Characterisation of carbohydrate transport in the solventogenic organism *Clostridium acetobutylicum* ATCC 824.

Eve Joss Bird

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

Edinburgh Napier University

September, 2011

Abstract

Fossil fuels are a finite resource, which are subject to erratic price fluctuations and are contributing to disruption of the global carbon cycle. There is a requirement to establish a renewable, sustainable and secure fuel source that can reduce worldwide dependence on oil. Industry and agriculture contribute to the vast quantity of carbohydrate-rich waste biomass continuously generated. Certain *Clostridium spp.* can convert the sugars found in biomass into solvents which can be used as biofuel, but there is a need to learn more about their metabolic capabilities in order to introduce them efficiently on an industrial scale.

The phosphotransferase system (PTS) is a predominant method of carbohydrate transport in Firmicutes. An *in silico* analysis of the complete phosphotransferase complement of *C. acetobutylicum* ATCC 824 revealed that the genome encodes thirteen permease proteins and a lone EIIA domain, the properties of which are discussed. The genes encoding the general PTS energy coupling proteins have an unusual genome context. In the majority of Gram positive bacteria the two genes are in an operon; in *C. acetobutylicum* the two genes appear to be monocistronic. Interestingly in *C. acetobutylicum* the gene *hprK,* encoding the fbp-stimulated HPr regulator HPr kinase/phosphorylase (HprK/P), is found directly upstream from a gene *glpX*, which appears to encode a fructose-1,6-bisphosphatase (FBPase). Here, two genes were identified encoding FBPases in *C. acetobutylicum* - a class II, GlpX-like, FBPase and a class III FBPase.

In silico analysis revealed two systems involved with β-glucoside utilisation. System I is comprised of three open reading frames (ORFs), which encode all functions necessary for the regulated uptake and utilisation of β-glucosides; a transcriptional antiterminator (*bglG*), an EII (*bglA*) and a phospho-β-glucosidase (*bglB*). Gene spacing, regulatory elements and similarity with other PTS's are consistent with the expression of *bglGAB* as an operon. Molecular cloning of the permease revealed it was specific for arbutin transport. The domains of the permease from System II were encoded on three separate ORFs, interjected by a phospho-β-glucosidase. A divergently expressed transcriptional activator was identified upstream from the EII. Heterologous expression revealed System II was specific for salicin. Preliminary investigation of the transport and utilisation mechanisms of the monomers present in hemicellulose revealed two putative pentose isomerases, as well as two mannose phosphotransferases. Characterisation of carbohydrate transport, metabolism and regulation could provide information to optimise future biosolvent production facilities.

Acknowledgements

I would like to express special thanks to my director of studies Professor Martin Tangney, who provided me with the opportunity to undertake this research project and supported me throughout this adventure.

I would also like to convey my gratitude to my second supervisor Dr Fiona Stainsby and Dr Wilfrid J. Mitchell, Dr Douglas Fraser-Pitt and Dr Jane White for their helpful advice and discussions throughout this project. Thanks to Dr Renata Riha at the Department of Sleep Medicine for valuable advice on managing study and narcolepsy.

I am extremely grateful to my friends and family, especially my household (Tom, Dickie and Buzzy) for their support. Thanks also to: Ron Bird, Heather Oliphant, Joe Milton, Katja Sievers, Adam Crawford, Nicky Hamilton-Smith, Thomas Harrison, Richard Clark, Stuart Galloway, Jill Martin, Mary-Jane Anderson, Will Power, Sarah Dalrymple, Jennifer Bell, Ron O. Badger, Jen Moss Urquart, Sarah and Paddy Pink, Gavin Smith, Alex Harrison, Nils Betzler, Magdalena Rosinska, Magdalena Klyszewska, Sandra Sharples, Kenneth Leiper, Ali Al-Mohaish, Julie Watson, Gillian Kane and Anna Tangney.

A massive thanks to the Biofuel Research Centre and my sponsors Green Biologics Ltd and Edinburgh Napier University, who made this all possible.

Publications in preparation arising from this research

Bird, E., W. J. Mitchell, & M. Tangney. The PTS complement of *Clostridium acetobutylicum* ATCC 824: A review.
Manuscript in preparation.
Chapter 3

Bird, E., W. J. Mitchell, & M. Tangney. β-glucoside transport in *Clostridium acetobutylicum* ATCC 824.
Manuscript in preparation.
Chapter 4

Al-Mohaish, A., **E. Bird,** W. J. Mitchell, & M. Tangney. Cloning and analysis of the *glpx* like class II fructose 1,6-biphosphatase in *Clostridium acetobutylicum* ATCC 824.
Manuscript in preparation.
Chapter 3

Watson, J., **E. Bird,** W. J. Mitchell, & M. Tangney. Xylose utilisation in *Clostridium acetobutylicum* ATCC 824.
Manuscript in preparation.
Chapter 5

Abbreviations

A	Adenine
ABE	Acetone butanol ethanol
ADP	Adenosine diphosphate
Amp	Ampicillin
ATP	Adenosine 5'-triphosphate
BfRC	Biofuel Research Centre
Bgl / *bgl*	(referring to protein / operon or gene) β-glucoside associated
BLAST	Basic local alignment search tool
BOD	Biological oxygen demand
bp	base pairs
BuOH	Butanol
C	Cytosine
cac	(referring to gene on chromosome) in *C. acetobutylicum*
cap	(referring to gene on plasmid) in *C. acetobutylicum*
CAT	Co-antiterminator
CBA	Clostridial basal agar
CBM	Clostridial basal medium
CbU	β-glucanase units
CcpA	Catabolite control protein
CCR	Carbon catabolite repression
Cel / *cel*	(referring to protein / operon or gene) cellobiose associated
DNA	Deoxyribonucleic acid
DOE	Department of energy (US)
EGU	Endoglucanase units
EI	Enzyme one (PTS)
EII	Enzyme two (PTS)
EPA	Environmental protection agency
EtOH	Ethanol
EU	European Union
F.O.B	Free on Board
fbp	Fructose 1,6 – bisphosphate
FBPase	Fructose 1,6 – bisphosphatase
FXU	Fungal xylanase units
G	Guanine
g	Grams
g/c	Gas chromotography
GC	Guanine cytosine content
GIB	Genome information broker
Glc / *glc*	(referring to protein / operon or gene) glucose associated
H	Hours
HI	Hyperl[illegible] 1 (Bioline)
HII	Hyperladder 2 (Bioline)
HPLC	High performance liquid chromatography
HPr	Histidine containing protein (heat stable protein)
Kbp	Kilo base pairs
KEGG	Kyoto Encyclopaedia of genes and genomes
L	Litres
Lac / *lac*	(referring to protein / operon or gene) lactose associated
LB	Luria-Bertoni medium
LCA	Life cycle assessment / analysis
Lev / *lev*	(referring to protein / operon or gene) fructose / levanase associated
mg	Milligrams

µg	Micrograms
min	Minutes
ml	Millilitres
µl	Microlitres
mRNA	Messenger ribonucleic acid
mSec	Milliseconds
MSW	Municipal solid waste
Mtl / *mtl*	(referring to protein / operon or gene) maltose associated
NAG	*N*-acetylglucosamine
NCBI	National centre for biotechnology information
NCIMB	National collection of industrial, marine and food bacteria
NCP	National chemical products
nm	Nanometers
OAPEC	Organization of Arab petroleum exporting countries
OD	Optical density
ORF	Open reading frame
pEB	Plasmid constructed by E. Bird
PEP	Phosphoenolpyruvate
pJA	Plasmid constructed by J. Aduse-Opoku
pJW	Plasmid constructed by J. Watson
PRD	PTS regulation domain
PTS	Phosphotransferase system
PYG	Peptone, yeast, glucose broth
RAT	Ribonucleic antiterminator target
rbs	Ribosome binding site
RCM	Reinforced clostridial medium
RegA	CcpA homologue
rRNA	Ribosomal ribonucleic acid
RTFO	Renewable transport fuel obligation
Scr / *scr*	(referring to protein / operon or gene) sucrose associated
SOC	Super optimal broth with catabolic repressor
SOP	Standard operating procedure
spp.	Species
SV	Spin or vacuum
T	Thymine
Tm	Melting temperature
TMHMM	Transmembrane helices:Hidden Markov Model
TNT	2,4,6-trinitrotoluene
TRAP	Tripartite ATP-independent periplasmic - transporter
URL	Uniform resource locator
USA	United States of America
V	Volts
VOC	Volatile organic compounds
WWI	World War I
x g	G force
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Index

Index of tables and figures

Unless stated otherwise all figures were drawn in Microsoft PowerPoint.

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Table of Contents

Chapter 1

1 Introduction

Solventogenic clostridia are a group of bacteria which can utilise a range of sugars to form solvents, such as butanol. Fermentative solvent production by clostridia was at one time the second largest biotechnological process carried out worldwide (Jones & Woods, 1986). However, many scientific and political factors influenced the decline of solvent fermentation, and today solvents are almost exclusively produced from fossil fuels (Gibbs, 1983).

A great deal of progress has been made to overcome many of the factors that make the fermentative route of solvent production unfavourable. Nevertheless, success of future commercial biobutanol production relies in part on the complete and consistent conversion of an economically competitive substrate, such as the sugars found in plant waste materials, into solvents. Characterisation of the carbohydrate transport systems of clostridia could provide vital information for the optimisation of the fermentation process.

1.1 Clostridia

The class Clostridia is comprised of three orders, Clostridiales, Halanaerobiales, and Thermoanaerobacterales. The type genus *Clostridium* is the second largest bacterial genus, presently containing 198 recognised species (Euzéby, 1997; Euzéby, 2009). Many of these strains are validly published, although the number of species will be subject to change as 16S rRNA phylogenetic information becomes available for further members (Wiegel, 2009).

Clostridia are a highly heterogeneous group both in their physiology and metabolic capabilities. The polyphyletic nature of the genus may, in part, reflect the incorrect assignment of a number of organisms to this genus. Proposals for the reorganisation of the genus have been discussed in great detail previously (Collins *et al.,* 1994; Rainey *et al* 1993; Stackebrandt *et al.,* 1999; Wiegel *et al.,* 2005; Rainey *et al.*, 2009) and will not be debated here. Most physiological characteristics of the clostridia are highly variable. In spite of this, *Clostridium spp.* generally stain Gram positive (at least in the early stages of growth), are obligately anaerobic (although oxygen tolerance can vary between species), and generate energy via substrate level phosphorylation (Mitchell & Tangney, 2005). Most clostridia form endospores and have a rod like morphology in their vegetative state (Figure 1.1).

Superkingdom: Bacteria

Phylum: Firmicutes

Class: Clostridia

Order: Clostridiales

Family: *Clostridiaceae*

Genus: *Clostridium*

Species: *Clostridium acetobutylicum*

Strain: ATCC 824

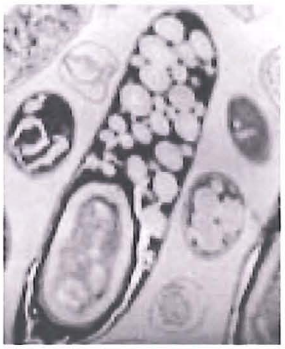


Figure 1.1: Taxonomic classification and morphology of *Clostridium acetobutylicum*

Taxonomy complied from Genbank and Bergey's Manual of Systematic Bacteriology (Volume III, 2nd Edition). Image of *C. acetobutylicum* ATCC 824 captured by Andrew Goldenkranz from URL:

http://www.genomenewsnetwork.org/resources/sequenced_genomes/genome_guide_p1a.shtml

Other phylogenetic traits include, fermentative metabolism without the capacity for dissimulatory sulphate reduction and genomic DNA with low GC content of between 22-55 mol%. The majority of clostridial species form a cluster around 28% GC genome content (Cato & Stackebrandt, 1989; Cato *et al.,* 1986; Keis *et al.,* 2001b; Jones & Keis, 2005).

Bacterial endospores are the most robust cell form known. Endospores can guard a cell against heat, radiation, dehydration, oxygen and alcohol stress, and are produced by members of the *Bacillus* and *Clostridium* genus (Dürre, 2005). This fact may help to account for the diverse habitats that these anaerobes are found in, ranging from soils and sediments to rumens and rhizospheres. Ecological niches in which clostridia have been isolated also vary greatly with respect to temperature, such as the psychrophilic *C. vincentii* which was isolated from pond sediment in Antartica (Mountfort *et al.,* 1997) and the thermophilic *C. thermohydrosulfuricum* isolated from a sewage plant in Georgia (Wiegel *et al.,* 1979).

Clostridia are versatile in utilisation of carbon and energy sources. Some species are able to excrete extracellular enzymes to degrade polysaccharides to simple sugars (saccharolytic), others are able to hydrolyse proteins and catabolise amino acids (proteolytic), while some species can do both (Cato & Stackebrandt, 1989).

Clostridia characteristically produce a mixture of organic acids, alcohols and gas from carbohydrates or peptones.

1.1.1 Prominent members of the genus

Several species of *Clostridium* can arise as food spoilage organisms or as pathogens. Some *Clostridium* can produce powerful toxins, for example botulinal neurotoxins, which are the most toxic proteins known to humankind (Montecucco & Molgó, 2005). The recent sharp increase in Healthcare-Associated Infections caused by *C. difficile* has attracted public scrutiny and presents an ongoing management issue for healthcare professionals (Warny *et al.,* 2005; McDonald *et al.,* 2005). Other pathogenic members of the genus include *C. tetani,* the causative agent of tetanus, *C. perfringens* and *C. septicum* that can cause gas gangrene, and *C. botulinum* which causes botulism (Popoff & Stiles, 2005).

Despite the pathogenic members of the group there are species of *Clostridium* which are proven to be of great biotechnological value. Several species have been used in the biodegradation of hazardous materials such as TNT, nitroaromatic explosive compounds, chlorinated hydrocarbons and toxic metals (Ahmad *et al.,* 2005). Collagenases isolated from *C. histolyticum* have been used successfully to aide wound repair (Brett, 2005). Research has also been conducted, in which species of *Clostridium* are manipulated to produce therapeutic proteins for tumour treatment (Minton *et al.,* 2001; Minton & Anné, 2005). Of particular interest are the *Clostridium* species that can perform large-scale industrially significant fermentations.

1.1.2 Solventogenic clostridia

Industrially important solvents such as butanol, acetone, ethanol and isopropanol can be naturally formed by a number of clostridia (Shaheen *et al.*, 2000; Dabrock *et al.*, 1992). Type and ratio of solvents produced can vary depending on the species. The solvent range and production ratio for the model solventogenic organism, *C. acetobutylicum* is typically 6:3:1 butanol, acetone and ethanol (Jones & Woods, 1986). Other clostridia have been isolated which produce a different range of solvents, such as *C. aurantibutyricum* which produces acetone, isopropanol and butanol (George & Chen, 1983), or *C. tetanomorphum* which produces only butanol and ethanol (Gottwald *et al.,* 1984). Many species of clostridia are saccharolytic and can efficiently convert a wide range of sugars found in organic waste materials to solvents. This interesting ability has provided the stimulus for the investment of a considerable amount of research to

elucidate the solvent production pathways and the mechanisms and regulation of carbon catabolism in these organisms.

1.2 Acetone, butanol and ethanol (ABE) fermentation

1.2.1 History of ABE production

Harnessing the metabolic capabilities of microbes to produce alcohol for human use is not a new phenomenon. The ancient use of yeast to produce potable ethanol throughout history is well documented (Bud, 1993; Dixon, 1996). In 1862, Louis Pasteur reported the bacterial production of butanol (Pasteur, 1862). Several years later, in 1905, the isolation of *Bacillus macerans* which, fermented various carbohydrates with the formation of ethanol, acetone and acetic acid was reported (Schardinger, 1905). During the early part of the 20th century, several patents concerning fermentative ABE production were filed on new production strains, improvements to the growth media and fermentation processes (Fernbach & Strange, 1911a; Fernbach & Strange, 1911b; Fernbach & Strange 1912a; Fernbach & Strange, 1912b).

During World War I (1914-1918) acetone (a colloidal solvent for nitrocellulose) was required in vast quantities for the manufacture of cordite, a smokeless propellant, used by the British army (Jones & Woods, 1986). This previously unprecedented demand for acetone provided the opportunity for the advancement and expansion of the fermentative ABE industry. In 1915, Chaim Weizmann isolated a superior organism (originally termed *B. granulobacter pectinovorum*, later termed *C. acetobutylicum*) and patented an improved fermentation process (Weizmann, 1915; Gibbs, 1983; Jones & Woods, 1986).

The Weizmann process, as it became known, delivered greatly enhanced solvent yields compared to previous efforts and in February 1915 the British Government adopted the process to provide solvents for the War effort (Gibbs, 1983). A timeline of these and other significant events affecting the development, decline and subsequent regrowth of the large-scale ABE fermentation process are presented in Figure 1.2.

During WWI, there was no economical use for butanol; considered a by-product of acetone formation it was stored in huge vats (Killeffer, 1926). When prohibition was introduced in the United States of America (USA) in 1920, it brought about a shortage of amyl alcohol (a by-product of yeast ethanol fermentation used in quick drying lacquers) required for the growing car industry. The excess butanol turned out to be the perfect replacement (Dürre, 2008).

The Weizmann process made such a positive contribution to the War effort that Lloyd George (then British Minister of Munitions) offered to ask the Prime Minister to recommend Chaim Weizmann for an appropriate honour. Weizmann declined anything for himself and instead spoke of his involvement in the Zionist movement and "the aspirations of the Jewish people for repatriation in their sacred land" (Reinharz, 1985). Shortly after Lloyd George became Prime Minister and Earl Balfour, the then Foreign Secretary, proclaimed the 1917 Balfour Declaration, which established Palestine as a home for the Jewish people (Dixon, 1996). This eventually led to the foundation of The Independent State of Israel, on the 14th of May 1948, with Chaim Weizmann as the first president. On this day, the first in a succession of Arab-Israeli Wars erupted (Upshall, 1994).

At this time biotechnological production of ABE by solventogenic clostridia was the second largest industrial fermentation process in the world, second only to yeast derived alcohol produced for the alcoholic beverage industry (Dürre, 1998). However, around 1950 the ABE fermentation process began to decline in Britain and the USA, so that by 1960 ABE was almost exclusively produced by the petrochemical route (Gibbs, 1983; Jones & Woods, 1986). Some major exceptions to this were ABE fermentation plants in the former USSR and South Africa which were in production until the late 1980s (Zverlov *et al.*, 2006; Gapes, 2000), and production plants in China, which in one case was not stepped down to complete closure until 2004 (Chiao & Sun, 2007). The rapid expansion and economical competitiveness of the maturing petrochemical industry coupled with a number of key factors, which are discussed in the next section, played a major role in the worldwide decline of fermentative ABE production (Hastings, 1971).

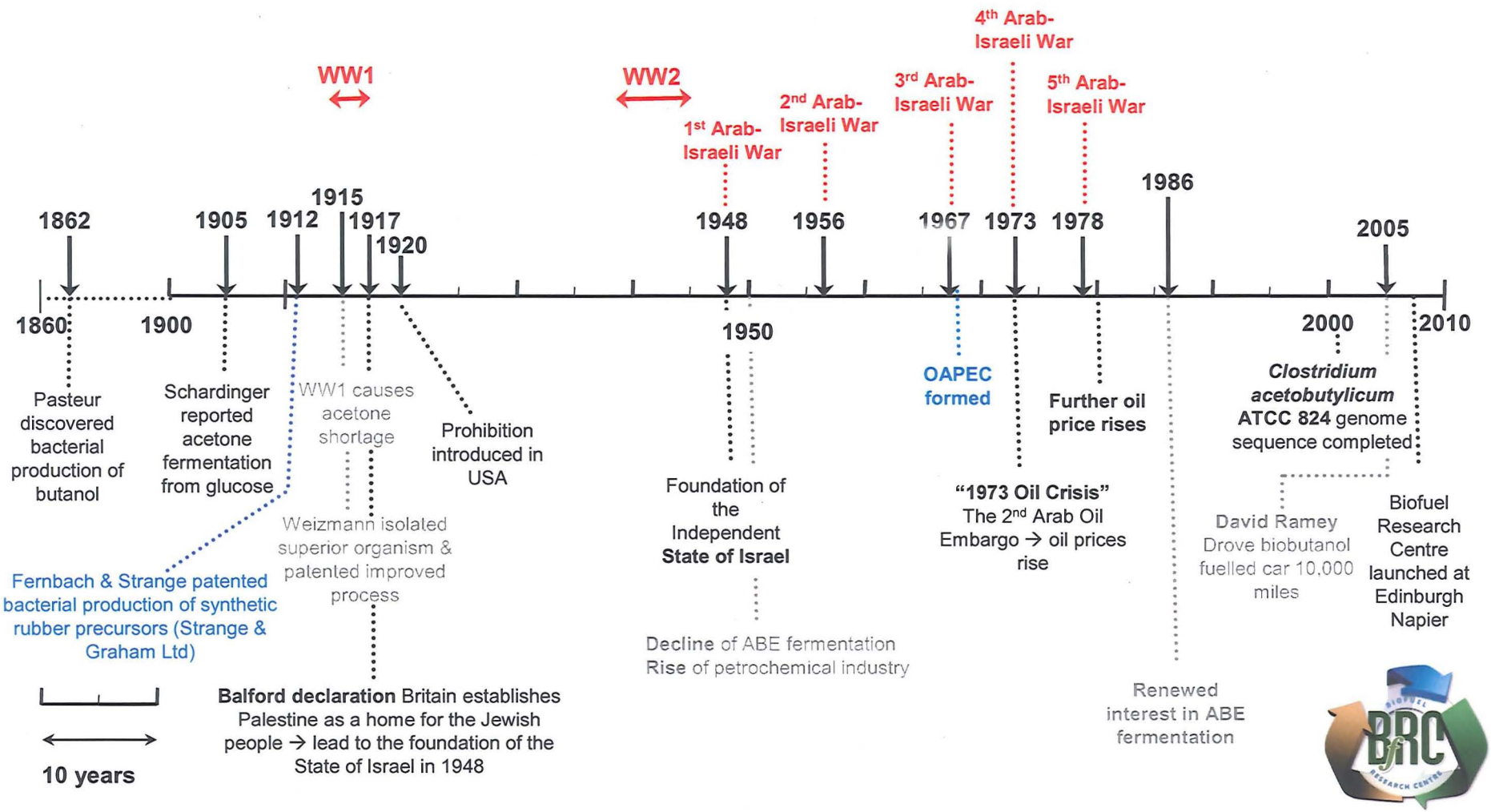


Figure 1.2: Time line of significant discoveries and events

The Organization of Arab Petroleum Exporting Countries (OAPEC) was established in 1968 (Nakhleh, 2010). During the fourth Arab-Israeli War (Yom Kippur War), in response to the U.S decision to re-supply the Israeli military, OAPEC proclaimed an oil embargo, which led to a drastic increase in oil prices.

The ongoing instability in oil prices and concerns over energy security has become a major factor in the renewed interest in ABE fermentation (Jones & Woods, 1986). The economic and practical limitations of the traditional ABE fermentation process need to be addressed before revival of this technology for biofuel production is possible.

1.2.2 Economic considerations

A major economic consideration for the revival of the ABE fermentation process is the initial capital cost of the production plant (Gapes, 2000). Previously, it has been calculated, that in the majority of cases the initial cost of a fermentation plant is lower than the cost of a plant for the petrochemical synthesis of solvents (Solomons, 1976). However, solventogenic clostridial fermentation requires sterile feedstocks and reactors, which necessitates a lot of energy, and thus adds a significant economic burden to the overall process.

Butanol concentrations achieved by traditional fermentation are usually low (maximum 2%), which results in high recovery costs, as it requires a large input of energy (Gapes, 2000). Traditionally ABE was recovered by distillation after a batch fermentation process, which was expensive (Dürre, 1998). Since the days of large-scale ABE fermentation, many advances have been made in solvent recovery and reactor technology. Methods such as pervaporation (Groot *et al.,* 1992; Geng & Park, 1994) gas stripping (Qureshi *et al.,* 1992) and membrane evaporation (Gapes *et al.,* 1996) have proven to vastly reduce the energy input needed to recover solvents and thus the associated economic burden.

The fermentation process also generates a large volume of effluent which has high biological oxygen demand (BOD). This initially posed an expensive waste disposal problem for the traditional plants (Jones, 2001). National Chemical Products (NCP), based in South Africa, operated a commercial ABE fermentation process from 1935 until 1982 and, developed several methods for dealing with the high BOD waste. The most successful method of mitigating the large volumes of waste, transpired to be evaporation of the distillation slops under vacuum and then resultant concentrate was spray dried. Around 97% of the resultant waste was sold on as animal feed (Jones, 2001).

Enzymatic hydrolysis or another pretreatment, such as steam explosion or acid hydrolysis of a complex fermentation feedstock to individual mono or disaccharides can be expensive (Claassen *et al.*, 2000). Feedstock price has the largest impact on the economic viability of the ABE fermentation process, often accounting for around 60% of the total cost of the procedure (Lenz & Morelra, 1980; Jones & Woods, 1986). Cost of purchase, transport and processing of a substrate must be considered in the life cycle analysis (LCA) of a chosen feedstock. The economic considerations are summarised below.

- Initial capital cost of the plant
- Sterilisation of fermentors and feedstocks
- Low butanol concentration, high cost of recovery
- Effluent processing
- Enzymatic hydrolysis / pre-treatment of feedstock
- Price of feedstock

1.2.3 Practical considerations

A practical limitation which has a huge economic impact is the incomplete utilisation of the fermentation feedstock. As substrate can account for over 60% of the cost of the fermentation process the factor of incomplete sugar utilisation needs to be addressed. Bacteria are subject to a phenomenon referred to as carbon catabolite repression (CCR - the mechanisms of this is discussed in detail in Section 1.5.3.1). A better understanding of the genes involved in carbon transport and utilisation, as well as the regulation they are subject to, may help to improve greatly our ability to manipulate fermentation conditions and improve strains to obtain more complete substrate utilisation and thus greatly reduce cost.

A major practical limitation of the traditional batch ABE fermentation process is that butanol is highly toxic to *C. acetobutylicum* causing; inhibition of growth and glucose uptake, disturbing the cells ability to maintain internal pH and lowering the intracellular ATP (Bowles & Ellefson, 1985). This can be mitigated to an extent by using one of the solvent stripping techniques described above. Another and often devastating, practical limitation is that bacterial fermentations can be subject to bacteriophage infection (Jones *et al.,* 2000). Typical symptoms of phage infection include slow fermentations and reduced solvent yields. Continuous research into phage resistant strains and reactor design has helped to mitigate this issue (Jones *et al.*, 2000).

Strain degeneration can also have a detrimental effect on the fermentation process. When this occurs clostridia can lose the ability to form solvents or develop spores, in the case of *C. acetobutylicum* degeneration is linked to the loss of the pSOL megaplasmid (Ezeji *et al.,* 2005). Interestingly, *C. beijerinckii* does not harbour such a plasmid, yet is still susceptible to degeneration of solvent and spore forming abilities (Kashket & Cao, 1993). The practical considerations are summarised below.

- Incomplete utilisation of feedstock
- Butanol highly toxic (maximum 2%)
- Fermentations subject to contamination or phage infection
- Strain degeneration

An international effort is underway to address many of the old limitations of the ABE fermentation process. The next section will examine some of the drivers which have brought about this renewed market for bacterially produced ABE.

1.3 Carbon, cars and carbohydrates

Carbon is the chemical basis of life on Earth, and it is the most abundant element in any living cell. Every living organism is composed of analogous arrangements of carbon atoms linked with oxygen, hydrogen and other carbon atoms.

The global carbon cycle is the circulation of carbon between living organisms, terrestrial ecosystems, the oceans and the atmosphere (Figure 1.3). Several significant carbon reservoirs exist in the form of buried coal and oil (fossil fuels), calcium carbonate ($CaCO_3$) deposits at the bottom of the oceans, and mature forests (Freeman, 2002).

Concerns about climate change and industrial production of CO_2 causing deleterious effects to the atmosphere were raised as far back as 1957 (Revelle & Suess, 1957). Modern production of CO_2 continues to be a major source of debate. However, it has been demonstrated that the use of fossil fuels contributes greatly to the radiatively active trace gases (CO_2, methane, nitrous oxide and ozone) in the atmosphere (Prinn *et al.,* 2000, Watson *et al.,* 1996).

There has been a dramatic rise in atmospheric CO_2 levels since the increase in the use of fossil fuels in the 1950's (Figure 1.4). The vast majority of fossil fuels which are consumed annually are used in the transport sector (Demirbas, 2007). Substitution of transport fuels produced by the petrochemical route, with biofuels derived from renewable waste biomass, could potentially present an opportunity to avert further increase in atmospheric CO_2. This is due to the source of carbon which the fuels are derived from. Biofuels are formed from carbon which has been recently captured by the growth of new biomass, whereas fossil fuels are produced from carbon which has been locked away from the atmosphere for millions of years (Sims *et al.,* 2006).

In 2005 Dr David Ramey conducted a 10,000 mile demonstration run across the USA in a unmodified 1992 Buick, fuelled solely on biobutanol. The car's emissions and performance on this fuel were independently tested in 10 states by the Environmental Protection Agency (EPA) and production of hydrocarbons, carbon monoxide and oxides of nitrogen were found to be significantly reduced. Additionally, the cars mileage was increased by 9%, averaging 24 miles per gallon. Ramey's patented biobutanol procedure is a two-stage, dual-path process. Utilising, in the first stage, *C. tyrobutyricum* to ferment biomass feedstock to butyric acid and hydrogen, and in the second stage *C. acetobutylicum* assimilates the butyric acid and converts it into butanol (Ramey, 1998).

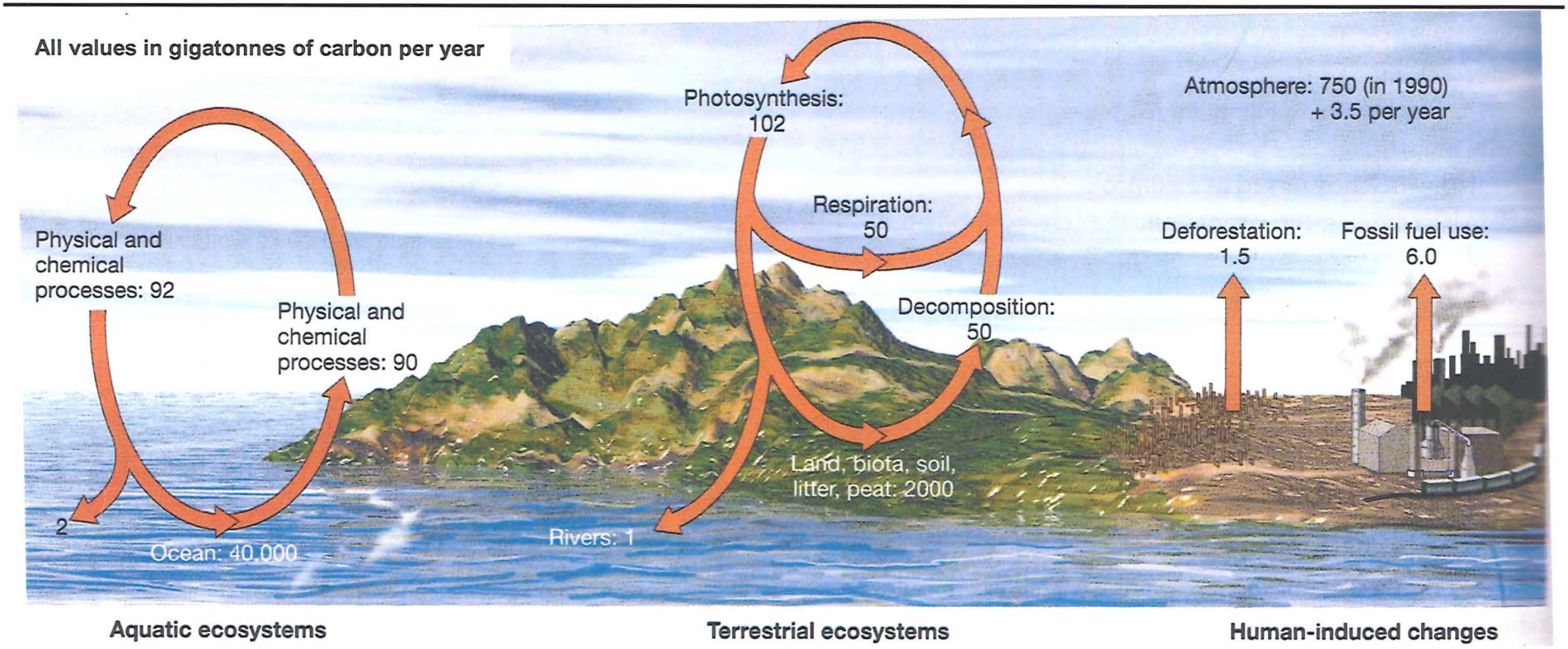


Figure 1.3: The global carbon cycle

Taken from Freeman, 2002. The arrows indicate how carbon moves into and out of ecosystems. Deforestation and fossil fuel use are major human induced changes to the cycle. The values next to the labels are in gigatonnes of carbon per year.

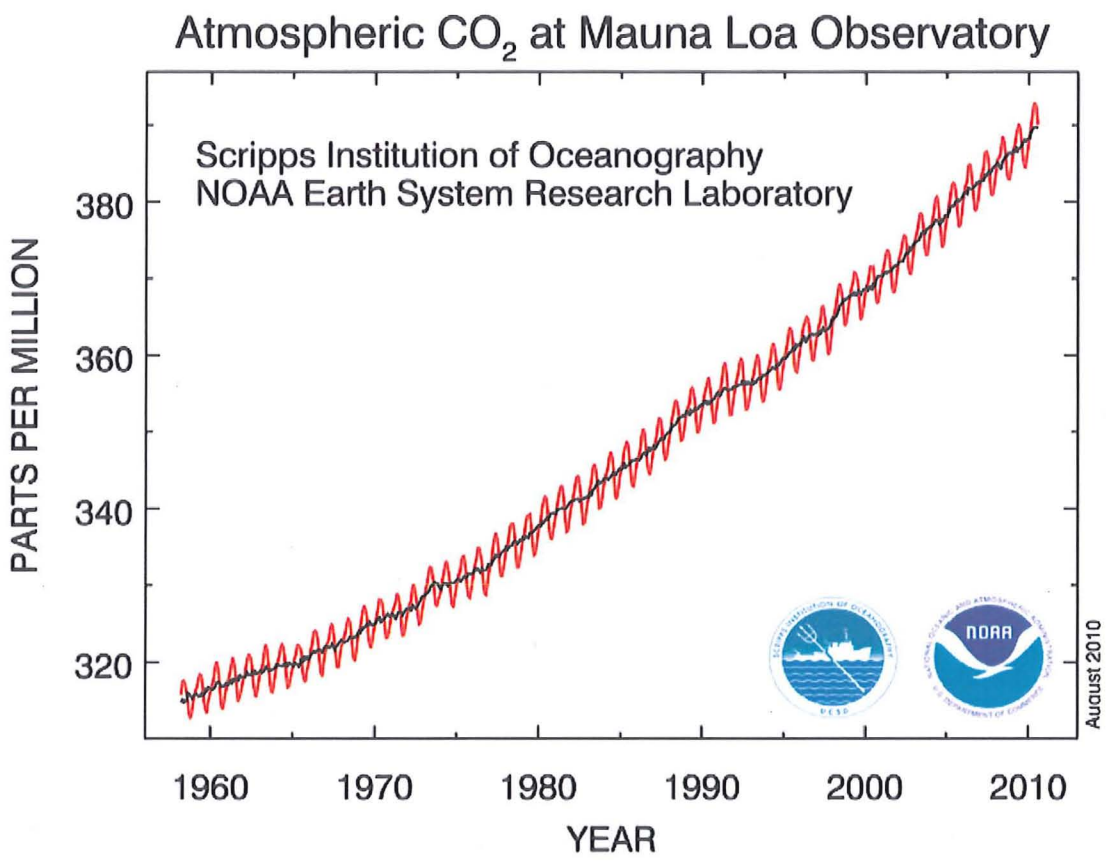


Figure 1.4: The atmospheric CO_2 concentrations recorded in the Mauna Loa observatory

Atmospheric CO_2 concentrations have been recorded in the Mauna Loa Observatory since 1958. Dr. Pieter Tans, NOAA/ESRL (www.esrl.noaa.gov/gmd/ccgg/trends/)

1.3.1 Drivers to identify an alternative fuel source

There have been several recent severe fluctuations in crude oil prices (Appendix 7.11). The spot price of oil remained relatively stable, at under $30 per barrel until April 2004, by April 2006 the price had risen to $69.44 and by July 2008 it had nearly doubled to $133.37, which has had a major influence on the price consumers pay at the pump.

As mentioned in the last section, the global carbon cycle is being affected because of disruption of the major carbon reservoirs. In addition to erratic oil prices and increased CO_2 production, energy security and rural economic development are also governmental drivers which have accelerated biofuel policy (Wooley *et al.*, 1999; Charles *et al.*, 2007).

1.3.2 Commercialisation of biofuels

Numerous venture companies have been established with the aim of commercialising biofuels and many of the large oil companies are now likewise developing biofuel subsidiaries. A joint venture between DuPont and BP is looking to develop a fermentative butanol process (BP, 2006). Shell and Virent Energy Systems are working on production of biogasoline and hydrogen from biomass (Shell, 2007). Another joint venture is being undertaken by Virgin Fuels, Khosla Ventures and Gevo, which is also endeavouring to commercialise biobutanol production (Virgin, 2007). The Biofuel Research Centre was launched at Edinburgh Napier University in 2007, with the bold mission statement "to operate as an internationally recognised portal to serve and supply the diverse sectors that will collectively underpin the development and large-scale production of the next generation of biofuel in the UK and beyond".

1.3.3 Biofuels

There are a variety of microbial-based transport biofuels currently available on the commercial market, as well as some proposed molecule types which are still in development (Figure 1.5). Most available biofuels are in liquid form - excluding hydrogen, methane and propane that are gaseous at 20°C. Hydrogen has the highest gravimetric energy density of any known fuel and does not produce carbon-based emissions during energy conversion, which makes it a promising prospect for the future of biofuels. Several major technical hurdles need to be overcome before hydrogen becomes an economical fuel, ranging from production volumes to storage and distribution issues (Levina *et al.,* 2004).

Biofuel research is expanding so fast that the terminology can often vary in meaning. A 1st generation biofuel generally refers to a liquid biofuel produced from a specifically grown biomass feedstock. In many cases, the feedstock could potentially be used for human nutrition and the upstream processing is often wasteful and utilises only part of the plant. An example of a 1st generation biofuel is bioethanol production from sugar cane (Bomb *et al.,* 2007). The term 2nd generation biofuel generally refers to a liquid or gas biofuel produced from either specifically grown non-edible crops, which utilises the vast majority of the plant or from waste agricultural residue (such as biobutanol production from waste cellulosic biomass).

Not all biofuels are beneficial when their full environmental impacts are assessed (Scharlemann & Laurance, 2008), therefore there needs to be a clear distinction between 1st generation, 2nd generation and truly sustainable biofuels, to avoid losing public support and confusing governmental policy. A model, sustainable biofuel would be produced from a biomass feedstock that does not compete for land, require fertilisers or need to be transported long distances.

A model sustainable biofuel would be produced from a regional, waste biomass feedstock, negating the need to transport the feedstock huge distances or to compete for land with food crops or incur the use of additional fertilisers. Ideally, a biofuel would be produced on, or near to, the source of the biomass residue (Anderson & Fergusson, 2006).

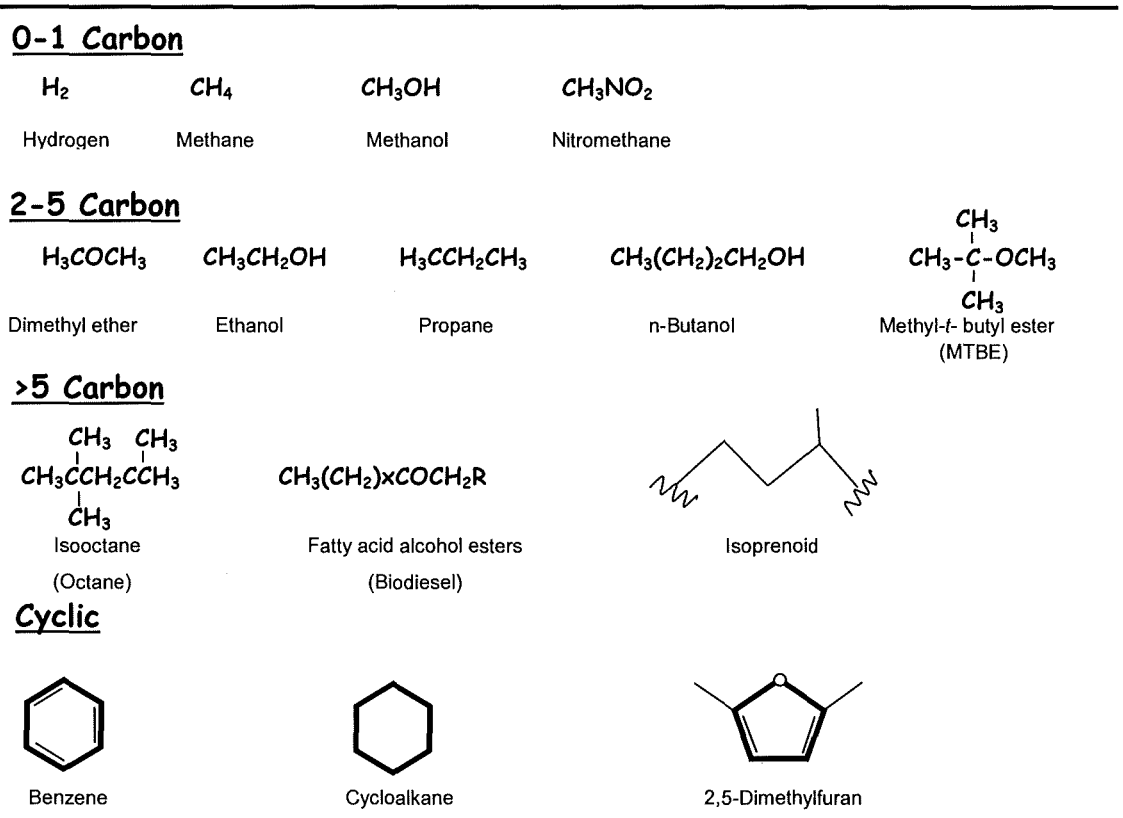


Figure 1.5: Gallery of current or proposed microbial biofuel molecules

Figure adapted from Wackett (2008). *C. acetobutylicum* can naturally from ethanol, butanol and hydrogen.

1.3.4 Butanol as a biofuel

Butanol is a four carbon, primary alcohol with the molecular formula C_4H_9OH (MW 74.12). It is a flammable, colourless liquid and butanol's main properties are summarised in Table 1.1. Applications of butanol include use as a solvent for a range of products or as a chemical intermediate for producing many chemicals and plastics, including safety glass, hydraulic fluids and detergent formulations (Table 1.1). Butanol has received a great deal of interest recently as a potential fuel or fuel additive and has been the subject of several reviews (Dürre, 2007; Dürre, 2008; Ezeji *et al.,* 2004; Ezeji *et al.,* 2007; Lee *et al.,* 2008). Between 2000 and 2005 the demand for butanol grew by 1.9% per year and has a projected market expansion of 3% per year thereafter (Kirschner, 2006).

There are three main methods of chemically synthesising butanol; 1) Reppe process 2) oxo synthesis and 3) crotonaldehyde hydrogenation. In the Reppe process propylene, carbon monoxide and water are reacted in the presence of a catalyst (tertiary ammonium salt of polynuclear iron carbonlyl hydrides). The second method, oxo synthesis or hydroformylation, carbon monoxide and hydrogen are added to a double bond in propylene (in the presence of a catalyst) to form aldehyde mixtures (n-butyralehyde and isobutyraldehyde), which are then hydrogenated to produce n-butanol and isobutanol. Crotonaldehyde hydrogenation is a process which consists of aldol condensation, dehydration and hydrogenation. This differs from the other two methods in that the precursor is ethanol rather than petroleum (Kirschner, 2006; Bochman *et al.,* 1999; Falbe, 1970).

The first report of microbial butanol production was in 1862 by the pioneering microbiologist Louis Pasteur (Pasteur, 1862). He referred to the bacteria which carried out this synthesis as "*Vibrion butyrique*", now thought of as consisting of a mixed culture of containing *Clostridium butyricum* or a closely related species (Dürre, 2008). Several species of *Clostridium* have the ability to produce butanol and other solvents from simple sugars or complex substrates. The most extensively characterised solventogenic clostridia is *Clostridium acetobutylicum* ATCC 824 (Dürre, 2008; Nölling *et al.,* 2001; Girbal *et al.,* 1995; Desai *et al.,* 1999; Girbal & Soucaille, 1995).

Table 1.1: The properties of n-butanol
Adapted from the technical data sheets for n-butanol by the Dow Chemical Company, BASF and Sigma-Aldrich.

Properties of Butanol	
Melting point	-89.3 (°C)
Boiling point	117.7 (°C)
Ignition temperature	35 (°C)
Flash point	365 (°C)
Density at 20°C	0.8098 (g/mL)
Critical pressure	48.4 (hPa)
Critical temperature	287 (°C)

1.3.5 Carbohydrates

As mentioned in the previous section, clostridia can convert simple sugars into solvents such as butanol. Basic carbohydrates are made up of carbon, hydrogen and oxygen atoms. Carbohydrates are utilised for an abundance of major functions in living systems, such as deoxribose which is a building block of DNA. Carbohydrates are a vast and diverse group of organic compounds, which, on the basis of molecular structure, can be classified into four major groups: monosaccharides, disaccharides, glycosides and polysaccharides.

Simple sugars, containing from three to seven carbon atoms, are referred to as monosaccharides, and can be grouped according to the number of carbon atoms they contain. For example; trioses (three carbon sugars), tetroses (four carbon sugars), pentoses (five carbon sugars), hexoses (six carbon sugars) and heptoses (seven carbon sugars). Figure 1.6 contains Fischer and abbreviated Haworth projections of some common monosaccharides (Sturgeon, 2002).

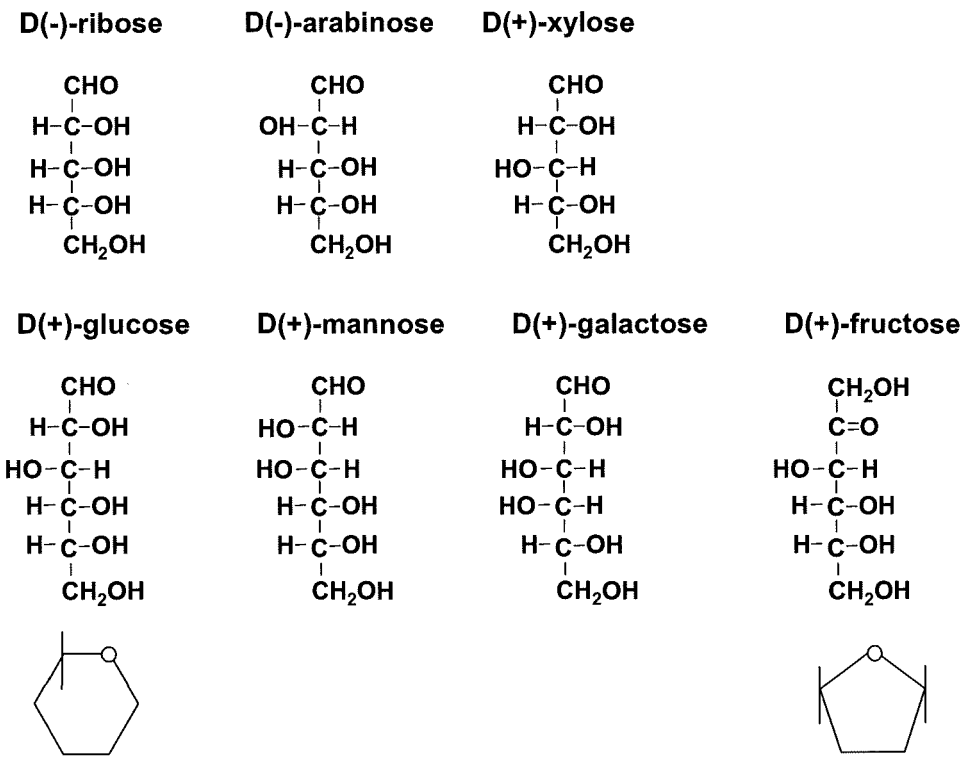


Figure 1.6: Fischer and abbreviated Haworth projections of biologically important monosaccharides

Many of these monosaccharides are found in large quantities in agricultural waste streams.

Disaccharides are formed when two monosaccharides bond in a dehydration synthesis reaction. Examples of common disaccharides are sucrose (table sugar), consisting of a molecule of glucose and a molecule of fructose, and lactose (milk sugar) which is made up of glucose and galactose. Glycosides are a group of saccharides, abundant in plants, which usually contain a non-carbohydrate moiety. The structure of the common glucosides, arbutin, salicin and cellobiose are depicted in Figure 1.7.

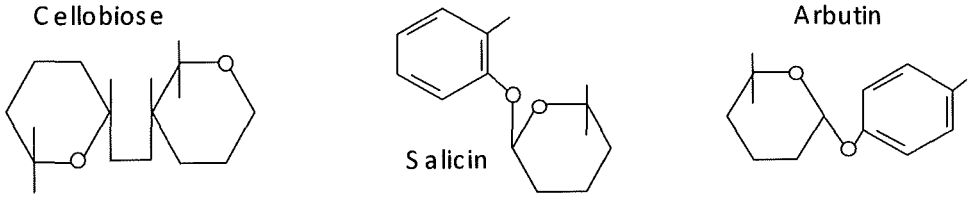


Figure 1.7: Abbreviated Haworth projections of the major plant β-glucosides

Cellobiose, arbutin and salicin are plant sugars belonging to a group of saccharides called β-glucosides. They are comprised of a glucose moiety joined to another unit by a glucosidic bond. Cellobiose, the most abundant β-glucoside consists of two β-1,4 linked glucose residues.

1.3.5.1 The structure of polymers

There are three broad classifications of polymer. Biopolymers, the first group, consists mainly of proteins, nucleic acids and polysaccharides; The second group consists of chemically modified natural polymers (such as cellulose acetate fibres); Third are the entirely synthetic polymers (such as polyethylene, polyvinyl chloride and nylon). A polymer consisting of a single monomer unit is termed a homopolymer, and a polymer consisting of different species of monomer is termed a heteropolymer (Dawber & Moore, 1985).

1.3.5.2 Polysaccharides

Polysaccharides are biopolymers, which may contain many hundreds of monosaccharides. Polysaccharides are widely distributed in nature and are important for energy storage and cell structure in living organisms. Accessing the sugars contained within polysaccharides is one of the first steps required in the fermentation of biomass to solvents.

Cellulose is the most abundant biopolymer on earth, the main component of plant cell walls (Albertsson & Huang, 1995): it is a linear glucose homopolymer joined by β-1,4-glucosidic linkages (Figure 1.8). Glucosidic bonds are very stable and cellulose forms long chains, which are arranged in parallel bundles. Cellulose fibres then have increased stability due to inter-fibre hydrogen bonds, thus crystalline cellulose is particularly resistant to breakdown. A schematic diagram of enzymatic hydrolysis of cellulose is depicted in Figure 1.9. Cellulose hydrolysis by endo- and exo-glucanases and β-glucosidases results in a mixture of cello-oligosaccharides, cellobiose and glucose.

Starch is a homopolymer of glucose, which is made up of amylose (linear chains of α-1,4-linked glucose residues) and amylopectin (branched at α-1,6-linked glucose residues). Starch is found in roots and seeds, an important energy storage for[illegible]gure 1.8). Chitin, a homopolymer of β-N-acetoglucosamine, is the second most abundant biopolymer on earth, and it is found in exoskeletons of insects, shells of crustaceans and the cell wall of many fungi and algae.

Cellulose, starch and many similar polymers are found in large quantities in plant waste material and may provide a plentiful and potentially economically viable fermentation feedstock for solvent production. Specific prospective renewable, inexpensive fermentation substrates will be discussed below.

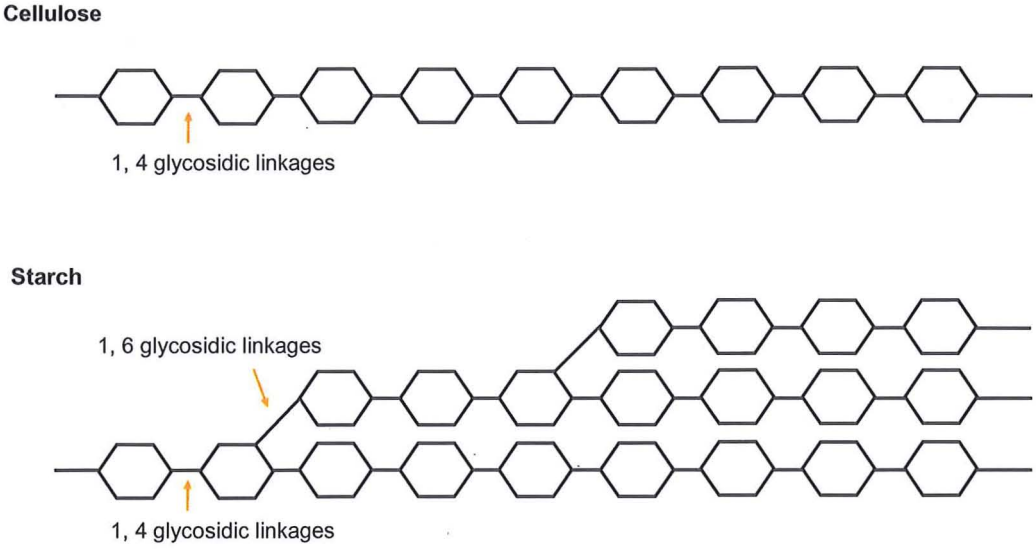


Figure 1.8: Two polymers present in large quantities in plant biomass

Schematic diagram of two important polymers isolated from plant biomass. Cellulose, the main component of plant cell walls, is a homopolymer of glucose connected by β-1,4 glucosidic linkages. Starch, also a homopolymer of glucose, is found as a storage carbohydrate in roots and seeds. Starch is made up of amylose (linear chains of α-1,4-linked glucose residues) and amylopectin (branched at α-1,6-linked glucose residues).

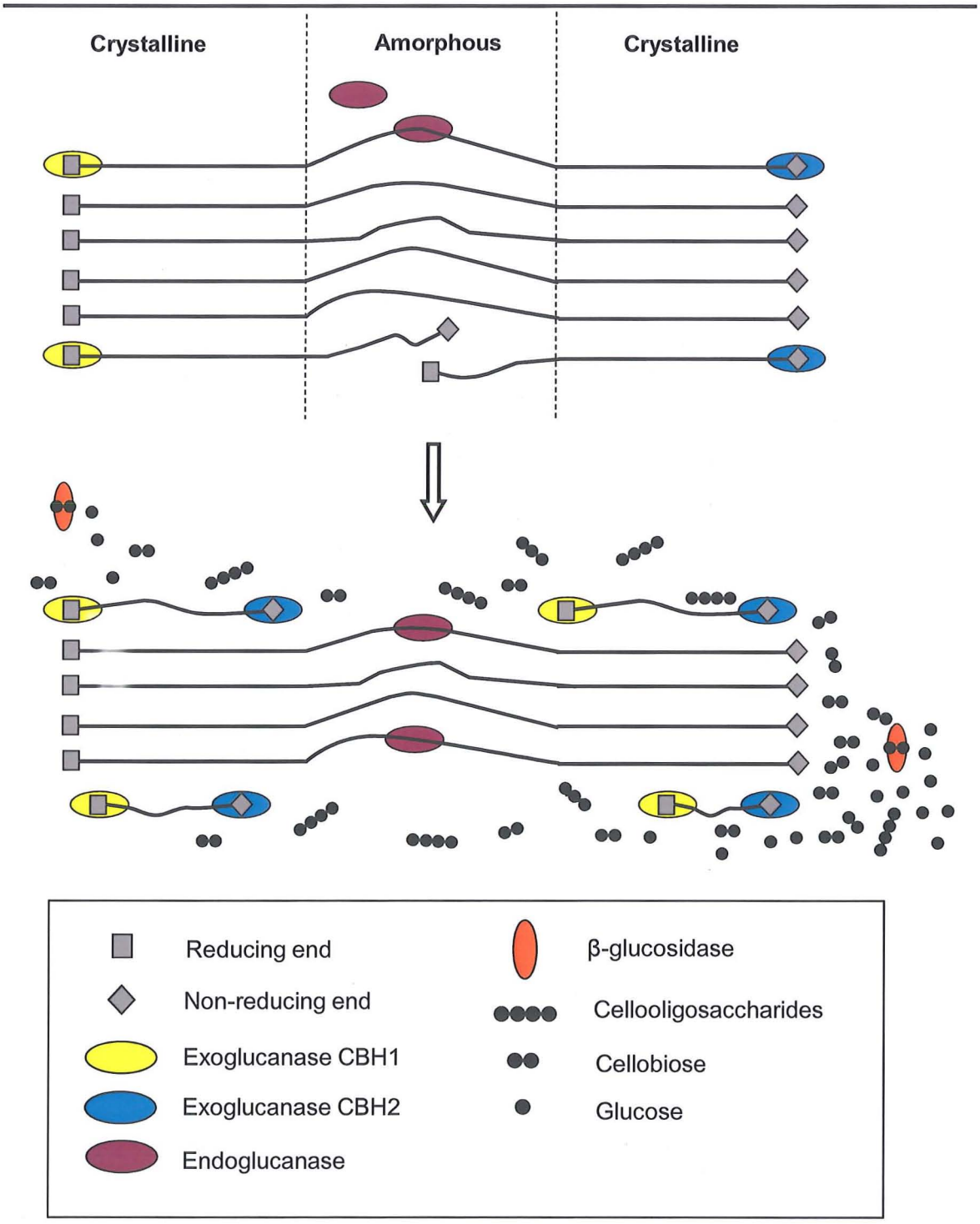


Figure 1.9: Schematic representation of the enzymatic hydrolysis of amorphous & crystalline cellulose.

Cellulose, enzyme and hydrolytic products are not shown to scale. Cellulose is hydrolysed in to cellooligosaccharides, cellobiose and glucose. Adapted from Lynd *et al.*, 2002.

1.3.6 Fermentation feedstocks

Traditional fermentation substrates for ABE production include molasses and starch from materials such as maize, wheat, millet and rye. During the 1950s, molasses (a by-product from the sugar industry) became economically unfavourable as a fermentation feedstock. This was partly due to an increase in the efficiency of the extraction process, which resulted in low sugar content and partly due to the inclusion of molasses in cattle-feed which consequently lead to an elevated market price (Dürre, 2008; Jones & Woods, 1986).

Potential food scarcity, caused by biofuel crops competing for arable land, which could be used for human nutrition, has polarised sections within the environmental lobby. However, it is now generally accepted that poverty, not food availability, is the most widespread cause of food insecurity (Smith, *et al.*, 2000; Cassman & Liska, 2007).

Fermentation feedstocks produced from specifically grown crops have a large environmental impact due to their requirement for land, water and fertiliser. In addition there is associated air pollution caused in the harvest of the feedstock crop. For a biofuel product to be considered carbon neutral, the entire LCA must be taken into account: from land re-appropriation, fertiliser use, energy used in harvest, transportation and processing of the feedstock right to the actual fermentation process. Utilisation of agricultural waste products or alternative abundant biological waste streams greatly cuts down on many of these factors. It is the responsibility of scientific researchers to inform and educate the policy makers to make the distinction between contrasting methods of production, and different types of biofuels (Gibbs, 1983; Ezeji *et al.*, 2007; Cassman & Liska, 2007; Dürre, 1998).

Market price of the chosen feedstock is one of the major contributing factors negatively affecting the economic viability of fermentative ABE production (Gapes, 2000; European Commission, 2005; Hill *et al.,* 2006). Therefore, for a fermentation process to be economically attractive there should be as close to complete sugar utilisation as possible. Hence the importance of understanding the regulation and genetics of sugar utilisation. Many potential substrates have been evaluated as feedstocks for fermentative solvent production with varying degrees of success. Some of the more interesting cases are discussed below.

1.3.6.1 Apple pomace as a fermentation feedstock

Butanol production by *C. acetobutylicum* and *C. butylicum* using fresh apple pomace as a feedstock have been reported. Apple pomace is the pulpy material remaining after the juice removal process is complete. Butanol yields of up to 2.2% from fresh apple pomace were obtained (Voget *et al.,* 1985). Interestingly it has also been reported that after fermentation, apple pomace becomes a more nutritional animal feed than unfermented apple pomace (Sandhu & Joshi, 1997; Hang *et al.,* 1981), thus adding to its merit as a fermentation feedstock.

1.3.6.2 Spent grains as a fermentation feedstock

Spent grains (SG) is a biomass by-product produced by the potable alcohol industry. Up to 80% of SG is composed of lignocellulosic materials, which can be hydrolysed to its component sugars and used for subsequent fermentation. SG were investigated for their suitability as a biomass source for bioethanol production by yeast (White *et al.,* 2008). SG sourced from Scottish distilleries and breweries, were hydrolysed to the component pentose and hexose sugars and fermented by the yeasts *Pichia stipitis* and *Kluyveromyces marxianus.* Ethanol yields of 4.2g and 3.0g of ethanol per 100g of SG were obtained, for *P. stipitis* and *K. marxianus*, respectively (White *et al.,* 2008).

1.3.6.3 Glycerol as a fermentation feedstock

Crude glycerol is a significant by-product of the biodiesel industry. This by-product is not however pure glycerol, but a contaminated mixture of glycerol (50-80%), water, methanol, catalyst, un-reacted oils and both organic and inorganic salts (Thomson & He, 2006). Due to the nature of the crude glycerol produced, it is economically prohibitive to purify it for use in cosmetic or pharmaceutical industries.

The large quantities produced and limited economically viable uses for the by-product, has led to market saturation and an artificially low price for the product (Yazdani & Gonalez, 2007). Crude glycerol could potentially provide a cheap, abundant, sustainable feedstock for solvent production, especially if linked with biodiesel production facilities. It has been reported that *Clostridium* species can ferment glycerol to 1, 3-propanediol (Forsberg, 1987). *C. butyricum* was found to be the best producer of 1, 3-propanediol and the only microorganism to use a coenzyme B12-independent glycerol dehydratase (Gonzalez-Pajuelo *et al.,* 2005). The team set out to improve the strain for a more economical process of 1, 3-propanediol production by metabolically engineering the strain. They were unsuccessful in genetically manipulating *C. butyricum,* but

successful in metabolically engineering *C. acetobutylicum*, by introducing the 1, 3-propanediol pathway of *C. butyricum* resulting in a *C. acetobutylicum* strain which could produce 1, 3-propanediol from glycerol on an industrial scale.

When provided with glycerol as a sole source of carbon and energy *C. pasteurianum* LMG 3285 was reported to convert more than half the glycerol provided to n-butanol with significant concurrent H_2 production (Heyndrickx *et al.,* 1991). Shortly after that Dabrock *et al.,* (1992) investigated the parameters (such as pH and iron & phosphate limitation) affecting the profile of neutral solvent formation by *C. pasteurianum* utilising glycerol as a feedstock.

They found that in continuous culture of *C. pasteurianum* under phosphate limitation the major fermentation products were ethanol, butanol and 1, 3-propanediol and that acetate and butyrate were only produced in trace amounts. Another finding was that increasing glycerol concentrations favoured the formation of 1, 3-propanediol, and that *C. pasteurianum* could tolerate high concentrations of glycerol, and growth was not inhibited by concentrations of up to 17% (wt/vol). Glycerol has been reported to be co-utilised with glucose by *C. acetobutylcum* (Vasconcelos *et al.,* 1994).

1.3.6.4 Algae as a fermentation feedstock

Algae are an exciting potential fermentation feedstock for solvent production. There are many reasons why algae make a superior feedstock over terrestrial biomass, most importantly is the fact that farming algae would not compete for land, which is used to grow food crops. Furthermore the photosynthetic efficiency of aquatic biomass is much higher than that of terrestrial plants (Aresta *et al.,* 2005). Horn *et al.,* (2000) successfully used seaweed extract as a fermentation substrate for bioethanol production by *Pichia angophorae.* The best yield obtained was 0.43 g ethanol per g substrate $^{-1}$.

Microalgae can also be used for biodiesel production (Chisti, 2007). They are comprised of significant quantities of oils, proteins, carbohydrates and other nutrients (Mirón *et al.,* 2003). Once the oils are extracted for biodiesel production, the algal residue could potentially be utilised as a fermentation feedstock for solventogenic clostridial fermentation. Microalgae culture systems require less water than traditional oilseed crops and produce more oil on an area basis (Schenk *et al.,* 2008). As microalgae are grown in aqueous suspension they have more efficient access to water, CO_2 and nutrients and are more efficient converters of solar energy due to their simplified cellular structure (Sheehan *et al.,* 1998).

Reducing CO_2 emissions has become high priority for most developed nations. CO_2 mitigation by microalgae has been evaluated in several reviews (Brown & Zeiler, 1993; Chelf *et al.,* 1993; Benemann, 1993). A demonstration of open pond mass production of microalgae biomass was run in Roswell, New Mexico, utilising the flue gas emitted from a coal fired power plant (containing up to 13% CO_2) with up to 90% capture efficiency (Sheehan *et al.,* 1998). Coal emits more CO_2 than either petroleum or natural gas and microalgal mitigation offers a potentially superior solution than some of the other options that have been proposed, such as deep ocean disposal (Cole *et al.,* 1993).

1.3.6.5 Whey as a fermentation feedstock

A potential economically attractive fermentation feedstock is whey, a by-product in the manufacture of cheese. Cheese whey can represent a significant environmental problem due to the high volumes which are manufactured - around half of the whey produced each year remains unused (Mostafa, 2001). An environmental factor of disposal of this waste stream is the high lactose (a disaccharide consisting of a glucose and galactose unit Figure 1.6) content, of around 4-5%, which gives this waste a high BOD. *C. saccharobutylicum* (formally *C. acetobutylicum* P262) was investigated for its ability to utilise whey for ABE production in a continuous bioreactor. However, the continuous fermentations were characterised by cyclic solventogenic and acidogenic behaviour ultimately degenerating into an acidogenic state (Ennis & Maddox, 1989).

In instances where batch fermentations were carried out the main problem with the fermentation was found to be incomplete utilisation of the sugar in the whey, which affected the overall bioreactor productivity (Ennis & Maddox, 1985) - again, highlighting why it is crucial to elucidate the regulation and mechanisms of carbohydrate transport and assimilation.

1.4 *Clostridium acetobutylicum* ATCC 824

As discussed in Section 1.2.1 many saccharolytic, solventogeneic *Clostridium* spp. were utilised in large-scale fermentation of ABE. However, the type strain *C. acetobutylicum* ATCC 824 remains the most extensively biochemically and genetically characterised solventogenic species. The name *C. acetobutylicum* was first introduced and validly published in 1926 (McCoy *et al.,* 1926), however vigilance must be used when considering the older literature characterising the species, as several other organisms referred to as "*acetobutylicum*" have now been reclassified. For example, previously, *C. acetobutylicum* P262 has now been reclassified as *C. saccharobutylicum* (Keis *et al.,* 2001a).

C. acetobutylicum is mesophilic and produces acetic, butyric and lactic acids, butanol, acetone and ethanol and CO_2 as well as H_2 gases during fermentation. Small amounts of succinic acid may also be formed. Growth requirements are relatively straight forward, requiring only a fermentable carbon source, nitrogen source, biotin, and *p*-aminobenzoic acid, (Cato & Stackebrandt, 1989). *C. acetobutylicum* cells in exponential growth phase are straight rods, which are highly motile and peritrichous (with flagella uniformly distributed over the surface of the cell). Cells range in size between 0.5-0.9μm in width and 1.6-6.4μm in length, and during exponential growth, the main fermentation products are acetate and butyrate (Schuster *et al.*, 1998). However *C. acetobutylicum* undergoes morphological changes during different growth phases (Figure 1.10; Jones *et al.*, 1982).

Exponentially growing cells are rod shaped and mainly produce acids. As the carbohydrates are utilised and the acid concentration increases motility decreases and the cells accumulate the glycogen-like storage polymer, granulose and become swollen. This is then followed by the initiation of endospore formation. There is production of an extracellular capsule and a forespore is formed with the spores usually appearing oval and subterminal (Jones & Woods, 1986). Production of butanol is highest during stationary growth phase. *C. acetobutylicum* cells stain Gram positive initially, yet often appearing Gram negative in older cultures (Cato *et al.,* 1986; Jones & Keis, 2005; Wiegel, 2009; Schuster *et al.*, 1998).

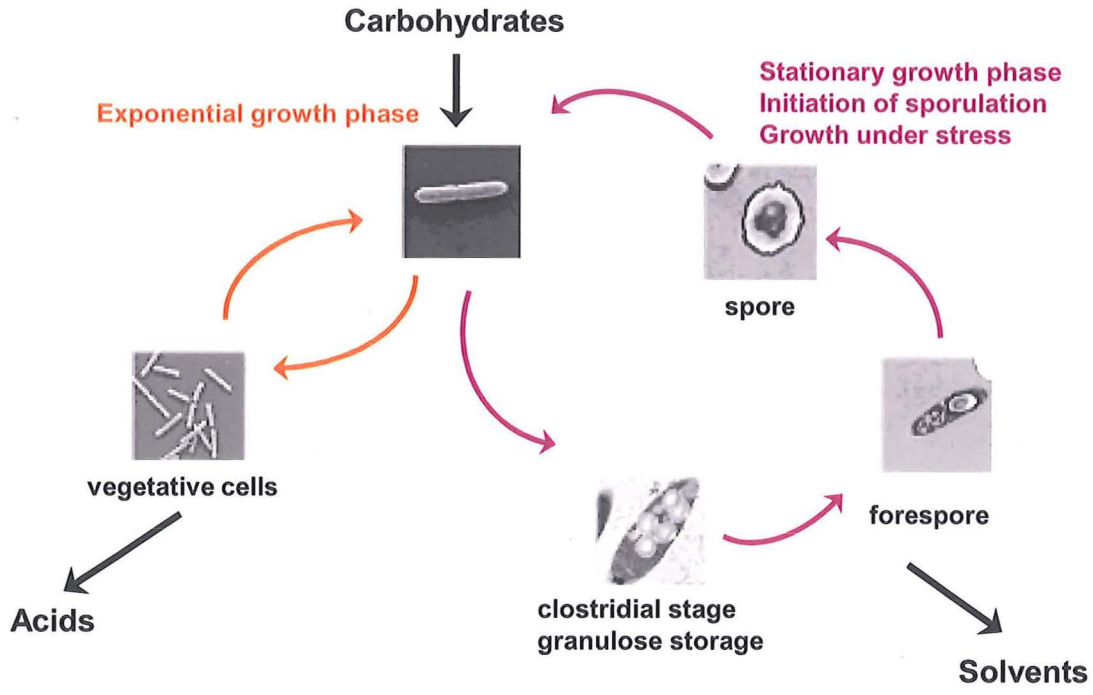


Figure 1.10: Growth phases of *C. acetobutylicum*

Picture adapted from Schuster *et al.*, 1998. C. *acetobutylicum* undergoes several morphological states throughout its biphasic growth cycle, from straight highly motile rods to heat resistant spores.

During exponential growth phase, mainly acids are produced, whereas during the stationary phase of growth mainly solvents are produced.

When cultivated in a batch culture, solventogenic clostridia, such as *C. acetobutylicum,* undergo a biphasic growth pattern. In the first phase (acidogenesis) there is rapid formation of hydrogen, carbon dioxide, acetate and butyrate, resulting in a decrease in the pH of the medium. When the culture reaches the stationary growth phase and the pH reaches a critical point, the second stage begins (solventogenesis). The acetic and butyric acids are reassimilated, this occurs alongside the continued utilisation of carbohydrates, and butanol and acetone are produced, normally resulting in an increase of the pH of the medium (Jones & Woods, 1986; Lee *et al.,* 2008). The biochemical pathways and enzymes are depicted in Figure 1.11.

This relationship between the pH shift and the formation of solvents as well as the relationship between the regulation of endospore development and solventogenesis, is discussed in detail elsewhere (Long *et al.,* 1983; Girbal *et al.,* 1995; Girbal & Soucaille, 1995), and is beyond the scope of this thesis.

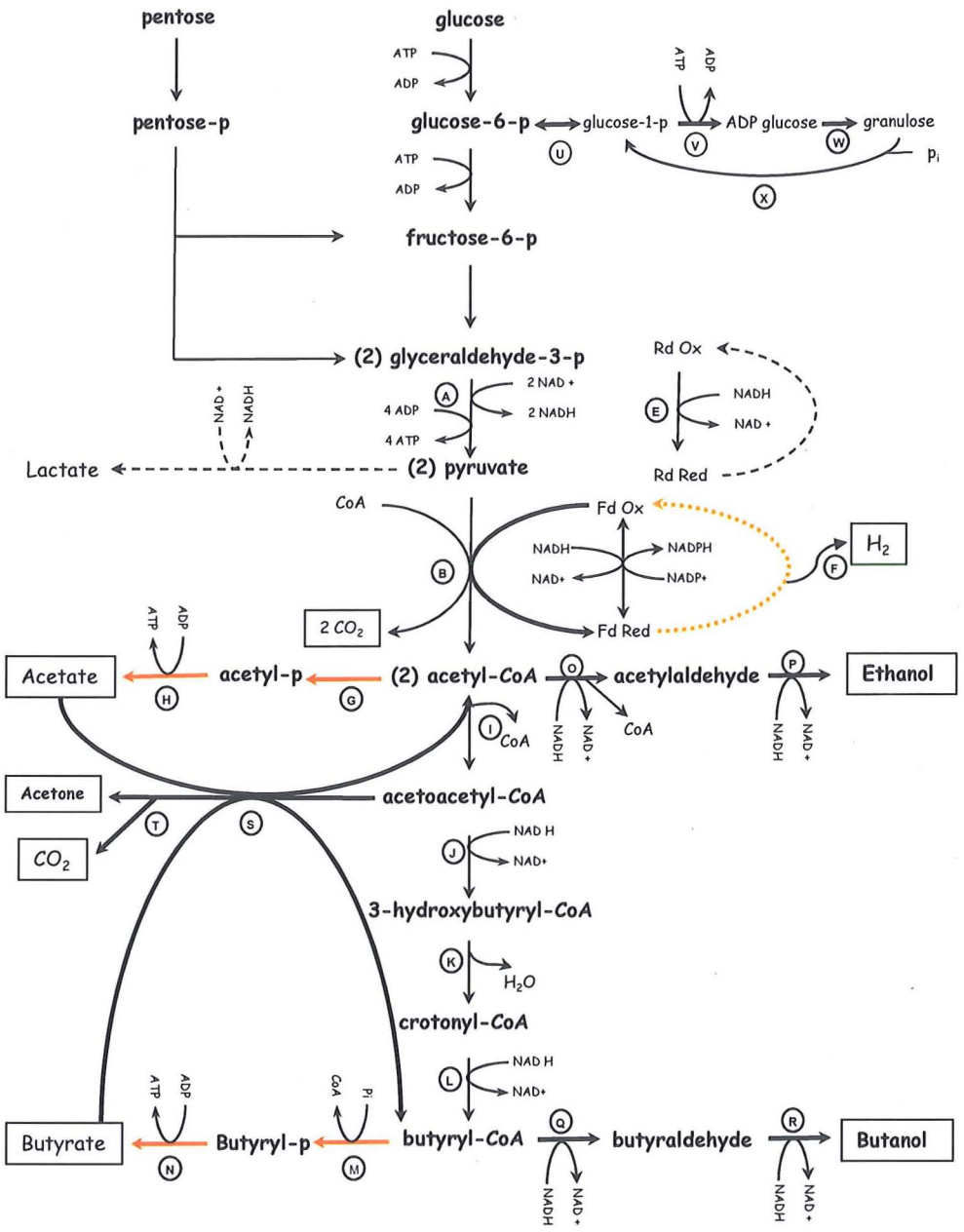


Figure 1.11: Biochemical pathways in *C. acetobutylicum*
Reactions which predominate during the acidogenic phase have **Orange** arrows. Reactions which predominate during the solventogenic phase have **Black** arrows. Figure adapted from Jones & Woods, 1986. Enzymes are indicated by letters as follows: (A) glyceraldehydes 3-phosphate dehydrogenase; (B) pyruvate-ferredoxin oxidoreductase; (C) NADH-ferredoxin oxidoreductase; (D) NADPH-ferredoxin oxidoreductase; (E) NADH-rubredoxin oxidoreductase; (F) hydrogenase; (G) phosphate acetyltransferase (phosphotransacetylase); (H) acetate kinase; (I) thiolase (acetyl-CoA acetyltransferase); (J) 3-hydroxybutyryl-CoA dehydrogenase; (K) crotonase; (L) butyryl-CoA dehydrogenase; (M) phosphate butyltransferase (phosphotransbutyrylase); (N) butyrate kinase; (O) acetaldehyde dehydrogenase; (P) ethanol dehydrogenase; (Q) butyraldehyde dehydrogenase; (R) butanol dehydrogenase; (S) acetoacetyl-CoA:acetate/butyrate:CoA transferase; (T) acetoacetate decarboxylase; (U) phosphoglucomutase; (V) ADP-glucose pyrophosphorylase; (W) granulose (glycogen) synthase; (X) granulose phosphorylase.

1.4.1 System level analysis of *C. acetobutylicum* ATCC 824

The complete genome sequence of *C. acetobutylicum* was published in 2001 (Nölling *et al.,* 2001). The genome consists of a 3,940,880bp chromosome and a 192,000bp megaplasmid, pSOL1. A circular representation of the chromosome and megaplasmid can be found in Figure 1.12. The GC content of the genome is 28-29%.

Genes for acetone and butanol formation reside in the *sol* operon on the pSOL1 megaplasmid (Nölling *et al.,* 2001). Studies have shown that continuous culture of *C. acetobutylicum* can lead to loss of the megaplasmid and thus solvent formation (Cornillot *et al.,* 1997). Many large scale transcriptome (Alsaker *et al.,* 2004; Paredes *et al.,* 2004; Tomas *et al.,* 2004; Tummala *et al.,* 2003), proteome (Schaffer *et al.,* 2005; Sullivan & Bennet, 2006) and metabolome (Senger & Papoutsakis, 2008) studies of *C. acetobutylicum* are completed or underway.

Spo0A is a transcriptional factor found in Gram positive, spore forming, bacteria which controls sporulation. Spo0A acts on gene expression as both an activator and repressor, and is most extensively characterised in *B. subtilis* (Sonenshein, 2000). Spo0A is a response regulator of a phosphorelay cascade, where upon phosphorylation represses or activates expression of most early stationary phase and sporulation genes. Spo0A homologues have been identified in *C. acetobutylicum, C. beijerinckii, C. butyricum, C. thermoaceticum, C. pasteurianum, C. innocuum, C. cellulolyticum. In silico* analysis of these sequences and others from the *Bacillus* genus revealed a highly conserved putative DNA recognition helix (Ravagnani *et al.,* 2000). It was therefore proposed that Spo0A recognises the same target DNA binding site in bacilli and clostridia. Spo0A binds to sequences referred to as "0A boxes". Putative 0A boxes were identified associated with genes involved in solventogenesis in *C. acetobutylicum*, suggesting that Spo0A may regulate more than just sporulation in these bacteria. The *sol* operon (*aad-ctfA/*B) has been shown to be directly regulated by Spo0A (Harris *et al.,* 2002). Indeed, *C. acetobutylicum* Spo0A knockout mutants were deficient in solvent production and sporulation, whereas, a *C. acetobutylicum* mutant which overexpressed Spo0A produced higher butanol concentrations than the parent strain (Harris *et al.,* 2002).

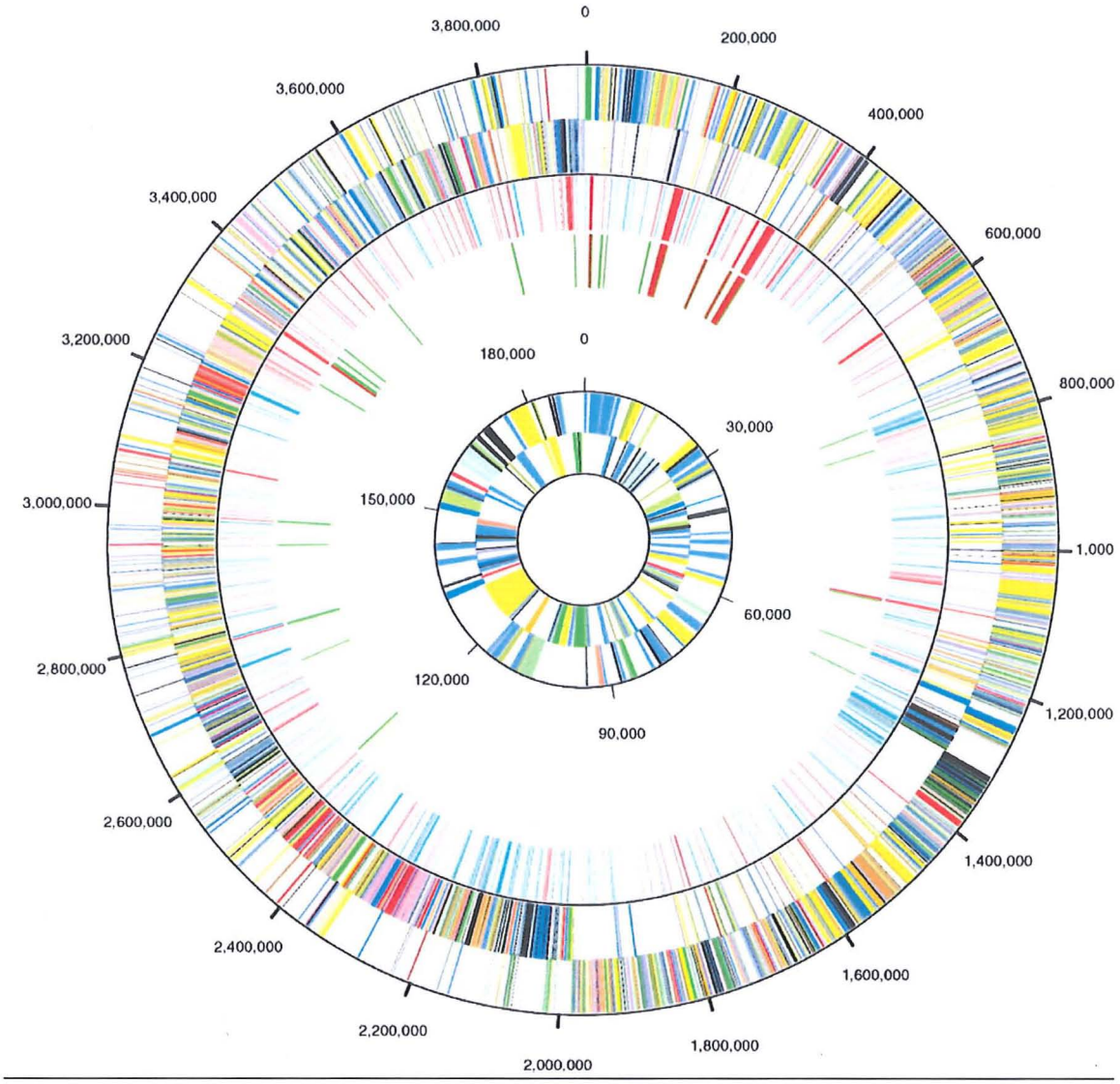


Figure 1.12: Circular representation of the *C. acetobutylicum* genome & megaplasmid.

Taken from Nölling *et al.*, 2001. The outer two rings indicate the positions of genes on the forward and reverse strands of the genome, respectively, color-coded by function. Moving inward, the third ring indicates the G+C content of each putative gene: turquoise (≤27%), gray (27 to 35%), pink-red (>35%); the fourth ring indicates the positions of tRNA (green) and rRNA genes (dark red). The inner rings show the positions of genes on the forward and reverse strands of pSOL1, respectively, color-coded by function (the distance scale for the inner rings differs from the scale of the outer rings, as indicated). The functional color-coding is as follows: energy production and conversion, dark olive; cell division and chromosome partitioning, light blue; amino acid transport and metabolism, yellow; nucleic acid transport and metabolism, orange; carbohydrate transport and metabolism, gold; coenzyme metabolism, tan; lipid metabolism, salmon; translation, ribosome structure, and biogenesis, pink; transcription, olive drab; DNA replication, recombination, and repair, forest green; cell envelope biogenesis, outer membrane, red; cell motility and secretion, plum; posttranslational modification, protein turnover, and chaperones, purple; inorganic ion transport and metabolism, dark sea green; general function prediction only, dark blue; conserved protein, function unknown, medium blue; signal transduction mechanisms, light purple; predicted membrane protein, light green; hypothetical protein, black.

C. acetobutylicum can utilise a vast range of carbohydrates, and this is reflected by the presences of at least 90 genes involved in the degradation of carbohydrate substrates, including fourteen distinct families of glycosyl hydrolases (Nölling *et al.,* 2001). A range of ATP-driven transporters and permeases were identified from the genome sequence, however, the phosphoenolpyruvate (PEP) dependent phopsphotransferase system (PTS) appears to be the dominant mechanism for carbohydrate accumulation.

As discussed earlier, feedstock accounts for approximately 60% of the cost of a large-scale ABE fermentation process, incomplete sugar utilisation is one of the major factors affecting the economic viability of this bioprocess. As carbohydrate uptake is the first step in the conversion of a sugar into solvent, it may be a rate-limiting step in the solventogenesis procedure. Thus, the elucidation of the regulation and mechanisms of carbohydrate transport and degradation is vitally important to understand and optimise the fermentation process in future biofuel plants.

1.5 Bacterial transport systems

Prokaryotic and eukaryotic cells depend on an influx of ions and nutrients such as amino acids and carbohydrates, to ensure the cytoplasm sustains a composition of materials conducive to growth, development and cell division. The bacterial cytoplasm is contained by a highly selective semi-permeable plasma membrane. The plasma membrane consists of a phospholipid bilayer, which is impermeable to large, polar or ionic molecules, such as amino acids and carbohydrates. However, embedded within the phospholipid bilayers are enzymatically active proteins that act as substrate specific permeases, which can transport required materials into the cell (Maloney, 2002).

The Transport Classification DataBase is an ever-increasing, interactive database, which undertakes the immense task of listing all known transporters (Saier, 2000). However, despite the vast number of characterised transporters, they can be classified into several families and, as the review by Nikaido & Saier (1992) indicates, there appears to be a common theme in the design of many transporters, (containing similar transmembrane domains which makes up a membrane spanning channel). They suggest that the energy-coupling mechanism has evolved as a secondary feature of a modified simple transporter. Bacterial transport systems can be broadly divided into two groups; passive transport, which requires no energy input, and active transport, which requires the cell to expend energy.

1.5.1 Passive transport

Passive processes can be further subdivided into passive (or simple) diffusion and facilitated diffusion. Passive diffusion occurs when a concentration gradient is present. Molecules and ions can naturally travel from an area of high concentration to an area of low concentration, until equilibrium is obtained. For example, acids such as butyric and acetic, are transported across the plasma membrane by passive diffusion in *C. pasteurianum* (Kell *et al.*, 1981). Facilitated diffusion also occurs over a concentration gradient, however the molecule is transported through a specific permease. Facilitated diffusion takes place by inducing a conformational change in the permease which transfers the molecule to inside the cell. For example, it has been demonstrated in ruminal clostridial species that the amino acid lysine is predominantly transported by facilitated diffusion (Kessel & Russell, 1992). This effectively accelerates the rate of diffusion, yet the concentration of the molecules within the cell will never exceed the concentration of the surrounding environment.

1.5.2 Active transport

Bacterial cells live a lifestyle of feast or famine, where they are often subjected to environments which contain very dilute concentrations of the molecules needed for survival. Active transporters can accumulate molecules against a concentration gradient. Examples of active transporters are; electrochemical potential-driven transporters (such as porters; uniporters, symporters, antiporters, ATP binding cassette (ABC)-type transporters, TRAP transporters and group translocators, such as the phosphotransferase system (Figure 1.13). Within the Clostridia many examples of these porters have been identified, such as the *C. pasteurianum* K^+ uniporter (Clarke *et al.*, 1982), or the serine / Na^+ symporter in the thermophilic *C. fervidus* (Speelmans *et al.*, 1993). However, the most interesting transport system from the point of view of understanding global carbon regulation and acquisition is the phosphoenolpyruvate (PEP) dependent phosphotransferase system (PTS), which is discussed in the next section. It has been proposed that the complex sophistication of the PTS has evolved to facilitate tighter control over all carbon assimilation process within the cell (Saier, 1977).

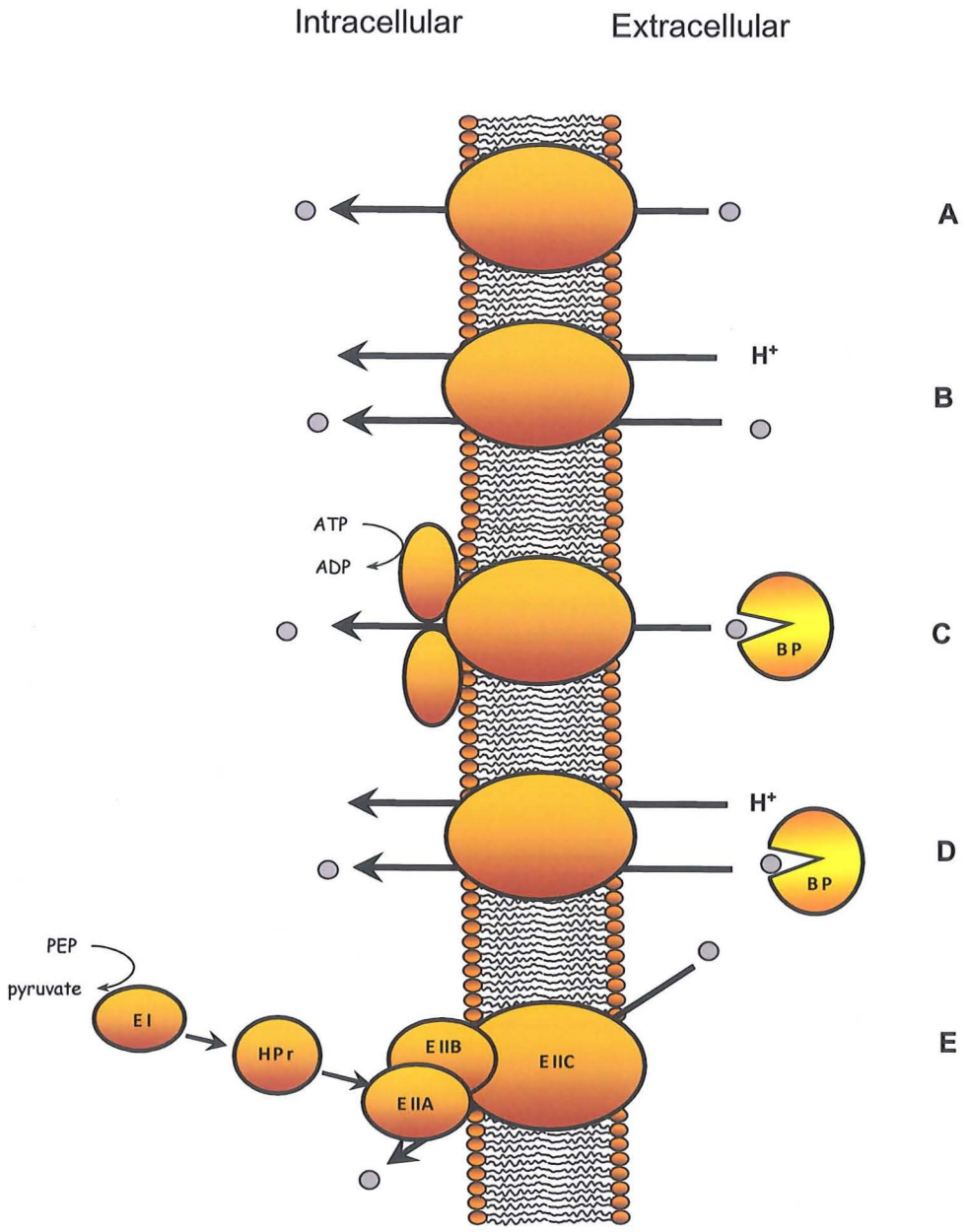


Figure 1.13: Models of the major classes of membrane transporters

Adapted from Nikaido & Saier, 1992 and Rabus *et al.*, 1999. Passive transporters require a concentration gradient to enable translocation across the membrane. However, many transporters exist which can accumulate solutes against a concentration gradient. (A) Passive Facilitator, (B) Symporter, (C) ABC-type transporter, (D) TRAP transporter, (E) Phosphotransferase system. BP – binding protein, EI – enzyme I, HPr – heatstable protein, EII – enzyme II, ● solute to be translocated.

1.5.3 The phosphoenolpyruvate dependent phosphotransferase system

The bacterial phosphoenolpyruvate (PEP) dependent phosphotransferase system (PTS) is a prevalent method of carbohydrate accumulation and is often the dominant method of carbohydrate transport in facultatively and obligately anaerobic bacteria (Meadow *et al.,* 1990; Postma *et al.,* 1993; Mitchell, 1998). The PTS performs two energy dependent steps in a single reaction i.e. with only a single input of energy (Postma *et al.,* 1993). The PTS is a multi-protein phosphotransfer chain. The PTS consists of two general cytoplasmic proteins, the enzyme I (EI) and HPr and a transmembrane, multi-domain substrate specific enzyme II (EII). A phosphate group is transferred from PEP to the incoming sugar (yielding sugar-6-phosphate) via the obligatory phospho intermediates of EI, HPr, EIIA and EIIB (Postma *et al.,* 1993) (Figure 1.14).

The EI protein has been well characterised in many Gram negative and positive bacteria (listed in Postma *et al.,* 1993), the functional organisation appears highly conserved regardless of the organism. At room temperature, in the presence of PEP and Mg^{2+} EI forms a dimer, which is autophosphorylated on each subunit at a histidine residue (His-189 in *E. coli*; Alpert *et al.,* 1985) at the N-terminal of the molecule. The dimer then dissociates (especially at low temperatures) and the phosphorylated monomer transfers its phosphoryl group to a conserved histidine residue on HPr (Deutscher *et al.*, 2006).

HPr has also been extensively characterised and its structure determined in many bacteria. The amino acid sequence of HPr proteins from Gram negative and Gram positive bacteria differ considerably, yet there is a great deal of conservation around the His-15 phosphorylation site (Postma *et al.,* 1993). In Gram positive bacteria the HPr also contains a second site which is highly conserved, Ser-46, which is phosphorylated by an ATP dependent kinase (Herzberg *et al.,* 1992), this will be discussed in detail later.

The EII component can be present as a single multi-domain protein or as individual polypeptides, for example the sucrose I *C. acetobutylicum* (627aa) is encoded on a single ORF (Tangney & Mitchell, 2000), whereas the domains of mannitol EII of *C. acetobutylicum* are split over two ORFs, yet still totalling 627aa (Behrens *et al.*, 2001). The EII complexes generally have similar molecular weights of about 68,000Da comprising around 620-700 total aa residues. The protein usually consists of three domains EIIA (~100aa), EIIB (~100aa) and EIIC (~350aa), apart from the mannose class of EII permeases which contains a fourth EIID domain. For example, the *Streptococcus salivarius* mannose EII operon is composed of three genes *manL, manM,*

and *manN* encoding IIAB, IIC, and IID domains adding up to 904aa (Lortie *et al.*, 2000).

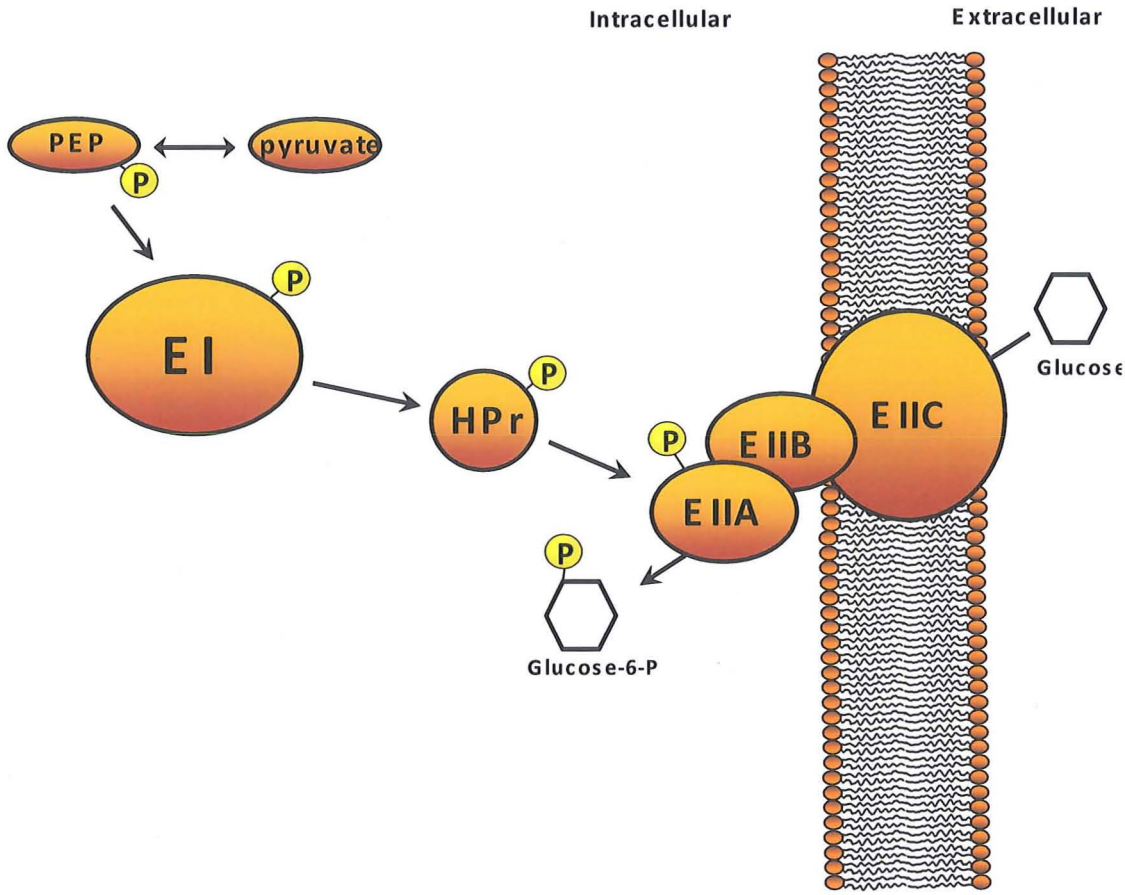


Figure 1.14: The phosphoenolpyruvate dependent phosphotransferase system
Organisation of the phosphoenolpyruvate dependent phosphotransferase system (PEP:PTS). The enzyme I (EI) and HPr are the general PTS proteins. The substrate specific enzyme II (EII) can have highly variable domain structure, present as a single polypeptide or individual proteins. A phosphate is donated from PEP to enzyme I and then transferred to HPr, enzyme II, and then the substrate.

The A domain of EII permeases are usually hydrophillic in nature and contain a highly conserved histidine residue which is phosphorylated by HPr-P. The B domain also tends to be hydrophilic and contain a well-conserved cysteine residue, which accepts a phosphate from EIIA-His-P. The large C domain contains, on average, nine transmembrane helices, at least in the Glc superfamily (Nguyen *et al.,* 2006) and at least one large hydrophilic loop containing a conserved histidine residue (Postma *et al.,* 1993).

The PTS is a complex system which interacts with many other proteins and systems in the cell. Some of the additional regulatory roles (other than the primary function of sugar transport and phosphorylation), that have been proposed for the PTS are summarised in (Table 1.2). PTS components can regulate the activities of non-PTS transport systems, such as carbohydrate symporters or ATP catalysed permeases (Saier *et al.*, 1995). PTS components have also been reported to take part in changes in the bacteria's behaviour, in the form of chemotaxis towards PTS sugars (Garrity *et al.*, 1998). A recent review proposed a link between the PTS and biofilm formation (Sutrina *et al.,* 2007).

Table 1.2: Functional complexity of PTS

Table summarised from the papers of: Barabote & Saier, 2005; Garrity *et al.*, 1998; Sutrina *et al.,* 2007.

PTS function
Regulation of non-PTS sugar transport and metabolism
Regulation of carbon storage
Regulation of fermentation verses respiration
Regulation of cellular motility (chemotaxis)
Coordination of nitrogen and carbon metabolism
Regulation of non-carbon-compound transport
Regulation of gene expression
Regulation of pathogenesis
Regulation of cell physiology
Regulation of cell division
Regulation of biofilm formation

1.5.3.1 Carbon catabolite repression

Carbon catabolite repression (CCR) is a regulatory mechanism that helps to maintain the metabolic/catabolic balance of the cell, by ensuring the sequential utilisation of carbohydrates (Brückner & Titgemeyer, 2002). CCR can cause the expression of catabolic genes to be repressed in the presence of a readily utilised carbon source such as glucose. CCR could play an important role in incomplete sugar utilisation during ABE fermentation from waste biomass substrates, because such biomass can contain a diverse assortment of sugars and sugar alcohols alongside glucose.

CCR has been extensively studied in *E. coli* and the mechanisms of regulation in Gram negative bacteria are well established (Postma *et al.,* 1993). In the low GC Gram positive bacteria, the elements of CCR have been largely studied in *B. subtilis* (Hueck & Hillen, 1995; Galinier *et al.,* 1997; Miwa *et al.,* 2000; Stülke & Hillen, 2000). *Cis*-acting sequences called catabolite responsive elements (CREs) mediate CCR of several genes (Hueck & Hillen, 1995). The catabolite control protein CcpA, which belongs to the LacI family of transcriptional regulators (Weickert & Adhya, 1992), represses transcription of catabolic genes by binding to the CRE sequence (Rodionov *et al.,* 2001). CcpA contains a typical helix-turn-helix motif DNA binding domain that facilitates binding at a specific CRE target site (Tangney & Mitchell, 2005). RegA is a CcpA homologue which was characterised from *C. saccharobutylicum* (formerly *C. acetobutylicum* P262), that can complement a *ccpA B. subtilis* mutant (Davison *et al.,* 1995). A putative *ccpA* gene, previously identified from the *C. acetobutylicum* genome, encoding a CcpA homolog which has 62% identity to the *C. saccharobutylicum* RegA (Tangney *et al.,* 2003) will be discussed in Chapter 3..

In low GC Gram positive bacteria, the PTS proteins, in particular HPr, play a crucial role in CCR (Figure 1.15). HPr is subject to phosphorylation at a histidine residue (His-15) by enzyme I (Postma *et al.,* 1993) and at a serine residue (Ser-46) by a bifunctional enzyme HPr kinase/phosphorylase (HPrK/P; Deutscher & Saier, 1983; Deutscher *et al.,* 1984). HPr kinase activity is stimulated by the glycolytic intermediate, fructose 1,6-bisphosphate (fbp), when a readily metabolisable substrate such as glucose is available, high levels of fbp are produced (Figure 1.15). HPrK/P can phosphorylate HPr(Ser-46), in the presence low concentrations of free inorganic phosphate and high concentrations of ATP and also dephosphorylate HPr(Ser-46) in the presence of high concentrations of inorganic phosphate (Brückner & Titgemeyer, 2002; Tangney & Mitchell, 2005).

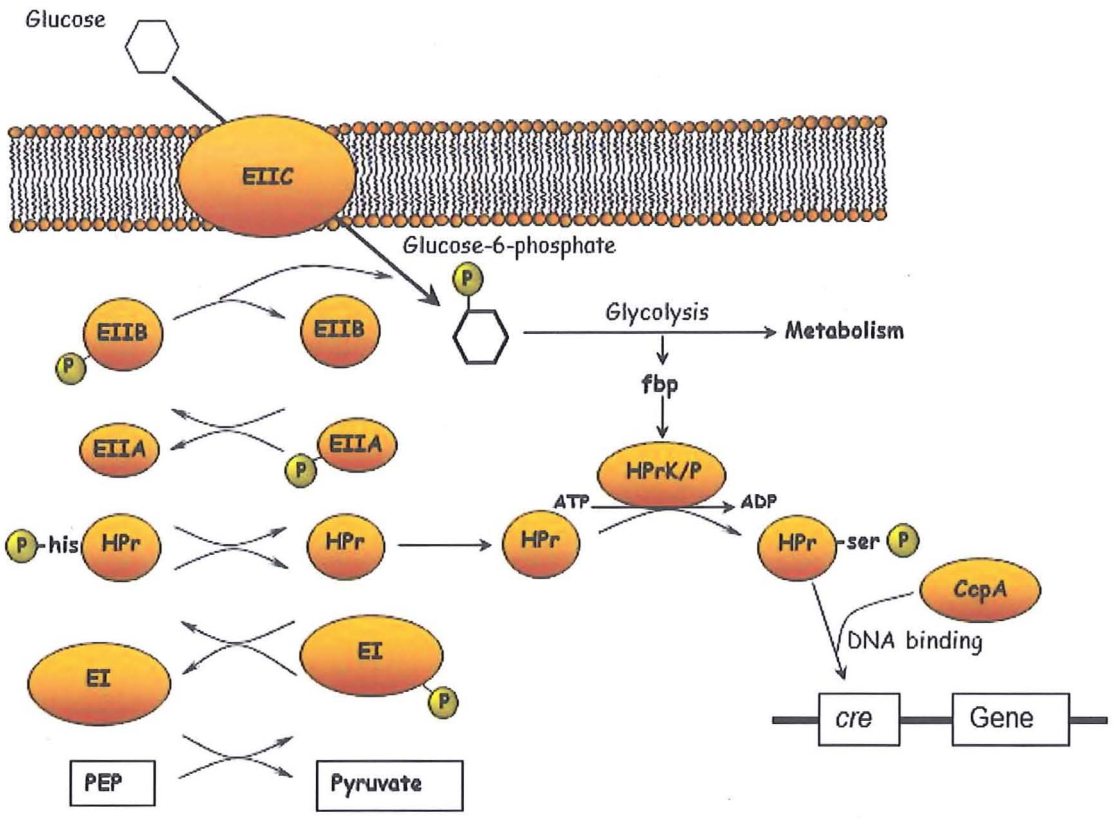


Figure 1.15: Model of HPr/CcpA mediated catabolite repression.

This figure, adapted from Tangney & Mitchell (2005), depicts the model of carbon catabolite repression (CCR) in low G/C-Gram positive bacteria, and its interaction with the PEP dependent PTS. Glycolytic activity generates fructose-1,6-bisphosphate (fbp) and other glycolytic intermediates, the presence of fbp stimulates the enzyme HPrK/P to phosphorylate HPr at residue Ser-46. HPr(Ser-P) can interact with the catabolite control protein (CcpA) to form a complex which binds to regulatory DNA sequences called catabolite responsive elements (*cre's*) to effect gene regulation. This results in repression of a target system, even in the presence an inducer of the system.

Abbreviations are as follows: EI, enzyme I; HPr, histidine protein; EII, enzyme II; PEP, phosphoenolpyruvate; HPrK/P, HPr kinase/phosphorylase, CcpA, catabolite control protein; ATP, adenosine-5'-triphosphate; ADP, adenosine 5'-diphosphate; fbp, fructose-1,6-bisphosphate; *cre*, catabolite responsive element.

1.5.3.2 Antitermination

Gene expression is regulated by a variety of mechanisms, such as transcription activation or repression, however bacteria have also evolved 'fine-tuning' mechanisms based on transcription termination/antitermination which enables tighter control of gene expression (Deutscher *et al.,* 2006). Transcription attenuation was first described in 1976 in the *Escherichia coli* tryptophan biosynthetic operon (Bertrand *et al.,* 1976). There are two categories of transcription termination mechanisms in prokaryotes: factor dependent termination and intrinsic termination (Yanofsky, 2000). Intrinsic, transcriptional terminators are comprised of a rho-independent terminator structure and an associated ribonucleic antiterminator target (RAT) sequence. Rho-independent terminators are 30-50bp of sequence with a dyad symmetry, followed by a run of thymine residues, which signals the termination of transcription, independent of the rho protein, by formation of a hairpin loop structure (Abe *et al.,* 1999; Hoon *et al.,* 2005). RAT sequences are the conserved binding sites for PTS regulated antiterminators. A key feature of the RAT sequences is that they partially overlap the associated terminator (Figure 1.16). Transcriptional antiterminators belong to a family of proteins which regulate gene expression by preventing premature transcription termination upstream of the genes they control (Yang *et al.,* 2002). The BglG protein in *E. coli* is the most thoroughly characterised antiterminator from the Gram negative bacteria, and the SacY protein from *Bacillus subtilis* is the most studied from the low GC Gram positive group of bacteria (Declerck *et al.,* 2002; Tortosa *et al,* 1997; Manival, 1997).

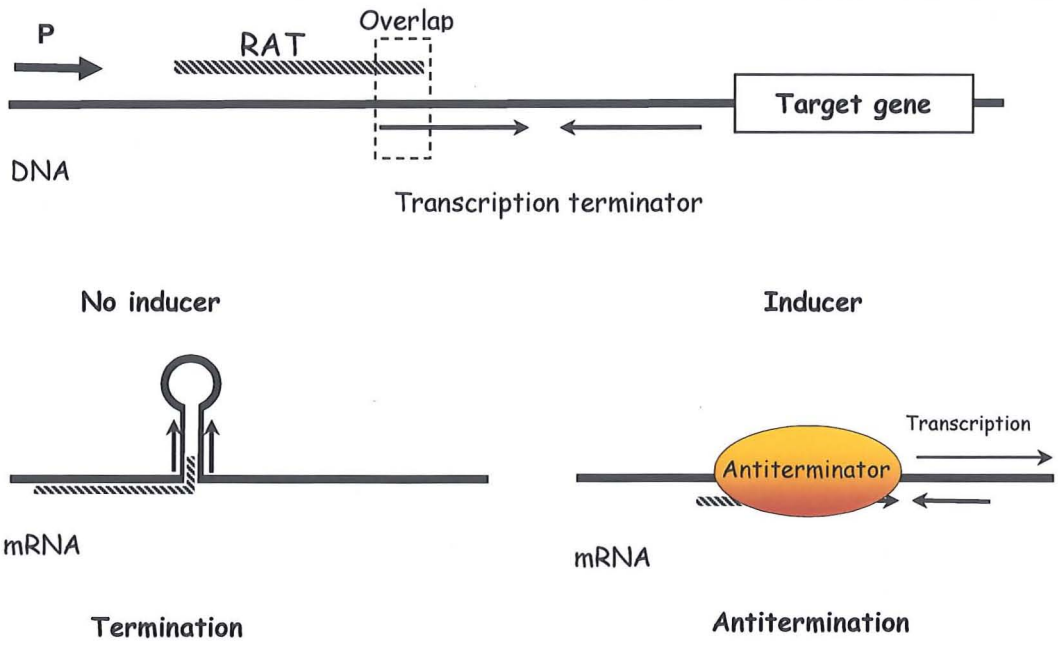


Figure 1.16: Regulation of transcription by BglG-type antiterminator proteins
This figure, adapted from Tangney & Mitchell (2005), depicts the regulation of transcription of a target gene by interaction with a BglG-type antiterminator protein. Characteristically, genes which are subject to regulation by BglG-type antiterminator proteins, are preceded by a ρ-independent transcriptional terminator which is partially overlapped by a ribonucleic antiterminator (RAT) sequence. Transcription is initiated constitutively at a promoter (P) upstream of these regulatory sequences.

When no inducer is present the transcription terminator structure forms and prevents transcription of the subject gene. When an inducer is present, the antiterminator protein is activated and interacts with the RAT sequence to preclude the formation of the transcription terminator, which permits transcription of the gene.

1.6 Project aim

The overall aim of this project is to characterise the regulation and genes involved in transport and utilisation of selected carbohydrates, commonly found in waste biomass feedstocks, in the solventogenic organism *C. acetobutylicum* ATCC 824.

- In particular focus of this thesis will be on analysing the phosphotransferase complement of *C. acetobutylicum.*

The literature will be reviewed to identify what is known of the EI and HPr proteins in *C. acetobutylicum* and searched to identify other HPr sequences from recently sequenced clostridial genomes. The additional clostridial HPr sequence data collected will be analysed and aligned with the HPr protein sequence from *C. acetobutylicum.* The domain structure and the genome context of all known and putative EII permeases will be identified as well as an investigation of the PTS associated regulators found in the genome of this organism. *In silico* investigation will be utilised to identify the elements required for the *B. subtilis* model of global carbon regulation to operate in *C. acetobutylicum.* In addition the fructose 1,6-bisphosphatase genes present in this organism will be characterised

- Characterise the system(s) responsible for the regulated transport and utilisation of β-glucosides in *C. acetobutylicum.*

Initially *in silico* investigation will be used to identify the putative β-glucoside specific EII permeases, metabolic genes and regulatory regions. Followed by further experimental characterisation of the identity of the EII permeases by investigation of the heterologous expression of the putative β-glucoside specific EII permeases in a β-glucoside utilisation deficient host.

- Characterise the mechanisms responsible for the transport and metabolism of hemicellulose monomers in *C. acetobutylicum.*

Firstly *in silico* investigation will be used to identify the putative pentose isomerases found in *C. acetobutylicum.* Followed by investigation of the heterologous expression of the putative pentose isomerases in a pentose isomerase deficient host.

In silico will be used to initiate the characterisation of the mannose specific EII permeases present in the *C. acetobutylicum* genome.

Chapter 2

2 Experimental procedures

2.1 Media and reagents

Reinforced Clostridial Medium (RCM) and Reinforced Clostridial Agar (RCA) (Oxoid, Hampshire, UK) was used for general maintenance of Clostridial cultures and prepared according to the manufacturer's instructions.

For generating spore stocks synthetic Reinforced Clostridial Agar (sRCA) was used, which contained (per litre): Yeast extract 3g; Lab-lemco powder 10g; Peptone 10g; Glucose 5g; Cysteine hydrochloride 0.5g; Sodium chloride 5g; Sodium acetate 3g.

Sugar utilisation experiments were carried out in Clostridial Basal Medium (CBM), which contained (per litre): Carbon source 10g (unless otherwise stated); Casein hydrolysate 4g; $MgSO_4 \cdot 7H_2O$ 0.2g; $MnSO_4 \cdot 4H_2O$ 10mg; $FeSO_4 \cdot 7H_2O$ 10mg; *p*-aminobenzoic acid 1mg; Thiamine HCL 1mg; d-biotin 2µg; KH_2PO_4 0.5g; K_2HPO_4 0.5g. Phosphates and carbon sources were sterilised separately and added after cooling. CBA agar plates were prepared by including 12g/L technical agar number 3 (Oxoid) in the medium.

For general cloning and screening the following media and reagents were used:

Luria-Bertani Medium (LB) contained (per litre); Tryptone 10g; Yeast extract 5g; NaCl 10g; adjusted to pH 7 with 1M NaOH.

For screening of the transformants, LB plates were prepared by preheating the plates at 37°C for 30 minutes before spreading 40µl of a 40mg/ml X-gal solution (Bioline) over the surface of the medium and incubated until required. SOC Outgrowth Medium was included within the Invitrogen TOPO cloning kit and consisted of 2% tryptone; 0.5% yeast extract; 10mM NaCl; 2.5mM KCl; 10mM $MgCl_2$; 10mM $MgSO_4$; 20mM glucose.

MacConkey Agar Plates contained (per litre); Carbon source 10g (unless otherwise stated); Peptone 20g; Bile salts 5g; NaCl 5g; 1ml 0.5 % Neutral red solution; 0.1ml 1% Crystal violet solution.

M9 Minimal Agar Plates prepared according to the Handbook of Microbiological Media (Atlas & Parks, 1993) with the addition of arginine (50mg/ml) and 2g/L of β-glucoside.

Agar plates were prepared by including 12g/L technical agar number 3 (Oxoid) in the medium.

Ampicillin for selective screening of the recombinants: 50µg/ml of ampicillin was added to the M9 Minimal Agar after autoclaving, once the medium cooled to approximately 50°C. The ampicillin was made as a stock solution of 50mg/ml which was filter sterilised and stored at -20°C in 1ml aliquots.

2.2 Bacterial strains and vectors

The focus of this study was on *C. acetobutylicum* ATCC 824, which will be referred to throughout as simply *C. acetobutylicum*. This strain was provided by Dr Wilfrid J. Mitchell at Heriot Watt University who obtained the strain from Dr Philippe Soucaille of the Institut National des Sciences Appliquées-DGBA, Toulouse, France, and was originally acquired from American Type Culture Collection, Rockville, Maryland.

Details of the *Escherichia coli* strains used in this study are listed in Table 2.1. Details of the plasmids used in this study are listed in Table 2.2.

Table 2.1: Details of the *Escherichia coli* strains used in this study

E. coli strain	Genotype
DH5α	F- recA1 endA1 hsdR17(rk -, mk+) supE44 λ- thi-1 gyrA96 relA1
TOP10	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(araleu) 7697 galU galK rpsL (StrR) endA1 nupG
DS941	thr1 leu6 hidG4 thi1 ara14 proA2 argE3 galK2 sup37 xyl5 mtl1 tsx33 str31 recF143 supE44 *lacI*^{q} *lacZ*ΔM15
ZSC113	F- gpt-2 mpt-2glk-7strA (UV mutant of ZSC103, Curtis & Epstein, 1975)

Table 2.2: Plasmids used in this study

Plasmid	Details	Source
pEB100	pCR2.1 + cac1407 putative cellobiose EII #1	This study
pEB200	pCR2.1 + cac1342 putative pentose isomerase	This study
pEB300	pCR2.1 + cac1346 putative pentose isomerase	This study
pCELL	pCR2.1 + cac0383-6 putative cellobiose EII #2	This study
pMAN/C	pCR2.1 + cac1457-60 putative mannose EII	This study
pMAN/P	pCR2.1 + cap0066-8 putative mannose EII	This study
pUC19	No insert	Invitrogen
pJA1	pUC19 + xylose isomerase (*C. thermosaccharolyticum*)	Heriot Watt University

2.3 Bacterial growth conditions

Anaerobic manipulations and incubations were carried out in a Don Whitley Scientific Limited MACS Anaerobic Work Station under a N_2-H_2-CO_2 atmosphere (at an 80:10:10 ratio) and unless otherwise stated were incubated at 37°C. Media used for anaerobic fermentations were anaerobised overnight within the workstation at ~23°C.

C. acetobutylicum was maintained as a spore suspension in water at 4°C. The spores were revived by transfer of a 1ml aliquot to a sterile glass test tube and heat shocked at 80°C for 10min, transferred into 20ml of RCM, and incubated at 37°C overnight in the Anaerobic Work Station.

Aerobic incubations were carried out at 37°C in either New Brunswick Scientific INNOVA44 shaking incubator or in the case of static aerobic incubation an Infors Unitron incubator.

For long term storage *E. coli* strains were cultured in LB broth at 37°C and 200 rpm, to mid-log phase. The culture (5ml) was then mixed with 5ml of sterile 80 % glycerol, then aliquotted and stored at -80°C. Aliquots were defrosted on ice when needed.

Agar plates were allowed to set and then dried in a Gelaire BSB 4A Laminar airflow cabinet. Plates were stored inverted and refrigerated.

2.3.1 Preparation of *Clostridium* spore stocks

A starter culture of the desired *Clostridium* was prepared as described above, 1% (v/v) of the starter culture was used as an inoculum and transferred to 500ml of synthetic RCM (sRCM, a modified RCM which does not contain agar or soluble starch) and incubated for four weeks. The culture was aseptically transferred into sterile 50ml centrifuge tubes and centrifuged (12,000xg, 15min). The supernatant was then discarded and the pellets were resuspended in sterile dH_2O and washed three times. The combined spore pellet was finally resuspended in 50-100ml of sterile dH_2O and stored at 4°C.

2.4 Fermentation analysis

2.4.1 Assay of sugar and butanol concentration in culture supernatants

Culture samples (1ml) were removed and centrifuged at stated intervals. Cells were harvested by centrifugation (20,000xg, 10min); the supernatant was collected and stored at -20°C. The supernatants were then defrosted and centrifuged (20,000xg, 10min) and then filtered through a 0.20μM cellulose acetate syringe filter (Anachem Supatop, ALG422A). One ml of the filtered supernatant was then aliquotted into a GC vial. The sugar concentration was determined by High Performance Liquid Chromatography (HPLC) and the solvent concentration was determined by gas chromatography (G/C).

2.4.2 Optical density and pH determination

Culture samples were removed at stated intervals and the OD measurements were read at 650nm with a Jenway 6300 spectrophotometer. The pH was determined using a Jenway 3310 pH meter.

2.5 Molecular manipulations

2.5.1 Chromosomal DNA extraction

The chromosomal DNA extractions were preformed with the Promega Wizard® Genomic DNA Purification Kit. The specific details and amendments are listed below. *C. acetobutylicum* spores were revived by heat shock, as described above and 1ml inoculated into 20ml of RCM and incubated at 37°C overnight in the anaerobic workstation. 1.5ml of overnight culture was pipetted into a microcentrifuge tube and centrifuged for 2min at 13,000xg. The supernatant was removed and a further 1.5ml of overnight culture was added and the centrifugation repeated. The supernatant was removed and the cells thoroughly resuspended in 480µl of 50mM EDTA. 120µl of lysozyme was gently added to the resuspended cell pellet and incubated at 37°C for 30-60 min. The microcentrifuge tube was centrifuged at 13,000xg and the supernatant removed. The cells were resuspended in 600µl of Nuclei Lysis Solution and incubated at 80°C for 5min and cooled to room temperature. 3µl of RNase Solution was added to the cell lysate and the contents mixed by inverting 3 times. This was then incubated at 37°C for 30min and cooled to room temperature. 200µl of Protein Precipitation Solution was added to the RNase treated cell lysate, vortexed for 20 seconds and incubated on ice for 5min. This was then centrifuged at 13,000xg for 3min.

The supernatant was transferred to a clean microcentrifuge tube containing 600µl of room temperature isopropanol, inverted 3 times and centrifuged at 13,000xg for 2min. The supernatant was poured off and drained on absorbent paper and 600µl of room temperature 70% ethanol was added and the tube inverted 3 times. The microcentrifuge tube was centrifuged at 13,000xg for 2min and the ethanol aspirated and left to air dry for 15min. Finally 100µl of DNA Rehydration Solution was added and either incubated at 65°C for 1 hour or overnight at room temperature. The extracted DNA was stored at 4°C.

2.5.2 PCR conditions

The water used in the PCR reactions was prepared by transferring distilled dH_2O in 5ml aliquots and autoclaving twice. The PCR reaction mix contained the following: 25µl of Biomix (Bioline), 21µl of dH_2O, 100pmol of the forward primer, 100pmol of the reverse primer and 200ng of DNA template. However, in most cases to obtain a PCR product, an optimised reaction mixture was required, this contained various quantities (details of each given in the results section) of the following: 10x NH_4 Buffer, $MgCl_2$ 50mM, 40mM dNTP, DNA, Fwd Primer, Rev Primer, BIOTAQ (Bioline), dH_2O.

The general cycle (Figure 2.1), using a Thermo electronic Px2 Thermal cycler, was as follows (unless otherwise stated): 1 hold at 94°C for 5 minutes, before addition of the Biomix or the BIOTAQ to the PCR reaction mix, followed by: 30 cycles of denaturation (94°C for 1 minute), annealing temperatures varied for each PCR (XX°C for 1 minute), extension (72°C for 1 minute) followed by a single hold of 10 minutes at 72°C. The thermalcycler then held the PCR products at 4°C until analysis. The PCR products were analysed by agarose gel electrophoresis.

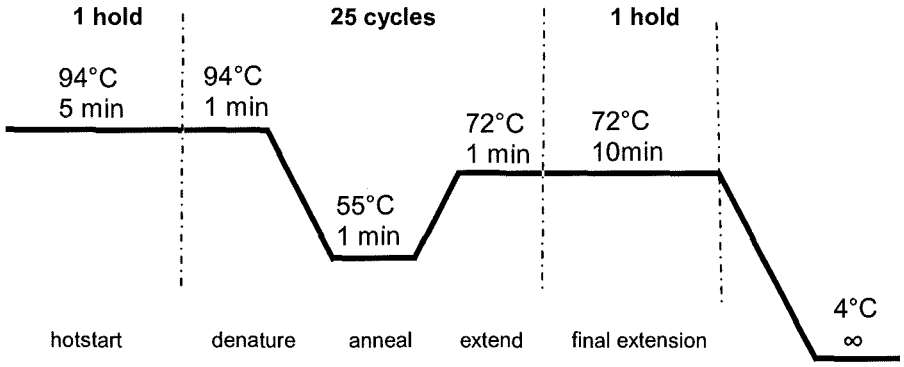


Figure 2.1: Example PCR reaction cycle.

A model PCR reaction cycle, specific varations in this cycle are detailed in the relevent results sections.

2.5.3 Primer details and sequences

The sequences of the forward and reverse PCR primers used in this study are presented in Table 2.3. All primers were obtained from MWG Biotech. Reverse primers are reverse complements of the DNA sequence (Table 2.3). Primers were designed, where possible, to be about 20bp long with an annealing temperature between 50°C and 60°C. Primers were rehydrated with DNase & RNase free 18.2MΩ PCR Grade Water (Bioline) to a concentration of 100pmol/µl, aliquotted and stored at -20°C.

Table 2.3: Primer details

Amplified region	Expected size (bp)	Sequence (5'- 3')
cac1342	1843	Fwd: GCTCCTTTTACCTGGGGAACTGACCCT Rev: ATACTCCGGTGTGATTTTTCCTTCATCC
cac1346	1616	Fwd: TCTATTGGAAAAGTTAGAGG Rev: GTTATATATGCTCTAAAATCC
cac1407	1941	Fwd: AAGGAGAGATTATCATGAAATATGAAAAGTTGG Rev: TTTGAGCACAAGTCAACCTTATGC
cac1457-1460	2740	Fwd: TTAGGCTAGGAGGGGAAAG Rev: GCCTACTTTGCTGCTTTGCCTC
cap0066-0068	2917	Fwd: TCAAAAGTTCCAGGGGGTCG Rev: AGCCAAAGCCCAGGCTAAG
cac0383-0386	3612	Fwd: GTGTTTGGCATGATGTTTGC Rev: CAGTACATATTCTTTACAGGG
cac1406 (internal probe primers)	635	Fwd: GGAGAAAAAGTAGATGAGAGGC Rev: GTACCTTAAGCATACATTGGTATGC
cac1407 (internal probe primers)	627	Fwd: TAGGTGCATCACTTGTTTACCC Rev: GTGTAACACCATATATAGCTGGC
cac1408 (internal probe primers)	651	Fwd: GTGAAAACAGTAGGTTCCTGG Rev: AAGCGGATCTATTTGCCATCC
TOPO control	750	Fwd: GTAAAACGACGGCCAG Rev: CAGGAAACAGCTATGAC

2.5.4 DNA gel electrophoresis

DNA samples were analysed by agarose gel electrophoresis. A 1% gel contained 1g of molecular grade agarose and 3µl Ethidium Bromide (10mg/ml stock solution) per 100ml of [x1] TAE buffer (Tris-base, glacial acetic acid, 1mM EDTA). Gels were loaded with 5µl of PCR product and 1µl of loading buffer dye. DNA HyperLadders and Loading Dye Buffer were used as supplied by Bioline.

2.5.5 Cloning reaction and transformants

Fresh PCR products were cloned using the TOPO TA Cloning kit (Invitrogen) following the "One Shot Chemical Transformation Protocol", which works as follows;

Taq polymerase has a non-template dependent terminal transferase activity that adds a single deoxyadenosine (A) to the 3' ends of the PCR product. The vector supplied with the TOPO TA cloning kit possesses single, overhanging 3' deoxythymidine (T) residues.

When these are mixed in the presence of salt (which prevents topoisomerase I from rebinding and potentially nicking the DNA after ligating the PCR product and dissociating from the DNA) the result is the PCR product of interest ligated into the TOPO vector. Some of the features of TOPO vector include an ampicillin resistance gene, which allows for selection and maintains the plasmid in *E. coli*, a Lac promoter (which allows bacterial expression of the LacZ fragment for blue-white screening) and a T7 promoter (which allows in vivo or in vitro transcription of antisense RNA).

A cloning reaction was set up which included 4µl of fresh PCR product, 1µl of salt solution (provided in the TOPO kit) and 1µl of TOPO cloning vector. Details of all the plasmids used in this study are in Table 2.2. These were mixed gently and incubated for 30 minutes at room temperature after which the reaction was kept on ice.

The TOPO cloning reaction (2µl) was added to a vial of TOP 10 chemically competent *E. coli* (provided within the kit). This was mixed gently and incubated on ice for 10 minutes. The reaction was then heat shocked at 42°C for 30 seconds then transferred back to ice. 250µl of room temperature S.O.C medium was added and the vial was shaken horizontally (~200rpm) at 37°C for one hour of outgrowth. Chemically competent cells are cells that have been chemically pre-treated and are thus susceptible to DNA transformation. When cool cells are heat shocked in the presence of external DNA, the DNA is taken into the cell. The genes coded by the DNA which has only just been introduced to the heat-shocked cells have not been expressed. The inclusion of an outgrowth period is essential for the new genes to be transcribed and translated.

The transformed cells were then plated out in three different aliquots (20µl, 50µl and 130µl) onto pre-warmed LB ampicillin, X-gal plates. 30µl S.O.C medium was used to facilitate the spreading of the 20µl aliquot. These plates were incubated at 37°C overnight.

2.5.6 Colony PCR

A loopful of this biomass was dissolved into a PCR tube containing 50µl of dH_2O. A Special PCR cycle was used to lyse the cells and release the DNA. This cycle was repeated three times and consisted of: 95°C for 5min followed by 4°C for 1min. The tubes were then centrifuged for 2min at 13,000xg. The supernatant (around 23µl) was removed and used for a PCR with the relevant primers and parameters as stated in results section.

2.5.7 Plasmid DNA extraction

The plasmid DNA extractions were preformed with the Promega PureYield Plasmid Midiprep System. The specific details and amendments are listed below.

A single, well isolated colony from a fresh LB ampicillin agar plate was used to inoculate 20ml of LB ampicillin medium. This was incubated overnight at 37°C in a shaking incubator. Overnight culture (1.5ml) was pipetted into a microcentrifuge tube and centrifuged for 2 minutes at 10,000xg. The supernatant was removed and a further 1.5ml of overnight culture was added and the centrifugation repeated. The supernatant was removed and the cells thoroughly resuspended in 250µl of Cell Resuspension Solution.

Cell Lysis Solution (250µl) was added, the tube inverted 4 times and incubated for 3 minutes at room temperature. Alkaline Protease Solution (10µl) was added, the tube inverted 4 times and incubated for 5 minutes at room temperature. 350µl of Neutralisation Solution was added and the tube inverted 4 times. The tube was then centrifuged at 13,000xg for 10 minutes at room temperature.

The cleared lysate was decanted into a spin column, prepared by inserting a Spin Column into a 2 ml Collection Tube, and centrifuged at 13,000 x g for 1min at room temperature. The Spin column was removed and the flowthough discarded. 750µl of Column Wash Solution, previously diluted with 95% ethanol, was added to the Spin Column and centrifuged at 13,000xg for 1min at room temperature. The Spin Column was removed and the flowthough discarded. The wash procedure was repeated with 250 µl of Column Wash Solution and centrifuged at 13,000xg for 2min at room temperature. The Spin Column was transferred to a new sterile microcentrifuge tube. The plasmid DNA was eluted by the addition of 100µl of Nuclease Free Water and collected in the microcentrifuge tube by centrifugation (13,000xg, 1min). The extracted plasmid DNA was then stored at -20°C.

2.5.8 Restriction analysis

Unless otherwise stated; 1µl of restriction enzyme and 2µl of the appropriate restriction buffer (Promega) were added to 200ng of DNA (~7µl) and made up to a final volume of 20µl with sterile dH_2O in a PCR tube. The tube was then centrifuged and then incubated at 37°C in the Thermocycler for 3 hours. 4µl of loading buffer was added to the reaction and 20µl was analysed by gel electrophoresis.

2.5.9 Electrocompetent *E. coli*

Sixteen hour cultures were used to inoculate fresh LB broth, these were then grown at 37°C and 200rpm in a shaking incubator to mid-log phase. The culture was then immediately placed on ice for 30min and then transferred into 50ml pre-chilled screwcap vials and centrifuged (4°C, 4000rpm, 15min) in a pre-chilled Heraeus multifuge 3L-R centrifuge. The pellet was washed 4 times with dH_2O at 4°C, by centrifugation (4000rpm, 15min), and the final pellet was resuspended in 5ml of 10% glycerol added to obtain a final volume of 45ml. This was then centrifuged as previously described, the supernatant discarded and pellet resuspended in 400µl 10% glycerol. The 400µl suspension was then transferred into 40µl aliquots in 1.5ml microcentrifuge tubes and stored at -80°C.

2.5.10 Transformation of *E. coli*

A 40µl aliquot of competent cells were defrosted on ice to be used for electroporation. 2µl of plasmid DNA was added and mixed with the competent cells, the mixture was incubated on ice for 1min and then transferred into a chilled 2mm electroporation cuvette (Cell Projects, EP120) so that the mixture completely covered the bottom of the cuvette. The cuvette was then placed in the shockpod of a Biorad gene pulser Xcell™ electroporator and was pulsed once, for 5mSec using 2.5kV. The cuvette was immediately removed from the shockpad and 1ml SOC was added and mixed gently using a pipette. The cell suspension was then transferred to a 1.5ml microcentrifuge tube and incubated for 1 hour at 37°C and 200rpm in a shaking incubator. Then 50µl and 100µl were plated onto LB plates with 50µg/ml ampicillin and incubated overnight at 37°C.

2.6 DNA and RNA computational analysis methods

2.6.1 Open Reading Frame tool

The Open Reading Frame (ORF) finder service at the National Centre for Biotechnology Information was used to identify the putative ORFs of each gene studied. This bioinformatics tool can be found at:
http://www.ncbi.nlm.nih.gov/gorf/gorf.html

2.6.2 Basic Local Alignment Search Tool

Basic Local Alignment Search Tool (BLAST) was used to search sequence databases for homology between sequences (Altschul *et al*, 1997). This bioinformatics tool can be found at: http://www.ncbi.nlm.nih.gov/blast/Blast.cgi

2.6.3 Genome Information Broker

Genome Information Broker (GIB) was used to search for putative gene systems (Fumoto *et al,* 2002). This bioinformatics tool can be found at:
http://gib.genes.nig.ac.jp/

2.6.4 Entrez Cross Database Search Engine

GenBank accession numbers were entered into the Entrez database to retrieve sequence data (Benson *et al*, 2008). Database can be found at:
http://www.ncbi.nlm.nih.gov/sites/gquery

2.6.5 ClustalW2

ClustalW2 (Larkin *et al,* 2007) on the EMBL-EBI server, was used to align multiple DNA or protein sequences. Multiple alignments conducted in ClustalW2 were visualised using the programmes GeneDoc (Nicholas *et al,* 1997) or TreeView (Page, 1996). This bioinformatics tool can be found at:
http://www.ebi.ac.uk/Tools/clustalw2/index.html

2.6.6 Construction of a multiple alignment

To construct a multiple alignment, the DNA or protein sequences (either putative sequences returned from a BLAST search or published sequences identified from the literature) were complied and input into ClustalW2. Standard settings were used except Output Order was set to Input.

The Alignment File was imported into GeneDoc, unless otherwise stated Similarity Groups were disabled, Shading Mode was set to Conserved and Conserved Percentages were set to Primary 100%. Secondary 80% and Tertiary 60%. The subsequent output was then directly transferred to Word.

2.6.7 Construction of a phylogenetic tree

To construct a phylogenetic tree, the DNA or proteins sequences were collected and input into ClustalW2. The Alignment File was imported into TreeView and the resulting output was imported into PowerPoint and the file was then transferred into Word.

2.6.8 PlasMapper version 2.0

Was used to visually represent the important features within the constructed plasmids (Dong *et al*, 2004). This bioinformatics tool can be found at: http://wishart.biology.ualberta.ca/PlasMapper/

2.6.9 RestrictionMapper

Was used to locate restriction sites within the studied sequences and to virtually digest the plasmid DNA prior to actual digest. This bioinformatics tool can be found at: http://www.restrictionmapper.org/

2.6.10 The RNA Mfold server version 3.2

Was used to predict secondary structures of RNA (Zuker, 2003; Mathews *et al*, 1999). This bioinformatics tool can be found at: http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1-2.3.cgi

2.6.11 The DINAMelt server

The DINAMelt server (Markham & Zuker, 2005) was used to predict the melting profile and secondary structures of nucleic acids. This bioinformatics tool can be found at: http://dinamelt.bioinfo.rpi.edu/quikfold.php

2.6.12 KEGG: The Kyoto Encyclopaedia of Genes and Genomes

The Kyoto Encyclopaedia of Genes and Genomes (Kanehisa & Goto, 2000) was used to examine the areas surrounding putative genes. This bioinformatics tool can be found at: http://www.genome.ad.jp/kegg/

2.6.13 CDD: A Conserved Domain Database for protein classification

The Conserved Domain Database (CDD) (Marchler-Bauer *et al*, 2005; Marchler-Bauer *et al*, 2007; Marchler-Bauer & Bryant, 2004) was used to identify conserved domain footprints in the studied sequences. This bioinformatics tool can be found at: http://www.ncbi.nlm.nih.gov/sites/entrez?db=cdd

2.6.14 Transmembrane helices:Hidden Markov Model (TMHMM)

TMHMM uses a Hidden Markov Model to predict transmembrane helices within protein sequences. This model can discriminate between membrane and soluble protein at greater than 99% accuracy (Krogh *et al*, 2001; Sonnhammer *et al*, 1998). This bioinformatics tool can be found at: http://www.cbs.dtu.dk/services/TMHMM/

2.6.15 Kyte Doolittle hydrophobicity plot

Kyte Doolittle ExPASy Proteomics Server is used to display the hydropathic character of a protein (Kyte & Doolittle, 1982). This bioinformatics tool can be found at: http://expasy.org/tools/protscale.html

Chapter 3

3 *In silico* analysis of the phosphotransferase system in *Clostridium acetobutylicum* ATCC 824

3.1 Introduction

The worldwide consumption of finite fossil resources (such as petroleum, coal and natural gas) is rapidly increasing. As introduced in Chapter 1, there is a global demand for a renewable fuel source which is secure and not subject to erratic price fluctuations. The main renewable technologies currently being developed are solar, hydro, wind, biomass, nuclear fission and fusion (Kamm & Kamm, 2004). Current agricultural and food processing techniques produce vast quantities of waste biomass, rich in sugars which bacteria can utilise. These by-products are often sent to landfill, where they can cause management problems due to their high BOD (Mostafa, 2001). *C. acetobutylicum* is metabolically flexible and can convert many of the sugars found in plant waste biomass into ethanol, acetone, butanol and hydrogen (Jones & Woods, 1986).

Biobutanol produced by clostridial fermentation of waste biomass is being hailed as the next generation sustainable biofuel. As well as having less impact on the global carbon cycle, the fermentation helps mitigate the vast amounts of waste biomass entering our landfill systems. However, several practical considerations require further research to make the large scale fermentation economically viable that is; butanol toxicity, bacteriophage infection, strain degeneration and incomplete utilisation of biomass feedstock.

Incomplete utilisation of feedstock can be caused by the complex regulatory cascades bacteria use to regulate carbohydrate transport. Characterisation of carbohydrate transport and regulation mechanisms could provide vital information to alleviate this particular practical obstacle. The phosphotransferase system (PTS) is often the principal mechanism for carbohydrate accumulation in anaerobic bacteria (Kotrba *et al.*, 2001).

A number of the different individual PTS genes have been studied within this laboratory, however since the release of the complete *C. acetobutylicum* genome sequence, no genome-wide analysis of the phosphotransferase systems present has been conducted. The objective of this work was to analyse the PTS components and the associated regulatory elements, using the genome mining tools described in Chapter 2 and present this in context of what is already known about this organism, in the style of

a review (to be submitted for publication). An example of the general bioinformatics approach adopted in this research is detailed below.

The genome of *C. acetobutylicum* was retrieved for analysis using the GenBank accession number AE001437 from the National Centre for Biotechnology Information (NCBI) server (Nölling *et al.,* 2001; http://www.ncbi.nlm.nih.gov/nuccore/AE001437). The DNA sequence was then translated into all the six possible reading frames, using the Open Reading Frame (ORF) finder tool (Section 2.6.1), with the settings optimised for translating bacterial code. The possible reading frames, three on the plus strand and three on the minus strand, are designated by the triplet of codons which potentially specify an amino acid. The ORF algorithm works by identifying potential start codons in the sequence, in bacteria usually an ATG codon specifying the amino acid methionine, followed by termination codons, such as TAA. The ORFs then need to be analysed to pick out potential valid ORFs. The presence of several large ORFs in the same direction and reading frame usually indicates the presence of an operon. Each bacterial species has a characteristic pattern of codon preference, as a check of the reliability of the predicted ORFs the sequence can be entered into a more specialised prokaryote ORF finder tool such as the EasyGene 1.2 Server (from the Technical University of Denmark server), which has specific settings optimised for gene prediction in low GC Gram positive bacteria.

The DNA sequence of the predicted genes of interest can then be translated into the deduced amino acid sequence using ExPASy - DNA to protein translation tool. As an additional check, a BLAST homology search is carried out with the sequence of the predicated protein against the protein data base to ensure that the predicted protein is homologous to other proteins of similar function. Additionally, where relevant, the conserved domains, molecular weight, hydrophobicity profile of the protein of interest were identified. For comparison with the predicted protein, several examples of model proteins, which have had their functions experimentally confirmed and their sequences analysed and published were identified from the literature. Their sequences can then be retrieved from a protein database and used for comparison with the predicted protein sequence of interest (deduced from the DNA sequence of the putative ORF).

To identify the degree of conservation between the characterised proteins and the predicted protein, a multiple sequence alignment can be conducted using a program such as ClustalW2, which determines the best match for lining up the entered sequences. The ClustalW2 file can then be entered into a visualisation program such as GeneDoc where the settings can be altered to optimally visualise the studied sequences.

To analyse the evolutionary relationship between the protein sequences, ClustalW2 files can be viewed in TreeView. The regions upstream of ORFs of interest were also investigated for putative regulatory sequences, prokaryotes have highly conserved promoter regions, -10 and -35 RNA polymerase interactions sites and ribosome binding sites.

The phosphotransferase complement and associated regulatory proteins of *C. acetobutylicum* will be analysed. Information on the PTS complement and global carbon regulation mechanisms of this organism could provide information for industrial strain development.

3.2 The PTS complement of *C. acetobutylicum*

Without membrane transport cellular life could not exist. As introduced in the first Chapter (section 1.5.3) the phosphotransferase system is a predominant, energy efficient mechanism of transporting sugars and sugar alcohols across bacterial membranes (Lengeler *et al.*, 1990). To recapitulate, the PTS is a multi-protein sugar transport and phosphorylation system, comprised of two general cytoplasmic proteins, the enzyme I (EI) and the histidine containing, heat-stable protein (HPr), and the (normally membrane associated) carbohydrate specific enzyme II (EII). A diagram of the organisation of the PTS is presented in Chapter 1, Figure 1.14. All phosphotransferases that have been characterised so far catalyse the same overall reaction (Kotrba *et al.*, 2001).

$$\text{P-enolpyruvate}_{(in)} + \text{carbohydrate}_{(out)} \xrightarrow{\text{PTS}} \text{pyruvate}_{(in)} + \text{carbohydrate-P}_{(in)}$$

PTS activity can be determined by directly monitoring PEP-dependent phosphorylation of substrate in permeabilized cells or cell-free extracts (Mitchell *et al.,* 1995). Many model PTS proteins have been characterised at the biochemical and genetic levels in the clostridia (Mitchell & Booth, 1984; Mitchell, 1996; Brown & Thompson, 1998; Mitchell *et al.*, 2007).

A wealth of physiological data exists for *C. acetobutylicum* and several PTS proteins have now been cloned and expressed in *E. coli*. The purpose of this chapter is to summarise the existing research relating to the PTS in *C. acetobutylicum* and to conduct *in silico* analysis using the genome sequence data.

A genomic approach can provide an overview of the entire PTS complement of an organism. For example, the genomes of *C. tetani*, *C. perfringens* and *C. botulinum* appear to possess one, eleven and fifteen phosphotransferases respectively (Mitchell & Tangney, 2005; Mitchell *et al.,* 2007). *C. acetobutylicum* possesses a complement of thirteen phosphotransferases.

3.3 Enzyme I and HPr

As illustrated earlier (Chapter 1, Figure 1.13) the PTS is a multi protein phosphotransfer chain, which transfers a phosphate from PEP, through enzyme I (EI), HPr, EII to the transported substrate. EI and HPr are the general energy coupling proteins, shared between all of the PTS's regardless of the substrate (Saier & Reizer, 1992). HPr (from low GC Gram positive bacteria) can be phosphorylated at two sites. Enzyme I can phosphorylate HPr at a histidine residue (His-15) and HPr

kinase/phosphorylase (HPrK/P) can phosphorylate HPr at a serine residue (Ser-46) (Stülke & Hillen, 1999).

The HPr protein is encoded by the *ptsH* gene which has been characterised in *C. acetobutylicum* ATCC 824 (Tangney *et al.*, 2003). Tangney *et al.*, (2003) searched the *C. acetobutylicum* genome for an HPr homolog and the identified gene was amplified and cloned into the pCR2.1 TOPO TA cloning vector. The vector was then transformed into *E. coli* LBG1605 which harbours three separate mutations in the *ptsH* gene (thus is unable to ferment PTS sugars). The introduction of the clostridial HPr gene complemented the phosphotransferase defect and endowed the transformant with the ability to ferment PTS sugars, confirming the identity of the gene. In addition to cloning, the deduced protein sequence of the HPr was aligned with thirteen other HPr sequences from well characterised low GC Gram positive bacteria. They found that the clostridial sequence differed slightly from the consensus, particularly around the phosphorylatable Ser-46 site (Tangney *et al.*, 2003).

Since 2003 many more clostridial genomes have been sequenced, it was of interest to investigate whether the divergence from the consensus around the Ser-46 site is a unique feature of *C. acetobutylicum* or a Genus wide attribute. Using the protein sequence of the clostridial HPr an alignment was carried out against a selected group of low GC Gram positive bacteria – including several other clostridial sequences. The results of this alignment are presented in Figure 3.1. The sequences have large homologous regions throughout, however of particular significance, the clostridial sequences surrounding (the HPrK/P phosphorylatable) Ser-46 site share the consensus deviations from the rest of the group (residues which are identical in only the clostidial proteins are underlined by blocks (Figure 3.1).

Figure 3.1: Multiple alignment of HPr sequences

The HPr sequences from several well characterised low GC Gram positive bacteria are compared with clostridial HPr proteins. The phosphorylation sites at His-15 and Ser-46are indicated by arrows. Residues that are identical in the clostridial proteins are underlined by blocks (▭); residues that are identical in all sequences are within black boxes; conserved residues are shaded. Abbreviations (with UniProt accession numbers are in brackets) are as follows: B.su, *Bacillus subtilis* (P08877); E.fa, *Enterococcus faecalis* (P07515); L.mo, *Listeria monocytogenes* (P0A438); L.sa, *Lactobacillus sakei* (O07125); S.au, *Staphylococcus aureus* (P0A0E3); S.mu, *Streptococcus mutans* (P45596); C.ac, *Clostridium acetobutylicum* (Q97I34); C.pe, *C. perfringens* (Q8XJT4); C.kl, *C. kluyveri* ATCC 8527 (A5N863); C.lj, *C. ljungdahlii* ATCC 55383 (D8GS50); C.bo, *C. botulinum* ATCC 19397 (A7FVX2); C.be, *C. beijerinckii* NCIMB 8052 (A6LSS1); *C. tetani* (DAA02130).

The EI protein is encoded by the *ptsI* gene and in most low GC Gram positive bacteria the *ptsH* and *ptsI* genes are present together as an operon (Vadeboncoeur *et al.*, 2000). However, in *C. acetobutylicum* the two genes appear to be monocistronic and are separated by over 1Mbp (Table 3.1). This phenomenon seems to be consistent throughout the clostridial branch of the Firmicutes, as the *ptsH* and *ptsI* genes in *C. beijerinckii* (Tangney & Mitchell, 2005), *C. tetani* (Tangney & Mitchell, 2004) and *C. perfringens* (Shimizu *et al.*, 2002) are all monocistronic.

Table 3.1: Genes encoding general energy coupling proteins in *C. acetobutylicum*

Locus tag	Gene	Encoded protein	Residues	protein id / db xref	UniProt Accession	Reference
cac1820	*ptsH*	HPr	86	AY196477.1 / GI:37781365	Q97I34	Tangney *et al.*, 2003
cac3087	*ptsI*	Enzyme I	539	NP_349687.1 / GI:15896338	Q97EM3	Mitchell & Tangney, 2005

3.4 PTS permeases of *C. acetobutylicum*

The *C. acetobutylicum* genome was screened for genes encoding sugar specific PTS permeases (EII) using the bioinformatics tools described in Chapter 2. These results are presented in Table 3.2, unless a reference is given, the sugar substrate indicated in the table is inferred by homology. The domain structure of each EII was determined using the Conserved Domain Database (CDD) (Marchler-Bauer *et al.*, 2005; Marchler-Bauer *et al.*, 2007; Marchler-Bauer & Bryant, 2004). Protein id, db xref and UniProt numbers are also presented in the table, as well as the predicted number of amino acid residues in each protein.

It was found that six of the thirteen permeases belong to the glucose-glucoside (Glu) family. There are two permeases in each of the fructose-mannitol (Fru), lactose-N-N'-diacetylchitobiose-β-glucoside (Lac) and the mannose (Man) families. There is also a single representative of the Gat family. A lone EIIA domain belonging to the Glu family is also encoded within the genome (Table 3.2).

Table 3.2: Genes encoding sugar specific PTS permeases in *C. acetobutylicum*

Locus tag	Sugar Substrate	Domain structure	Residues	protein id / db xref	UniProt Accession	Family	Reference
cac0154	Mannitol	IICF	481	AAK78138.1 / GI:15022983	O65989	Fru	Behrens *et al.*, 2001
cac0156	Mannitol	IIA	146	AAK78140.1 / GI:15022985	Q97MN7	Fru	Behrens *et al.*, 2001
cac0233	Fructose	IIA	147	AAK78214.1 / GI:15023067	Q97MG5	Fru	Inferred from homology
cac0234	Fructose	IIBC	458	AAK78215.1 / GI:15023068	Q97MG4	Fru	Inferred from homology
cac0383	β-glucoside	IIA	106	AAK78363.1 / GI:15023233	Q97M17	Lac	This work - Chapter 4, System II
cac0384	β-glucoside	IIB	102	AAK78364.1 / GI:15023234	Q97M16	Lac	This work - Chapter 4, System II
cac0386	β-glucoside	IIC	405	AAK78366.1 / GI:15023236	Q97M14	Lac	This work - Chapter 4, System II
cac0423	Sucrose	IIBCA	627	AAK78403.1 / GI:15023277	Q9L8G6	Glu	Tangney & Mitchell, 2000
cac0532	Maltose	IICB	531	AAK78511.1 / GI:15023396	Q97LM5	Glu	Tangney *et al.*, 2001
cac0570	Glucose	IICBA	665	AAK78549.1 / GI:15023438	Q97LJ0	Glu	Tangney & Mitchell, 2007
cac1353	N-acetylglucosamine	IIC	488	AAK79321.1 / GI:15024287	Q97JD3	Glu	Inferred from homology
cac1354	N-acetylglucosamine	IIA	159	AAK79322.1 / GI:15024288	Q97JD2	Glu	Inferred from homology
cac1407	β-glucoside	IIBCA	628	AAK79375.1 / GI:15024346	Q97J79	Glu	This work - Chapter 4, System I
cac1457	Fructose / Mannose	IIA	142	AAK79425.1 / GI:15024401	Q97J29	Man	This work - Chapter 6
cac1458	Fructose / Mannose	IIB	164	AAK79426.1 / GI:15024402	Q97J28	Man	This work - Chapter 6
cac1459	Fructose / Mannose	IIC	281	AAK79427.1 / GI:15024403	Q97J27	Man	This work - Chapter 6
cac1460	Fructose / Mannose	IID	277	AAK79428.1 / GI:15024404	Q97J26	Man	This work - Chapter 6
cac2956	Galactose	IIC	470	AAK80898.1 / GI:15026009	Q97EZ9	Gat	Inferred from homology
cac2957	Galactose	IIB	101	AAK80899.1 / GI:15026010	Q97EZ8	Gat	Inferred from homology
cac2958	Galactose	IIA	164	AAK80900.1 / GI:15026011	Q97EZ7	Gat	Inferred from homology
cac2964	Lactose	IICB	560	AAK80906.1 / GI:15026018	Q97EZ1	Lac	Yu *et al.*, 2007
cac2965	Lactose	IIA	104	AAK80907.1 / GI:15026019	Q97EZ0	Lac	Yu *et al.*, 2007
cac2995	Glc?	IIA	157	AAK80936.1 / GI:15026051	Q97EW1	Glu	Inferred from homology
cac3425	α-glucoside	IICB	518	AAK81355.1 / GI:15026513	Q97DP7	Glu	Thompson *et al.*, 2004
cac3427	α-glucoside	IIA	165	AAK81357.1 / GI:15026516	Q97DP5	Glu	Thompson *et al.*, 2004
cap0066	Fructose / mannose	IIAB	325	AAK76812.1 / GI:14994382	Q97TN2	Man	This work - Chapter 6
cap0067	Fructose / mannose	IIC	268	AAK76813.1 / GI:14994383	Q97TN1	Man	This work - Chapter 6
cap0068	Fructose / mannose	IID	303	AAK76814.1 / GI:14994384	Q97TN0	Man	This work - Chapter 6

3.5 Genetic organisation of gene clusters encoding sugar-specific PTS permeases and potential regulatory proteins

3.5.1 Glu family

The glucose–glucoside family is normally concerned with the transport of substrates such as glucose, *N*-acetylglucosamine and various α- and β-glucosides. Investigation of the genome revealed that this family has the largest number of representatives within the *C. acetobutylicum* genome, operon diagrams representing the genes which encode each of these permeases and the surrounding genes have been calculated and are presented in Figure 3.2.

The sucrose permease (Figure 3.2A), has been previously characterised (Tangney & Mitchell, 2000). It was established that the EII permease has the domain order of BCA and is found in an operon *scrTAKB,* encoding an antiterminator, an EIIBCA, a fructokinase and a sucrose-6-phosphate hydrolase respectively. The genes were induced by sucrose and transcribed in a single transcript. ScrT was found to be homologous to the BglG family of antiterminator proteins and a highly conserved ribonucleic antiterminator target (RAT) sequence was identified upstream of *scrT*, and a more unusual RAT was identified upstream of *scrA* (Tangney & Mitchell, 2000). In contrast, the gene arrangement of the sucrose-induced operon found in *C. beijerinckii* is *scrARBK,* encoding an EIIBC, a transcriptional repressor, a sucrose-6-phosphate hydrolase and a fructokinase (Reid *et al.,* 1999). The two sucrose-6-phosphate hydrolase proteins share 37% identity (data not shown). Apart from the different gene order in *C. beijerinckii* the EII appears to only encode the B and C domains, and instead of antiterminator mediated regulation transcription is controlled by a repressor.

The results revealed that there is also a β-glucoside operon with similar gene order to the sucrose operon, encoding an antiterminator, EIIBCA and a hydrolase which would carry out the first steps in breaking down the substrate to enter central metabolism (Figure 3.2B). This operon is discussed in detail in Chapter 4, Part I.

Upstream of the glucose EIICBA permease is a BglG type antiterminator (Figure 3.2C). Glucose utilisation has previously been investigated (Tangney & Mitchell, 2007) PTS activity was detected in extracts of glucose grown cells. Fractionation of cell extract revealed that both membrane and soluble fractions were required for PTS activity (Tangney & Mitchell, 2007).

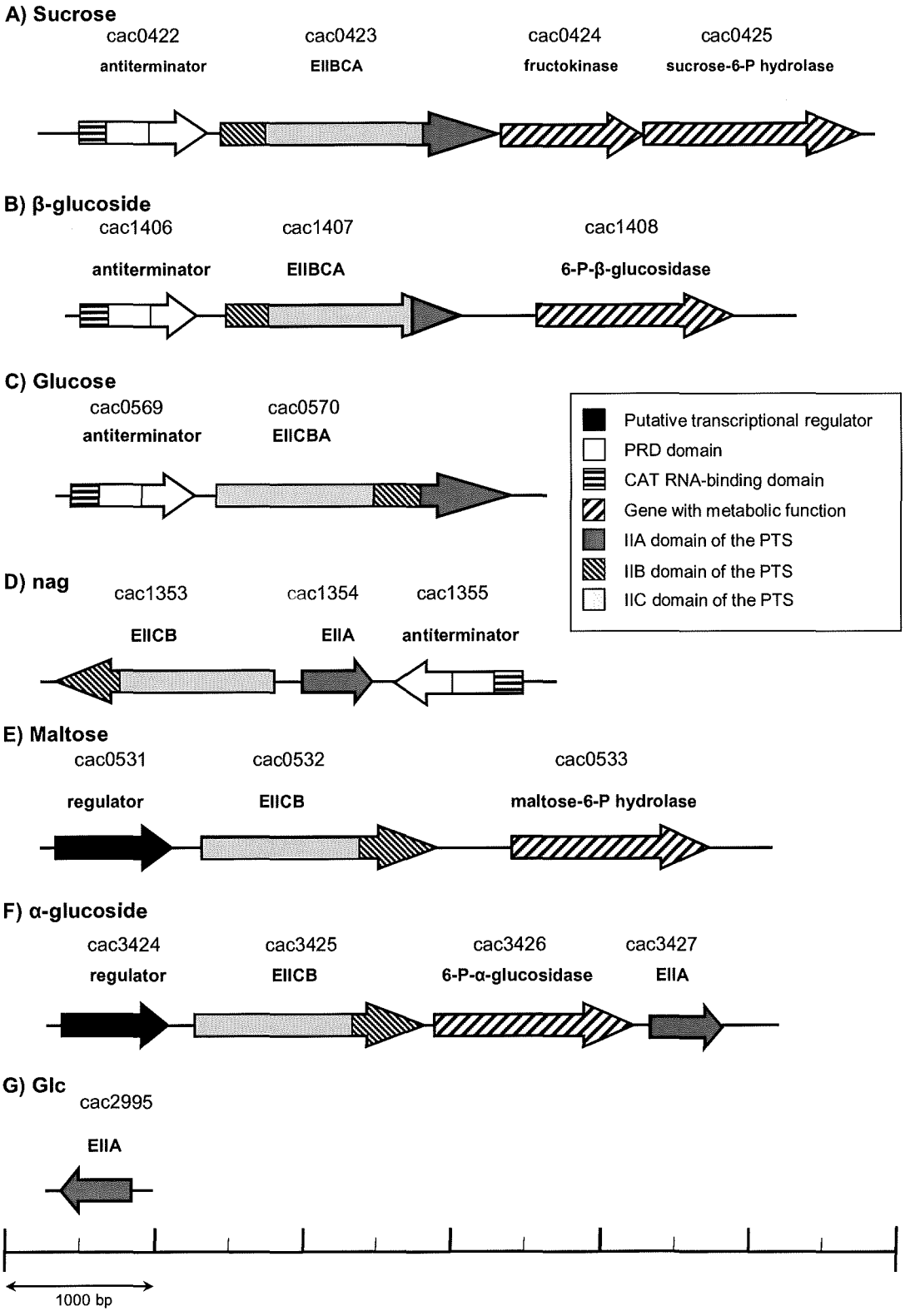


Figure 3.2: Genetic organisation of gene clusters encoding sugar-specific PTS permeases of the Glu family

It is of interest that in the related low GC Gram positive bacterium *Bacillus subtilis* control of the phosphotransferase-dependent accumulation and subsequent metabolism of the substrates sucrose (Debarbouille *et al.*, 1990; Zukowski *et al.,* 1990), glucose (Bachem & Stülke, 1998) and β-glucosides (Schnetz *et al.*, 1996) is apparently also mediated by a BglG-type antiterminator.

Examination of the *C. acetobutylicum* genome indicated the presence of a fourth putative antiterminator associated with an unusual gene cluster of divergently transcribed genes encoding EIICB and EIIA proteins (Figure 3.2D). In *B. subtilis* the *nagP* gene, encoding the putative *N*-acetylglucosamine-specific EIICB, appears to form a monocistronic transcriptional unit (Reizer *et al.*, 1999), with the genes involved in metabolism of *N*-acetylglucosamine metabolism located elsewhere on the chromosome (Kunst *et al.,* 1997). Homologs of *N*-acetylglucosamine-6-phosphate deacetylase (*nagA*; cac0188) and glucosamine-6-phosphate isomerase (*nagB*; cac1087) are similarly found elsewhere in the *C. acetobutylicum* chromosome, suggesting that *N*-acetylglucosamine potentially accumulated by the phosphotransferase system could be metabolised.

Maltose PTS activity was detected in maltose grown extracts of *C. acetobutylicum* (Tangney *et al.*, 2001). They also found that glucose was utilised preferentially to maltose and proposed the operon *malPH,* encoding an EIICB and a 6-phospho-α-D-glucosidase respectively. Upstream from the *malPH* operon is a gene encoding a putative RpiR type regulator (Figure 3.2E). Interestingly, if compared to the similar *pag* (phospho-α-glucoside) operon, a RpiR type regulator is also encoded upstream of an EIICB component, followed by a phospho-α-glucoside (designated *pagL*). However unlike the *mal* operon the *pag* operon contains an additional ORF encoding an EIIA domain (Figure 3.2F). MalH and PagL share 50% residue identity. Both proteins were expressed during growth on maltose but not on sucrose or glucose (Thompson *et al.*, 2004).

An orphan monocistronic $EIIA^{glc}$ domain was identified (Figure 3.2G). Lone $EIIA^{glc}$ domains is however not unique, *B. subtilis* possesses two individual lone $EIIA^{glc}$ domains; *ypqE* and the truncated *yyzE* (Reizer *et al.*, 1999). The sequences of the *C..acetobutylicum* $EIIA^{glc}$ and the *B. subtilis* YpqE were compared and found to share 39% identity.

3.5.2 Lac and Gat families

A second system which appears to be associated with β-glucoside utilisation was found in this analysis, which belongs to the Lac family. The system is comprised of an EIIA, EIIB, phospho-β-glucosidase, EIIC and a divergently expressed upstream regulator (Figure 3.3A). This system is discussed in detail in Chapter 4, Part II.

Of the Lac family, the most comprehensively characterised permease is in the lactose system (Figure 3.3C). Lactose is the major sugar found in industrial whey waste, which presents a significant environmental disposal problem (Mostafa, 2001). Previous studies have demonstrated that several clostridial species are able to utilise the sugar present in whey to form solvents (Maddox, 1980; Maddox *et al.*, 1993). Lactose PTS activity was determined in *C. acetobutylicum* cells by monitoring PEP dependent phosphorylation of radiolabelled lactose (Yu *et al.,* 2007). Lactose accumulated by the PTS would be phosphorylated on entering the cell, thus a phospho-β-galactosidase would be required for hydrolysis of lactose-6-phosphate to galactose-6-phosphate and glucose (Alpert & Chassys, 1990). Lactose grown cells of *C. acetobutylicum* were assayed, and found to possess phospho-β-galactosidase activity. The EII^{lac} was found to consist of both membrane associated and cytosolic substrate specific components. An operon was proposed *lacRFEG,* encoding a transcriptional regulator of the DeoR family, an EIIA, an EIICB and a phospho-β-galactosidase. The *lac* operon was found to be induced in the presence of lactose, but not glucose. Indeed, in cells grown in the presence of both lactose and glucose the *lac* operon was not expressed until all of the glucose was exhausted from the medium (Yu *et al.,* 2007). However, several aspects of the regulation of this system remain to be elucidated and as discussed below it would be of great interest to investigate whether the tagatose 6-phosphate pathway is induced by growth on lactose.

The single representative of the galactitol (Gat) family was found to be spread over three ORFs, associated with genes encoding enzymes of the tagatose 6-phosphate pathway (Figure 3.3B). A comparative genomic study of phosphotransferases from 174 bacterial genomes found only 4% possessed representatives of the Gat family of permeases, whereas 30% of the permeases observed belonged to the Glc family. Interestingly the Glc, Fru, Man, and Lac family systems often occur in multiple copies, whereas the Asc, Gat, and Gut systems are usually only found as a single copy in each organism (Barabote & Saier, 2005).

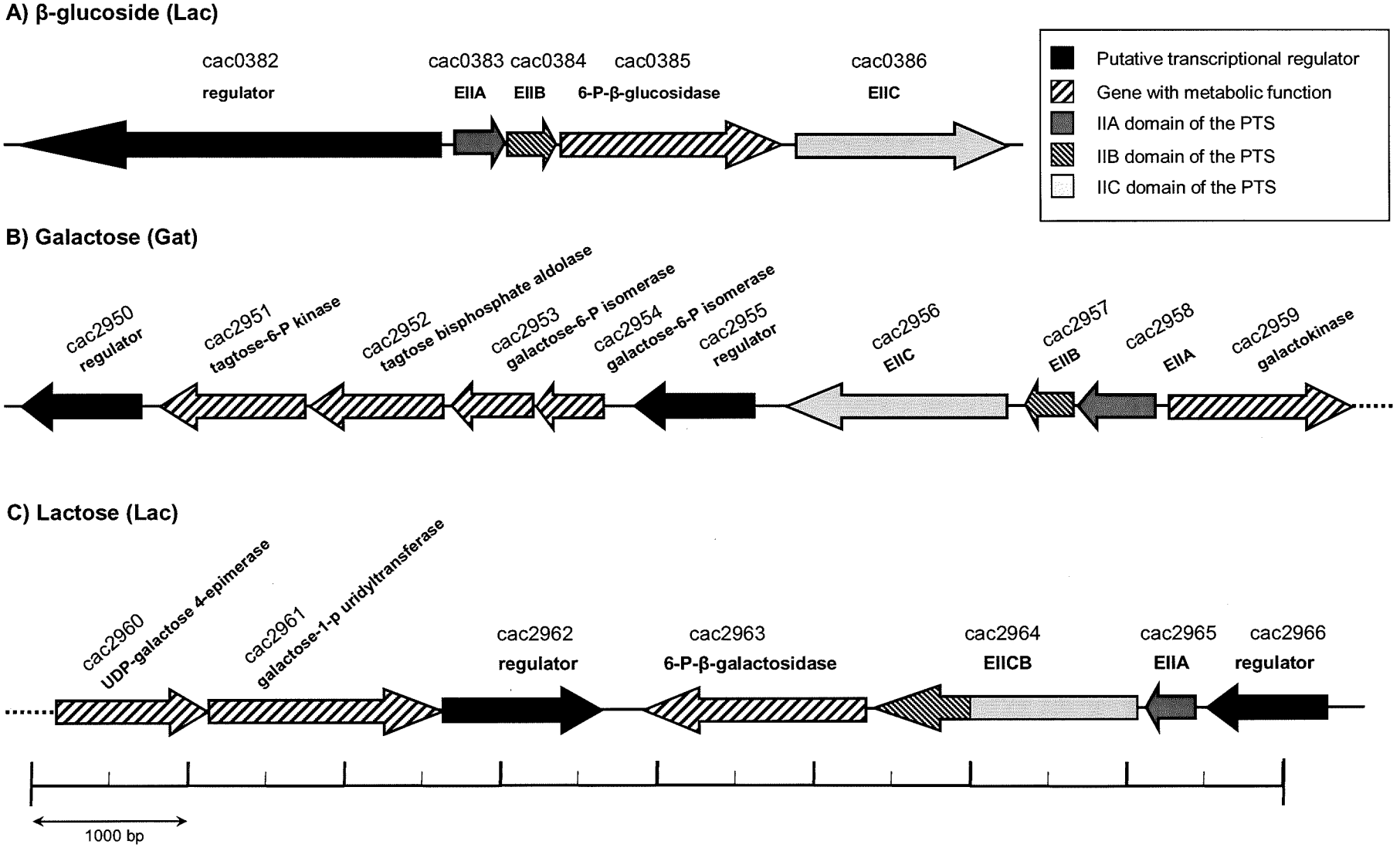


Figure 3.3: Genetic organisation of gene clusters encoding sugar-specific PTS permeases of the Lac and Gat families

Members of the Gat family usually transport galactitol and D-arabitol (Barabote & Saier, 2005). Typically, genes encoding galactitol permeases are associated with a galactitol-1-phosphate dehydrogenase (Nobelmann & Lengeler, 1996). However, the Gat family permease in *C. acetobutylicum* was found to be associated with a galactose-6-phosphate isomerase, rather than a galactitol-1-phosphate dehydrogenase, as would be expected for a functional galactitol PTS, suggesting that the permease is galactose specific.

However, galactose grown extracts of *C. acetobutylicum* were shown to have inconclusive PTS activity for galactose (Mitchell & Tangney, 2005). Little research has been conducted on the characterization of the EII^{gat} and the surrounding gene cluster. It is possibly significant that in several other Gram positive organisms, genes for the metabolism of galactose-6-phosphate are included within the lactose PTS operons (Rooijen *et al.,* 1991; Rosey *et al.,* 1991; Rosey & Stewart, 1992). In *C. acetobutylicum* the genes between the lactose PTS and the putative metabolic genes for galactose-6-phosphate are only separated by eight ORFs. Yu *et al.*, (2007) found that galactose-6-phosphate isomerase is induced by growth on lactose. Therefore it would be interesting to investigate whether the entire tagatose 6-phosphate pathway is induced by galactose and lactose. Also further investigation of the substrate specificity of the Gat family permease needs to be conducted.

3.5.3 Man and Fru families

Analysis of the *C. acetobutylicum* genome and megaplasmid (pSOL1) found two permeases from the Man family (Figure 3.4A & B). The domains of the EII on the chromosome are split over four ORFs (EIIA, EIIB, EIIC and EIID), whereas the EII on pSOL1 is spread over three ORFs (IIAB, IIC and IID). These systems are discussed in detail in Chapter 6.

Further examination of the genome revealed two members of the fructose-mannitol (Fru) family present on the *C. acetobutylicum* chromosome. Permeases of the Fru family usually phosphorylate the substrate on the 1-position (Saier *et al.*, 2006). The mannitol utilisation system had been characterised in *C. acetobutylicum* DSM 792 (Behrens *et al.*, 2001). The mannitol induced operon was found to contain four genes in the order *mtlARFD,* encoding; an EIICB, a PRD-containing transcriptional activator, an EIIA, and a mannitol-1-phosphate dehydrogenase. A highly similar operon was found to exist within the *C. acetobutylicum* ATCC 824 genome. When the deduced amino acid sequences were compared, they were found to share 99%, 99%, 90% and 100% identity respectively (data not shown). Even the truncated *orfP* located upstream of the *C. acetobutylicum* DSM 792 *mtl* operon matches 100% to the coordinate ORF found in *C. acetobutylicum* ATCC 824 (data not shown). The similarities suggest that this operon is concerned with the regulated transport and utilisation of mannitol.

The second permease of the Fru family is split between two ORFs encoding an EIIA domain and an EIIBC domain. Directly upstream is a 1-phosphofructokinase (Figure 3.4D). This suggests that fructose may be transported by the PTS, phosphorylated and subsequently phosphorylated by the 1-phosphofructokinase, yielding fructose-1,6-bisphosphate. Upstream from the 1-phosphofructokinase lays a putative regulator of the DeoR family. The fructose phosphotransferase system in the high GC Gram positive *Corynebacterium glutamicum* is associated with a DeoR regulator. When the deduced amino acid sequence of the regulators from *C. acetobutylicum* and *C. glutamicum* are compared, they are found to share 31% identity (49% positive).

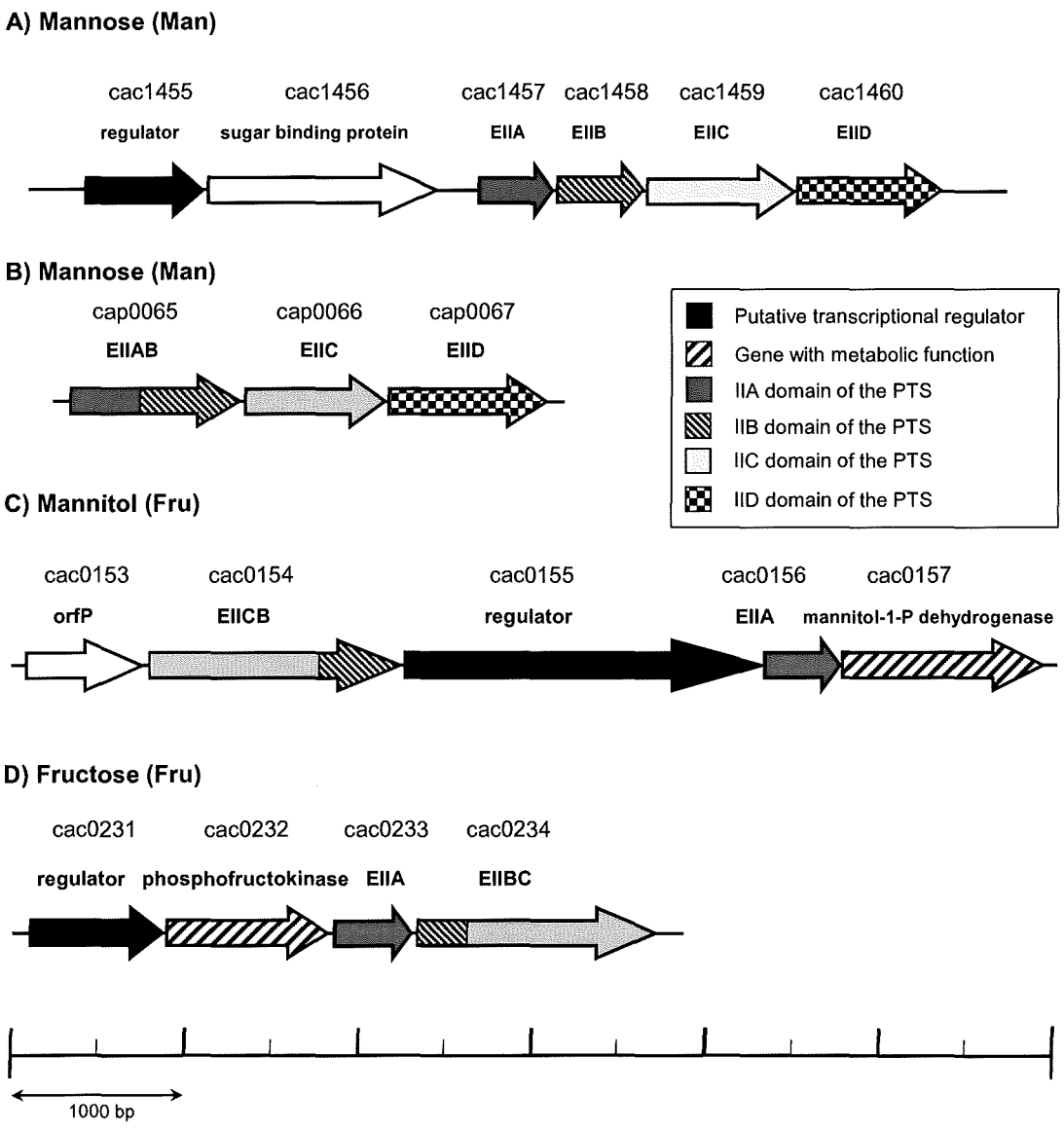


Figure 3.4: Genetic organisation of gene clusters encoding sugar-specific PTS permeases of the Man and Fru families

3.6 Regulators of *pts* operons

PTS operons can be controlled by RNA-binding antiterminators, such as the *B. subtilis licT* antiterminator (Tilbeurgh & Declerck, 2001) or DNA-binding transcriptional activators, for example the *B. subtilis licR* activator (Tobisch *et al.*, 1999). The mechanisms of these PRD containing regulators are discussed in Chapter 1, Section 1.5.

Analysis reveals the *C. acetobutylicum* genome encodes four putative transcriptional antiterminators (cac0422, cac1406, cac0569 and cac1355) and two transcriptional activators (cac0382 and cac0155). There are also seven other putative regulators associated with *pts* operons, the results are presented in Table 3.3.

Transcriptional antiterminators consist of an N-terminal RNA binding domain followed by two PRDs each containing two well conserved histidine residues (Manival *et al.*, 1997). An alignment of the four antiterminators from *C. acetobutylicum* presented in Figure 3.5.

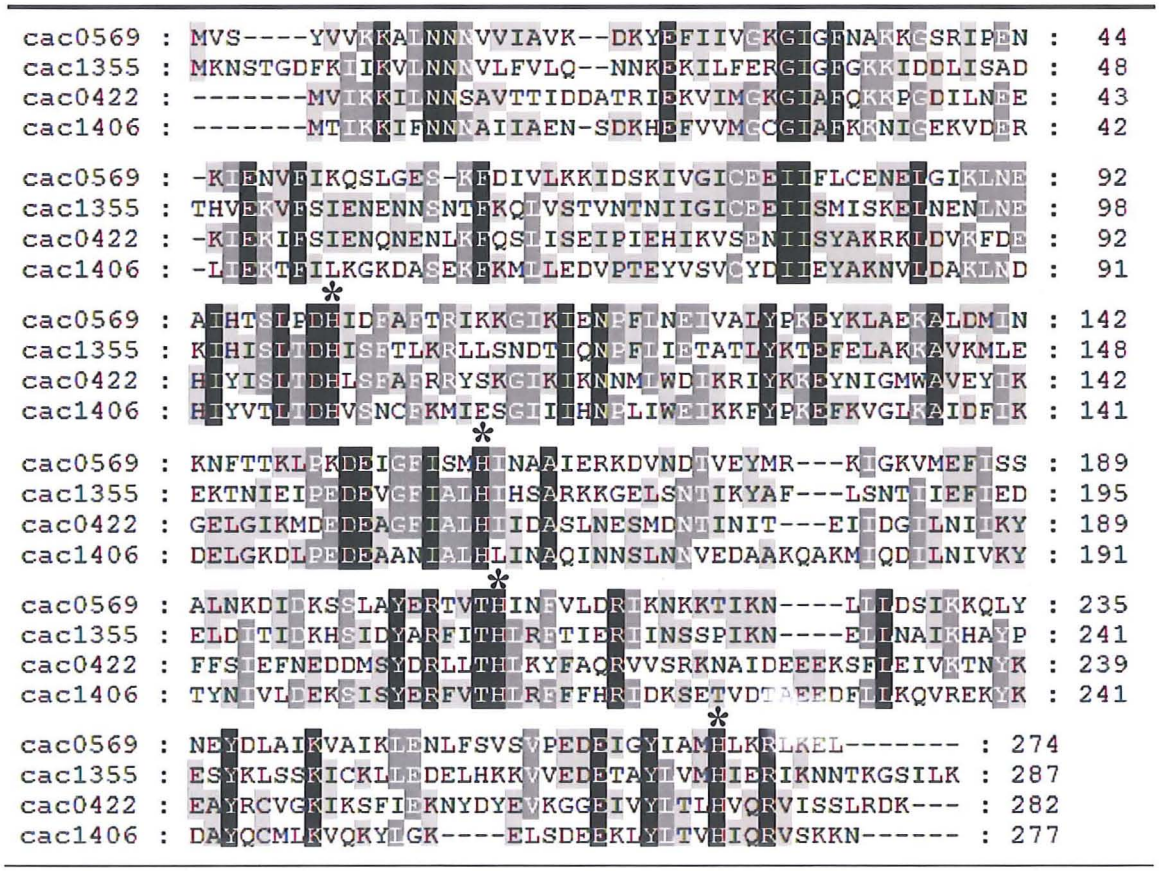


Figure 3.5: Alignment of the antiterminator complement of *C. acetobutylicum*

Dashes within the sequences indicate gaps giving optimal alignment. Identical residues are within black boxes and conserved residues are shaded. Conserved histidine residues, associated with the regulation of antiterminator proteins, are indicated by an asterix above.

The domain order of transcriptional antiterminators are consistent (Figure 3.6A). In contrast, transcriptional activators have much more variation in the domain order of the protein (Figure 3.6B), however in general contain a DNA binding helix-turn-helix motif, two PRDs and an EIIA-like domain are usually present.

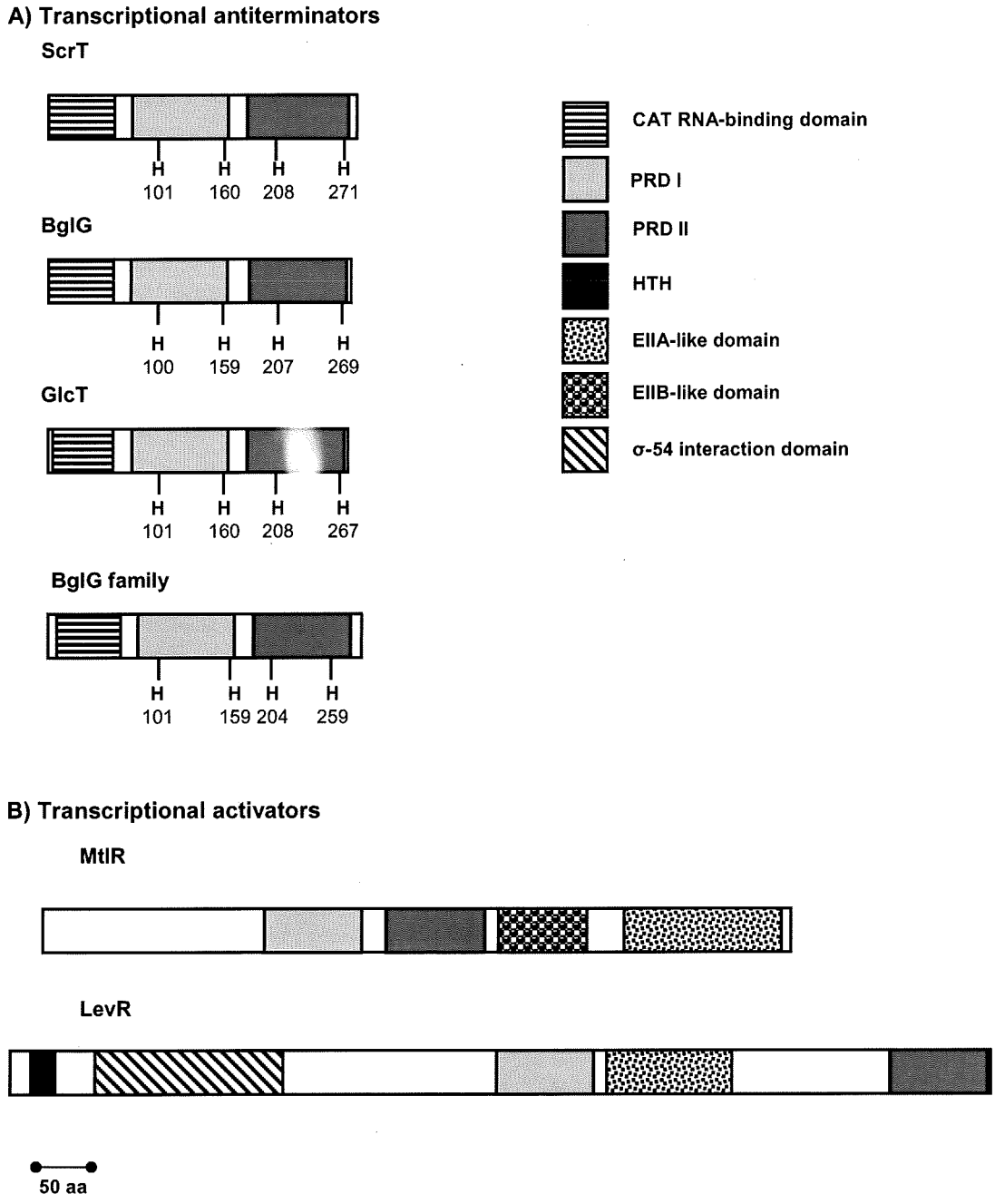


Figure 3.6: PRD containing transcriptional regulators of *C. acetobutylicum*
Conserved histidine residues associated with the operation of antiterminator proteins are indicated below the relevant positions.

Table 3.3: Genes encoding PTS associated transcriptional regulator proteins in *C. acetobutylicum*

Gene	Substrate	Residues	Locus tag	protein id / db xref	UniProt Accession	Reference
Transcriptional antiterminators						
scrT	Sucrose	282	cac0422	AAK78402.1 / GI:15023276	Q9L8G7	Tangney & Mitchell, 2000
bglG	β-glucoside	277	cac1406	AAK79374.1 / GI:15024345	Q97J80	This work - Chapter 4, System I
glcT	Glucose	274	cac0569	AAK78548.1 / GI:15023437	Q97LJ1	Tangney & Mitchell, 2007
bglG family	NAG	287	cac1355	AAK79323.1 / GI:15024289	Q97JD1	Inferred by homology
Transcriptional activators						
celR	Cellobiose	896	cac0382	AAK78362.1 / GI:15023232	Q97M18	This work - Chapter 4, System II
mtlR	Mannitol	684	cac0155	AAK78139.1 / GI:15022984	Q97MN8	Behrens *et al.*, 2001
Putative regulators						
RpiR	Maltose	257	cac0531	AAK78510.1 / GI:15023395	Q97LM6	Inferred by homology
Transcriptional regulator, RpiR family	α-glucoside	235	cac3424	AAK81354.1 / GI:15026512	Q97DP8	Inferred by homology
2-component system regulator-CheY domain & HTH DNA-binding domain	Mannose	225	cac1455	AAK79423.1 / GI:15024399	Q97J31	Inferred by homology
Transcripcional regulator of sugar metabolism	Fructose	254	cac0231	AAK78212.1 / GI:15023065	Q97MG7	Inferred by homology
Transcriptional regulators of the LacI family	Lactose	337	cac2962	AAK80904.1 / GI:15026016	Q97EZ3	Inferred by homology
Lactose phosphotransferase system repressor lacR	Galactose	254	cac2950	AAK80892.1 / GI:15026003	Q97F05	Inferred by homology
Lactose phosphotransferase system repressor lacR	Lactose	254	cac2966	AAK80908 / GI:15026020	Q97EY9	Inferred by homology

3.7 PTS associated proteins of *C. acetobutylicum*

3.7.1 Catabolite control protein (CcpA)

The mechanisms for carbon catabolite repression (CCR) in Gram negative bacteria, such as *E. coli*, differ significantly from the proposed mechanisms for low GC Gram positive bacteria (Saier *et al.*, 1995). The major transcription factor mediating CCR in Gram positive bacteria, such as *B. subtilis,* is the catabolite control protein (CcpA) (Moreno *et al.*, 2001). CcpA belongs to the LacI-GalR family of transcriptional repressors/activators (Deutscher *et al.*, 1995). Using a highly specific polyclonal antiserum, CcpA-like proteins were identified in a large number of low and high GC Gram positive bacteria, suggesting that the mechanism of CCR proposed for *B. subtilis* and *B. megaterium* could be widespread in Gram positive bacteria (Küster *et al.,* 1996). A number of *ccpA* genes have been cloned and their involvement in CCR confirmed, for example those from *Streptococcus mutans* (Simpson & Russell, 1998), *Staphylococcus xylosus* (Egeter & Bruckner, 1996), *Lactobacillus casei* (Monedero *et al.*, 1997) and *L. pentosus* (Lokman *et al.*, 1997).

C. saccharobutylicum (previously classified as *C. acetobutylicum* P262) was found to encode a gene referred to as *regA* which complemented a *B. subtilis ccpA* mutant (Davison *et al.*, 1995). From the genome of *C. acetobutylicum* ATCC 824 a *ccpA* homologue was found that encodes a protein which has 62% identity to the *C. saccharobutylicum* RegA (Tangney *et al.,* 2003).

The *ccpA* homolog identified in *C. acetobutylicum* (locus tag cac3037) by *in silico* methods was later cloned and expressed in a *B. subtilis ccpA* deficient mutant (Ren *et al.*, 2010). The clostridial CcpA was found to complement the *B. subtilis* deletion, furthermore, when the *ccpA* gene was deleted from *C. acetobutylicum,* CCR of xylose by glucose was abolished. Thus, confirming the original proposal that cac3037 encodes a CcpA protein involved in global carbon regulation.

Table 3.4: Genes encoding PTS associated proteins in *C. acetobutylicum*

Gene	Encoded protein	Residues	Locus tag	UniProt Accession	Reference
ccpA	CcpA (member of the LacI/GalR)	334	cac3037	P58258	Tangney *et al.*, 2003
glpX	fructose bis-phosphatase	324	cac1088	Q97K33	Inferred from homology Figure 3.9
fbp	fructose bis-phosphatase	665	cac1572	Q97IR8	Inferred from homology Figure 3.9
Metabolic enzymes					
glpK	glycerol kinase	502	cac1321	Q97JG4	Inferred from homology
pykA	pyruvate kinase	473	cac0518	O08309	Belouski *et al.*, 1998
pykA	pyruvate kinase	472	cac1036	Q97K83	Inferred from homology
Protein kinases/phosphorylases					
prkA	serine protein kinase	646	cac0579	Q97LI2	Inferred from homology
hprK	HPr (Ser) kinase/phosphorylase	304	cac1089	Q97K32	Tangney *et al.*, 2003

3.7.2 Fructose 1,6-bisphosphatase in *Clostridium acetobutylicum*

The enzyme fructose 1,6-bisphosphatase (EC 3.1.3.11; FBPase) has been characterised in a wide range of bacteria, yeasts and higher eukaryotes. FBPase is responsible for the second irreversible step in gluconeogenesis, it catalyses the hydrolysis of fructose-1,6-bisphosphate (fbp) to fructose-6-phosphate and inorganic phosphate. Under glycolytic conditions the opposite reaction is catalysed by 6-phosphofructokinase (Figure 3.7).

From the extensive characterisation of the known FBPases four main classes have emerged. FBPase I (class I) in *E. coli* is encoded by the *fbp* gene (Sedivy *et al.*, 1984) and is highly similar to eukaryotic FBPases. FBPase II is encoded by the *glpX* gene in *E. coli* (Donahue *et al.*, 2000); homologs of this gene are widespread throughout the prokaryotic genera. FBPases in class III are structurally distinct from other classes of FBPases and are mainly found in *Firmicutes* as exemplified by the *Bacillus subtilis* encoded *yydE,* now renamed as *fbp* (Fujita & Freese, 1979; Fujita *et al.*, 1998). Enzymes of the fourth class FBPase IV are present in archaea, as illustrated by *Pyrococcus furiosus* (Verhees *et al.*, 2002), and have been found to show higher sequence similarity to inositol-monophosphatases than to FBPases classes I-III. Indeed, analysis of the FBPase IV in *P. furiosus* and *Methanococcus jannaschii* reveals they also possess inositol-monophosphatase activity (Verhees *et al.*, 2002; Johnson *et al.*, 2001). An emerging fifth class of FBPases has also been proposed (Rashid *et al.*, 2002).

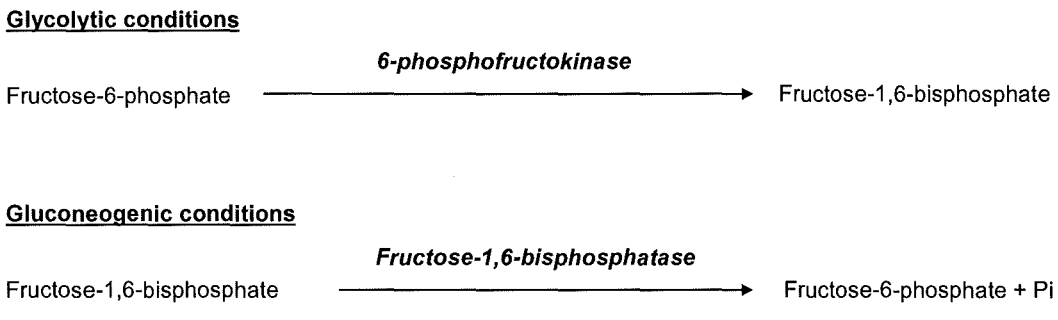


Figure 3.7: The enzymes involved in the interconversion of fructose-6-phosphate and fructose-1,6-bisphosphate.

The genome of *C. acetobutylicum* was examined for potential homologues of fructose-1,6-bisphosphatase encoding genes. Two potential homologues were recognized. The first gene identified (cac1088) putatively encodes a GlpX-like protein of 324aa (Figure 3.8; *glpX*). The gene directly upstream from *glpX* is the *hprK* gene, which encodes the HPr kinase/phosphorylase (HprK/P).

The second homolog (cac1572) putatively encodes an Fbp-like protein of 665aa (Figure 3.8; *fbp*). Upstream of cac1572 are two genes encoding putative glutathione peroxidases. Immediately downstream from cac1572 is a gene which encodes a *B. subtilis* ortholog of *yjdF*. In *B. subtilis* this gene is associated with the mannose specific PTS, however when the gene was knocked out it had no effect on mannose utilisation (Sun & Altenbuchner, 2010).

The amino acid sequences of experimentally verified and published FBPase proteins were obtained from a search of the UniProt database and used for phylogenetic comparison with the two putative clostridial sequences. The results are presented in Figure 3.9. The clostridial sequence cac1088 clearly falls on a branch with class II GlpX-like sequences, whereas the second clostridial sequence cac1572 lies on a branch of class III FBPase sequences, suggesting that *C. acetobutylicum* possesses a class II and a class III FBPase.

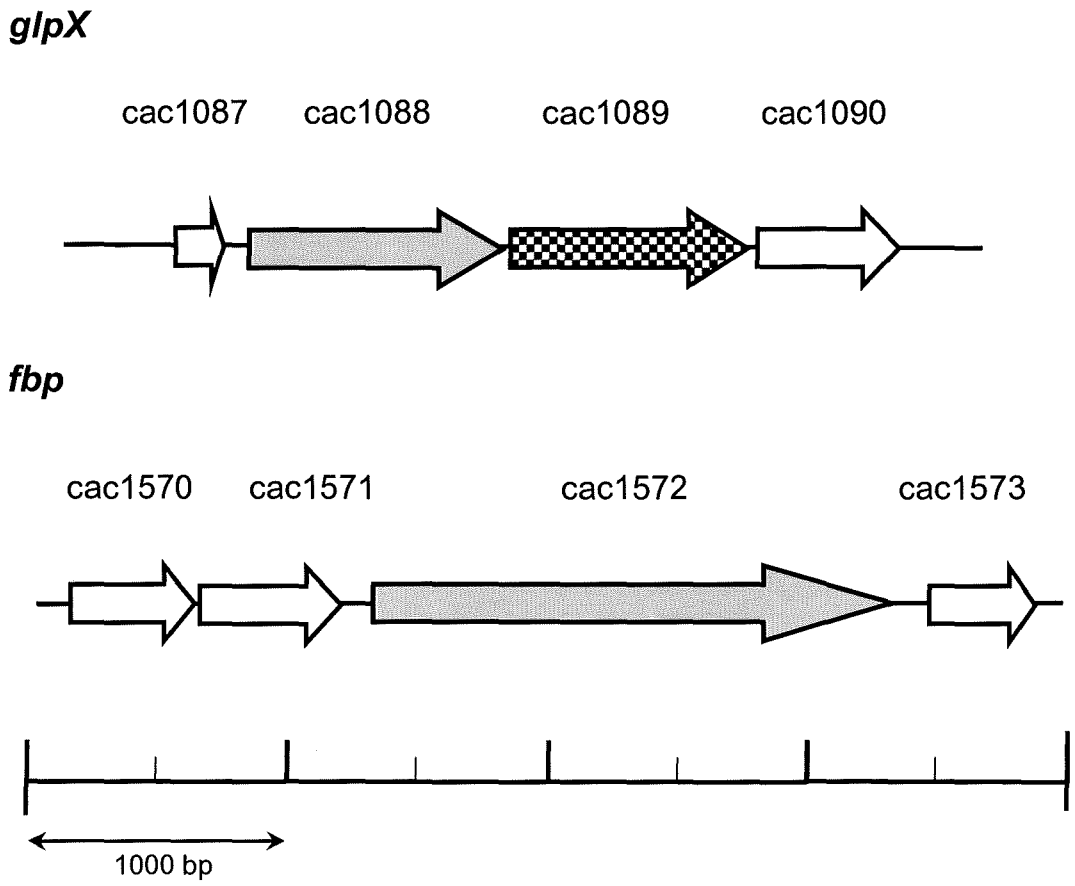
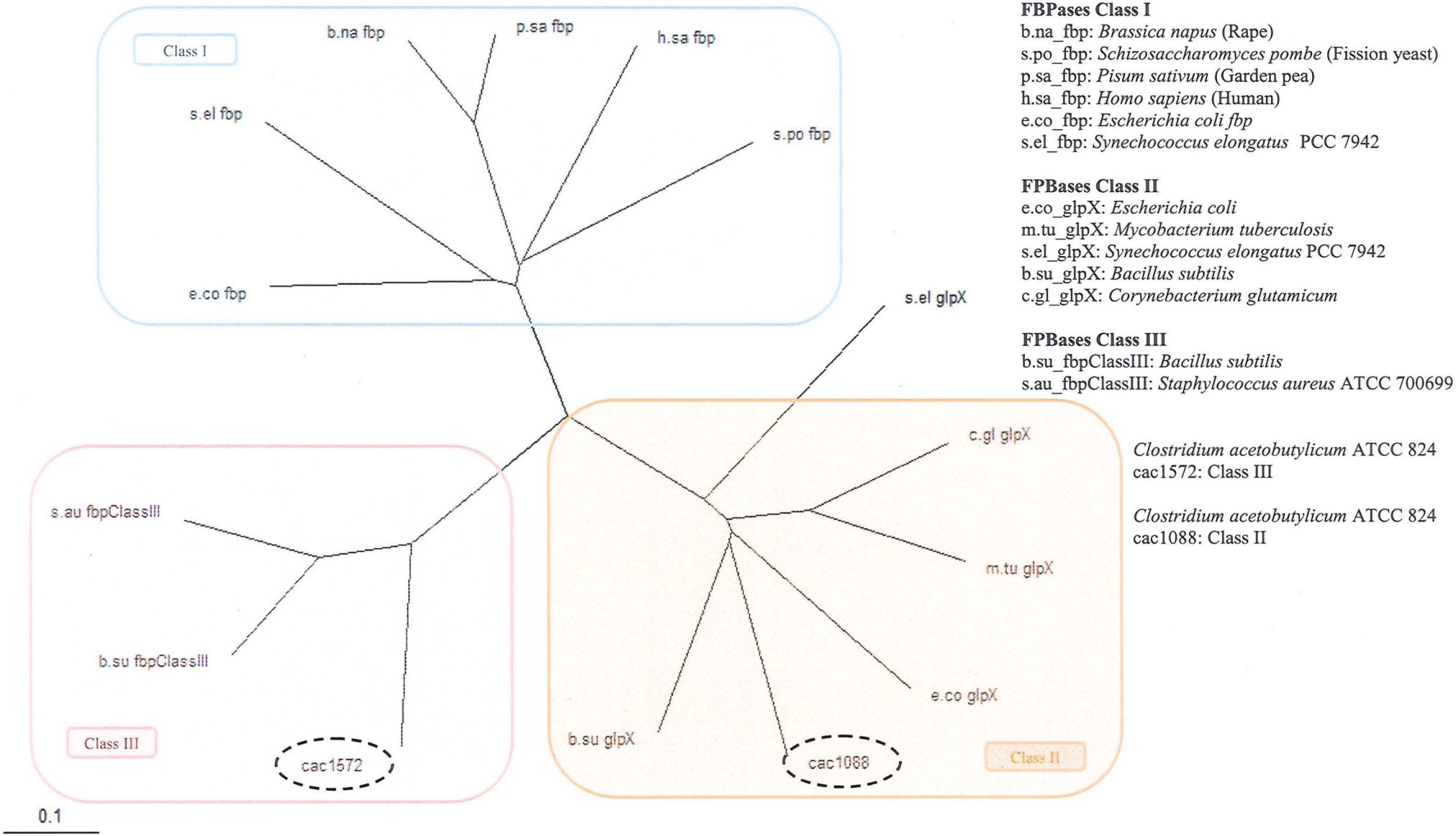


Figure 3.8: Organisation of the gene clusters encoding putative fructose-1,6-bis phosphatase genes

Scale bar is given at the bottom of the figure. The FBPase genes are highlighted in grey and the putative HPr kinase/phosphorylase encoding gene *hprK* is checkered.



FBPases Class I
b.na_fbp: *Brassica napus* (Rape)
s.po_fbp: *Schizosaccharomyces pombe* (Fission yeast)
p.sa_fbp: *Pisum sativum* (Garden pea)
h.sa_fbp: *Homo sapiens* (Human)
e.co_fbp: *Escherichia coli fbp*
s.el_fbp: *Synechococcus elongatus* PCC 7942

FPBases Class II
e.co_glpX: *Escherichia coli*
m.tu_glpX: *Mycobacterium tuberculosis*
s.el_glpX: *Synechococcus elongatus* PCC 7942
b.su_glpX: *Bacillus subtilis*
c.gl_glpX: *Corynebacterium glutamicum*

FPBases Class III
b.su_fbpClassIII: *Bacillus subtilis*
s.au_fbpClassIII: *Staphylococcus aureus* ATCC 700699

Clostridium acetobutylicum ATCC 824
cac1572: Class III

Clostridium acetobutylicum ATCC 824
cac1088: Class II

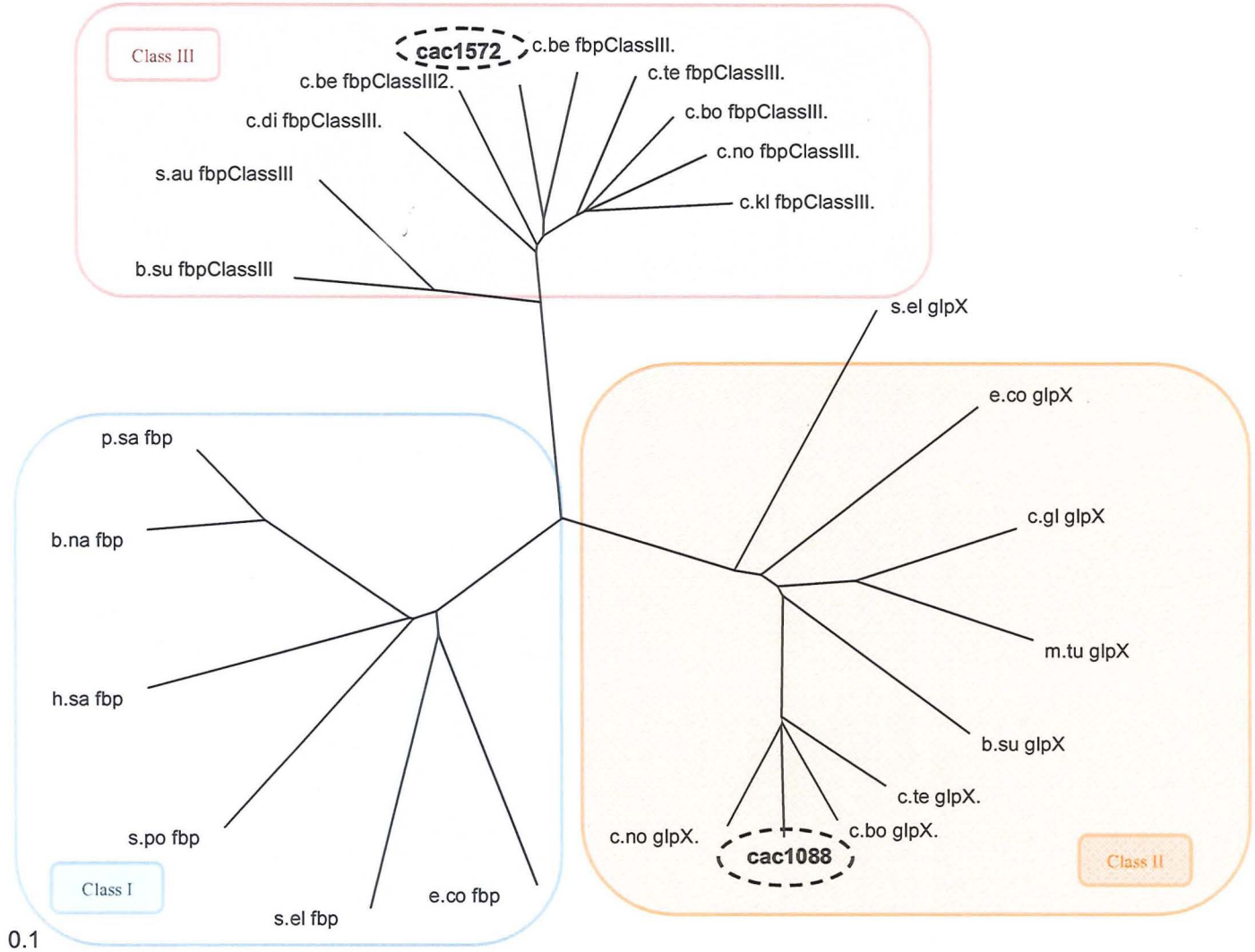
Figure 3.9: Unrooted phylogenetic tree of characterised FBPases and the two putative clostridial FBPases

3.7.2.1 FBPase complement of other organisms

The best studied Gram negative organism, *Escherichia coli* possesses both a class I FBPase (Sedivy *et al.,* 1984) and a class II GlpX-like FBPase (Donahue *et al.,* 2000). The cyanobacterium *Synechococcus elongatus* PCC 7942 also possesses both a class I and class II GlpX FBPase (Tamoi *et al.,* 1996). *Mycobacterium tuberculosis* Rv1099c possesses a class II GlpX-like FBPase (Movahedzadeh *et al.,* 2004) and no other FBPase encoding gene has so far been identified in this organism. The Gram positive, high GC bacterium *Corynebacterium glutamicum* ATCC 13032 also has only one gene encoding a class III FBPase (Rittmann *et al,* 2003). In contrast the well characterised representative of low GC Gram positive bacteria, *Bacillus subtilis* possesses both a class II FBPase [originally *ywjI* now *glpX*] (Jules *et al.,* 2009) and a class III FBPase [originally *yydE* now classed as *fbp*] (Fujita *et al.,* 1998). This may reflect the differing lifestyle of these bacteria.

3.7.2.2 Do other clostridia possess FBPases?

No other FBPase encoding gene has been characterised and published in any other clostridial species. However, from the sequences available on UniProtKB server (http://www.uniprot.org/) several predicted FBPase's (inferred from homology) were collected and added to the analysis. *C. botulinum, C. novyi,* and *C. tetani* all have a predicted class II GlpX-like and class III FBPases. *C. difficile, C. kluyveri,* and *C. beijerinckii* all possess putative class III FBPases, and interestingly *C. beijerinckii* has two copies of this gene which share 66% identity (and 81% positive) with each other (Figure 3.10). As can also be seen from Figure 3.10, the clostridial proteins are all closely related.



FBPases Class I
b.na_fbp: *Brassica napus* (Rape)
s.po_fbp: *Schizosaccharomyces pombe* (Fission yeast)
p.sa_fbp: *Pisum sativum* (Garden pea)
h.sa_fbp: *Homo sapiens* (Human)
e.co_fbp: *Escherichia coli fbp*
s.el_fbp: *Synechococcus elongatus* PCC 7942

FPBases Class II
e.co_glpX: *Escherichia coli*
m.tu_glpX: *Mycobacterium tuberculosis*
s.el_glpX: *Synechococcus elongatus* PCC 7942
b.su_glpX: *Bacillus subtilis*
c.gl_glpX: *Corynebacterium glutamicum*
c.no_glpX: *Clostridium novi*
c.bo_glpX: *Clostridium botulinum*
c.te_glpX: *Clostridium tetani*

FPBases Class III
b.su_fbpClassIII: *Bacillus subtilis*
s.au_fbpClassIII: *Staphylococcus aureus* ATCC 700699
c.di_fbpClassIII: *Clostridium difficile*
c.be_fbpClassIII: *Clostridiumbeijerinckii*
c.te_fbpClassIII: *Clostridium tetani*
c.bo_fbpClassIII: *Clostridium botulinum*
c.no_fbpClassIII: *Clostridium novi*
c.kl_fbpClassIII: *Clostridium kluyveri*

Clostridium acetobutylicum ATCC 824 cac1572: Class III

Clostridium acetobutylicum ATCC 824 cac1088: Class II

Figure 3.10: Unrooted phylogenetic tree of characterised FBPases and several putative clostridial FBPases

3.7.3 Metabolic enzymes

Glycerol kinase is involved in the regulation, transport and metabolism of glycerol, and carries out the ATP dependent phosphorylation of glycerol to yield glycerol-3-phosphate. The glycerol kinase (*glpK*) gene of *B. subtilis* has been sequenced and characterised (Holmberg *et al.*, 1990). Components of the PTS have been shown to interact directly with glycerol kinase to regulate glycerol (a non-PTS substrate) uptake and metabolism (Wehtje *et al*, 1995). By searching the *C. acetobutylicum* genome with the sequence of the *B. subtilis* GlpK protein one significant match was identified, which shares 60% identity and 78% positives (Table 3.4).

Pyruvate kinase catalyses the ATP dependent phosphorylation of pyruvate to phosphoenolpyruvate. The *pykA* gene in *C. acetobutylicum* has been characterised (Belouski *et al.*, 1998). However, it was indicated from the analysis that also encoded within the genome is potentially a second pyruvate kinase which share together 63% identity and 79% positive (Table 3.4).

3.7.4 HPr kinase phosphorylase

The amino acid sequence of HprK/P from *C. acetobutylicum* has previously been aligned with HprK/P sequences from a selection of organisms (Fieulaine *et al.*, 2001), and a conservation of a HprK/P protein family signature sequence and the Walker A motif were observed (Reizer *et al.*, 1989; Walker *et al.*, 1982).

To confirm that HprK/P operated in this organism, extracts of glucose grown cells were assayed for ATP dependent phosphorylation of HPr. ATP dependent kinase activity was identified only in the presence of fructose 1, 6-bisphosphate (Tangney *et al.*, 2003).

As mentioned above the *hprK* gene lies directly downstream from the *glpX* fructose-1,6 bisphosphatase gene (Figure 3.8). A selection of sequenced clostridial genomes (*C. saccharolyticum WM1, C. cellulolyticum H10, C. botulinum A str. ATCC 19397, C. beijerinckii NCIMB 8052*), were screened for homologs of the *C. acetobutylicum* HprK/P all of these clostridia were found to harbour a HprK/P homolog, however none possessed the same genome context arrangement as *C. acetobutylicum* (*glpX* followed by *hprK*; as shown in Figure 3.8).

The *hprK* gene encodes the enzyme HPr kinase/phosphorylase (HprK/P), which can - in the presence of fructose-1,6-bisphosphate (fbp), phosphorylate HPr at a serine residue (Ser-46). When phosphorylated at Ser-46 HPr can complex with the catabolite control protein (CcpA), to effect gene regulation by binding to DNA sequences called catabolite

responsive elements (CREs). Directly upstream from *hprK* lies *glpX* which encodes GlpX, a fructose 1,6-biophosphatase (FBPase). FBPase catalyses the hydrolysis of fbp to fructose-6-phosphate and inorganic phosphate.

3.8 Conclusions and summary

In bacterial industrial solvent formation, carbohydrate accumulation is often a major bottleneck in the fermentation process. Bacteria have multiple carbohydrate transport systems and are subject to complex regulatory cascades, thus identification of the transport proteins and elucidation of the regulatory mechanisms of carbon accumulation can provide vital information for the manipulation and optimisation of the bottleneck point.

The PEP dependent PTS is often the predominant system of carbohydrate assimilation in anaerobic and facultatively anaerobic bacteria. The genome of *C. acetobutylicum* encodes a complement of thirteen phosphotransferases, which have similar domain structure to the phosphotransferases found in other bacteria and include representatives of six families of permeases. Genes encoding the general PTS proteins EI and HPr were identified along with genes encoding proteins implicated in PTS associated global carbon regulation mechanisms. The majority of these findings were confirmed by the use of several different bioinformatic algorithms to analyse the data sets, to ensure that no spurious results were obtained. For example, to predict the ORFs studied in this research two *in silico* tools were utilised; the NCBI server, which can be set for translation of bacterial code, and the EasyGene1.2 server, which can be set specifically for gene prediction in clostridia. However, it was found that each predicted ORF studied was presented by both algorithms. The specific EasyGene tool is more likely to be required when characterising unusual groups of bacteria or extremophiles.

Experimental confirmation of many of these *in silico* findings are completed or underway. Using techniques such as: cloning the permease genes in relevant mutants to look for complementation of the defect, biochemical assays of the metabolic enzymes and RNA expression studies of the operons as a whole.

The rest of this thesis is dedicated to a more detailed examination of the phosphotransferases associated with the utilisation of the carbohydrates present in plant waste biomass. The next chapter will focus on the two systems, identified from the genome linked with β-glucoside utilisation.

Chapter 4

4 β-glucoside transport in *C. acetobutylicum* ATCC 824

A significant aspect to the success of future bioenergy production facilities will be the consistent and complete conversion of one or more economically competitive substrates into product. Therefore, detailed knowledge of the genes which encode the PTS permeases and carbohydrate utilisation proteins, as well as the regulation mechanisms which govern them, may present targets for future metabolic engineers to improve bioreactor efficiency. Thus, enhancing the overall economics of the fermentation process. Potential fermentation feedstocks are discussed in detail in Section 1.3.6.

A potential major source of inexpensive substrate is lignocellulosic plant waste biomass, which is comprised of large quantities of plant sugars such as cellobiose, esculin, arbutin and salicin. This group of saccharides are referred to as β-glucosides, disaccharides containing (at least) one glucose moiety linked via a β-glucosidic bond (Wrolstad *et al.,* 1989). Hydrolysis of plant biomass releases cellobiose and other plant sugars, which may either be further hydrolysed to glucose, or transported into the cell intact.

Prokaryotes possess a range of membrane bound mechanisms for energizing the transport of sugars and other nutrients across their cell membrane, such as the ATP-dependent binding protein systems (part of the ABC superfamily of transporters) and the phosphoenolpyruvate-dependent phosphotransferase systems (PTS). The operation of the former is energised by the hydrolysis of ATP and is comprised of two components inside the cytoplasm that contain an ATP binding site and a substrate specific binding protein exposed on the cell surface (Schlösser *et al.,* 1999). The phosphotrasferase systems are dependent on the transfer of a phosphate group from PEP. The PTS is comprised of two general, cytoplasmic, proteins enzyme I (EI) and histidine containing or heat-stable protein (HPr) together with a transmembrane, multi-domain, substrate specific enzyme II (EII; Postma *et al.,* 1993). The mechanisms of carbohydrate transport and regulation systems are discussed in detail in Chapter 1.

Little is currently known about β-glucoside transport in *Clostridium acetobutylicum.* In the related organism *C. thermocellum* cellobiose uptake is energised directly by ATP hydrolysis (Nochur *et al.,* 1992), while in contrast *C. longisporum* possesses a PTS specific for the aryl-β-glucosides arbutin and salicin (Brown & Thomson, 1998).

Aim

Characterisation of the uptake and utilisation of β-glucosides would provide fundamental information that could be exploited by metabolic engineers for optimisation of sugar utilisation and thus biofuel production. There are no reports of research characterising the β-glucoside transport and utilisation genes in *C. acetobutylicum.*

In this chapter several important aspects of the utilisation of β-glucosides as substrate for biofuel production will be investigated based on the following studies.

1. Investigation of cell growth, β-glucoside utilisation and solvent production by *C. acetobutylicum.*
2. Search of the *C. acetobutylicum* genome to identify potential β-glucoside transport systems and analysis of the deduced amino acid sequence of the encoded proteins.
3. *In silico* analysis of putative regulatory regions of importance in carbon catabolite repression.
4. Cloning of the putative β-glucoside transport genes and investigating the protein products interaction *in vivo* with the *Escherichia coli* PTS.

4.1 Transport and solvent production

4.1.1 Culture identification

A Gram stain was routinely performed on cells of *Clostridium acetobutylicum* ATCC 824, revived from spore stocks, prior to the commencement of any experimental work. Spore revival and Gram stain procedures were carried out as stated in the experimental procedures section. Revived cells and Gram stained preparations were observed using an Olympus BX51 (bright field) microscope (magnification x1000) and images captured with a JVC 3CCD digital camera attachment. Stained cells appeared Gram positive with a rod-shaped morphology, as exemplified in Figure 4.1.

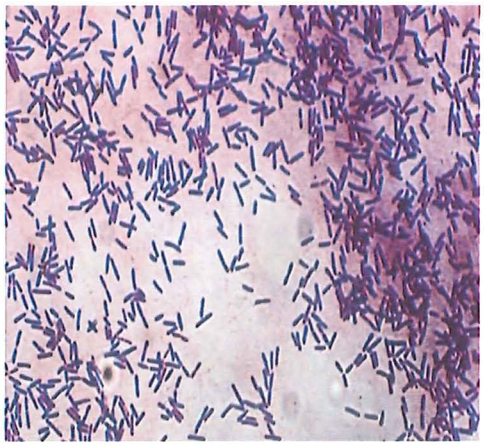


Figure 4.1: Gram stain of an overnight culture of *C. acetobutylicum* ATCC 824
C. acetobutylicum ATCC 824 spores were heat shocked and inoculated into RCM and incubated anaerobically overnight. A loopful of culture was then Gram stained by SOP and magnified x1000 by an Olympus BX51 microscope and photographed with a JVC 3CCD digital camera attachment. Gram positive rods can be observed.

4.1.2 Sugar utilisation by *C. acetobutylicum* ATCC 824

Hydrolysed plant waste material contains considerable quantities of β-glucosides. It is of interest to identify the range of these sugars *C. acetobutylicum* can utilise. *C. acetobutylicum* was cultured on Clostridial Basal Agar (CBA), supplemented with a range of individual β-glucosides as the sole carbon source, to determine which β-glucosides could support growth. CBA plates were prepared containing 20mM of carbon source. The results are presented in Table 4.1.

Table 4.1: β-glucoside utilisation on CBA by *C. acetobutylicum*

C. acetobutlyicum cultured on CBA, incubated at 37°C overnight containing either 20mM of cellobiose, arbutin, salicin, esculin, glucose or no carbon source.

Carbon Source	Growth / No Growth
Cellobiose	+
Arbutin	+
Salicin	+
Esculin	-
Glucose	+
No C-source	-

The partial hydrolysis of cellulose results in a mixture of cellooligosaccharides, cellobiose and glucose. In order to determine whether *C. acetobutylicum* could ferment both sugars simultaneously or if accumulation was subject to carbon catabolite repression, a dual sugar fermentation was employed. *C. acetobutylicum* cells grown initially with cellobiose as a sole carbon source were diluted into a medium supplemented with both cellobiose and glucose. Sugar concentrations were chosen based on past growth experiments (data not shown), which successfully demonstrate glucose depletion from the medium and cellobiose utilisation by *C. acetobutylicum* taking place within a suitable sampling time period. The optical density and utilisation of each substrate was monitored and results are presented in Figure 4.2. Sequential utilisation of sugars was observed. Glucose was initially depleted from the medium, concomitant with cell growth, whereas there was no perceptible utilisation of cellobiose until around hour 5, at which point most of the glucose has been depleted.

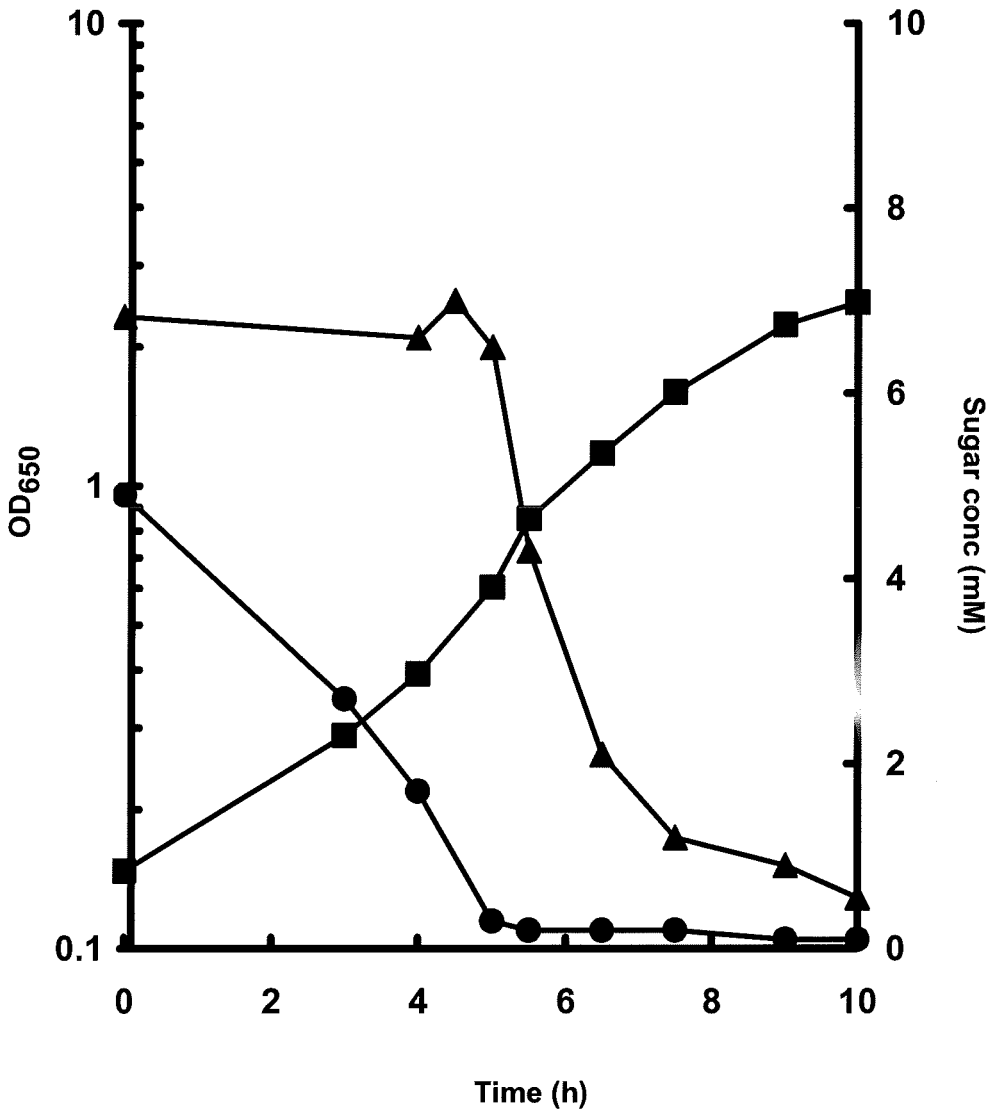


Figure 4.2: Culture growth and carbohydrate utilisation by *C. acetobutylicum*. Cells pregrown on cellobiose were inoculated into clostridial basal medium containing both glucose, and cellobiose. The OD_{650} was monitored over time to record growth. The glucose and cellobiose in the culture supernatant was determined at intervals. Glucose (●), cellobiose (▲), OD_{650} (■).

4.1.3 Solvent production from β-glucosides by *C. acetobutylicum* ATCC 824

As *C. acetobutylicum* can utilise glucose and cellobiose as a source of carbon and energy, experiments were conducted to determine solvent production from *C. acetobutylicum* when cultured in a medium with cellobiose as the sole carbon source. Overnight cultures of *C. acetobutylicum* were inoculated into 20ml universals of tryptone yeast extract medium (as per Ezeji & Blaschek, 2008) containing either cellobiose or glucose as the main carbon source at a variety of concentrations. Universals were incubated, without agitation, at 37°C within the anaerobic workstation. Samples were taken before inoculation for background sugar and solvent concentrations, and at hours 72, 96 and 120. Experiments were conducted in triplicate.

The acetone, butanol, ethanol and total ABE concentrations present in the culture supernatants at sample time-point 120 hours are presented in Figure 4.3. The averages of the triplicates were calculated and the standard deviation was calculated accordingly.

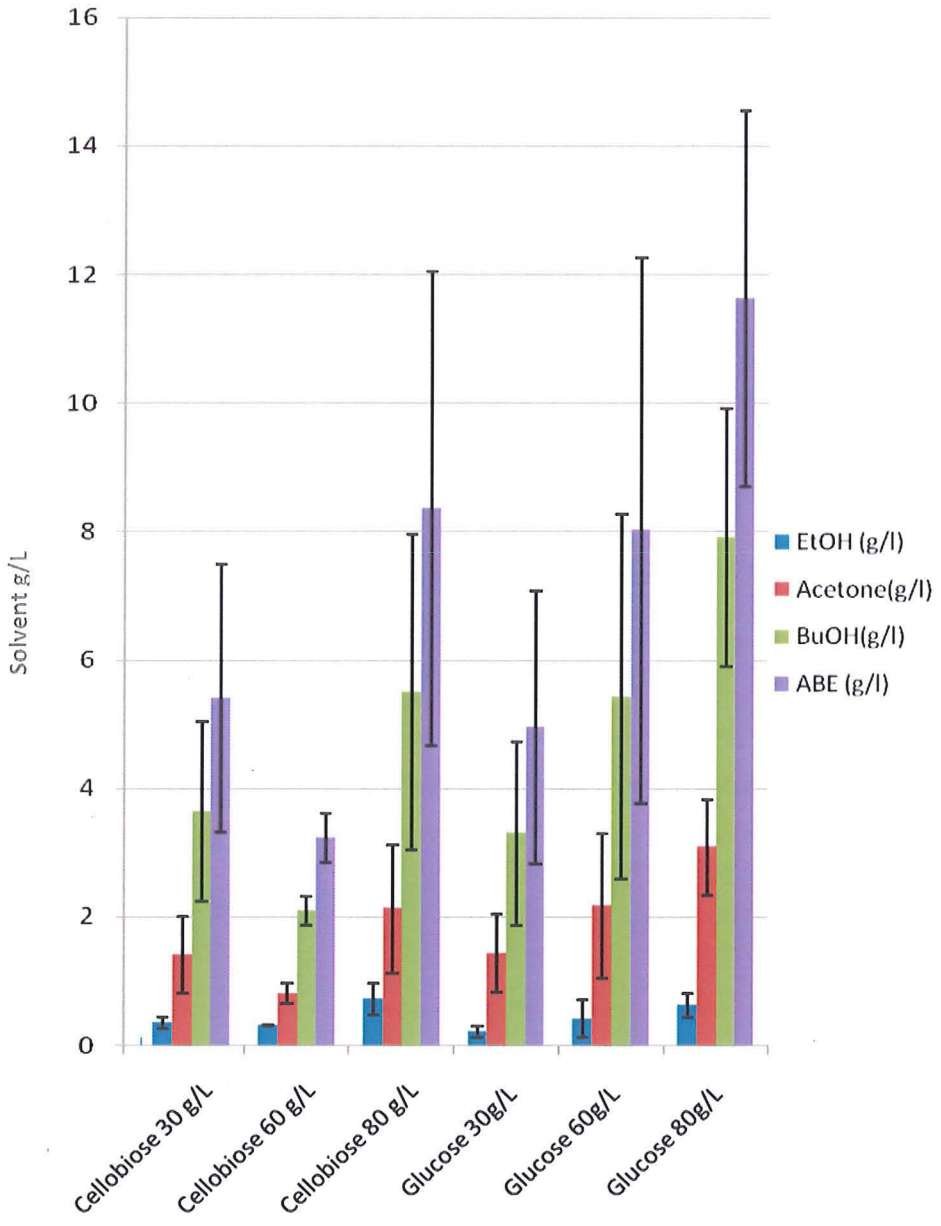


Figure 4.3: Production of ABE from cellobiose and glucose by *C. acetobutylicum*
The acetone, butanol, ethanol and total ABE concentration achieved by *C. acetobutylicum* from 30, 60 & 80g/L of carbon source (either glucose or cellobiose), after 120 hours of incubation. Bars are coloured as follows: ethanol – (■), acetone – (■), butanol – (■) and the total solvents – (■). Error bars show the standard deviation between the triplicate samples (n=3 +/- SE).

4.2 Sequence analysis of the putative β-glucoside transport systems of *C. acetobutylicum*

Having established that *C. acetobutylicum* can utilise the β-glucosides; cellobiose, arbutin, and salicin as a sole carbon source, the genes responsible for their transport and utilisation remains to be deduced. The chromosome and megaplasmid of *C. acetobutylicum* have been sequenced and the sequence data are available under the accession number AE001437 (Nölling *et al.*, 2001: http://www.ncbi.nlm.nih.gov/nuccore/AE001437). To identify putative transport genes the genome was analysed using the Genome Information Broker (GIB) server (Fumoto *et al.*, 2002), the open reading frame (ORF) finder tool at the National Centre for Biotechnology Information (NCBI) and the Basic Local Alignment Search Tool (BLAST) service at NCBI (Altschul *et al.*, 1997).

Analysis revealed the presence of two candidate systems encoded within the chromosome, potentially involved in β-glucoside utilisation. The first system (System I) covers a 4380bp region containing three ORFs with the locus tags cac1406, cac1407 and cac1408, which will be discussed in Part I. The second system (System II) consists of a 3900bp region containing four ORFs with the locus tags cac0383, cac0384, cac0385 and cac0386. A potential regulator cac0302, reading in the opposite direction, lies upstream from this cluster. System II will be discussed in Part II.

Part I, System I:

4.3 *In silico* characterisation of putative β-glucoside system cac1406-cac1408

Examination of the 4380bp region revealed the presence of three clustered ORFs, all in the same orientation and reading frame, with the locus tags cac1406, cac1407 and cac1408 (Figure 4.4). The deduced amino acid sequence of each putative ORF was queried against sequence databases using the BLAST to search for sequence homology. On the basis of this homology the putative genes were tentatively named *bglGAB,* putatively encoding an antiterminator protein (BglG), a PTS enzyme IIBCA (BglA) and a phospho-β-glucosidase (BglB). Evidence presented in this section.

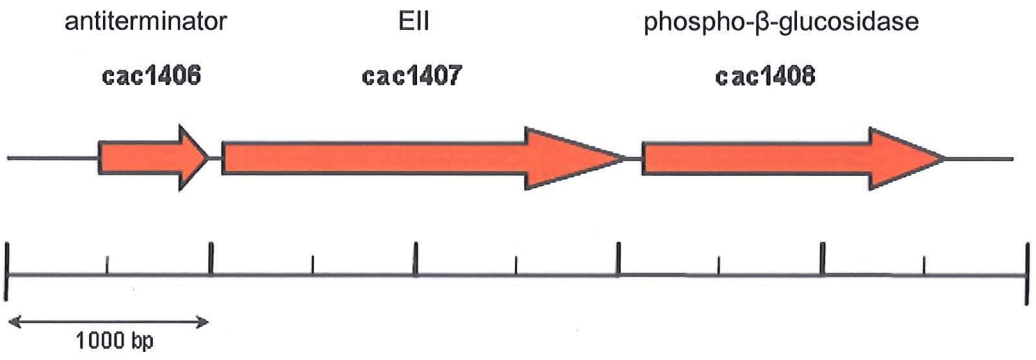


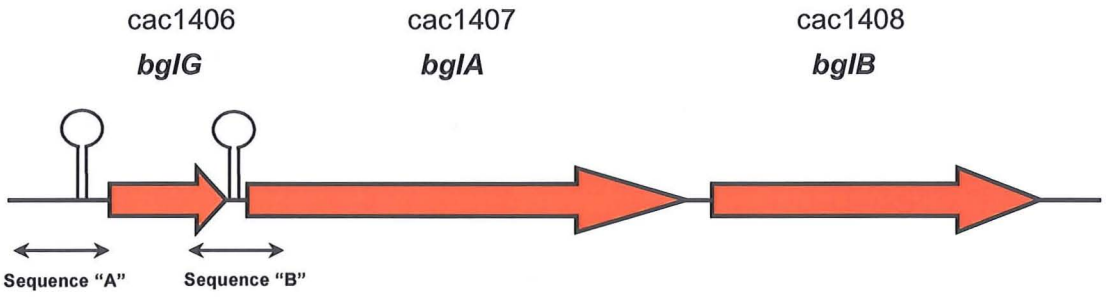
Figure 4.4: Arrangement of the ORFs of the putative β-glucoside operon
Predicted ORFs are represented by block arrows (indicating direction of transcript), locus tags are given above, scale (bp) is given below. The genome was examined using GIB server and the ORFs were predicted using the ORF finder tool at the NCBI server. Genes were tentatively named on the basis of the results of BLAST homology searches.

4.3.1 Regulatory features of the putative *C. acetobutylicum bgl* operon

The DNA sequence upstream of the putative *bglG* gene (stretching approximately 200nt) was analysed and found to contain a number of discrete regions which resemble known regulatory features. There is a putative promoter region which is in good agreement with the consensus clostridial promoter (Young *et al.,* 1989). The putative – 35 (TTGACA) and – 10 (TATAAT) are separated by 15 nucleotides. Overlapping the putative – 35 region is a sequence which closely resembles a catabolite responsive element (CRE). Nucleotide sequence and features are presented in Figure 4.5A. Downstream of the putative promoter is a region of dyad symmetry, followed by several T residues, which resembles a prokaryotic ρ-independent transcription terminator (Platt, 1986) as indicated by inverted arrows in Figure 4.5B. The secondary configuration of the terminator structure was predicted using RNA Mfold Server Version 3.2 (Zuker, 2003; Mathews *et al.,* 1999) and is presented in Figure 4.6.

A characteristic of operons controlled by antiterminators of the BglG family is the presence of a short highly conserved sequence which precedes, and partially overlaps, the terminator region, which is referred to as a ribonucleic antiterminator target (RAT) sequence (Schnetz *et al.,* 1996). An alignment of the sequence preceding, and partially overlapping, this terminator region (Figure 4.5A, underlined by •) with published RAT sequences from a variety of organisms clearly shows the high sequence similarity to these regulatory elements (Figure 4.7, RAT-1).

Situated in the intergenic region between *bglG* and *bglA* lies another potential transcriptional terminator which could prevent transcription of *bglAB*. The region was compared to the characterised RAT sequences, and a sequence with high homology to the other sequences was identified and is presented in Figure 4.7, RAT-2.



Sequence A

AAAAAAATTAATTCATATTAAGATTGGATTGAATAGTTAGTATGTATAAAACAAAAAACGACTAATTACGATATTT

TTTAAAAAAGTTA**TTGACA**ATTAATATAAAAGGG**TATAAT**ATTAAGACATGGTTAAACGATTACAAATTAATATCT
-35 -10

RAT-1

GGGATTGTTACTGATTCGATCAGGCGAGACCCGAGCTTTTGAAAGTAATATTTATATTATTTTCATTAGTTGGGGT

TTTTATTTTTAGATCAATGTTGTAAAT*GGGG*TAATTCTG **ATG ACG** ...
M T

Sequence B

RAT-2

ATAAAAATTT**GGATTGTTACTGGTAATGCAGGCGAGACC**AAAATAGACAGTATAAAATAAAATTATGCTATCTATG

GGGTCTCGCTTTTTTTATATTATTTATACTAATGAACTTATACATTATTCTTATATTTTATAAATCAAAAGAAA*AG*

*GAGA*GATTATC **ATG AAA** ...
M K

Figure 4.5: Nucleotide sequence of the DNA upstream and downstream of *bglG*

A. Nucleotide sequence of the regulatory region upstream of *bglG*.

B. Nucleotide sequence of the intergenic region between *bglG* and *bglA*. The potential –35 and –10 regions, marked accordingly, are underlined in bold text, and a putative CRE sequence is shaded. Putative transcription terminator sequences are indicated by arrows. Putative RAT sequences upstream of *bglG* and *bglA* are in bold text and marked as RAT-1 and RAT-2 respectively (● underneath). Potential ribosome binding site sequences are underlined and in italic text. Stem loops structures are located upstream and downstream of *bglG*.

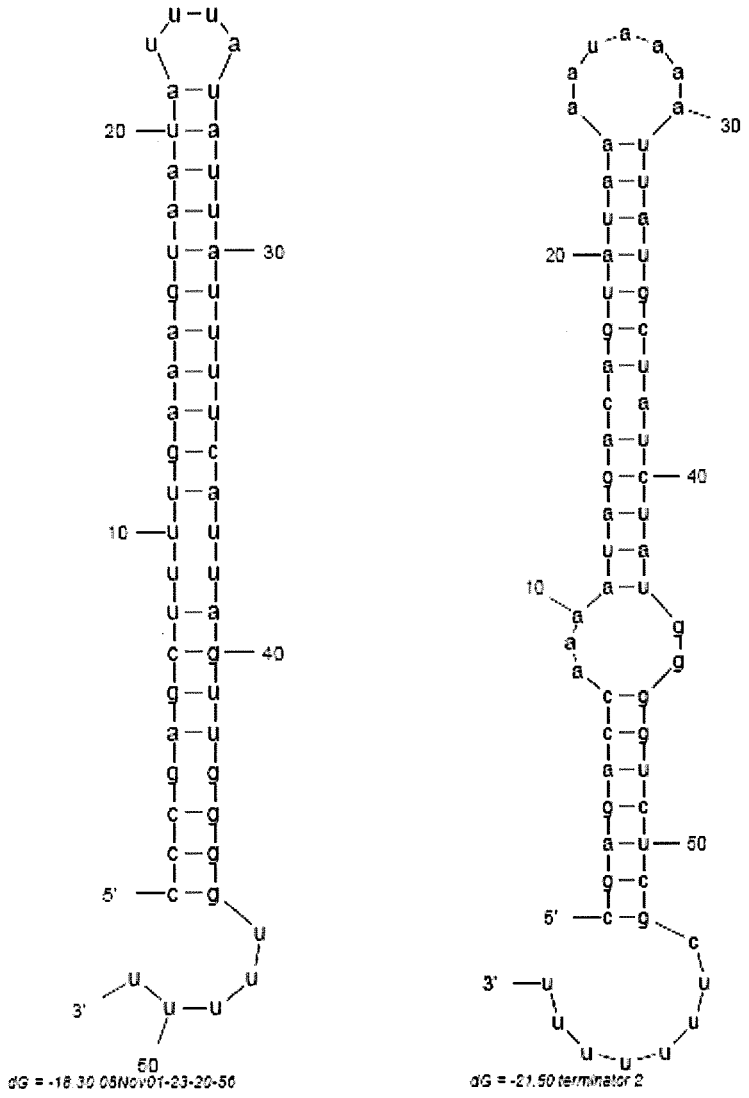


Figure 4.6: Rho-independent transcriptional terminator structures

Rho-independent transcriptional terminator structures are regions of dyad symmetry, followed by a run of several thymine residues. Secondary structures presented above were predicted using the RNA Mfold Server Version 3.2 (Zuker, 2003; Mathews *et al.*, 1999). The positions of the above secondary structures are indicated by inverted arrows in Figure 4.5.

RAT-1

```
C.ac_RAT-1  :  GGATTGTTACTGTTAA-GCAGGCAAAACC  :  28
C.lo_abgG2  :  GGATTGTTACTGATAATGCAGGCAAGACC  :  29
B.su_sacP   :  GGATTGTGACTGGTAAAGCAGGCAAGACC  :  29
B.su_sacB   :  GGTTTGTTACTGATAAAGCAGGCAAGACC  :  29
C.ac_scrT   :  GGATTGTTACTGTTAA-GCAGGCAAAACC  :  28
B.su_licT   :  GGATTGTTACTGCTAAAGCAGGCAAAACC  :  29
E.co bglG   :  GGATTGTTACTGCATTCGCAGGCAAAACC  :  29
```

RAT-2

```
C.ac_RAT-2  :  GGATTGTTACTGGTAATGCAGGCGAGACC  :  29
C.lo_abgG2  :  GGATTGTTACTGATAATGCAGGCAAGACC  :  29
B.su_sacP   :  GGATTGTGACTGGTAAAGCAGGCAAGACC  :  29
B.su_sacB   :  GGTTTGTTACTGATAAAGCAGGCAAGACC  :  29
C.ac_scrT   :  GGATTGTTACTGTTAA-GCAGGCAAAACC  :  28
B.su_licT   :  GGATTGTTACTGCTAAAGCAGGCAAAACC  :  29
E.co bglG   :  GGATTGTTACTGCATTCGCAGGCAAAACC  :  29
```

Figure 4.7: Sequence alignment of ribonucleic anti terminator target (RAT) sequences of the BglG family

Presented above is a multiple alignment of the putative RAT sequences of the *C. acetobutylicum bgl* operon with published RAT sequences from a variety of organisms. The clostridial RAT sequences are indicated by • underneath in Figure 4.5. C.ac_RAT-1 is the putative sequence located upstream of the *C. acetobutylicum bglG* gene. C.ac_RAT-2 is located upstream of the *C. acetobutylicum bglA* gene. Dashes within the sequences indicate gaps giving optimal alignment. Identical residues are within black boxes and conserved residues are shaded.

The RAT sequences aligned with C.ac_RAT-1 & -2 are the targets of the antiterminator proteins aligned with BglG (Figure 4.10) and lie upstream of the associated genes as follows; C.lo abgG2, *C. longisporum bglG* (Brown & Thomson, 1998); B.su sacP, *B. subtilis sacP* (Debarbouille *et al.,* 1990); B.su sacB, *B. subtilis sacB* (Crutz & Steinmetz, 1992); C.ac scrT, *C. acetobutylicum scrT* (Tangney & Mitchell, 2000); B.su licT, *B. subtilis licT* (Schnetz *et al.,* 1996); E.co bglG, *E. coli bglG* (Schnetz *et al.,* 1987).

4.3.2 Identification of the putative antiterminator protein BglG

The first ORF in the *bgl* operon (cac1406), encodes a putative 277 amino acid protein (834bp), which has a calculated molecular mass of 32,124Da. A possible rbs (GGGG) was identified 8bp upstream of the initiation codon. The amino acid sequence of cac1406 was used for a BLAST homology search and the results are presented in Table 4.2. The sequence exhibits significant identity (69% identity and 85% similarity) to a putative BglG type transcriptional antiterminator encoded by *Clostridium beijerinckii* NCIMB 8052 and to a transcriptional antiterminator protein of *Clostridium longisporum* (65% identity and 82% similarity, Brown & Thomson, 1998). Identity describes the percentage of amino acids that are identical to the input sequence and percentage similarity refers to the amino acids with similar properties to those in the query sequence.

Figure 4.8 presents the nucleotide and deduced amino acid sequence of the putative antiterminator protein, along with the upstream regulatory region showing the positions of the discrete regulatory features. Conserved histidine residues are highlighted and in bold. Utilising the top 10 sequences returned from the BLAST (Table 4.2), a multiple alignment was constructed using ClustalW2 (Larkin *et al.,* 2007) and GeneDoc (Nicholas *et al.,* 1997), as described in Chapter 2. GeneDoc similarity groups were disabled. The results of the multiple alignment is shown in Figure 4.9. Conserved histidine residues, associated with the regulation of antiterminator proteins, are marked with asterixes.

The sequences of the BglG antiterminator proteins which target the RAT sequences presented in Figure 4.7 were also aligned with the putative clostridial BglG protein (Figure 4.10). Significantly, consistent with the initial alignment of BglG with the BLAST results, conserved histidine residues (*) are found. Conservation is in distinct areas, even though the overall percentage identity of the protein is relatively low.

To examine the phylogenetic relationships between the query sequence and those returned by the BLAST, a phylogenetic tree was constructed. For completeness the sequences presented in Figure 4.7 were also included and presented in Figure 4.11. The sequences discussed in other papers are marked with a full stop after the protein name and the reference is given in the figure legend. It can be seen from the tree that the deduced amino acid sequence of cac1406 clusters on a deep rooted branch shared with an antiterminator protein associated with a characterised β-glucoside utilisation operon in *C. longisporum* (Brown & Thomson, 1998) and two other putative clostridial antiterminator proteins from *C. butyricum* and *C.beijerinckii*.

Table 4.2: BLAST homology results for the deduced amino acid sequence of the putative *bglG* (cac1406)

Results of a BLAST homology search with the deduced amino acid sequence of the putative ORF (with the locus tag cac1406), the putative antiterminator, BglG.

Species	Putative Gene	% Identity	% Similarity	Predicted molecular mass (Da)	ORF Locus tag	Accession
Clostridium beijerinckii NCIMB 8052	Transcriptional antiterminator, BglG	69	85	32337	Cbei_3274	YP_001310359
Clostridium longisporum	Transcriptional antiterminator	65	82	-	CLOABGL 4933 6.1	AAC05712
Clostridium butyricum 5521	Transcription antiterminator, LicT	60	77	32720	CBY1094	ZP_02948880
Bacillus thuringiensis serovar	Transcription antiterminator, BglG	45	67	32725	BT9727_08 76	YP_035217
Bacillus pumilus ATCC 7061	Transcription antiterminator, LicT	48	67	32614	BAT1009	ZP_03054861
Listeria monocytogene EGD-e	Transcription antiterminator, LicT	45	69	33003	lmo2436	NP_465959
Bacillus licheniformis ATCC 14580	Transcriptional antiterminator, LicT	47	66	32418	BL02652	YP_081273
Clostridium beijerinckii NCIMB 8052	Transcriptional antiterminator, BglG	46	67	32702	Cbei2834	YP_001309938
Bacillus subtilis	Transcription antiterminator, LicT	46	65	-	BSU39080	P39805
Clostridium botulinum B str. 17B	Transcription antiterminator, LicT	41	65	32494	CLL_A152 0	YP_001885714

```
acctgacttttcaaaaaaattaattcatattaagattggattgaatagttagtatgtat
aaaacaaaaaacgactaattacgatatttttaaaaaagttattgacaattaatataaaa
                                                -35
gggtataatattaagacatggttaaacgattacaaattaatatctgggattgttactgat
    -10                                          ●●●●●●●●●●●●●●
tcgatcaggcgagacccgagcttttgaaagtaatatttatattattttcattagttgggg
●●●●●●●●●●●●●●●●
ttttttatttttagatcaatgttgtaaatggggtaattctg
                              rbs?

  1 - ATGACGATAAAAAAAGATATTTAATAACAATGCAATAATAGCTGAAAATTCAGATAAACAT - 60
  1 - M  T  I  K  K  I  F  N  N  N  A  I  I  A  E  N  S  D  K  H   - 20

 61 - GAATTTGTTGTCATGGGGTGTGGAATTGCATTTAAGAAAAATATAGGAGAAAAAGTAGAT - 120
 21 - E  F  V  V  M  G  C  G  I  A  F  K  K  N  I  G  E  K  V  D   - 40

121 - GAGAGGCTAATAGAAAAAACTTTTATTCTTAAAGGAAAGGATGCATCCGAAAAGTTTAAA - 180
 41 - E  R  L  I  E  K  T  F  I  L  K  G  K  D  A  S  E  K  F  K   - 60

181 - ATGCTGCTGGAAGATGTTCCCACAGAGTATGTTTCGGTATGTTATGACATTATTGAATAT - 240
 61 - M  L  L  E  D  V  P  T  E  Y  V  S  V  C  Y  D  I  I  E  Y   - 80

241 - GCTAAAAATGTTTTAGATGCAAAATTAAATGACCATATATATGTTACTTTAACTGATCAT - 300
 81 - A  K  N  V  L  D  A  K  L  N  D  H  I  Y  V  T  L  T  D [H]  - 100

301 - GTTAGCAATTGTTTTAAAATGATTGAATCAGGAATTATTATACACAATCCATTAATTTGG - 360
101 - V  S  N  C  F  K  M  I  E  S  G  I  I  I  H  N  P  L  I  W   - 120

361 - GAAATCAAA[illegible]GTTTTATCCTAAGGAATTTAAAGTTGGACTTAAAGCTATAGACTTTATA - 420
121 - E  I  K  K  F  Y  P  K  E  F  K  V  G  L  K  A  I  D  F  I   - 140

421 - AAGGATGAATTAGGAAAAGACTTGCCAGAAGATGAGGCAGCTAATATAGCTCTACATTTG - 480
141 - K  D  E  L  G  K  D  L  P  E  D  E  A  A  N  I  A  L [H] L   - 160

481 - ATAAATGCACAAATTAACAATTCCTTAAATAATGTTGAAGATGCTGCAAAGCAAGCAAAA - 540
161 - I  N  A  Q  I  N  N  S  L  N  N  V  E  D  A  A  K  Q  A  K   - 180

541 - ATGATACAGGATATTTTAAATATAGTTAAGTATACTTATAATATAGTACTTGACGAAAAG - 600
181 - M  I  Q  D  I  L  N  I  V  K  Y  T  Y  N  I  V  L  D  E  K   - 200

601 - TCGATAAGTTATGAAAGGTTTGTCACACATTTAAGATTCTTCTTTCATAGAATTGATAAG - 660
201 - S  I  S  Y  E  R  F  V  T [H] L  R  F  F  F  H  R  I  D  K   - 220

661 - TCAGAAACAGTAGATACAGCAGAGGAAGATTTTTTATTAAAGCAAGTTAGAGAAAAATAT - 720
221 - S  E  T  V  D  T  A  E  E  D  F  L  L  K  Q  V  R  E  K  Y   - 240

721 - AAAGATGCATACCAATGTATGCTTAAGGTACAAAAGTATCTTGGAAAAGAATTGTCAGAT - 780
241 - K  D  A  Y  Q  C  M  L  K  V  Q  K  Y  L  G  K  E  L  S  D   - 260

781 - GAAGAAAAGCTTTATTTAACAGTTCACATTCAAAGGGTCTCAAAGAAAAATTAA       - 834
261 - E  E  K  L  Y  L  T  V [H] I  Q  R  V  S  K  K  N  *  X     - 280
```

Figure 4.8: Nucleotide and deduced amino acid sequence of the putative antiterminator

The deduced amino acid sequence is placed below the first nucleotide of the corresponding codon. Putative Shine-Dalgarno sequences for ribosomal binding are underlined and labelled. Stop codons are indicated by asterisks. Putative -35 and -10 promoter sequences are in bold text, underlined and labelled. A putative CRE sequence is shaded. Arrows indicate a palindromic sequence in the promoter region, immediately downstream from the proposed antiterminator binding site (marked with ●). Conserved histidine residues are in bold and highlighted.

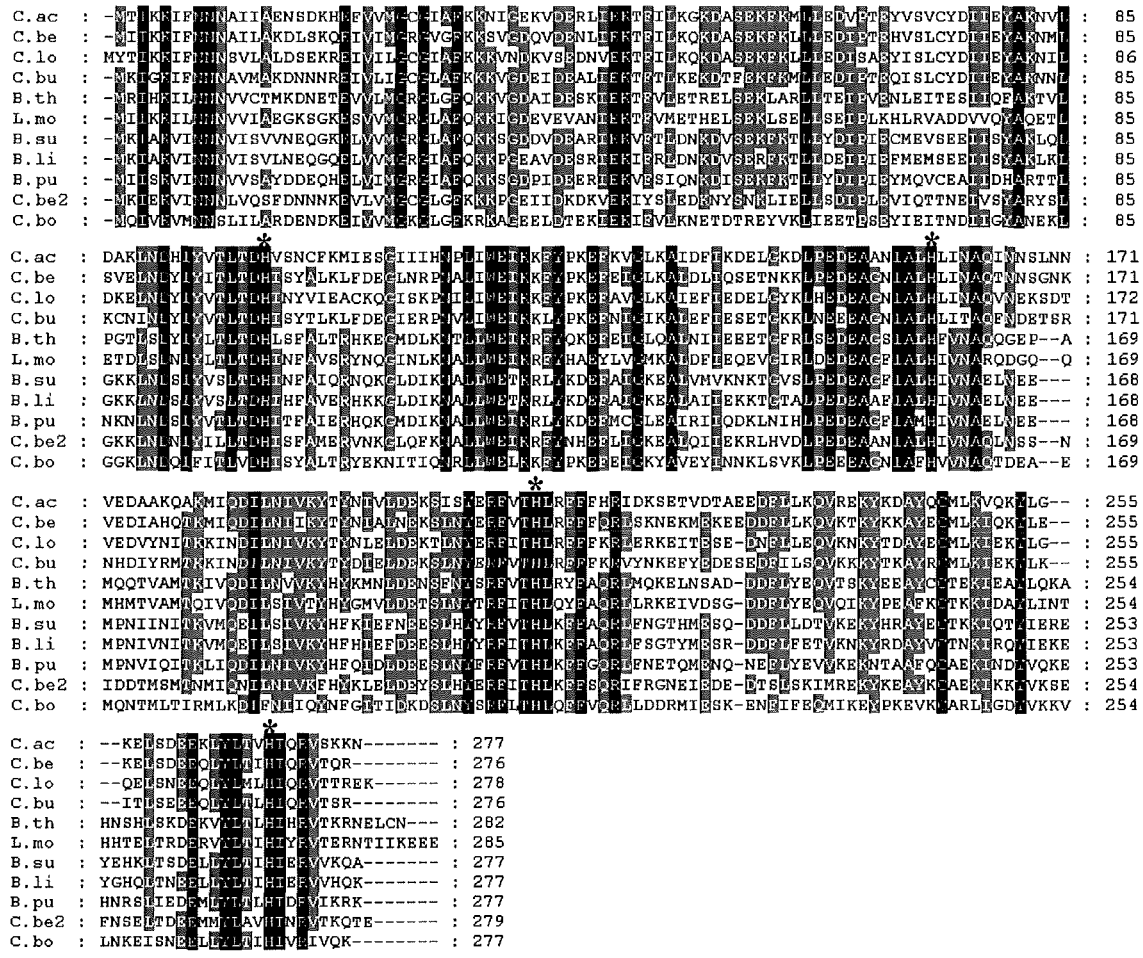


Figure 4.9: Multiple alignment of the *C. acetobutylicum* BglG sequence with other antiterminators

The deduced amino acid sequence of the putative *C. acetobutylicum* BglG protein is aligned with the antiterminator proteins with the highest percentage identities from the BLAST output. Dashes within the sequences indicate gaps giving optimal alignment. Identical residues are within black boxes and conserved residues (present in at least seven of the eleven sequences) are shaded. Conserved histidine residues, associated with the regulation of antiterminator proteins, are indicated by an asterix above.

Abbreviations; C.ac, *Clostridium acetobutylicum*; C.be, *Clostridium beijerinckii* NCIMB 8052; C.lo, *Clostridium longisporum*, C.bu, *Clostridium butyricum* 5521, B.th, *Bacillus thuringiensis* serovar, L.mo *Listeria monocytogenes* EGD-e, B.su, *Bacillus subtilis,* B.li, *Bacillus licheniformis* ATCC 14580; B.pu, *Bacillus pumilus* ATCC 7061; C.be2, *Clostridium beijerinckii;* NCIMB 8052, C.bo, *Clostridium botulinum* B str. 17B.

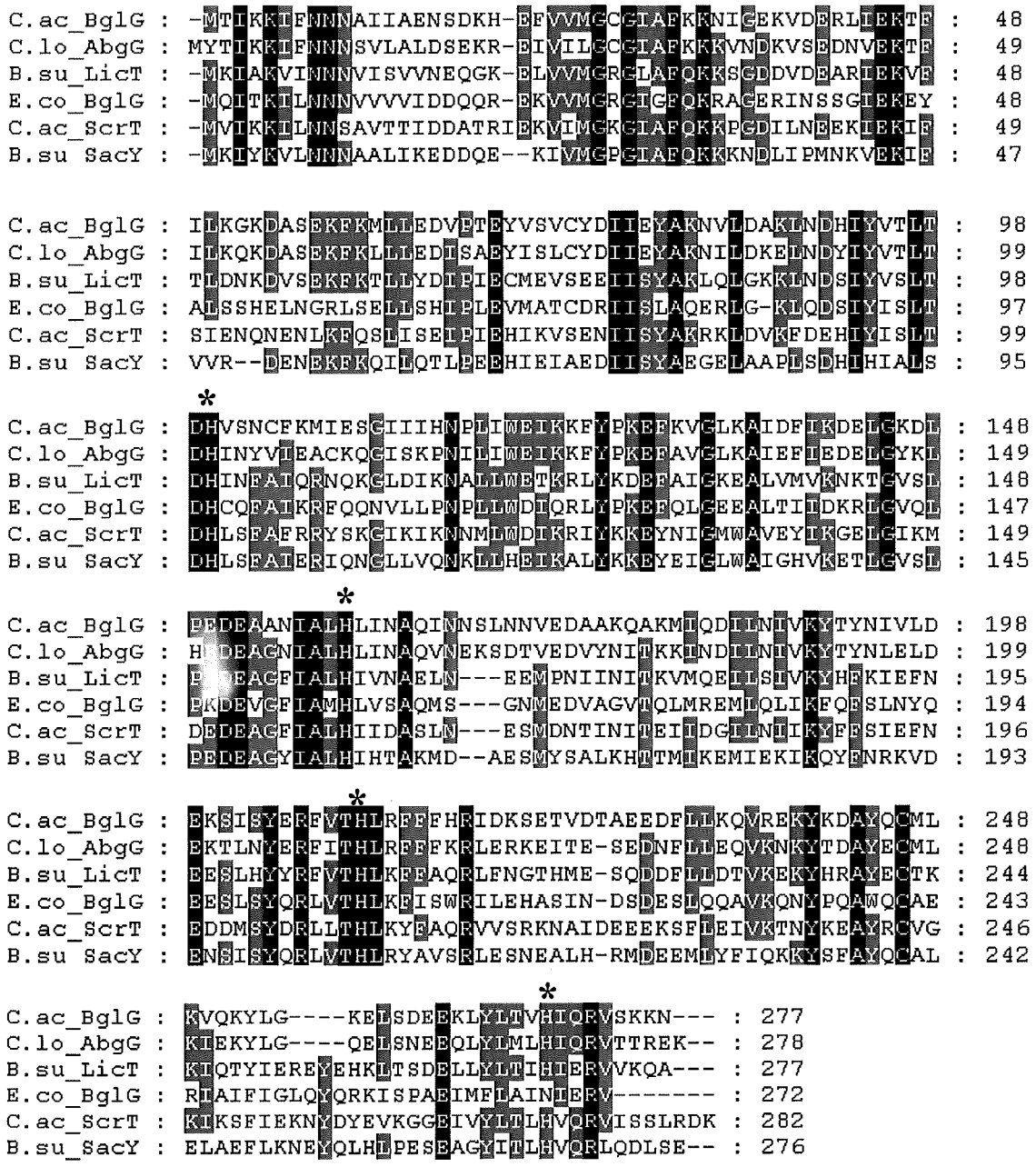


Figure 4.10: Sequence alignment of protein sequences of the BglG family

Comparison of the deduced amino acid sequence of the *C. acetobutylicum* putative BglG protein with the regulatory proteins of the BglG family. Dashes within the sequences indicate gaps giving optimal alignment. Identical residues are within black boxes and conserved residues (present in at least four of the six sequences) are shaded. Conserved histidine residues, associated with the regulation of antiterminator proteins, are indicated by an asterix above.

Abbreviations: C. lo_AbgG, *C. longisporum* AbgG (Brown & Thomson, 1998); B.su_LicT, *Bacillus subtilis* LicT (Schnetz *et al.*, 1996); E.co_BglG, *Escherichia coli* BglG (Schnetz *et al.*, 1987); C.ac_ScrT, *C. acetobutylicum* ScrT (Tangney & Mitchell, 2000); B.su_SacY, *B. subtilis* SacT (Debarbouillé *et al.*, 1990); C.ac_BglG, *C. acetobutylicum bglG* gene product.

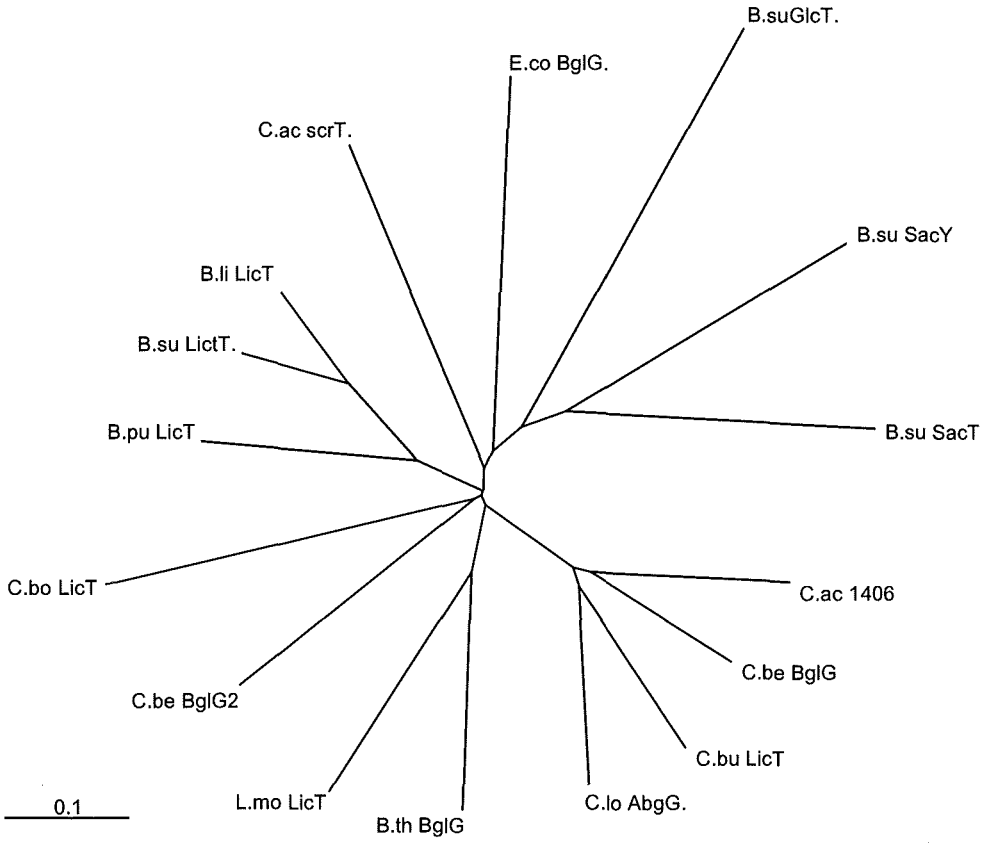


Figure 4.11: Unrooted phylogenetic tree for the putative antiterminator, BglG.

The amino acid sequences of the putative antiterminator were compared with selected Gram positive and Gram-negative sequences returned from BLAST search and analysed using ClustalW and TreeView. The first letter of the genus and the first two letters of the species are listed, followed by the protein. The full designations are given below. Sequences discussed in other papers are marked with a full stop after the protein name and the reference is provided in parentheses.

Abbreviations: C.ac_1406, *C. acetobutylicum* cac1406; C.be_BglG, *Clostridium beijerinckii* NCIMB 8052; C.bu_LicT, *Clostridium butyricum* 5521; B.th_BglG, *Bacillus thuringiensis serovar konkukian str. 97-27*; B.pu_LicT, *Bacillus pumilus* ATCC 7061; L.mo_LicT, *Listeria monocytogenes* EGD-e; C.be_BglG2, *Clostridium beijerinckii* NCIMB 8052; B.su_LicT, *Bacillus subtilis*; B.li_LicT, *Bacillus licheniformis* ATCC 14580; C.bo_LicT, *Clostridium botulinum B str. Eklund 17B*; C. lo_AbgG, *C. longisporum* AbgG (Brown & Thomson, 1998); B.su_LicT, *Bacillus subtilis* LicT (Schnetz *et al.*, 1996); E.co_BglG, *Escherichia coli* BglG (Schnetz *et al.*, 1987); C.ac_ScrT, *C. acetobutylicum* ScrT (Tangney & Mitchell, 2000); B.su_ SacY, *B. subtilis* SacT (Debarbouillé *et al.*, 1990).

4.3.3 Identification of the putative (PTS EII) *bglA* gene

The second ORF in the *bgl* operon (cac1407) is 1887bp in length and encodes a 628 amino acid protein, BglA which has a calculated molecular mass of 67,330Da. The deduced protein sequence of the putative BglA was used to carry out a BLAST homology search; the results are presented in Table 4.3. The outcome reveales that the sequence bears considerable identity to a PTS β-glucoside specific enzyme IIBCA component produced by *Clostridium beijerinckii* NCIMB 8052 (68% identity and 83% similarity) and to the arbutin and salicin specific EII found in *C. longisporum* (63% identity and 79% similarity).

The nucleotide and deduced amino acid sequence of the putative EII is presented in Figure 4.12. The *bglA* gene is preceded, at a distance of 7 nucleotides, by a putative ribosome binding site (rbs) AGGAGA which is labelled and underlined.

A consensus sequence for disaccharide binding has previously been proposed: [Flm][Gn][Iv][Tsn]**EP**[Aiv][ILmv][Fy]**G**[Vilma][Npt][Li] the highly conserved residues are in boldface type, the most abundant residue in uppercase, and lowercase residues in order of frequency (Lai *et al.*, 1997). The clostridial sequence (FGVTEPAIYGVTL) is in good agreement with this consensus. The domain signature sequences for EIIA (GTEILIHVGMDTV) and EIIB (NVNSLTHCITRLRFKLKD) are underlined and labelled in Figure 4.12, along with the conserved EIIC domain sequence for disaccharide binding.

Table 4.3: BLAST homology results for the deduced amino acid sequence of cac1407

Results of a BLAST homology search with the deduced amino acid sequence of the putative ORF with the locus tag cac1407.

Species	Putative Gene	% Identity	% Positive	Predicted molecular mass (Da)	ORF Locus tag	Accession
Clostridium beijerinckii NCIMB 8052	PTS β-glucoside specific EIIABC	68	83	67785	Cbei3273	YP_001310358
Clostridium longisporum	PTS enzyme II	63	79	-	CLOABG	AAC05713
Clostridium butyricum 5521	PTS β-glucoside specific EIIBCA	59	76	66399	CBY_1093	ZP_02948899
Listeria innocua Clip11262	PTS β-glucoside specific EIIABC	58	79	67146	lin0026	NP_469373
Listeria welshimeri serovar 6b	PTS β-glucoside specific EIIABC	58	79	67246	lwe0030	YP_848233
Listeria monocytogenes str. 4b F2365	PTS β-glucoside specific EIIBC	58	78	67297	LMOf23650030	YP_848233
Carnobacterim sp. AT7	PTS β-glucoside specific EIIABC	55	76	67707	CAT705169	NP_463560
Bacillus cereus E33L	PTS β-glucoside specific EIIABC	53	72	67139	BCZK0859	YP_082463
Bacillus halodurans C-125	PTS β-glucoside specific EIIABC	57	74	68307	BH0595	NP_241461
Bacillus thuringiensis serovar	PTS β-glucoside specific EIIIIABC	53	72	67124	BT9727_0877	YP_035218

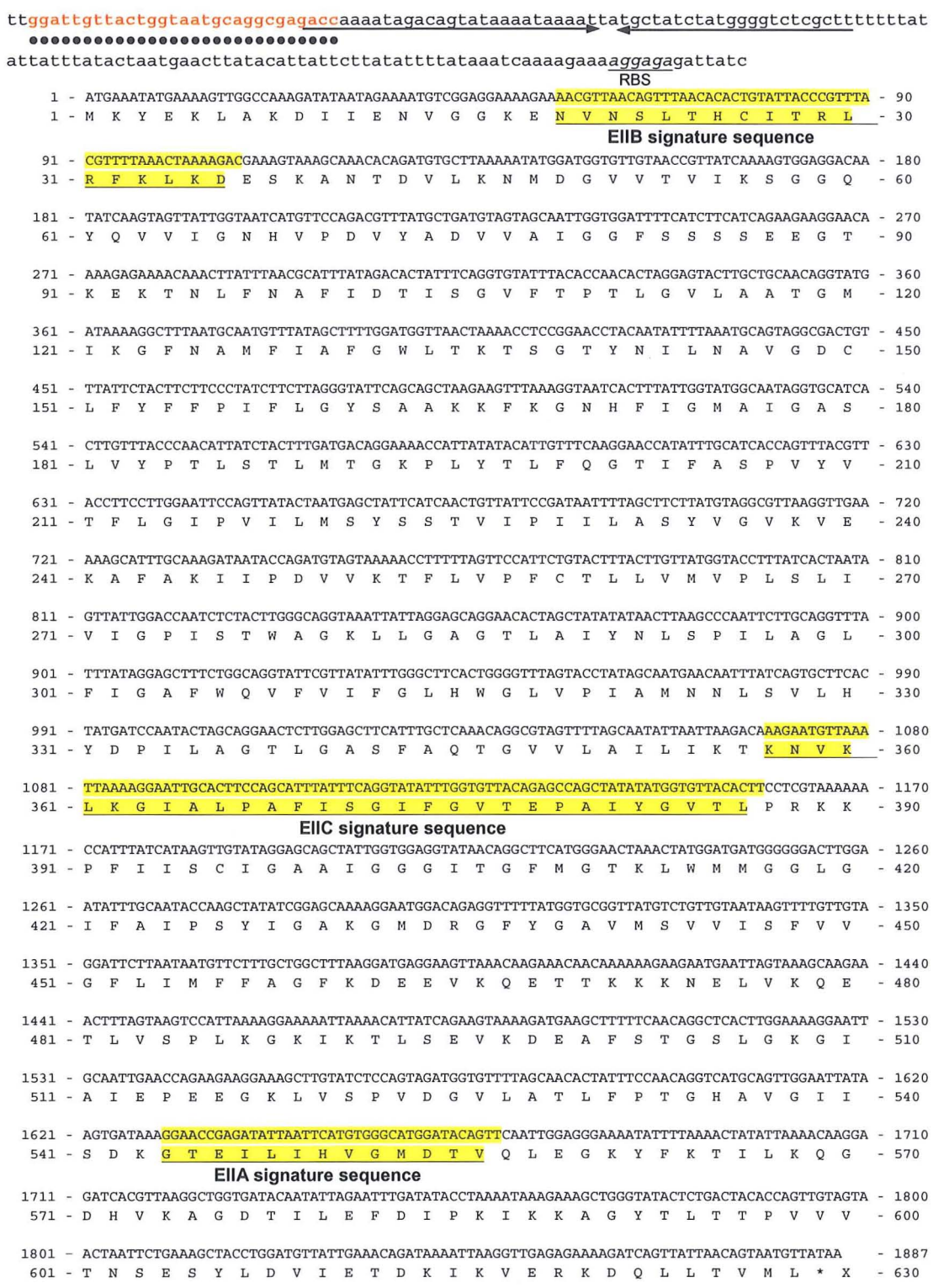


Figure 4.12: Nucleotide and deduced amino acid sequence of the putative EII

The deduced amino acid sequence is placed below the first nucleotide of the corresponding codon. Putative Shine-Dalgarno sequences for ribosomal binding (RBS) are underlined and labelled. Stop codons are indicated by asterisks. Arrows indicate a palindromic sequence, immediately downstream from the proposed antiterminator binding site (marked with a closed circle beneath). The signature sequences for EIIA and EIIB are underlined and labelled. A conserved region for disaccharide EIIC is also underlined and labelled.

The signature sequences from the well characterised *Klebsiella oxytoca* disaccharide transporting EII (Lai *et al.*, 1997) was aligned with the clostridial EII permease domains for comparison and are presented in Figure 4.13.

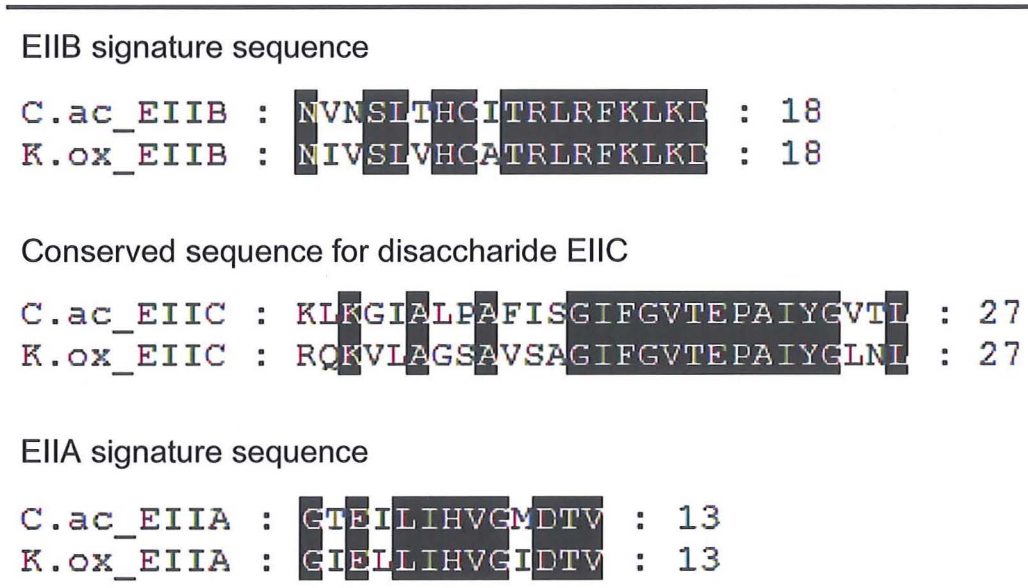


Figure 4.13: Enzyme II signature sequences of *C. acetobutylicum* and *Klebsiella oxytoca*

K. oxytoca signature sequences from the individual EII domains aligned with the clostridial EII

Enzyme II permeases consist of several domains, which can either be fused as a single protein or present as individual polypeptides. There is a hydrophilic EIIA domain, a hydrophilic EIIB domain and a hydrophobic EIIC domain which usually possesses several transmembrane helices (Postma *et al.*, 1993). To investigate the nature of the putative *bglA* product, the deduced amino acid sequence of the putative EII was processed in a bioinformatics tool (TMHMM), which uses a Hidden Markov Model to predict transmembrane helices within protein sequences (Krogh *et al.*, 2001; Sonnhammer *et al.*, 1998). The results of this analysis presented in Figure 4.14A revealed that the putative EII is predicted to contain a membrane spanning domain, containing 11 transmembrane helices, flanked by two hydrophilic regions. By comparison, the protein sequence of the *C. longisporum* arbutin and salicin specific EII was found to contain 10 transmembrane helices (Figure 4.14B).

A protein hydrophobicity plot can be used to determine the overall hydropathicity profile of a protein (Kyte & Doolittle, 1982). The deduced amino acid sequence of the clostridial EII was analysed and the results presented in Figure 4.15.

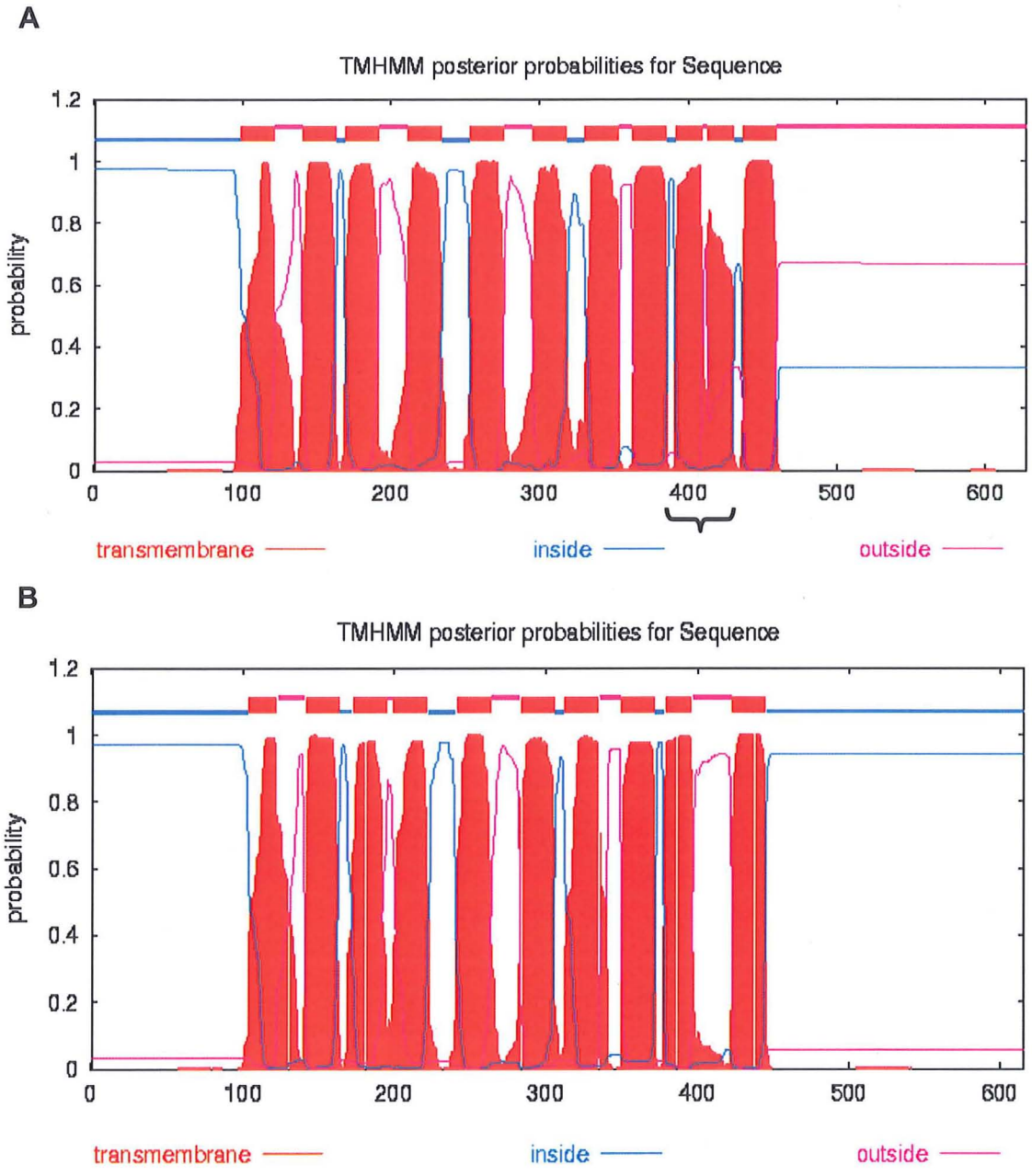


Figure 4.14: Transmembrane helices:hidden Markov Model (TMHMM)

The amino acid sequences of the putative β-glucoside EII from *C. acetobutylicum* (A) and the arbutin and salicin specific EII from *C. longisporum* (B) were input into TMHMM to identify the position and number of transmembrane helices present within the protein.

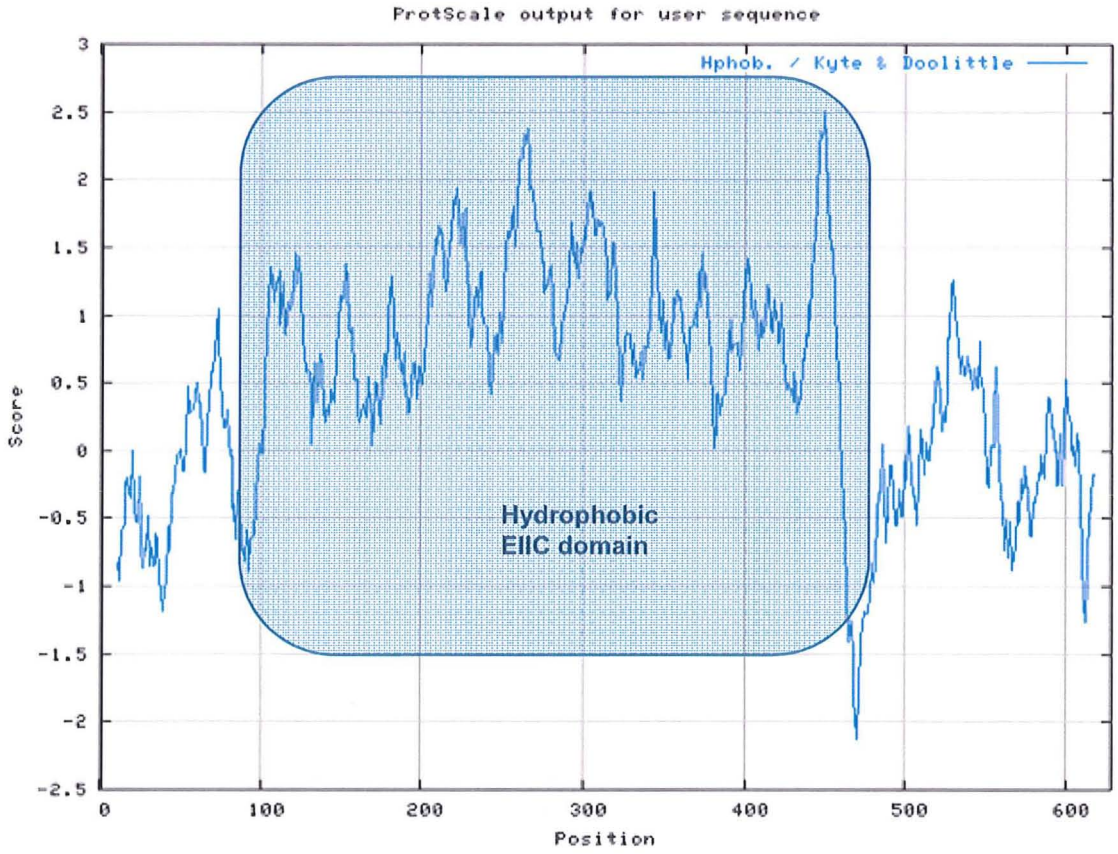


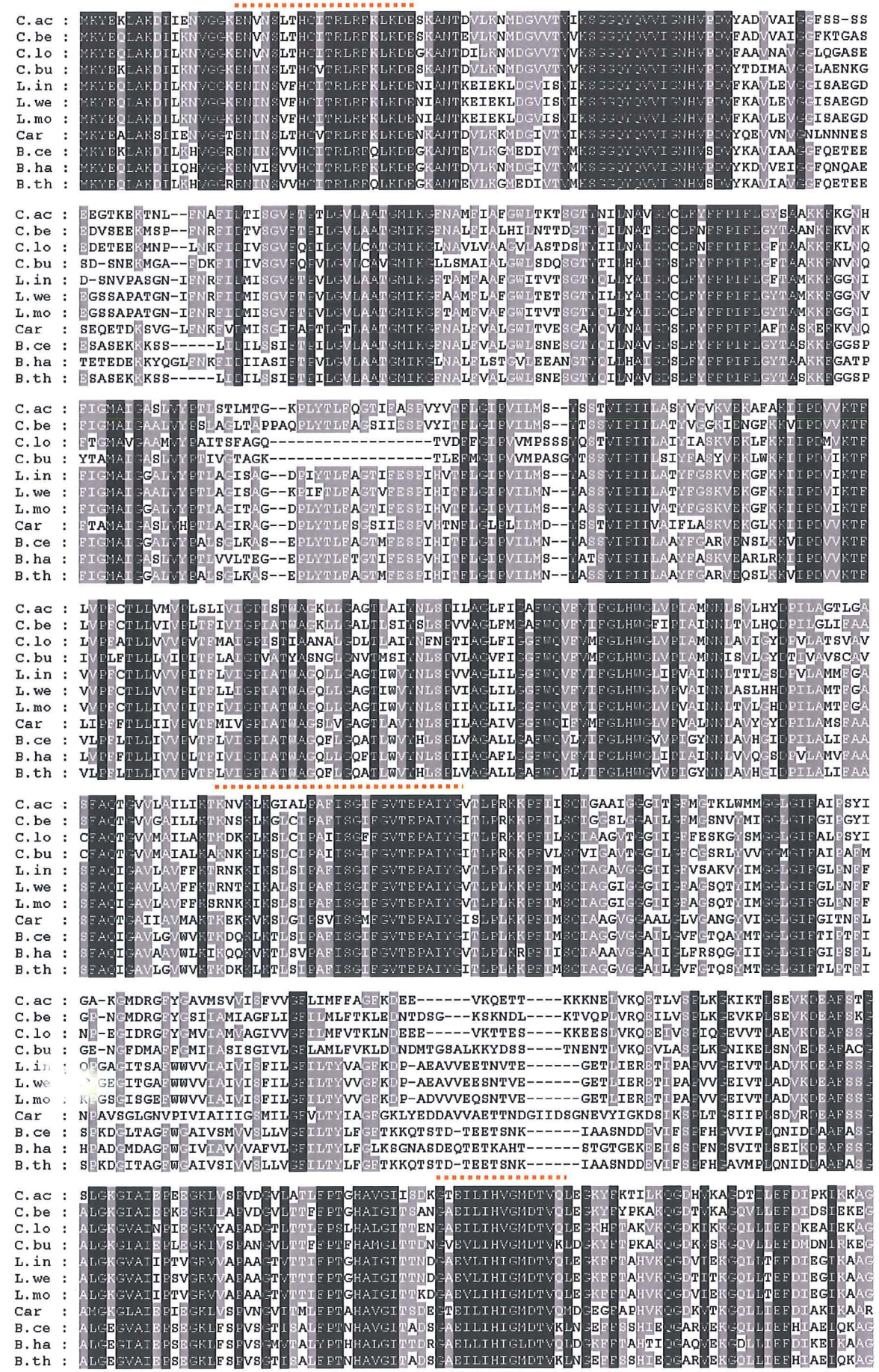
Figure 4.15: Protein hydrophobicity plot of the putative EII

Kyte Doolittle ExPASy proteomics server was used to display the overall hydropathicity profile of the deduced amino acid sequence of cac1407, the putative β-glucoside specific EII. Window size = 21. Window size refers to the number of amino acids examined at a time to determine a point of hydrophobic character, 21 is optimum for identifying hydrophobic regions within proteins (Kyte & Doolittle, 1982). The central hydrophobic EIIC domain is shaded in blue.

A multiple alignment of the putative EII proteins, presented in Figure 4.16, reveal large consecutive regions of conserved amino acids throughout the entire length of the sequence. The EIIBCA signature sequences are indicated by dashed orange lines in this figure.

The phylogenetic tree presented in Figure 4.17 clearly shows that the four *Clostridium* spp. possess EIIs with high sequence diversity, unlike the *Listeria* spp. which possess highly conserved EII sequences falling together on a deep rooted monophyletic branch.

To investigate if the putative *bglA* gene is phylogenetically related to any of the published, experimentally verified β-glucoside specific EIIs, a phylogram of the amino acid sequences of the EII's was constructed and the results are presented in Figure 4.18. The sugar specificity of each is noted next to the gene name. A sister relationship is seen between the putative EII (cac1407) and the arbutin and salicin specific EII from the related *C. longisporum* which fall together in a clade, in the most derived position in the phylogeny (Figure 4.18).



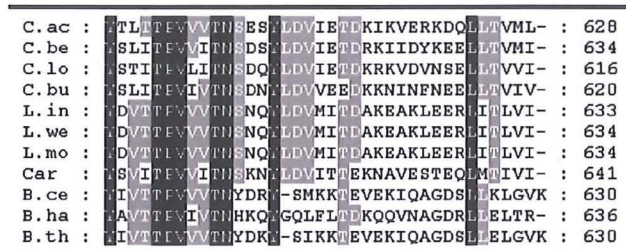


Figure 4.16: Multiple alignment of the *C. acetobutylicum* BglA sequence with other EII components

The deduced amino acid sequence of the *C. acetobutylicum* BglA is aligned with the enzyme II components with the highest percentage identities from the BLAST output. Dashes within the sequences indicate gaps giving optimal alignment. Identical residues are within black boxes and conserved residues (present in at least seven of the eleven sequences) are shaded.

Positions of signature sequences are indicated by (·····) above. Abbreviations; C.ac - *Clostridium acetobutylicum* ATCC 824; C.be - *Clostridium beijerinckii* NCIMB 8052; C.lo - *Clostridium longisporum;* C.bu - *Clostridium butyricum* 5521; L.in - *Listeria innocua* Clip11262; L.we - *Listeria welshimeri* serovar 6b str. SLCC5334; L.mo - *Listeria monocytogenes* str. 4b F2365; C.ar - *Carnobacterium* sp. AT7; B.ce - *Bacillus cereus* E33L; B.ha - *Bacillus halodurans* C-125; B.th - *Bacillus thuringiensis* serovar konkukian str. 97-27

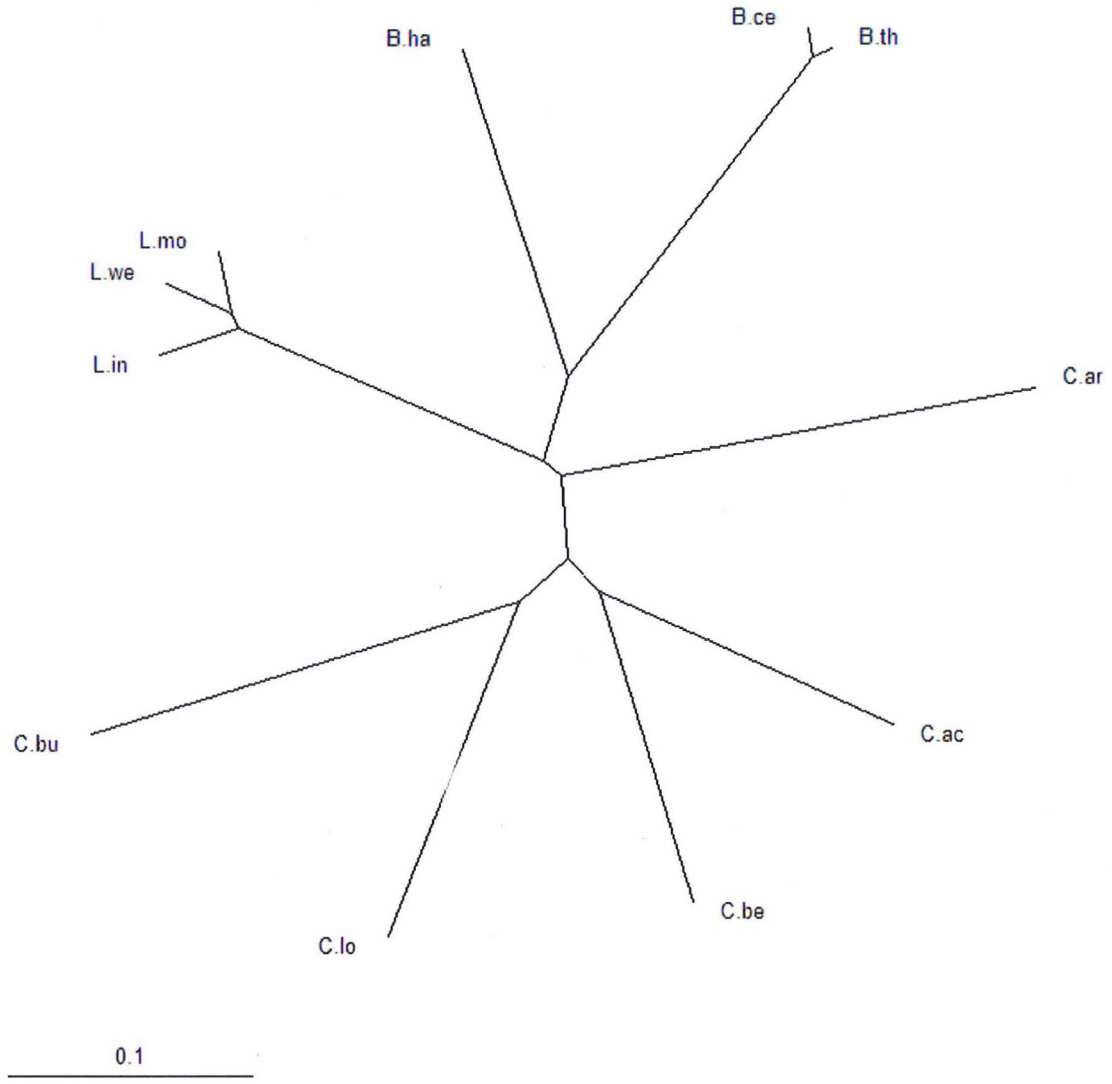


Figure 4.17: Unrooted phylogenetic tree for the putative enzyme II, BglA.

The amino acid sequences of the putative enzyme II were compared with selected Gram positive sequences returned from BLAST search and analysed using ClustalW and TreeView.

C.ac - *Clostridium acetobutylicum* ATCC 824; C.be - *Clostridium beijerinckii* NCIMB 8052; C.lo - *Clostridium longisporum;* C.bu - *Clostridium butyricum* 5521; L.in - *Listeria innocua* Clip11262; L.we - *Listeria welshimeri* serovar 6b str. SLCC5334; L.mo - *Listeria monocytogenes* str. 4b F2365; C.ar - *Carnobacterium* sp. AT7; B.ce - *Bacillus cereus* E33L; B.ha - *Bacillus halodurans* C-125; B.th - *Bacillus thuringiensis* serovar konkukian str. 97-27

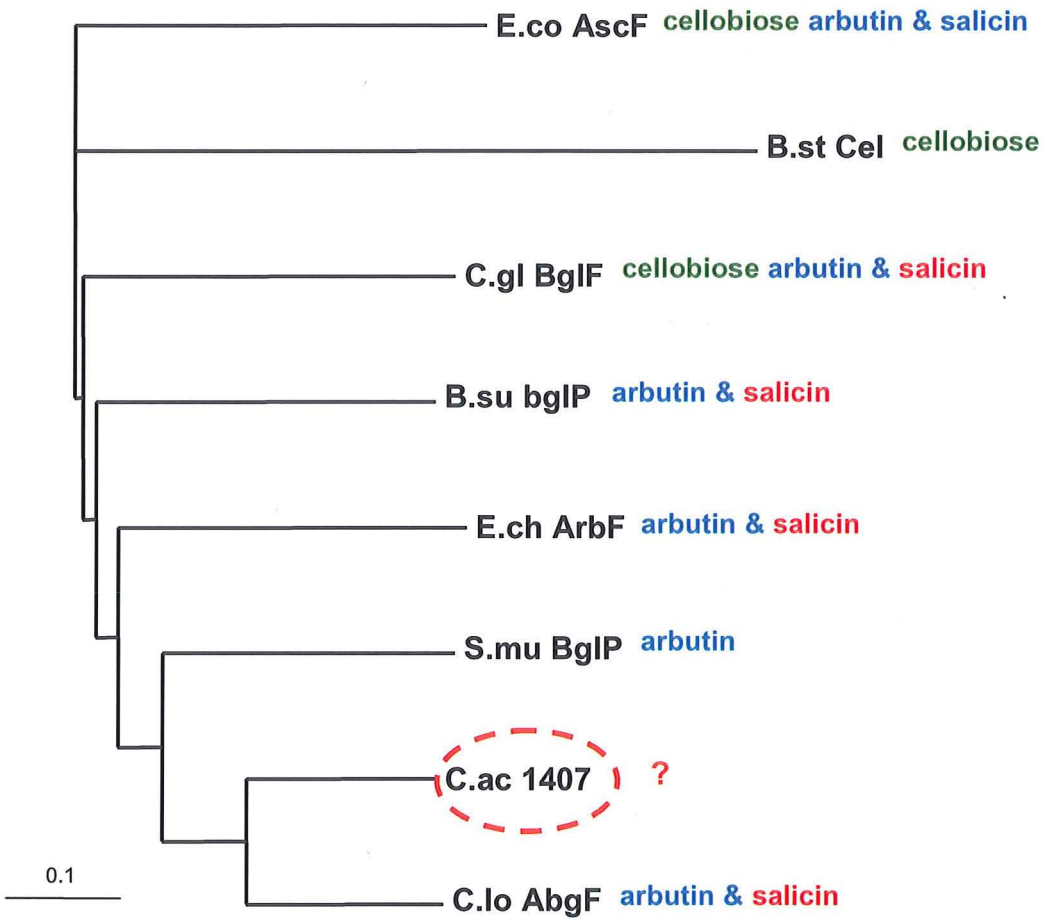


Figure 4.18: Phylogram of experimentally verified β-glucoside specific EII's

The first letter of the genus and the first two letters of the species, identify the organism, followed by the protein. The β-glucosides which are transported are listed adjacent to the protein name. Full designations and accessions are given below.

E.co - *Escherichia coli* [AscF] (P24241) Hall & Xu, 1992; B.st - *Bacillus stearothermophilus* [Cel] Lai & Ingram, 1993; C.gl - *Corynebacterium glutamicum* [BglF] (Q8GGK3) Kotrba *et al.*, 2003; B.su - *Bacillus subtilis* [BglP] (P40739) Le Coq *et al.*, 1995; E.ch - *Erwinia chrysanthemi* [ArbF] (P26207) Hassouni, *et al.*, 1992; S.mu - *Streptococcus mutans* [BglP] (Q9KJ80) Cote *et al.* 2000; C.ac - *Clostridium acetobutylicum* ATCC 824; C.lo - *Clostridium longisporum* [AbgF] (AAC05713) Brown & Thomson, 1998.

4.3.4 Identification of the putative (phospho-β-glucosidase) *bglB* gene

The third ORF in the *bgl* operon (cac1408) is 1434bp in length. Seven nucleotides upstream of the putative start codon there is a putative rbs (AGGAGAG). The putative 477 amino acid protein, *bglB,* has a calculated molecular mass of 54,629Da. The nucleotide and deduced amino acid sequence of the putative phospho-β-glucosidase, along with the upstream region are presented in Figure 4.19.

Previously, the phospho-β-glucosidases and β-glucosidases from a range of organisms were studied and a highly conserved family 1 hydrolase signature sequence (LFIVENGLG) was identified (Lai *et al.*, 1997). The putative phospho-β-glucosidase encoded by cac1408 (*bglB*) has a signature sequence which matches this 100%. The position of the signature sequence is marked in Figure 4.19.

The sequence shows significant identity to a *C. beijerinckii* NCIMB 8052 protein, in this case a glycoside hydrolase family protein, with 85% identity and 94% similarity. However, it also has close similarity to a 6-phospho-β-glucosidase from *C. difficile* 630 (80% identity and 87% similarity; Table 4.4). To explore the distribution of sequence similarity between the BLAST output results, a multiple alignment was carried out. The results of the multiple alignment showed extensive areas of identity (Figure 4.20). The hydrolase signature sequence is 100% conserved in all of the clostridial species in the alignment, apart from in *C. beijerinckii*, where there is one residue difference (an isoleucine substituted for a valine residue). The ClustalW2 output was also analysed in TreeView and the results are presented in Figure 4.21. The *C. acetobutylicum* phospho-β-glucosidase encoding sequence shares a branch with the related *C. beijerinckii* phospho-β-glucosidase encoding sequence.

```
                                                                    RBS
      ataacttttactttgagcataaggttgacttgtgctcaaagttttataaggagagatgttta
    1 - ATGTCAAAAGGTTTTAGTAAAGAATTTTTATGGGGAGGCGCTACGGCGGCTAATCAGTGTGAAGGTGCTTATTTAGAGGATAATAAAGGGTT - 92
    1 - M  S  K  G  F  S  K  E  F  L  W  G  G  A  T  A  A  N  Q  C  E  G  A  Y  L  E  D  N  K  G    - 30

   93 - ATCTACAGTTGATGTAATTCCAAAGGGAAAGGACCGTTTTCCCGTAGGCTTAGGAAAAATGAAAATGCTTGAGTGTGACAGTGAACATTATT - 184
   31 - L  S  T  V  D  V  I  P  K  G  K  D  R  F  P  V  G  L  G  K  M  K  M  L  E  C  D  S  E  H    - 60

  185 - ATCCAAGTCATGAGGCTATTGATTTTTATCACAGATATAAAGAAGATATAGCTCTATTTCATGAAATGGGCTTTAAATGTTTTAGATTATCT - 276
   61 - Y  P  S  H  E  A  I  D  F  Y  H  R  Y  K  E  D  I  A  L  F  H  E  M  G  F  K  C  F  R  L    - 90

  277 - CTTGCTTGGTCACGAATATTTCCAAATGGGGACGATGAAATACCAAATGAGGAGGGATTGAAATTTTACGATGCGGTATTTGATGAATGTTT - 368
   91 - L  A  W  S  R  I  F  P  N  G  D  D  E  I  P  N  E  E  G  L  K  F  Y  D  A  V  F  D  E  C    - 120

  369 - AAAATACGGAATAGAACCATTAGTTACTATAACTCATTTTGATGTTCCGGTGAATTTAGTGAAAACAGTAGGTTCCTGGAGAAGCAGTAAAA - 460
  121 - L  K  Y  G  I  E  P  L  V  T  I  T  H  F  D  V  P  V  N  L  V  K  T  V  G  S  W  R  S  S    - 150

  461 - TGGTAGAGTATTACGAAAAACTGTGTAAGGTTATATTTAGTCGTTATAAAAATAAGGTTAAGTACTGGCTTACCTTTAATGAAATAAATATG - 552
  151 - M  V  E  Y  Y  E  K  L  C  K  V  I  F  S  R  Y  K  N  K  V  K  Y  W  L  T  F  N  E  I  N    - 180

  553 - CTATTGCATTTACCTTTTATAGGTGCAGGTTTAGTTTTTGAAGAGGGTGAAAATGAGGAAGCTATAAAATATCAAGCAGCACATCATCAGCT - 644
  181 - L  L  H  L  P  F  I  G  A  G  L  V  F  E  E  G  E  N  E  E  A  I  K  Y  Q  A  A  H  H  Q    - 210

  645 - TATTGCAAGTGCAAAGGCAACAAAAATAGCTCGTGAAATAAATCCAAATTTTAAAATAGGGTGTATGCTAGCTGCTGGAAATAATTATGCAA - 736
  211 - L  I  A  S  A  K  A  T  K  I  A  R  E  I  N  P  N  F  K  I  G  C  M  L  A  A  G  N  N  Y    - 240

  737 - ACACATGTGCACCAGAAGATGTGTGGAGATCTATAGAAAAGGATAGAGAAAACTATTTCTTTATTGATGTACAATCTCGTGGAGAGTATCCA - 828
  241 - N  T  C  A  P  E  D  V  W  R  S  I  E  K  D  R  E  N  Y  F  F  I  D  V  Q  S  R  G  E  Y    - 270

  829 - AACTATGCTAAAAAGATGCTTAAAAGGAAAGGTATAGAGCTTCAAGTAGAGGATGAAGATGCTGGAATTTTAAAGAATAATACTGTTGATTT - 920
  271 - N  Y  A  K  K  M  L  K  R  K  G  I  E  L  Q  V  E  D  E  D  A  G  I  L  K  N  N  T  V  D    - 300

  921 - TATATCATTTAGTTACTATTCTTCACGCCTAACTAGTGCCAATCCTAATATAAATGGAAATACTGAAGGAAATGTCTTTGCTACCTTGAAAA - 1012
  301 - F  I  S  F  S  Y  Y  S  S  R  L  T  S  A  N  P  N  I  N  G  N  T  E  G  N  V  F  A  T  L    - 330

 1013 - ATCCTTATCTAAAAGCCAGTGAATGGGGATGGCAAATAGATCCGCTTGGACTTCGAATTACATTAAATTCTTTATACGATCGTTATCAAAAA - 1104
  331 - N  P  Y  L  K  A  S  E  W  G  W  Q  I  D  P  L  G  L  R  I  T  L  N  S  L  Y  D  R  Y  Q    - 360

 1105 - CCATTATTTATAGTTGAAAATGGTTTGGGAGCCGTTGATACTCCAGATGAAAATGGATATGTAGAAGATGACTATCGAATTGAGTATTTAAG - 1196
  361 - P  L  F  I  V  E  N  G  L  G  A  V  D  T  P  D  E  N  G  Y  V  E  D  D  Y  R  I  E  Y  L    - 390
        Family 1 hydolase signature sequence

 1197 - AGAGCACATTAAGGCTATGAGAGATGCAGTTAATGTTGATGGTGTAGAACTTATGGGATACACACCTTGGGGATGTATAGATTTAGTAAGTG - 1288
  391 - R  E  H  I  K  A  M  R  D  A  V  N  V  D  G  V  E  L  M  G  Y  T  P  W  G  C  I  D  L  V    - 420

 1289 - CTTCAACTGGAGAAATGAAAAAGAGATATGGCTTTATATATGTTGATAAGGATAATGATGGGAATGGAACATTAAAGCGTTCTAAGAAAAAG - 1380
  421 - A  S  T  G  E  M  K  K  R  Y  G  F  I  Y  V  D  K  D  N  D  G  N  G  T  L  K  R  S  K  K    - 450

 1381 - AGTTTTTATTGGTATAAGAAGGTAATAGAATCAAATGGTATGGATATAGAGTAA                                        - 1434
  451 - S  F  Y  W  Y  K  K  V  I  E  S  N  G  M  D  I  E  *  X                                       - 480
```

Figure 4.19: Nucleotide and deduced amino acid sequence of the putative p-β-glucosidase

The deduced amino acid sequence is placed below the first nucleotide of the corresponding codon. Putative sequences for ribosomal binding (RBS) are underlined and labelled. Stop codons are indicated by asterisks. The signature sequence for family 1 hydrolases was also found, 100% conserved in the putative β-glucosidase sequence.

Table 4.4: BLAST homology results for the deduced amino acid sequence of cac1408.

Results of a BLAST homology search with the deduced amino acid sequence of the putative ORF with the locus tag cac1408.

Species	Putative Gene	% Identity	% Positive	Predicted molecular mass (Da)	ORF Locus tag	Accession
Clostridium beijerinckii NCIMB 8052	Glycoside hydrolase family protein	85	94	55446	Cbei_3271	YP_001310356
Clostridium difficile 630	6-phospho-β-glucosidase	80	87	54564	CD3096	YP_001089610
Clostridium difficile QCD-63q42	6-phospho-β-glucosidase	79	87	54591	CdifQCD6_010100007374	ZP_02745620
Streptococcus gordonii str.	6-phospho-β glucosidase	72	86	54845	SGO0297	YP_001449616
Enterococcus faecalis V583	6-phospho-β-glucosidase	74	87	54319	EF0271	NP_814064
Streptococcus sanguinis SK36	6-phospho-β-glucosidase	71	86	54817	SSA0395	YP_001034397
Streptococcus agalactiae COH1	6-phospho-β-glucosidase	71	84	55124	SAN1223	ZP_00785422
Streptococcus agalactiae A909	6-phospho-β-glucosidase	71	85	55146	SAK1188	YP_329809
Clostridium butyricum 5521	6-phospho-β-glucosidase AscB	75	85	54230	CBY4097	ZP_02947929
Clostridium botulinum B1 str. Okra	Glycosyl hydrolase family 1	69	84	54233	CLD3684	YP_001780509

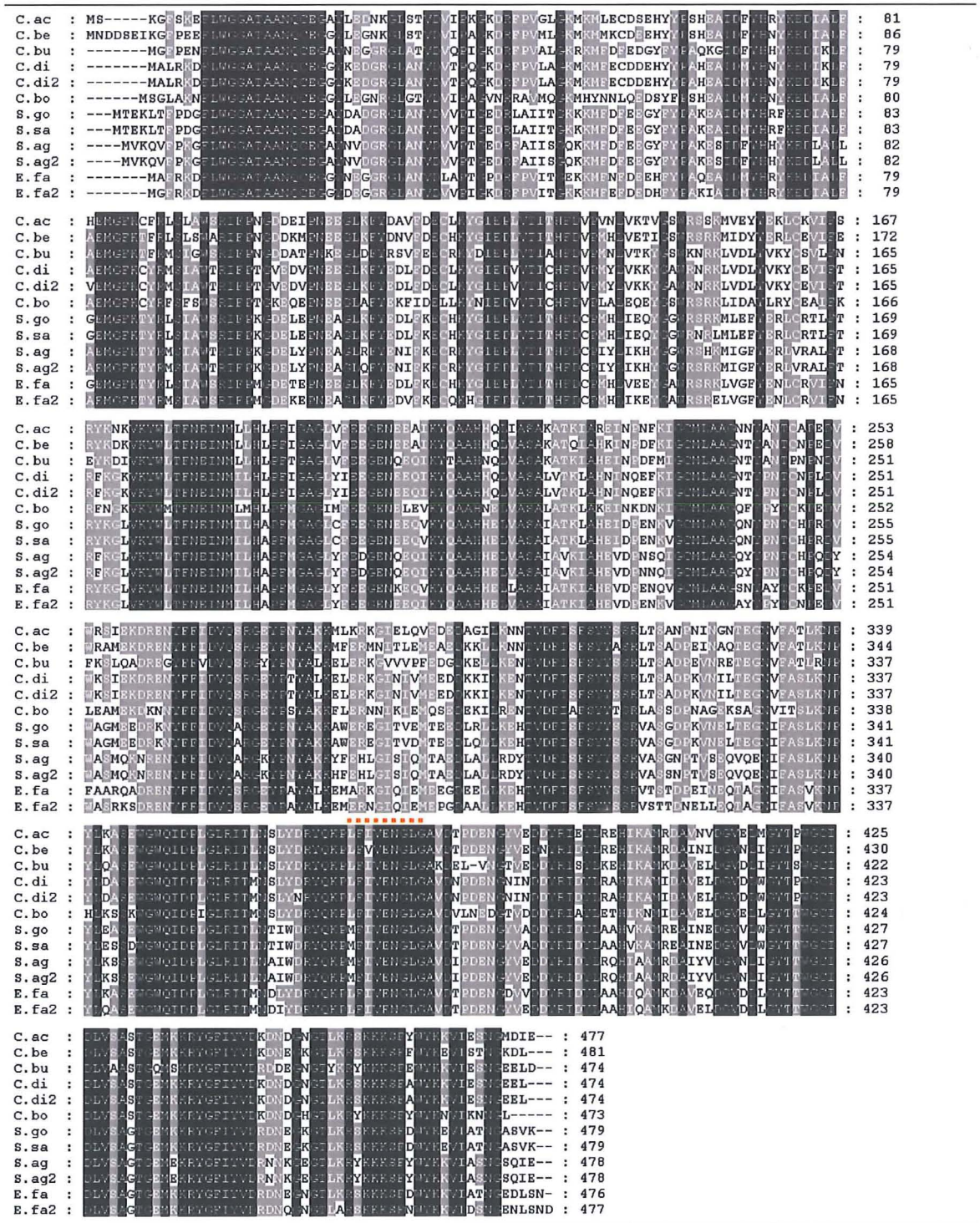


Figure 4.20: Alignment of the *C. acetobutylicum* BglG sequence with other phospho-β-glucosidases

The deduced amino acid sequence of the *C. acetobutylicum bglG* is aligned with the phospho-β-glucosidases with the highest percentage identities from the BLAST output. Dashes within the sequences indicate gaps giving optimal alignment. Identical residues are within black boxes and conserved residues (present in at least seven of the eleven sequences) are shaded. Position of the hydrolase signature sequence is indicated by (▪▪▪▪▪) above

Abbreviations; C.ac - *Clostridium acetobutylicum* ATCC 824; C.be - *Clostridium beijerinckii*; NCIMB 8052; C.bu *Clostridium butyricum* 5521; C.di - *Clostridium difficile* 630; C.di2 - *Clostridium difficile* QCD-63q42; C.bo - *Clostridium botulinum* B1 str. Okra; S.go - *Streptococcus gordonii* str. Challis substr. CH1; S.sa - *Streptococcus sanguinis* SK36; S.ag - *Streptococcus agalactiae COH1*; S.ag2 - *Streptococcus agalactiae* A909; E.fa - *Enterococcus faecalis* V583; E.fa2 - *Enterococcus faecium* DO.

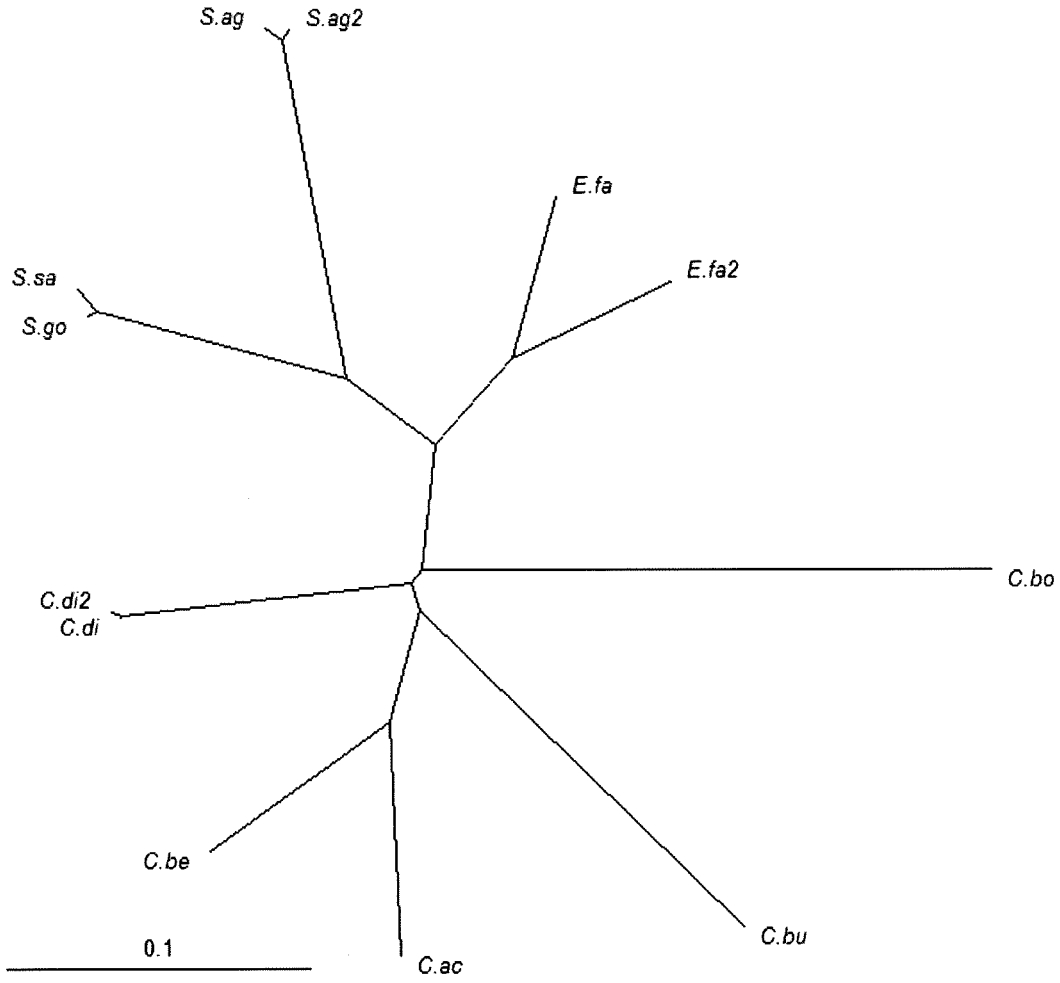


Figure 4.21: Unrooted phylogenetic tree for the putative phospho-β-glucosidase, BglB.

The amino acid sequences of the putative phospho-β-glucosidase were compared with selected Gram positive and Gram-negative sequences returned from BLAST search and analysed using ClustalW and TreeView.

Abbreviations; C.ac - *Clostridium acetobutylicum* ATCC 824; C.be - *Clostridium beijerinckii*; NCIMB 8052; C.bu *Clostridium butyricum* 5521; C.di - *Clostridium difficile* 630; C.di2 - *Clostridium difficile* QCD-63q42; C.bo - *Clostridium botulinum* B1 str. Okra; S.go - *Streptococcus gordonii* str. Challis substr. CH1; S.sa - *Streptococcus sanguinis* SK36; S.ag - *Streptococcus agalactiae COH1;* S.ag2 - *Streptococcus agalactiae* A909; E.fa - *Enterococcus faecalis* V583; E.fa2 - *Enterococcus faecium* DO.

4.4 Cloning of the putative *bglA* gene of *C. acetobutylicum* ATCC 824

Genomic DNA was extracted from an overnight RCM culture of *C. acetobutylicum* ATCC 824 and PCR reactions were performed, as described in the Experimental Procedures section 2.6. Primers were designed to target the region containing the putative *bglA* gene, encoding the sugar specific enzyme II component of the β-glucoside PTS. Details of the primers are in Chapter 2, Table 2.3.

The PCR reaction parameters were optimised (data not shown) and an annealing temperature of 55°C was established, which yielded a single band of the expected size. For cloning a fresh PCR reaction was performed, using the optimised annealing temperature, and the product was analysed by agrose gel electrophoresis using DNA molecular weight markers to estimate the size of the product (Figure 4.22). The product obtained was estimated at just under 2000bp (expected 1941bp) which is consistent with the successful amplification of the putative EII region.

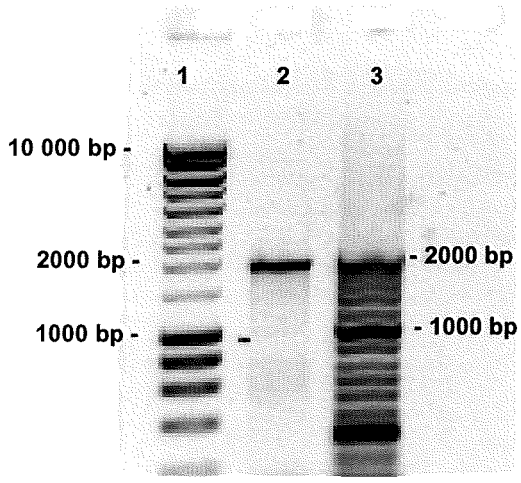


Figure 4.22: PCR amplification of the putative β-glucoside enzyme II

Agrose gel electrophoresis image of the PCR product produced with primers specific for amplification of the region cac1407, expected size of band is 1941bp.

Lane 1, Hyperladder I (Bioline); lane 2, PCR product; lane 3, Hyperladder II (Bioline).

The PCR product was added to the linearized pCR2.1-TOPO TA cloning vector (details and DNA sequence of pCR2.1 can be found in appendix I & II) and transformed into chemically competent *E. coli* TOP10 cells, using the "One Shot Chemical Transformation Protocol" as stated in Chapter 2.

Plasmid was added to an aliquot of TOP 10 chemically competent *E. coli*, mixed and incubated on ice. The reaction was then heat shocked and S.O.C medium (Bioline) was added and the vial was incubated for one hour of outgrowth. After the outgrowth period the resulting transformants were analysed by spreading various volumes (to ensure some of the plates contain single isolated colonies) onto prewarmed LB plates supplemented with ampicillin and with X-gal spread on the surface.

After an overnight incubation many white and fewer blue colonies were observed, indicating that the cloning was successful (Figure 4.23). Isolated colonies were selected and re-streaked, for single colonies, onto individual LB / ampicillin plates (for short term storage during analysis), and spread on to ampicillin / X-gal plates in 1cm lines, to amass sufficient biomass for colony PCR.

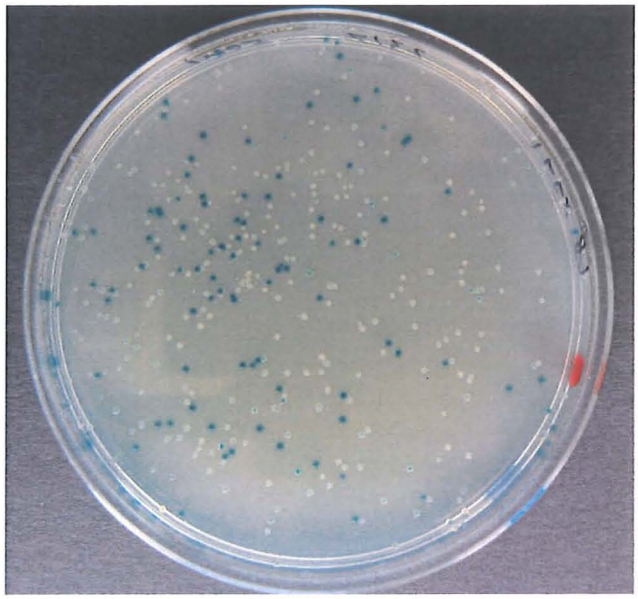


Figure 4.23: Blue/white screening for positive transformants

Screening of TOPO TA cloning reaction on a LB / ampicillin / X-gal agar plate. Positive transformants containing a vector with an insert appeared white.

Colony PCR was performed on the amassed cells of the 1cm streaks, as a rapid method of ascertaining if the white transformants contained a plasmid carrying a DNA insert. The resultant supernatant from the colony PCR procedure was collected and used as the template in a PCR conducted with the primers initially used to target the putative EII region, as described in Chapter 2, Section 2.5.6. PCR products were analysed by agarose gel electrophoresis, lane 7 of the gel contains a band consistent with the amplification of DNA from a colony harbouring the recombinant plasmid (Figure 4.24).

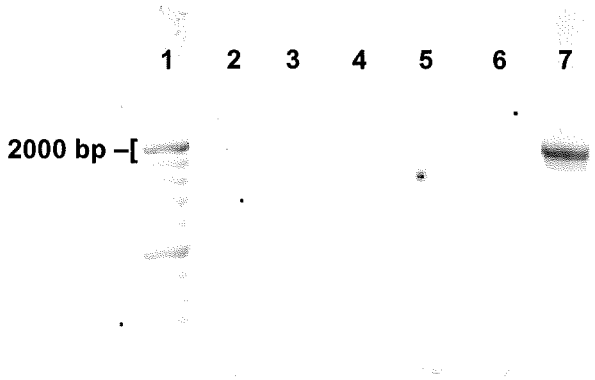


Figure 4.24: Screening for cell lines which contain recombinant plasmid
Agrose gel electrophoresis image of the PCR product produced after the colony PCR procedure was carried out on the *E. coli* TOP10 transformants.
Lane 1, 5µl of DNA Hyperladder II (Bioline); lanes 2-7, 5µl of colony PCR product.

The selected transformants were inoculated into LB / ampicillin broth and incubated overnight. For long term storage 850µl of the broth culture was mixed with 150µl of sterile glycerol and stored at -80°C. Plasmids were extracted from the overnight LB / ampicillin broth cultures as described in Chapter 2, Section 2.5.7. The plasmids were examined by agarose gel electrophoresis before further analysis, the gel image is presented in Figure 4.25).

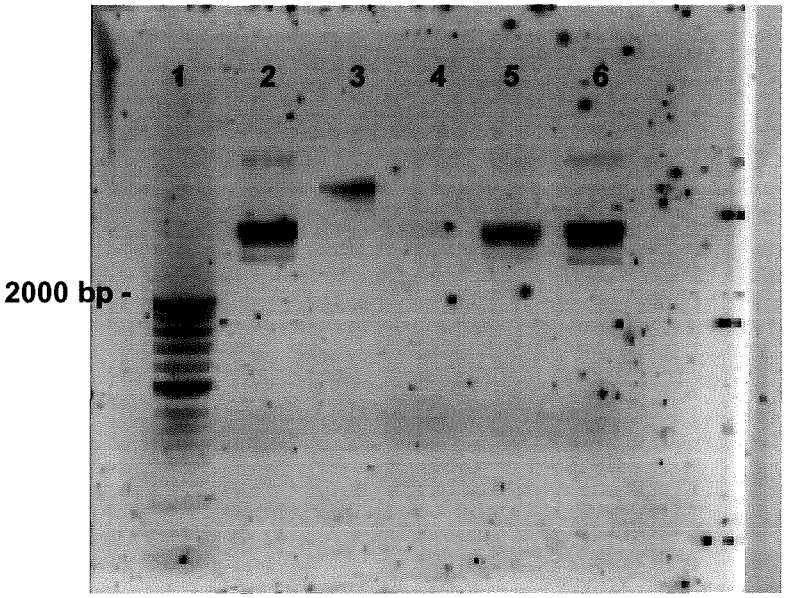


Figure 4.25: Purified preparations of recombinant plasmids

Purified plasmid preparations from the transformed *E. coli* TOP10 cells. Lane 1: 5µl DNA Hyperladder II, Lanes 2 – 6: 5µl purified plasmid preparation. Black dots on the image are due to dust within the image analyser.

The plasmid and vector combination was calculated to be 5872bp in total. Using the RestrictionMapper programme three HindIII sites were identified, two cutting within the vector, and the other cutting asymmetrically within the DNA insert. The theoretical restriction patterns resulting from insert ligating within the vector under the control of the T7 promoter would result in three fragments of 5385, 431 and 56 (bp). Whereas the restriction pattern of the insert under the control of the *Lac* promoter would result in three fragments of 4246, 1570 and 56 (bp). A restriction digest was carried out using HindIII and purified plasmid preparation as stated in the Experimental Procedures Chapter. The resultant digest was separated on a 1% agrose gel by electrophoresis presented in Figure 4.26. The size of the fragments obtained indicate the digested plasmid in lane 1 contains the insert under the control of the Lac promoter and the digested plasmid in lane 2 contains the insert under the control of the T7 promoter. Figure 4.27 shows a visual representation of the plasmid pEB100 (the DNA sequences of the amplified EII region, under the control of the Lac promoter, and vector pCR2.1) generated using Plasmapper (Dong *et al.*, 2004). The plasmid was designated pEB100.

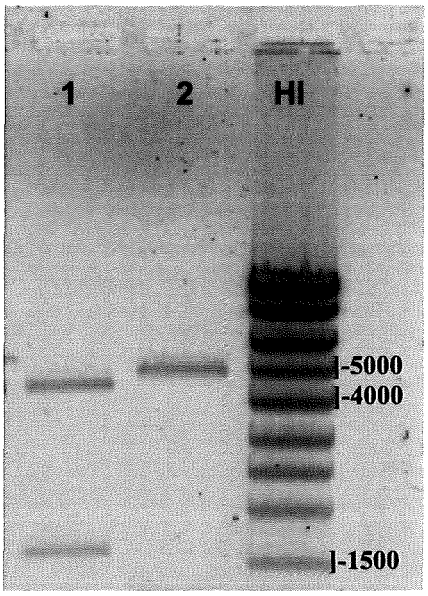


Figure 4.26: Restrication digest of with HindIII

Image of a 1% agarose gel. The DNA molecular weight marker is Bioline Hyperladder I. Lanes 1 & 2 contains recobinantant plasmids which harbour the putative β-glucoside digested with HindIII (Promega).

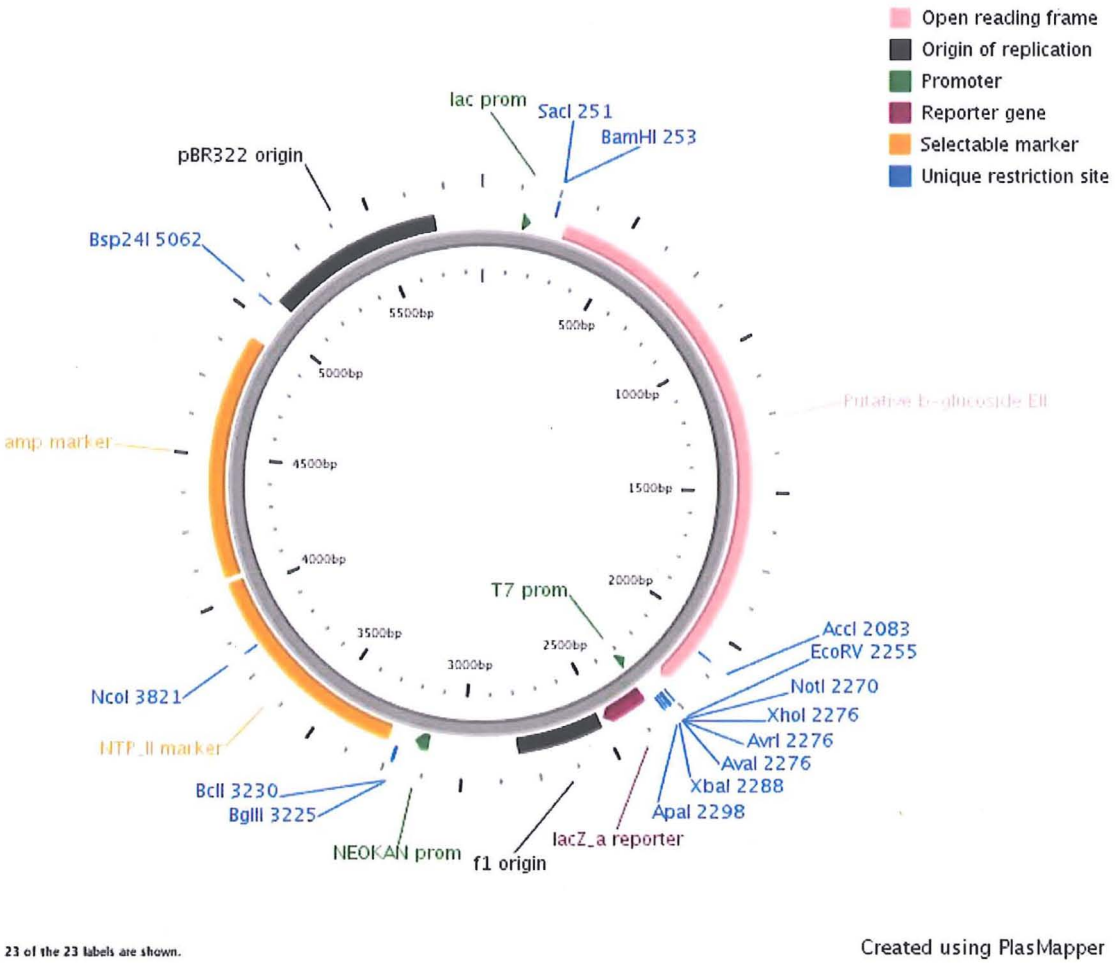


Figure 4.27: Features of pEB100

Important features of pEB100 comprising the pCR2.1 TOPO TA cloning vector and the putative β-glucoside specific enzyme II. The plasmid is 5872 bp and has many unique restriction sites, also there are two HindIII sites, one cutting within the vector the other cutting asymmetrically within the inserted DNA region (pink). The *lac* and T7 promoters are indicated by green triangles.

4.4.1 Complementation of the *E. coli* cryptic *bgl* PTS

E. coli TOP10 cannot transport the β-glucosides; cellobiose, arbutin and salicin, therefore when grown on MacConkey agar the *E. coli* colonies appear white/yellow and the surrounding agar appears yellow. *E. coli* TOP10 does however encode the general PTS proteins and a phospho-β-glucosidase. Thus, if a gene encoding a β-glucoside permease was introduced and expressed, the fermentation profile of *E. coli* TOP10 on MacConkeys agar would subsequently switch to red colonies on red agar.

Two *E. coli* transformants, one containing the cloned fragment of the *C. acetobutylicum* putative β-glucoside specific EII under the control of the Lac promoter (designated pEB100), were streaked onto MacConkey agar containing either glucose, cellobiose or the related β-glucosides salicin or arbutin. Both transformants displayed a positive fermentation phenotype for glucose (red colonies). Neither transformant could ferment cellobiose or salicin, thus the colonies and surrounding agar appear white/yellow. However, when screened on MacConkey's agar plates containing the β-glucoside arbutin, the colonies produced by the transformant *E. coli* EB100 were dark brown/red, in stark contrast to the white/yellow colonies produced by the other (Figure 4.28, summarised in Table 4.5). Therefore, the *E. coli* transformant EB100 (which harbours the vector pEB100 that contains the clostridial EII permease) produced a positive fermentation phenotype for arbutin.

Table 4.5: MacConkey Screening Results

	Fermentation Phenotype + / -			
Transformant	Glucose	Cellobiose	Salicin	Arbutin
EB10A	+	-	-	-
EB100	+	-	-	+

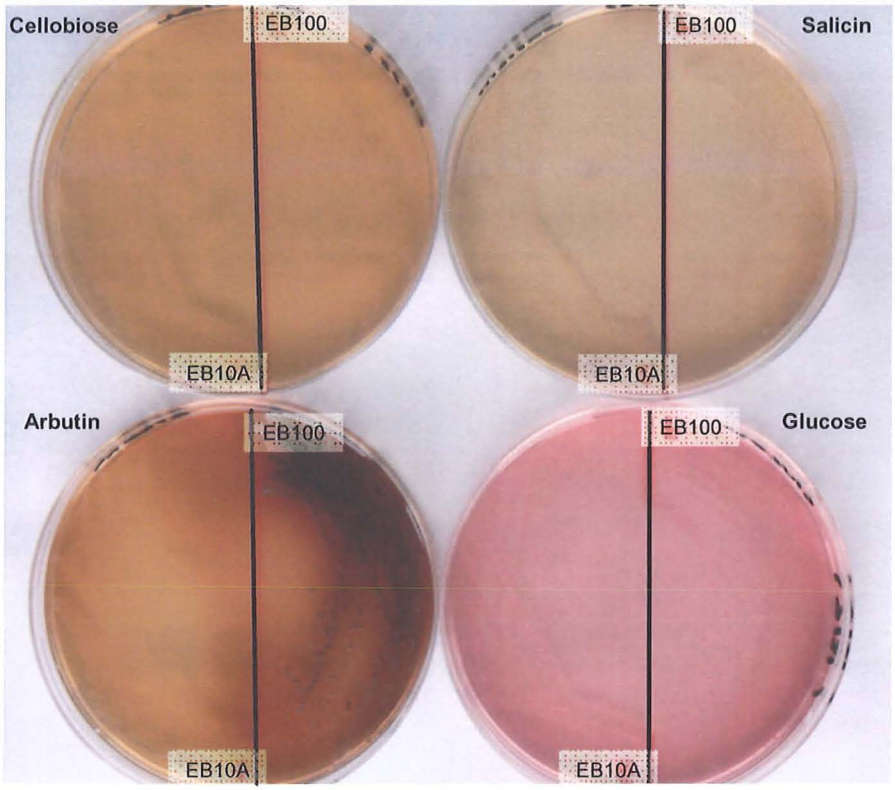


Figure 4.28: Screening complementation analysis

E. coli TOP10 cells transformed with the putative β-glucoside specific enzyme II streaked onto MacConkey agar plates containing either; cellobiose, salicin, arbutin or glucose. Transformants labelled EB10A and EB100 (harbours the pEB100 plasmid).

Plate is split in half with transformants streaked on both sides.

Part II:

4.5 Examination of putative β-glucoside system cac0383-cac0386

The research reported in Part I described the investigation of genes putatively encoding: a BglG type transcriptional terminator, an EII^{bgl}, and a phopsho-β-glucosidase. However, heterologous expression of the sugar-specific EII component in *E. coli* TOP10 only conferred the ability to transport the β-glucoside arbutin, and not the related sugars salicin or cellobiose. *C. acetobutylicum* has the ability to utilise salicin and cellobiose (Table 4.1) and PTS activity for cellobiose has been observed in this organism (Mitchell & Tangney, 2005). Therefore, research continued to identify the salicin and cellobiose specific permeases.

Examination of the 3900bp region revealed the presence of four clustered ORFs, all in the same orientation, with the locus tags cac0383, cac0384, cac0385 and cac0386 (Figure 4.29). The deduced amino acid sequence of each putative ORF was queried against sequence databases using the BLAST service at NCBI (Altschul *et al.,* 1997) to search for sequence homology. On the basis of this homology the putative proteins encoded by these genes were (in the interim) designated an EIIA domain, EIIB domain, phospho-β-glucosidase and an EIIC domain.

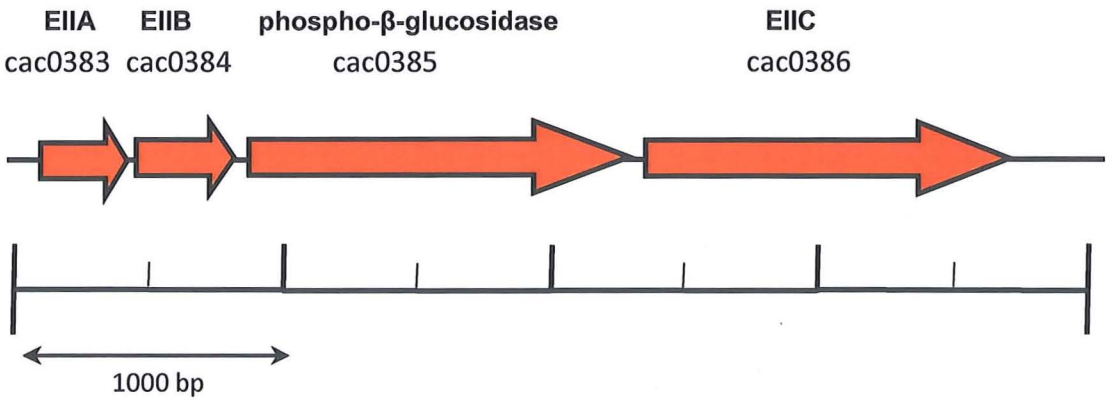


Figure 4.29: Arrangement of the ORFs of a putative β-glucoside utilisation operon of *C. acetobutylicum*

Representative diagram of the 3.9kb region containing a putative β-glucoside utilisation operon. ORFs are represented by a block arrows (indicating direction of transcript), locus tags are given above, scale (bp) is given below. The genome was examined using GIB server and the ORFs were predicted using the ORF finder tool at the NCBI server. Genes were tentatively named on the basis of the results of BLAST homology searches an; EIIA domain, EIB domain, phospho-β-glucosidase and an EIIC domain.

By comparing and analysing the protein sequences of experimentally validated published proteins with putative proteins, inferences about their function can be made with a great deal of confidence. An extensive search of the current literature was carried out to identify organisms with characterised genes for β-glucoside transport and utilisation. Gram positive and Gram negative organisms, which possess characterised β-glucoside phosphotransferase systems are listed in Table 4.6 and Table 4.7 respectively.

The 3.9 kb region containing the putative β-glucoside utilisation and transport genes was subjected to further scrutiny. The EII domains are spread over three separate ORFs, unlike the system investigated in Part I, in which all three domains were encoded in a single ORF. The deduced amino acid sequence of cac0383, cac0384 and cac0386 were analysed for the hydropathicity profile of the protein, the results are presented in Figure 4.30. Descriptions of the individual domains are presented in the next section.

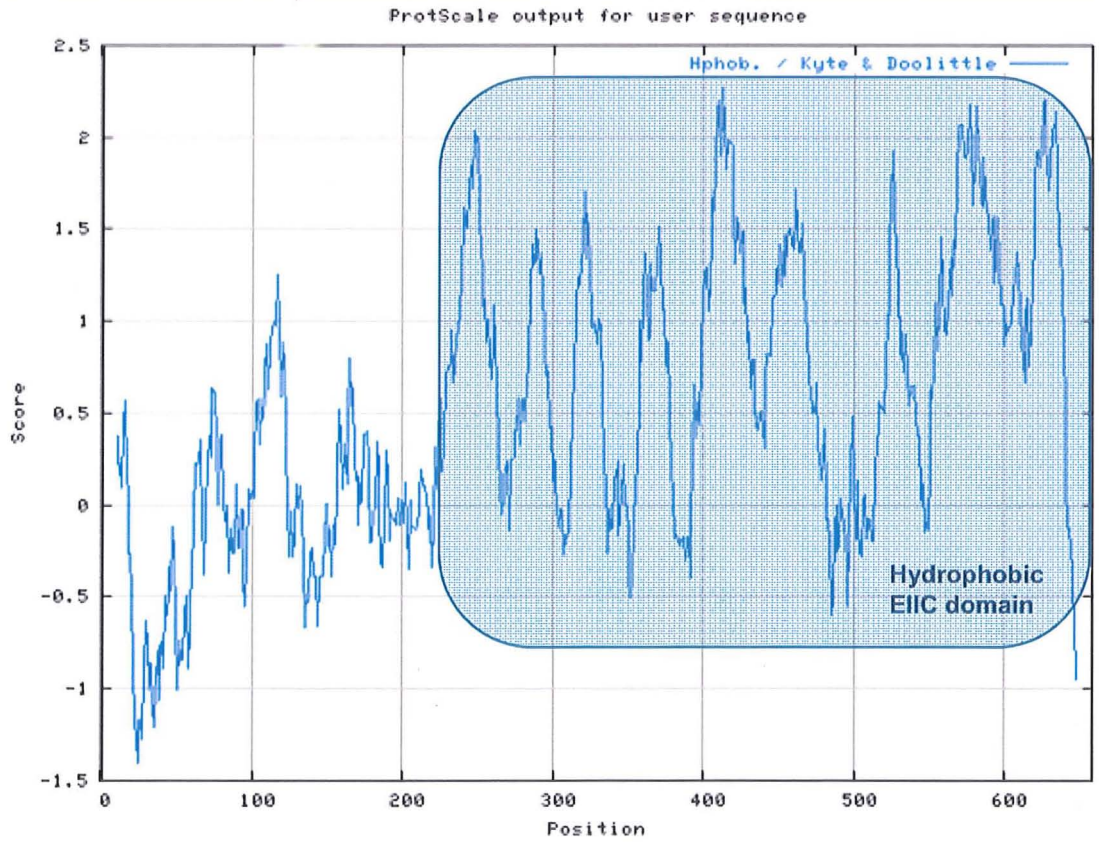


Figure 4.30: Protein hydrophobicity plot of the putative EII

Kyte Doolittle ExPASy proteomics server was used to display the overall hydropathicity profile of the deduced amino acid sequence of cac0383, cac0384 and cac0386, the putative β-glucoside specific EII. Window size = 21. Window size refers to the number of amino acids examined at a time to determine a point of hydrophobic character, 21 is optimum for identifying hydrophobic regions within proteins (Kyte & Doolittle, 1982).

Table 4.6: Experimentally verified genes and operons involved in β-glucoside utilisation from selected Gram positive organisms

Organism	Operon	Accession No. Operon	Accession No. Regulator	Accession No. Permease	Accession No. β-glucosidase	Reference
Clostridium longisporum	AbgGFA	L49336	AAC05712	AAC05713	AAC05714	Brown & Thomson, 1998
Bacillus subtilis	BglA	-	-	-	P42973	Zhang & Aronson, 1994
Bacillus subtiis	LicRBC AH	Z49992	P46321	P46318 P46317 P46319	P46320*	Tobisch *et al.*, 1997
Bacillus subtilis	LicST	-	P39805	-	P04957*	Schnetz *et al.*, 1996; Murphy *et al.*, 1984
Bacilus subtilis	BglPH	Z34526	-	P40739	P40740	Le Coq *et al.*, 1995
Bacilus subtilis	YckE (BglC)	-	-	-	P42403	Setlow *et al.*, 2004
Bacilus subtilis	YdhP (GmuD BglD)	-	-	-	O05508	Setlow *et al.*, 2004
Bacillus stearothermop -hilus	CelRAB CD	No accession – copy seq from paper				Lai & Ingram, 1993
Lactobacillus delbrueckii	ArbXZ	Z86115	-	-	O53046	Weber *et al.*, 1998
Lactobacillus plantarum	BglH	Y15954	-	-	O86291	Marasco *et al.*, 1998
Lactobacillus plantarum	BglGPT	AJ250202	Q9L462	Q9L461	Q9L460	Marasco *et al.*, 2000
Corynebacteri -um glutamicum	BglFAG	AF508972	Q8GGK1	Q8GGK3	Q8GGK2	Kotrba *et al.*, 2003
Streptococcus mutans	BglPCA	AF206272	Q9KJ78	Q9KJ80	Q9KJ76	Cote *et al.*, 2000
Streptococcus mutans	LicT	-	Q8DUF3	-	-	Cote & Honeyman, 2003

Table 4.7: Experimentally verified genes and operons involved in β-glucoside utilisation from selected Gram negative organisms

Organism	Operon	Accession No. Operon	Accession No. Regulator	Accession No. Permease	Accession No. β-glucosidase	Reference
Pectobacterim carotovorum	BglTPB	AY542524	Q6QGY7	Q6QGY6	Q6QGY5	An *et al.*, 2004
Pectobacterim carotovorum	BglEFIA	AY769096	-	Q5UAU5 Q5UAU7 Q5UAU6	Q5UAU4	Hong *et al.*, 2006 (EFI=IIBCA)
Pectobacterim carotovorum	AscGFB	AY622309	Q6IV16	Q6IV15	Q6IV14	An *et al.*, 2005
Erwinia chrysanthemi	ArbGFB	M81772	P26211	P26207	P26206	Hassouni *et al.*, 1992
Azospirillum irakense	SalAB	AF090429	-	-	Q9RH02 Q9RH03	Faure *et al.*, 1999
Klebsiella aerogenes	BglFBG	AY124799 /800	Q8KP25	Q0PWF9	Q8KP24	Raghunand & Mahadevan, 2003
Klebsiella oxytoca	CasRAB	U61727	Q48407	Q48408	Q48409	Lai *et al.*, 1997
Escherichia coli K-12	AscGFB	M73326	P24242	P24241	P24240	Hall & Xu, 1992
Escherichia coli K-12	ChbBCA RF CelABC DF	X52890	P17410	P17335/ P69791; P17409/ P69795; P17334	P17411	Parker & Hall, 1990; Keyhani & Roseman, 1997; Thompson *et al.*, 1999
Escherichia coli K-12	BglGFB / BglGCB	M16487	AAA23509	P08722 / AAA23510	AAA23511	Schnetz *et al.*, 1987; Houman *et al.*, 1990
Escherichia coli K-12	BglA	-	-	-	Q46829	Parker & Hall, 1988; Prasad, *et al.*, 1973
Escherichia coli K-12	BglX	-	-	-	U15049	Yang *et al.*, 1996
Escherichia coli K-12	GlvBC	-	-	P69789 P31452	-	Reizer *et al.*, 1994

4.5.1 EIIA domain cac0383

As revealed by the BLAST homology search, the (321bp) ORF cac0383 appears to encode an EIIA domain (106aa) and has a calculated molecular weight of 11,808Da. The Conserved Domain Database (CDD) (Marchler-Bauer *et al.,* 2005; Marchler-Bauer *et al.,* 2007; Marchler-Bauer & Bryant, 2004) was screened against the deduced amino acid sequence of the putative EIIA. Several conserved domains were recognised within the sequence and are presented in Figure 4.31. Five out of five of the residues which compose the conserved active site for EIIA^{lac} domains were found within the putative EIIA domain. Four out of five of the residues which constitute the conserved stabilising methionine cluster, found in the EIIA^{lac} (Sliz *et al.,* 1997), were identified in the cac0383 sequence. A conserved Mg binding site was also located just downstream from a conserved histidine phosphorylation site (Figure 4.31).

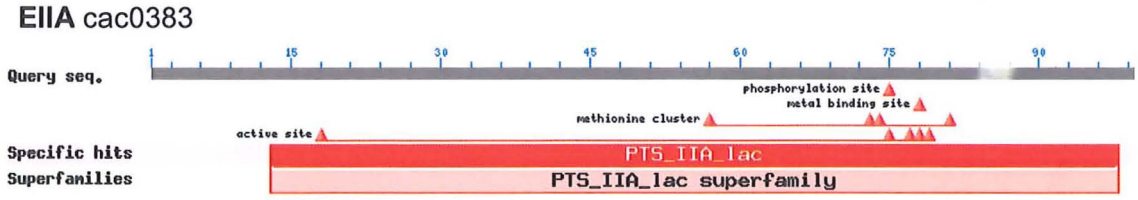


Figure 4.31: Comparison of the EIIA domain from the CDD visual output
The deduced amino acid sequence of the putative β-glucoside specific EIIA domain (cac0383) was entered into the BLASTP suite and the conserved domain output captured.

The deduced amino acid sequence of the putative EIIA cac0383 along with other EIIA domains from experimentally verified EIIA domains from β-glucoside specific EIIs were aligned in ClustalW2 and the output analysed in TreeView, the results are presented in Figure 4.32. The tree clearly forms two main clusters, with the EIIA^{lac} in one cluster and the EIIA^{glu} domain in the other. The cac0383 (EIIA^{lac}) sequence shares only limited percentage identity with the *Bacillus stearothermophilus* CelEIIA (41%), *Bacillus subtilis* LicEIIA (38%), *Pectobacterium carotovorum* BglEIIA (38%) and *Escherichia coli* (38%) ChbEIIA domains.

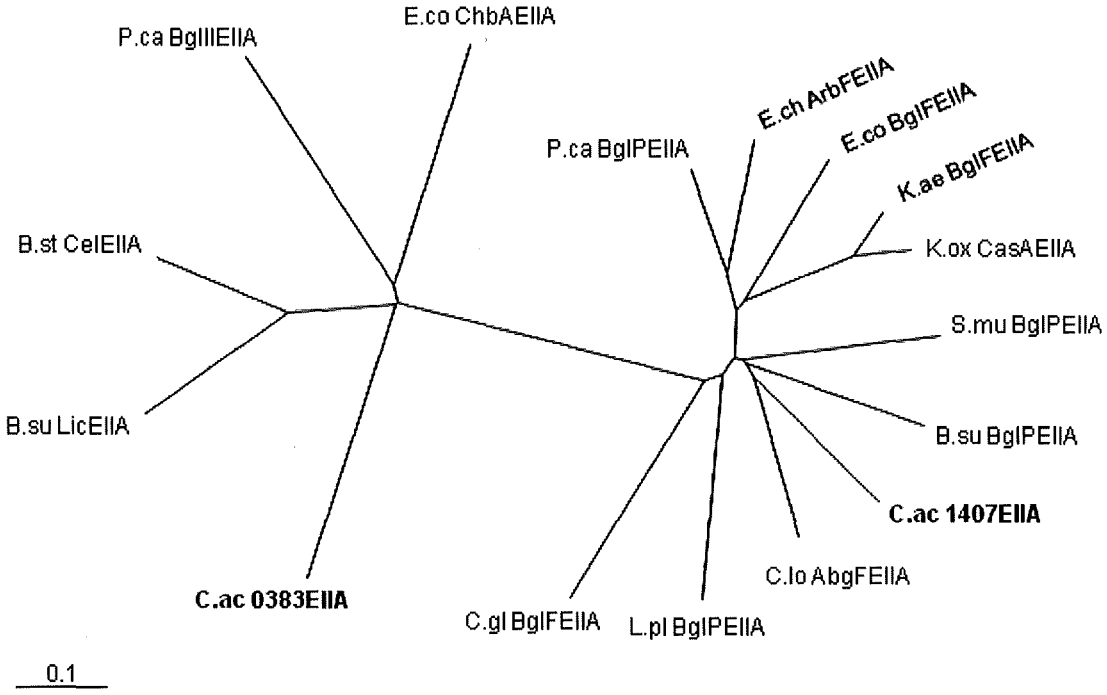


Figure 4.32: Unrooted phylogenetic tree of EIIA domains from characterised EII PTS permeases

The amino acid sequences of the putative EIIA cac0383 were compared with the EIIA domain of cac1407 alongside selected, experimentally verified EIIA domains and analysed using ClustalW and TreeView. The first letter of the genus and the first two letters of the species are listed, followed by the protein. The full designations are given below.

c.ac, *Clostridium acetobutylicum*; c.lo, *Clostridium longisporum*; b.su, *Bacillus subtilis*; b.su, *Bacillus stearothermophilus*; l.de, *Lactobacillus delbrueckii*; l.pl, *Lactobacillus plantarum*; c.gl, *Corynebacterium glutamicum*; s.mu, *Streptococcus mutans*; p.ca, *Pectobacterium carotovorum* subsp. *carotovorum* LY34; e.ch, *Erwinia chrysanthemi*; a.ir, *Azospirillum irakense*; k.ae, *Klebsiella aerogenes*; k.ox, *Klebsiella oxytoca*; e.co, *Escherichia coli*

4.5.2 EIIB domain cac0384

The ORF cac0384 putatively encodes an EIIB domain. The protein is predicted to have a molecular weight of 10,883Da and is 117aa in length (354bp). As with the putative EIIA domain, many conserved features were identified from the CD database, presented in Figure 4.33. A conserved phosphorylation site and active site found in EIIB domains (Eiso *et al.*, 2001), were identified (Figure 4.33).

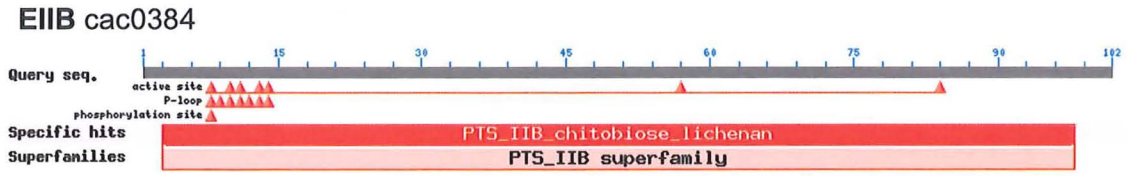


Figure 4.33: EIIB (cac0384) domain from the CDD visual output

The deduced amino acid sequence of the putative β-glucoside specific EIIB domain (cac0384) was entered into the BLASTP suite and the conserved domain output captured.

A phylogenetic tree was constructed to investigate the relatedness of the putative EIIB sequence to characterised published EIIB domains associated with β-glucoside utilisation operons, the results are presented in Figure 4.34. The sequence shares 52% identity with the *B. stearothermophilus* cellobiose EIIB and 48% with the *E. coli* chitobiose and cellobiose specific EIIB.

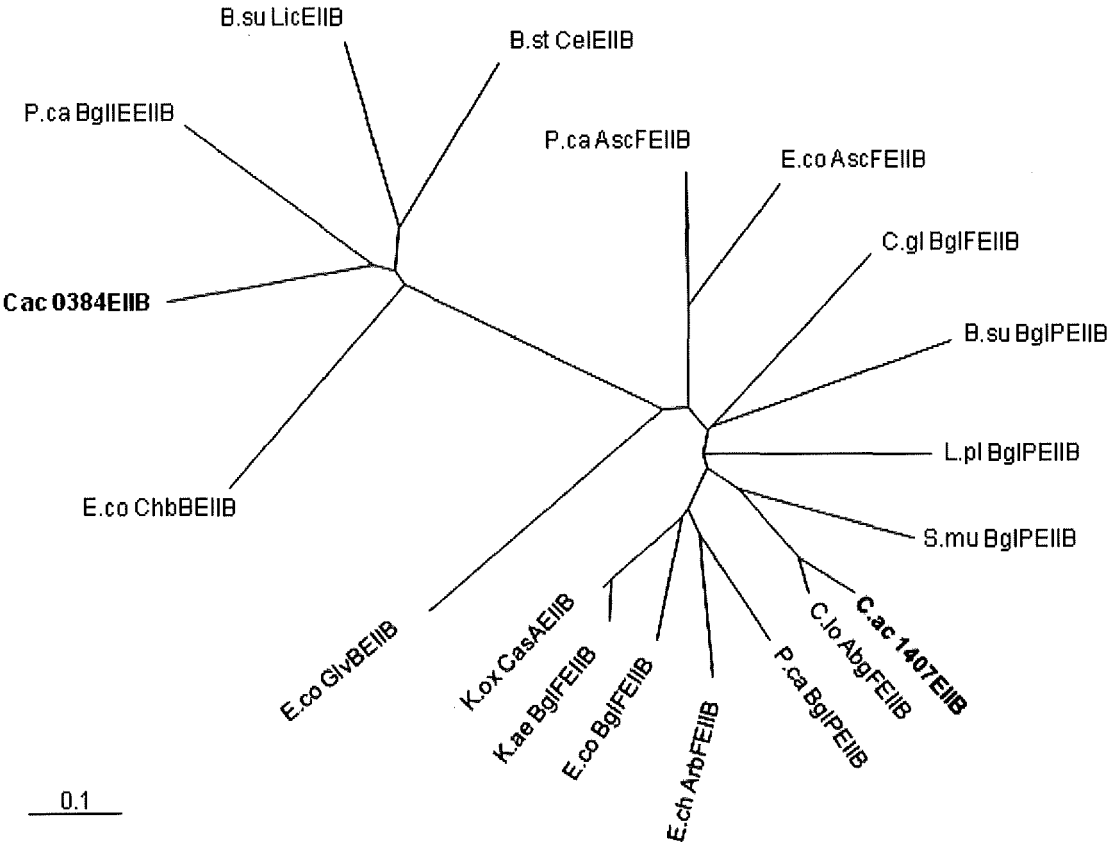


Figure 4.34: Unrooted phylogenetic tree of EIIB domains from characterised EII PTS permeases

The amino acid sequences of the putative EIIB (cac0384) were compared with the EIIB domain of cac1407 alongside selected, experimentally verified EIIB domains and analysed using ClustalW and TreeView. The first letter of the genus and the first two letters of the species are listed, followed by the protein. The full designations are given below.

The first letter of the genus and the first two letters of the species are listed, followed by the protein. The full designations are given below.

c.ac, *Clostridium acetobutylicum*; c.lo, *Clostridium longisporum*; b.su, *Bacillus subtilis*; b.su, *Bacillus stearothermophilus*; l.de, *Lactobacillus delbrueckii*; l.pl, *Lactobacillus plantarum*; c.gl, *Corynebacterium glutamicum*; s.mu, *Streptococcus mutans*; p.ca, *Pectobacterium carotovorum* subsp. *carotovorum* LY34; e.ch, *Erwinia chrysanthemi*; a.ir, *Azospirillum irakense*; k.ae, *Klebsiella aerogenes*; k.ox, *Klebsiella oxytoca*; e.co, *Escherichia coli*.

4.5.3 EIIC domain cac0386

The putative EIIC domain (cac0386) was identified just downstream of a phospho-β-glucosidase. The protein was predicted to have a molecular weight of 48 177Da, and to be 450aa in length (1353bp). The consensus sequence for disaccharide binding was aligned against the putative EIIC domains of cac1407 (discussed in Part I) and cac0386, the results are presented in Figure 4.35. The sequence of cac0386 was found to only share limited identity with the other two sequences.

Figure 4.35: Conserved sequence for disaccharide binding

Alignment of the conserved EIIC sequence for disaccharide binding from *K. oxytoca.*

The C domain provides the EII component with substrate specificity (Postma *et al.*, 1993). Therefore, isolating and compiling the sequences of EIIC domains of published, verified permeases specific for β-glucoside transport could provide useful information in the characterisation of EIIs with unknown substrate specificity. The unrooted phylogenetic tree is presented in Figure 4.36; the substrate specificity of each permease is labelled next to the organism and gene name. The branch which the putative β-glucoside EII falls in contains two proteins from the Gram positive *Lactococcus lactis* PtcC and YidB. Both proteins are specific for cellobiose transport and they share 43% and 40% identity respectively. Lower down on that branch there is an offshoot branch which carries two *Bacillus* proteins, once specific for lichenan and the other specific for cellobiose transport.

Additional information for the legend of Figure 4.36: Unrooted phylogenetic tree of EIIC domains from characterised EII PTS permeases correlated with substrate specificity.

The first letter of the genus and the first two letters of the species are listed, followed by the protein. The full designations are given below.

c.ac, *Clostridium acetobutylicum*; c.lo, *Clostridium longisporum*; b.su, *Bacillus subtilis*; b.su, *Bacillus stearothermophilus*; l.de, *Lactobacillus delbrueckii*; l.pl, *Lactobacillus plantarum*; l.la, *Lactococcus lactis*; c.gl, *Corynebacterium glutamicum*; s.mu, *Streptococcus mutans*; p.ca, *Pectobacterium carotovorum* subsp. *carotovorum* LY34; e.ch, *Erwinia chrysanthemi*; a.ir, *Azospirillum irakense*; k.ae, *Klebsiella aerogenes*; k.ox, *Klebsiella oxytoca*; e.co, *Escherichia coli.*

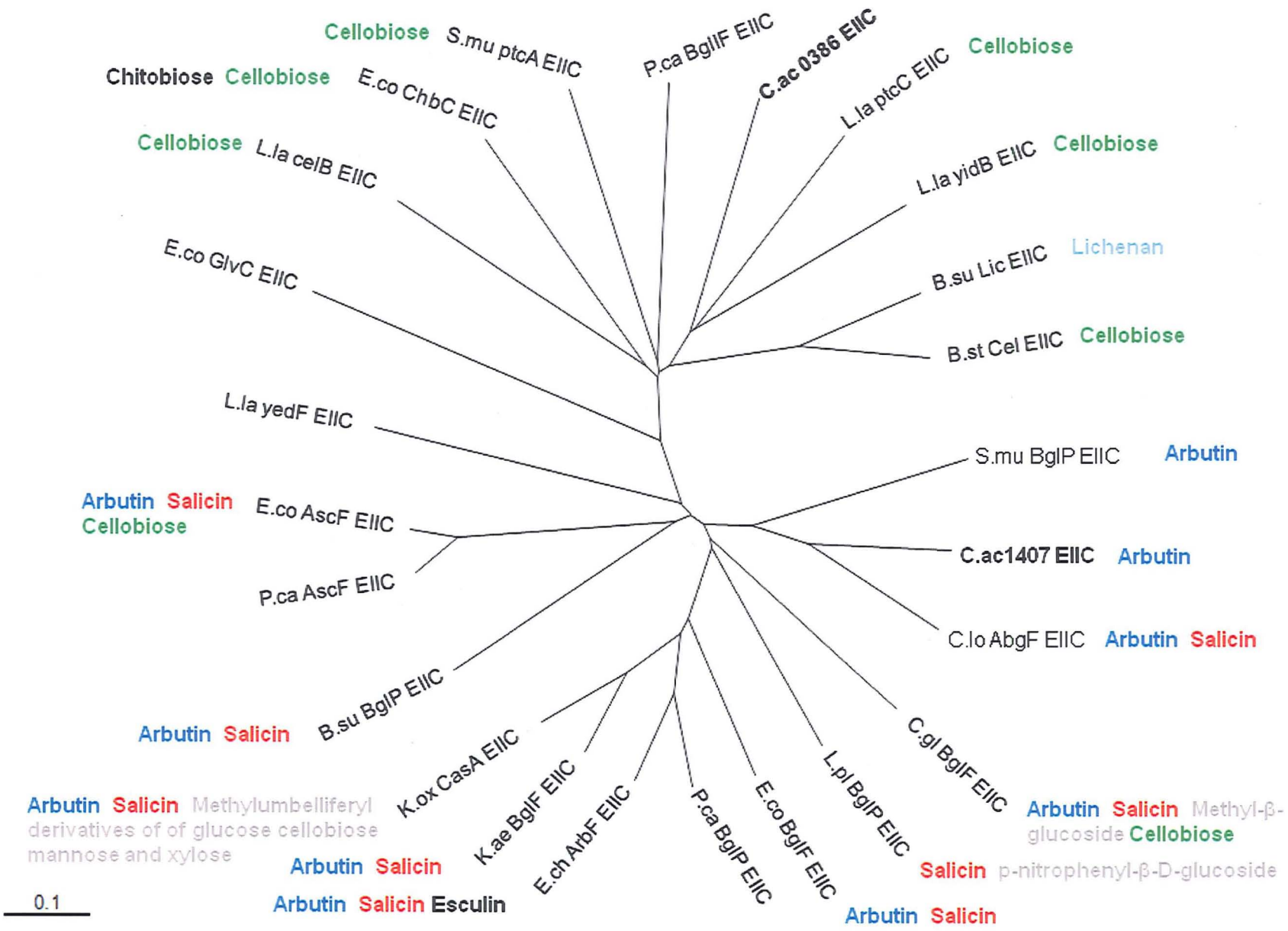


Figure 4.36: Unrooted phylogenetic tree of EIIC domains from characterised EII PTS permeases correlated with substrate specificity

The first letter of the genus and the first two letters of the species are listed, followed by the protein.

4.5.4 Phospho-β-glucosidase cac0385

The EII domains are interrupted by ORF cac0385, which appears to encode a 6-phospho-β-glucosidase. The ORF encodes a 469aa protein (1410bp), with a calculated molecular weight of 54,133Da. Phospho-β-glucosidases and β-glucosidases contain a highly conserved signature sequence (LFIVENGLG; Lai *et al.*, 1997). The glycosyl hydrolase family 1 signature sequence was identified (370-379aa) and six out of nine residues matched perfectly (IYITENGLG). In the three positions where there were deviations from the signature sequence, the residues in place were greatly similar (for example the replacement of leucine with isoleucine – both small hydrophobic amino acids).

From the studied operons all of the sequences which encode phospho-β-glucosidase proteins were aligned in ClustalW2 and viewed in TreeView. The results are presented in Figure 4.37. If stated in the literature, the family designation of each of the proteins are noted next to the gene names. The family 1 glycosyl hydrolase group seems to form a main cluster, with a second offshoot branch further down. The clostridial protein appears to fall into the family 1 group. The cac0385 sequence groups with the two *B. subtilis bgl* operon proteins BglC and BglD, sharing 44% and 38% respectively.

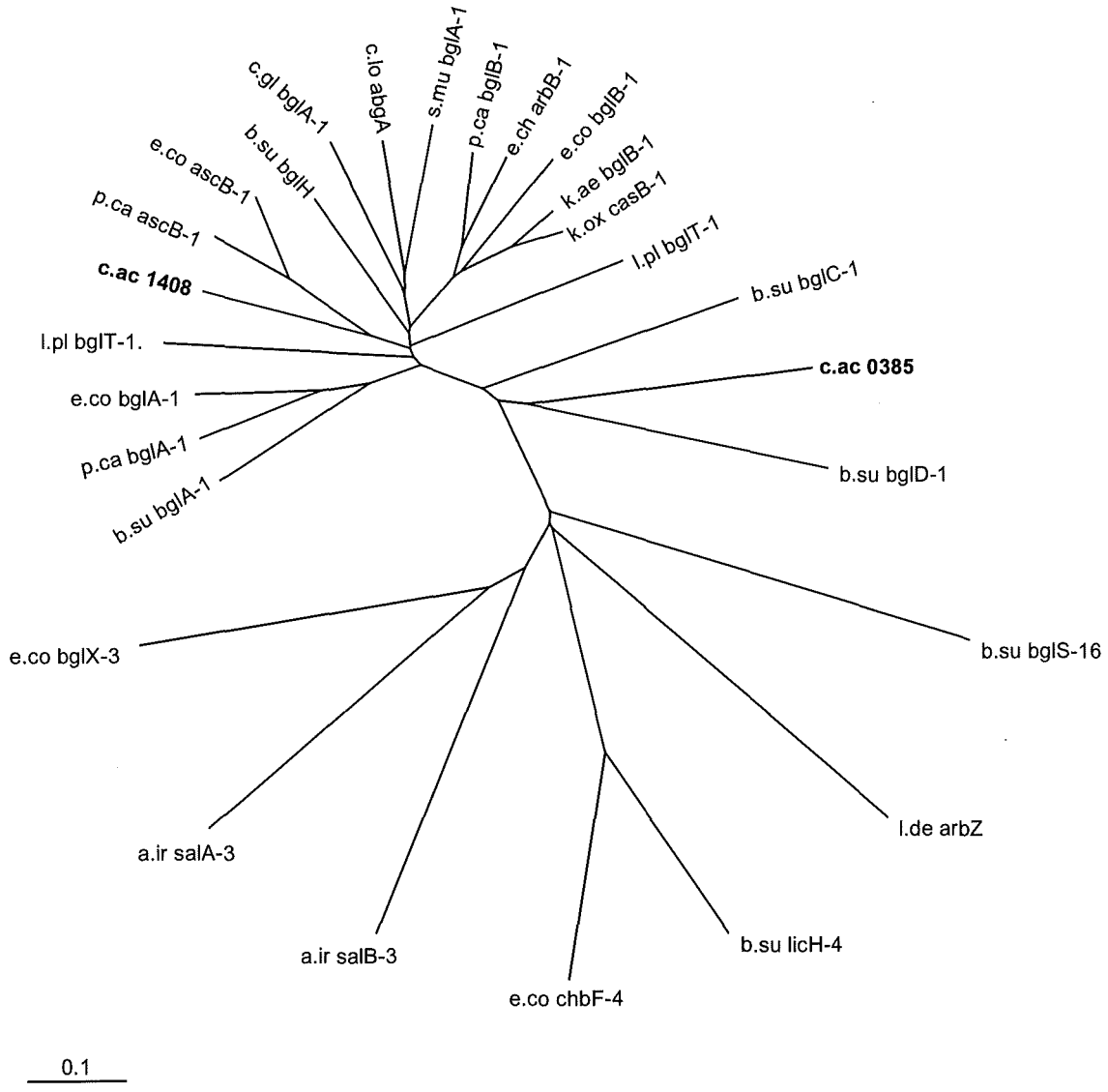


Figure 4.37: Unrooted phylogenetic tree of phospho-β-glucosidases

The amino acid sequences of the two putative phospho-β-glucosidases found in *C. acetobutylicum* were compared with selected experimentally verified published phospho-β-glucosidases and analysed using ClustalW and TreeView. The number after the gene name denotes the glycosyl hydrolase family the protein belongs to, where no number is present, no family designation was indicated in the literature. The *C. acetobutylicum* proteins are highlighted in **bold**. The full designations are given below.

c.ac, *Clostridium acetobutylicum*; c.lo, *Clostridium longisporum*; b.su, *Bacillus subtilis*; l.de, *Lactobacillus delbrueckii*; l.pl, *Lactobacillus plantarum*; c.gl, *Corynebacterium glutamicum*; s.mu, *Streptococcus mutans*; p.ca, *Pectobacterim carotovorum*; e.ch, *Erwinia chrysanthemi*; a.ir, *Azospirillum irakense*; k.ae, *Klebsiella aerogenes*; k.ox, *Klebsiella oxytoca*; e.co, *Escherichia coli* K-12

4.5.5 Putative regulator cac0382

A putative 896aa regulator (cac0382) was identified upstream of the EIIA domain. As a first step to obtaining the identity of the putative ORF a BLAST search was carried out, the results are summarised in Table 4.8. The majority of the top 10 hits were σ-54 factor interaction domain containing proteins from *C. beijerinckii* or *Bacillus* species, other hits include a transcriptional activator LevR from *C. butyricum*.

To investigate the phylogenetic relationship between the *C. acetobutylicum* β-glucoside associated regulators and the sequences of the characterised β-glucoside associated regulators gathered from the literature, and alignment was conducted in ClustalW2 and the resultant information viewed in TreeView, and the results are presented in Figure 4.38. In agreement with the results from Part I the antiterminator cac1406 forms a cluster with the antiterminator from *C. longisporum* (blue dashed box). However, the putative regulator cac0382 appears to be most closely related to the PRD-containing transcriptional regulator LicR from *Bacillus subtilis* (red dashed box). Northern Blot experiments, in *B. subtilis*, indicated that LicR was monocistronic and positively regulates the *licBCAH* operon which encodes an EIIB, EIIC, EIIA and hydrolase specific for lichenan and lichenan hydrolysis products (Tobisch *et al.*, 1997; Tobisch *et al.*, 1999). To investigate this further these two sequences were aligned in ClustalW2 and viewed in GeneDoc (data not shown). Overall the sequences were not significantly homologous and no blocks of conservation were observed.

LevR is a PRD-containing transcriptional activator also from *B. subtilis* (Martin-Verstraete *et al.*, 1998); the sequence was aligned with the putative clostridial regulator (Figure 4.39A). Extensive blocks of conservation were observed in the N-terminus of the protein.

A search was carried out to identify additional transcriptional activators associated with β-glucoside utilisation. Lai & Ingram (1993) identified the CelR protein *Bacillus stearothermophilus*. The truncated sequence is available on UniProt, and was used for an alignment with the clostridial sequence (Figure 4.39B).

Each of the three complete regulators were input into BLAST (Débarbouillé *et al.*, 1991) and the putative domain organisation of each protein was ascertained by comparison with CDD (Figure 4.40). LevR and the putative clostridial transcriptional activator have similar domain organisation.

Table 4.8: BLAST homology results for the deduced amino acid sequence of the putative regulator cac0382

Results of a BLAST homology search with the deduced amino acid sequence of the putative ORF (with the locus tag cac0382), the putative regulator.

Species	Putative Gene	% Identity	% Similarity	ORF Locus tag	Accession
Clostridium beijerinckii NCIMB 8052	σ-54 factor interaction domain-containing protein	58	78	Cbei_4641	YP_001311705
Thermoanaerobacterium thermosaccharolyticum DSM 571	PTS system transcriptional activator	59	77	TtheDRAFT_2142	ZP_05336916
Clostridium butyricum 5521	transcriptional regulatory protein LevR	56	75	CBY_0873	ZP_02949216
Thermoanaerobacter mathranii subsp. *mathranii* str. A3	PTS system transcriptional activator	48	69	TmathDRAFT_1847	ZP_05379879
Clostridium beijerinckii NCIMB 8052	σ -54 factor interaction domain-containing protein	46	65	Cbei_4686	YP_001311750
Brevibacillus brevis NBRC 100599	transcriptional regulatory protein	43	64	BBR47_56680	YP_002775149
Bacillus cereus Rock3-44	σ -54 factor interaction domain-containing protein	41	63	bcere0022_42400	ZP_04219805
Bacillus cereus subsp. cytotoxis NVH 391-98	σ-54 factor interaction domain-containing protein	42	63	Bcer98_3748	YP_001376935
Bacillus pseudomycoides DSM 12442	σ -54 factor interaction domain-containing protein	41	63	bpmyx0001_46340	ZP_04153814
Bacillus mycoides Rock1-4	σ -54 factor interaction domain-containing protein	41	63	bmyco0002_43780	ZP_04165095

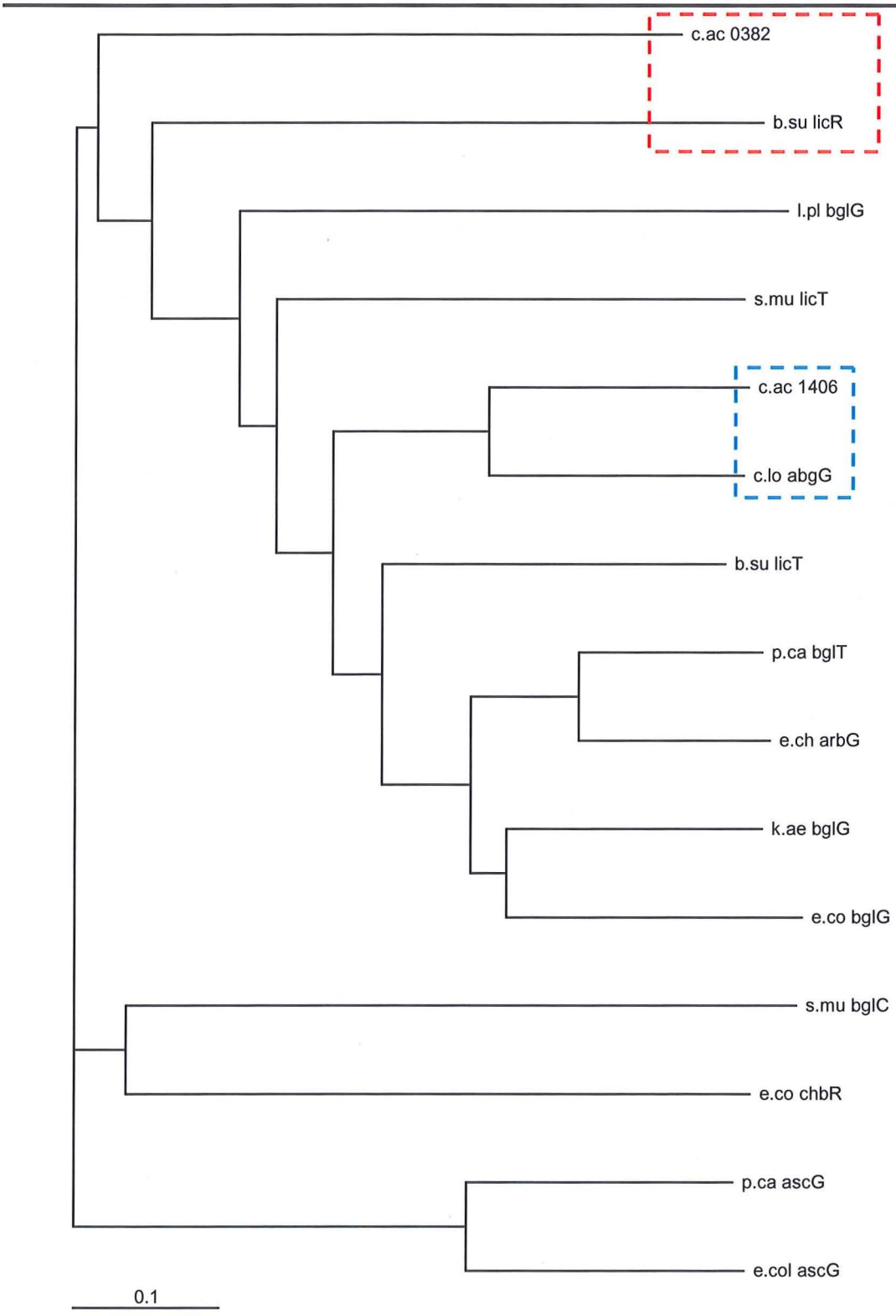


Figure 4.38: Rooted tree of regulators associated with β-glucoside utilisation genes

The amino acid sequences of the putative regulator cac0382 were compared with experimentally verified regulators associated with β-glucoside permeases and analysed using ClustalW and TreeView. The first letter of the genus and the first two letters of the species are listed, followed by the protein. The full designations are given below.

c.ac, *Clostridium acetobutylicum*; c.lo, *Clostridium longisporum*; b.su, *Bacillus subtilis*; l.pl, *Lactobacillus plantarum*; s.mu, *Streptococcus mutans*; p.ca, *Pectobacterium carotovorum* subsp.

carotovorum LY34; e.ch, *Erwinia chrysanthemi*; k.ae, *Klebsiella aerogenes*; e.co, *Escherichia coli*

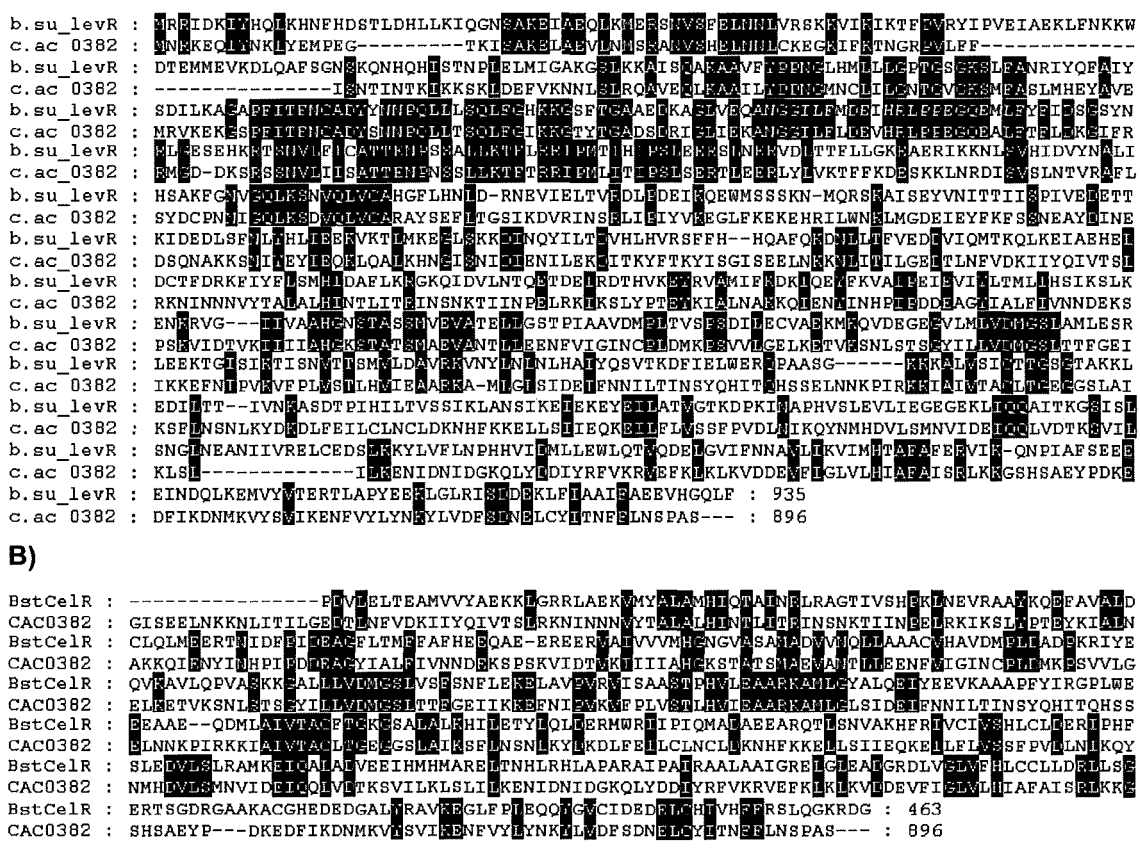


Figure 4.39: Alignment of the *C. acetobutylicum* putative transcriptional regulator against A) LevR from *B. subtilis* and B) the truncated CelR from *B. stearothermophilus*

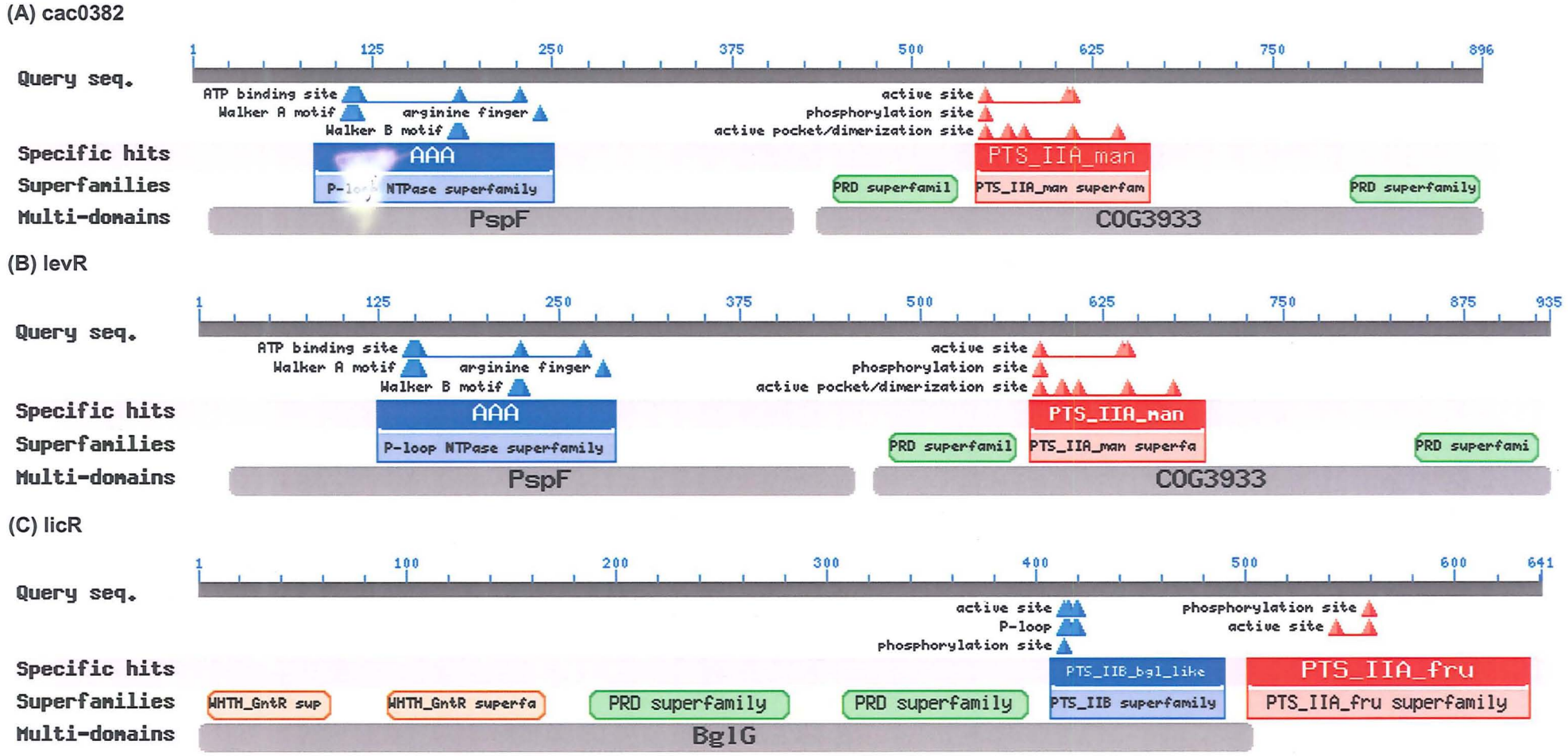


Figure 4.40: CDD output for (A) *C. acetobutylicum* cac0382 (B) *Bacillus subtilis* LevR and (C) *B. subtilis* LicR

The Conserved Domain Database (CDD) (Marchler-Bauer *et al*, 2005; Marchler-Bauer *et al*, 2007; Marchler-Bauer & Bryant, 2004) was used to make this output. Note the different scales above the domains.

4.6 Cloning of the putative β-glucoside specific operon

The bioinformatics evidence strongly suggests that the ORFs of the second *bgl* operon encode a β-glucoside specific EIIA (cac0383), EIIB (cac0384) and EIIC (cac0386) domains (Figure 4.29). To investigate this further, the 3612bp region of cac0383 to cac0386 was PCR amplified using primers detailed in Chapter 2, Table 2.3, and a long range taq polymerase. The analysed gel image is presented in Figure 4.41A. After initial PCR optimisation, a fresh PCR reaction was performed and the ~3612bp product was added to linearized pCR2.1-TOPO TA cloning vector (details and sequence of pCR2.1 can be found in Appendix 7.1 & 7.2) and used to transform chemically competent *E. coli* TOP10 cells. The resultant mixture was plated out on LB X-gal plates, as stated in Chapter 2, Section 2.5.10. After plating and a 37°C overnight incubation, a total of 39 white colonies were obtained. Six white colonies were selected for colony PCR and further analysis.

To identify cell lines which contain required recombinant plasmid harbouring the cac0383-cac0386 amplified region, colony PCR was performed as described in Chapter 2, Section 2.5.6. The results of the colony PCR are presented in Figure 4.41B. Lanes 1, 2 and 6 all contain strong bands of ~3500bp, consistent with the amplification of the cac0383-cac0386 region. Lanes 3 and 5 contain weaker bands of a similar size and lane 4 contains no PCR product.

Cell lines of *E. coli* TOP10 containing plasmids; pEB400/1, pEB400/2 and pEB400/6 were further analysed. For more accurate analysis of the cell lines purified plasmid preparations were required. Plasmids pEB400/1, pEB400/2 and pEB400/6 were extracted as stated in Chapter 2, Section 2.5.7. A restriction enzyme was then chosen with would asymmetrically cut the plasmid, once in the vector and once in the insert.

The bioinformatics tool RestrictionMapper was used to perform a "virtual" digest using SacI, to obtain the theoretical values for the restriction digest. The theoretical restriction pattern, which would result from the insert ligating within the vector under the control of the T7 promoter, would result in two fragments of 5150 and 2390 (bp). Whereas the restriction pattern obtained if the insert ligated under the control of the Lac promoter would result in two fragments of 6229 and 1312bp.

A restriction digest was carried out using SacI and purified plasmid preparation as stated in Chapter 2. The resultant digest was separated on a 1% agrose gel by electrophoresis the results of the restriction digest are presented in Figure 4.42.

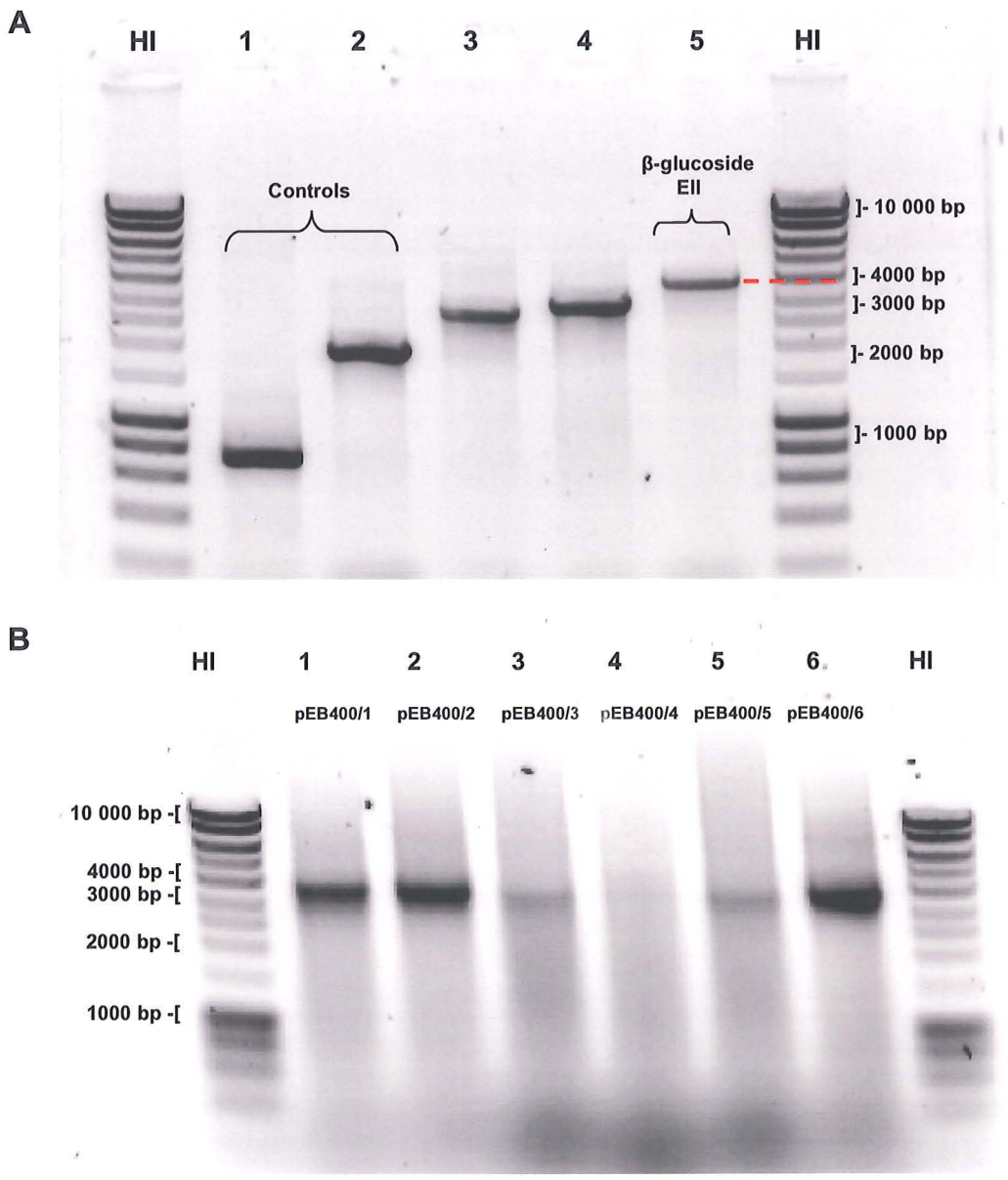


Figure 4.41: A) PCR amplification of the putative β-glucoside EII B) Colony PCR on cells containing pEB400/1-6

Analysed on a 1% agarose gel, using a RunOne electrophoresis unit. Hyperladder I [HI] (Bioline) is the DNA molecular size marker.

A) Lane 1 contains the TOPO control reaction, lane 2 contains the arbutin EII, lane 5 contains the amplified region covering the putative β-glucoside specific (cac0383-cac0386) EII [3612bp].

B) Lanes 1 - 6 contain amplification products from colony PCR of TOP10 cells containing pEB400/1- 6.

The size of the fragments obtained indicate that the digested plasmid in lanes 1 and 2 contain the insert under the control of the Lac promoter and the digested plasmid in lane 3 contains the insert under the control of the T7 promoter. Features of plasmid pEB400/1 are presented in Figure 4.43.

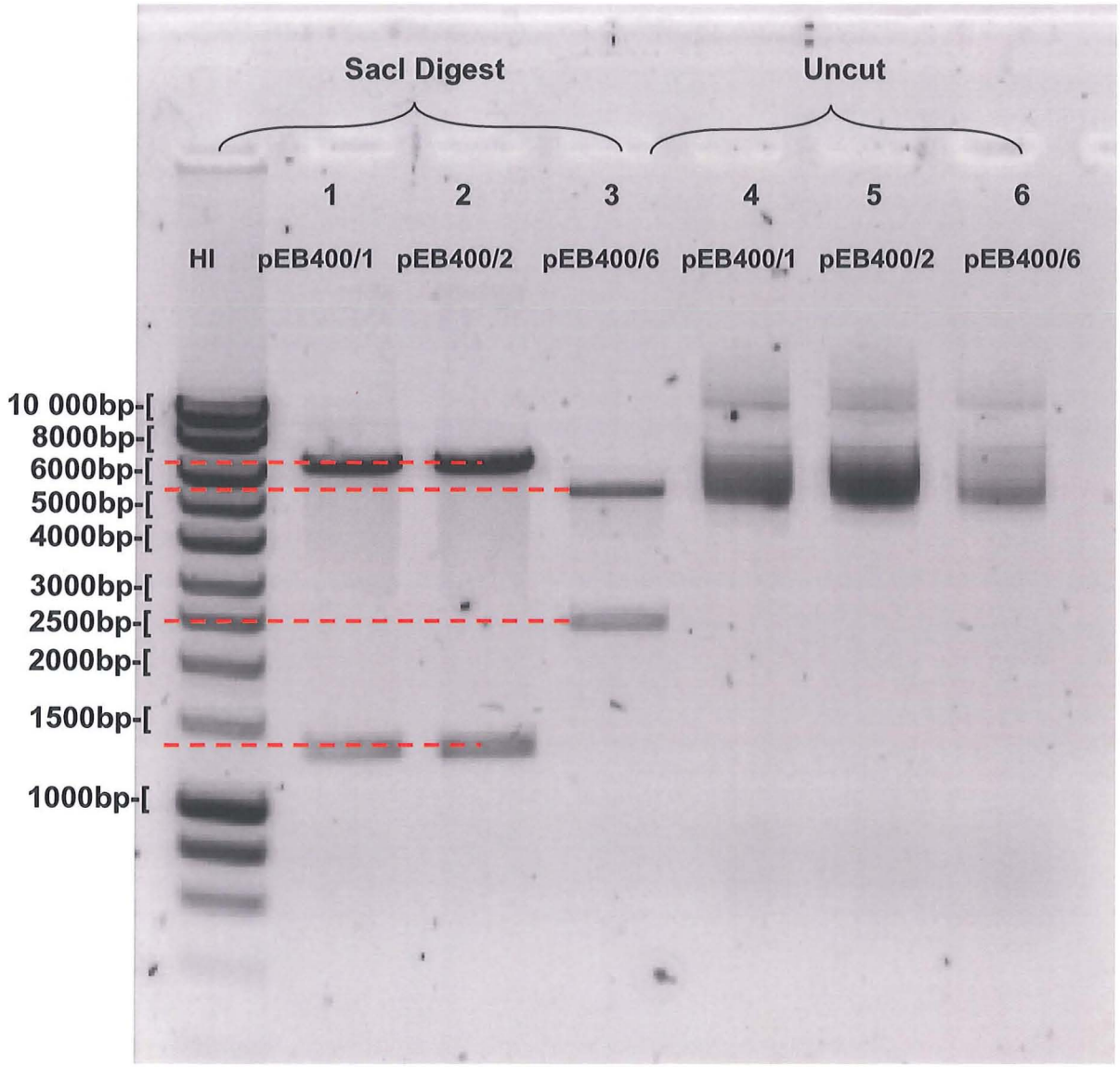


Figure 4.42: Restriction digest of pEB400/1, 2, 6 with SacI

Image of a 1% agarose gel. The DNA molecular weight marker is Bioline Hyperladder I. Lanes 1-3 contains pCELL/1, 2 & 6 digested with SacI (Promega) and lanes 4-6 contains uncut pEB400/1, 2 & 6.

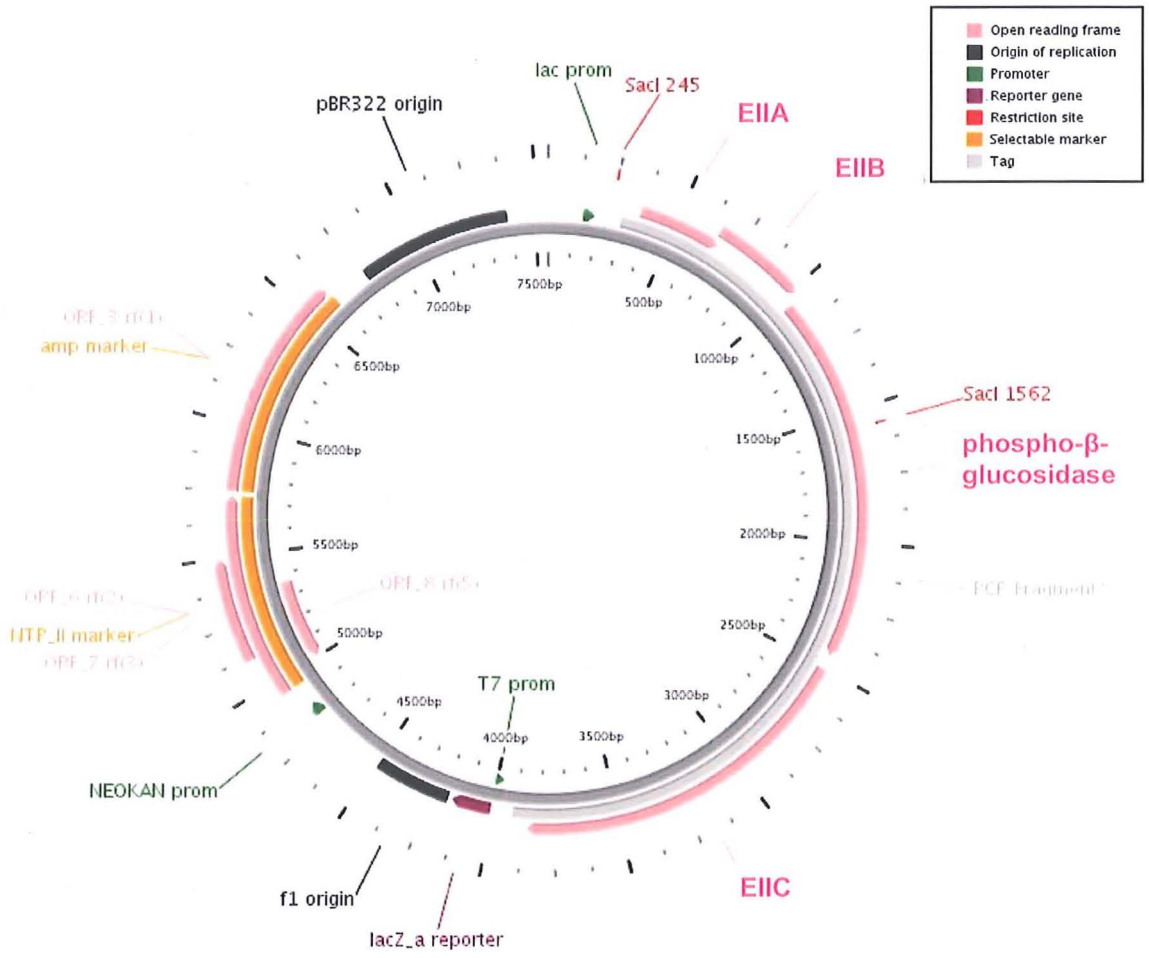


Figure 4.43: Features of plasmid pEB400/1

Important features of the pCR2.1 TOPO TA cloning vector and the amplified fragment containing the putative β-glucoside specific EII domains and a phospho-β-glucosidase.

To identify if any of the plasmids carried the gene for β-glucose transport, transformants of *E. coli* harbouring plasmids pEB400/1, pEB400/2 and pEB400/6 were screened on MacConkey agar supplemented with either glucose, cellobiose, arbutin or salicin. The complementation results are summarised in Table 4.9 and presented in Figure 4.44.

Strain EB400/6 did not produce any colonies. Strains EB400/1 and EB400/2, which both harboured the inserted EII under the control of the *lac* promoter, produced a negative fermentation phenotype for arbutin and cellobiose and a positive fermentation phenotype for salicin.

Table 4.9: EB400/1, 2 & 6 Screening on MacConkey's agar – After 24 Hours

	Arbutin	Salicin	Cellobiose	Glucose
EB400/1	Yellow	Red	Yellow	Red
EB400/2	Yellow	Red	Yellow	Red
EB400/6	-	-	-	-

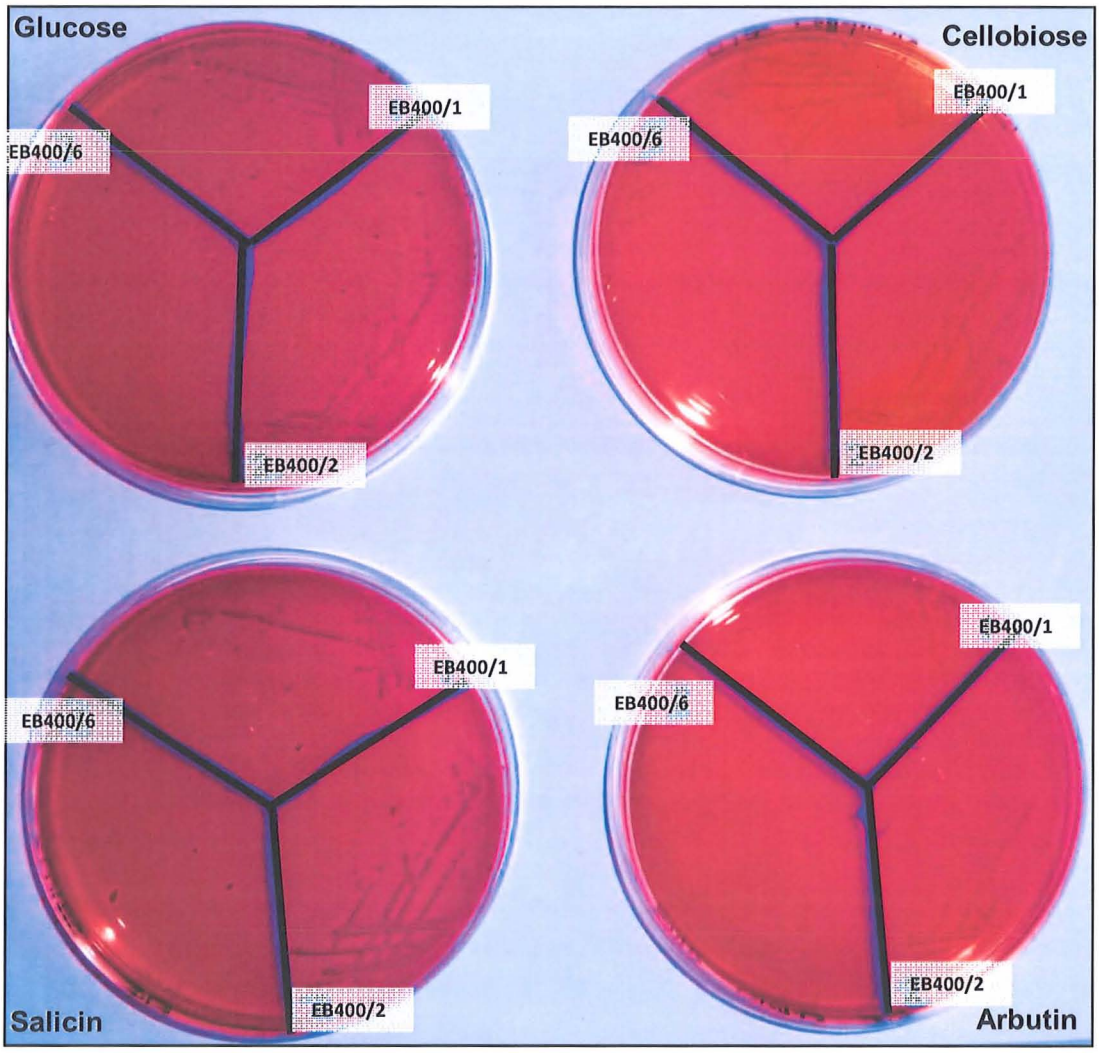


Figure 4.44: Screening complementation analysis

E. coli TOP10 cells transformed with the putative β-glucoside specific EII streaked onto MacConkey agar plates containing either; cellobiose, arbutin, salicin or glucose. Transformants are labelled 1 (pEB400/1), 2 (pEB400/2), & 3 (pEB400/6).

4.7 Discussion

The market price of fermentation feedstock is one of the major considerations in the economics of fermentative biofuel production (Gapes, 2000; European Commission, 2005; Hill *et al.,* 2006). Lignocellulosic plant waste material is an abundant and potentially inexpensive fermentation feedstock. Lignocellulosic plant waste material is composed of lignin, hemicellulose and cellulose. Hydrolysis of lignocellulosic biomass would result in the release of large quantities of plant sugars, such as cellobiose and related β-glucosides. Cellulose is comprised of β-1,4-linked glucose units, therefore hydrolysis of cellulose produces a mixture of glucose, cellobiose and cellooligsaccharides. It has previously been demonstrated that *C. acetobutylicum* has the ability to utilise a wide range of mono- and disaccharides, such as glucose and maltose, which could be present in such plant waste material (Tangney & Mitchell, 2005; Tangney *et al.*, 2001). In fact, it has been demonstrated that *C. acetobutylicum* has the ability to produce solvents by directly utilising the hydrolysate of lignocellulosic biomass (Ezeji & Blaschek, 2007). As demonstrated for the first time in this work *C. acetobutylicum* has the ability to convert cellobiose into acetone, butanol and ethanol (Figure 4.3). Enzymatic hydrolysis, or other pretreatments, of complex fermentation feedstocks to monosaccharides can add a great economic burden to the overall cost of the fermentation process. Cellulolytic enzyme preparations are often used to breakdown cellulose to cellobiose and glucose (Figure 1.9). A second enzyme preparation is frequently applied to breakdown cellobiose to glucose. As *C. acetobutylicum* can directly utilise cellobiose the need for a second expensive enzyme preparation is eliminated.

Carbohydrate accumulation can be viewed as a major bottleneck in the formation of solvents from biomass. As found in Chapter 3, the phosphotransferase system is the predominant method of mono- and disaccharide accumulation in *C. accetobutylicum.* Characterisation of the carbohydrate transport systems of solventogenic clostridia could provide vital information for future strain improvement and maximal feedstock conversion.

The first aim of this chapter was to biochemically characterise aspects of β-glucoside utilisation by *C. acetobutylicum.* Clostridial basal agar, supplemented with a range of individual β-glucosides, were inoculated with *C. acetobutylicum.* Colonies were observed on plates supplemented with cellobiose, arbutin and salicin (Table 4.1). This result is perhaps not surprising as phospho-β-glucosidases and β-glucoside permeases

are widespread in Gram positive bacteria (Schaefler *et al.,* 1969; Faure *et al.,* 2001). However, this is the first report of *C. acetobutylicum* utilising arbutin and salicin.

Upon hydrolysis of plant biomass, the resulting hydrolysate is likely to (amongst other plant sugars) form large quantities of cellooligosacchardes, cellobiose and glucose (Chapter 1, Figure 1.9). Complex gene regulation transpires when bacteria are presented with multiple carbon sources potentially causing certain carbohydrates to be used preferentially (Deutscher *et al.*, 2006). In the artificial environment of a bioreactor, the aim for industry is to convert the maximal amount of feedstock into product, as incomplete utilisation of carbohydrate negatively effects the economic viability of the fermentation process. Bacteria possess the capacity to preferentially utilise one carbon source over another, in a phenomenon called carbon catabolite repression (the proposed mechanism of CCR in low GC Gram positive bacteria is discussed in Chapter 1, Section 1.5.3.1). Solventogenic clostridia appear to preferentially utilise glucose over a range of other carbon sources (Mitchell, 1996; Mitchell *et al.,* 1995). In *C. acetobutylicum,* utilisation of several PTS sugars is subject to repression by glucose, for example sucrose (Tangney & Mitchell, 2000), maltose (Tangney *et al.,* 2001) and lactose (Yu *et al.,* 2007).

The utilisation by *C. acetobutylicum* of a mixture of cellobiose and glucose in minimal broth medium was investigated. In this study utilisation of cellobiose by *C. acetobutylicum* was found to be subject to carbon catabolite repression (CCR) by glucose (Figure 4.2). The cellobiose in the media remained unutilised, until the majority of the glucose was depleted from the media. Preferential utilisation in this manner demonstrates that glucose can regulate cellobiose metabolism in this organism, which has industrial significance, as sequential utilisation of sugars can detrimentally affect bioreactor productivity. All of the proteins required for CCR have been identified in *C. acetobutylicum* including HPrK/P, CcpA and the general PTS proteins, along with many CRE sites (Chapter 3; Rodionov *et al.*, 2001; Behrens *et al.*, 2001; Brückner & Titgemeyer, 2002; Tangney *et al.*, 2003).

The conversion of cellobiose to solvents by *C. acetobutylicum* was investigated (Figure 4.3). Cellobiose was successfully converted to butanol, acetone and ethanol. The fermentations were carried out in triplicate, however several of the cultures lagged and there was a great deal of variation in the replicates. This may be due to the fact that a static fermentation was carried out i.e. no agitation was applied to the vessels other than during sampling. Future fermentations will be carried out in a shaking incubator or in bioreactors with inbuilt mixing. When the error bars were taken into account the

cellobiose and glucose (80g/L) datasets were compatible to the expected ratio of 6:3:1 (Jones & Woods, 1986).

Having established, for the first time that *C. acetobutylicum* can transport and utilise a range of β-glucosides as a carbon and energy source, the mechanism of transport was investigated in this work. In other organisms β-glucosides are predominantly accumulated by the PTS (Brown & Thomson, 1998; Zhang & Aronson, 1994; Tobisch *et al.,* 1997; Schnetz *et al.,* 1996; Murphy *et al.,* 1984; Le Coq *et al.,* 1995; Setlow *et al.,* 2004; Hall & Xu, 1992; Schnetz *et al.,* 1987; Houman *et al.,* 1990). Accumulation in this manner would necessitate the presence of the general PTS proteins, a substrate specific EII and a phospho-β-glucosidase. The general PTS proteins, EI and HPr, have been identified in *C. acetobutylicum* (this work; Mitchell & Tangney, 2005).

In this work, *in silico* analysis of the *C. acetobutylicum* genome revealed two systems involved with β-glucoside utilisation (Chapter 4, Parts I and II). To avoid confusion in this discussion the results presented in Part I will be referred to as System I, the results presented in Part II will be referred to as System II. System I is comprised of three ORFs, with the EII domains all present on a single protein. Whereas, the EII of System II is split over three separate ORFs. The genome context and regulator elements of System I are more analogous to arbutin and salicin utilisation systems in similar organisms. System II shares more features with cellobiose transporting systems of related organisms. Both Systems possess a family 1 p-β-glucosidase encoding gene. A β-glucoside transported via a PTS enters the cell and central metabolism as depicted in Figure 4.45.

Genes encoding phosphotransferase EII permeases are often found associated with genes encoding enzymes required in the first stages of metabolism of the PTS product. For example, in *C. acetobutylicum* the maltose specific EII is found associated with a 6-phospho-α-D-glucosidase, an enzyme capable of hydrolysing the product (maltose-6-phosphate) of maltose transport by PTS (Tangney *et al.,* 2001), while the sucrose specific permease is associated with a sucrose-6-P hydrolase and a frucktokinase (Tangney & Mitchell, 2000) and the lactose specific permease is associated with a 6-P-β-galactokinase (Yu *et al.*, 2007). Additionally, the metabolic and transport genes are often also associated with some sort of regulator, for example upstream from the sucrose permease and metabolic genes, is a transcriptional antiterminator (Chapter 3, Figure 3.2).

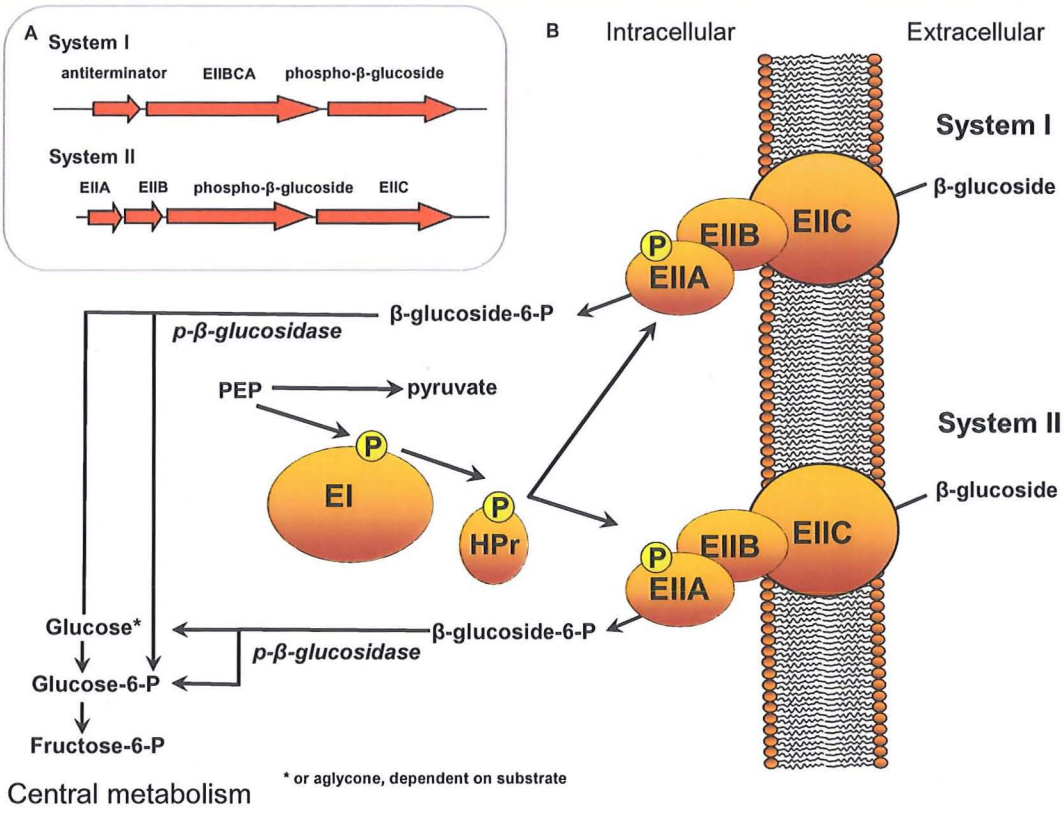


Figure 4.45: Transport and metabolism of β-glucosides by the phosphotransferase

(A) The *C. acetobutylicum* genome was analysed and two potential operons for β-glucoside transport and utilisation were identified (B) the proposed β-glucoside uptake and utilisation pathways encoded by *C. acetobutylicum.*

Accumulation of β-glucosides by the phosphotransferase would most likely proceed as follows; cellobiose, for example, would be translocated across the membrane by the substrate specific EIIC domain and phosphorylated to form cellobiose-6-P. A P-β-glucosidase could hydrolyse the cellobiose-6-P into glucose and glucose-6-P both of which could directly enter the glycolysis pathway. In the case of another β-glucoside with contains a non-glucose moiety, the result would be glucose-6-P and an aglycone.

As detailed in Chapter 3, analysis in this work of the *C. acetobutylicum* genome indicated the presence of thirteen potential PTS permeases and an additional lone EIIA domain. Two candidate systems involved in β-glucoside assimilation and degradation were identified. The two β-glucoside systems were subjected to a more detailed study and the results were presented in this chapter.

PTS permeases can be split into six families (Saier & Tseng 1999). The permease investigated in the System I was found to belong to the glucose-glucoside (Glc) family, located within a 4380bp region containing 3 ORFs (Figure 4.4). The second system's permease belongs to the lactose-N,N'-diacetychitobiose-β-glucoside (Lac) family and comprises of 4 ORFs and covers 3900bp (Figure 4.29), this system (results presented in Part II) will be referred to as System II.

Firstly, before looking in depth at the evidence supporting the identity of each putative protein in system I, a wider look at the operon as a whole is beneficial. This system contains three consecutive ORFs. Putative ribosome binding sites (rbs) with homology to established rbs found in clostridia were identified upstream of each ORF. Bioinformatic analysis revealed that the three ORFs, presented in Figure 4.4, encode an antiterminator protein (BglG), a PTS EII (BglA) and a phospho-β-glucosidase (BglB). These represent all of the proteins required for the regulated transport and initial metabolism of β-glucosides by the phosphotransferase. This arrangement of genes has been previously observed in two other β-glucoside operons, the *C. longisporum abg* operon (Brown & Thomson, 1998) and the *E. coli bgl* operon (Schnetz & Rak, 1988). The three operons are presented together in Figure 4.46 for comparison. The aryl-β-D-glucoside (*abg*) uptake and utilisation system has been characterised in the ruminal *Clostridium* species *C. longisporum* (Brown & Thomson, 1998). The operon consists of three genes *abgG, abgF* and *abgA*. Functions for AbgF and AbgA were assigned on the basis of functional analysis in *E. coli* LP100 and MK120, which revealed that the *abgF* encodes an aryl-β-glucoside (arbutin and salicin) specific PTS EII component and *abgA* encodes a phospho-β-glucosidase specific for hydrolysis of phospho-arbutin and phospho-salicin. An antiterminator (AbgG) was found to be responsible for regulation of the operon, and ribonucleic antiterminator target (RAT) sequences were detected preceding and partially overlapping two rho-independent transcriptional terminators, which were located upstream and downstream of *abgG,* similar to what was found in this work in the *C. acetobutylicum bgl* operon.

The *E. coli* β-glucoside (*bgl*) operon has also been characterised (Schnetz & Rak, 1988). As with the *C. longisporum abg* operon the *E. coli bgl* the operon consists of three consecutive ORF; *bglG, bglF* and *bglB,* encoding an antiterminator protein, an EII (specific for arbutin, salicin and cellobiose) and a phospho-β-glucosidase. However under laboratory conditions these genes are not expressed and the system is cryptic. Significantly, similar regulatory features are present in all three operons, in the form of the presence of ρ-independent terminator structures and ribonucleic antiterminator target sequences (RATs) up- and downstream from the antiterminator gene (Figure 4.46).

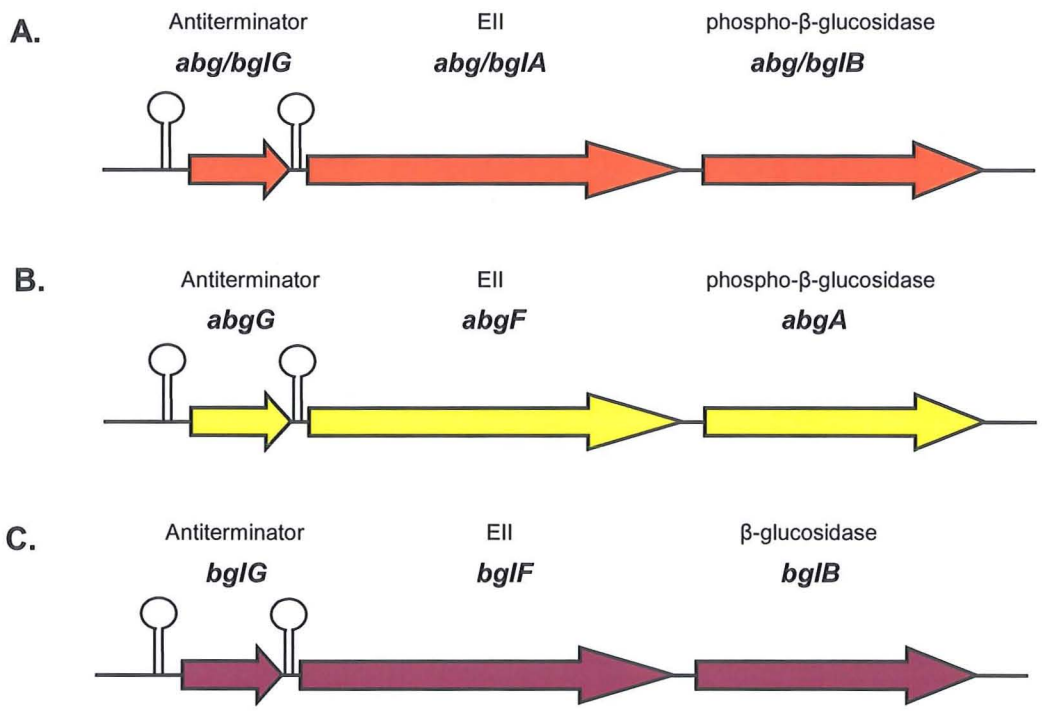


Figure 4.46: Comparison of the arrangements of the ORFs of β-glucoside utilisation operons

ORFs are represented by block arrows (indicating direction of transcript), gene and protein names are given above. Stem loops structures are located upstream and downstream of *bglG/abgG*. A) putative *bgl* operon (this study), B) arbutin & salicin *abg* operon of *C. longisporum* (Brown & Thomson, 1998), C) arbutin, salicin and cellobiose cryptic *E. coli bgl* operon (Schnetz & Rak, 1988).

In most habitats bacteria exist in a constantly changing environment, where conditions can alternate between an abundance of nutrients to near starvation conditions (Roszak & Colwell, 1987). Therefore, for a bacterium to adapt to its dynamic surroundings, delicate sophisticated mechanisms are required to detect the carbon sources available and to the express the relevant array of genes to transport and metabolise them. One level of regulation is mediated by a group of regulatory proteins, called transcriptional activators and transcriptional antiterminators, which can stimulate the expression of genes and operons concerned with carbohydrate transport and metabolism. These proteins contain duplicated PTS regulation domains (PRDs) whose activity is modulated by phosphorylation of conserved histidine residues in response to the availability of carbon sources (Stulke *et al.,* 1998). An exception is the GutR transcriptional activator from *B. subtilis,* required for the expression of the glucitol operon, which has no similarities to any other known transcriptional activator (Ye *et al.*, 1994).

PRD containing regulators can vary in function, size and structure, however all PRD containing regulators possess a short N-terminal effector domain which in the case of antiterminators binds RNA and in the case of activators binds DNA and two regulatory modules referred to as PRD1 and PRD2 (Reizer & Saier, 1997). Many PRD containing regulators have been characterised, such as the transcriptional antiterminators associated with glucose and sucrose specific phosphotransferases in *B. subtilis* GlcT (Bachem & Stulke, 1998), SacY (Tortosa *et al.,* 1997) and SacT (Arnaud *et al.*, 1996). Antiterminators have also been identified associated with β-glucoside specific PTS's for instance, the BglG from *E. coli* (Chen *et al.,* 1997; Gorke & Rak 1999) and LicT from *B. subtilis* (Tortosa *et al.*, 2001; Lindner *et al.*, 1999). Examples of transcriptional activators include MtlR from *B. stearothermophilus* (Henstra *et al.*, 2000), and the LevR (Martin-Verstraete *et al.*, 1998) and LicR activators from *B. subtilis* (Tobisch *et al.*, 1999).

From the analysis of the regulatory features of the *C. acetobutylicum bgl* operon, the BLAST results for the putative antiterminator protein (cac1406) revealed that the highest homology was with transcriptional antiterminator proteins of the BglG family, all belonging to low G/C Gram positive organisms (Table 4.2). The calculated molecular mass of the proteins was very similar and ranged from 32,124Da in *C. acetobutylicum* to 33,003Da in *L. monocytogenes.*

The *E. coli* BglG protein carries two PTS regulation domains (PRD-1 and PRD-2) which each carry conserved histidine residues (PRD-1 His-101 & His-160; PRD-2 His-

208; Tortosa *et al.,* 1997). The multiple alignment of BLAST results revealed considerable homology throughout the sequences of the antiterminator proteins. Unless otherwise stated the similarity groups were disabled while analysing the sequences, this enables a more accurate visualisation of areas of conservation within the sequence. Conserved histidine residues were found at; His-100, His-159, His-210 and His-269. The region upstream of the putative *bglG* gene, the putative antiterminator protein, was examined for regulatory elements. This region was found to contain several discrete regions which resemble known regulatory features. Putative -35 and -10 regions were identified. Overlapping the -35 region is a sequence which shows considerable homology to the consensus CRE sequence of other low GC Gram positive bacteria (Hueck & Hillen, 1995; Miwa *et al.,* 2000). In other low GC Gram positive bacteria the CRE sequence is the target site for catabolite repression via the catabolite control protein CcpA (Hueck & Hillen, 1995). A CcpA homologue has been identified from the *C. acetobutylicum* (Chapter 3; Tangney & Mitchell, 2005).

Upstream from *bglA* a Rho-independent terminator was located (Figure 4.5). Overlapping the ρ-independent terminator is a sequence which demonstrates substantial identity to RAT sequences found in other low GC Gram positive bacteria (Figure 4.7). The 5' and 3' ends of the RAT sequences were found to be well conserved, with most variation occurring in the middle of the sequence. This region has been thought to play a role in the individual specificity of RAT structures (Aymerich & Steinmetz, 1992). A putative ρ-independent terminator and RAT sequence is also found upstream of *bglA* (Figure 4.6).

The associated antiterminator proteins for each of the RAT sequences which were analysed were aligned against the putative BglG, antiterminator protein. Again regions of conservation were observed throughout the length of the protein. Fully conserved histidine residues were discovered at positions; His-100, His-159, His-210 as well as a well conserved His found in five of the six sequences, His-269. The fourth conserved histidine residue was only found in the low GC gram positive bacteria. In conclusion the most reasonable interpretation of the sequence data is that these sequences are RATs which can interact with BglG to prevent termination of transcription in this region.

In order to identify the glucose-glucoside (Glc) family permease of system I, the putative clostridial EII (cac1407) was queried against sequence databases using the BLAST server and the top ten results were examined. All ten results were β-glucoside specific EII components, significantly all belonging to low GC Gram positive organisms (Table 4.3). A phylogram was constructed to investigate the relationships between the clostridial putative EII sequence and published, experimentally verified EIIs, where substrate specificity was known (Figure 4.18). The *C. acetobutylicum* sequence was most closely related to the arbutin specific *Streptococcus mutans* EII (Cote *et al.*, 2000) and the arbutin and salicin specific *C. longisporum* EII (Brown & Thomson, 1998). Both organisms EII proteins have the domain structure BCA, and the overall arrangement to the *C. longisporum* operon and the *C. acetobutylicum* genes are very similar, although gene arrangement is not necessarily significant in terms of functional relationship.

The KyteDoolittle hydrophobicity profile (Kyte & Doolittle, 1982) of the EII was analysed. A large hydrophobic region was identified, flanked by two smaller hydrophilic regions (Figure 4.15). The C domain of a Glc family permease normally possesses between 8 and 12 transmembrane helices (Nguyen *et al.*, 2006) and their presence can be predicted by computer algorithm (Krogh *et al.*, 2001; Sonnhammer *et al.*, 1998). The bioinformatics tool TMHMM uses a Hidden Markov Model to predict transmembrane helices within protein sequences. This model can discriminate between membrane and soluble protein at greater than 99% accuracy. It can also correctly predict the position of 97-98% of transmembrane helices (Krogh *et al.*, 2001). When analysed by TMHMM the putative EII was predicted to be a membrane protein possessing a large hydrophobic region flanked by two hydrophilic domains (Figure 4.14A). The 11 predicted transmembrane helices are contained within the central EIIC domain (residues 90-480), flanked by two hydrophilic regions covering the B and A domains (residues 1-80 and 480-608 respectively). This fully supports the proposal that the domain architecture of the protein is EIIBCA. For comparison, the characterised *C. longisporum* (Brown & Thomson, 1998) EII protein sequence was subjected to the same analysis (Figure 4.14B) and was found to possess 10 transmembrane helices and similar hydrophilic periphery domains.

However, if the *C. acetobutylicum* protein truly possessed an odd number (11) of transmembrane helices, the IIB and IIA domains (which flank the IIC domain) would reside on opposite sides of the membrane, hence phospho-transfer would be impossible. A more logical conclusion would be that the *C. acetobutylicum* protein possesses the

same even number of transmembrane helices as the *C. longisporum* protein. On closer examination of the visual output from TMHMM between amino acids 400 & 440, the programme may have calculated two individual helices to be a single helix (Figure 4.14A, underlined by }). Therefore it is proposed that *C. acetobutylicum* EII permease in fact contains 10 transmembrane helices.

The EII of *Klebsiella oxytoca* has been characterised (Lai *et al.,* 1997). Comparative analysis of the protein sequence of the *K. oxytoca* EII with a large number of other characterised EII permeases has revealed signature sequences for each of the EII domains. Furthermore, a consensus sequence for disaccharide binding was proposed. *K. oxytoca* EII permease has an unusually broad substrate range including cellobiose, arbutin, salicin and methylumbelliferyl derivatives of glucose, cellobiose, mannose and xylose. Homologous sequences to each of the EII domain signatures and the consensus sequence for disaccharide binding were identified in the expected locations of the clostridial EII permease (Lai *et al.,* 1997).

A third gene was found in the operon – a putative 6-phospho-β-glucosidase. As depicted in Figure 4.45, transport of a substrate by the PTS deposits the substrate into the cell in a phosphorylated form. For a phosphorylated β-glucoside to enter central metabolism, the β-1,4 linkage must be hydrolysed, by a substrate specific phospho-β-hydrolase. The result of a BLAST search of the last ORF in the operon indicated that all of the homologous matches were with phospho-β-glucosidases (Table 4.4). In addition, when viewed in a multiple alignment a high degree of conservation is found throughout the entire length of the sequence (Figure 4.16). A phylogenetic tree was constructed with a large selection of experimentally verified, published phospho-β-glucosidases found in association with other β-glucoside EII permeases (Figure 4.37). The clostridial sequence clustered with proteins of the glycosyl hydrolase family 1. Closer investigation of the protein revealed the position of a highly conserved family 1 hydrolase signature sequence (Figure 4.19). The entire operon therefore, contains all of the elements necessary for the regulated transport of β-glucosides.

The PTS is a complex multi protein system, previous studies have demonstrated that PTS components can interact *in vitro* (Mitchell & Tangney, 2005; Tangney & Mitchell, 2005). To investigate whether the identified clostridial permease could indeed transport β-glucosides, it was decided to clone the EII. Cryptic genes are defined as phenotypically silent DNA sequences not normally expressed but in a population. Wild type *E. coli* possesses a cryptic β-glucoside EII, and it cannot utilise any of the β-

glucoside sugars, including cellobiose, arbutin and salicin, as carbon and energy sources under laboratory conditions (Kricker & Hall, 1984). *E. coli* does possess the general enzymes of the PTS (common to all PTS carbohydrates) HPr and EI, which are encoded by the genes *ptsH* and *ptsI* (Reuse & Danchin, 1988), as well as an active phospho-β-glucosidase. Consequently *E. coli* can be transformed with a carbohydrate specific EII component to complete the phosphotransfer chain. This was the rationale for the decision to use *E. coli* as a host for cloning the EII.

The gene was cloned by PCR and ligated in to the vector (pCR2.1). Blue/white screening was used to identify transformed *E. coli* TOP10 cells containing the cloned gene (Figure 4.23). TOPO cloning is however, non-directional – i.e. an inserted gene may be cloned into the vector in either orientation. The orientation of the DNA inserted in the vector can be readily determined by two methods; (i) a restriction digest using an enzyme which cuts once within the vector and once asymmetrically within the inserted gene (and the orientation can then be deduced by the sizes of the resultant bands) or (ii) by a directional PCR utilising the original cloning primers and a primer specific for part of the vector, where the orientation is determined by the presence or absence of a band. The plasmid pEB100 possesses two HindIII sites, one in the vector and one asymmetrically placed within the inserted gene. Figure 4.26 shows a restriction digest of plasmids harbouring the inserted gene in each orientation i.e. under the control of the Lac and T7 promoters. The plasmid containing the EII permease under the control of the Lac promoter was designated pEB100 and a map of the plasmid can be found in Figure 4.27.

The strain *E. coli* EB100 was screened for complementation of the *E. coli* cryptic *bgl* PTS. MacConkey agar was used as the basis of a complementation screen. MacConkey's usually contains lactose and peptone as carbon sources, supplemented with neutral red as a pH indicator and crystal violet to inhibit the growth of Gram positive organisms. Utilisation of the peptone results in the production of a small amount of ammonium, thus raising the pH, causing the agar to adopt a yellow/white colour and the colonies to appear whitish. However, fermentation of the lactose results in acid production and a drop in the pH, causing the colonies to adopt a red colouration and the agar remains red. Substituting the lactose for β-glucoside provides a medium for detection of β-glucoside utilisation.

E. coli strains EB100 and EB101, were streaked onto MacConkey agar supplemented with either glucose, cellobiose, arbutin or salicin (Figure 4.28). *E. coli* EB100 (which

harbours the plasmid pEB100 containing the putative *bglA* gene) had a negative fermentation phenotype for cellobiose and salicin and a positive fermentation phenotype with glucose and arbutin. Unlike the glucose positive control the colonies on the arbutin containing plate appeared dark red/brown and the surrounding agar darkened to deep terracotta.

A possible source of a false positive fermentation phenotype from the transformants could arise from unintentional hydrolysis of the arbutin molecule during sterilisation, which would result in production of glucose which could support the growth of the organism. To prevent this, the arbutin solution was filter sterilised instead of autoclaving and transformants were plated next to transformants negative for expression of the putative EII. This ensured that any discolouration observed was due to the active bacterial hydrolysis of arbutin rather than the natural break down of the substrate.

The dark red/brown colonies and discolouration of the agar (rather than the red colonies expected for a positive fermentation phenotype) may be explained by the release of hydroquinone. As the hydrolysis of arbutin would most likely result in glucose and hydroquinone (Figure 4.47). Brown colonies on MacConkey-arbutin agar plates have previously been reported as a method of qualitatively estimating arbutin utilisation by *E. coli* (Monderer-Rothkoff & Amster-Choder, 2007).

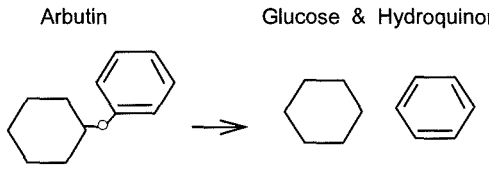


Figure 4.47: Hydrolysis of Arbutin

The hydrolysis of arbutin would most likely result in glucose and hydroquinone – a chromogenic substance (adapted from Clifford, 2000).

The results of the complementation screening suggests that the clostridial permease is interacting with the *E. coli* PTS to transport arbutin and that the permease specificity is active for arbutin and not the related β-glucosides cellobiose and salicin. It therefore appears that System I is arbutin specific. System II has a different composition.

It is widely accepted that the PTS is a highly mosaic system (Nguyen *et al.*, 2006), EII proteins possess at least three (or four as in the case of the Man family of permeases) functionally distinct domains, which may be all encoded in a single ORF or spread over

several independent ORF. The hydrophilic A and B domains are involved in phosphotransfer while the hydrophobic transmembrane C domain is responsible for substrate translocation (Barabote & Saier, 2005). The order of these domains is highly variable, making direct comparison of permeases with dissimilar domain structure problematic. Unlike System I the Lac family permease (cac0383, cac0384 and cac0386) is spread over three ORFs (Figure 4.29). Therefore, in a second set of analyses, a selection of experimentally verified β-glucoside specific EII permeases (Table 4.6 & Table 4.7), as well as the clostridial EII (cac1407) investigated in Part I, were collected and split into their individual domains – allowing the putative protein sequence to be compared to other EIIs with similar substrate specificity but different domain order, presented in Figure 4.32, Figure 4.34 & Figure 4.36.

The EIIC domain endows the permease with its substrate specificity (Postma *et al.*, 1992; Barabote & Saier, 2005). As deduced from the available literature, the substrate(s) that each permease is known to transport was mapped against the sequences in the unrooted phylogenetic tree (Figure 4.36). The Lac family permease cac0386 falls on a branch with two cellobiose permeases from *Lactobacillus* species. An earlier offshoot of that branch harbours two *Bacillus* permeases capable of cellobiose and lichenan (a linear, 1,3:1,4-β-D glucan) transport. However, discretion must be used when drawing conclusions from Figure 4.36, as the literature may not state all of the substrates tested. Therefore, some of the permeases may be able to transport additional substrates not included in the reports. Additionally, substrates which were not transported may not be reported.

The operon structure of System II has similarities to that of other cellobiose transporting systems, and resides in the genome in association with a phospho-β-glucosidase, suggesting that this Lac family permease is not involved in lactose transport. These facts, coupled with the homology of individual domains to other characterised cellobiose permeases strongly suggest that these disparate ORFs together encode a cellobiose specific EII. As previously discussed permeases are often found associated with enzymes specific for the metabolism of the permease substrate. In the case of cellobiose (or any other β-glucoside) a phospho-β-glucosidase would be required to hydrolyse the phosphorylated disaccharide. *B. subtilis* possesses four characterised phospho-β-glucosides; BglA, BglH, YckE, YdhP, which account for >99.9% of the glucosidase activity on aryl-phospho-β-glucosides (Setlow *et al.*, 2004). The *B. subtilis bglA* gene does not appear to be associated with any phosphotransferase genes (Kasahara *et al.*, 1997; Zhang & Aronson, 1994), whereas, BglH is in an operon *bglPH*,

which encodes an aryl-β-glucosidase specific EII permease and a phospho-β-glucosidase. This operon is subject to carbon catabolite repression and is subject to both termination and antitermination (Le Coq *et al.*, 1995; Krüger & Hecker, 1995). Expression of *yckE* and *ydhP* were not induced by aryl-β-glycosides and while *yckE* was expressed at a constant low level during growth, sporulation and spore germination, *ydhP* was expressed at high levels at the initiation of stationary phase and at very low levels during growth and germinating spores (Setlow *et al.*, 2004). A fifth phospho-β-glucosidase (*licH*) has been identified from the *B. subtilis* genome which is induced by cellobiose, lichenan hydrolysate and lichenan. It resides in the operon *licRBCAH*, encoding a regulator, PTS permease domains EIIB, EIIC and EIIA and finally a phospho-β-glucosidase (Yoshida *et al.*, 1996; Tobisch *et al.*, 1997). Interestingly a *licH* mutant could still grow on lichenan hydrolysis products and cellobiose, indicating one of the other phospho-β-glucosidases has overlapping substrate specificity (Tobisch *et al.*, 1997).

Phylogeny of the cac0385 sequence puts it loosely in a cluster with the family 1 hydrolase group of enzymes (Figure 4.37). Analysis of the genome revealed that *C. acetobutylicum* has only two phospho-β-glucosidases associated with PTS proteins, cac1408 and cac0385. Akin with the *B. subtilis* genome there is a third putative phospho-β-glucosidase (cac0742; Q97L23) not associated with any PTS proteins. It would be of interest to determine the substrate specificity of this lone phospho-β-glucosidase and also to ascertain if it is constitutively expressed, or regulated in the presence of β-glucosides.

A putative regulator, which resembles a transcriptional activator, was identified upstream of the EII and phospho-β-glucosidase encoding genes of system II. Compared to transcriptional antiterminators, transcriptional activators have much more variation in the modular structure of their domains and possess functional domains homologous to other protein families (Tilbeurgh & Declerck, 2001). For example, the MtlR transcriptional activator possesses a C-terminal domain similar to an EIIA PTS component. This domain was found to be phosphorylated by $EIICB^{mtl}$ (Henstra *et al.*, 2000). The large LevR protein from *B. subtilis* contains a NifA-like domain, situated between the HTH DNA binding motif and the PRDs, which can interact with the σ^{54}-factor of the RNA polymerase (Martin-Verstraete *et al.*, 1994).

The region upstream of the putative β-glucoside EII was examined for potential regulatory sequences. An ORF (cac0382), reading in the opposite direction, was discovered, which upon a BLAST search was found to bear homology to a possible transcriptional activator containing two fused BglG-like domains (Table 4.8). A phylogenetic tree was constructed which included the putative regulator (cac0382) and a representative selection of sequences of characterised regulators found associated with β-glucoside utilisation operons (Figure 4.38). The sequence of cac0382 was found to be most closely associated with the *B. subtilis* PRD-containing LicR, although cac0382 and LicR do not appear to have any conserved domains in common. However, it has been previously stated that the domain structure of LicR is unusual (Tobisch *et al.*, 1997), which could explain why the sequences do not align. As the BLAST results suggested that there may be homology to a transcriptional activator, the sequence was then aligned with the transcriptional activator LevR from *B. subtilis* (Figure 4.39B). Large blocks of conservation, especially in the N-terminal of the protein, can be observed. The domain structure of the three proteins sequences were identified and presented in Figure 4.40. It can be clearly seen that cac0382 and LevR share similar domain structure.

A search was then carried out to identify additional transcriptional activators with associated with cellobiose utilisation. The truncated CelR from *B. stearothermophilus* (Lai & Ingram, 1993) was aligned with cac0382. Large blocks of conservation are found in the available sequence, especially around the PRDs (Figure 4.39C). It would be extremely interesting to clone and sequence the entire *B. stearothermophilus celR* gene to compare the full sequence with the clostridial sequence. The evidence suggests that cac0382 encodes a transcriptional activator, which can respond to the presence of substrate to induce gene expression.

The region which contains the Lac family β-glucoside EII of system II spans 3495bp, and each of the three EII domains is present on separate ORFs, with the ORF encoding the putative phospho-β-glucosidase interjected between the EIIB and EIIC domains (Figure 4.29). This contrasts with the putative Glc family permease of system I, in which all three domains are present in a single ORF with the phospho-β-glucosidase residing downstream (Figure 4.4).

The size of the fragment to be amplified presented specific problems. Specific long-range Taq polymerase was employed to produce a PCR product for cloning (Figure 4.41). Cloning of large PCR fragments can reduce the transformation efficiency of TOPO TA cloning reactions (Invitrogen, 2006). Following optimisation, transformants were nevertheless obtained.

The transformants were screened on MacConkeys agar supplemented with either glucose, cellobiose, salicin or arbutin. *E. coli* EB400/6 produced no colonies on the minimal media. EB400/1 and EB400/2 produced a positive fermentation phenotype for salicin (Figure 4.44), yet a negative fermentation phenotype for arbutin and cellobiose. The inability of the transformant to transport cellobiose was surprising. System II is homologous to other cellobiose transporting systems. *C. acetobutylicum* can transport and utilise cellobiose and no *in silico* evidence supports the presence of a different transport system for cellobiose encoded within the genome (data not shown). Possible reasons for this are discussed in Chapter 6.

C. acetobutylicum possesses two distinctive transport systems for β-glucoside accumulation, allowing the organism to ferment complex β-glucoside rich waste biomass streams - such as pear or berry pulp left over from the juice industry. Genetic manipulation of the identified regulatory domains could provide co-current utilisation of β-glucosides and glucose thus potentially optimising the carbon accumulation step of the industrial solvent fermentation process.

Chapter 5

5 Transport of hemicellulose monomers

5.1 Introduction

A vitally important factor in the success of future bioenergy plants is identifying economically attractive feedstocks for bioconversion. A great deal of research has been focused on identifying cost effective candidates, such as lignocellulosic biomass which is cheap and abundant (Rani, *et al.*, 1998; Krishna, *et al.*, 1998; Ballesteros, *et al.*, 2004; Zaldivar, *et al.*, 2001). Lignocellulosic plant waste biomass is composed of up to 50% cellulose, 30% hemicellulose and 20% lignin (Gray *et al.,* 2006). Hemicellulose has been widely acknowledged as a potentially important substrate for ABE fermentation (Hazelwood & Gilbert, 1993; Saha, 2003). Hemicelluloses are heteropolysaccharides, consisting of pentoses (xylose, arabinose), hexoses (mannose, glucose, galactose) and sugar acids (Saha, 2003). *C. acetobutylicum* can ferment a wide range of the monomers present in hemicellulose to biosolvents and hydrogen. Glucose regulation, transport and metabolism has previously been characterised in *C. acetobutylicum* (Tangney & Mitchell, 2007; Nölling *et al.*, 2001). However, the transport systems for the critical industrial substrates including; xylose, arabinose, mannose and galactose remain to be elucidated.

Pentose utilisation and transport genes have been characterised in some low G/C Gram positive bacteria, such as the xylose uptake genes in *Bacillus megaterium* (Schmiedel *et al.,* 1997), xylose utilisation in *B. subtilis* (Wilhelm & Hollenberg, 1985) and arabinose utilisation in *B. subtilis* (Sa-Nogueira & De Lencastre, 1989; Sa-Nogueira *et al.,* 1997). Generally, in *Bacillus,* pentose transport into the cell occurs via proton-sugar symporter. D-xylose is converted to D-xylulose by xylose isomerase which is then converted to D-xylulose-5P by **xylulose kinase.** L-arabinose is converted to L-ribulose by **arabinose isomerase** which is then converted to L-ribulose-6P by ribulokinase, and then to D-xylulose-5P by **L-ribulose 5 phosphate 4 epimerase.** D-xylulose-5P is converted to glyceraldehyde 3-phosphate, ribose-5P and sedoheptulose 7-phosphate by **transketolase**, then to erythrose4-P and fructose 6-phosphate by **transaldolase** (enzymes which are in bold also appear in Table 5.1). The xylose isomerase (XylA)-encoding gene (*xylA*) from the thermophile *C. thermosaccharolyticum* NCIB 9385 has been characterised (Meaden *et al.,* 1994). D-xylose isomerase (EC 5.3.1.5) catalyzes the reversible isomerization of D-xylose to D-xylulose in the first step of xylose metabolism to fructose-6P, before entry into the cells central metabolism (Sanchez & Smiley, 1975). The proposed *Bacillus* pathways for xylose and arabinose utilisation are depicted in Figure 5.1.

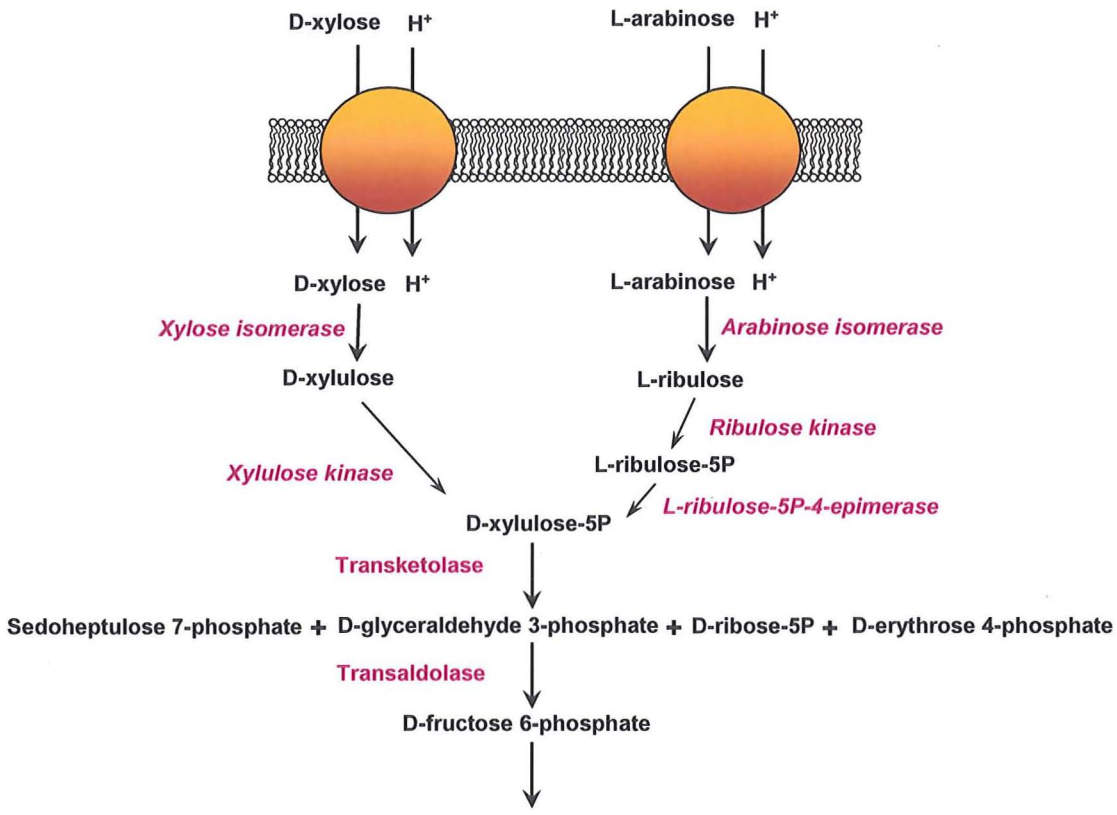


Figure 5.1: The pathway of xylose and arabinose transport and catabolism in *Bacillus subtilis*

Information collated from: Schmiedel *et al.,* 1997; Wilhelm & Hollenberg, 1985; Sa-Nogueira & De Lencastre, 1989; Sa-Nogueira *et al.,* 1997.

Clostridium acetobutylicum ATCC 824 has been shown to metabolise xylose and arabinose resulting in the production of solvents, mostly consisting of butanol (Ounine *et al.,* 1985). A division of ongoing collaborative research conducted within the BfRC has been focused on elucidating the uptake and utilisation mechanisms of xylose and arabinose in *C. acetobutylicum* ATCC 824.

The pathway for xylose and arabinose metabolism has been elucidated in the related Firmicute *B. subtilis* and is presented in Figure 5.1. The isomerisation of xylose and arabinose by their relevant isomerases is the first step in the interconversion of the pentoses into central metabolism. Therefore, the aim of this subsection of investigation is to seek out and characterise potential new putative pentose isomerases from the *C. acetobutylicum* genome. This will be initiated by *in silico* analysis of the clostridial homologues found in the model *Bacillus spp.* xylose and arabinose utilisation pathways, followed by heterologous expression of the putative genes in relevant *Escherichia coli* mutants.

5.2 Pentose transport in *C. acetobutylicum*

The *C. acetobutylicum* genome was analysed as stated in Chapter 2 using the GIB server (Fumoto *et al.*, 2002) and the ORF finder tool at the NCBI. A keyword search was performed for xylose isomerase in the *C. acetobutylicum* genome annotation, which yielded no positive matches. To ascertain whether an unannotated xylose isomerise homologue existed within the *C. acetobutylicum* genome, a BLAST2 search was carried out using the characterised XylA sequence from *C. thermosaccharolyticum* (Meaden *et al.*, 1994; accession number M91248). No significant sequences matches were identified (data not shown). It has been reported previously that arabinose isomerases can also exhibit activity for D-xylose and D-fucose (Patrick & Lee, 1968). A search for arabinose isomerase yielded two putative genes with the locus tags cac1342 and cac1346.

The genome region containing the two genes was examined using the KEGG Genome Map of *C. acetobutylicum* (Kanehisa & Goto, 2000) and the ORFs surrounding the two candidate genes were investigated and presented in Table 5.1. The surrounding genes putatively encode homologs of several of the components involved in xylose / arabinose metabolism in *Bacillus* (highlighted in bold, Figure 5.1). A topological map of the genome region is presented in Figure 5.2.

Table 5.1: Examination of the area surrounding the putative isomerase genes

Locus tag		Putative protein
cac1339	*araE*	Possible sugar-proton symporter
cac1340	*araR*	Transcriptional regulator of the LacI family
cac1341	*araD*	Ribulose-5-phosphate 4-epimerase family protein
cac1342	***araA***	**L-arabinose isomerase**
cac1343	TK	Transketolase
cac1344	*xylB*	Sugar kinase, possible xylulose kinase
cac1345	*xylT*	D-xylose-proton symporter
cac1346	***araA***	**L-arabinose isomerase**
cac1347	TA	Transaldolase
cac1348	TK	Transketolase
cac1349	*galM*	Aldose-1-epimerase

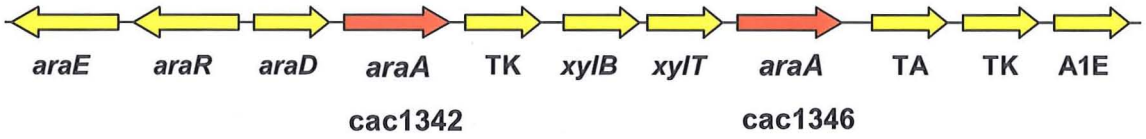


Figure 5.2: Organisation of the pentose isomerase containing genome region

The putative identities of the genes are as follows: ***araE*** - sugar-proton symporter, ***araR*** - arabinose operon transcriptional repressor, ***araD*** - ribulose 5-phosphate 4-epimerase, ***araA*** and ***araA2*** - L-arabinose isomerase, **PK**- phosphoketolase, ***xylB*** - sugar kinase xylulose kinase, ***xylE*** - D-xylose proton symporter, **TA** - transaldolase, **TK** - transketolase, **A1E** - aldose-1 epimerase.

The regions upstream of the ORFs cac1342 and cac1346 were examined for putative regulatory regions, by aligning the upstream regions with CRE sequences found in other clostridia. Consensus CRE query sequence is as follows: WTGNAANCGNWNNCW (N and W stand for any base and A or T, respectively; Miwa *et al.*, 2000). Sequences which bear homology to CRE sequences found in other solventogenic clostridia catabolic operons were identified just upstream of each ORF. These were aligned and the results are presented in Figure 5.3.

Figure 5.3: Comparison of putative CRE sequences found in clostrida

CRE sequences identified in other solventogenic clostridia catabolic operons were aligned against the regions upstream of cac1342 and cac1436.

Clostridium acetobutylicum DSM 792 mannitol operon (cacManCRE; Behrens *et al.*, 2001), *C. acetobutylicum* ATCC 824 sucrose operon (cacScrCRE; Tangney & Mitchell, 2000), *C. acetobutylicum* ATCC 824 maltose operon (cacMalCRE; Tangney *et al.*, 2001).

5.2.1 *In silico* analysis of the putative pentose isomerase gene (cac1342)

Further analysis was conducted on the *araA/xylA* sequence (cac1342). The sequence was input into the ORF finder tool at the NCBI and the results are shown in (Figure 5.4). The deduced amino acid sequence of the putative ORF was placed below the first nucleotide of the corresponding codon. Putative Shine-Dalgerno sequences for ribosomal binding (RBS) are underlined and labelled. Stop codons are indicated by asterisks. A putative CRE sequence was identified (TGCAAATGCATATTA) and is highlighted. A putative ribosome binding site (GAGGA) was identified 7 nucleotides upstream from the putative start codon of cac1342. The gene is 1467bp long and encodes a protein of 488aa with a predicted molecular weight of 55,206Da.

To determine if the sequence shared homology with other sequences deposited in the NCBI database, the deduced amino acid sequence was input into BLAST and the results tabulated in Table 5.2. The deduced amino acid sequence of cac1342 was found to share considerable homology (83% identity, 91 % positive) to cac1346. All of the proteins encoded by the genes returned in the top n BLAST results were putative L-arabinose isomerase genes.

Figure 5.4: Nucleotide and deduced amino acid sequence of the *C. acetobutylicum* putative xylose isomerase gene *xylA* (cac1342)

The deduced amino acid sequence of the putative ORF was placed below the first nucleotide of the corresponding codon. Putative Shine-Dalgerno sequences for ribosomal binding (RBS) are underlined and labelled. Stop codons are indicated by asterisks. A possible CRE sequence is highlighted.

```
acatggtgcaaatgcatattatggtcaaaaaacaaagtaagaaaagaggatataaaa
                                            rbs

   1 - ATGCTAAAGAATAAAAAATTAGAATTTTGGTTTGTAGTAGGTAGTCAGAATTTATACGGTGAAGAAGCATTAAAT - 75
   1 - M  L  K  N  K  K  L  E  F  W  F  V  V  G  S  Q  N  L  Y  G  E  E  A  L  N   - 25

  76 - GCAGTAAAAAAAGATTCTAAAGAAATTGTGGATTCTTTAAATGAAAGTGGAAAATTACCATATCCTATTGTATTT - 150
  26 - A  V  K  K  D  S  K  E  I  V  D  S  L  N  E  S  G  K  L  P  Y  P  I  V  F   - 50

 151 - AAAACTCTAGCTACATCAGCAGATGAAATTAAGAATATAGTTAAAGAAATAAATTATAGAGATGAAGTAGCAGGT - 225
  51 - K  T  L  A  T  S  A  D  E  I  K  N  I  V  K  E  I  N  Y  R  D  E  V  A  G   - 75

 226 - GTTATTACTTGGATGCATACATTTTCTCCTGCTAAAATGTGGATAGCAGGAACAAAGCTTTTACAAAAACCACTA - 300
  76 - V  I  T  W  M  H  T  F  S  P  A  K  M  W  I  A  G  T  K  L  L  Q  K  P  L   - 100

 301 - TTGCATCTAGCAACACAATTTAACGAAAATATTCCTTGGAAAACTATTGATATGGATTATATGAATTTACATCAA - 375
 101 - L  H  L  A  T  Q  F  N  E  N  I  P  W  K  T  I  D  M  D  Y  M  N  L  H  Q   - 125

 376 - AGTGCACATGGTGATAGGGAATATGGATTTATTAATGCTAGATTAAACAAAAATAACAAAGTTGTTGTAGGATAT - 450
 126 - S  A  H  G  D  R  E  Y  G  F  I  N  A  R  L  N  K  N  N  K  V  V  V  G  Y   - 150

 451 - TGGAAGGATAATCAAGTTCAAAAGGAAATTGCAGAGTGGATGCAGGTTGCTTATGGGTACGTTGCAAGTGAAAAT - 525
 151 - W  K  D  N  Q  V  Q  K  E  I  A  E  W  M  Q  V  A  Y  G  Y  V  A  S  E  N   - 175

 526 - ATAAAGGTTGCAAGATTTGGGGATAACATGCGTAATGTTGCCGTAACAGAAGGAGATAAAGTAGAAGCCCAAATT - 600
 176 - I  K  V  A  R  F  G  D  N  M  R  N  V  A  V  T  E  G  D  K  V  E  A  Q  I   - 200

 601 - CAATTTGGATGGACAGTTGATTATTTTGCTATTGGTGATTTAGTGGCTGAAATGAACAAAGTTTCACAAAAAGAT - 675
 201 - Q  F  G  W  T  V  D  Y  F  A  I  G  D  L  V  A  E  M  N  K  V  S  Q  K  D   - 225

 676 - ATAGATGCTACTTATGAAGAGTTTAAGGATATTTATATATTAGATATTGGAGATAATGACCCTGAATTTTATGAG - 750
 226 - I  D  A  T  Y  E  E  F  K  D  I  Y  I  L  D  I  G  D  N  D  P  E  F  Y  E   - 250

 751 - AATCATGTAAAAGAACAAATTAAGATAGAGATAGGTTTACGTAATTTTCTAGAGGCAGGTAATTATACAGCATTT - 825
 251 - N  H  V  K  E  Q  I  K  I  E  I  G  L  R  N  F  L  E  A  G  N  Y  T  A  F   - 275

 826 - ACAACAAACTTTGAGGATCTTTATGGAATGAAGCAATTACCTGGACTTGCAGTTCAACGTTTAAATGCTGAAGGT - 900
 276 - T  T  N  F  E  D  L  Y  G  M  K  Q  L  P  G  L  A  V  Q  R  L  N  A  E  G   - 300

 901 - TATGGTTTTGCAGGTGAAGGCGATTGGAAAACAGCAGCACTTAATCGTTTATTTAAAATTATGACTGACAACAAG - 975
 301 - Y  G  F  A  G  E  G  D  W  K  T  A  A  L  N  R  L  F  K  I  M  T  D  N  K   - 325

 976 - AAAACTGGATTTATGGAAGACTATACTTATGAGCTAAGTGCTGGAAATGAAAGAATTTTGGGAGCACATATGCTA - 1050
 326 - K  T  G  F  M  E  D  Y  T  Y  E  L  S  A  G  N  E  R  I  L  G  A  H  M  L   - 350

1051 - GAAGTTGATCCAACACTTGCAGCTAGTAAACCAAGAGTTGTAGTTAAACCACTTGGAATTGGAGATAAGGAAGCA - 1125
 351 - E  V  D  P  T  L  A  A  S  K  P  R  V  V  V  K  P  L  G  I  G  D  K  E      - 375

1126 - CCAGCACGTTTAATATTTGATGGAGTTGTAGGTGATGGAGTAGTTGTATCTATGCTTGATTTAGGGACACACTAT - 1200
 376 - P  A  R  L  I  F  D  G  V  V  G  D  G  V  V  V  S  M  L  D  L  G  T  H  Y   - 400

1201 - CGTTTACTTATTAACGAAGTAAAGGCAGTTAAACCTACTGAGGATGCTCCTAATTTACCTGTAGCAAAGCTAGTA - 1275
 401 - R  L  L  I  N  E  V  K  A  V  K  P  T  E  D  A  P  N  L  P  V  A  K  L  V   - 425

1276 - TGGCAGCCACAACCAAACTTTAAAGATGCAGTTAAAGCATGGATTTATGCTGGAGGTGGACATCATACTGTTGCA - 1350
 426 - W  Q  P  Q  P  N  F  K  D  A  V  K  A  W  I  Y  A  G  G  G  H  H  T  V  A   - 450

1351 - ACCTTAGAATTAACAGTTGAGCAAGTTTATGACTGGAGTCGTATGGTTGGTTTAGAAACAATAGTTATTGATCAT - 1425
 451 - T  L  E  L  T  V  E  Q  V  Y  D  W  S  R  M  V  G  L  E  T  I  V  I  D  H   - 475

1426 - AATACCAATTTAAGAGATATTATAAAAGAGACTTCAAGATAA                                    - 1467
 476 - N  T  N  L  R  D  I  I  K  E  T  S  R  *  X                                  - 500
```

Table 5.2: BLAST homology results for the deduced amino acid sequence of the putative *araA/xylA* (cac1342).

Results of a BLAST homology search with the deduced amino acid sequence of the putative pentose isomerase cac1342.

Species	Putative gene	% Identity	% Positive	Predicted molecular mass	ORF Locus tag	Accession
Clostridium acetobutylicum ATCC 824	L-arabinose isomerase	83	91	55357	CAC1346	NP_347974
Carnobacterium sp. AT7	L-arabinose isomerase	78	89	53088	CAT7_10260	T7_10260
Bacillus coagulans 36D1	L-arabinose isomerase	75	86	53439	BcoaDRAFT_0992	ZP_01695782
Bacillus licheniformis ATCC 14580	L-arabinose isomerase	72	83	53374	BL01182	YP_079230
Bacillus pumilus ATCC 7061	L-arabinose isomerase	67	79	53387	BAT_1675	ZP_03055726
Lactobacillus brevis ATCC 367	L-arabinose isomerase	64	76	53162	LVIS_1740	YP_795838
Pediococcus pentosaceus ATCC 25745	L-arabinose isomerase	64	76	53925	PEPE_0165	YP_803714
Bacillus licheniformis ATCC 14580	L-arabinose isomerase	61	74	55517	BL00352	YP_080173
Oenococcus oeni ATCC BAA-1163	L-arabinose isomerase	61	76	53668	OENOO_52035	ZP_01544136
Clostridium beijerinckii NCIMB 8052	L-arabinose isomerase	61	73	54182	Cbei_4457	YP_001311522

To visually explore the similarities of the sequences from the BLAST output the sequences were aligned in ClustalW2 and then analyzed in GeneDoc, the results are presented in Figure 5.5. Significant blocks of conservation were observed, especially around residues 75-90 and 279-320. The ClustalW alignment was then input into the TreeView programme to identify the phylogenetic groups the sequences fall into (Figure 5.6). The cac1342 and cac1346 form a cluster together. At the root of that branch is a cluster of two *Bacillus* sequences and the *Carnobacterium* sp. AT7 sequence.

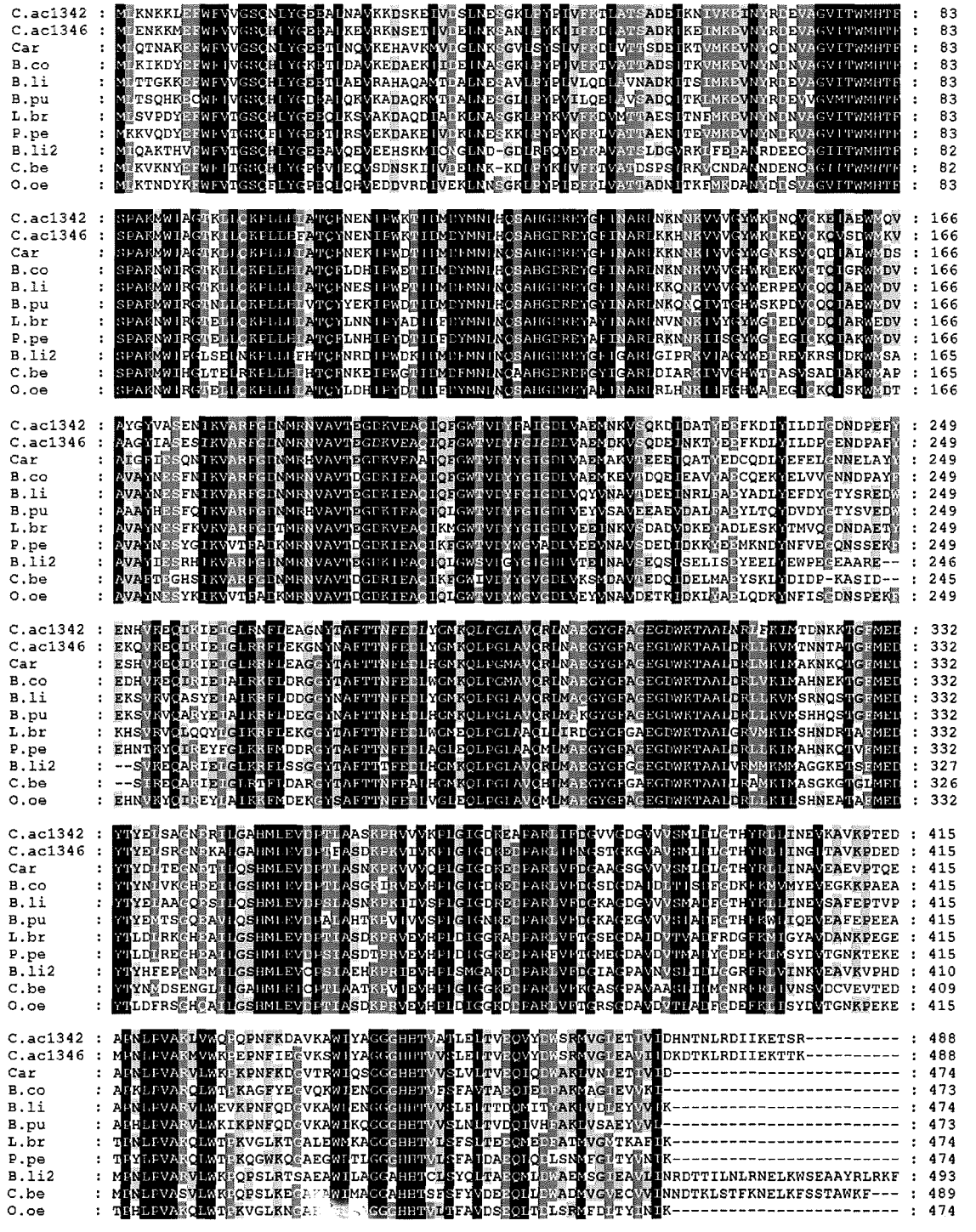


Figure 5.5: Multiple alignment of the putative *C. acetobutylicum* AraA/XylA sequence with other pentose isomerases

The deduced amino acid sequence of the putative *C. acetobutylicum* AraA/XylA is aligned with the pentose isomerases with the highest percentage identities from the BLAST output. Dashes within the sequences indicate gaps giving optimal alignment. Identical residues are within black boxes and conserved residues (present in at least seven of the eleven sequences) are shaded.

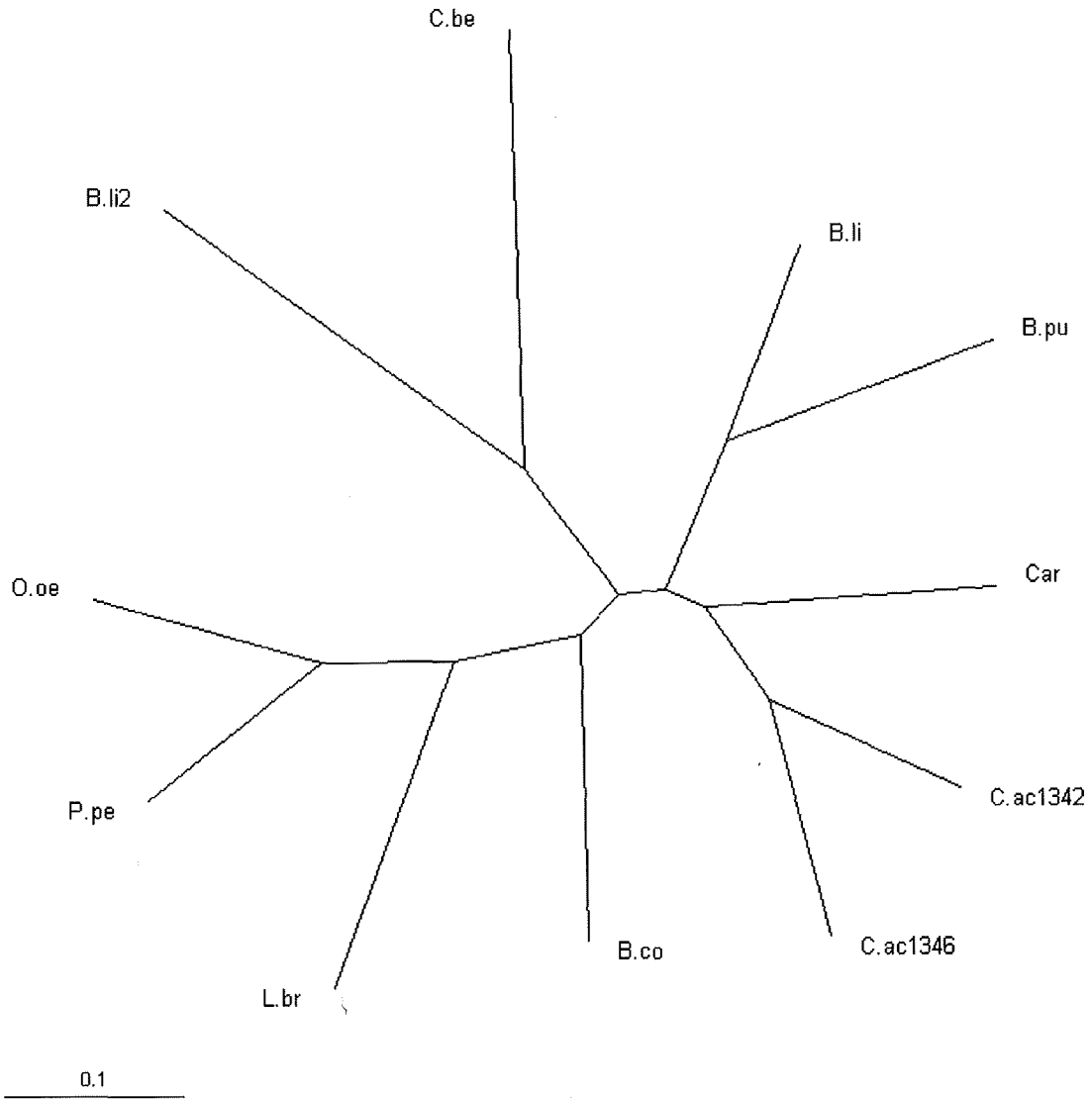


Figure 5.6: Unrooted phylogenetic tree for the putative pentose isomerase, cac1432.

The amino acid sequences of the putative pentose isomerase were compared with selected Gram positive and Gram-negative sequences returned from BLAST search and analysed using ClustalW and TreeView.

C.ac1342, C.ac1346, *Clostridium acetobutylicum* ATCC 824; Car, *Carnobacterium* sp. AT7; B.co, *Bacillus coagulans* 36D1; B.li, *Bacillus licheniformis* ATCC 14580; B.pu, *Bacillus pumilus* ATCC 7061; L.br, *Lactobacillus brevis* ATCC 367; P.pe, *Pediococcus pentosaceus* ATCC 25745; B.li2, *Bacillus licheniformis* ATCC 14580; C.be, *Clostridium beijerinckii* NCIMB 8052; O.oe, *Oenococcus oeni* ATCC BAA-1163.

5.2.2 *In silico* analysis of the putative pentose isomerase gene (cac1346)

The sequence of cac1346 and the regions upstream and downstream were entered into the ORF finder tool on the NCBI server. A possible rbs (GGGG) was identified 9 nucleotides upstream from the putative start codon (Figure 5.7). The putative ORF is 1467bp long and encodes a protein of 488aa in length with a molecular weight of 55,488Da. A putative CRE sequence was identified (TATAAATGTTTTT).

The deduced amino acid sequence of the putative isomerase protein was queried against sequence databases using the BLAST server (Altschul *et al.,* 1997) to search for sequence homology (Table 5.3). The sequence demonstrated significant similarity to *Bacillus licheniformis* ATCC 14580, another low GC Gram positive bacteria, L-arabinose isomerase (68% identity, 82% similarity).

Figure 5.7: Nucleotide and deduced amino acid sequence of the *C. acetobutylicum* putative pentose isomerase gene *araA/xylA* (cac1346)

The deduced amino acid sequence of the putative ORF was placed below the first nucleotide of the corresponding codon. Stop codons are indicated by asterisks. A possible CRE sequence is highlighted.

```
                agtagtttataaatgtttttagaataaataaaaaatagataaggggagatttatc
                                                                rbs
   1 - ATGTTAGAAAATAAAAAGATGGAATTTTGGTTTGTAGTAGGAAGTCAACATTTATATGGTGAAGAGGCTTTAAAA - 75
   1 - M  L  E  N  K  K  M  E  F  W  F  V  V  G  S  Q  H  L  Y  G  E  E  A  L  K  - 25

  76 - GAAGTAAGAAAAAATTCTGAGACAATTGTAGATGAATTAAATAAAAGTGCTAATCTTCCATATAAAATAATATTT - 150
  26 - E  V  R  K  N  S  E  T  I  V  D  E  L  N  K  S  A  N  L  P  Y  K  I  I  F  - 50

 151 - AAAGATTTAGCAACTTCTGCTGATAAAATAAAGGAAATAATGAAGGAAGTTAACTATAGAGATGAAGTAGCAGGA - 225
  51 - K  D  L  A  T  S  A  D  K  I  K  E  I  M  K  E  V  N  Y  R  D  E  V  A  G  - 75

 226 - GTTATAACTTGGATGCATACGTTTTCTCCAGCTAAAATGTGGATAGCAGGTACAAAGATATTACAAAAACCTTTA - 300
  76 - V  I  T  W  M  H  T  F  S  P  A  K  M  W  I  A  G  T  K  I  L  Q  K  P  L  - 100

 301 - CTTCATTTTGCAACTCAATATAATGAAAATATTCCATGGAAAACAATAGATATGGATTATATGAACTTACATCAA - 375
 101 - L  H  F  A  T  Q  Y  N  E  N  I  P  W  K  T  I  D  M  D  Y  M  N  L  H  Q  - 125

 376 - AGTGCTCATGGAGATAGAGAGTATGGATTTATTAATGCAAGACTTAAAAAGCATAATAAAGTTGTTGTAGGATAT - 450
 126 - S  A  H  G  D  R  E  Y  G  F  I  N  A  R  L  K  K  H  N  K  V  V  V  G  Y  - 150

 451 - TGGAAGGATAAAGAAGTTCAAAAACAAGTTTCAGATTGGATGAAGGTTGCTGCAGGATATATTGCAAGTGAAAGC - 525
 151 - W  K  D  K  E  V  Q  K  Q  V  S  D  W  M  K  V  A  A  G  Y  I  A  S  E  S  - 175

 526 - ATAAAAGTTGCACGTTTTGGTGATAACATGCGTAATGTTGCAGTTACAGAGGGAGATAAAGTAGAAGCTCAAATA - 600
 176 - I  K  V  A  R  F  G  D  N  M  R  N  V  A  V  T  E  G  D  K  V  E  A  Q  I  - 200

 601 - CAATTCGGATGGACAGTAGATTACTTTGGTATAGGTGATTTAGTTGCTGAAATGGACAAGGTAAGCCAAGATGAA - 675
 201 - Q  F  G  W  T  V  D  Y  F  G  I  G  D  L  V  A  E  M  D  K  V  S  Q  D  E  - 225

 676 - ATAAATAAAACTTATGAAGAATTTAAAGATTTATATATTTTAGATCCAGGTGAAAATGATCCTGCTTTCTATGAG - 750
 226 - I  N  K  T  Y  E  E  F  K  D  L  Y  I  L  D  P  G  E  N  D  P  A  F  Y  E  - 250

 751 - AAACAAGTTAAAGAACAAATCAAAATTGAAATAGGATTAAGAAGGTTCTTAGAAAAAGGAAATTATAATGCATTT - 825
 251 - K  Q  V  K  E  Q  I  K  I  E  I  G  L  R  R  F  L  E  K  G  N  Y  N  A  F  - 275

 826 - ACAACAAACTTTGAAGATCTTTATGGAATGAAACAGTTACCTGGACTTGCAGTACAACGTTTAAATGCTGAAGGC - 900
 276 - T  T  N  F  E  D  L  Y  G  M  K  Q  L  P  G  L  A  V  Q  R  L  N  A  E  G  - 300

 901 - TATGGCTTTGCAGGCGAAGGAGACTGGAAAACTGCAGCTTTAGATAGATTATTAAAGGTTATGACTAATAATACT - 975
 301 - Y  G  F  A  G  E  G  D  W  K  T  A  A  L  D  R  L  L  K  V  M  T  N  N  T  - 325

 976 - GCTACAGGTTTTATGGAAGATTACACATATGAACTTAGTCGTGGAAATGAGAAGGCATTAGGAGCTCATATGCTT - 1050
 326 - A  T  G  F  M  E  D  Y  T  Y  E  L  S  R  G  N  E  K  A  L  G  A  H  M  L  - 350

1051 - GAAGTTGACCCAACTTTTGCTTCAGATAAACCAAAGGTTATTGTTAAACCACTAGGAATTGGAGATAAAGAAGAT - 1125
 351 - E  V  D  P  T  F  A  S  D  K  P  K  V  I  V  K  P  L  G  I  G  D  K  E  D  - 375

1126 - CCAGCACGTTTAATCTTTAATGGTTCAACAGGAAAAGGTGTAGCAGTTTCAATGCTTGATTTAGGAACACATTAT - 1200
 376 - P  A  R  L  I  F  N  G  S  T  G  K  G  V  A  V  S  M  L  D  L  G  T  H  Y  - 400

1201 - CGTTTAATAATAAACGGACTTACAGCAGTGAAACCAGATGAAGACATGCCAAACCTACCAGTTGCTAAAATGGTA - 1275
 401 - R  L  I  I  N  G  L  T  A  V  K  P  D  E  D  M  P  N  L  P  V  A  K  M  V  - 425

1276 - TGGAAACCAGAACCAAACTTCATTGAAGGAGTTAAATCTTGGATTTATGCAGGTGGCGGACATCATACAGTGGTT - 1350
 426 - W  K  P  E  P  N  F  I  E  G  V  K  S  W  I  Y  A  G  G  G  H  H  T  V  V  - 450

1351 - TCACTAGAATTAACAGTAGAACAGGTTTATGATTGGAGTCGTATGGTAGGCTTGGAAGCTGTAATAATAGATAAG - 1425
 451 - S  L  E  L  T  V  E  Q  V  Y  D  W  S  R  M  V  G  L  E  A  V  I  I  D  K  - 475

1426 - GATACTAAATTAAGAGATATAATAGAAAAGACAACAAAATAA                                  - 1467
 476 - D  T  K  L  R  D  I  I  E  K  T  T  K  *  X                               - 500
```

Table 5.3: BLAST homology results for the deduced amino acid sequence of the putative AraA/XylA (cac1346).

Results of a BLAST homology search with the deduced amino acid sequence of the putative pentose isomerase cac1346.

Species	Putative Gene	% Identity	% Positive	Predicted molecular mass	ORF Locus tag	Accession
Clostridium acetobutylicum ATCC 824	L-arabinose isomerase araA	80	90	55075	CAC1342	NP_347970
Carnobacterium sp. AT7	L-arabinose isomerase	71	85	53088	CAT7_10260	ZP_02184463
Bacillus licheniformis ATCC 14580	L-arabinose isomerase	68	82	53374	BL01182	YP_079230
Oceanobacillus iheyensis HTE831	L-arabinose isomerase	67	81	55548	OB2797	NP_693719
Bacillus pumilus ATCC 7061	L-arabinose isomerase	62	79	53387	BAT_1675	ZP_03055726
Lactobacillus plantarum WCFS1	L-arabinose isomerase	59	75	53443	lp_3554	NP_786756
Bacillus halodurans C-125	L-arabinose isomerase	56	76	56189	BH1873	NP_242739
Lactobacillus pentosus	L-arabinose isomerase	58	74			AAQ82779
Lactobacillus brevis ATCC 367	L-arabinose isomerase	58	73	53162	LVIS_1740	YP_795838
Bacillus licheniformis ATCC 14580	L-arabinose isomerase	57	74	55517	BL00352	YP_080173

The sequences returned from the BLAST database were aligned and analysed in ClustalW2 and then in GeneDoc (Figure 5.8), consecutive strings of conserved sequences were observed. The ClustalW alignment was then entered into TreeView to observe the genetic relatedness of the sequences (Figure 5.9). The sequence of the putative arabinose isomerase from *Carnobacterium* sp. AT7 forms a cluster with the two putative clostridial isomerases, cac1342 and cac1346. All the *Lactobacillus* sequences form a tight cluster together.

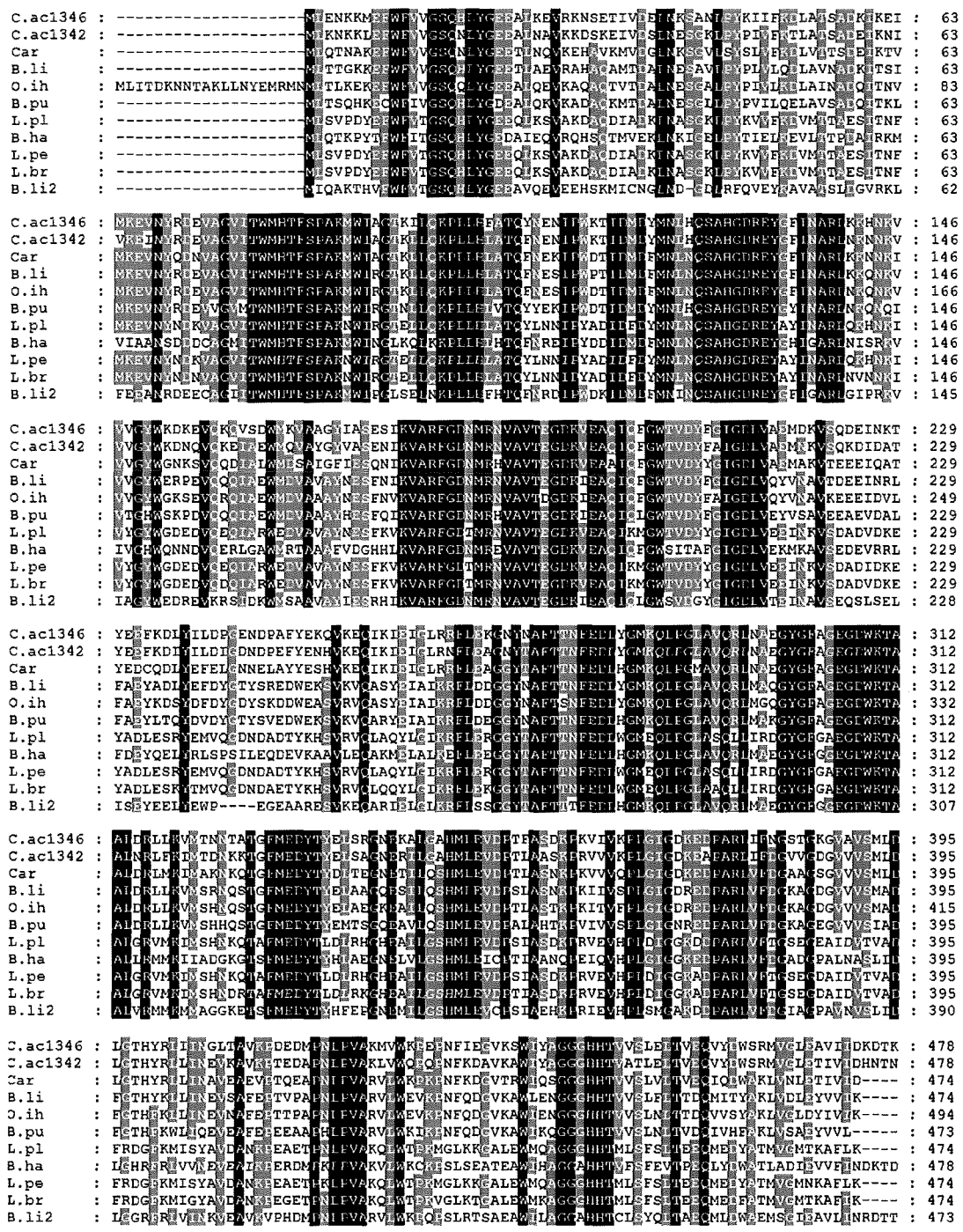


Figure 5.8: Multiple alignment of the putative *C. acetobutylicum* AraA/XylA sequence with other pentose isomerases (cac1346).

The deduced amino acid sequence of the putative *C. acetobutylicum* AraA/XylA is aligned with the pentose isomerases with the highest percentage identities from the BLAST output. Dashes within the sequences indicate gaps giving optimal alignment. Identical residues are within black boxes and conserved residues (present in at least seven of the eleven sequences) are shaded.

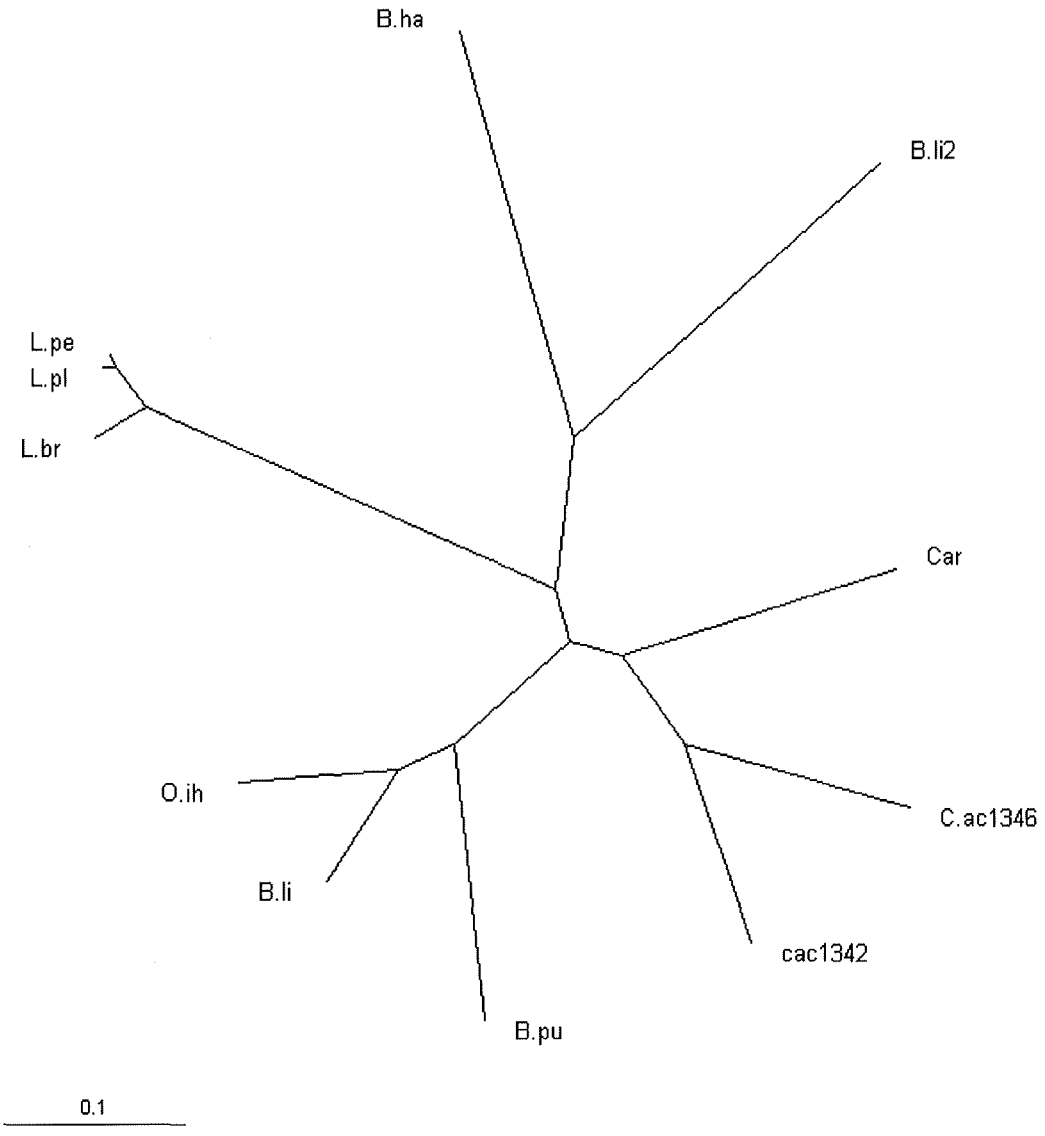


Figure 5.9: Unrooted phylogenetic tree for the putative pentose isomerase, cac1346.

The amino acid sequences of the putative pentose isomerase were compared with selected Gram positive and Gram-negative sequences returned from BLAST search and analysed using ClustalW and TreeView.

C.ac1346, cac1342, *Clostridium acetobutylicum* ATCC 824; Car, *Carnobacterium* sp. AT7; B.li, *Bacillus licheniformis* ATCC 14580; O.ih, *Oceanobacillus iheyensis* HTE831; B.pu, *Bacillus pumilus* ATCC 7061; L.pl, *Lactobacillus plantarum* WCFS1; B.ha, *Bacillus halodurans* C-125; L.pe, *Lactobacillus pentosus;* L.br, *Lactobacillus brevis* ATCC 367; B.li2, *Bacillus licheniformis* ATCC 14580.

To add further validity to the investigation of the two putative clostridial isomerases, sequences of experimentally verified, published xylose (XylA) and arabinose (AraA) isomerases were collected and aligned in ClustalW2 and visualised in TreeView. The results of this phylogenetic analysis are presented in Figure 5.10. The tree of experimentally verified pentose isomerases clearly forms two main clusters. The xylose isomerases forms one cluster and the second cluster is formed by the arabinose isomerases. Interestingly, the two clostridial sequences fall within the arabinose isomerase cluster (indicated by a blue highlighted square).

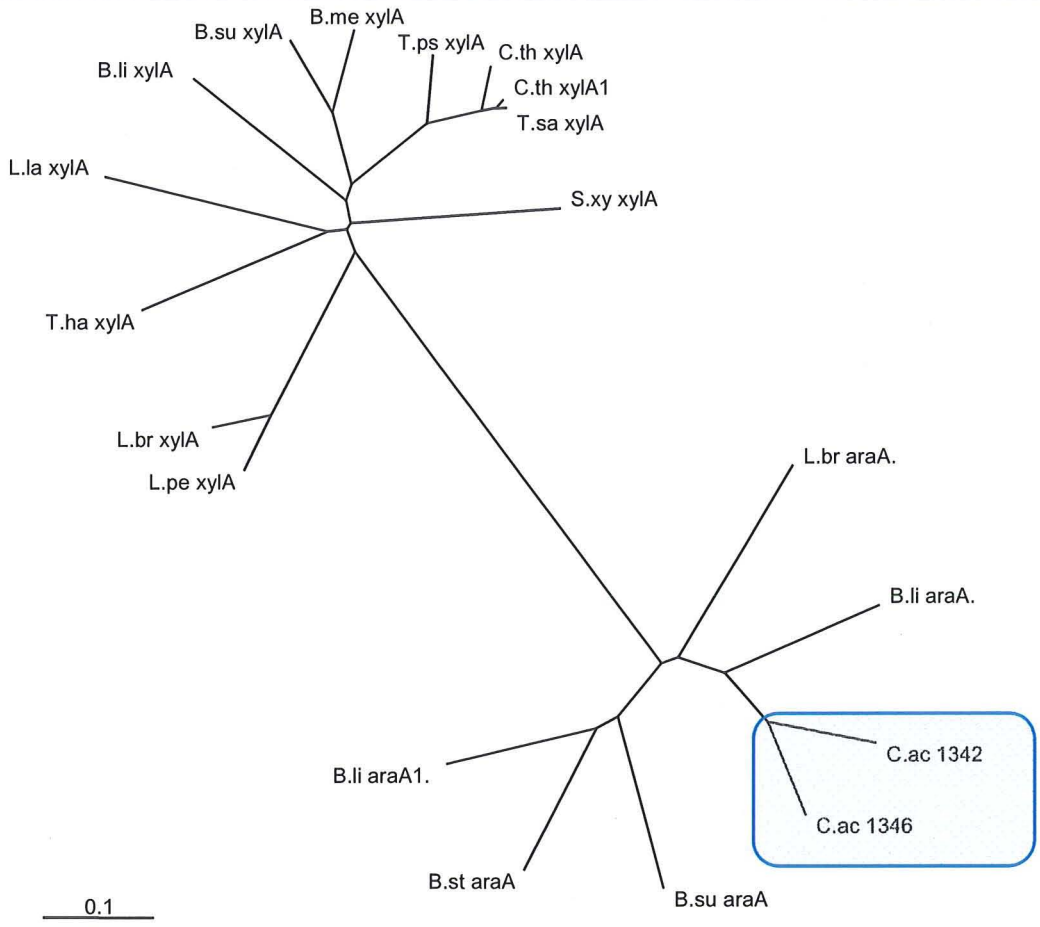


Figure 5.10: Unrooted phylogenetic tree of experimentally verified AraA and XylA proteins compared with the putative clostridial pentose isomerase sequences.

The tree of experimentally verified pentose isomerases clearly forms two main clusters, with the xylose isomerases forming one cluster and the arabinose isomerases forming another. The clostridial proteins fall within the arabinose isomerase cluster.

C.ac_1342, C.ac_1346, *Clostridium acetobutylicum* ATTC 824; B.su_xylA, *Bacillus subtilis* BSU17600 xylose isomerase; T.ha_xylA, *Tetragenococcus halophilus* O82845 xylose isomerase; L.la_xylA, *Lactococcus lactis subsp. lactis* LL1509 xylose isomerase; C.th_xylA1, *Clostridium thermosulfurogenes* P19148 xylose isomerase; B.li_xylA, *Bacillus licheniformis* (strain DSM 13 / ATCC 14580) P77832 xylose isomerase; T.ps_xylA, *Thermoanaerobacter pseudethanolicus* P22842 xylose isomerase; B.me_xylA, *Bacillus megaterium* O08325 xylose isomerase; L.pe_xylA, *Lactobacillus pentosus* P21938 xylose isomerase; S.xy_xylA, *Staphylococcus xylosus* P27157 xylose isomerase; T.sa_xylA, *Thermoanaerobacter saccharolyticum* P30435 xylose isomerase; C.th_xylA, *Clostridium thermosaccharolyticum* P29441 xylose isomerase; L.br_xylA, *Lactobacillus brevis* P29443 xylose isomerase; B.su_araA, *Bacillus subtilis* BSU28800 arabinose isomerase; B.st_araA, *Bacillus stearothermophilus* Q9S467 arabinose isomerase; B.li_araA1., *Bacillus licheniformis* Q65GC0 arabinose isomerase; B.li_araA., *Bacillus licheniformis* Q65J10 arabinose isomerase; L.br_araA., *Lactobacillus brevis* (strain ATCC 367 / JCM 1170) Q03PR5 arabinose isomerase.

5.2.3 Cloning of the putative pentose isomerase genes

The genes *xylA* and *araA* encode the xylose and arabinose isomerase proteins respectively. The genome of *C. acetobutylicum* ATCC 824 contains two candidate ORFs which potentially encode a pentose isomerase. Genomic DNA was extracted from an overnight RCM culture of *C. acetobutylicum* ATCC 824 and PCR reactions were performed as stated in Chapter 2, using the primers specific for a 1843bp region containing the ORF cac1342, and a 1616bp region containing the ORF cac1346 (details of the primers are contained in Chapter 2, Table 2.3). The PCR reaction parameters were optimised (data not shown) until an optimal annealing temperature of 58.2°C was established for both genes. PCR products were analysed by gel electrophoresis (Figure 5.11) and the products were of sizes consistent with the successful amplification of cac1342 and cac1346.

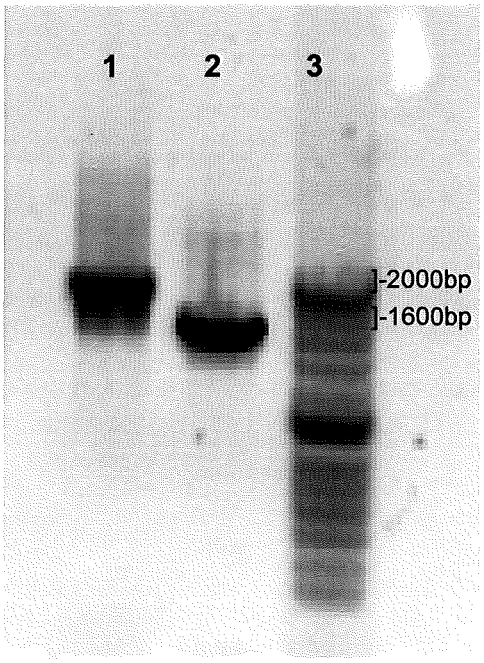


Figure 5.11: PCR amplification of the putative pentose isomerases of *C. acetobutylicum*

The PCR products obtained are consistent with the successful amplification of the putative ORFs cac1342 (lane 1) and cac1346 (lane 2). Analysed on a 1% agarose gel, using a RunOne electrophoresis unit. Hyperladder II (Bioline) is the DNA molecular size marker (lane 3).

Fresh PCR products, obtained from a reaction using the above optimised conditions, were individually ligated into pCR2.1-TOPO cloning vectors to produce plasmids harbouring the clostridial DNA sequences (details of the cloning vector are in Chapter 7 - 7.1 & 7.2). The constructed plasmids were then individually transformed into *E. coli* TOP10 cells to obtain transformants with either cac1342 or cac1346 insertions. After the outgrowth period the resulting transformants were analysed by spreading in various volume aliquots onto prewarmed LB plates containing ampicillin and with 40 mg/ml X-gal spread on the surface. After an overnight incubation many blue and white colonies resulted, with positive transformants appearing white. The colonies on the master plate were coded and replica plated onto a second LB ampicillin / x-gal agar plate. Five isolated colonies, for each putative gene, were selected and inoculated into LB / ampicillin broth (for plasmid preparation & glycerol storage) and spread on to ampicillin / X-gal agar plates in 1 cm lines (to amass enough cells for colony PCR).

To determine the presence of the cac1432 or cac1436 inserts within the cells purified plasmid was extracted from the overnight broth cultures. Plasmid preparations were preformed on broth cultures and the results are shown in Figure 5.12.

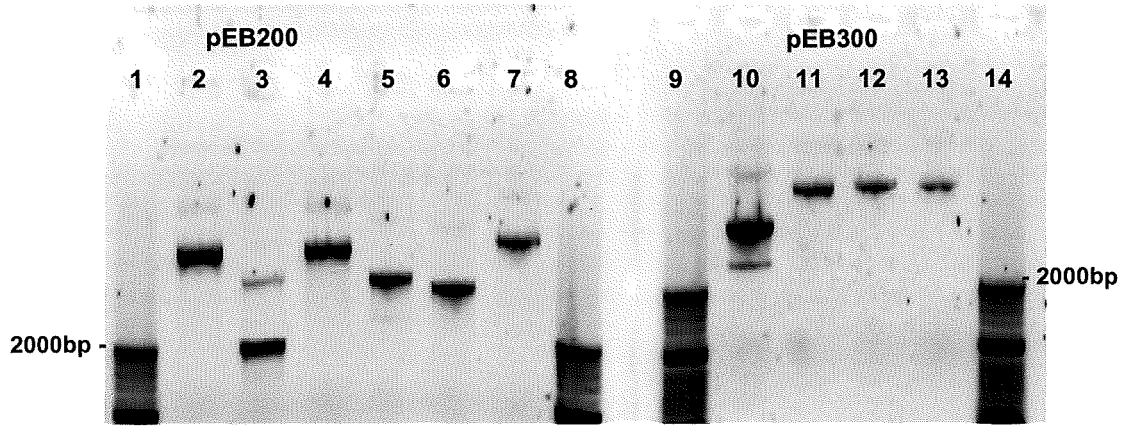


Figure 5.12: Purified plasmid preparations of pEB200 & pEB300

Purified recombinant plasmids were extracted with Promega PureYield mid Midiprep System from *E coli* TOP10 transformants (lanes 2 – 7 and 10 – 13). DNA analysed on a 1% agarose gel, using a RunOne electrophoresis unit. Hyperladder II (Bioline) is the DNA molecular size marker (lanes 1, 8, 9 and 14).

Using the purified plasmid preparations as template DNA, PCR reactions were conducted under the same conditions originally used to produce the PCR products from the chromosomal DNA. The presence of the gene within the plasmids was confirmed by gel electrophoresis (Figure 5.13). Lanes 1, 3, 4 and 5 contain cac1346 amplified from the purified plasmid extracted from the cac1346 transformants. Lanes 10, 12, 13 and 14 contain cac1342 amplified from the cac1342 transformants. Once the presence of the inserted gene within the purified plasmid preparations was confirmed, the next step was to determine the orientation of the putative gene within the plasmid.

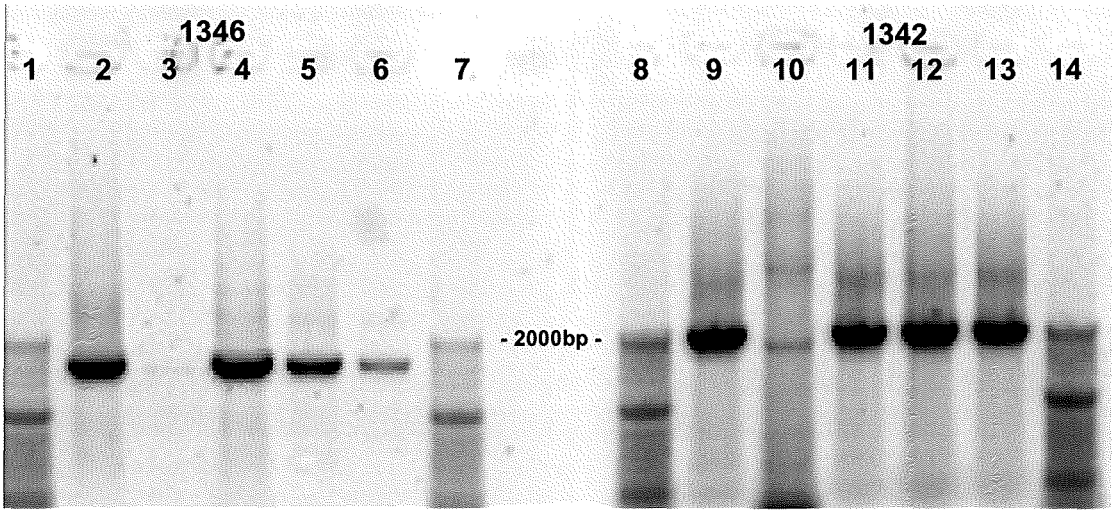


Figure 5.13: PCR to determine presence of the *xylA* insert

PCR products produced using purified plasmid preparations as template DNA to determine the presence of the *xylA* insert (lanes 2 – 6 and 9 – 13). DNA analysed on a 1% agarose gel, using a RunOne electrophoresis unit. Hyperladder II (Bioline) is the DNA molecular size marker (lanes 1, 7, 8 and 14).

5.2.4 Analysis of plasmid containing cac1342

Directional PCRs were performed on the plasmid preparations containing the cac1342 gene region, to determine the orientation of the inserted gene within the vector. However, all plasmids obtained possessed the putative gene under the control of the T7 promoter (designated pEB200); the results are presented in Figure 5.14. Lanes 1, 8 & 15 contain DNA molecular weight marker Hyperladder II (Bioline). Control reactions were performed with TOPO A & B, the M13 sequencing primers, (Lane 2) and with the forward & reverse cloning primers (Lane 3). Transformants were selected to perform a double test to demonstrate that the putative genes were certainly under the control of the T7 promoter.

The TOPO A primer and the reverse cloning primer were used (Lanes 4 & 6). Also the TOPO B primer and the reverse cloning primer were used (Lane 5 & 7). As none of the colonies selected for the plasmid preparations contained the cac1342 under control of the Lac promoter, thirty-five transformants were screened by colony PCR, all were found to be under the control of the T7 promoter.

5.2.5 Analysis of plasmid containing cac1346

Plasmids carrying the cac1346 gene region under the control of the Lac promoter was identified - designated pEB300 (Figure 5.14). A map of the features of this plasmid is presented in Figure 5.15. The purified plasmid preparation was then used for transformation of the *xylA* mutant *E. coli* DS941.

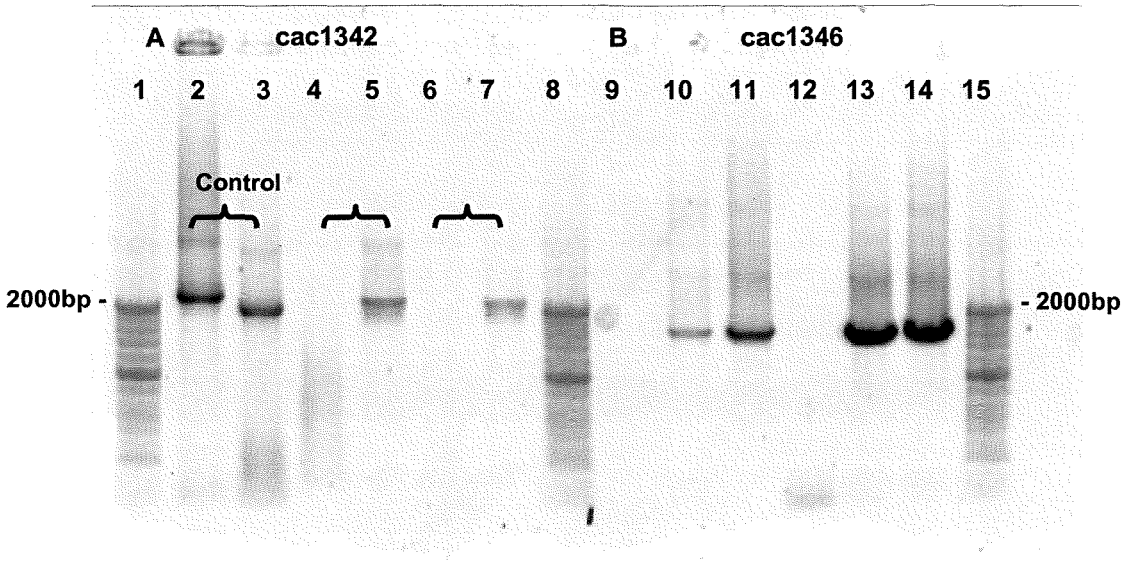


Figure 5.14: PCR to determine the orientation of the *xylA*

PCR products produced using purified plasmid preparations as template DNA to determine the orientation of the insert (lanes 2 – 7 and 9 – 14). DNA analysed on a 1% agarose gel, using a RunOne electrophoresis unit. Hyperladder II (Bioline) is the DNA molecular size marker (lanes 1, 8 and 15).

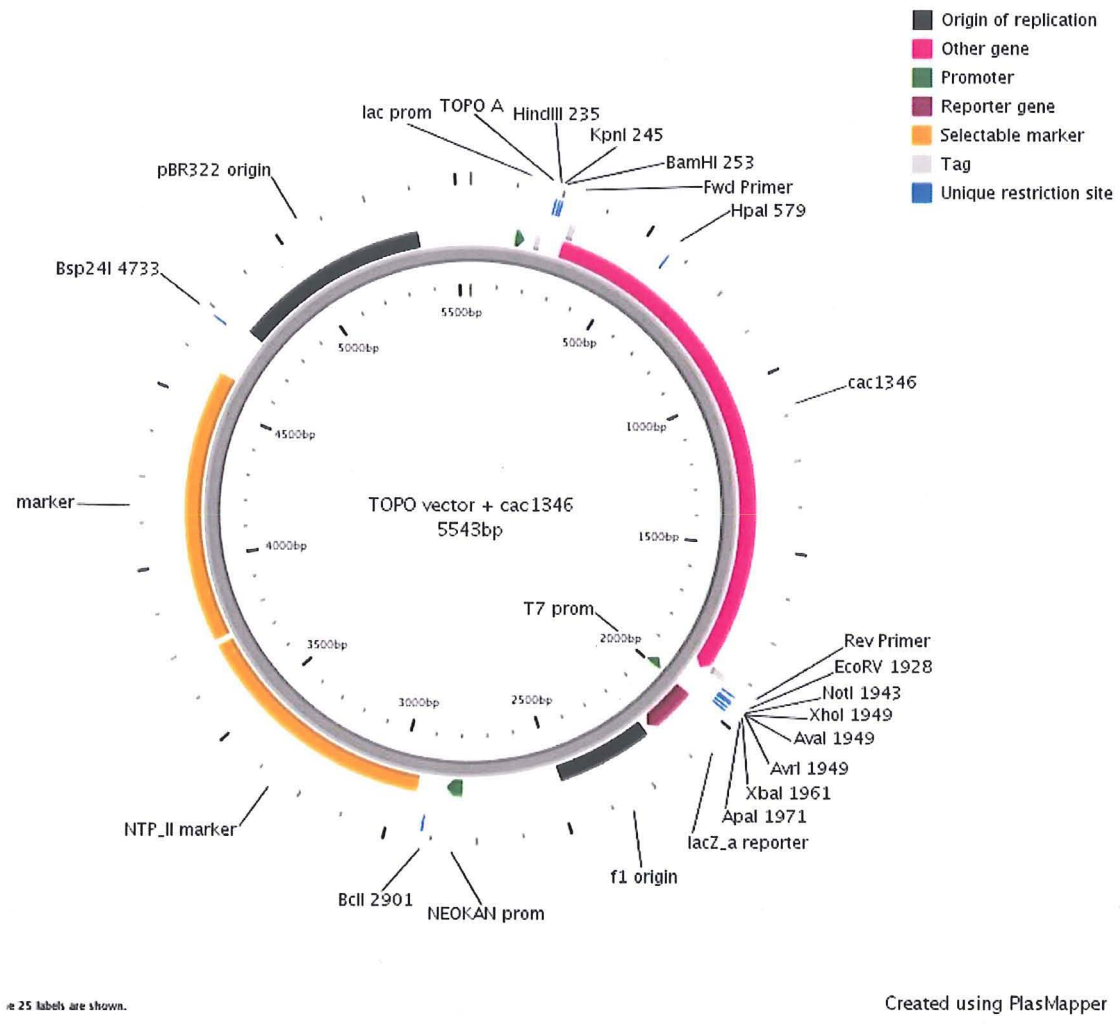


Figure 5.15: Features of pEB300

Important features of the pCR2.1 TOPO TA cloning vector and the putative xylose isomerase, cac1346. The plasmid is 5543bp and has many unique restriction sites. The Lac and T7 promoters are indicated by green triangles.

5.2.6 Complementation of *E. coli xylA* mutant with pEB300

A suitable mutant and positive control were sought for complementation screening to confirm the identity of the isomerase. *E. coli* DS941 carries the *xyl5* mutation, therefore has no active xylose isomerase and is unable to metabolise xylose. Previously, the XylA from *C. thermosaccharolyticum* has been cloned and expressed in *E. coli* DS941 (Meaden *et al.*, 1994). Purified plasmid preparations carrying the *xylA* from *C. thermosaccharolyticum* (pJA1) and cac1346 (pEB300) from *C. acetobutylicum* were used to transformed competent *E. coli* DS941 cells.

Transformants were then screened on MacConkey agar containing either xylose, arabinose or glucose as a sole carbon source. A section of the agar plate was left uninoculated for comparison. The vector pUC19 with no insert was used as a negative control and pJA1 (pUC19 + xylose isomerase gene from *C. thermosaccharolyticum)* was used as a positive control. The results are presented in Figure 5.16 and summarised in Table 5.4.

Table 5.4: Screening results

	Phenotype		
Plasmid	Glucose	Xylose	Arabinose
pUC19	+	-	-
pJA1	+	+	-
pEB300	+	-	-

The *E. coli* DS941 transformant containing the pEB300 plasmid demonstrated a negative fermentation phenotype (yellow) for both xylose and arabinose and a positive fermentation phenotype for glucose (red). The strain carrying pUC19 also produced a similar fermentation pattern. The stain harbouring the pJA1 plasmid was positive for xylose fermentation. Results suggest that cac1346 does not encode a xylose specific isomerase.

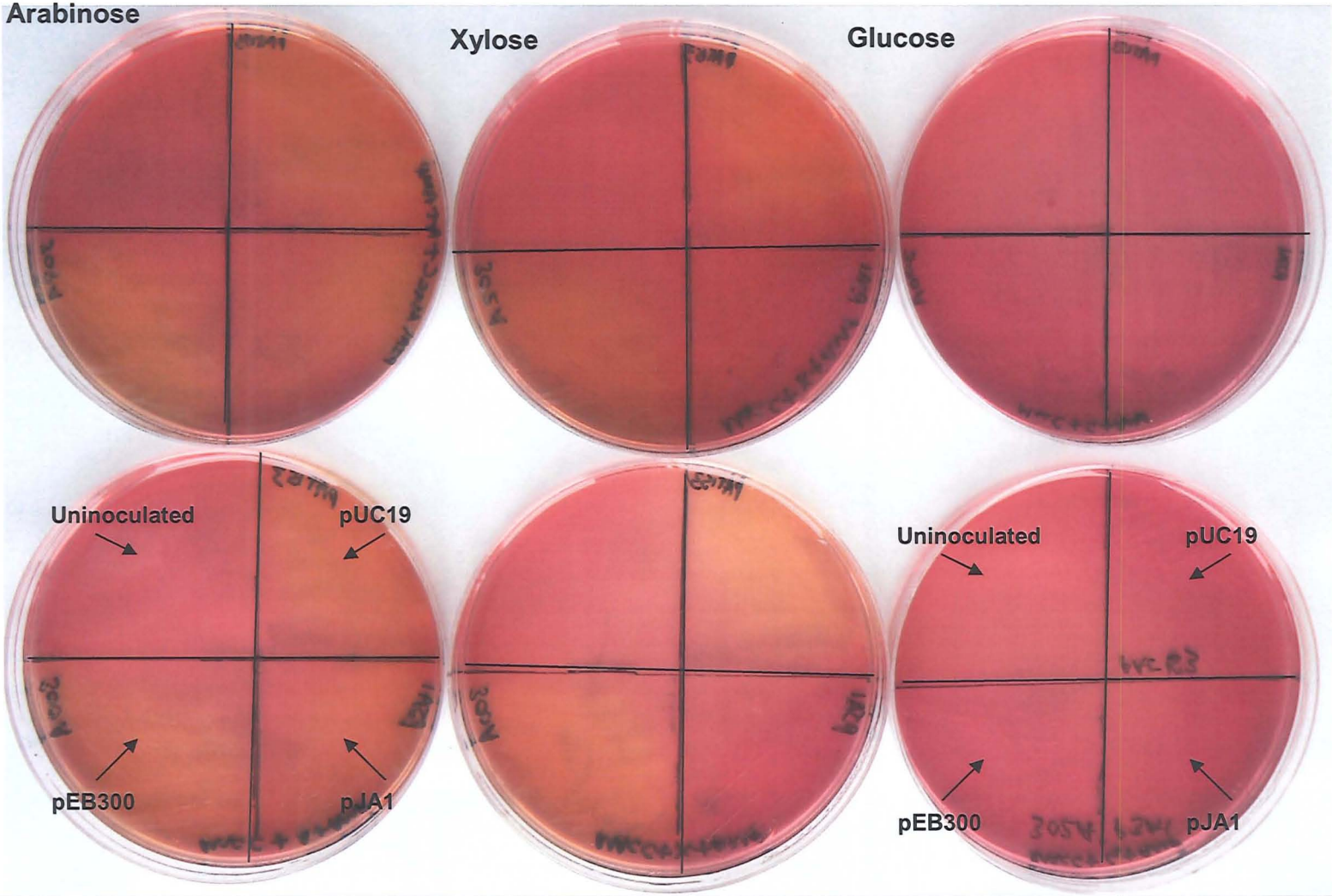


Figure 5.16: Screening complementation analysis

In a clockwise direction; the 1st quadrant of each plate is uninnoculated. The 2nd quadrant is streaked with DS941 cells which contain pUC19 (negative control). The 3rd quadrant is streaked with DS941 cells which contain pJA1 (XylA from *C. thermosaccharolyticum,* positive control). The 4th quadrant is DS941 cells transformed with pEB300.

5.3 Mannose transport in *C. acetobutylicum*

The monosaccharide mannose is a six carbon (hexose) sugar found in large quantities in hemicellulose. Acid hydrolysis of lignocellulosic wood biomass obtained from coniferous trees results in a hydrolysate containing a monosaccharide fraction consisting mainly of glucose and mannose (Maddox & Murray, 1983; Tang *et al.*, 2006). It has been demonstrated that *C. acetobutylicum* can produce significant volumes of solvents from mannose containing hydrolysates (Ezeji & Blaschek, 2008). Mannose has been shown to be transported by the PTS in a number of low GC Gram positive bacteria (Lortie *et al.*, 2000; Cochu *et al.*, 2003; Asanuma *et al.*, 2004; Gravesen *et al.*, 2002; Erni *et al.*, 1987) and initial investigations, presented in Chapter 3, suggested the *C. acetobutylicum* genome contained mannose PTS permeases.

Utilising the genetic information of biochemically and functionally defined proteins is an efficient means of initiating characterisation of unannotated proteins. A search of the current literature was carried out to identify organisms that possess representative PTS genes for mannose transport.

Gram positive organisms with mannose specific phosphotransferase systems are listed in Table 5.5, alongside the publication reference and operon accession number. The model Gram negative *Escherichia coli* is also included for comparison. Specific details (UniProt accession number, length, gene and protein name) of each protein are contained in Table 5.6).

Table 5.5: Characterised mannose PTS's from selected Gram positive organisms

Organism	Reference	Accession Number
Streptococcus salivarius	Lortie *et al.*, 2000	AF130465
Streptococcus thermophilus	Cochu *et al.*, 2003	AY253328
Streptococcus bovis	Asanuma *et al.*, 2004	AB114606
Listeria monocytogenes	Gravesen *et al.*, 2002	AF397145
Escherichia coli	Erni *et al.*, 1987	J02699

Table 5.6: Details of mannose PTS's from selected organisms

Organism	Gene Name	Protein Name	Length aa	UniProtKB accession
S. salivarius	*manL*	$IIAB^{man}$	330	Q9S4L5
	manM	IIC^{man}	271	Q9S4L4
	manN	IID^{man}	303	Q9S4L3
	manO	ManO	124	Q9S4L2
S. thermophilus	*manL*	$IIAB^{man}$	330	Q841T4
	manM	IIC^{man}	275	Q841T3
	manN	IID^{man}	303	Q841T2
	manO		124	Q841T1
S. bovis	*manL*	$IIAB^{man}$	330	Q7WYV3
	manM	IIC^{man}	267	Q7WYV2
	manN	IID^{man}	303	Q7WYV1
	manO	manO	123	Q7WYV0
L. monocytogenes	*mptA*	$IIAB^{man}$	321	Q8YAM2
	mptC	IIC^{man}	268	Q8YAM1
	mptD	IID^{man}	303	Q8YAM0
E. coli	manX	$EIIAB^{man}$	323	P69797
	manY	IIC^{man}	266	P69801
	manZ	IID	286	P69805

5.3.1 Identification of the *C. acetobutylicum* mannose transport genes

The *C. acetobutylicum* genome was examined using the GIB server and the ORFs and domain assignments were predicted using the ORF finder tool and BLAST on the NCBI server. Investigation revealed two potential mannose specific EIIs. The first gene cluster was found residing on the megaplasmid pSOL1, consisting of three ORF's putatively encoding EII domains, IIAB, IIC and IID (Figure 5.17a). The second gene cluster is on the chromosome spread over four ORFs (cac1457, cac1458, cac1459 and cac1460) which appear to encode EIIA, EIIB, EIIC and EIID domains respectively (Figure 5.17b).

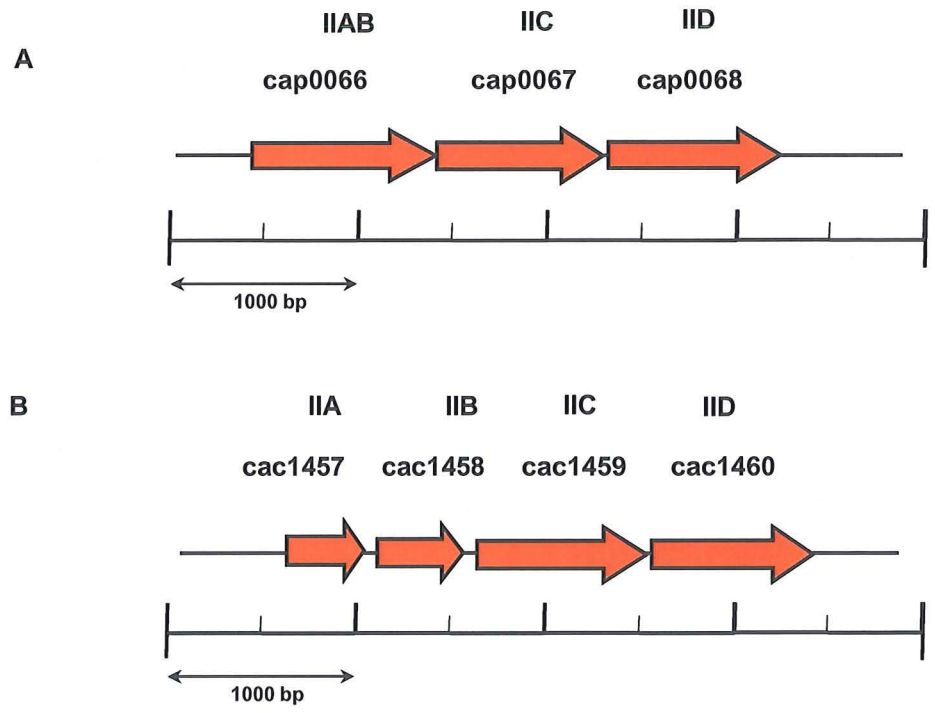


Figure 5.17: Arrangement of the ORFs of the putative mannose phosphotransferase system

Putative ORFs are represented by block arrows (indicating direction of transcript), locus tags are given above, scale (bp) is given below.

A) Plasmid borne genes. A putative mannose specific EII was identified residing on the mega plasmid pSOL1.

B) Chromosome borne genes. A putative mannose specific EII was identified residing on the chromosome.

To investigate the phylogenetic relatedness of the two clostridial EII^{man} proteins, the deduced amino acid sequences were aligned using the blastp algorithm (Altschul *et al.*, 1997). The sequences were found to share only 25% identity and 45% similarity (E value of 9e-64). To ensure that this was not a spurious effect caused by the differing domain architecture of the two permeases, the fused IIAB ORF was checked against the conserved domain database (CDD, Marchler-Bauer *et al*, 2005; Marchler-Bauer *et al*, 2007; Marchler-Bauer & Bryant, 2004), and then the individual IIA and IIB domains were entered into the blastp algorithm. The values did not change significantly (data not shown) and hence the sequence dissimilarity was indeed not caused by the domain architecture. As the sequences share only limited identity they were further analysed to determine if significant blocks of conservation were present. The sequences were aligned using ClustalW2 (Larkin *et al.*, 2007) multiple sequence alignment programme and to visualise the areas of homology the output file was input into the GeneDoc (Nicholas *et al.*, 1997) programme and the results are presented in Figure 5.18.

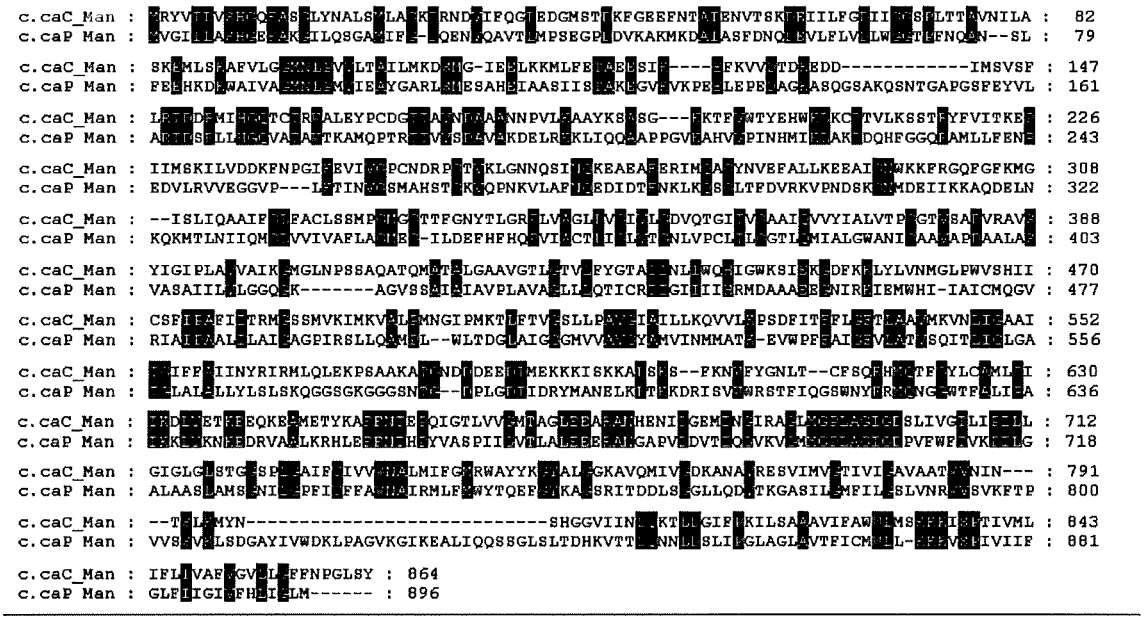


Figure 5.18: Sequence alignment of the two putative mannose EIIs of *C. acetobutylicum*

The deduced amino acid sequence of the two putative mannose EII components from *C. acetobutylicum* are aligned with each other, to check for regions of homology. The sequences were aligned using ClustalW2 programme and visualised using GeneDoc. Similarity groups were disabled. Dashes within the sequences indicate gaps giving optimal alignment. Identical residues are within black boxes.

Abbreviations are as follows: **c.ac_EII_ch,** *Clostridium acetobutylicum* chromosome borne EII; **c.ac_EII_pl,** *C. acetobutylicum* plasmid borne EII.

As can be observed from the alignment presented above in Figure 5.18, the only stretches (more than five consecutive matches) of homologous residues, are between the residues 95 – 100 and residues 691 – 700, falling within the EIIA and EIID domains respectively.

The sequences of the putative EII^{man} permeases were then aligned with the protein sequences of experimentally verified, published EII^{man}'s from a selected group of Gram negative and Gram positive bacteria. The alignment results are presented in Figure 5.19. The sequences show areas of significant homology thoughout.

In order to obtain an enhanced understanding of the relationships between the proteins the ClustalW2 alignment was visualised using the TreeView programme (Page, 1996). As can be seen from the phylogenetic tree presented in Figure 5.20, the *streptococcal* EII^{man} proteins unsurprisingly form a monophyletic group, whereas, the EII^{man} from *L. monocytogenes* and the putative EII^{man} from the *C. acetobutylicum* plasmid form a cluster together. The EII^{man} from the *C. acetobutylicum* chromosome is out on a distant branch on its own.

Due to the dissimilarity of the chromosomal EII^{man} the sequence was not included in a new alignment of the putative plasmid borne EII^{man} conducted against the selected characterised proteins (Figure 5.21). Far greater blocks of conservation can be observed from the alignment in Figure 5.21 compared to the alignment presented in Figure 5.19. This raises the question. Are both of these proteins functional for mannose transport? Cloning and heterologous expression in a relevant mutant host could answer this question.

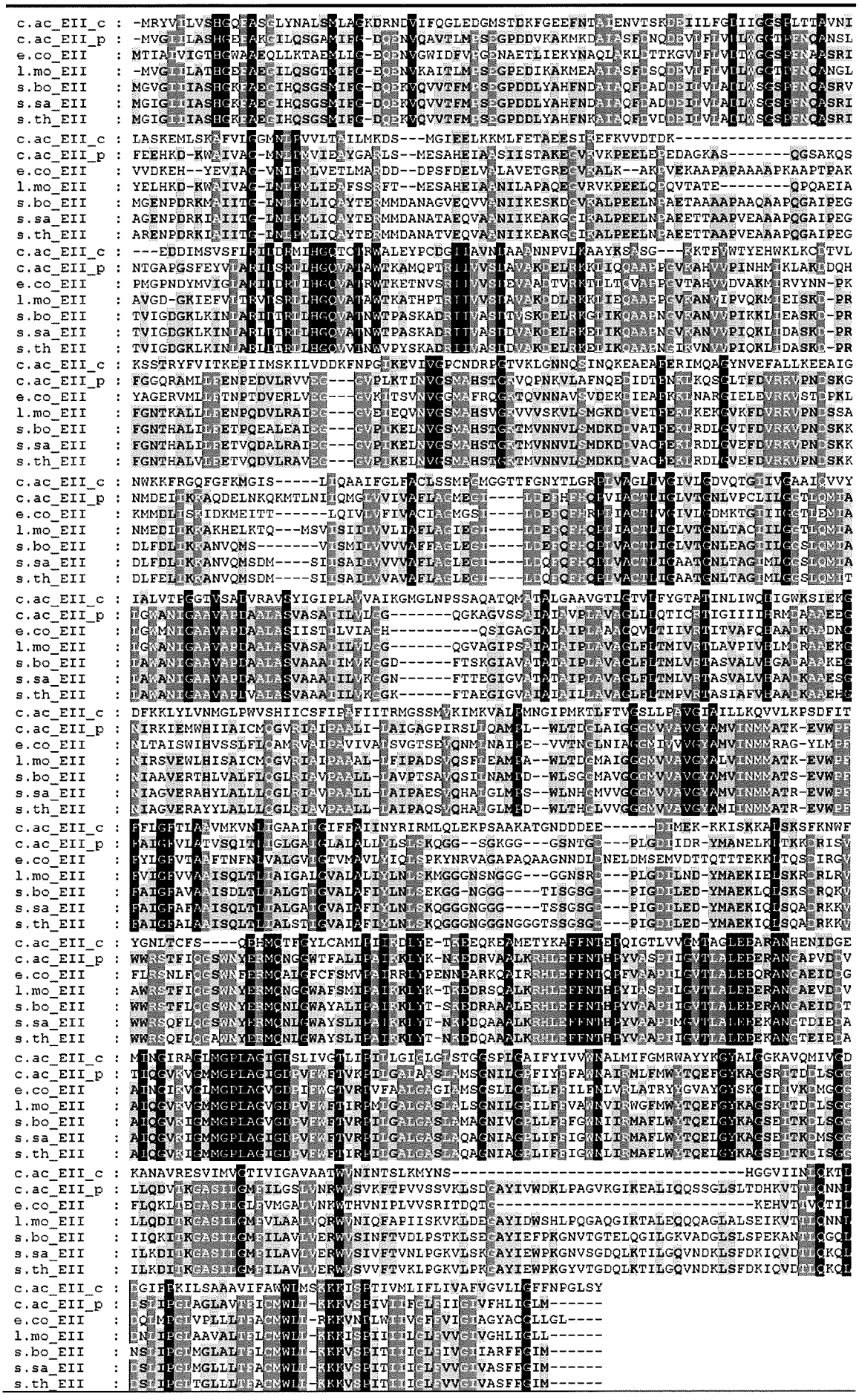


Figure 5.19: Comparison of the *C. acetobutylicum* and the characterised mannose EIIs

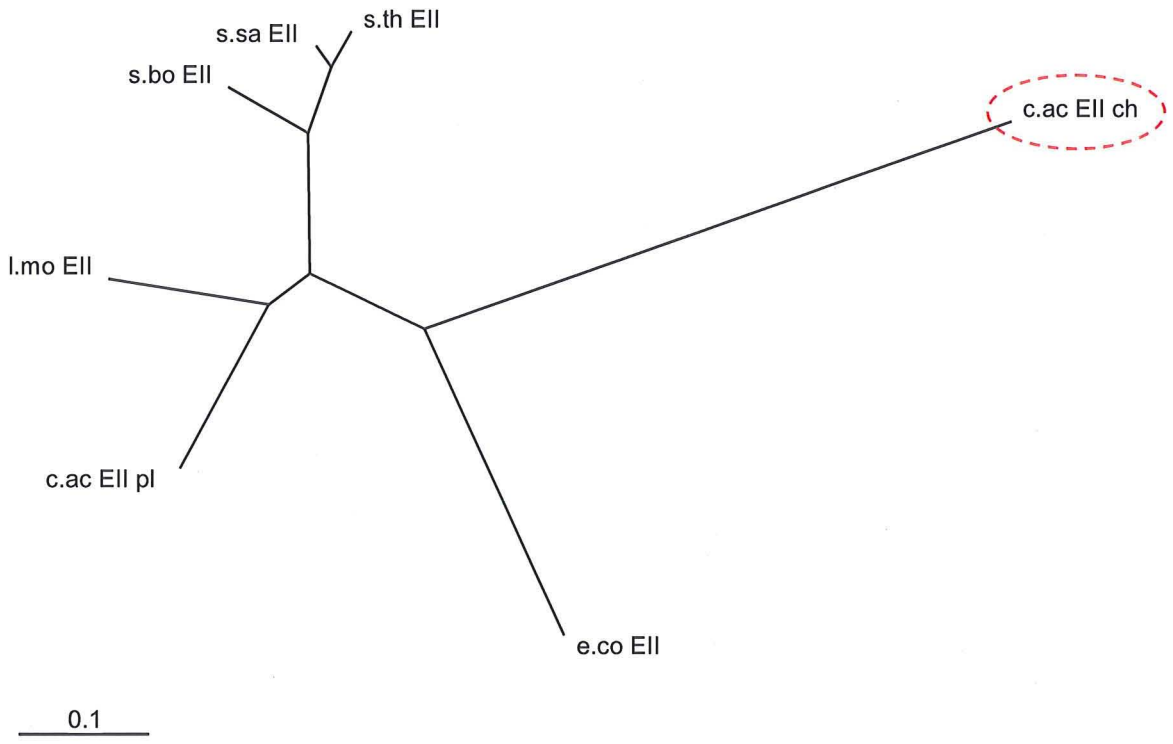


Figure 5.20: Unrooted phylogenetic tree of putative and characterised mannose EIIs.

The amino acid sequences of the putative clostridial mannose EIIs were compared with selected experimentally verified Gram positive and Gram negative mannose specific sequences and analysed using ClustalW and TreeView.

Abbreviations are as follows: c.ac_EII_ch, *Clostridium acetobutylicum* chromosome borne EII; c.ac_EII_pl, *C. acetobutylicum* plasmid borne EII; e.co_EII, *Escherichia coli*; l.mo_EII, *Listeria monocytogenes*; s.bo_EII, *Streptococcus bovis*; s.sa_EII, *S. salivarius*; s.th_EII, *S. thermophilus*.

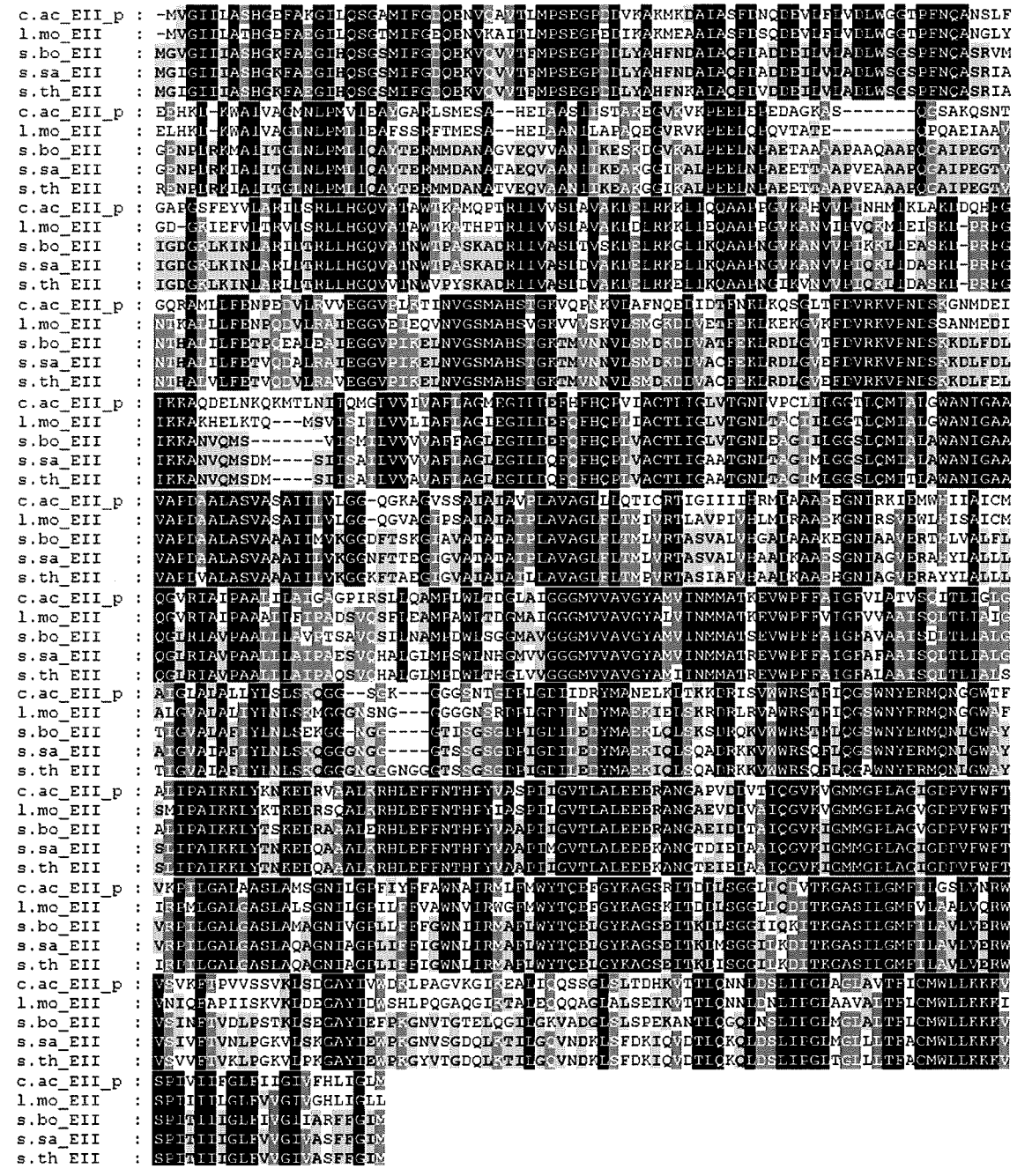


Figure 5.21: Multiple alignment of the putative mannose EII (plasmid borne) candidate against characterised EIIs

Abbreviations are as follows: c.ac_EII_pl, *C. acetobutylicum* plasmid borne EII; e.co_EII, *Escherichia coli*; l.mo_EII, *Listeria monocytogenes*; s.bo_EII, *Streptococcus bovis*; s.sa_EII, *S. salivarius*; s.th_EII, *S. thermophilus*.

5.3.2 Cloning of the putative EII^{man}

Strain *E. coli* ZSC113 is deficient in mannose transport, but with functional general PTS proteins see Chapter 2 for full phenotypic details. To confirm the phenotype, the ZSC113 mutant was streaked onto MacConkeys agar supplemented with either xylose, sucrose, glucose or mannose as the main carbon source. A positive fermentation phenotype was only observed for xylose. Glucose, sucrose and mannose all demonstrated a negative fermentation phenotype (Figure 5.22). Competent cells were prepared as stated in Experimental procedures and stored at -80°C until used.

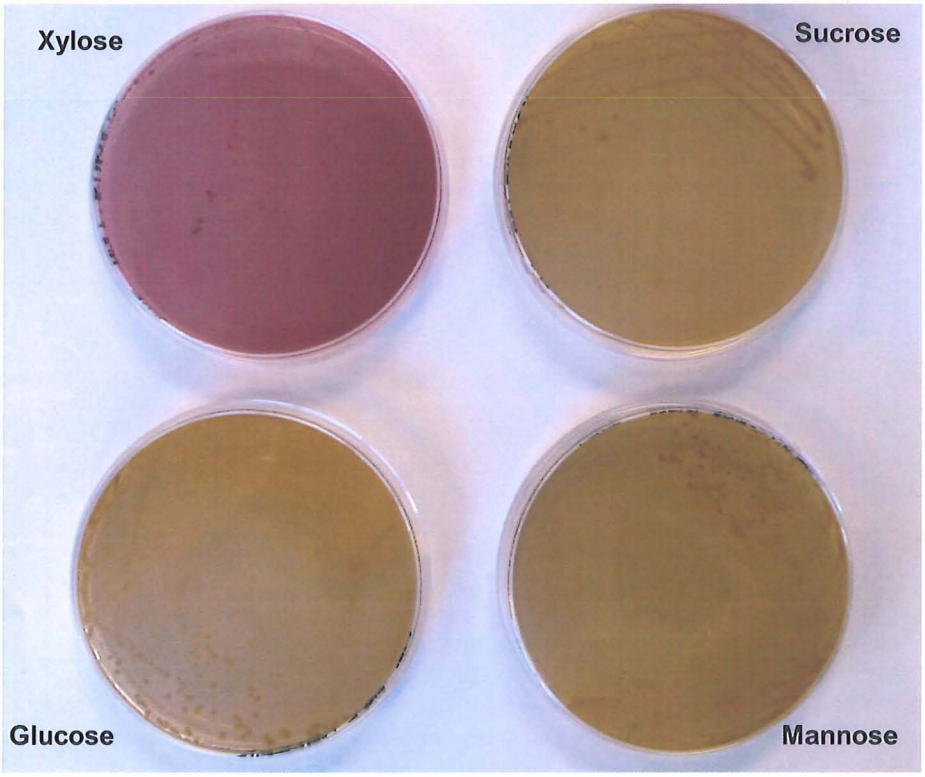


Figure 5.22: Phenotype of *E. coli* ZSC113

E. coli ZSC11 streaked onto MacConkey agar plates containing either; mannose, sucrose, xylose or glucose as the main carbon source.

Primers were designed as stated in Chapter 2 to amplify both the plasmid and chromosome borne putative mannose EIIs. Optimisation of the PCR was carried out using a range of annealing temperatures to obtain a single strong band of the expected size for each of the putative EII's (Figure 5.23).

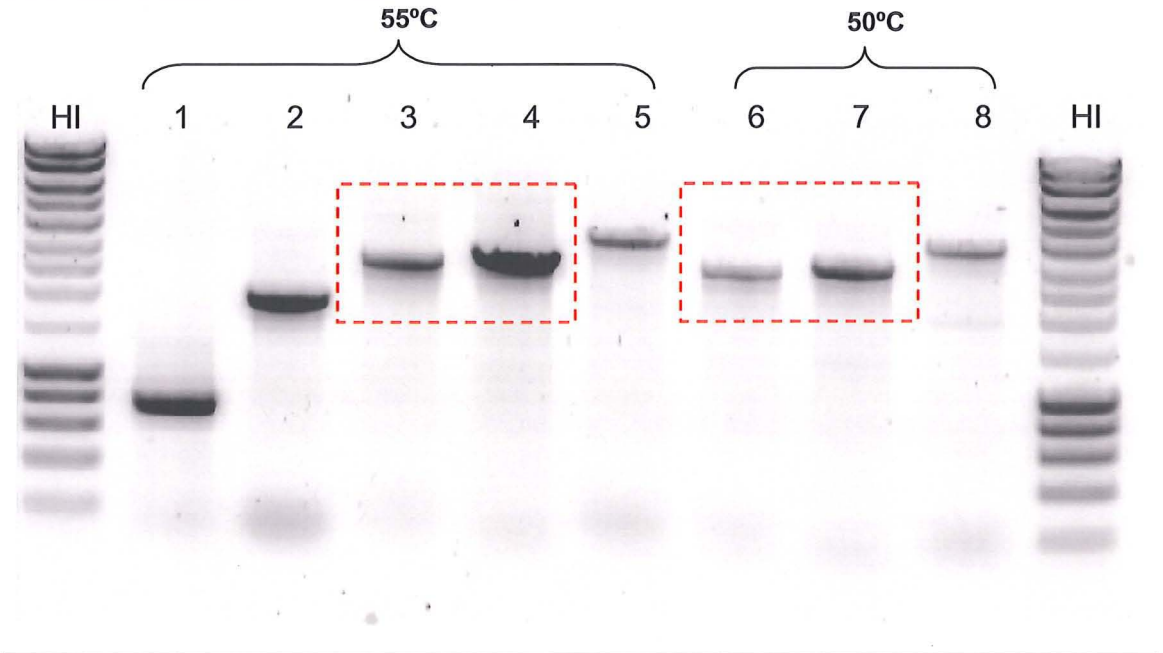


Figure 5.23: Optimisation PCR amplification of putative mannose EIIs

Optimisation PCR analysed on a 1% agrose gel, using a RunOne electrophoresis unit. Hyperladder I [HI] (Bioline) is the DNA molecular size marker. Lanes 1, 2, 5 and 8 are control reactions. PCR products run in lanes 1 to 5 had an annealing temperature of 55°C and PCR products run in lanes 6 to 8 had an annealing temperature of 50°C. Lanes 3 and 6 contain the chromosome borne putative mannose EII. Lanes 4 and 7 contain the plasmid borne EII.

For ligation into the cloning vector, fresh PCR products were obtained (Figure 5.24). Transformation of *E. coli* ZSC113 with the plasmids containing the two individual mannose EII permeases was conducted as stated in Chapter 2. Several transformants selected and glycerol stocked for future analysis in a separate project at the BfRC.

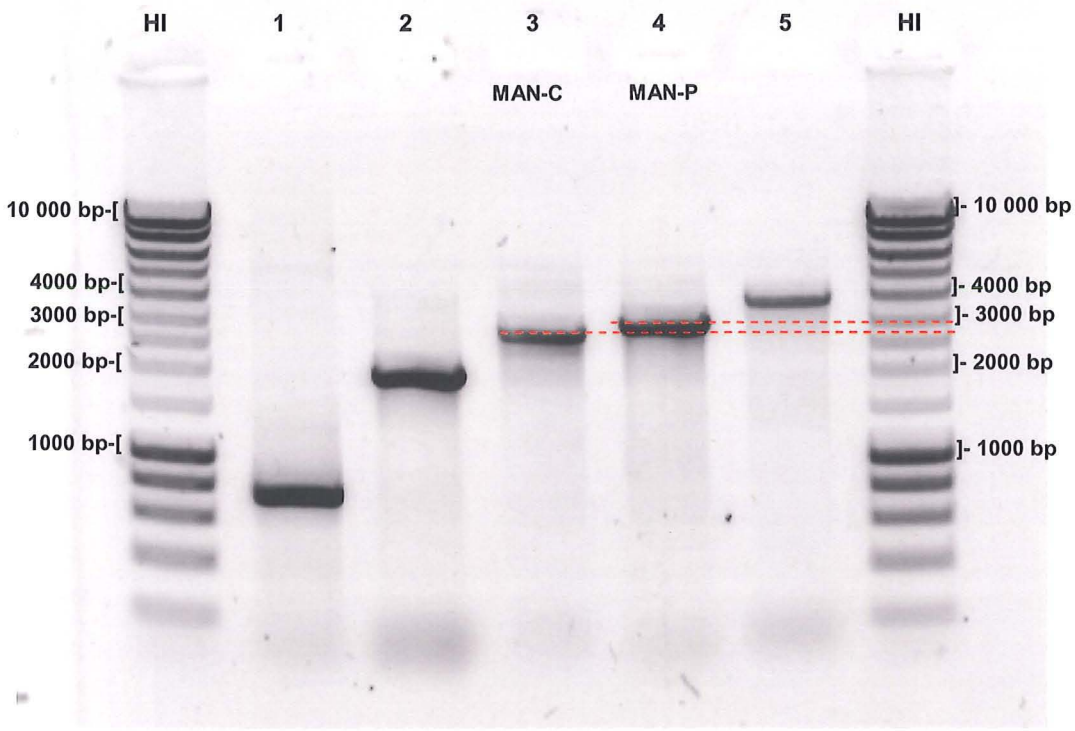


Figure 5.24: PCR amplification of mannose EIIs

Analysed on a 1% agarose gel, using the RunOne unit. Hyperladder I (Bioline) is the DNA molecular size marker. Lanes 1, 2 and 5 contain control reactions. Lane 3 contains the chromosome borne mannose EII (2740bp), lane 4 contains the plasmid borne mannose EII (2917bp).

5.4 Discussion

Xylose containing hemicelluloses are abundant in nature (Saha, 2003). Many bacteria can utilize xylose and because of this the genes for xylose transport and metabolism have been characterized in a variety of Gram negative and Gram positive bacteria, including a number of industrially important species including; *Bacillus licheniformis, B. megaterium, Escherichia coli, Lactobacillus brevis, L. pentosus* (Scheler *et al.*, 1991; Rygus *et al.*, 1991; Lawlis *et al.*, 1984; Bor *et al.*, 1992; Lokman *et al.*, 1991).

The 3,940,880bp chromosome and 192,000bp megaplasmid (pSOL1) of *C. acetobutylicum* has been sequenced and is available for analysis (Nölling *et al.*, 2001). *C. acetobutylicum* has the ability to utilise xylose as a sole source of carbon and energy (Ounine *et al.*, 1985). However, to date, the xylose utilisation pathway has not been characterised in this organism.

Xylose isomerase is the first enzyme involved in the bacterial metabolism of xylose. In the genome annotation of *C. acetobutylicum*, no xylose isomerase gene was identified and no candidate could be identified via BLAST. As xylose is metabolised by *C. acetobutylicum,* rationally, the enzymes for the metabolism of xylose are likely to be encoded by the cell. Arabinose isomerases have been shown to exhibit activity for D-xylose and D-fucose (Patrick & Lee, 1968). Within the genome annotation two hits for arabinose isomerases were found, locus tags cac1342 and cac1346. These two putative genes and the surrounding gene clusters were subjected to further investigation initially using the KEGG Genome Map (Kanehisa & Goto, 2000) of *C. acetobutylicum* (Table 5.1). On investigation of the genes surrounding the putative pentose isomerases, homologs of a number of the genes involved in xylose and arabinose metabolism in other bacteria were identified (Table 5.1).

The first gene (cac1339) and the sixth gene (cac1345) in the cluster both appear to be sugar-proton symporters. When the phylogeny of the sequences were analysed against characterised pentose-proton symporters obtained from UniProt, cac1339 clusters on a branch with the *araE B. subtilis* arabinose transporter (Sa-Nogueira & Ramos, 1997), whereas, the cac1345 sequence lies with the *xylT* proton symporter from *Lactobacillus brevis* (phylogenetic tree not shown; Chaillou *et al.*, 1998).

The second gene in the cluster (*araR,* cac1340) is homologous to a transcriptional regulator of the LacI family. A XylR homolog cac3673 exists elsewhere in the chromosome and is a much more likely candidate for a xylose repressor. Both cac1339 and cac1340 are divergently transcribed from the other genes in the cluster.

The next gene in the sequence (cac1341) encodes a putative ribulose-5-phosphate 4-epimerase family protein. This enzyme converts L-ribulose-5-P to D-xylulose-5-P, in the breakdown of arabinose. The locus tags cac1342 and cac1346 both encode putative arabinose isomerases. A putative XylB xylulose kinase resides at locus tag cac1344. Putative transketolases are found at locus tags cac1343 and cac1348 and a putative transaldolase is found at cac1347. All of these putative genes are associated with pentose metabolism and transport.

The deduced amino acid sequences of cac1342 and cac1346 were analysed by BLAST searches, all of the top 10 results were found to be arabinose isomerases (Table 5.2 & Table 5.3). Currently, no xylose isomerise annotation exists in the *C. acetobutylicum* genome. An arabinose isomerase characterised in *E. coli* was found to have slight activity towards xylose (Patrick & Lee, 1968). Therefore a possibility exists that a novel isomerase may operate within *C. acetobutylicum*, which can isomerise xylose, arabinose and possibly other pentoses, such as the case with the pentose isomerase from *Klebsiella pneumoniae* (Menavuvu *et al.*, 2006).

In *C. acetobutylicum*, the transport of xylose was found to be repressed by the presence of glucose (Fond *et al.*, 1986 a & b). In the related *B. subtilis* the expression of the *xyl* operon is induced by xylose and repressed by glucose (Gartner *et al.*, 1988). A global mechanism of carbon catabolite repression of the genus *Bacillus,* which may also apply to the whole Firmicute phylum, consists of ser-46 phosphorylared HPr complexing with the CcpA protein and binding to DNA sequences called catabolite-responsive elements (CRE) which are associated with the target genes. This type of negative regulation can result in repression of a target system, even in the presence an inducer. CRE sequences were identified upstream of both of the isomerases (Figure 5.3).

Both putative isomerases were amplified and ligated into a TOPO cloning vector. Only cac1346 was found to be ligated under the control of the *lac* promoter (all transformants analysed of cac1342 were under control of the T7 promoter).

Plasmid containing cac1346, under the control of the *lac* promoter was isolated and used to transform DS941, a *xyl5* mutant. The transformant demonstrated a negative fermentation phenotype on the MacConkey-xylose agar, which suggests that the putative gene is not a xylose isomerase. However, a negative fermentation phenotype on MacConkeys-xylose agar is not definitive evidence that cac1436 is not a xylose isomerase. The clostridial gene may simply not be expressed within the *E. coli* background. An arabinose isomerase deficient mutant needs to be sourced, to screen the putative genes for arabinose isomerase activity.

This work provides a firm basis for developing an understanding of xylose and arabinose metabolism in *C. acetobutylicum.* At this stage the plasmid containing cac1342 under the control of the T7 promoter and the plasmid pEB300 was handed over to a colleague for further characterisation.

Unlike the pentoses present in hemicellulose mannose is usually transported by PTS. The mannose PTS has been characterised many important Gram positive bacteria. The *Streptococcus salivarius* mannose PTS is composed of four genes *manL, manM, manN* and *manO* encoding $IIAB^{man}$, a IIC^{man}, a IID^{man} and a protein of unknown function (Lortie *et al.,* 2000). In *S. thermophilus* the glucose/mannose PTS proteins were identified, also encoded in four genes *manL, manM, manN* and *manO* which represent the $IIAB^{man}$, IIC^{man} and IID^{man} domains. To assess the phosphotransferase competence of these PTS proteins *in vitro* PEP-dependent phosphorylation investigations were conducted with purified HPr, EI, and $IIAB^{man}$, as well as membrane fractions containing IIC^{man} and IID^{man} domains (Cochu *et al.,* 2003). They found that the PTS components efficiently transferred a phosphate from PEP to glucose, mannose, 2-deoxyglucose, and to a lesser extent fructose. However, whole cells were unable to catalyse the uptake of mannose and 2-deoxyglucose, thus the *S. thermophilus* mannose PTS proteins are unable to function as a competent transport system. In *S. bovis* the *man* operon is transcribed constitutively and consists of four gene: *manL, manM, manN* and *manO* encoding a $IIAB^{man}$, IIC^{man}, IID^{man} and a putative regulator, respectively (Asanuma *et al.*, 2004). A similar gene arrangement to the operons found in *S. salivarius* and *S. thermophilus. Listeria monocytogenes* possesses two mannose PTS permeases, *mpo* (mannose permease one) and *mpt* (mannose permease two), however only expression of *mpt* has been demonstrated (Gravesen *et al.,* 2002; Dalet *et al.*, 2001). Expression of the *mtp* operon was induced by the presence of glucose and mannose (Dalet *et al.*, 2001).

Akin to the *L. monocytogenes* genome, analysis of the *C. acetobutylicum* genome revealed the presence of two separate permeases of the Man family (Figure 5.17). The first permease has four domains spread over three ORFs (cap0066, cap0067 and cap0068), is encoded in pSOL1. The second permease is encoded within the genome rather than plasmid borne, and encodes each of the EII domains in a separate ORF (cac1457, cac1458, cac1459 and cac1460). The ORF cap0066 encodes a fused EIIAB domain, this arrangement is usually found in permeases of the Man family, however in a few cases this arrangement has been identified in permeases of the Asc (L-ascorbate)

family (Barabote & Saier, 2005). Nevertheless, EIID domains are only found in the Man family. The two proteins only shared limited homology (Figure 5.18) and when the phylogenetic relationships were examined the chromosome borne EII was a clear outlier (Figure 5.20). A multiple alignment conducted, without including the chromosome borne EII, clearly demonstrated a high degree of conservation between the clostridial plasmid borne EII and the experimentally verified EII proteins (Figure 5.21).

To identify if both EII permeases were functional for mannose transport, a mannose PTS deficient mutant was sought. *E. coli* ZSC113 is unable to transport mannose or glucose (Curtis & Epstein, 1975), the phenotype was checked (Figure 5.22) and competent cells were prepared (as per Chapter 2). Both of the regions encoding the EIIs were amplified and ligated into vectors awaiting transformation into the competent cells as part of an on-going project at the BfRC.

There is a vast, currently mostly sent to landfill, supply of hemicellulosic biomass materials, which could be utilised for solvent formation. *C. acetobutylicum* can ferment the monomers present in such biomass to produce ABE, this is in stark contrast to ethanol production by yeast, which are unable to utilise pentose monomers.

Chapter 6

6 Summary and future research

Fossil fuels are a limited resource, and historically the price of oil has been subject to severe price fluctuations. Furthermore, the continued reliance on fossil fuels is affecting the global carbon cycle. Biofuels could potentially provide a stable, renewable, carbon neutral alternative to fossil fuels. Currently large-scale production of bio-solvents from biomass is mainly carried out by fermentation of sugarcane or maize. However, utilisation of specifically grown feedstock, which could otherwise be used for human nutrition has been deemed irresponsible by lobbyists. Moreover, the reappropriation of farmland and the additional use of fertilisers and pesticides to grow and maintain biofuel crops are contentious issues. Nevertheless, most westernised countries continually generate substantial quantities of sugar rich waste biomass material, which is often disposed of by landfill, which is becoming an increasing environmental problem.

The UK has apparent potential for diversion and reappropriation of renewable / sustainable biomass streams such as municipal solid waste (MSW), as well as an abundance of large-scale agricultural waste residues. The metabolically diverse clostridia can ferment the sugars present in such biomass to produce butanol, acetone, ethanol and hydrogen (Jones & Woods, 1986). Transport fuels are responsible for a large proportion CO_2 equivalent emissions produced worldwide (Demirbas, 2007). Biobutanol can be directly substituted for gasoline in an unmodified car engine (Ramey, 2005). Therefore, fermentation of waste biomass streams by clostridia could not only reduce waste entering landfill, but also reduce CO_2 equivalent emissions.

The UK and Germany produce similar quantities of MSW per capita per year. In an analysis of the carbon emissions associated with treatment and disposal of both countries MSW, it was found that the associated CO_2 equivalent emissions were five times higher in the UK than in Germany (Mühle *et al.*, 2010). This was partly due to the heavy reliance the UK has on landfill disposal and partly due to the composition of the materials being sent to landfill (such as biodegradable biomass and other biogenic substances), rather than disposal by a more ecological method. In Germany over 60% of MSW is recycled and only 1% landfilled, whereas in the UK only around 30% of the MSW is recycled and approximately 55% is sent to landfill (Mühle *et al.*, 2010).

Biomass streams such as described above could be diverted from landfill and utilised as a fermentation feedstock for bio-solvent production. However, sequential and incomplete utilisation of feedstock seriously effects the efficiency and economics of the biomass-to-solvent production process (Nichols *et al.*, 2001). Characterisation of the

carbohydrate transport and regulation systems involved in the regulated uptake of the sugars present in waste biomass streams, will assist in the future optimisation of biofuel production facilities.

Clostridia such as *C. acetobutylicum* can utilise the sugars present in such biomass streams to produce solvents. Waste biomass feedstocks are not homogenous and a large variety of sugars may be present. Bacteria employ complex regulatory mechanisms to control expression of catabolic operons, determined by the carbon sources available their environment. Within a bioreactor, accumulation of fermentable carbohydrates is the first step in the conversion of sugar into solvents.

Many mono- and disaccharides which are present in biomass are accumulated by the phosphoenolpyruvate (PEP) phosphotransferase system (PTS). The PTS is often the principal mechanism of carbohydrate accumulation in low GC Gram positive bacteria, and plays a role in global carbon regulation (Mitchell, 1998). An *in silico* analysis of the phosphotransferase complement of *C. acetobutylicum* was therefore carried out and was presented in Chapter 3. Several systems were then selected for more in depth characterisation (Chapters 4 & 5).

C. acetobutylicum has thirteen complete permeases and a single lone EIIA domain. Six of the permeases and the lone EIIA belong to the glucose-glucoside (Glu) family of permeases, there is only a single representative of the galactitol (Gat) family and two members of each of the fructose-mannitol (Fru), lactose-N-N'-diacetylchitobiose-β-glucoside (Lac) and the mannose (Man) families. The permease genes were often identified associated with genes encoding regulators. Four transcriptional antiterminators were discovered associated with the EII permeases, all of which belong to the Glu family (glucose, sucrose, β-glucoside, and N-acetylglucosamine). Two transcriptional activators were found, one in an operon specific for mannitol utilisation (Fru family) and the other upstream from the β-glucoside permease belonging to the Lac family.

The *C. acetobutylicum* genome was found to possess two genes encoding fructose-1,6-bisphosphatases (FBPases), a class II, GlpX-like, FBPase (cac1088) and a class III FBPase (cac1572). FBPase catalyses the hydrolysis of fructose-1,6-bisphosphate (fbp) to fructose-6-phosphate and inorganic phosphate. The HPr kinase/phosphorylase (HprK/P) encoding gene *hprK* is found directly upstream from *glpX*. In the presence of fbp, HprK/P can phosphorylate HPr at serine-46. Ser-46 HPr can complex with the catabolite control protein, CcpA, to effect gene regulation by binding to catabolite

responsive elements, CRE's (Stülke & Hillen, 1999). HprK/P and CcpA have been shown to operate in *C. acetobutylicum* (Tangney *et al.*, 2003; Ren *et al.*, 2010). Based on the preliminary *in silico* analysis presented in this work, subsequent analysis was undertaken in another project at the BfRC that confirmed the identity of both FBPases. The class II and class III FBPase homologs were cloned into *E. coli* JB108 (*Δfbp*287; previously described by Donahue *et al.*, 2000) and both were found to possess FBPase activity (*Manuscript in preparation*).

A particular focus of this work was on β-glucoside metabolism, investigations revealed that *C. acetobutylicum* can transport and metabolise cellobiose, arbutin and salicin. Experiments with pairs of substrates revealed that the presence of glucose regulates transport of the β-glucoside cellobiose. Sequential utilisation of glucose before PTS disaccharides is consistent with other studies in this organism, such as is found with preferential glucose utilisation over sucrose and lactose (Tangney & Mitchell, 2000; Yu *et al.*, 2007).

In silico analysis of the *C. acetobutylicum* genome revealed two systems involved with β-glucoside utilisation (Chapter 4). System I is comprised of three ORFs, which appear to encode all functions necessary for the regulated uptake and utilisation of β-glucosides; cac1406 (*bglG*, a trasnscriptional antiterminator), cac1407 (*bglA*, an EII) and cac1408 (*bglB*, a phospho-β-glucosidase). Gene spacing, regulatory elements and similarity with other PTS's are consistent with the expression of *bglGAB* as an operon. Ribonucleic antiterminator target (RAT) sequences were identified upstream of both *bglG* and *bglA*. In the presence of an inducer antiterminator proteins bind to RATs preventing transcription termination, thus enabling transcription of the gene (Tangney & Mitchell, 2005). The information compiled from a succession of computer analyses (CCD, BLAST, TMHMM, hydrophobicity plots and signature sequence homology) of the BglA sequence were all consistent with the sequence encoding an EII of the Glc family, which has the domain organisation of BCA, with ten transmembrane helices. A highly conserved signature sequence was identified for disaccharide binding, as well as signature sequences for EIIA and EIIB domains. BglB contains a glycosyl hydrolase family 1 signature sequence and when the phylogenetic relationships of a large group of phospho-β-glucosidases were examined, BglB fell within a cluster all belonging to family 1 hydrolases.

In silico analysis of the second system revealed that it is comprised of four clustered ORF's; cac0383 (EIIA), cac0384 (EIIB), cac0385 (phospho-β-glucosidase) and cac0386 (EIIC). System II shares similarities with several cellobiose utilising operons and when

the EIIC domain was examined phylogenetically it clustered on a branch with two cellobiose specific EII's from *Lactococcus lactis*. Hydrophobicy profiling and domain analysis of the multi peptide protein, supported the presence of a large hydrophobic EIIC domain and hydrophilic EIIA and EIIB domains. The deduced amino acid sequence of cac0385 was homologous to experimentally verified phospho-β-glucosidases found associated with other β-glucoside operons. When examined phylogenetically the sequence also fell within the cluster of family 1 hydrolase proteins (however, was not as closely related to the others in the group as the phospho-β-glucosidase in System I). A divergently expressed transcriptional activator was identified upstream from the cac0383. The sequence was homologues to other transcriptional activators found associated with catabolic operons.

Multiple systems for β-glucoside utilisation, with overlapping substrate specificity are not uncommon. For example; *E. coli* (Keyhani, & Roseman, 1997), *Streptococcus mutans* (Old *et al.*, 2006) and *B. subtilis* (Setlow *et al.*, 2004) all have several systems for the regulated uptake, and metabolism of a range of β-glucosides. The *in silico* evidence presented in Chapter 4 strongly suggests that System I encodes all the components necessary for the regulated transport and metabolism of arbutin and salicin akin to the homologous operon found in *C. longisporum* (Brown & Thomson, 1998), or resembling the arbutin, salicin and cellobiose specific operon found in *E. coli* (Schnetz & Rak, 1988). The bioinformatics evidence also strongly indicated that System II encodes the components for regulated transport and metabolism of cellobiose, akin to the cellobiose system in *L. lactis* (Linares *et al.*, 2010).

Biochemical analysis of the cloned clostridial EII proteins revealed that the permease from System I only provided a positive fermentation type for arbutin and the permease from System II only provided a positive fermentation type for salicin. It is unexpected that neither System I nor System II could transport cellobiose.

However, these results require caveat, as there are several factors to be considered with the cloning experiments. Cloning in *E. coli* is limited; expression of a gene from a low GC Gram positive bacteria in a Gram negative host background, may lead to a protein that behaves in a slightly different manner than in its native background. Specialised gene knockout systems exist for genus clostridium (Heap *et al.*, 2007). It would be of great interest to construct strains of *C. acetobutylicum* with targeted inactivation of the β-glucoside permeases genes. The strains could then be screened on a range of β-glucosides to identify which proteins are specific for which β-glucoside, and deduce if overlapping substrate specify exists. If a knockout stain constructed with the β-

glucoside permeases inactivated was unable to transport cellobiose, it would support the bioinformatic evidence presented in Chapter 4. Which, suggests that cellobiose is transported by at least one of these permeases and that the protein cloned in the *E. coli* host background was not behaving as it would in its native background.

Another point which must be considered with cloning in *E. coli* TOP10 is whether or not a phospho-β-glucosidase specific for phospho-cellobiose and phospho-salicin were expressed in this strain. *E. coli* TOP10 has a constitutively expressed phospho-arbutin specific phospho-β-glucosidase (Parker & Hall, 1988). Therefore the permease from System I, may also be able to transport cellobiose and salicin, but possess no way to hydrolyse the phosphorylated disaccharide. The permease from System II was cloned with the associated phospho-β-glucosidase, which could potentially be specific for phospho-salicin. Due to this, future cloning of β-glucoside permeases, will be undertaken in more suitable *E. coli* mutants. For example, *E. coli* LP100, which has both the cryptic *bgl* and *cel* operons deleted (Parker & Hall, 1990). This would ensure any observation of a positive fermentation phenotype to be solely due to the expression of the cloned gene, and not spontaneous activation of cryptic systems. Additionally, for conformation of the identity of phospho-β-glucosidases, *E. coli* MK120 could be employed. This strain does not encode either of the wildtype *E. coli* phospho-β-glucosidases (named *bglA* and *bglB* in *E. coli*), but does possess a constitutively expressed EII (*bglF*) specific for transport of arbutin and salicin (Parker & Hall, 1990).

The complementation experiments also lacked a positive control. An ideal positive control for this could be the arbutin, salicin and cellobiose specific EII from *Corynebacterium glutamicum* (Kotrba *et al.,* 2003) or the arbutin and salicin specific EII from *C. longisporum* (Brown & Thomson, 1998). Work is underway to obtain a suitable positive control with *E. coli* LP100 and MK120 for future cloning experiments.

Extracts of *C. acetobutylicum* cells grown on glucose, and separately cellobiose, as the sole carbon sources were prepared as per Mitchell *et al.,* (1991). The extra ere subjected to high throughput LC MS/MS with monolithic columns (Batycka *et al.*, 2006) however, the results remains to be analysed and will be presented elsewhere.

Breakdown of the cellulose fraction of lignocellulose yields mainly cellobiose and glucose. Whereas breakdown of the hemicellulose fraction results in mainly the pentoses xylose and arabinose, and the hexose mannose. In clostridia xylose and arabinose are not normally PTS sugars and instead are usually transported by protein symporter. After a pentose sugar is transported the first step of its metabolism is

isomerisation. Interestingly, it is long established that *C. acetobutylicum* can utilise pentoses to yield solvents (Ounine *et al.*, 1985), yet despite extensive characterisation of the catabolic gene systems of this organism, no xylose or arabinose isomerase has been characterised. Therefore, research to characterise the pentose isomerase genes of *C. acetobutylicum* was initiated at the BfRC. *In silico* analysis of the *C. acetobutylicum* genome revealed the presence of two pentose isomerases associated with other genes related to the metabolism of pentoses by the pentose phosphate pathway. CRE sequences were identified upstream of cac1342 and cac1346. BLAST homology searches indicated that the deduced sequences of both the ORFs in fact shared the highest identity (%) with a range of arabinose isomerases. When the phylogenetic relationship of the isomerase encoding ORFs and a large group of published, experimentally verified xylose and arabinose isomerases were examined, the two ORFs fell in a cluster of arabinose isomerases. However, the two proteins were most closely related to each other (83% identity, 91% positive). The XylA from *C. thermosaccharolyticum* has been cloned and expressed in *E. coli* DS941 (Meaden *et al.*, 1994). *E. coli* DS941 carries the *xyl5* mutation, therefore has no active xylose isomerase. Plasmid pEB300 did not complement the *xylA* mutation in *E. coli* DS941 and produced a negative fermentation phenotype. Despite extensive efforts to clone cac1342 under the control of the *lac* promoter, only plasmids which were obtained were under the control of the T7 promoter.

I propose three strands of future research on this project. Firstly transform *E. coli* DS941 with pEB200, secondly transform *E. coli* K12 with pEB200 and pEB300, and finally knockout cac1342 and cac1346. Transformation of the *xylA* deficient *E. coli* DS941 with pEB200, would be beneficial as there may be a chance that the cac1342 may still be expressed, despite having ligated under the control of the T7 promoter. To identify if pEB200 and pEB300 have any arabinose isomerase activity, the first step would be procurement of an arabinose positive control and an arabinose utilisation deficient host strain, such as the arabinose isomerase from *Bacillus licheniformis* ATCC 14580 (Prabhu *et al.*, 2008), and the AraA negative *E. coli* K12 (Kim *et al.*, 2002). The absence of cac1342 and cac1346 in the host background, could provide more information about their function. Therefore targeted inactivation of the two genes and examination of the effect on the *C. acetobutylicum* phenotype, is recommended. The plasmids pEB200 and pEB300 are now the subject of a different research project at the BfRC.

Unlike xylose and arabinose, mannose is a PTS sugar. The *C. acetobutylicum* genome encodes two Man family EII proteins. One is encoded within the pSOL1 megaplasmid, the other is encoded within the genome. The first is comprised of EIIAB (cap0066), EIIC (cap0067) and EIID (cap0068) domains, the second is composed of four separate ORFs; EIIA (cac1457), EIIB (cac1458), EIIC (cac1459), EIID (cac1460). The two permeases share little homology, however the plasmid borne EII is homologous to a selection of experimentally verified mannose EIIs. A mannose PTS deficient mutant was sought - *E. coli* ZSC113 (constructed previously by Curtis & Epstein, 1975). The phenotype of this strain was confirmed, competent cells were made, additionally the regions containing the EIIs were amplified and ligated into separate vectors, awaiting transformation into the mannose PTS mutant.

The availability of the *C. acetobutylicum* genome sequence permitted a systematic investigation of the complete complement of phosphotransferases and associated metabolic and regulatory genes. This research has prepared the foundation for the genetic characterisation of the of PTS proteins in *C. acetobutylicum.* Future experiments have been proposed and targets for the generation of knockout mutants identified. Discussion of the PTS function and its relation to the suggested model for global carbon regulation in *C. acetobutylicum* was also presented. The knowledge gained from this research could provide important information towards improving carbohydrate accumulation rates and maximal feedstock consumption. This could be of great use to industry, as carbohydrate accumulation can be a rate-limiting step in solvent production, and incomplete feedstock utilisation damages the economic efficacy of the process. The identification of the complement of PTS proteins could provide targets for cloning and overexpression of the transporters, potentially leading to increased carbon accumulation. Classification of the proteins of the global carbon regulation system could provide targets to knock out sequential utilisation of carbon sources. The characterisation of two functioning β-glucoside utilisation systems points towards solvent performance testing on a β-glucoside rich feedstocks. Furthermore, the identification of the regulatory domains which induce carbon catabolite repression provides sites for genetic manipulation to knock out preferential utilisation of glucose. Two pentose isomerases and two mannose transport proteins were identified, indicating that hemicellulosic rich feedstocks could be efficiently utilised by *C. acetobutylicum.*

A wider implication of this research is that fermentative production of solvents, provides a carbon neutral fossil fuel alternative, which can help reduce our reliance on fossil fuels, use of which has been linked to disruption of the global carbon cycle.

Another benefit of the optimisation of clostridial fermentation is that it can divert biomass waste away from landfill, which is a significant environmental issue, as high BOD waste emits greenhouse gases and can pollute ground water, leading to eutrophication. Thus, this research has the potential to combat short and long-term environmental problems. Additionally, at a time when oil prices are particularly erratic, the ability to produce cheap, clean, renewable fossil fuel substitutes on demand could provide energy security, reduce dependence on oil and inject some much-needed stability into the global economy.

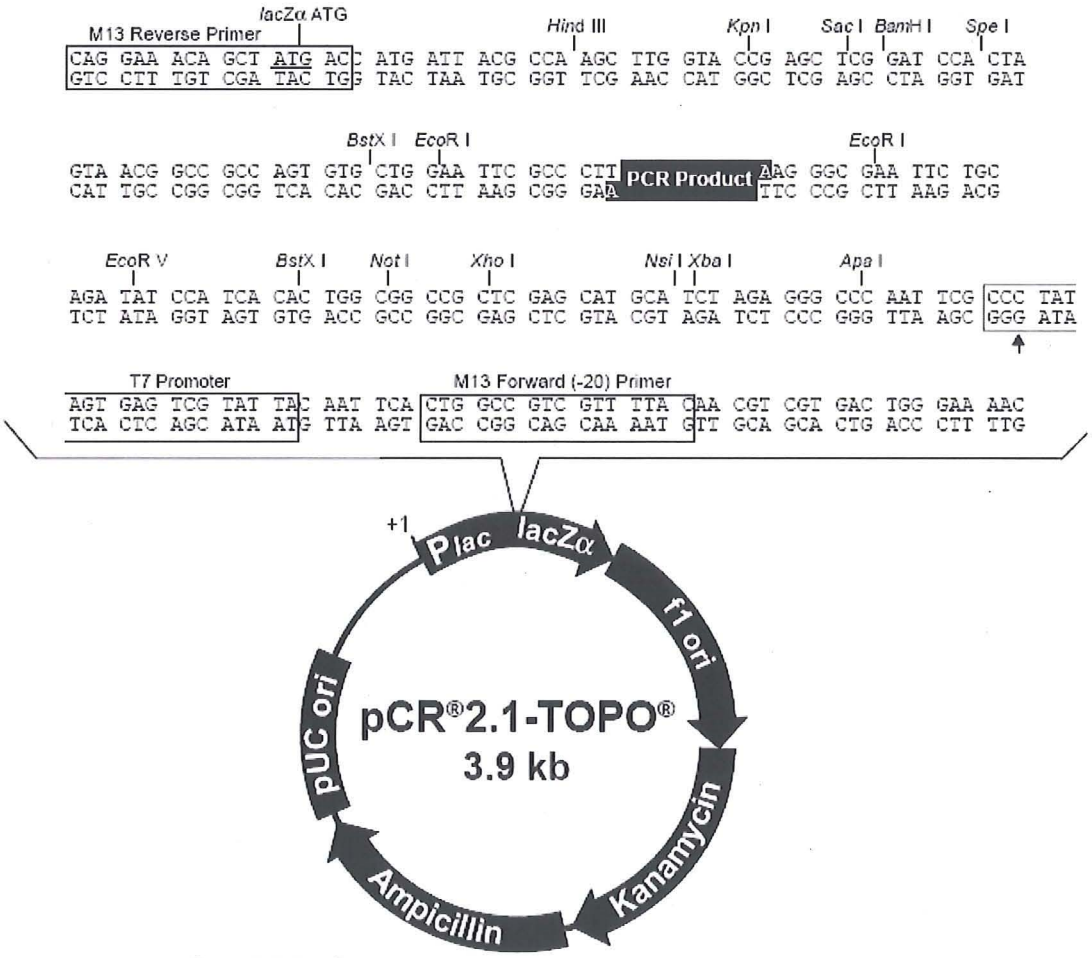
Chapter 7

7 Appendixes

7.1 pCR 2.1 TOPO cloning vector

Obtained from Invitrogen (2006) TOPO TA® Cloning User Manual.

http://www.invitrogen.com/content/sfs/manuals/topota_man.pdf



Comments for pCR®2.1-TOPO®
3931 nucleotides

LacZα fragment: bases 1-547
M13 reverse priming site: bases 205-221
Multiple cloning site: bases 234-357
T7 promoter/priming site: bases 364-383
M13 Forward (-20) priming site: bases 391-406
f1 origin: bases 548-985
Kanamycin resistance ORF: bases 1319-2113
Ampicillin resistance ORF: bases 2131-2991
pUC origin: bases 3136-3809

7.2 pCR 2.1 TOPO cloning vector sequence

Obtained from Invitrogen (2006) http://www.invitrogen.com

AGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTT
TCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAG
GCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAA
ACAGCTATGACCATGATTACGCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGC
TGGAATTCGCCCTTAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGCATGCATCTAG
AGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGG
GAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCG
AAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGC
GGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGC
CCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCG
GGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGAT
GGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTA
ATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGG
GATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAC
AAAATTCAGGGCGCAAGGGCTGCTAAAGGAAGCGGAACACGTAGAAAGCCAGTCCGCAGAAACGGTGCTG
ACCCCGGATGAATGTCAGCTACTGGGCTATCTGGACAAGGGAAAACGCAAGCGCAAAGAGAAAGCAGGTA
GCTTGCAGTGGGCTTACATGGCGATAGCTAGACTGGGCGGTTTTATGGACAGCAAGCGAACCGGAATTGC
CAGCTGGGGCGCCCTCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAG
GATCTGATGGCGCAGGGGATCAAGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAA
GATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGA
CAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGAC
CGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGC
GTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGC
CGGGGCAGGATCTCCTGTCATCCCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCG
GCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCA
CGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAG
CCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGC
CTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTG
GCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTG
ACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGA
CGAGTTCTTCTGAATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTT
TTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCA
GTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCC
GAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACG
CCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCAC
AGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAAC
ACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGG
GGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGA
CACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCT
TCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTC
CGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACT
GGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAA
CGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACT
CATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGA
TAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATC
AAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTAC
CAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGC
GCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCG
CCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCG
GGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACA
GCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACG
CTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGG
AGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCG
ATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTC
CTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTA
TTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGA
GGAAGCGGAAG

7.3 Sequence of the putative β-glucoside operon

The unannotated DNA sequence from the *C. acetobutylicum* ATCC 824 genome containing the region of the putative β-glucoside operon (cac1406-cac1408).

Obtained from: http://gib.genes.nig.ac.jp/single/index.php?spid=Cace_ATCC824

ATGACGATAAAAAAGATATTTAATAACAATGCAATAATAGCTGAAAATTCAGATAAACATGAATTTGTTGTCATGGGGTGTGGAATT
GCATTTAAGAAAAATATAGGAGAAAAAGTAGATGAGAGGCTAATAGAAAAAACTTTTATTCTTAAAGGAAAGGATGCATCCGAAAAG
TTTAAAATGCTGCTGGAAGATGTTCCCACAGAGTATGTTTCGGTATGTTATGACATTATTGAATATGCTAAAAATGTTTTAGATGCA
AAATTAAATGACCATATATATGTTACTTTAACTGATCATGTTAGCAATTGTTTTAAAATGATTGAATCAGGAATTATTATACACAAT
CCATTAATTTGGGAAATCAAAAAGTTTTATCCTAAGGAATTTAAAGTTGGACTTAAAGCTATAGACTTTATAAAGGATGAATTAGGA
AAAGACTTGCCAGAAGATGAGGCAGCTAATATAGCTCTACATTTGATAAATGCACAAATTAACAATTCCTTAAATAATGTTGAAGAT
GCTGCAAAGCAAGCAAAAATGATACAGGATATTTTAAATATAGTTAAGTATACTTATAATATAGTACTTGACGAAAAGTCGATAAGT
TATGAAAGGTTTGTCACACATTTAAGATTCTTCTTTCATAGAATTGATAAGTCAGAAACAGTAGATACAGCAGAGGAAGATTTTTTA
TTAAAGCAAGTTAGAGAAAAATATAAAGATGCATACCAATGTATGCTTAAGGTACAAAAGTATCTTGGAAAAGAATTGTCAGATGAA
GAAAAGCTTTATTTAACAGTTCACATTCAAAGGGTCTCAAAGAAAAATTAAATAAAAATTTGGATTGTTACTGGTAATGCAGGCGAG
ACCAAAATAGACAGTATAAAATAAAATTATGCTATCTATGGGGTCTCGCTTTTTTTATATTATTTATACTAATGAACTTATACATTA
TTCTTATATTTTATAAATCAAAAGAAAAGGAGAGATTATCATGAAATATGAAAAGTTGGCCAAAGATATAATAGAAAATGTCGGAGG
AAAAGAAAACGTTAACAGTTTAACACACTGTATTACCCGTTTACGTTTTAAACTAAAAGACGAAAGTAAAGCAAACACAGATGTGCT
TAAAAATATGGATGGTGTTGTAACCGTTATCAAAAGTGGAGGACAATATCAAGTAGTTATTGGTAATCATGTTCCAGACGTTTATGC
TGATGTAGTAGCAATTGGTGGATTTTCATCTTCATCAGAAGAAGGAACAAAAGAGAAAACAAACTTATTTAACGCATTTATAGACAC
TATTTCAGGTGTATTTACACCAACACTAGGAGTACTTGCTGCAACAGGTATGATAAAAGGCTTTAATGCAATGTTTATAGCTTTTGG
ATGGTTAACTAAAACCTCCGGAACCTACAATATTTTAAATGCAGTAGGCGACTGTTTATTCTACTTCTTCCCTATCTTCTTAGGGTA
TTCAGCAGCTAAGAAGTTTAAAGGTAATCACTTTATTGGTATGGCAATAGGTGCATCACTTGTTTACCCAACATTATCTACTTTGAT
GACAGGAAAACCATTATATACATTGTTTCAAGGAACCATATTTGCATCACCAGTTTACGTTACCTTCCTTGGAATTCCAGTTATACT
AATGAGCTATTCATCAACTGTTATTCCGATAATTTTAGCTTCTTATGTAGGCGTTAAGGTTGAAAAAGCATTTGCAAAGATAATACC
AGATGTAGTAAAAACCTTTTTAGTTCCATTCTGTACTTTACTTGTTATGGTACCTTTATCACTAATAGTTATTGGACCAATCTCTAC
TTGGGCAGGTAAATTATTAGGAGCAGGAACACTAGCTATATATAACTTAAGCCCAATTCTTGCAGGTTTATTTATAGGAGCTTTCTG
GCAGGTATTCGTTATATTTGGGCTTCACTGGGGTTTAGTACCTATAGCAATGAACAATTTATCAGTGCTTCACTATGATCCAATACT
AGCAGGAACTCTTGGAGCTTCATTTGCTCAAACAGGCGTAGTTTTAGCAATATTAATTAAGACAAAGAATGTTAAATTAAAAGGAAT
TGCACTTCCAGCATTTATTTCAGGTATATTTGGTGTTACAGAGCCAGCTATATATGGTGTTACACTTCCTCGTAAAAAACCATTTAT
CATAAGTTGTATAGGAGCAGCTATTGGTGGAGGTATAACAGGCTTCATGGGAACTAAACTATGGATGATGGGGGGACTTGGAATATT
TGCAATACCAAGCTATATCGGAGCAAAAGGAATGGACAGAGGTTTTTATGGTGCGGTTATGTCTGTTGTAATAAGTTTTGTTGTAGG
ATTCTTAATAATGTTCTTTGCTGGCTTTAAGGATGAGGAAGTTAAACAAGAAACAACAAAAAAGAAGAATGAATTAGTAAAGCAAGA
AACTTTAGTAAGTCCATTAAAAGGAAAAATTAAAACATTATCAGAAGTAAAAGATGAAGCTTTTTCAACAGGCTCACTTGGAAAAGG
AATTGCAATTGAACCAGAAGAAGGAAAGCTTGTATCTCCAGTAGATGGTGTTTTAGCAACACTATTTCCAACAGGTCATGCAGTTGG
AATTATAAGTGATAAAGGAACCGAGATATTAATTCATGTGGGCATGGATACAGTTCAATTGGAGGGAAAATATTTTAAAACTATATT
AAAACAAGGAGATCACGTTAAGGCTGGTGATACAATATTAGAATTTGATATACCTAAAATAAAGAAAGCTGGGTATACTCTGACTAC
ACCAGTTGTAGTAACTAATTCTGAAAGCTACCTGGATGTTATTGAAACAGATAAAATTAAGGTTGAGAGAAAAGATCAGTTATTAAC
AGTAATGTTATAAATAACTTTTACTTTGAGCATAAGGTTGACTTGTGCTCAAAGTTTTATAAGGAGAGATGTTTAATGTCAAAAGGT
TTTAGTAAAGAATTTTTATGGGGAGGCGCTACGGCGGCTAATCAGTGTGAAGGTGCTTATTTAGAGGATAATAAAGGGTTATCTACA
GTTGATGTAATTCCAAAGGGAAAGGACCGTTTTCCCGTAGGCTTAGGAAAAATGAAAATGCTTGAGTGTGACAGTGAACATTATTAT
CCAAGTCATGAGGCTATTGATTTTTATCACAGATATAAAGAAGATATAGCTCTATTTCATGAAATGGGCTTTAAATGTTTTAGATTA
TCTCTTGCTTGGTCACGAATATTTCCAAATGGGGACGATGAAATACCAAATGAGGAGGGATTGAAATTTTACGATGCGGTATTTGAT
GAATGTTTAAAATACGGAATAGAACCATTAGTTACTATAACTCATTTTGATGTTCCGGTGAATTTAGTGAAAACAGTAGGTTCCTGG
AGAAGCAGTAAAATGGTAGAGTATTACGAAAAACTGTGTAAGGTTATATTTAGTCGTTATAAAAATAAGGTTAAGTACTGGCTTACC
TTTAATGAAATAAATATGCTATTGCATTTACCTTTTATAGGTGCAGGTTTAGTTTTTGAAGAGGGTGAAAATGAGGAAGCTATAAAA
TATCAAGCAGCACATCATCAGCTTATTGCAAGTGCAAAGGCAACAAAAATAGCTCGTGAAATAAATCCAAATTTTAAAATAGGGTGT
ATGCTAGCTGCTGGAAATAATTATGCAAACACATGTGCACCAGAAGATGTGTGGAGATCTATAGAAAAGGATAGAGAAAACTATTTC
TTTATTGATGTACAATCTCGTGGAGAGTATCCAAACTATGCTAAAAAGATGCTTAAAAGGAAAGGTATAGAGCTTCAAGTAGAGGAT
GAAGATGCTGGAATTTTAAAGAATAATACTGTTGATTTTATAT[illegible]AGTTACTATTCTTCACGCCTAACTAGTGCCAATCCTAAT
ATAAATGGAAATACTGAAGGAAATGTCTTTGCTACCTTGAAAA[illegible]TATCTAAAAGCCAGTGAATGGGGATGGCAAATAGATCCG
CTTGGACTTCGAATTACATTAAATTCTTTATACGATCGTTATCAAAAACCATTATTTATAGTTGAAAATGGTTTGGGAGCCGTTGAT
ACTCCAGATGAAAATGGATATGTAGAAGATGACTATCGAATTGAGTATTTAAGAGAGCACATTAAGGCTATGAGAGATGCAGTTAAT
GTTGATGGTGTAGAACTTATGGGATACACACCTTGGGGATGTATAGATTTAGTAAGTGCTTCAACTGGAGAAATGAAAAAGAGATAT
GGCTTTATATATGTTGATAAGGATAATGATGGGAATGGAACATTAAAGCGTTCTAAGAAAAAGAGTTTTTATTGGTATAAGAAGGTA
ATAGAATCAAATGGTATGGATATAGAGTAA

7.4 Sequence of the putative cellobiose operon

The unannotated DNA sequence from the *C. acetobutylicum* ATCC 824 genome containing the region of the putative cellobiose operon (cac0383-cac0386).

Obtained from: http://gib.genes.nig.ac.jp/single/index.php?spid=Cace_ATCC824

CAACAAAAAAATCACTATTTCTATGTATTGTATAAAAGTGTTTGAACACTAAAATGTTTACATATAAACACTATGCTTTTAATTTAA
ACACTAAATAAAATTGTTGATAAATAAAGACTTTAAGAAGCAAAAAAGTGTTTGGCATGATGTTTGCTTTAATAGTTAAGTGAGAAA
AAAAGTATTTGACTGGAGGGGTACGAATGGATATAGAGCAAATAATAATGAATATAATAATTTATTCCGGCGATGCTAGGAGCTACA
TGTATGAAGCGCTTAGCAAGGCAAGAGAAGGAAAGTATGATGAGATAAATGGCTTAATAAAGCAAGCAAATGATGCTATAGGAAAGG
CTCATGATATACAGACTTCAATGCTTCAAAAAGAAGCAGAGGGCGAGGCACTAAAGGTTTCAATATTATTTGTACATGCGCAGGATC
AATTAATGACCACAATTTCAGAAAAAAATCTTATAACTGAAATGATTGAAATGACAAAGGATATAAATGCTTTGTCCAATAAATAAA
AAGAAAATGTTAGGAGGAACAATTATGGTTAGAATTTTATTATTTTGTAGTGCAGGAATGTCAACTAGTATGTTAGTTTCAAAAATG
AAAAAGGCAGCTGAAGCAAAAAATATAGAAGCATTTATAGAGGCTTATCCTCAAGCACAGATGGCTGATTATGCCGACAAGGCTGAT
ATAGCATTACTTGGGCCTCAAATTAAATTTCTTTTACCAAAAGCAAAAGGAGTATTTGATCCAAAGGGAGTACCAGTAGAAGTAATA
AATTCAGCTGATTATGGAATGCTCAATGGTGAAAAGGTTTTAGATCATGCATTAAAGGTTTTAGGAAAATAAAAGACAAACTTAAGG
AGGTAGAAAATGAAATTTCCAAAGGATTTCTTTTTAGGTGCTGCTTCTGCATCATACCAGGTAGAGGGTGCCTGGAATGAAGATGGA
AAGGGAGTATCTAACTGGGATGTATTTACCAAAATACCAGGAAAAACTTTTGAAGGTACTAATGGAGATGTTGCAGTAGACCACTAC
CATAGATACAAAGAAGACGTAAAACTGATGGCAGAAATGGGACTTGATTCATACAGATTCTCCGTTTCTTGGCCAAGAATAATACCA
GATGGTGATGGAGAAATAAATCAAAAGGGAATAGAGTTTTACAATAATTTGATCGATGAGTGCCTTAAATATGGCATTGTACCATTT
GTTACCTTGTATCACTGGGATATGCCAGAGGTACTGGAGAAGGCTGGAGGATGGACAAATAAAAAAACTGTAGATGCCTTTGTAAAG
TATGCTAAAGCGTGTTTTGAAGCGTTTGGAGATAGGGTAAAGCGCTGGATTACTTTTAATGAGACTATAGTGTTTTGCAGCAATGGA
TACTTAAGTGGAGCTCACCCGCCTGGTATAACAGGAGATGTTAAAAAGTATTTTCAAGCAACACACAATGTATTTACAGCTCATGCC
AGATCAGTTATTGAATATAAAAAATTAAAACAGTATGGCGAAATAGGAATAACGCATGTATTCTCACCAGCCTTCAGTGTTGATGAT
AAGGAAGAAAATAAGGCTGCTGCATATCATGCAAATCAATATGAGATTACCTGGTATTACGATCCAATATTAAAGGGAAAATACCCC
GAATATGTAATTAAAAATATAGAAAAACAAGGATTTTTACCAGATTGGACAGATGAGGAATTGAATACATTAAGGGAGGCAGCTCCG
CTTAATGACTTTATAGGACTTAACTACTATCAACCTCAAAGAGTAATTAAAAATCATGACACAGGGGAAAAAATTGAAAGAACAAGG
GAGAACTCTACAGGGGCACCTGGAAATGCATCCTTTGATGGATTTTATAGAACAGTAAAAATGGACGACAAAACTTACACTAAATGG
GGCTGGGAGATATCACCTGAATCATTAATTTTGGGGCTTGAGAAATTGAAAGAACAGTATGGTGATATCAAAATTTATATTACTGAA
AACGGTTTAGGCGATCAGGATCCAATAATAGAAGATGAAATTCTTGATATGCCTAGAATAAAATTCATAGAAGCTCATTTAAGAGCA
ATAAAAGAAGCTATAAGTAGGGGGATAAACCTTAAAGGATATTACGCTTGGTCAGTAATCGATCTTTTAAGTTGGCTTAACGGATAC
AAAAAACAGTATGGTTTTATATATGTAGACCATAAACACAATTTAGATAGAAAGAAGAAATTATCATTTTATTGGTATAAAAAAGTA
ATAGAAGAAAGAGGTAAAAATATATAGTTTAAACAGGACTTAATTTATATACCATATAAAATTATCGAGGAGGAATTATAATAATGA
GTAAAGTAGATACCATTCAAGAAAAACTCATACCGCCAATAATGAAATTTGCTAACTCTAAAATTATAACTGCCATAAAGGATGGAG
TTATAATCACAATGCCACTTACTCTAGTTGGATCAATATTTTTATTGATTGCTAACTTTCCAATTAAATCCTGGAACTCTTGGATGA
CAGCAATATTTGGACCAAATTGGGCTGTTCCATTAACTCAGGTTTCAGGTTCAACCTTCGATATAATGGCAATTGCGGCTGTAATAG
GAATAGCAGCTCAATATTCAAAAAATGAAGGCCAGGATCCAGTTACAGCAGGACTTTTAGGTGCAGTAGCATTTTTAATTATCATGC
CAAGTTATTCACCAATAAATATAGGTGATAAAGTTGGTGGAATTGTTATGAAACATGCCGACAAGGTTCAAAGTGTAATACCAAAAG
CTTGGACCGGCGGTAAAGGAATGATAGCCGCTTTAATTATTGGTGTACTTACAGGTTATATTTATTCGTGGTTCTTAAAACGTGATA
TAAGAATCAAACTTCCAGAAAGCGTACCATCAGGTGTTTCAAATGCTTTCTCAGCATTAATTCCAGGTGCTGTTATTATAACAGGTT
CTATGTTAATATTCATATTCTTTAATATTGTTACAGGTGCAACTGCACTTGAATGGATTTATAAAGTATTACAAATTCCACTACAAG
GAATAACGGACTCATTAGGTGGAGCAATAGCAATACCATTTGTTATATCTTTCTTCTGGTGGTTTGGTGTTCATGGTGCTACTCTTA
TCGGTGGTGTTATGAACTCTATATACCAAGCAAATACATTAGCAAACCAGGATGTTTTCAAAGCTGGACATAAATTAATAGCTTCAG
GAGCTGGAAAAAATGCTCACATAGTATGTCAGCAGTTCCAAGATAACTTTGTAACCTTAGGAGGTTCAGGAATAACTTTAGGACTAG
TTGCAGCATGTTTCTTATTCGCTAGATCACAACGTCTTAAACAGTTAGGTAAATTAGCCTTAGTTCCAGGTTGCTTTAATATAAATG
AGCCAGTTCTATTTGGTTTACCAATAGTTTTAAATCCACTTATGTTTATACCATTTTTGATTTGTCCTCTTGTTTCAGGAATACTAA
CTTACTTCGCAATATATACTGGTTTAGTACCTTGCTTTACAGCTGTTCAAGCAATATGGACTACACCGCCTGTACTATCGGGTCTTA
TAGTTGCAGGTTGGAGAGGAGCGGTACTCCAGCTTATAATAATTGCAATAGCTGCTATAGTTTACTTCCCATTCTTCAAAAGTCTTG
ATAAACAATACTATCAAGAAGAGATATCTGGAGAAGCAGAATAGAAATCTAAGGTTAATCTAAATTAAGCTTTGGATTAACCCTGTA
AAGAATATGTACTGAAGGAAAACTAAGGTAAAACTTAGCTTACCCATATTAGTATTCAGGGGGGAACAGTGTGGATAGTTAAGTTTG
CCTGGTTTTCATATACATAAACCCACAAACTATATTATTTTAGGAGGATTAAAATGAAATTATTTAAGATAAAAATCACAGGAAGCT
TAGAGGAATTTAAAATTGAGTACTCTTTCTCAACAGATTACTTTAACTACAAAGAGTGTACTTATGAAGGAACAGAACAAGAAAGAT
ACGATCAATTCTATGAGGATTTAAAAACAAATGGAGGACCTCAGCCACTAAACATAAAATTAAAAATGTCAAATGGTGTAATGGATA
GAGCGTTCCCTAAAAAGGATCTCCTAAAATTAAAGAATGTCCAAGACTTCGTCAAAAAAATGTACACCTAAAGAGGTTTGGGCGGAG
CCCATTATATTCATGAACTTAGAAAAAACACGTCATTCAAAAAAGAAAATAAACTAATTAAGGATGTCGTTAATAAAGAGGTTTGGG
CGGAGCCCATTATATTCATGAACTTAGAAAAAACACGTCATTCAAAAAAGAAAATAAACTAATTAAGGATGTCGTTAATAAAGAGGT
TTGGGCGGAGCCCATTATATTCATGAACTTAGAAAAAACACGTCATTCAAAAAAGAAAATAAACTAATTAAGGATGTCGTTAATAAA
GAGGTTTGGGCGGAGCCCATTATATTCATGAACTTAGAAAAAACACGTCATTCAAAAAAGAAAATAAACTAATTAAGGATGTCG

7.5 Sequence of the putative mannose operon

The unannotated DNA sequence from the *C. acetobutylicum* ATCC 824 genome containing the region of the putative mannose operon (cac1457-1460) residing on the chromosome.

Obtained from: http://gib.genes.nig.ac.jp/single/index.php?spid=Cace_ATCC824

AAAAATATAAGAAAATTACAAATGTCATATAGGAAATTACATTTGTAATGAGGTATTAATGACAAATTAT
AGGTGCCTATAAATTAAAATATTATTAAAATTAAAGATGTAAACGTTCGTTAGCGCTGATGAGGTTAGTT
GAATTTAAAAATTACGCATATAAGTTGTTGCTTAACTATATTTAGGCTAGGAGGGGAAAGATGAGATATG
TAATATTAGTAAGTCATGGACAATTTGCATCAGGACTTTATAATGCATTGTCAATGCTTGCAGGTAAAGA
CAGAAATGATGTAATATTTCAAGGATTAGAGGACGGAATGAGTACTGATAAATTTGGAGAAGAATTTAAT
ACAGCTATTGAAAACGTTACAAGTAAAGACGAGATTATTCTTTTCGGAGATATTATTGGTGGATCTCCTT
TGACTACAGCAGTAAATATATTAGCTAGTAAAGAAATGTTATCGAAAGCTTTTGTCCTTGGAGGAATGAA
TTTACCCGTTGTATTAACAGCAATTTTGATGAAGGATTCTATGGGGATAGAAGAGCTAAAGAAAATGTTA
TTTGAAACAGCAGAGGAGTCAATAAAAGAATTTAAAGTTGTAGATACAGATAAGGAAGATGATATTTAGG
AGGGATAAAAATGTCAGTTTCATTTTTAAGAATAGATGACAGAATGATACACGGACAAACATGTACAAGG
TGGGCTTTAGAGTATCCATGTGATGGTATAATAGCTGTTAATGATGCGGCAGCTAATAATCCTGTATTAA
AAGCAGCTTACAAAAGTGCTAGTGGCAAAAAAACATTTGTATGGACATATGAACATTGGAAATTAAAATG
CGATACAGTTTTGAAAAGTAGTACTAGATATTTTGTAATTACAAAAGAACCTATTATTATGTCTAAGATT
TTGGTGGATGATAAATTTAATCCTGGTATTAAAGAGGTTATTGTAGGACCATGCAATGATAGGCCGGGAA
CAGTAAAATTGGGAAATAATCAATCTATAAATCAAAAGGAAGCAGAAGCTTTTGAACGTATTATGCAAGC
CGGCTATAATGTGGAATTTGCTCTTCTTAAAGAAGAAGCAATAGGCAATTGGAAAAAGTTTAGAGGGCAA
TTTGGATTTAAATAGTTTTTAATAATATATTAGGGGGGAAAATAACTATGGGGATAAGTTTTGATTCAGGC
AGCTATATTTGGATTATTTGCATGCCTATCAAGTATGCCCGGTATGGGAGGAACTACCTTTGGTAACTAT
ACCTTAGGAAGACCTTTAGTAGCAGGTTTGCTAGTTGGAATTGTTTTAGGAGACGTACAAACTGGTATTA
TAGTTGGGGCAGCAATTCAAGTTGTTTATATTGCACTGGTTACTCCAGGTGGCACTGTATCTGCTGATGT
ACGTGCAGTTAGTTATATTGGTATTCCTCTAGCTGTAGTAGCAATTAAGGGAATGGGTTTAAATCCAAGT
TCTGCACAGGCAACTCAAATGGCAACTGCATTAGGGGCAGCAGTTGGAACCTTGGGGACTGTTTTATTTT
ATGGAACAGCAACTATAAACTTGATATGGCAGCATATAGGCTGGAAATCAATAGAAAAAGGTGATTTTAA
AAAGCTTTATTTAGTTAATATGGGATTGCCATGGGTTTCTCACATTATTTGTAGTTTTATTCCAGCATTT
ATTATTACAAGAATGGGTTCTAGCATGGTTAAAATTATGAAGGTAGCATTGCCAATGAATGGTATACCAA
TGAAAACTTTATTTACAGTAGGAAGTTTGCTTCCAGCAGTAGGAATTGCAATATTATTAAAACAGGTTGT
ATTAAAACCAAGTGATTTTATAACCTTCTTTTTAGGTTTTACTTTGGCAGCAGTTATGAAAGTTAACTTG
ATTGGAGCAGCAATAATAGGTATTTTCTTTGCAATAATTAATTATAGAATTAGAATGCTTCAACTAGAAA
AGCCATCTGCAGCTAAAGCGACGGGTAACGATGATGATGAGGAGGATATATAATATGGAAAAGAAAAAAA
TATCTAAAAAAGCTTTAAGTAAATCTTTTAAAAATTGGTTTTATGGAAATTTAACTTGTTTTTCTCAAGA
GCACATGCAAACTTTTGGTTATCTTTGTGCAATGCTTCCAATAATTAAAGATTTATATGAAACAAAAGAA
GAACAAAAGGAAGCTATGGAAACTTATAAAGCATTTTTTAATACAGAACCTCAGATTGGAACATTAGTTG
TTGGTATGACAGCCGGTCTTGAAGAGGCAAGAGCTAATCATGAAAATATTGATGGTGAAATGATAAATGG
TATTCGTGCTGGACTTATGGGGCCTTTAGCAGGCATTGGAGATTCATTAATAGTAGGTACTTTAATACCA
ATACTTTTAGGTATAGGGTTAGGACTTTCAACAGGGGGTTCTCCTCTAGGAGCAATATTTTATATTGTAG
TGTGGAATGCTTTAATGATTTTTGGAATGAGATGGGCCTATTATAAAGGATATGCGTTAGGTGGTAAAGC
TGTTCAGATGATTGTTGGAGATAAAGCTAATGCTGTCCGTGAATCTGTAATTATGGTTGGAACTATAGTG
ATTGGTGCAGTTGCTGCTACTTGGGTAAATATTAATACATCTTTGAAAATGTACAATTCACATGGTGGAG
TAATTATAAATTTGCAAAAAACTTTAGATGGCATTTTTCCTAAAATCCTTTCAGCGGCAGCTGTAATTTT
TGCATGGTGGCTTATGTCAAAGAAGAAAATATCCCCTACAATAGTAATGTTAATCTTTTTGATAGTAGCT
TTTGTAGGTGTGCTCTTAGGATTCTTTAATCCAGGACTATCTTATTAATTTGTATGAGGAATCAGTATGG
AAGATGAAGAAAAAAAGCAAATGTTTTATGAGGCAAAGCAGCAAAGTAGGCTCTTAAAAAATTTATCTAA
ATGGTCTAGAAATGTTATGGGATTATCTAGTATAGGTGTTGTAATAGCTTACTATGGTTTGTCACACTCT
GGAATAAAATTTGCTTTTGGAGTTTTTGGTATTTTATTTACTGTAATTTGTGCTTCAGCATGCTTTCTAA
TTAATTTAGCAATAAGGAACGGGAGAAGAAATGTTAATCATATTTTAGAGATGATTAACAGTAAGTAAGA
GAGCTATAGGTGTTATTTCAAACTAATGGTAAAAAGTAAGGAAATACTTTTATTCAAAAAAGACAGTAGA
TAAGCGAGTTTAACTATTGAAATTCATATGCACCACAATGTTCATGAAACAAAATTTTGGTATATATGAA
TTTCAAGTTTTAGTTTGTCAGTATGAGGTTATTAATTAAGAAGACAATGTAGAGTTGTAAAAATAAAATC
AGTGTTAGTAGAATGTATCATTTGAAAACGTTTTTATAAAATGAAGTAATTTAAAGTGTAAGGGGGATAA
ATATGAGATTAAAACAGATAGCTGTGTTATTGCTTAGCAGTATTGTTATTGTTACAGGTAATAATTGGGT
AAAAGCAAACACATCAATATCAAGTACTTCAGCATATAGACGTGAGGCATATCGTCCACAATATCATTTT
ACACCAGATATTGGCTGGATGAATGATCCAAATGGAATGGTTTATTATCATGGAGTATA

7.6 Sequence of the putative mannose operon

The unannotated DNA sequence from the *C. acetobutylicum* ATCC 824 genome containing the region of the putative mannose operon (cap0066-cap0068) residing on the plasmid.

Obtained from: http://gib.genes.nig.ac.jp/single/index.php?spid=Cace_ATCC824

TATAAGAAATACTCTAAGATATTTTTAAAGTTAAAAGAACGTGACATTGAATTAGAAATTTCAAAAGTTCCAGGGGGT
CGAAAAAATAATCTTTTTAGTTGATTCAGAAAGAAATTCACAATAATTAATTTTGGTTACAGGTTTTCCAGAACTTTT
TTATAAATTTTTCTGTAACAAATAAAAAAATAGGAGGTATAAAGATGGTAGGAATTATTCTTGCTAGTCACGGAGAAT
TTGCTAAGGGTATCTTGCAATCGGGTGCGATGATTTTTGGAGACCAAGAGAATGTGCAAGCGGTTACATTAATGCCTA
GTGAAGGCCCTGATGATGTTAAAGCAAAAATGAAAGACGCAATTGCATCCTTTGACAACCAGGATGAGGTTTTATTCT
TAGTTGATCTTTGGGGTGGTACACCATTCAACCAGGCTAATAGTCTATTTGAAGAACATAAAGATAAATGGGCAATCG
TAGCTGGTATGAATCTACCAATGGTAATTGAAGCTTATGGTGCACGTCTTTCAATGGAATCTGCACATGAGATTGCAG
CTAGCATTATAAGCACAGCTAAAGAAGGAGTTAAAGTTAAGCCTGAAGAATTAGAACCAGAAGATGCTGGTAAAGCTT
CTCAGGGTTCTGCAAAGCAATCTAATACAGGTGCACCTGGATCATTCGAATATGTTTTAGCTCGTATTGATTCTCGTT
TACTTCATGGTCAAGTAGCAACTGCTTGGACAAAAGCTATGCAACCTACAAGAATTATTGTAGTATCAGATGCAGTAG
CTAAAGACGAGCTTCGTAAGAAATTGATTCAACAAGCTGCTCCTCCAGGAGTTAAAGCACATGTTGTACCAATTAATC
ACATGATTAAACTTGCAAAAGACGATCAACACTTTGGAGGACAACGTGCAATGCTTCTTTTTGAGAATCCAGAAGATG
TACTAAGAGTAGTAGAAGGTGGCGTACCTCTTAAAACAATCAATGTTGGTTCTATGGCTCACTCTACTGGTAAGGTTC
AACCAAATAAAGTACTTGCTTTCAATCAAGAAGATATTGATACCTTCAATAAGCTTAAACAATCTGGGTTAACTTTTG
ATGTCCGTAAGGTTCCAAATGATTCAAAAGGAAATATGGACGAAATAATTAAAAAGGCCCAAGACGAACTAAATAAAC
AAAAATAATCTAACTATGTATATAGTAAAGGAGGATTAACCGTGACTCTAAATATAATTCAAATGGGATTAGTAGTTA
TTGTAGCGTTTCTAGCTGGTATGGAAGGTATATTGGACGAATTCCATTTCCATCAACCAGTAATTGCTTGTACTTTAA
TCGGATTAGTTACAGGTAACTTAGTACCTTGCTTAATATTAGGTGGTACTCTTCAAATGATTGCCTTAGGTTGGGCAA
ATATAGGTGCTGCTGTAGCGCCTGATGCAGCTTTAGCATCTGTTGCATCCGCAATTATTTTAGTTCTTGGAGGACAAG
GAAAAGCAGGAGTTTCTTCAGCTATTGCTATTGCTGTTCCACTAGCAGTTGCAGGGCTATTATTACAAACTATTTGTC
GTACAATTGGTATAATCATTATACATCGTATGGATGCTGCTGCTGAAGAAGGAAATATAAGAAAAATTGAAATGTGGC
ATATTATTGCTATTTGCATGCAGGGTGTACGTATTGCAATTCCAGCAGCTTTGATTTTAGCAATTGGTGCTGGTCCTA
TTCGTTCATTACTTCAAGCTATGCCTCTTTGGTTGACAGATGGTTTAGCAATAGGTGGTGGAATGGTTGTAGCTGTTG
GTTATGCAATGGTAATCAATATGATGGCTACAAAAGAAGTATGGCCATTCTTCGCAATTGGTTTTGTGTTAGCAACAG
TTTCACAAATTACACTTATCGGACTAGGTGCAATTGGTTTAGCTTTAGCTCTTCTTTACTTATCGCTTTCTAAACAAG
GCGGCTCAGGTAAGGGTGGTGGATCAAATACTGGTGATCCATTGGGCGATATTATCGATAGATACTAAGAAAGGAGAG
AACACGAAAATGGCAAATGAATTAAAATTAACAAAGAAGGATCGTATTTCTGTTTGGTGGCGTTCAACTTTCATTCAA
GGTTCTTGGAACTATGAAAGAATGCAAAACGGTGGTTGGACATTTGCATTAATTCCTGCAATCAAAAAATTATATAAG
AATAAAGAAGATCGTGTAGCTGCATTGAAACGTCACTTAGAGTTTTTTAACACTCACCCATATGTAGCTTCACCAATC
ATTGGTGTAACATTAGCCCTAGAAGAAGAACGTGCAAATGGTGCACCAGTTGATGATGTAACTATTCAAGGTGTTAAA
GTTGGTATGATGGGACCTTTAGCAGGTATCGGAGATCCAGTTTTCTGGTTTACTGTTAAGCCAATTTTAGGAGCATTA
GCAGCTTCACTTGCTATGAGTGGTAACATTCTTGGACCATTTATATATTTCTTTGCTTGGAATGCCATCCGTATGTTA
TTTATGTGGTATACACAAGAGTTTGGTTACAAAGCAGGCTCTCGTATTACCGATGATCTATCAGGTGGTTTACTACAA
GACGTTACAAAAGGAGCATCCATCCTTGGTATGTTTATTTTGGGATCGTTAGTTAATAGATGGGTATCTGTTAAATTT
ACACCGGTAGTATCCTCTGTTAAGTTAAGTGATGGTGCATATATTGTTTGGGATAAACTTCCGGCTGGAGTTAAAGGT
ATTAAAGAAGCTTTGATTCAGCAGTCATCTGGTTTGTCGTTAACGGATCATAAAGTTACAACACTACAAAATAACCTG
GATTCATTAATTCCTGGACTTGCAGGATTAGCAGTAACATTTATTTGTATGTGGTTACTTAAGAAGAAAGTATCTCCA
ATTGTTATTATTTTTGGTTTATTCATAATCGGTATTGTTTTCCATTTAATTGGTTTAATGTAATATTTCTTACTTAGC
CTGGGCTTTGGCTAAACATATTATTGATTTTAATTTAATATGTTTAACTTAAGCCTAGGCTTTTGTTAAGAAATAGAT
TGAGAAATAATTTTTATGTATCTTGAGACGCTTAATTATTTTATTAATAAAAATGTATGAAAAAACTGCGATAATGCA
TTCTAATATAAGAAGTGTTATATTATCCTATTTCTCTTACAATGGCTTCTTAATTTCTAAAAAAAACATATAATAAAA
ACAAATGATTGTATACATTAGAAAAGAGGTAGATCGTATGGTTCAATCACTCAATACAAAGGTTGACTTAGTAATTAA
TGCAACATCTTTTACAGGGTTTTCAGATTACGGCAGAATCATGATTGGAGACAAAGGTTTTGAGTTTTATAATTCCCG
TGATTCTCGCAAATTTATTCAAATTCCCTGGGAAGAAGTAGACTATGTTATTGCATCCGTATTACTAAAAGGAAAGTG
GATTCCACGATATGCAATTAAAACTAAAAAAAAATGGTACATATACTTTTGCTTCTAAGGATGCAAAAAAAGTGCTTCG
TGTTATTCGAAAATATGTTGATCCAAATCATATGGTTCATTCGTTAAGCTTTTTTGATGTAGTCAAACGATTTGTAAA
ATCAAGGTTTAAGAAGAGCTGATGAATTAATATGTAAATATAATTAAAGACATATCAATGAGATATGTCTTTAAGAAA
GCCCAAAAGTACATAAATATTTATTTTTTCATGAGCTTATTTATATAGTTAGAGCTCTAAACCCTTCATATATTTGCT
TAATGTATAAGCTATGCCATCTTCTTCGTTTGATAGGGTAACTTCGTCTGCTGCATTCTTTAACTCCTGCACAGCATT
ATCCATAGCTACTCCTATACCAGCATATTTAACCATTGAAATATCATTATGTCCATCTCCAAAGGCAATCATTTCTTC
TTTTTTATACCCCATAGGAATTAGTACCGTATCCAGTGCTTTAGCTTTATCAATTCCTTCAGCAGTAAATTCAAAATA
GAAATCTGCTGTAAACATACAGTTTAAGCTTTCCTTAAAGGGTTCCATCATTTCTTTATAATGTTTTTTTAAATATTC
AGGCTCCCCAGCTGTTAAAATTTTATTTAAAGGATAGTCTGCAAAGGCAGCCAAATCATGTTTCTCACATAACTTAAA
TTTATTATTACGTGACTCATATTGAATTATGTTAAAAGGTTCAC

7.7 Sequence of the putative pentose isomerase

The unannotated DNA sequence from the *C. acetobutylicum* ATCC 824 genome containing the region of the putative pentose isomerase (cac1342)

Obtained from: http://gib.genes.nig.ac.jp/single/index.php?spid=Cace_ATCC824

ACCTGGGGAACTGACCCTAAGAATGCAGTGCATAATTCTGTTGTATTAGAAGAGGTTGCT
AAAATGACATATCATTCTTTACAGCTTAATCCACATAACATTGAAATGAGTCAGGATTTG
TTAGATAAGCATTTTAAAAGAAAACATGGTGCAAATGCATATTATGGTCAAAAAACAAAG
TAAGAAAAGAGGATATAAAAATGCTAAAGAATAAAAAATTAGAATTTTGGTTTGTAGTAG
GTAGTCAGAATTTATACGGTGAAGAAGCATTAAATGCAGTAAAAAAAGATTCTAAAGAAA
TTGTGGATTCTTTAAATGAAAGTGGAAAATTACCATATCCTATTGTATTTAAAACTCTAG
CTACATCAGCAGATGAAATTAAGAATATAGTTAAAGAAATAAATTATAGAGATGAAGTAG
CAGGTGTTATTACTTGGATGCATACATTTTCTCCTGCTAAAATGTGGATAGCAGGAACAA
AGCTTTTACAAAAACCACTATTGCATCTAGCAACACAATTTAACGAAAATATTCCTTGGA
AAACTATTGATATGGATTATATGAATTTACATCAAAGTGCACATGGTGATAGGGAATATG
GATTTATTAATGCTAGATTAAACAAAAATAACAAAGTTGTTGTAGGATATTGGAAGGATA
ATCAAGTTCAAAAGGAAATTGCAGAGTGGATGCAGGTTGCTTATGGGTACGTTGCAAGTG
AAAATATAAAGGTTGCAAGATTTGGGGATAACATGCGTAATGTTGCCGTAACAGAAGGAG
ATAAAGTAGAAGCCCAAATTCAATTTGGATGGACAGTTGATTATTTTGCTATTGGTGATT
TAGTGGCTGAAATGAACAAAGTTTCACAAAAAGATATAGATGCTACTTATGAAGAGTTTA
AGGATATTTATATATTAGATATTGGAGATAATGACCCTGAATTTTATGAGAATCATGTAA
AAGAACAAATTAAGATAGAGATAGGTTTACGTAATTTTCTAGAGGCAGGTAATTATACAG
CATTTACAACAAACTTTGAGGATCTTTATGGAATGAAGCAATTACCTGGACTTGCAGTTC
AACGTTTAAATGCTGAAGGTTATGGTTTTGCAGGTGAAGGCGATTGGAAAACAGCAGCAC
TTAATCGTTTATTTAAAATTATGACTGACAACAAGAAAACTGGATTTATGGAAGACTATA
CTTATGAGCTAAGTGCTGGAAATGAAAGAATTTTGGGAGCACATATGCTAGAAGTTGATC
CAACACTTGCAGCTAGTAAACCAAGAGTTGTAGTTAAACCACTTGGAATTGGAGATAAGG
AAGCACCAGCACGTTTAATATTTGATGGAGTTGTAGGTGATGGAGTAGTTGTATCTATGC
TTGATTTAGGGACACACTATCGTTTACTTATTAACGAAGTAAAGGCAGTTAAACCTACTG
AGGATGCTCCTAATTTACCTGTAGCAAAGCTAGTATGGCAGCCACAACCAAACTTTAAAG
ATGCAGTTAAAGCATGGATTTATGCTGGAGGTGGACATCATACTGTTGCAACCTTAGAAT
TAACAGTTGAGCAAGTTTATGACTGGAGTCGTATGGTTGGTTTAGAAACAATAGTTATTG
ATCATAATACCAATTTAAGAGATATTATAAAAGAGACTTCAAGATAAATTGTATTTTTAG
AGAATAAAAGTAACAACTATGAAGTTAAAGGACTAATTAACTTCATAGTTGTCAAAATAT
AATTAAATTTATACGTACAAATTACTAAAAAGTCATTGACATATGCGTACAAATATTTTA
CTATACGTATAGTACATATAATATTTGTTTTAAGACATTTAATTAAAAATTAGTAGAAAA
TCTAGAA

7.8 Sequence of the putative pentose isomerase

The unannotated DNA sequence from the *C. acetobutylicum* ATCC 824 genome containing the region of the putative pentose isomerase (cac1346)

Obtained from: http://gib.genes.nig.ac.jp/single/index.php?spid=Cace_ATCC824

TAATCTTCTTCGGTTTTGCTGCAATTTGTATCATAGGATTTTTATTTGCAAAATATGTTC
TTTATGAAACAAAAGGAAAATCTTTAGAAGAAATTGAGACATATTTGTACAATCGTTCTA
TTGGAAAAGTTAGAGGATTAAATGAGTAGTTTATAAATGTTTTTTAGAATAAATAAAAAA
TAGATAAGGGGAGATTTATCATGTTAGAAAATAAAAAGATGGAATTTTGGTTTGTAGTAG
GAAGTCAACATTTATATGGTGAAGAGGCTTTAAAAGAAGTAAGAAAAAATTCTGAGACAA
TTGTAGATGAATTAAATAAAAGTGCTAATCTTCCATATAAAATAATATTTAAAGATTTAG
CAACTTCTGCTGATAAAATAAAGGAAATAATGAAGGAAGTTAACTATAGAGATGAAGTAG
CAGGAGTTATAACTTGGATGCATACGTTTTCTCCAGCTAAAATGTGGATAGCAGGTACAA
AGATATTACAAAAACCTTTACTTCATTTTGCAACTCAATATAATGAAAATATTCCATGGA
AAACAATAGATATGGATTATATGAACTTACATCAAAGTGCTCATGGAGATAGAGAGTATG
GATTTATTAATGCAAGACTTAAAAAGCATAATAAAGTTGTTGTAGGATATTGGAAGGATA
AAGAAGTTCAAAAACAAGTTTCAGATTGGATGAAGGTTGCTGCAGGATATATTGCAAGTG
AAAGCATAAAAGTTGCACGTTTTGGTGATAACATGCGTAATGTTGCAGTTACAGAGGGAG
ATAAAGTAGAAGCTCAAATACAATTCGGATGGACAGTAGATTACTTTGGTATAGGTGATT
TAGTTGCTGAAATGGACAAGGTAAGCCAAGATGAAATAAATAAAACTTATGAAGAATTTA
AAGATTTATATATTTTAGATCCAGGTGAAAATGATCCTGCTTTCTATGAGAAACAAGTTA
AAGAACAAATCAAAATTGAAATAGGATTAAGAAGGTTCTTAGAAAAAGGAAATTATAATG
CATTTACAACAAACTTTGAAGATCTTTATGGAATGAAACAGTTACCTGGACTTGCAGTAC
AACGTTTAAATGCTGAAGGCTATGGCTTTGCAGGCGAAGGAGACTGGAAAACTGCAGCTT
TAGATAGATTATTAAAGGTTATGACTAATAATACTGCTACAGGTTTTATGGAAGATTACA
CATATGAACTTAGTCGTGGAAATGAGAAGGCATTAGGAGCTCATATGCTTGAAGTTGACC
CAACTTTTGCTTCAGATAAACCAAAGGTTATTGTTAAACCACTAGGAATTGGAGATAAAG
AAGATCCAGCACGTTTAATCTTTAATGGTTCAACAGGAAAAGGTGTAGCAGTTTCAATGC
TTGATTTAGGAACACATTATCGTTTAATAATAAACGGACTTACAGCAGTGAAACCAGATG
AAGACATGCCAAACCTACCAGTTGCTAAAATGGTATGGAAACCAGAACCAAACTTCATTG
AAGGAGTTAAATCTTGGATTTATGCAGGTGGCGGACATCATACAGTGGTTTCACTAGAAT
TAACAGTAGAACAGGTTTATGATTGGAGTCGTATGGTAGGCTTGGAAGCTGTAATAATAG
ATAAGGATACTAAATTAAGAGATATAATAGAAAAGACAACAAAATAAAAATTATATTTTA
AAAGAATATCAATAAAATTCAACACATAATTAGGATTTTAGAGCATATATAACTGTTTTT
AAATGTAACTTAGAGTAATTAAAGAAGCTTTGTTTTGAATAAGTATATAAAGAACACTTA
TTTAAAACAAAGCTAAAAAAAATATAGGAGGCATAAGCAATGAAACTTTTTATAGATACA
GCTAATG

7.9 BioLine hyperladders

DNA molecular weight markers obtained from:

http://www.bioline.com/h_ladderguide.asp

HyperLadder I

SIZE (bp)	ng/BAND
10000	100
8000	80
6000	60
5000	50
4000	40
3000	30
2500	25
2000	20
1500	15
1000	100
800	80
600	60
400	40
200	20

1% agarose gel

- Higher intensity bands: **1000bp and 10000bp**
- Supplied in a ready-to-use format
- Each lane (5µl) provides 720ng of DNA

HyperLadder II

SIZE (bp)	ng/BAND
2000	50
1800	20
1600	20
1400	20
1200	20
1000	100
800	30
700	30
600	30
500	30
400	30
300	100
200	40
100	40
50	40

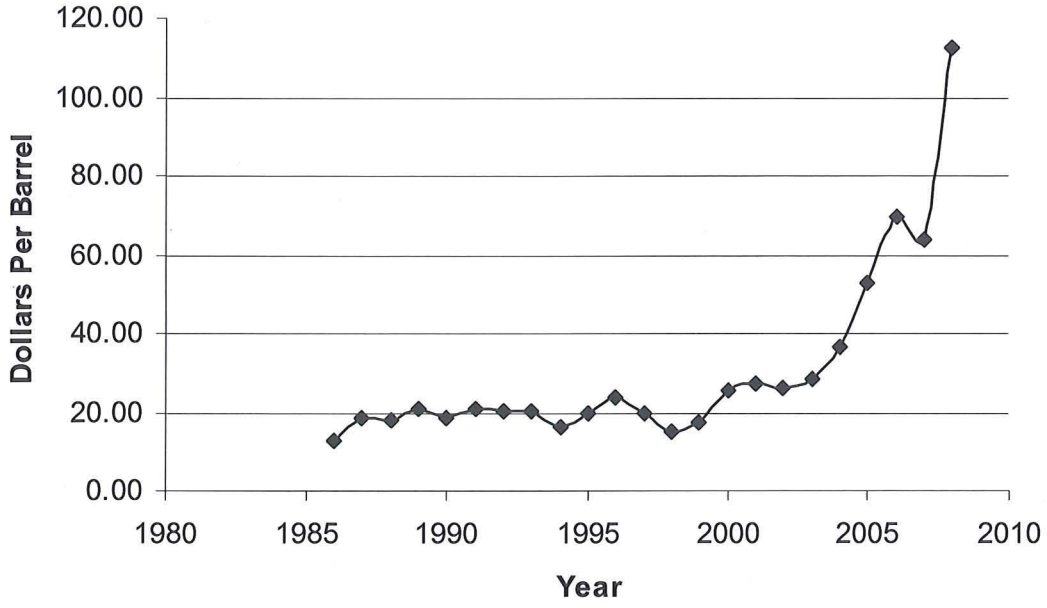
1.5% agarose gel

- Higher intensity bands: **300bp, 1000bp and 2000bp**
- Supplied in a ready-to-use format
- Each lane (5µl) provides 600ng of DNA

7.10 CBM recipes from selected papers

The defined medium contained the following (per litre):	**Glucose** (Tangney & Mitchell, 2007)	**Sucrose** (Tangney & Mitchell, 2000)	**Glucitol** (Tangney *et al.*,1998)	**Sucrose** (Tangney *et al.*, 1998)	**Mannitol** (Behrens *et al*, 2001)	**Glucose** (Tangney & Mitchell 2007)	**DDGS** (Ezeji & Blascheck, 2007)	**Various** (Ezeji *et al.*, 2004)
Sugar								
carbon source	10g	10g	5%	?	20g	10g	60g	60-100g
yeast extract	-	-	-	-	-	-	2g	1g
casein hydrolysate	4g	4g	-	-	-	4g	-	-
ammonium acetate	-	-	2.2g	2.2g	2.2g	-	220g	220g
Mineral								
$MgSO4.7H_2O$	0.2g	0.2g	0.2g	0.2g	0.4g	0.2g	20g	20g
$MnSO_4.4H_2O$	10mg	10mg	0.01g	0.01g	10mg	10mg	1g	1g
NaCl	10mg	10mg	0.01g	0.01g	10mg	10mg	1g	1g
$FeSO_4.7H_2O$	10mg	10mg	0.01g	0.01g	10mg	10mg	1g	1g
Vitamin								
para-aminobenzoic acid	1mg	1mg	0.01g	0.01g	10mg	1mg	1g	0.1g
thiamine HCL	1mg	1mg	-	-	-	1mg	0.1g	0.1g
d-biotin	2µg	2µg	0.001g	0.001g	1mg	2µg	0.001g	0.001g
Phosphates								
KH_2PO_4,	0.5g	0.5g	0.5g	0.5g	0.5g	0.5g	50g	50g
K_2HPO_4,	0.5g	0.5g	0.45g	0.45g	0.5g	0.5g	50g	50g

7.11 Crude oil prices



April F.O.B Crude Oil Spot Prices.

Crude oil spot prices (in April) between the years of 1986-2008 (dollars per barrel). The F.O.B Price is the price actually charged at the producing country's port of loading. Raw data obtained from:

http://tonto.eia.doe.gov/dnav/pet/hist/rwtcM.htm

Chapter 8

8 References

Abe, H., T. Abo, & H. Aiba (1999) Regulation of intrinsic terminator by translation in *Escherichia coli*: transcription termination at a distance downstream. *Genes to Cells* **4**:87-97.

Ahmad F., J. B. Hughes, & G. N. Bennett (2005) Biodegradation of hazardous materials by Clostridia. In: Handbook on Clostridia (Editor, P. Dürre) CRC Press. ISBN: 0-8493-1618-9.

Albertsson, A., & S. J. Huang (1995) Degradable polymers, recycling and plastics waste management. (Editors, A. C. Albertsson, & S. J. Huang) CRC Press. ISBN: 0-8247-9668-3.

Alpert, C. A., & B. M. Chassys (1990) Molecular cloning and DNA sequence of *lacE*, the gene encoding the lactose-specific enzyme II of the phosphotransferase system of *Lactobacillus casei*. Evidence that a cysteine residue is essential for sugar phosphorylation. *Journal of Biological Chemistry* **265**(36):22561-22568.

Alpert, C. A., M. Dörschug, D. Saffen, R. Frank, J. Deutscher, & W. Hengstenberg (1985) The bacterial phosphoenol pyruvate-dependent phosphotransferase system: Isolation of active site peptides by reversed-phase high-performance liquid chromatography and determination of their primary structure. *Journal of Chromatography* **19**(326):363-371.

Alsaker, K. V., T. R. Spitzer, & E. T. Papoutsakis (2004) Transcriptional analysis of *spo0A* overexpression in *Clostridium acetobutylicum* and its effect on the cell's response to butanol stress. *Journal of Bacteriology* **186**:1959-1971.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, & D. J. Lipman (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**(17):3389-3402.

An, C. L., J. L. Woo, S. Y. Hong, E. C. Shin, M. K. Kim, J. R. Lee, S. R. Park, J. G. Woo, Y. P. Lim, & H. D. Yun (2005) Structural and biochemical analysis of the *asc* operon encoding 6-phospho-β-glucosidase in *Pectobacterium carotovorum* subsp. *carotovorum* LY34. *Research in Microbiology* **156**(2):145-153.

An, C. L., W. J. Lim, S. Y. Hong, E. J. Kim, E. C. Shin, M. K. Kim, J. R. Lee, S. R. Park, J. G. Woo, Y. P. Lim, & H. D. Yun (2004) Analysis of *bgl* operon structure and characterization of β-glucosidase from *Pectobacterium carotovorum* subsp. *carotovorum* LY34. *Bioscience, Biotechnology, and Biochemistry* **68**(11):2270-

2278.

Anderson, G. Q. A., & M. J. Fergusson (2006) Energy from biomass in the UK: sources, processes and biodiversity implications. *The International Journal of Avian Science IBIS* **148**:180-183.

Aresta, M., A. Dibenedetto, & G. Barberio (2005) Utilisation of micro-algae for enhanced CO_2 fixation and biofuels production: development of a computing software for an LCA study. *Fuel Processing Technology* **86**:1679-1693.

Arnaud, M., M. Debarbouille, G. Rapoport, M. H. Saier, & J Reizer (1996) *In vitro* reconstitution of transcriptional antitermination by the SacT and SacY proteins of *Bacillus subtilis*. *Journal of Biological Chemistry* **271**:18966-18972.

Asanuma, N., T. Yoshii, & T. Hino (2004) Molecular characteristics of phosphoenolpyruvate: mannose phosphotransferase system in *Streptococcus bovis*. *Current Microbiology* **49**:4-9.

Atlas, R. M., & L. C. Parks (1993) Handbook of Microbiological Media. CRC Press. ISBN: 0-8493-1818-1.

Aymerich, S., & M. Steinmetz (1992) Specificity determinants and structural features in the RNA target of the bacterial antiterminator proteins of the BglG/SacY family. *Proceedings of the National Academy of Sciences* **89**(21):10410-10414.

Bachem, S., & J. Stülke (1998) Regulation of the *Bacillus subtilis* GlcT antiterminator protein by components of the phosphotransferase system. *Journal of Bacteriology* **180**(20):5319-5326.

Ballesteros, M., J. M. Oliva, M. J. Negro, P. Manzanares, & I. Ballesteros (2004) Ethanol from lignocellulosic materials by a simultaneous saccharification and fermentation process (SFS) with *Kluyveromyces marxianus* CECT 10875. *Process Biochemistry* **39**:1843-1848.

Barabote, R. D., & Jr. M. H. Saier (2005) Comparative genomic analyses of the bacterial phosphotransferase system. *Microbiology and Molecular Biology Reviews* **69**:608-634.

Batycka, M., N. F. Inglis, K. Cook, A. Adam, D. Fraser-Pitt, D. G. Smith, L. Main, A. Lubben, & B. M. Kessler (2006) Ultra-fast tandem mass spectrometry scanning combined with monolithic column liquid chromatography increases throughput in proteomic analysis. *Rapid Communications in Mass Spectrometry* **20**(14):2074-2080.

Behrens, S., W. J. Mitchell, & H. Bahl (2001) Molecular analysis of the mannitol operon of *Clostridium acetobutylicum* encoding a phosphotransferase system and a putative PTS-modulated regulator. *Microbiology* **147**:75-86.

Belouski, E., D. E. Watson, & G. N. Bennett (1998) Cloning, sequence, and expression of the phosphofructokinase gene of *Clostridium acetobutylicum* ATCC 824 in *Escherichia coli*. *Current Microbiology* **37**:17-22.

Benemann, J. R. (1993) Utilization of carbon dioxide from fossil fuel-burning power plants with biological systems. *Energy Conversion and Management* **34**:999-1004.

Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, & D. L. Wheeler (2008) GenBank. *Nucleic Acids Research* **36**(Database Issue):D25-D30.

Bertrand, K., C. Squires, & C. Yanofsky (1976) Transcription termination *in vivo* in the leader region of the tryptophan operon of *Escherichia coli*. *Journal of Molecular Biology* **103**:319-337.

Bochman, M., F. A. Cotton, C. A. Murillo, & G. Wilkinson (1999) Advanced Inorganic Chemistry – A Comprehensive Text. (Editor, F. A. Cotton) John Wiley & Sons, Inc. ISBN: 0-4711-9957-5.

Bomb, C., K. McCormick, E. Deurwaarder, & T. Kbergera (2007) Biofuels for transport in Europe: lessons from Germany and the UK. *Energy Policy* **35**:2256-2267.

Bor, Y.-C., C. Moraes, S.-P. Lee, W. Crosby, A. J. Sinskey, & C. A. Batt (1992) Cloning and sequencing the *Lactobacillus brevis* gene encoding xylose isomerase. *Gene* **114**:127-131.

Bowles, L. K., & W. L. Ellefson (1985) Effects of butanol on *Clostridium acetobutylicum*. *Applied and Environmental Microbiology* **50**(5):1165-1170.

Brett, D. W. (2005) Clostridial collagenase in wound repair. In: Handbook on Clostridia (Editor, P. Dürre) CRC Press. ISBN: 0-8493-1618-9.

Brown, G. D., & J. A. Thomson (1998) Isolation and characterisation of an aryl-β-d-glucoside uptake and utilisation system (*abg*) from the Gram-positive ruminal clostridium species *C. longisporum*. *Molecular and General Genetics* **257**:213-218.

Brown, L. M., & K. G. Zeiler (1993) Aquatic biomass and carbon dioxide trapping. *Energy Conversion and Management* **34**:1005-1013.

Brückner, R., & F. Titgemeyer (2002) Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization. *FEMS Microbiology Letters* **209**:141-148.

Bud, R. (1993) The origins of zymotechnology. In: The Uses of Life: A History of Biotechnology. Cambridge University Press. ISBN: 0-521-38240-8.

Cassman, K. G., & A. J. Liska (2007) Food and fuel for all: realistic or foolish? *Biofuels, Bioproducts and Biorefining* **1**(1):18-23.

Cato, E. P., & E. Stackebrandt (1989) Taxonomy and phylogeny. In: Biotechnology Handbook 3 Clostridia (Editors, N. P. Minton & D. J. Clarke) Plenum Press. ISBN 0-3064-3261-7.

Cato, E. P., W. L. George, & S. M. Finegold (1986) Genus Clostridium. In: Bergey's Manual of Systematic Bacteriology (Volume II) Gram-Positive Bacteria Other Than Actinomycetes (Editors, P. Sneath, N. Mair, M. Sharpe, & J. Holt) Williams and Wilkins. ISBN: 0-683-07893-3.

Chaillou, S., Y-C. Bor, C. A. Batt, P. W. Postma, & P. H. Pouwels (1998) Molecular cloning and functional expression in *Lactobacillus plantarum* 80 of *xylT*, encoding the D-xylose-H+ symporter of *Lactobacillus brevis*. *Applied and Environmental Microbiology* **64**:4720-4728.

Charles, M. B., R. Ryan, N. Ryan, & R. Oloruntoba (2007) Public policy and biofuels: the way forward? *Energy Policy* **35**:5737-5746.

Chelf, P., L. M. Brown, & C. E. Wyman (1993) Aquatic biomass resources and carbon dioxide trapping. *Biomass and Bioenergy* **4**(3):175-183.

Chen, Q., H. Engelberg-Kulka, & O. Amster-Choder (1997) The localization of the phosphorylation site of BglG, the response regulator of the *Escherichia coli bgl* sensory system. *Journal of Biological Chemistry* **272**:17263-17268.

Chiao, J. S., & Z. H. Sun (2007) History of the acetone-butanol-ethanol fermentation industry in China: development of continuous production technology. *Journal of Molecular Microbiology and Biotechnology* **13**:12-14.

Chisti, Y. (2007) Biodiesel from microalgae. *Biotechnology Advances* **25**:294-306.

Clarke, D. J., C. D. Morley, D. B. Kell, & J. G. Morris (1982) On the mode of action of the bacteriocin butyricin 7423: effects on membrane potential and potassium-ion accumulation in *Clostridium pasteurianum*. *European Journal of Biochemistry* **127**:105-116.

Claassen, P. A. M., M. A. W. Budde, A. M. Lopez-Contreras, A. M. (2000) Acetone, butanol and ethanol production from domestic organic waste by solventogenic clostridia. *Journal of Molecular Microbiology and Biotechnology* **2**(1)39 -44.

Clifford, M. N. (2000) Miscellaneous phenols on food and beverages – nature, occurrence and dietary burden. *Journal of the Science of Food and Agriculture* **80**:1126-1137.

Cochu, A., C. Vadeboncoeur, S. Moineau, & M. Frenette (2003) Genetic and biochemical characterization of the phosphoenolpyruvate: glucose/mannose phosphotransferase system of *Streptococcus thermophilus. Applied and Environmental Microbiology* **69**:(9)5423-5432.

Cole, K. H., G. R. Stegen, & D. Spencer (1993) The capacity of the deep oceans to absorb carbon dioxide. *Energy Conversion and Management* **34**:991-998.

Collins, M. D., P. A. Lawson, A. Willems, J. J. Cordoba, J. Fernandez-Garayzabal, P. Garcia, J. Cai, H. Hippe, & J. A. E. Farrow (1994) The phylogeny of the Genus *Clostridium*: proposal of five new genera and eleven new species combinations. *International Journal of Systematic and Evolutionary Microbiology* **44**:812-826.

Cornillot, E., R.V. Nair, E. T. Papoutsakis, & P. Soucaille (1997) The genes for butanol and acetone formation in *Clostridium acetobutylicum* ATCC 824 reside on a large plasmid whose loss leads to degeneration of the strain. *Journal of Bacteriology* **179**:5442-5447.

Cote, C. K., & A. L. Honeyman (2003) The LicT protein acts as both a positive and a negative regulator of loci within the *bgl* regulon of *Streptococcus mutans. Microbiology* **149**:1333-1340.

Cote, C. K., D. Cvitkovitch, A. S. Bleiweis, & A. L. Honeyman (2000) A novel β-glucoside-specific PTS locus from *Streptococcus mutans* that is not inhibited by glucose. *Microbiology* **146**:1555-1563.

Crutz, A. M., & M. Steinmetz (1992) Transcription of the *Bacillus subtilis sacX* and *sacY* genes, encoding regulators of sucrose metabolism, is both inducible by sucrose and controlled by the DegS-DegU signalling system. *Journal of Bacteriology* **174**:6087-6095.

Curtis, S. J., & W. Epstein (1975) Phosphorylation of D-glucose in *Escherichia coli* mutants defective in glucosephosphotransferase, mannosephosphotransferase, and glucokinase. *Journal of Bacteriology* **122**:1189-1199

Dabrock, B., H. Bahl, & G. Gottschalk (1992) Parameters affecting solvent production by *Clostridium pasteurianum*. *Applied and Environmental Microbiology* **58**(4):1233-1239.

Dalet, K., Y. Cenatiempo, P. Cossart, The European Listeria Genome Consortium, & Y. Héchard (2001) A σ54-dependent PTS permease of the mannose family is responsible for sensitivity of *Listeria monocytogenes* to mesentericin Y105. *Microbiology* **147**:3263-3269.

Davison, S. P, J. D. Santangelo, S. J. Reid, & D. R. Woods (1995) A *Clostridium acetobutylicum* regulator gene (*regA*) affecting amylase production in *Bacillus subtilis*. *Microbiology* **141**(4):989-996.

Dawber J. D., & A. T. Moore (1985) The structure and properties of biopolymers. Chemistry for the Life Sciences 2nd Edition. MacMillan Education Ltd. ISBN: 0-333-25821-5.

Débarbouillé, M., I. Martin-Verstraete, A. Klier, & G. Rapoport (1991) The transcriptional regulator LevR of *Bacillus subtilis* has domains homologous to both Sigma 54- and phosphotransferase system-dependent regulators. *Proceedings of the National Academy of Sciences of the USA* **88**(6):2212-2216.

Debarbouille, M., M. Arnaud, A. Fouet, A. Klier, & G. Rapoport (1990) The *sacT* gene regulating the *sacPA* operon in *Bacillus subtilis* shares strong homology with transcriptional antiterminators. *Journal of Bacteriology* **172**:3966-3973.

Declerck, N., L. N. Minh, Y. Yang, V. Bloch, M. Kochoyan, & S. Aymerich (2002) RNA recognition of by transcriptional antiterminators of the *bglG/sacY* family: mapping of *sacY* RNA binding site. *Journal of Molecular Microbiology* **319**:1035-1048.

Demirbas, A. (2007) Progress and recent trends in biofuels. *Progress in Energy and Combustion Science*. **33**:1-18.

Desai, R. P., L. M. Harris, N. E. Welker, & E. T. Papoutsakis (1999) Metabolic flux analysis elucidates the importance of the acid-forming pathways in regulating solvent production by *Clostridium acetobutylicum*. *Metabolic Engineering* **1**:206-213.

Deutscher, J. & Jr. M. H. Saier (1983) ATP-dependent protein kinase-catalyzed phosphorylation of a seryl residue in HPr, a phosphate carrier protein of the phosphotransferase system in *Streptococcus pyogenes*. *Proceedings of the*

National Academy of Sciences of the United States of America **80**:6790-6794.

Deutscher, J., C. Francke, & P. W. Postma (2006) How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiology and Molecular Biology Reviews* **70**(4):939-1031.

Deutscher, J., E. Küster, U. Bergstedt, V. Charrier, & W. Hillen (1995) Protein kinase-dependent HPr/CcpA interaction links glycolytic activity to carbon catabolite repression in Gram-positive bacteria. *Molecular Microbiology* **15**(6):1049-1053.

Deutscher, J., U. Kessler, C. A. Alpert, & W. Hengstenberg (1984) Bacterial phosphoenolpyruvate-dependent phosphotransferase system: P-Ser-HPr and its possible regulatory function? *Biochemistry* **23**:4455-4460.

Dixon, B. (1996) The makers – microbes that shaped the world. In: Power Unseen: How Microbes Rule the World. W. H. Freeman & Company Ltd. ISBN: 0-7167-4550.

Donahue, J. L., J. L. Bownas, W. G. Niehaus, & T. J. Larson (2000) Purification and characterization of *glpX*-encoded fructose 1,6-bisphosphatase, a new enzyme of the glycerol 3-phosphate regulon of *Escherichia coli*. *Journal of Bacteriology* **182**(19):5624-5627.

Dong, X., P. Stothard, I. J. Forsythe, & D. S. Wishart (2004) PlasMapper: a web server for drawing and auto-annotating plasmid maps. *Nucleic Acids Research* **1**:(32)Web Server Issue:W660-4.

Dürre, P. (1998) New insight and novel developments in clostridial acetone/butanol/isopropanol fermentation. *Applied Microbiology & Biotechnology* **49**:639-648.

Dürre, P. (2005) Sporulation in clostridia (Genetics) In: Handbook on Clostridia (Editor, P. Dürre) CRC Press. ISBN: 0-8493-1618-9.

Dürre, P. (2007) Biobutanol: an attractive biofuel. *Biotechnology Journal* **2**:1525-1534.

Dürre, P. (2008) Fermentative butanol production: bulk chemical and biofuel. *Annals of the New York Academy of Sciences* **1125**:353-362.

Egeter, 0., & R. Bruckner (1996) Catabolite repression mediated by the catabolite control protein CcpA in *Staphylococcus xylosus*. *Molecular Microbiology* **21**:739-749.

Eiso A. B., G. K. Schuurman-Wolters, D. Nijlant, K. Dijkstra, M. H. Saier, G. T. Robillard, & R. M. Scheek (2001) NMR structure of cysteinyl-phosphorylated enzyme IIB of the N,N'-diacetylchitobiose-specific phosphoenolpyruvate-dependent phosphotransferase system of *Escherichia coli*. *Journal of Molecular Biology* **308**(5):993-1009.

Ennis, B. M., & I. S. Maddox (1985) Use of *Clostridium acetobutylicum* P262 for production of solvents from whey permeate. *Biotechnology Letters* **7**(8):610-606.

Ennis, B. M., & I. S. Maddox (1989) Production of solvents (ABE fermentation) from whey permeate by continuous fermentation in a membrane bioreactor. *Bioprocess and Biosystems Engineering* **4**(1):27-34.

Erni, B., B. Zanolari, & H. P. Kocher (1987) The mannose permease of *Escherichia coli* consists of three different proteins. *Journal of Biological Chemistry* **262**:(11)5236-5247.

European Commission (2005) Biomass Action Plan. Commission of the European Communities, Brussels. Accessed on 20.10.07 from the following URL: http://ec.europa.eu/energy/res/biomass_action_plan/doc/2005_12_07_comm_biomass_action_plan_en.pdf

Euzéby, J. P. (1997) List of bacterial names with standing in nomenclature: a folder available on the Internet. *International Journal of Systematic and Evolutionary Microbiology* **47**:590-592.

Euzéby, J. P. (2009) List of bacterial names with standing in nomenclature. [Last updated October 08 2009]. Accessed on 05.03.10 from the following URL: http://www.bacterio.cict.fr/c/clostridium.html

Ezeji T. C., N. Qureshi, & H. P. Blaschek (2004) Butanol fermentation research: upstream and downstream manipulations. *The Chemical Record* **4**:305-314.

Ezeji T. C., N. Qureshi, & H. P. Blaschek (2007a) Bioproduction of butanol from biomass: from genes to bioreactors. *Current Opinion in Biotechnology* **18**:220-227.

Ezeji, T. C., N. Qureshi, & H. P. Blaschek (2005) Process for continuous solvent production. *US patent* 2005/0089979.

Ezeji, T. C., N. Qureshi, & H. P. Blaschek (2007b) Butanol production from agricultural residues: Impact of degradation products on *Clostridium beijerinckii* growth and butanol fermentation. *Biotechnology and Bioengineering* **97**(6):1460-

1469.

Ezeji. T. C., & H. Blaschek (2008) Fermentation of dried distillers grains and solubles (DDGS) hydrolysates to solvents and value added products by solventogenic clostridia. *Bioresource Technology* **99**(12):5232-5242.

Falbe, J. (1970) Carbon monoxide in organic synthesis. Springer Verlag, Berlin-Heidelberg-New York. ISBN: 4-9998-2870.

Faure, D., J. Desair, V. Keijers, M. A. Bekri, P. Proost, B. Henrissat, & J. Vanderleyden (1999) Growth of *Azospirillum irakense* KBC1 on the aryl β-glucoside salicin requires either *salA* or *salB*. *Journal of Bacteriology* **181**(10):3003-3009.

Faure, D., M. H. Saier, & J. Vanderleyden (2001) An evolutionary alternative system for aryl beta-glucosides assimilation in bacteria. *Journal of Molecular Microbiology and Biotechnology* **3**(3):467-470.

Fernbach, A., & E. H. Strange (1911a) Improvements in the manufacture of products of fermentation. *Br Patent* 15,203.

Fernbach, A., & E. H. Strange (1911b) Improvements in the manufacture of higher alcohols. *Br Patent* 15,204.

Fernbach, A., & E. H. Strange (1912a) Fermentation process for the production of acetone and higher alcohols from starch, sugars and other carbohydrates. *US Patent* 772,120.

Fernbach, A., & E. H. Strange (1912b) Manufacture and production of a nutrient medium for the uses in fermentation processes. *US Patent* 724,990.

Fieulaine, S., S. Morera, S. Poncet, V. Monedero, V. Gueguen-Chaignon, A. Galinier, J. Janin, J. Deutscher, & S. Nessler (2001) X-ray structure of HPr kinase: A bacterial protein kinase with a P-loop nucleotide-binding domain. *The EMBO Journal* **20**:3917-3927.

Fond, O., J.-M. Engasser, G. Matta-El-Amouri, & H. Petitdemange (1986) The acetone butanol fermentation on glucose and xylose. I. Regulation and kinetics in batch cultures. *Biotechnology and Bioengineering* **28**(2):160-166.

Fond, O., J.-M. Engasser, G. Matta-El-Amouri, & H. Petitdemange (1986) The acetone butanol fermentation on glucose and xylose. II. Regulation and kinetics in fed-batch cultures. *Biotechnology and Bioengineering* **28**(2):167-175.

Forsberg, C. W. (1987) Production of 1, 3-propanediol from glycerol by *Clostridium*

acetobutylicum and other *Closrtidium* species. *Applied and Environmental Microbiology* **53**(4):639-643.

Freeman, S. (2002) Ecosystems. In: Biological Science (Editor, S. Freeman) Prentice Hall. ISBN: 0-1308-1923-9.

Fujita, Y., & E. Freese (1979) Purification and properties of fructose-1,6-bisphosphatase of *Bacillus subtilis*. *Journal of Biological Chemistry* **254**:5340-5349.

Fujita, Y., K. Yoshida, Y. Miwa, N. Yanai, E. Nagakawa, & Y. Kasahara (1998) Identification of the *Bacillus subtilis* fructose 1,6-bisphosphatase gene (*fbp*). *Journal of Bacteriology* **180**(16):4309-4313.

Fumoto, M., S. Miyazaki, & H. Sugawara (2002) Genome information broker (GIB): data retrieval and comparative analysis system for completed microbial genomes and more. *Nucleic Acids Research* **30**(1):66-68.

Galinier, A., J. Haiech M-C. Kilhoffer, M. Jaquinod, J. Stülke, J. Deutscher, & I. Martin-Verstraete (1997) The *Bacillus subtilis crh* gene encodes a HPr-like protein involved in carbon catabolite repression. *Proceedings of the National Academy of Sciences* USA **94**:8439-8444.

Gapes J. R. (2000) The economics of acetone-butanol fermentation: theoretical and market considerations. *Journal of Molecular Microbiology and Biotechnology* **1**:27-32.

Gapes, J. R., D. Nimcevic, & A. Friedl (1996) Long-term continuous cultivation of *Clostridium beijerinckii* in a two-stage chemostat with on-line solvent removal. *Applied and Environmental Microbiology* **62**:3210-3219.

Garrity, L. F., S. L. Schiel, R. Merrill, J. Reizer, M. H. Saier, & G. W. Ordal (1998) Unique regulation of carbohydrate chemotaxis in *Bacillus subtilis* by the phosphoenolpyruvate-dependent phosphotransferase system and the methyl-accepting chemotaxis protein McpC. *Journal of Bacteriology* **180**:4475-4480.

Gartner, D., M. Geissendorfer, & W. Hillen (1988) Expression of the *Bacillus subtilis xyl* operon is repressed at the level of transcription and is induced by xylose. *Journal of Bacteriology* **170**:3102-3109.

Geng, Q., & C. Park (1994) Pervaporative butanol fermentation by *Clostridium acetobutylicum* B18. *Biotechnology and Bioengineering* **43**(10):978-986.

George H. A., & J. S. Chen (1983) Acidic conditions are not obligatory for onset of

butanol formation by *Clostridium beijerinckii* (Synonym, *C. butylicum*). *Applied and Environmental Microbiology* **46**(2):321-327.

Gibbs, D. F. (1983) The rise and fall (...and rise?) of acetone/butanol fermentations. *Trends in Biotechnology* **1**(1):12-15.

Giel, M., M. Desnoyer, & J. Lopilato (1996) A mutation in a new gene, *bglJ* activates the *bgl* operon in *Escherichia coli* K-12. *Genetics* **143**:627-635.

Girbal, L., & P. Soucaille (1995) Regulation of solvent production in *Clostridium acetobutylicum*. *Trends in Biotechnology* **16**:11-16.

Girbal, L., C. Croux, I. Vasconcelos, & P. Soucaille (1995) Regulation of metabolic shifts in *Clostridium acetobutylicum* ATCC 824. *FEMS Microbiology Reviews* **17**:87-297.

Gonzalez-Pajuelo, M., I. Meynial-Salles, F. Mendes, J. C. Andrade, I. Vasconcelos, & P. Soucaille (2005) Metabolic engineering of *Clostridium acetobutylicum* for the industrial production of 1, 3-propanediol from glycerol. *Metabolic Engineering* **7**:329-336.

Gorke, B., & B. Rak (1999) Catabolite control of *Escherichia coli* regulatory protein BglG activity by antagonistically acting phosphorylations. *The EMBO Journal* **18**:3370-3379.

Gottwald, M., H. Hippe, & G. Gottschalk (1984) Formation of n-butanol from D-glucose by strains of the "*Clostridium tetanomorphum*" group. *Applied and Environmental Microbiology* **48**:573-576.

Gravesen, A., M. Ramnath, K. B. Rechinger, N. Andersen, L. Jansch, Y. Héchard, J. W. Hastings, & S. Knøchel (2002) High-level resistance to class IIa bacteriocins is associated with one general mechanism in *Listeria monocytogenes*. *Microbiology* **148**:2361-2369.

Gray, K. A., L. Zhao, & M. Emptage (2006) Bioethanol. *Current Opinion in Chemical Biology* **10**:141-146.

Groot, W. J., M. R. Kraayenbrink, R. H. Waldram, R. G. J. M. Vanderlans, & K. C. A. M. Luyben (1992) Integration of pervaporation and continuous butanol fermentation with immobilized cells: I. Experimental results. *Bioprocess Engineering* **8**:99-112.

Hall, B. G., & L. Xu (1992) Nucleotide sequence, function, activation, and evolution of the cryptic *acs* operon of *Escherichia coli* K12. *Molecular Biology and*

Evolution **9**(4):688-706.

Hang, Y. D., C. Y. Lee, E. E. Woodams, & H. J. Cooley (1981) Production of alcohol from apple pomace. *Applied and Environmental Microbiology* **42**(6):1128-1129.

Harris, L. M., N. E. Welker, & E. T. Papoutsakis (2002) Northern, morphological, and fermentation analysis of *spo0A* inactivation and overexpression in *Clostridium acetobutylicum* ATCC 824. *Journal of Bacteriology* **184**(13):3586-597.

Hassouni, M. el, B. Henrissat, M. Chippaux, & F. Barras (1992) Nucleotide sequences of the *arb* genes, which control β-glucoside utilization in *Erwinia chrysanthemi*: comparison with the *Escherichia coli bgl* operon and evidence for a new β-glycohydrolase family including enzymes from Eubacteria, Archeabacteria, and Humans. *Journal of Bacteriology* **174**(3):765-777.

Hastings, J. J. H (1971) Development of the fermentation industries in Great Britain. In: Advances in Applied Microbiology Volume 14. (Editor, D. Perlman) Academic Press Inc. ISBN: 0-120-02614-7.

Hazelwood, G. P., & H. J. Gilbert (1993) Xylose and cellulose utilisation by the clostridia. In: The Clostridia and Biotechnology (Editor, D. R. Woods) Butterworth-Heinemann. ISBN: 9-780-750690-041.

Heap, J. T., O. J. Pennington, S. T. Cartman, G. P. Carter, & N. P. Minton (2007) The ClosTron: a universal gene knock-out system for the Genus Clostridium. *Journal of Microbiological Methods* **70**:452-464.

Henstra, S. A., R. H. Duurkens, & G. T. Robillard (2000) Multiple phosphorylation events regulate the activity of the mannitol transcriptional regulator MtlR of the *Bacillus stearothermophilus* phosphoenolpyruvate-dependent mannitol phosphotransferase system. *Journal of Biological Chemistry* **275**:7037-7044.

Herzberg, O., P. Reddy, S. Sutrina, M. H. Saier, J. Reizer, & G. Kapadia (1992) Structure of the histidine-containing phosphocarrier protein HPr from *Bacillus subtilis* at 2.0-A resolution. *Proceedings of the National Academy of Sciences* **89**:2499-2503.

Heyndrickx, M., P. De Vos, M. Vancanneyt, & J. De Ley (1991) The fermentation of glycerol by *Clostridium butyricum* LMG $1212t_2$ and $1213t_1$ and *C. pasteurianum* LMG 3285. *Applied Microbiology Biotechnology* **34**:637-642.

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

Proceedings of the National Academy of Sciences **103**(30):11206-11210.

Holmberg C., L. Beijer, B. Rutberg, & L. Rutberg (1990) Glycerol catabolism in *Bacillus subtilis*: nucleotide sequence of the genes encoding glycerol kinase (*glpK*) and glycerol-3-phosphate dehydrogenase (*glpD*). *Journal of General Microbiology* **136**:2367-2375.

Hong, S. Y., C. L. An, K. M. Cho, S. M. Lee, Y. H. Kim, M. K. Kim, S. J. Cho, Y. P. Lim, H. Kim, & H. D. Yun (2006) Cloning and comparison of third beta-glucoside utilization (*bglEFIA*) operon with two operons of *Pectobacterium carotovorum* subsp. *carotovorum* LY34. *Bioscience, Biotechnology and Biochemistry* **70**(4):798-807.

Hoon, M. J. L. de, Y. Makita, K. Nakai, & S. Miyano (2005) Prediction of transcriptional terminators in *Bacillus subtilis* and related species. *Public Library of Science, Computational Biology* **1**(3):e25.

Horn, S. J., I. M. Aasen, & K. Ostgaard (2000) Ethanol production from seaweed extract. *Journal of Industrial Microbiology and Biotechnology* **25**:249-254.

Houman F., M. R. Diaz-Torres, & A. Wright (1990) Transcriptional antitermination in the *bgl* operon of *Escherichia coli* is modulated by a specific RNA binding protein. *Cell* **62**(6):1153-63.

Hueck C. J., & W. Hillen (1995) Catabolite repression in *Bacillus subtilis*: a global regulatory mechanism for the Gram-positive bacteria? *Molecular Microbiology* **15**(3):395-401.

Invitrogen (2006) Invitrogen TOPO TA® Cloning User Manual. Invitrogen. Accessed on 10.04.08 from the following URL: http://www.invitrogen.com/content/sfs/manuals/topota_man.pdf

Johnson, K. A., L. Chen, H. Yang, M. F. Roberts, & B. Stec (2001) Crystal structure and catalytic mechanism of the MJ0109 gene product: a bifunctional enzyme with inositol monophosphatase and fructose 1,6-bisphosphatase activities. *Biochemistry* **40**(3):618-630.

Jones D. T., M. Shirley, X. Wu, & S. Keis (2000) Bacteriophage infections in the industrial acetone butanol (AB) fermentation process. *Journal of Molecular Microbiology and Biotechnology* **2**(1):21-26.

Jones, D. T. (2001) Applied acetone-butanol fermentation. In: Clostridia – Biotechnology and Medical Applications (Editors, H. Bahl & P. Dürre) Wiley-

VCH, Weinheim. ISBN: 3-527-30175-5.

Jones, D. T., & D. R. Woods (1986) Acetone-butanol fermentation revisited. *Microbiological Reviews*. **50**(4):484-524.

Jones, D. T., & S. Keis (2005) Species and strain identification methods. In: Handbook on Clostridia (Editor, P. Dürre) CRC Press. ISBN: 0-8493-1618-9.

Jones, D. T., A. Westhuizen, S. Long,, E. R. Allcock, S. J. Reid, & D. R. Woods (1982) Solvent production and morphological changes in *Clostridium acetobutylicum*. *Applied and Environmental Microbiology* **43**:1434-1439

Jørgensen, H., J. B. Kristensen, & C. Felby (2007) Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. *Biofuels, Bioproducts and Biorefining* **1**(2):119-134.

Jules, M., L. Le Chat, S. Aymerich, & D. Le Coq (2009) The *Bacillus subtilis ywjI* (*glpX*) gene encodes a class II fructose-1, 6-bisphosphatase, functionally equivalent to the class III Fbp enzyme. *Journal of Bacteriology* **191**(9):3168-3171.

Kamm, B. & M. Kamm (2004) Principles of biorefineries. *Applied Microbiology and Biotechnology* **64**:137-145.

Kanehisa, M., & S. Goto (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28**(1):27-30.

Kasahara, Y., S. Nakai, & N. Ogasawara (1997) Sequence analysis of the 36-kb region between *gntZ* and *trnY* genes of *Bacillus Subtilis* genome. *DNA Research* **4**:155-159.

Kashket, E. R., & Z. Y. Cao (1993) Isolation of a degeneration-resistant mutant of *Clostridium acetobutylicum* NCIMB 8052. *Journal of Bacteriology* **59**:4198-4202.

Keis, S., J. T. Sullivan, & D. T. Jones (2001a) Physical and genetic map of the *Clostridium saccharobutylicum* (formerly *Clostridium acetobutylicum*) NCP 262 chromosome. *Microbiology* **147**:1909-1922.

Keis, S., R. Shaheen, & D. T. Jones (2001b) Emended descriptions of *Clostridium acetobutylicum* and *Clostridium beijerinckii*, and descriptions of *Clostridium saccharoperbutylacetonicum* sp. nov. and *Clostridium saccharobutylicum* sp. Nov. *International Journal of Systematic and Evolutionary Microbiology* **51**:2095-2103.

Kell, D. B., M. W. Peck, G. Rodger, & J. G. Morris (1981) On the permeability to weak acids and bases of the cytoplasmic membrane of *Clostridium pasteurianum*.

Biochemical and Biophysical Research Communications **99**:81-88.

Kessel, Van J. S., & J. B. Russell (1992) Energetics of arginine and lysine transport by whole cells and membrane vesicles of strain SR, a monensin-sensitive ruminal bacterium. *Applied and Environmental Microbiology* **58**:969-975.

Keyhani, N. O., & S. Roseman (1997) Wild-type *Escherichia coli* grows on the chitin disaccharide, N,N'-diacetylchitobiose, by expressing the *cel* operon. *Proceedings of the National Academy of Science USA* **94**(26):14367-14371.

Killeffer, D. H. (1926) Butanol and acetone from corn a description of the fermentation process. *Industrial and Engineering Chemistry* **19**(1):46-50.

Kim, B. C., Y. H. Lee, H. S. Lee, D. W. Lee, E. A. Choe, & Y. R. Pyun (2002) Cloning, expression and characterization of L-arabinose isomerase from *Thermotoga neapolitana*: bioconversion of D-galactose to D-tagatose using the enzyme. *FEMS Microbiology Letters* **212**(1):121-126.

Kirschn[illegible], M. (2006) Chemical profile: n-butanol. *Chemical Market Reporter* **269**:42-43.

Kotrba, P., M. Inui, & H. Yukawa (2001) Bacterial phosphotransferase system (PTS) in carbohydrate uptake and control of carbon metabolism. *Journal of Bioscience and Bioengineering* **92**(6):502-517.

Kotrba, P., M. Inui, & H. Yukawa (2003) A single V317A or V317M substitution in enzyme II of a newly identified β-glucoside phosphotransferase and utilization system of *Corynebacterium glutamicum* R extends its specificity towards cellobiose. *Microbiology* **149**:1569-1580.

Kricker, M., & B. G. Hall (1984) Directed evolution of cellobiose utilisation in *Escherichia coli* K12. *Molecular Biology and Evolution* **1**(2):171-182.

Krishna, H. S., K. Prasanthi, G. V. Chowdary, & C. Ayyanna (1998) Simultaneous saccharification and fermentation of pretreated sugar cane leaves to ethanol. *Process Biochemistry* **33**:825-830.

Krogh, A.. B. Larsson, G. von Heijne, & E. L. L. Sonnhammer (2001) Predicting transmembrane protein topology with a Hidden Markov Model: application to complete genomes. *Journal of Molecular Biology* **305**(3):567-580.

Krüger, S., & M. Hecker (1995) Regulation of the putative *bglPH* operon for aryl-b-glucoside utilization in *Bacillus subtilis*. *Journal of Bacteriology* **177**:5590-5597.

Kunst, F., N. Ogasawara, I. Moszer, & 147 others (1997) The complete genome

sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* **390**(6657):249-256.

Küster, E., E. J. Luesink, W. M. de Vos, & W. Hillen (1996) Immunological cross reactivity to the catabolite control protein CcpA from *Bacillus megaterium* is found in many Gram-positive bacteria. *FEMS Microbiological Letters* **139**:109-115.

Kyte, J. & R. Doolittle (1982) A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**(1):105-132.

Lai, X., & L. O. Ingram (1993) Cloning and sequencing of a cellobiose phosphotransferase system operon from *Bacillus stearothermophilus* XL-65-6 and functional expression in *Escherichia coli*. *Journal of Bacteriology* **175**:6441-6450.

Lai, X., F. C. Davis, R. B. Hespell, & L. O. Ingram (1997) Cloning of cellobiose phosphenolpyruvate-dependent phosphotransferase genes: functional expression in recombinant *Escherichia coli* and identification of a putative binding region for disaccharides. *Applied and Environmental Microbiology* **63**(2):355-363.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, & D. G. Higgins (2007) ClustalW and ClustalX Version 2. *Bioinformatics* **23**(21):2947-2948.

Lawlis, V. B., M. S. Dennis, E. Y. Chen, D. H. Smith, & D. J. Henner (1984) Cloning and sequencing of the xylose isomerase and xylulokinase genes of *Escherichia coli*. *Applied and Environmental Microbiology* **47**:15-21.

Le Coq, D., C. Lindner, S. Kruger, M. Steinmetz, & J. Stulke (1995) New β-glucoside (*bgl*) genes in *Bacillus subtilis*: the *bglP* gene product has both transport and regulatory functions similar to those of BglF, its *Escherichia coli* homolog. *Journal of Bacteriology* **177**:1530-1535.

Lee, S. Y., J. H. Park, S. H. Jang, L. K. Nielsen, J. Kim, & K. S. Jung (2008) Fermentative butanol production by clostridia. *Biotechnology and Bioengineering* **101**:209-228.

Lengeler J. W., F. Titgemeyer, A. P. Vogler, & B. M. Wöhrl (1990) Structures and homologies of carbohydrate: phosphotransferase system (PTS) proteins. *Philosophical Transactions of the Royal Society B: Biological Sciences* **326**(1236):489-504.

Lenz, T. & A. R. Morelra (1980) Economic evaluation of the acetone-butanol fermentation. *Industrial Engineering Chemistry Product Research Development* **19**:478-483.

Levina, D. B., L. Pitt, & M. Love (2004) Biohydrogen production: prospects and limitations to practical application. *International Journal of Hydrogen Energy* **29**:173-185.

Linares, D. M., J. Kok, & B. Poolman, Bert (2010) Genome sequences of *Lactococcus lactis* mg1363 (revised) and nz9000 and comparative physiological studies. *Journal of Bacteriology* **192**:5806-5812.

Lindner, C., A. Galinier, M. Hecker, & J. Deutscher (1999) Regulation of the activity of the *Bacillus subtilis* antiterminator LicT by multiple PEP-dependent, enzyme I- and HPr-catalysed phosphorylation. *Molecular Microbiology* **31**:995-1006.

Lokman, B. C., M. Heerikhuisen, R. J. Leer, A. van den Broek, Y. Borsboom, S. Chaillou, P. W. Postma, & P. H. Pouwels (1997) Regulation of expression of the *Lactobacillus pentosus xylAB* operon. *Journal of Bacteriology* **179**:5391-5397.

Lokman, B. C., P. Van Santen, J. C. Verdoes, J. Kruse, R. J. Leer, M. Posno, & P. H. Pouwels (1991) Organization and characterization of three genes involved in D-xylose catabolism in *Lactobacillus pentosus. Molecular and General Genetics* **230**:161-169.

Long, S., D. T. Jones, & D. R. Woods (1983) Sporulation of *Clostridium acetobutylicum* P262 in a defined medium. *Applied Environmental Microbiology* **45**(4):1389-1393.

Lortie, L., M. Pelletier, C. Vadeboncoeur, & M. Frenette (2000) The gene encoding IIABManL in *Streptococcus salivarius* is part of a tetracistronic operon encoding a phosphoenolpyruvate: mannose / glucose phosphotransferase system. *Microbiology* **146**:677-685.

Lynd, L. R., P. J. Weimer, W. Zyl, P. H. Willem, & S. Isak (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews* **66**:506-577.

Maddox, I. S., & A. E. Murray (1983) Production of n-butanol by fermentation of wood hydrolysate *Biotechnology Letters* **5**(3):175-178.

Maddox, I. S., N. Qureshi, & N. A. Gutierrez (1993) Utilization of whey by clostridia and process technology. In: The Clostridia and Biotechnology (Editor, D. R.

Woods) Butterworth-Heinemann. ISBN: 9-780-750690-041.

Maddox, S. (1980) Production of n-butanol from whey filtrate using *Clostridium acetobutylicum* NCIB 2951. *Biotechnology Letters* **2**:493-498.

Maloney, P. C. (2002) Bacterial membrane transport: superfamilies of transport proteins. In: Encyclopedia of Life Science (ELS). John Wiley & Sons, Ltd Chichester.

Manival, X., Y. Yang, M. P. Stub, M. Kochoyan, M. Steinmetz, & S. Aymerich (1997) From genetic to structural characterisation of a new class of RNA-binding domain within the SacY/BglG family of antiterminator proteins. *The European Molecular Biology Organization Journal* **16**:5019-5029.

Marasco, R., I. Salatiello, M. De Felice, & M. Sacco (2000) A physical and functional analysis of the newly-identified *bglGPT* operon of *Lactobacillus plantarum*. *FEMS Microbiology Letters* **186**:269-273.

Marasco, R., L. Muscariello, M. Varcamonti, M. De Felice, & M. Sacco (1998) Expression of the *bglH* gene of *Lactobacillus plantarum* is controlled by carbon catabolite repression. *Journal of Bacteriology* **180**(13):3400-3404.

Marchler-Bauer A., & S. H. Bryant (2004) CD-search: protein domain annotations on the fly. *Nucleic Acids Research* **32**(Web Server Issue):W327-W331.

Marchler-Bauer, A., J. B. Anderson, M. K. Derbyshire, C. DeWeese-Scott, N. R. Gonzales, M. Gwadz, L. Hao, S. He, D. I. Hurwitz, J. D. Jackson, Z. Ke, D. Krylov, C. J. Lanczycki, C. A. Liebert, C. Liu, F. Lu, S. Lu, G. H. Marchler, M. Mullokandov, J. S. Song, N. Thanki, R. A. Yamashita, J. J. Yin, D. Zhang, & S. H. Bryant (2007) CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Research* **35**(Database Issue):D237-D240.

Marchler-Bauer, A., J. B. Anderson, P. F. Cherukuri, C. DeWeese-Scott, L. Y. Geer, M. Gwadz, S. He, D. I. Hurwitz, J. D. Jackson, Z. Ke, C. J. Lanczycki, C. A. Liebert, C. Liu, F. Lu, G. H. Marchler, M. Mullokandov, B. A. Shoemaker, V. Simonyan, J. S. Song, P. A. Thiessen, R. A. Yamashita, J. J. Yin, D. Zhang, & S. H. Bryant (2005) CDD: a conserved domain database for protein classification. *Nucleic Acids Research* **33**(Database Issue):D192-D196.

Markham, N. R., & M. Zuker (2005) DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Research* **33**(Web Server Issue):W577-W581.

Martin-Verstraete, I., M. Debarbouille, A. Klier, & G. Rapoport (1994) Interactions

of wild-type and truncated LevR of *Bacillus subtilis* with the upstream activating sequence of the levanase operon. *Journal of Molecular Biology* **241**:178-192.

Martin-Verstraete, I., V. Charrier, J. Stulke, A. Galinier, B. Erni, G. Rapoport, & J. Deutscher (1998) Antagonistic effects of dual PTS-catalysed phosphorylation on the *Bacillus subtilis* transcriptional activator LevR. *Molecular Microbiology* **28**:293-303.

Mathews, D. H., J. Sabina, M. Zuker, & D. H. Turner (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology* **288**:911-940.

McCoy, E., E. B. Fred, W. H. Peterson, & E. G. Hastings (1926) A cultural study of the acetone butyl alcohol organism. *Journal of Infectious Diseases* **39**(6):457-483.

McDonald, C., G. E. Killgore, A. Thompson, R. C. Owens, S. V. Kazakova, S. P. Sambol, S. Johnson, & D. N. Gerding (2005) An epidemic, toxin gene–variant strain of *Clostridium difficile*. *New England Journal of Medicine* **353**(23):2433-2441.

Meaden, P. G., J. Aduse-Opoku, J. Reizerb, A. Reizerb, Y. A. Lanceman, M. F. Martin, & W. J. Mitchell (1994) The xylose isomerase-encoding gene (*xylA*) of *Clostridium thermosaccharolyticum*: cloning, sequencing and phylogeny of XylA Enzymes. *Gene* **141**:97-101.

Meadow, N. D., D. K. Fox, & S. Roseman (1990) The bacterial phosphoenol-pyruvate: glycose phosphotransferase system. *Annual Review of Biochemistry* **59**:497-542.

Menavuvu, B. T., W. Poonperm, K. Takeda, K. Morimoto, T. B. Granström, G. Takada, & K. Izumor (2006) Novel substrate specificity of d-arabinose isomerase from *Klebsiella pneumoniae* and its application to production of d-altrose from d psicose. *Journal of Bioscience and Bioengineering* **102**(5):436-441.

Minton, N. P., & J. Anné (2005) Clostridia as production systems for prokaryotic and eukaryotic proteins of therapeutic value in tumor treatment. In: Handbook on Clostridia (Editor, P. Dürre) CRC Press. ISBN: 0-8493-1618-9.

Minton, N. P., J. M. Brown, P. Lambin, & J. Anné (2001) Clostridia in cancer therapy. In: Clostridia – Biotechnology and Medical Applications (Editors, H. Bahl & P. Dürre) Wiley-VCH, Weinheim. ISBN: 3-527-30175-5.

Mirón, A. S., M. C. Cerón García, A. C. Gómez, F. G. Camacho, E. M. Grima, & Y. Chisti (2003) Shear stress tolerance and biochemical characterization of *Phaeodactylum tricornutum* in quasi steady-state continuous culture in outdoor photobioreactors. *Biochemical Engineering Journal* **16**(3):287-297.

Mitchell, W. J. (1996) Carbohydrate uptake and utilisation in *Clostridium beijerinckii* NCIMB 8052. *Anaerobe* **2**:379-384.

Mitchell, W. J. (1998) Physiology of carbohydrates to solvent conversion by clostridia. *Advances in Microbiology Physiology* **39**:31-130.

Mitchell, W. J., & I. R. Booth (1984) Characterisation of the *Clostridium pasteurianum* phosphotransferase system. *Journal of General Microbiology* **130**:2193-2200.

Mitchell, W. J., & M. Tangney (2005) Carbohydrate uptake by the phosphotransferase system and other mechanisms. In: Handbook on Clostridia (Editor, P. Dürre) CRC Press. ISBN: 0-8493-1618-9.

Mitchell, W. J., J. E. Shaw, & L. Andrews (1991) Properties of the glucose phosphotransferase system of *Clostridium acetobutylicum* NCIB 8052. *Applied and Environmental Microbiology* **57**(9):2534-2539.

Mitchell, W. J., K. A. Albasheri, & M. Yazdanian (1995) Factors affecting utilization of carbohydrates by clostridia. *FEMS Microbiology Reviews* **17**:317-329.

Mitchell, W. J., P. Tewatia, & P. G. Meaden (2007) Genomic analysis of the phosphotransferase system in *Clostridium botulinum. Journal of Molecular Microbiology and Biotechnology* **12**:33-42.

Miwa, Y., A. Nakata, A. Ogiwara, M. Yamamoto, & Y. Fujita (2000) Evaluation and characterisation of catabolite-responsive elements (*cre*) of *Bacillus subtilis. Nucleic Acids Research* **28**(5):1206-1210.

Monderer-Rothkoff, G., & O. Amster-Choder (2007) Genetic dissection of the divergent activities of the multifunctional membrane sensor BglF. *Journal of Bacteriology* **189**(23):8601-8615.

Monedero, V., M. J. Gosalbes, & G. Perez-Martinez (1997) Catabolite repression in *Lactobacillus casei* ATCC 393 is mediated by CcpA. *Journal of Bacteriology* **179**:6657-6664.

Montecucco, C., & J. Molgó (2005) Botulinal neurotoxins: revival of an old killer. *Current Opinion in Pharmacology* **5**(3):274-279.

Moreno, M. S., B. L. Schneider, R. R. Maile, W. Weyler, & M. H. Saier (2001) Catabolite repression mediated by the CcpA protein in *Bacillus subtilis*: novel modes of regulation revealed by whole-genome analyses. *Molecular Microbiology* **39**(5):1366-1381.

Mostafa, N. A. (2001) Production of acetic acid and glycerol from salted and dried whey in a membrane cell recycle bioreactor. *Energy Conservation and Management* **42**(9):1133-1142.

Mountfort, D. O., F. A. Rainey, J. Burghardt, H. F. Kaspar, & E. Stackebrandt (1997) *Clostridium vincentii* sp. nov., a new obligately anaerobic, saccharolytic, psychrophilic bacterium isolated from low-salinity pond sediment of the McMurdo Ice Shelf, Antarctica. *Archives of Microbiology* **167**(1):54-60.

Movahedzadeh, F., S. C. G. Rison, P. R. Wheeler, S. L. Kendall, T. J. Larson, & N. Stoker (2004) The *Mycobacterium tuberculosis* Rv1099c gene encodes a GlpX-like class II fructose 1,6-bisphosphatase. *Microbiology* **150**:3499-3505.

Mühle, S., I. Balsam, C. R. Cheeseman (2010) Comparison of carbon emissions associated with municipal solid waste management in Germany and the UK. *Resources, Conservation and Recycling* **54**:793-801.

Murphy, N., D. J. McConnell, & B. A. Cantwell (1984) The DNA sequence of the gene and genetic control sites for the excreted *B. subtilis* enzyme β-glucanase. *Nucleic Acids Research* **12**(13):5355-5367.

Nakhleh, E. A. (2004) Organization of Arab Petroleum Exporting Countries (OAPEC). In: Encyclopaedia of the Modern Middle East and North Africa (Editor, P. Mattar) Macmillan Library Reference. ISBN: 0028657691.

Neelakanta, G., S. Sankar, & K. Schnetz (2009) Characterization of a β-Glucoside operon (*bgc*) prevalent in septicemic and uropathogenic *Escherichia coli* strains. *Applied and Environmental Microbiology* **75**(8):2284-2293.

Nguyen T. X., M. R. Yen, R. D. Barabote, & M. H. Saier (2006) Topological predictions for integral membrane permeases of the phosphoenolpyruvate:sugar phosphotransferase system. *Journal of Molecular Microbiology and Biotechnology* **11**:345-360.

Nicholas, K. B., H. B. Jr. Nicholas, & D. W. Deerfield (1997) GeneDoc: analysis and visualization of genetic variation. EMBNEW.NEWS **4**:14.

Nichols, N. N., B. S. Dien & R. J. Bothast (2001) Use of catabolite repression

mutants for fermentation of sugar mixtures to ethanol. *Applied Microbiology and Biotechnology* **56**(1-2):120-125.

Nikaido, H., & M. H. Saier (1992) Transport proteins in bacteria: common themes in their design. *Science* **258**(5084):936-942.

Nobelmann, B., & J. W. Lengeler (1996) Molecular analysis of the *gat* genes from *Escherichia coli* and of their roles in galactitol transport and metabolism. *Journal of Bacteriology* **178**(23):6790-6795.

Nochur, S. V., G. R. Jacobson, M. F. Roberts, & A. L. Demain (1992) Mode of sugar phosphorylation in *Clostridium thermocellum*. *Applied Biochemistry & Biotechnology* **33**:33-41.

Nölling, J., G. Breton, M. V. Omelchenko, K. S. Makarova, Q. Zeng, R. Gibson, H. M. Lee, J. Dubois, D. Qiu, J. Hitti, Y. I. Wolf, R. L. Tatusov, F. Sabathe, L. Doucette-Stamm, P. Soucaille, M. J. Daly, G. N. Bennett, E. V. Koonin, & D. R. Smith (2001) Genome sequence and comparative analysis of the solvent producing bacterium *Clostridium acetobutylicum*. *Journal of Bacteriology* **183**(16):4823-4838.

Old, L. A., S. Lowes, & R. R. B. Russell (2006) Genomic variation in *Streptococcus mutans*: deletions affecting the multiple pathways of β-glucoside metabolism. *Oral Microbiology and Immunology* **21**(1):21-27.

Ounine, K., H. Petitdemange, G. Raval, & R. Gay (1985) Regulation and butanol inhibition of D-xylose and D-glucose uptake in *Clostridium acetobutylicum*. *Applied and Environmental Microbiology* **49**(4):874-878.

Page, R. D. M. (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**:357-358.

Paredes, C. J., I. Rigoutsos, & E. T. Papoutsakis (2004) Transcriptional organization of the *Clostridium acetobutylicum* genome. *Nucleic Acids Research* **32**(6):1973-1981.

Parker, L. L., & B. G. Hall (1988) A fourth *Escherichia coli* gene system with the potential to evolve β-glucoside utilization. *Genetics* **119**:485-490.

Parker, L. L., & B. G. Hall (1990) Characterization and nucleotide sequence of the cryptic *cel* operon of *Escherichia coli* K12. *Genetics* **124**(3):455-71.

Pasteur, L. (1862) Quelques résultats nouveaux relatifs aux fermentations acétique et butyrique. (Some new results concerning the fermentations acetic and butyric).

Bulletin de la Société chimique de Paris 52-53.

Patrick, J. W., & N. Lee (1968) Purification and properties of an L-arabinose isomerase from *Escherichia coli*. *The Journal of Biological Chemistry* **243**(16):4312-4318.

Platt, T. (1986) Transcription termination and the regulation of gene expression. *Annual Review of Biochemistry* 55:339-372.

Popoff, M. R., & B. G. Stiles (2005) Clostridial toxins vs. other bacterial toxins. In: Handbook on Clostridia (Editor, P. Dürre) CRC Press. ISBN: 0-8493-1618-9.

Postma, P. W., G. J. G. Ruijter, J. van der Wag, & K. van Dam (1992) Control of carbohydrate metabolism in enteric bacteria: qualitative and quantitative aspects. In: Molecular Mechanisms of Transport (Editors, E. Quagliarello & F. Palmieri) Elsevier Science Publishers. ISBN: 0-0444-89408-X

Po na, P. W., J. W. Lengeler, & G. R. Jacobson (1993) The phosphoenolpyruvate (PEP) dependent phosphotransferase system. *Microbiology Reviews* **57**(3):543-594.

Prabhu, P., M. K. Tiwari, M. Jeya, P. Gunasekaran, I. W. Kim & J. K. Lee (2008) Cloning and characterization of a novel L-arabinose isomerase from *Bacillus licheniformis*. *Applied Microbiology and Biotechnology* **81**(2):283-290.

Prasad, I., B. Young, & S. Schaefler (1973) Genetic determination of the constitutive biosynthesis of phospho-β-glucosidase A in *Escherichia coli* K-12. *Journal of Bacteriology* **114**:909-915.

Prinn, R. G., R. F. Weiss, P. J. Fraser, P. G. Simmonds, D. M. Cunnold, F. N. Alyea, S. O'Doherty, P. Salameh, B. R. Miller, J. Huang, R. H. J. Wang, D. E. Hartley, C. Harth, L. P. Steele, G. Sturrock, P. M. Midgley, & A. McCulloch (2000) A history of chemically and radiatively important gases in air deduced from ALE/GAGE/AGAGE. *Journal of Geophysical Research* **105**(D14):17751-17792.

Qureshi, N., I. S. Maddox, & A. Friedl (1992) Application of continuous substrate feeding to the ABE fermentation: relief of product inhibition using extraction, perstraction, stripping, and pervaporation. *Biotechnology Progress* **8**:382-390.

Rabus, R. J., L. Donald, D. J. Kelly, & Jr. M. H. Saier (1999) TRAP transporters: an ancient family of extracytoplasmic solute- receptor-dependent secondary active transporters. *Microbiology* **145**:3431-3445.

Raghunand, T. R., & S. Mahadevan (2003) The beta-glucoside genes of *Klebsiella aerogenes*: conservation and divergence in relation to the cryptic *bgl* genes of *Escherichia coli*. *FEMS Microbiology Letters* **223**:267-274.

Rainey, F. A., & E. Stackebrandt (1993) 16S rDNA analysis reveals phylogenetic diversity among the polysaccharolytic clostridia. *FEMS Microbiology Letters* **113**(2):125-128.

Rainey, F. A., B. J. Hollen, & A. Small (2009) Clostridium. In: Bergey's Manual of Systematic Bacteriology Volume III, The Firmicutes (Editors, P. De Vos, G. Garrity, D. Jones, N. R. Krieg, W. Ludwig, F. A. Rainey, K. Schleifer & W. B. Whitman). Williams and Wilkins. ISBN: 978-0-387-95041-9.

Ramey, D. (1998) Continuous two stage, dual path anaerobic fermentation of butanol and other organic solvents using two different strains of bacteria. *US Patent* 5753474.

Ramey, D. (2005) Butylfuel. Accessed on 14.11.09 from the following URL: http://www.butanol.com/

Rani, S. K., M. V. Swamy, & G. Seenayya (1998) Production of ethanol from various pure and natural cellulosic biomass by *Clostridium thermocellum* strains SS21 and SS22. *Process Biochemistry* **33**:435-440.

Rashid, N., H.Imanaka, T. Kanai, T. Fukui, H. Atomi, & T. Imanaka (2002) A novel candidate for the true fructose-1,6-bisphosphatase in archaea. *Journal of Biological Chemistry* **277**:30649-30655.

Ravagnani, A., K. C. B. Jennert, E. Steiner, R. Grünberg, J. R. Jefferies, S. R. Wilkinson, D. I. Young, E. C. Tidswell, D. P. Brown, P. Youngman, G. Morris, & M. Young (2000) Spo0A directly controls the switch from acid to solvent production in solvent-forming clostridia. *Molecular Microbiology* **37**(5):1172-1185.

Reid, S. J., M. S. Rafudeen, & N. G. Leat (1999) The genes controlling sucrose utilization in *Clostridium beijerinckii* NCIMB 8052 constitute an operon. *Microbiology* **145**:1461-1472.

Reinharz, J. (1985) Science in the service of politics: the case of Chaim Weizmann during the First World War. *The English Historical Review* **100**(396):572-603.

Reizer, J., & M. H. Saier (1997) Modular multidomain phosphoryl transfer proteins of bacteria. *Current Opinion in Structural Biology* **7**:407-415.

Reizer, J., S. Bachem, A. Reizer, M. Arnaud, M. Saier, & J. Stulke (1999) Novel phosphotransferase system genes revealed by genome analysis - the complete complement of PTS proteins encoded within the genome of *Bacillus subtilis*. *Microbiology* **145**:3419-3429.

Reizer, J., S. L. Sutrina, M. H. Saier, G. C. Stewart, A. Peterkofsky, & P. Reddy (1989) Mechanistic and physiological consequences of HPr(Ser) phosphorylation on the activities of the phosphoenolpyruvate phosphotransferase system in Gram-positive bacteria: studies with site-specific mutants of HPr. *The EMBO Journal* **8**:2111-2120.

Reizer, J., V. Michotey, A. Reizer, & M. H. Saier (1994) Novel phosphotransferase system genes revealed by bacterial genome analysis: unique, putative fructose- and glucoside-specific systems. *Protein Science* **3**(3):440-50.

Ren, C., Y. Gu, S. Hu, Y. Wu, P. Wang, Y. Yang, C. Yang, S. Yang, & W. Jiang (2010) Identification and inactivation of pleiotropic regulator CcpA to eliminate glucose repression of xylose utilization in *Clostridium acetobutylicum*. *Metabolic Engineering* **12**(5):446-454.

Reuse, H. D., & A. Danchin (1988) The *ptsH*, *ptsI*, and *crr* genes of the *Escherichia coli* phosphoenolpyruvate - dependent phosphotransferase system: a complex operon with several modes of transcription. *Journal of Bacteriology* **170**(9):3827-3837.

Revelle, R., & H. E. Suess (1957) Carbon dioxide exchange between atmosphere and ocean and the question of an increase of atmospheric CO_2 during past decades. *Tellus* **9**:18-27.

Rittmann, D., S. Schaffer, V. F. Wendisch, & H. Sahm (2003) Fructose-1,6-bisphosphatase from *Corynebacterium glutamicum*: expression and deletion of the *fbp* gene and biochemical characterization of the enzyme. *Archives of Microbiology* **180**(4):285-292.

Rodionov, D.A., A. A. Mironov, & M. S. Gelfand (2001) Transcriptional regulation of pentose utilisation systems in the *Bacillus/Clostridium* group of bacteria. *Federation of European Microbiological Societies: Microbiology Letters* **205**:305-314.

Rooijen, R. J., S. Schalkwijk, & W. M. Vos (1991) Molecular cloning, characterization, and nucleotide sequence of the tagatose 6-phosphate pathway

gene cluster of the lactose operon of *Lactococcus lactis*. *Journal of Biological Chemistry* **266**:7176-7181.

Rosey, E. L., & G. C. Stewart (1992) Nucleotide and deduced amino acid sequences of the *lacR*, *lacABCD*, and *lacFE* genes encoding the repressor, tagatose 6-phosphate gene cluster, and sugar-specific phosphotransferase system components of the lactose operon of *Streptococcus mutans*. *Journal of Bacteriology* **174**:6159-6170.

Rosey, E. L., B. Oskoulian, & G. C. Stewart (1991) Lactose metabolism by *Staphylococcus aureus*: characterization of *lacABCD*, the structural genes of the tagatose 6-phosphate pathway. *Journal of Bacteriology* **173**:5992-5998.

Roszak, D. B., & R. R. Colwell (1987) Survival strategies of bacteria in the natural environment. *Microbiological Reviews* **51**(3):365-379.

Rygus, T., A. Scheler, R. Allmansberger, & W. Hillen (1991) Molecular cloning, structure, promoters and regulatory elements for transcription of the *Bacillus megaterium* encoded regulon for xylose utilization. *Archives of Microbiology* **155**:535-542.

Saha, B. C. (2003) Hemicellulose bioconversion. *Journal of Industrial Microbiology and Biotechnology* **30**:279-291.

Saier, M. H. (1977) Bacterial phosphoenolpyruvate:sugar phosphotransferase systems: structural, functional, and evolutionary interrelationships. *Bacteriological Reviews* **41**(4):856-871.

Saier, M. H. (2000) Transporter Classification DataBase (TCDB). Accessed on 27.06.09 from the following URL: http://www.tcdb.org/

Saier, M. H., & J. Reizer (1992) Proposed uniform nomenclature for the proteins and protein domains of the bacterial phosphoenolpyruvate: sugar phosphotransferase system. *Journal of Bacteriology* **174**(5): 1433-1438.

Saier, M. H., & T. T. Tseng (1999) Evolutionary origins of transmembrane transport systems. In: Transport of molecules across microbial membranes. (Editors, J. K. Broome-Smith, S. Baumberg, C. J. Stirling & F. B Ward) University Press Cambridge. ISBN: 0-521-77270-2.

Saier, M. H., C. V. Tran, & R. D. Barabote (2006) TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Research* **34**(Database Issue):D181-D186.

Saier, M. H., S. Chauvaux, J. Deutscher, J. Reizer, & J. J. Ye (1995) Protein phosphorylation and regulation of carbon metabolism in Gram-negative versus Gram-positive bacteria. *Trends in Biochemical Sciences* **20**(7):267-271.

Sanchez S., & K. L. Smiley (1975) Properties of D-xylose isomerase from *Streptomyces albus*. *Applied Microbiology* **29**(6):745-750.

Sandhu, D. K., & V. K. Joshi (1997) Solid-state fermentation of apple pomace for concomitant production of ethanol and animal feed. *Journal of Scientific & Industrial Research* **56**(2):86-90.

Sa-Nogueira, I., & H. De Lencastre, (1989) Cloning and characterisation of *araA*, *araB*, and *araD*, the structural genes for L-arabinose utilisation in *Bacillus subtilis*. *Journal of Bacteriology* **171**(7):4088-4091.

Sa-Nogueira, I., & S. S. Ramos (1997) Cloning, functional analysis, and transcriptional regulation of the *Bacillus subtilis araE* gene involved in L-arabinose utilisation. *Journal of Bacteriology* **179**:7705-7711.

Sa-Nogueira, I., T.V., Nogueira, S. Soares, & H. de Lencastre (1997) The *Bacillus subtilis* L-arabinose (*ara*) operon: nucleotide sequence, genetic organization and expression. *Microbiology* **143**:957-969.

Schaefler, S., A. Malamy, & I. Green (1969) Phospho-β-glucosidases and β-glucoside permeases in *Streptococcus, Bacillus*, and *Staphylococcus*. *Journal of Bacteriology* **99**(2):434-440.

Schaffer, S., K. Thormann, S. Hujer, & P. Dürre (2005) Quantitative proteome analysis of Clostridia. In: Handbook on Clostridia (Editor, P. Dürre) CRC Press. ISBN: 0-8493-1618-9.

Schardinger, F. (1905) *Bacillus macerans,* ein Aceton-bildender Rottebacillus. *Zentralbl Bakteriol Parasitenkd Infektionskr Hyg* **14**:772-781.

Scharlemann, J. P. W., & W. F. Laurance (2008) How green are biofuels? *Science* **319**(4):43-44.

Scheler, A., T. Rygus, R. Allmansberger, & W. Hillen (1991) Molecular cloning, structure, promoters and regulatory elements for transcription of the *Bacillus licheniformis*-encoded regulon for xylose utilization. *Archives of Microbiology* **155**:526-534.

Schenk, P. M., S. R. Thomas-Hall, E. Stephens, U. C. Marx, J. H. Mussgnug, C. Posten, O. Kruse, & B. Hankamer (2008) Second generation biofuels: high-

efficiency microalgae for biodiesel production. *Bioenergy Research* **1**:20-43.

Schlösser, A., J. Jantos, K. Hackmann, & H. Schrempf (1999) Characterization of the binding protein-dependent cellobiose and cellotriose transport system of the cellulose degrader *Streptomyces reticuli*. *Applied and Environmental Microbiology* **65**(6):2636–2643.

Schmiedel, D., M. Kintrup, E. Küster, & W. Hillen (1997) Regulation of expression, genetic organization and substrate specificity of xylose uptake in *Bacillus megaterium*. *Molecular Microbiology* **23**(5):1053-1062.

Schnetz, K., & B. Rak (1988) Regulation of the *bgl* operon of *Escherichia coli* by transcriptional antitermination. *The EMBO Journal* **7**(10):3271-3277.

Schnetz, K., C. Toloczyki, & B. Rak (1987) Beta-glucoside (*bgl*) operon of *Escherichia coli* K12 - nucleotide sequence, genetic organization, and possible evolutionary relationship to regulatory components of two *Bacillus subtilis* genes. *Journal of Bacteriology* **169**:2579-2590.

Schnetz, K., J. Stülke, S. Gertz, S. Krüger, M. Krieg, M. Hecker, & B. Rak (1996) LicT, a *Bacillus subtilis* transcriptional antiterminator protein of the BglG family. *Journal of Bacteriology* **178**:1971-1979.

Schuster, K. C., R. van den Heuvel, N. A. Gutierrez, & I. S. Maddox (1998) Development of markers for product formation and cell cycle in batch cultivation of *Clostridium acetobutylicum* ATCC 824. *Applied Microbiology and Biotechnology* **49**:669-676.

Sedivy, J. M., F. Daldal, & D. G. Fraenkel (1984) Fructose bisphosphatase of *Escherichia coli*: cloning of the structural gene (*fbp*) and preparation of a chromosomal deletion. *Journal of Bacteriology* **158**:1048-1053.

Senger, R. S., & E. T. Papoutsakis (2008) Genome-Scale model for *Clostridium acetobutylicum*: Part I. Metabolic network resolution and analysis. *Biotechnology and Bioengineering* **101**(5):1036-1052.

Setlow, B., A. Cabrera-Hernandez, R. M. Cabrera-Martinez, & P. Setlow (2004) Identification of aryl-phospho-beta-D-glucosidases in *Bacillus subtilis*. *Archives of Microbiology* **181**(1):60-67.

Shaheen, R., M. Shirley, & D. T. Jones (2000) Comparative fermentation studies of industrial strains belonging to four species of solvent-producing clostridia. *Journal of Molecular Microbiology and Biotechnology* **2**(1):115-124.

Sheehan, J., T. Dunahay, J. Benemann, & P. Roessler (1998) A look back at the U.S. Department of Energy's aquatic species program: biodiesel from algae. *National Renewable Energy Laboratory*. US Department of Energy's Office of Fuel Development. Accessed on 20.10.07 from the following URL: http://www.nrel.gov/docs/legosti/fy98/24190.pdf

Shimizu, T., K. Ohtani, H. Hirakawa, K. Ohshima, A. Yamashita, T. Shiba, N. Ogasawara, M. Hattori, S. Kuhara, & H. Hayashi (2002) Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *Proceedings of the National Academy of Sciences of the United States of America* **99**(2):996-1001.

Simpson, C. L., & R. R. B. Russell (1998) Identification of a homolog of CcpA catabolite repressor protein in *Streptococcus mutans*. *Infection Immunology* **66**:2085-2092.

Sims, R. E. H., A. Hastings, B. Schlamadinger, G. Taylor, & P. Smith (2006) Energy crops: current status and future prospects. *Global Change Biology* **12**:2054-2076.

Sliz, P., R. Engelmann, W. Hengstenberg, & E. F. Pai (1997) The structure of enzyme IIAlactose from *Lactococcus lactis* reveals a new fold and points to possible interactions of a multicomponent system. *Structure* **5**(6):775-788.

Smith, L. C., A. E. El Obeid, & H. H. Jensen (2000) The geography and causes of food insecurity in developing countries. *Agricultural Economics* **22**(2):199-215.

Solomons, G. L. (1976) Solvents from carbohydrates: some economic considerations. *Process Biochemistry* **11**:32-37.

Sonenshein, A. L. (2000) Control of sporulation initiation in *Bacillus subtilis*. *Current Opinion in Microbiology* **3**(6):561-566.

Sonnhammer, E. L. L., G. von Heijne, & A. Krogh (1998) A Hidden Markov Model for predicting transmembrane helices in protein sequences. In: Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology (Editors, J. Glasgow, T. Littlejohn, F. Major, R. Lathrop, D. Sankoff, & C. Sensen) AAAI Press. ISBN: 1-57735-053-7.

Speelmans, G., B. Poolman, & W. N. Konings (1993) Amino acid transport in the thermophilic anaerobe *Clostridium fervidus* is driven by an electrochemical sodium gradient. *Journal of Bacteriology* **175**(7): 2060-2066.

Stackebrandt, E., I. Kramer, J. Swiderski, & H. Hippe (1999) Phylogenetic basis for a taxonomic dissection of the genus *Clostridium*. *FEMS Immunology and Medical*

Microbiology **24**:253-258.

Stülke, J., & W. Hillen (1999) Carbon catabolite repression in bacteria. *Current Opinion in Microbiology* **2**(2):195-201.

Stülke, J., & W. Hillen (2000) Regulation of carbon catabolism in *Bacillus subtilis*. *Annual Review of Microbiology* **54**:849-880.

Stülke, J., M. Arnaud, G. Rapoport, & I. Martin-Verstraete (1998) PRD - a protein domain involved in PTS-dependent induction and carbon catabolite repression of catabolic operons in bacteria. *Molecular Microbiology* **28**:865-874.

Sturgeon, R. J. (2002) Carbohydrates. *Encyclopedia of Life Sciences*. 1-12.

Sullivan, L., & G. N. Bennet (2006) Proteome analysis and comparison of *Clostridium acetobutylicum* ATCC 824 and Spo0A strain variants. *Journal of Industrial Microbiology and Biotechnology* **33**(4):298

Sun, T., & J. Altenbuchner (2010) Characterization of a mannose utilization system in *Bacillus subtilis*. *Journal of Bacteriology* **192**(8):2128-2139.

Sutrina S. L., T. McGeary, & C. A. Bourne (2007) The phosphoenolpyruvate:sugar phosphotransferase system and biofilms in Gram-positive bacteria. *Journal of Molecular Microbiology and Biotechnology* **12**:269-272.

Tamoi, M., T. Ishikawa, T. Takeda, & S. Shigeoka (1996) Molecular characterization and resistance to hydrogen peroxide of two fructose-1,6-bisphosphatases from *Synechococcus* PCC 7942. *Archives of Biochemistry and Biophysics* **334**:27-36.

Tang, Y., M. An, K. Liu, S. Nagai, T. Shigematsu, S. Morimura, & K. Kida (2006) Ethanol production from acid hydrolysate of wood biomass using the flocculating yeast *Saccharomyces cerevisiae* strain KF-7. *Process Biochemistry* **41**(4):909-914.

Tangney M., A. Galinier J. Deutscher, & W. J. Mitchell (2003) Analysis of the elements of catabolite repression in *Clostridium acetobutylicum* ATCC 824. *Journal of Molecular Microbiology Biotechnology* **6**:6-11.

Tangney, M., & W. J. Mitchell (2000) Analysis of a catabolic operon for sucrose transport and metabolism in *Clostridium acetobutylicum* ATCC 824. *Journal of Industrial Microbiology & Biotechnology* **2**(1):71-80.

Tangney, M., & W. J. Mitchell (2004) *Clostridium tetani* encodes a phosphocarrier protein, HPr. *Microbiology* **150**:525-526.

Tangney, M., & W. J. Mitchell (2005) Regulation of catabolic gene systems. In: Handbook on Clostridia (Editor, P. Dürre) CRC Press. ISBN: 0-8493-1618-9.

Tangney, M., & W. J. Mitchell (2007) Characterisation of a glucose phosphotransferase system in *Clostridium acetobutylicum* ATCC 824. *Applied Genetics and Molecular Biotechnology* **74**:398-405.

Tangney, M., A. Galinier, J. Deutscher, & W. J. Mitchell (2003) Analysis of the elements of catabolite repression in *Clostridium acetobutylicum* ATCC 824. *Journal of Molecular Microbiology Biotechnology* **6**:6-11.

Tangney, M., G. T. Winters, & W. J. Mitchell (2001) Characterisation of a maltose transport system in *Clostridium acetobutylicum* ATCC 824. *Journal of Industrial Microbiology and Biotechnology* **27**:298-306.

Thompson, J., S. B. Ruvinov, D. I. Freedberg, & B. G. Hall (1999) Cellobiose-6-phosphate hydrolase (CelF) of *Escherichia coli*: characterization and assignment to the unusual family 4 of glycosylhydrolases. *Journal of Bacteriology* **181**(23):7339-7345.

Thompson, J., S. Hess, & A. Pikis (2004) Genes *malh* and *pagl* of *Clostridium acetobutylicum* ATCC 824 encode NAD+- and Mn2+-dependent phospho-α glucosidase(s). *Journal of Biological Chemistry* **279**:1553-1561.

Thomson, J. C., & B. B. He (2006) Characterisation of crude glycerol from biodiesel production from multiple feedstocks. *Applied Engineering in Agriculture* **22**(2):261-265.

Tilbeurgh, H., & N. Declerck (2001) Structural insights into the regulation of bacterial signalling proteins containing PRDs. *Current Opinion in Structural Biology* **11**:685-693.

Tobisch, S., J. Stulke, & M. Hecker (1999) Regulation of the *lic* operon of *Bacillus subtilis* and characterization of potential phosphorylation sites of the LicR regulator protein by site-directed mutagenesis. *Journal of Bacteriology* **181**:4995-5003.

Tobisch, S., P. Glaser, S. Kruger, & M. Hecker (1997) Identification and characterization of a new β-glucoside utilization system in *Bacillus subtilis*. *Journal of Bacteriology* **179**:496-506.

Tomas, C. A., J. Beamish, & E. T. Papoutsakis (2004) Transcriptional analysis of butanol stress and tolerance in *Clostridium acetobutylicum*. *Journal of*

Bacteriology **186**:2006-2018.

Tortosa P., N. Declerck, H. Dutartre, C. Lindner, J. Deutscher, & D. Le Coq (2001) Sites of positive and negative regulation in the *Bacillus subtilis* antiterminators LicT and SacY. *Molecular Microbiology* **41**:1381-1393.

Tortosa, P., S. Aymerich, C. Lindner, M. H. Saier, Jr., J. Reizer, & D. Le Coq (1997) Multiple phosphorylation of SacY, a *Bacillus subtilis* transcriptional antiterminator negatively controlled by the phosphotransferase system. *Journal of Biological Chemistry* **272**(27):17230-17237.

Tummala, S. B., S. G. Junne, C. J. Paredes, & E. T. Papoutsakis (2003) Transcriptional analysis of product-concentration driven changes in cellular programs of recombinant *Clostridium acetobutylicum* strains. *Biotechnology and Bioengineering* **84**(7):842-854.

Upshall, M. (1994) The Hutchinson Encyclopedia, Tenth Edition. Hutchinson Reference. ISBN: 0-0917-5393-7.

Vadeboncoeur, C., M. Frenette, & L.-A. Lortie (2000) Regulation of the *pts* operon in low G+C Gram-positive bacteria. *Journal of Molecular Microbiology Biotechnology* **2**(4):483-490.

Vasconcelos, I., L. Girbal, & P. Soucaille (1994) Regulation of carbon and electron flow in *Clostridium acetobutylicum* grown in chemostat culture at neutral pH on mixtures of glucose and glycerol. *Journal of Bacteriology* **176**(3):1443-1450.

Verhees, C. H., J. Akerboom, E. Schiltz, W. M. de Vos, & J. van der Oost (2002) Molecular and biochemical characterization of a distinct type of fructose-1,6-bisphosphatase from *Pyrococcus furious*. *Journal of Bacteriology* **184**:3401-3405.

Voget, C. E., C. F. Mignone, & R. J. Ertola (1985) Butanol production from apple pomace. *Biotechonology Letters* **7**(1):43-46.

Wackett, L. P. (2008) Microbial-based motor fuels: science and technology. *Microbial Biotechnology* **1**(3):11-25.

Walker, J. E., M. Saraste, M. J. Runswick, & N. J. Gay (1982) Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO Journal* **1**:945-951.

Warny, M., J. Pepin, A. Fang, A, G. Killgore, A. Thompson, J. Brazier, E. Frost, & L.

C. McDonald (2005) Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. *Lancet* **366**(9491):1079-1084.

Watson, R. T., M. C. Zinyowera, & R. H. Moss (1996) Climate Change 1995. Impacts, Adaptation and Mitigation of Climate Change: Scientific-technical analyses. Contribution of Working Group II to the Second Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge. ISBN: 0-5215-6437-9.

Weber, B. A., J. R. Klein, & B. Henrich (1998) The *arbZ* gene from *Lactobacillus delbrueckii* subsp. *lactis* confers to *Escherichia coli* the ability to utilize the β-glucoside arbutin. *Gene* **212**:203-211.

Wehtje C., L. Beijer, R. P. Nilsson, & B. Rutberg (1995) Mutations in the glycerol kinase gene restore the ability of a *ptsGHI* mutant of *Bacillus subtilis* to grow on glycerol. *Microbiology* **141**:1193-1198.

Weickert, M. J., & S. Adhya (1992) A family of bacterial regulators homologous to Gal and Lac repressors. *The Journal of Biological Chemistry* **267**(22):15869-15874.

Weizmann, C. (1915) Improvements in the bacterial fermentation of carbohydrates and in bacterial cultures for the same. *Br Patent* 4,845.

White, J., B. K. Yohannan, & G. M. Walker (2008) Bioconversion of brewer's spent grains to bioethanol. *FEMS Yeast Research* **8**(7):1175-1184.

Wiegel, J. (2009) Clostridiaceae. In: Bergey's Manual of Systematic Bacteriology Volume III, The Firmicutes (Editors, P. De Vos, G. Garrity, D. Jones, N. R. Krieg, W. Ludwig, F. A. Rainey, K. Schleifer and W. B. Whitman) Williams and Wilkins, Baltimore. ISBN: 0-683-07893-3.

Wiegel, J., L. G. Ljungdahl, & J. R. Rawson (1979) Isolation from soil and properties of the extreme thermophile *Clostridium thermohydrosulfuricum*. *Journal of Bacteriology* **139**(3):800-810.

Wiegel, J., R. Tanner, & F. A. Rainey (2005). An introduction to the Family Clostridiaceae. In: The Prokaryotes: a Handbook on the Biology of Bacteria Volume 4,: Bacteria: Firmicutes, Cyanobacteria, 3rd edition. release 3.20. New York: Springer. ISBN: 0-3872-5492-7.

Wilhelm, M., & C. P. Hollenberg (1985) Nucleotide sequence of the *Bacillus subtilis*

xylose isomerase gene: extensive homology between the *Bacillus* and *Escherichia coli* enzyme. *Nucleic Acids Research* **13**(15):5717-5722.

Wooley, R., M. Ruth, D. Glassner, & J. Sheehan (1999) Process design and costing of bioethanol technology: a tool for determining the status and direction of research and development. *Biotechnology Progress* **15**:797-803.

Wrolstad, R. E., D. A. Heatherbell, G. A. Spanos, R. W. Durst, J. Hsu, & B. M. Yorgey (1989) Processing and storage influences on the chemical composition and quality of apple, pear and grape juice concentrates (Editor, J. J. Jen) In: Quality Factors of Fruit and Vegetables. American Chemical Society. ISBN: 9-78084-12166-31.

Yang, M., S-M. Luoh, A. Goddard, D. Reilly, W. Henzel, & S. Bass (1996) The *bglX* gene located at 47.8 min on the *Escherichia coli* chromosome encodes a periplasmic β-glucosidase. *Microbiology* **142**:1659-1665.

Yang, Y., N. Declerck, X. Manival, S. Aymerich, & M. Kochoyan (2002) Solution structure of the LicT–RNA antitermination complex: CAT clamping RAT. *The EMBO Journal* **21**(8):1987-1997.

Yanofsky, C. (2000) Transcription attenuation: once viewed as a novel regulatory strategy. *Journal of Bacteriology* **182**(1):1-8.

Yazdani, S. S., & R. Gonalez (2007) Anaerobic fermentation of glycerol: a path to economic viability for the biofuels industry. *Current Opinion in Biotechnology* **18**:213-219.

Ye, R., S. N. Rehemtulla, & S-L. Wong (1994) Glucitol induction in *Bacillus subtilis* is mediated by a regulatory factor, GutR. *Journal of Bacteriology* **176**(11):3321-3327.

Yoshida, K., K. Shindo, H. Sano, S. Seki, M. Fujimura, N. Yanai, Y. Miwa, & Y. Fujita (1996) Sequencing of a 65 kb region of the *Bacillus subtilis* genome containing the *lic* and *cel* loci, and creation of a 177 kb contig covering the *gnt-sacXY* region. *Microbiology* **142**:3113-3123.

Young, M., N. P. Minton, & W. L. Staudenbauer (1989) Recent advanced in the genetics of clostridia. *FEMS Microbiology Reviews* **63**:301-326.

Yu, Y., M. Tangney, H. C. Aass, & W. J. Mitchell (2007) Analysis of the mechanism and regulation of lactose transport and metabolism in *Clostridium acetobutylicum* ATCC 824. *Applied and Environmental Microbiology* **73**(6):1842-1850.

Zaldivar, J., J. Nielsen, & L. Olsson (2001) Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration. *Applied Microbiology and Biotechnology* **56**:17-34.

Zhang, J., & A. Aronson (1994) A *Bacillus subtilis bglA* gene encoding phospho-β-glucosidase is inducible and closely linked to a NADH dehydrogenase-encoding gene. *Gene* **140**(1):85-90.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**(13):3406-3415.

Zukowski, M. M., L. Miller, P. Cosgwell, K. Chen, S. Aymerich, & M. Steinmetz (1990) Nucleotide sequence of the *sacS* locus of *Bacillus subtilis* reveals the presence of two regulatory genes. *Gene* **90**:153-155.

Zverlov, V. V., O. Berezina, G. A. Velikodvorskaya, & W. H. Schwarz (2006) Bacterial acetone and butanol production by industrial fermentation in the Soviet Union: use of hydrolyzed agricultural waste for biorefinery. *Applied Microbiology and Biotechnology* **71**:587-597.