecular Microbiology*, *69*(2), 466–78. doi:10.1111/j.1365-2958.2008.06295.x

Wang, E., Ikonen, T. P., Knaapila, M., Svergun, D., Logan, D. T., & Wachenfeldt, C. Von. (2011). Small-angle X-ray Scattering Study of a Rex Family Repressor : Conformational Response to NADH and NAD + Binding in Solution. *Journal of Molecular Biology*, *408*(4), 670–683. doi:10.1016/j.jmb.2011.02.050

Weber, J., & Senior, A. E. (2003). ATP synthesis driven by proton transport in F 1 F 0 -ATP synthase. *FEBS Letters*, *545*(1), 61–70. doi:10.1016/S0014-5793(03)00394-6

WEF. (2016). The global risks report 2016, 11th edition, 103. Retrieved from https://www.weforum.org/reports/the-global-risks-report-2016/

Weiner, J. H., Bilous, P. T., Shaw, G. M., Lubitz, S. P., Frost, L., Thomas, G. H., … Turner, R. J. (1998). A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins. *Cell*, *93*(1), 93–101. doi:10.1016/S0092-8674(00)81149-6

Wietzke, M., & Bahl, H. (2012). The redox-sensing protein Rex, a transcriptional regulator of solventogenesis in Clostridium acetobutylicum. *Applied Microbiology and Biotechnology*, *96*(3), 749–61. doi:10.1007/s00253-012-4112-2

Williams, P. R. D., Inman, D., Aden, A., & Heath, G. A. (2009). Environmental and sustainability factors associated with next-generation biofuels in the U.S.: what do we really know? *Environmental Science & Technology*, *43*(13), 4763–75. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19673263

Xu, P., Li, L., Zhang, F., Stephanopoulos, G., & Koffas, M. (2014). Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(31), 11299–304. doi:10.1073/pnas.1406401111

Yagi, T. (1993). The bacterial energy-transducing NADH-quinone oxidoreductases. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1141*(1), 1–17. doi:10.1016/0005-2728(93)90182-F

Yamamoto, S., Gunji, W., Suzuki, H., Toda, H., Suda, M., Jojima, T., … Yukawa, H. (2012). Overexpression of genes encoding glycolytic enzymes in Corynebacterium glutamicum enhances glucose metabolism and alanine production under oxygen deprivation conditions. *Applied and Environmental Microbiology*, *78*(12), 4447–57. doi:10.1128/AEM.07998-11

Yasueda, H., Kawahara, Y., & Sugimoto, S. (1999). Bacillus subtilis yckG and yckF Encode Two Key Enzymes of the Ribulose Monophosphate Pathway Used by Methylotrophs, and yckH Is Required for Their Expression. *J. Bacteriol.*, *181*(23), 7154–7160. Retrieved from http://jb.asm.org/content/181/23/7154?ijkey=ea11518321460e880a5d6a0e845157ef919f044f&keytype2=tf_ipsecsha

Yu, J., & Tao, M. (2010). [Effect of inorganic salts on the conjugation and heterologous expression of actinorhodin in Streptomyces avermitilis]. *Wei Sheng Wu Xue Bao = Acta Microbiologica Sinica*, *50*(11), 1556–61. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21268904

Yurimoto, H., Hirai, R., Matsuno, N., Yasueda, H., Kato, N., & Sakai, Y. (2005). HxlR, a member of the DUF24 protein family, is a DNA-binding protein that acts as a positive regulator of the formaldehyde-inducible hxlAB operon in Bacillus subtilis. *Molecular Microbiology*, *57*(2), 511–9. doi:10.1111/j.1365-2958.2005.04702.x

Zeigler, D. R. (2014). The Geobacillus paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet? *Microbiology (Reading, England)*, *160*(Pt 1), 1–11. doi:10.1099/mic.0.071696-0

Zeldes, B. M., Keller, M. W., Loder, A. J., Straub, C. T., Adams, M. W. W., & Kelly, R. M. (2015). Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. *Frontiers in Microbiology*, *6*, 1209. doi:10.3389/fmicb.2015.01209

Zhang, L., Nie, X., Ravcheev, D. a, Rodionov, D. a, Sheng, J., Gu, Y., … Yang, C. (2014). Redox-responsive repressor Rex modulates alcohol production and oxidative stress tolerance in Clostridium acetobutylicum. *Journal of Bacteriology*, *196*(22), 3949–63. doi:10.1128/JB.02037-14

Zhang, X., Shanmugam, K. T., & Ingram, L. O. (2010). Fermentation of glycerol to succinate by metabolically engineered strains of Escherichia coli. *Applied and Environmental Microbiology*, *76*(8), 2397–401. doi:10.1128/AEM.02902-09

Zheng, Y., Ko, T. P., Sun, H., Huang, C. H., Pei, J., Qiu, R.,Guo, R. T. (2014). Distinct structural features of Rex-family repressors to sense redox levels in anaerobes and aerobes. *Journal of Structural Biology*, *188*(3), 195–204. doi:10.1016/j.jsb.2014.11.001

8. Appendix

Abbreviations

AprR	Apramycin resistant
ATP	Adenosine 5'-triphosphate
BLAST	Basic local alignment tool
Bp	Base pair
B-Rex	*Bacillus subtilis* Rex
ChIP	Chromatin immunoprecipitation
C-Rex	*Clostridium thermocellum* Rex
Da	Daltons
dH2O	distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
EMSA	Electromobility shift assay
G-Rex	*Geobacillus thermoglucosidans* Rex
g	Grams
h	Hour(s)
Kan^R	Kanamycin resistant
Kan^S	Kanamycin sensitive
L	Litre
min	Minute(s)
NAD^+	Nicotinamide adenine dinucleotide, oxidised form
NADH	Nicotinamide adenine dinucleotide, reduced form
OD	Optical density (wavelength indicated in subscript)

PCR	Polymerase chain reaction
PDB	Protein data bank
PMF	Proton motive force
qPCR	Quantitative PCR
ROP	Rex operator
RT-qPCR	Reverse transcriptase qPCR
sec	Second(s)
T-Rex	Thermus aquaticus Rex
U	Unit(s)

DNA sequences

C-Rex

```
   1 ATGAACCTGG ACAAAAAAAT CAGCATGGCA GTTATTCGTC GTCTGCCTCG TTATTATCGT
  61 TATCTGAGCG ATCTGCTGAA ACTGGGTATT ACCCGTATTA GCAGCAAAGA ACTGAGCAGC
 121 CGTATGGGCA TTACCGCAAG CCAGATTCGT CAGGATCTGA ATTGTTTTGG TGGTTTTGGT
 181 CAGCAGGGTT ATGGTTATAA TGTGGAATAT CTGTACAAAG AGATCGGCAA TATTCTGGGT
 241 GTTAATGAGG CCTTCAAAAT CATTATTATC GGTGCCGGTA ATATGGGTCA GGCACTGGCA
 301 AATTATACCA ACTTTGAAAA ACGCGGTTTC AAACTGATCG GCATCTTTGA CATTAATCCG
 361 AACCTGATCG GGAAAAAAAT CCGTGATGTT GAAATCATGC ATCTGGATAG CCTGGATCGT
 421 TTTGTTGCAG AAAATCAGGT TGATATTGCC ATTCTGTGTG TGCCGTATGA AAATACACCG
 481 GCAGTTGCAG ATAAAGTTGC ACGTCTGGGT GTGAAAGGTC TGTGGAATTT TAGCCCGATG
 541 GATCTGAAAC TGCCTTATGA TGTGATTATC GAAAATGTGC ATCTGTCCGA TAGCCTGATG
 601 GTTCTGGGTT ATCGTCTGAA TGAAATGCGT AAAAGCCAGC GCAATAAAAG CGACTACAAA
 661 GATCATGACG GCGATTATAA GGACCACGAT ATCGACTACA AAGACGATGA CGATAAGTAA
 721 c
```

G-Rex

```
   1 ATGAACAATG AGCAGCCAAA GATCCCACAG GCAACTGCTA AACGTCTGCC TCTGTATTAT
  61 CGTTTTCTGA AGAATTTACA TGCTTCAGGA AAACAGCGTG TTAGCAGCGC GGAATTATCT
 121 GAAGCCGTCA AAGTAGATAG CGCAACAATT CGCCGCGACT TCTCATATTT TGGCGCTCTG
 181 GGTAAAAAGG GCTATGGCTA CAACGTTAAC TATCTGCTGT CTTTTTTCCG GAAAACACTG
 241 GATCAGGACG AAATCACGGA AGTTGCTCTT TTCGGCGTTG GCAACCTTGG TACGGCATTC
 301 TTAAATTACA ATTTCAGCAA AAATAACAAT ACAAAAATCG TAATTGCGTT TGACGTGGAT
 361 GAGGAGAAAA TTGGTAAAGA GGTTGGTGGG GTGCCTGTAT ATCACCTTGA CGAAATGGAA
 421 ACAAGACTCC ACGAAGGCAT TCCAGTTGCC ATTTTAACAG TTCCGGCACA TGTCGCGCAA
 481 TGGATTACTG ACCGACTTGT TCAAAAAGGC ATCAAAGGCA TTTTGAACTT CACACCGGCA
 541 AGACTTAATG TCCCGAAACA TATTAGAGTT CATCATATCG ACCTGGCAGT CGAATTGCAG
 601 TCGTTAGTTT ATTTTTTGAA AAACTATCCA GCGGAGGACT ACAAAGATCA TGACGGCGAT
 661 TATAAGGACC ACGATATCGA CTACAAAGAC GATGACGATA AGTAAc
```