Pentose Sugar Utilisation in *Clostridium beijerinckii* NCIMB 8052 for Biobutanol Production: Genetic and Physiological Studies

Julie Elizabeth Watson

Doctor of Philosophy

2012

Pentose Sugar Utilisation in *Clostridium beijerinckii* NCIMB 8052 for Biobutanol Production: Genetic and Physiological Studies

Julie Elizabeth Watson

A thesis submitted in partial fulfilment of the requirements of Edinburgh Napier University, for award of Doctor of Philosophy

May 2012

Abstract

The future of biofuel production hinges on a cheap, readily available feedstock. In terms of resources available, lignocellulose is the most abundant renewable resource on the planet, available from a plethora of sources such as agriculture, forestry, industry and municipals, therefore presenting an attractive resource. Cellulose, hemicellulose and lignin are the three main constituents of lignocellulose. The viability of such a feedstock requires as much of these constituents being converted to product as possible and therefore requires the suitable candidate organism to achieve this. Hemicellulose, an often pentose-rich portion of lignocellulose, can constitute as much as 35%. Traditionally yeasts, namely *Saccharomyces cerevisiae*, have been employed in biofuel production. However, yeasts are often unable to utilise pentose sugars found in the hemicellulose fraction.

In this study the ability of *Clostridium beijerinckii* to utilise the pentose sugars xylose and arabinose was investigated. The ability of *C. beijerinckii* to utilise these pentose sugars was established, total solvent yields were 0.34 gram/gram (g.g) of arabinose or xylose consumed, in comparison to 0.41 g.g of glucose consumed. The presence of glucose on low mixed-sugar concentrations (1%), but not on high (6%) hindered the use of both pentoses. The ability of *C. beijerinckii* to utilise pentose sugars in xylan. *C. beijerinckii* was capable of fermenting xylan and smaller hydrolysate units of xylan, however the solvent yield was poor. The ability of *C. beijerinckii* to utilise a pentose-rich waste stream, spent dried distillers' grains (DDGS) was also investigated. Simultaneous saccharification and fermentation of DDGS produced a total solvent yield of 0.25 g.g in comparison to 0.34 g.g on media containing the equivalent soluble sugars.

The ability of *C. beijerinckii* to utilise both pentose sugars led to *in-silico* studies to identify gene systems involved and implicated several genes organised in two distinct clusters, one for each pentose sugar, within the genome. Proteomic analyses by liquid chromatography electrospray ionisation tandem mass spectrometry of whole-cell proteins harvested from cultures of *C. beijerinckii* grown on either xylose or arabinose as the sole carbon source provided further evidence to implicate the gene systems identified by the *in silico* analyses. Genes were then targeted for characterisation *in-vivo* using a number of *E. coli* knock-out mutants, lacking one or more of key genes involved in pentose sugar use. This identified arabinose isomerase, xylose isomerase and xylulokinase genes.

Acknowledgements

"*What are you waiting for nobody's gonna show you how. Why wait for someone else to do what you can do right now. Got no limits and no boundaries...*" – Madonna 'Give it 2 me'. Inspired me throughout my research to independently and confidently venture down new avenues and advance.

Firstly, Thank you to my funders Green Biologics LTD and BBSRC without their funding this project would not have been possible.

Thanks to Martin and Fiona for believing in me and giving me this opportunity four years ago, the BfRC group 'meetings' at Bennets and the encouragement and support given to me. A special thanks to Dr Wilf Mitchell (Heriot-Watt University) for his selfless support and feedback, of which I am in debt to. Thanks also to Dr Douglas Fraser-Pitt (Edinburgh Napier University), Dr Neil Inglis (The Moredun Institute), Dr Edward Green, Dr Matthew Crow, Dr James Nicholl and Dr Prebben Krabben (Green Biologics LTD) for their support and expertise during my research.

Thank you to the technical staff, especially Sandra Dunbar for expertise and keeping the lab equipped as purple as possible. Dr Jane White, Dr Sandra Messenger and Dr David Messenger for their expertise and advice they shared with me so selflessly which made a real difference to my work. A special thanks to Eve and Gillian for the fun we shared amongst work, especially at conferences.

My appreciation also extends to Jen, Jenny, Amy, Fiona, Matt, Heli, Majid, Samson and the rest of the research students past and present all of whom made research life so endearing. To my mam and dad whose belief and support allowed me to pursue science and get where I am today, I cannot thank you enough. Thank you to Bill and Margaret Wallace whose very much appreciated support and encouragement also got me here. Finally, to Phil 'Hawkman' for his love and support in the twilight of my PhD.

Veni, *vidi*, *vici* - Caesar

In loving memory of my Grandmothers Gwladys & Elsie, their memory has pushed me to aspire & to sustain diligence

Rest in Peace

Contents

List of Figures

List of Tables

Abbreviations

AA	Amino Acid
AB	Acetone Butanol
ABC	ATP Binding Cassette
ABE	Acetone Butanol Ethanol
ADP	Adenosine diphosphate
AraA	L-Arabinose isomerase
AraB	L-Ribulokinase
AraC	Arabinose transcription regulator
AraD	L-ribulose 5-phosphate 4-epimerase
AraF	Arabinose binding protein
AraG	Arabinose ATPase
AraH	Arabinose permease
AraR	Arabinose repressor
ATP	Adenosine Triphosphate
ATCC	American Type Culture Collection
BCA	Binchinonic Acid
BP	British Petroleum
BLAST	Basic Local Alignment Search Tool
bp	Base Pair
BSA	Bovine Serum Albumin
C	Carbon
°C	Celsius
CBEI	Clostridium beijerinckii
CBM	Clostridium Basal Medium
CcpA	Catabolite Control Protein A
CCR	Carbon Catabolite Repressor
CMC	Carboxymethyl Cellulose
CRE	Catabolite Responsive Element
Da	Dalton
dH_2O	Distilled Water
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotides (ATP, CTP, GTP, TTP)
EC	European Community
EDTA	Ethylenediaminetetraacetic Acid
EMBL	European Molecular Biology Laboratory
ESI	Electrospray Ionisation
EU	European Union
G (g.g)	Gram (gram per gram)
GC	Gas Chromatogram
G/C	Guanine / Cytosine content
gDNA	Genomic DNA
GHG	Greenhouse Gas
GLU	Glutamine
H^+	Hydrogen ion
HC	Hydrocarbon
HIS	Histidine
HMF	hydroxymethyl furfural
HMM	Hidden Markov Model
HPLC	High Performance Liquid Chromatography
Hr	hour
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KEGG	Kyoto Encyclopaedia of Genes and Genomes
L	Litres
lac	Lactose
LB(A)	Luria-Bertoni (Agar)
lb/In^2	Pounds per square Inch
LC	Liquid Chromatography
LTD	Limited
MCMC	Markov chain Monte Carlo
µm	Micrometer
mm	Millimeter
mM	Millimolar
M	Mole
mg	Milligram
µg	Microgram
min	Minute

ml	Millilitre
µl	Microlitre
MOWSE	Molecular weight search
MS	Mass Spectrometry
N	Normality
NAD(PH)	Nicotinamide Adenine Dinucleotide (phosphate)
NCBI	National Centre for Biotechnological Information
NCIMB	The National Collection of Industrial, food and Marine Bacteria
nl	nanolitre
NM	Nanometer
NO_x	Nitrous Oxide gases
OD	Optical Density
ORF	Open Reading Frame
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
pDNA	Plasmid DNA
pMol	picomole
ppm	Parts Per Million
PPP	Pentose Phosphate Pathway
PRSF	Potential scale reduction factor
PTS	Phosphoenolpyruvate Transport System
RbsA	Ribose ATPase
RbsB	Ribose binding protein
RbsC	Ribose permease
RbsR	Ribose Repressor
RCM	Reinforced Clostridial Medium
RNase	Ribonuclease
ROA	Rezex Organic Acid
RPM	Revolutions per Minute
S	Second
SDS	Sodium Dodecyl Sulphate
SSF	Simultaneous Saccharification and fermentation
TAE	Tris Acetate EDTA buffer
TalB	Transaldolase
TE	Tris EDTA
TktB	Transketolase
TLC	Thin Layer Chromatography
TMHMM	Transmembrane helices Hidden Markov Model
TYA	Tryptone Yeast Extract
UK	United Kingdom
US	United States
UV	Ultra Violet
V	Volts
Vis	Visual
v/v	Volume per volume
VOC	Volatile Organic Compounds
w/v	Weight per Volume
w/w	Weight per Weight
xg	Relative Centrifugal Force
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
XylA	Xylose isomerase
XylB	Xylulokinase
XylF	Xylose binding protein
XylG	Xylose ATPase
XylH	Xylose permease
XylR	Xylose repressor
XynA	Endo 1,4-Xylanase
XynB	β-Xylosidase

Chapter 1:
Introduction

1 Introduction

1.1 A Requirement for Fuel Alternatives

Globally, energy demand has risen by seventeen fold over the past century (Demirbas, 2007). Fossil fuels currently meet 85% of the total energy demand. However, its combustion releases greenhouse gases such as CO_2, SO_2 and nitrous oxide (NO_X) into the atmosphere, to the detriment of public health and the environment (Ture *et al.*, 1997, Demirbas, 2007). Transport is the greatest contributor; petrol and diesel combustion results in CO_2, SO_2 and unburned hydrocarbons (HC), the latter of which results in smog formation encountered in urban areas (Galbe and Zacchi, 2002, Zaldivar *et al.*, 2001). Of the greenhouse gases (GHG), CO_2 is described as the major contributor to global warming. Its presence in the atmosphere has increased by over a third from 280 parts per million (ppm) to 365 ppm in 150 years (Galbe and Zacchi, 2002). Over a twenty year period extending from 1970s to 1990s a 27% CO_2 elevation, with an average 0.5 degrees Celsius (°C) global temperature increase has been observed. If no restrictions are enforced on fossil fuel combustion, this trend is likely to continue resulting in an average 2-5°C elevation in global temperature causing rise in sea levels by 2.4 metres (Ture *et al.*, 1997).

In 1997 the Kyoto protocol was introduced and six gases were targeted for mitigation. These include methane, hydrofluorocarbons, perfluorocarbons and sulphur hexafluorides, alongside CO_2 and N_2O. Members signed up to the protocol are required to increase and develop alternative methods to provide forms of energy to limit emissions of these GHG and to abide to the limits assigned to them. Over the 2008 to 2012 period a target 5% GHG abatement of the levels observed in 1990 is sought. Therefore, there is great importance placed on nations to find fuel alternatives that will not contribute to and aim towards mitigating global warming through GHG reduction (United Nations, 1998).

Globally the United States (US) is responsible for one of the highest CO_2 discharge rates. Along with Europe the US has set targets for biofuels to replace petroleum-based fuels (Demirbas, 2007). The US, under the Energy Policy Act, aims to replace over 75% of imported oil with alternatives by 2025 (Hahn-Hagerdal *et al.*, 2006) with an increase in bioenergy three-fold over the next decade (Demirbas, 2007, Hahn-Hagerdal *et al.*, 2006). Under the European Union (EU) directive 2003/30/EC, by the end of

2010 biofuels and other alternative fuels had to hold 5.75% of the market share. By 2020 this will increase to 20% substitution of fossil fuels by alternatives (European Parliament, 2003).

1.2 Biofuels

Biofuels are described as either liquid or gas fuels derived from biomass. Current liquid-based biofuels encompass biodiesel and bioethanol, which are described as first generation biofuels, generated from purpose grown crops. However, both bioethanol and biodiesel bear a number of limitations.

1.2.1 Biodiesel

Biodiesel is used as a diesel alternative. It is produced by transesterification of vegetable oil which can be carried out in the absence or presence of a catalyst such as KOH or NaOH, with methanol (Demirbas, 2007). This process reduces the viscosity of the oil by reducing the large branched structure into a smaller and linear form (Demirbas, 2007). Rapeseed oil is used in Europe and palm oil in tropical regions of the World. Germany is currently the largest producer of biodiesel in Europe, responsible for half of EU biodiesel production and some 2.2 billion litres were produced by the country in 2004 (Frondel and Peters, 2007). However, the advantages of using biodiesel over its fossil fuel counterpart are small. In its production, crops require large quantities of fertilisers and pesticides, this application causes an increase SO_2 and NO_x emissions as well as causing eutrophication of water courses due to surface run-off. In addition, on the claim that the combustion of biodiesel reduces SO_2 emissions by half, it is arguable that improvements to particle filters for fossil fuel based diesel could meet this reduction. Furthermore, it is not known whether biodiesel combustion produces more 'photochemical' smog (Frondel and Peters, 2007).

1.2.2 Bioethanol

Currently, bioethanol is the most widely used biofuel and under the EU quality standard EN228 a 5% ethanol blend in petrol does not require any vehicle engine modifications and is covered by vehicle warrantees (Demirbas, 2007, Frondel and Peters, 2007, Union, 2003).

Brazil and USA are the World's largest producers of ethanol fuels, using sugar cane and corn, respectively. Brazil produces 12 million $Metre^3$ $year^{-1}$ in comparison to Europe,

which produces about one sixtieth of this (Galbe and Zacchi, 2002). Bioethanol produces low CO_2, HC, CO, NO_x and volatile organic acid (VOC) emissions. The production of aldehydes such as acetaldehyde and formaldehyde is of concern as these may affect air quality in urban areas (Galbe and Zacchi, 2002). The use of ethanol provides many advantages, such as adding oxygen to the combustion mixture when added to petrol. Ethanol has a higher heat of vaporisation and a higher octane rating. This therefore replaces the need for toxic octane enhancing additives, and reduces the CO and unburned HCs produced during petrol combustion (Galbe and Zacchi, 2002, Hahn-Hagerdal *et al.*, 2006, Hansen *et al.*, 2005). Ethanol provides approximately two-thirds less energy in comparison to petrol, however, an ethanol-fuelled vehicle can still be expected to do up to 80% of the distance of that of a petrol-fuelled vehicle (Galbe and Zacchi, 2002).

Ethanol also holds a number of disadvantages as a biofuel. Firstly, it is not miscible with diesel. Diesel-ethanol blends are subject to phase separation under cold and damp conditions and therefore the addition of emulsifiers is required (Hansen *et al.*, 2005, Lenz and Moreira, 1980). It also lowers the cetane rating of diesel, which causes longer ignition delays while the fuel vaporises. In addition because of its chemical properties, there are problems with viscosity and lubricity which lead to pump and injector leakage, therefore decreasing fuel delivery and ultimately power to the vehicle. Ethanol is also corrosive to engine components causing swelling of seals and corrosion to injection pumps. Finally, there are issues with handling and storage of ethanol, because of its low flash point and highly flammable vapours (Hansen *et al.*, 2005).

1.2.3 Biobutanol

The problems associated with ethanol can be addressed whilst simultaneously retaining the advantageous qualities, by another alcohol, butanol. Butanol is seen as a superior biofuel to ethanol for many reasons. Butanol has a longer HC chain in comparison to ethanol and thus shares more chemical properties to petrol and diesel (Lenz and Moreira, 1980). Butanol is completely miscible with diesel. Having a low vapour pressure makes the handling and storage of butanol compatible with current infrastructure and this is much easier and safer in comparison to ethanol (Lenz and Moreira, 1980, Durre, 2007). Butanol can be blended to any concentration with petrol, whereas ethanol is limited to 85%. The combustion of butanol does not require any engine modifications and is less corrosive than ethanol (Durre, 2007). DuPont and

British Petroleum (BP) created a partnership in 2006 for developing biobutanol in United Kingdom (UK) (Durre, 2007), as a next generation biofuel. A summary of the advantages of biobutanol is shown in Table 1-1.

Table 1-1: The advantages of biobutanol

Advantages of Butanol
Not hygroscopic
Completely miscible with diesel
Low vapour pressure
Adds oxygen to petrol reducing GHG & unburned HC emissions
Higher octane rating than petrol/diesel
Blending with petrol to any per cent
No engine modifications required
Can be piped using current infrastructure

1.3 Substrates for Biofuel Production

1.3.1 Purpose grown crops

Bioethanol is primarily produced using purpose grown crops, such as corn and sugar cane; and biodiesel from rapeseed and palm oil. This has led to questions about the sustainability of biofuels as a fuel alternative. Large amounts of crops and thus large areas of land are required to meet the energy demand of the World. This could result in less food crop availability and/or land availability for food crop growth. Additionally, the use of intensive agricultural practices can result in environmental impacts, such as soil degradation, toxin accumulation from pesticide use, eutrophication of local water courses and increased NO_x gas emissions, from the use of fertilisers. The latter is most notable, as N_2O, a NO_x gas is a GHG and has more potency than CO_2 (Wheals *et al.*, 1999, Frondel and Peters, 2007). Further controversy over biofuels has come from deforestation in tropical countries in palm oil plantations, which not only has a massive impact on wildlife biodiversity and the environment, but results in yet further CO_2 emissions from human activity. Indeed, a reported average six billion tonnes of CO_2 is emitted annually by this activity alone (Ture *et al.*, 1997).

A further weakness of the current biofuels, bioethanol and biodiesel, comes during the production. Currently only the starch or oily fraction is utilised in each case, respectively, and the remainder is unfermentable. The whole use of the substrate biomass would allow more value to be extracted, making such a process more economically viable with more of the substrate being converted to product. This and along with the environmental impacts could be addressed by employing clostridia to wholly ferment sustainable waste biomass.

1.3.2 Waste biomass

Waste biomass is an alternative to purpose grown crops and may be sourced from agriculture, forestry, industry, and municipal wastes (Zaldivar *et al.*, 2001). These are cheap, readily available and renewable resources and the use of waste biomass can alleviate problems associated with their disposal.

Industries often require waste removal which can be expensive and have an impact on the environment. Organic waste is generally disposed of via landfill. However, this

route is no longer an option due to the EU landfill directive (1999/31/EC). Targets aimed for biodegradable municipal solid waste was reduced by 75% in 2010 and by 50% in 2013 and 35% by 2020, to that of the levels in 1995 (European commission, 1999). There are a number of other options such as incineration, anaerobic digestion and composting available. But another option is to extract further value from the organic fraction of the biomass by converting it to biobutanol.

1.4 Lignocellulose

For biofuel production lignocellulose represents the most abundant renewable resource on the planet (Galbe and Zacchi, 2002, Jones and Woods, 1986). Sources of lignocellulosic wastes are numerous; forestry residues, sawdust, recycled paper and paper mill residues, municipal waste and agricultural waste can all be used. Agricultural waste may consist of corn stover, bagasse, wheat and rice straw. Sources are cheap, plentiful and use materials which may be destined for disposal, which is an industrial expense as well as potentially polluting the soil, water courses and the atmosphere (Demain *et al.*, 2005, Gallert and Winter, 2002). Upon use of such a resource there are a number of factors to consider. Firstly, there may be great seasonal variation of lignocelluloses sourced from agriculture and forestry. Secondly, these resources are required to be transported and stored until use (Lin and Tanaka, 2006). Lastly, it is also important that over-utilisation of agricultural wastes is avoided, as this could potentially lead to environmental damage due to soil organic matter reduction and erosion (Kim and Dale, 2006).

Lignocellulose constitutes the cell wall of plants. It is made up of three main constituents: cellulose which accounts for 40-50%, hemicellulose 25-35% and lignin 15-20% (Gray *et al.*, 2006). The exact ratios of each are source dependent. Cellulose is made of long linear glucose polymer chains linked to one another by hydrogen bonding (Perez *et al.*, 2002). This polymer has a characteristic orderly and compact crystalline structure, thus the individual glucose units linked by β-1, 4 bonds, are highly resistant to enzymatic degradation, making it such a recalcitrant resource. Hemicellulose is associated with cellulose by hydrogen bonding and in turn hemicelluloses are covalently linked to lignin by ferulic acid and ester bonds (Demain *et al.*, 2005, Shallom and Shoham, 2003), together making a cross-linked network. The basic structure of lignocellulose is shown in Figure 1-1.

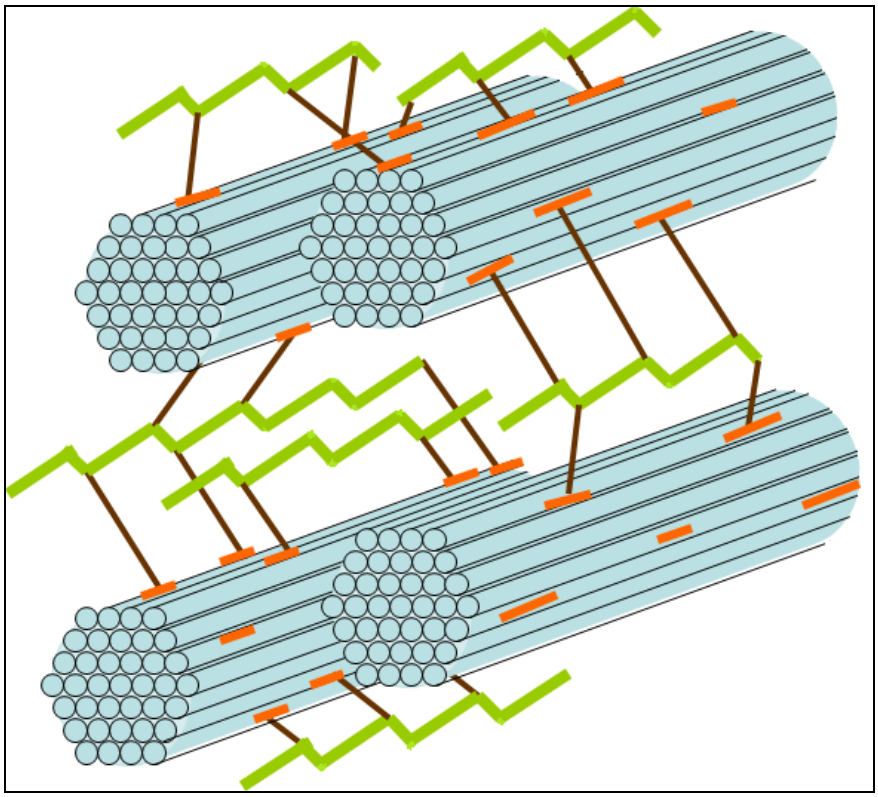


Figure 1-1: The structure of lignocellulose

Lignocellulose is composed of bundles of cellulose (light blue), hemicellulose (orange) is associated with cellulose via hydrogen bonding; and lignin (green) is covalently linked to hemicelluloses, forming a cross-linked network. Taken and adapted from Thostrup, 2006

1.4.1 Hemicellulose

After cellulose, hemicellulose comprises the next largest constituent of lignocellulose. Hemicellulose, in contrast to cellulose, is much easier to degrade. Hemicellulose is a heteropolysaccharide of hexose, pentose and acid sugars and the exact composition is source dependent (Saha, 2003). The most prevalent in nature is heteroxylan, which is composed of a backbone of xylan, a polymer of xylose units held together by β-1, 4 linkages (Saha, 2003, Shallom and Shoham, 2003, Gray *et al.*, 2006). The backbone has heterosaccharides branching off, with the extent and contents of the branching again being source dependent. Sugar units of glucose, galactose, xylose, arabinose, glucuronic acid and mannose may be found (Gray *et al.*, 2006, Perez *et al.*, 2002). The hemicellulose of hardwoods is composed mainly of heteroxylans, and in softwoods it consists mainly of glucomannans. Examples of a few plant resources and their sugar make-up are shown in

Table 1-2. Other than xylans, hemicelluloses may include arabinans, mannans and galactans. As their names suggest they consist of a backbone of arabinose, mannose or galactose, respectively (Perez *et al.*, 2002).

Table 1-2: Examples of plant-based resources and their hemicellulose constituent composition

Plant	Xylose (%)	Arabinose (%)	Glucose (%)	Galactose (%)	Mannose (%)	Glucouronate (%)	Anhydrouronate (%)
Birchwood	89.3	1	1.4	-	-	-	8.3
Rice bran	46	44.9	1.9	6.1	-	-	1.1
Wheat arabinoxylan	65.8	33.5	0.3	0.1	0.1	-	-
Corn fibre	48-54	33-35	-	5-11	-	3.6	-

Some examples of plant sources hemicellulose constituents (Taken from: Saha, 2003)

1.5 Pentose Sugars

Xylose and arabinose are also known as aldopentoses, five carbon sugars carrying an aldehyde group. The name xylose is derived from the Greek word for wood and arabinose from gum Arabic, so called because of the region where it was first isolated. Both sugars are structurally similar, but structural distinction lies with the orientation of the hydroxyl groups on C1-C4 (Figure 1-2). The D- form of xylose and L- form of arabinose are more common in nature than the other stereoisomers (Lee *et al.*, 1970). Pentose metabolism is distinct between prokaryotes and eukaryotes. Prokaryotes use isomerases and eukaryotes use redox reactions, to produce xylulose 5-phosphate which is then fed into the pentose phosphate pathway (PPP) (Bettiga *et al.*, 2008).

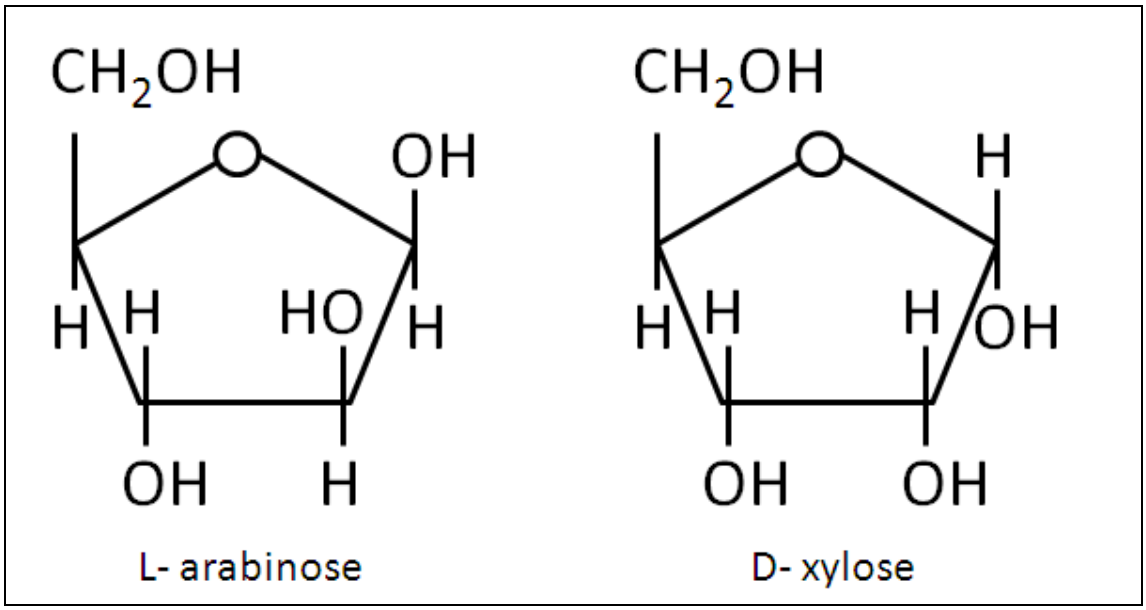


Figure 1-2: Structure of pentose sugars xylose and arabinose

There are two stereoisomers of each sugar, the L-arabinose and D-xylose forms are more prevalent in nature and their structures are shown above.

Pentose sugars have been shown to be fermentable by a wide variety of bacteria such as *Escherichia coli*, *Salmonella enterica*, bacilli, staphylococci, streptomycetes, and thermophiles such as geobacilli, thermoanaerobacteriae and *Thermus* sp. However, pentose sugars cannot be fermented by a large number of yeasts. The commonly used bioethanol producing organism *Saccharomyces cerevisiae*, preferred for its robustness, cannot ferment xylose and thus genetic engineering has been employed to enable xylose fermentation. However, use of such an organism at industrial scale is not feasible if antibiotics are required to maintain a selective pressure. Other yeasts cannot provide an

adequate alternative where *S. cerevisiae* fails. *Pichia stipitis* can ferment xylose, but is inhibited by compounds arising from pre-treatment, whereas, filamentous fungi are tolerant to such inhibitors, but ferment the sugars too slow for industrial use (Hahn-Hagerdal *et al.*, 2006). Some of the solventogenic strains of clostridia have been found to be also tolerant to some of the inhibitors. The presence of furfural, hydroxymethyl furfural (HMF) and glucuronic acid have been shown to have a stimulatory effect with *C. beijerinckii*, *C. butylicum*, *C. saccharolyticum*, with an increase in cell density and Acetone Butanol Ethanol (ABE) production (Ezeji *et al.*, 2007, Ezeji and Blaschek, 2008). Therefore may be advantageous over fungi in the fermentation of plant-based waste streams. However, other inhibitors such as ferulic and *p*-coumaric acids and syringaldehyde had an inhibitory effect on all strains and dramatically reduced the cell density and thus ABE production with increasing concentration in all the strains (Ezeji *et al.*, 2007, Ezeji and Blaschek, 2008).

1.6 The Genus Clostridium

Clostridium is an extensive genus comprising some medically important members which produce toxins. Examples include *C. tetani*, *C. botulinum* and *C. difficile*. Infection with such clostridia causes illness, due to the effect the toxins have on the host. Alternatively, members can be of industrial importance, producing solvents rather than toxins. Examples of industrially important members include *C. acetobutylicum, C. saccharobutylicum* and *C. beijerinckii*. Clostridia are Gram-positive, rod-shaped bacteria and are equipped with flagella allowing cell motility (Jones *et al.*, 1982). A noteworthy feature of clostridia is the inability to grow in the presence of oxygen, i.e. clostridia are obligate anaerobes. Moreover, clostridia are sporulators (Durre and Hollergschwandner, 2004). Sporulation permits survival over long periods of time in hostile environments. Spore formation occurs upon entry into the stationary phase of growth, when nutrients are becoming scarce and toxic metabolites start to accumulate (Durre *et al.*, 1995). Simultaneous with this phase is the toxin or solvent production associated with medical and industrial clostridia, respectively. Solvents are of high importance industrially as a renewable source of chemicals, which can be used as fuels. A selection of clostridial species and the individual solventogenic end-products generated are shown in Table 1-3 below. Clostridial genomes are of low Guanine/ Cytosine (G/ C) content (de Vos *et al.*, 1997), a feature which is also shared with a number of other Gram positive bacteria, such as bacilli, lactobacilli and staphylococci. The genomes of a number of *Clostridium* species have been sequenced, including *C. acetobutylicum* American type culture collection (ATCC) 824 (Nolling *et al.*, 2001), *C. beijerinckii* National culture collection of industrial food and marine bacteria (NCIMB) 8052 (Copeland *et al.*, 2007b) and *C. thermocellum* ATCC 27405 (Copeland *et al.*, 2007a).

Table 1-3: Solvent end-products of a range of clostridia

Clostridium Species	Solvents Produced
C. acetobutylicum	Acetone Butanol Ethanol
C. beijerinckii	Acetone Butanol Ethanol Isopropanol
C. pasteurianum	Acetone Butanol Ethanol
C. saccharobutylicum	Acetone Butanol Ethanol
C. thermocellum	Ethanol

The solvent profile produced is strain dependent in clostridia capable of production.

1.7 Solvent Production in Clostridia

1.7.1 The history of clostridial solvent production

During the former part of twentieth Century Acetone Butanol (AB) fermentation was a very important biotechnological process, surpassed only by ethanol fermentation, as the most successful industrial process. Its origins sprung from a high demand for natural rubber, where a shortage and expense had prompted research for a cheaper synthetic substitute (Durre, 1998). This work was led by Chaim Weizmann whose work led to the isolation of a strain of *Clostridium acetobutylicum*, a strain which he referred to as BY (Gabriel, 1928). BY had the ability to produce acetone, butanol and ethanol as fermentation end-products. However, the work of Weizmann became invaluable for another reason. The onset of World War I demanded an abundance of munitions, smokeless gunpowder – cordite, was required in vast quantities. Cordite manufacture required acetone and with the supply from Austria and Germany interrupted, left only a short supply from the US. Because of the capability of Weizmann's BY strain to produce acetone, Weizmann made his work known and he was later commissioned to produce acetone using the ABE fermentation or Weizmann process, as it was also called (Gabriel, 1928).

Over the next twenty years expansion of AB fermentation continued on a global scale. In addition, improvements were made through the use of molasses, a cheap and plentiful substrate; and new strains, boosting the efficiency and solvent yield of the process (Jones and Woods, 1986, Nolling *et al.*, 2001). Demand for butanol continued and coupled with the requirement for acetone during World War II the AB industry was at its peak. In the Post–war years there was a very rapid decline in AB fermentation, which eventually led to all but a few plants closing.

The decline of the AB fermentation industry was due to the process being no longer economically viable and reasons for this were two-fold. Firstly, butanol could be synthesised by a more efficient and cheaper method, from crude oil. Secondly, competition had arose for the feedstock used at the time, molasses, which were being used in cattle feed. Leading to this once cheap and plentiful carbon source becoming too expensive for the AB industry to compete with the petrochemical industry (Jones and Woods, 1986, Nimcevic and Gapes, 2000).

Research and plant operation continued in very few places post Second World War. Plants in South Africa, China and Russia remained operational for decades afterwards. Eventually research was rekindled again in a wider field, including western societies, prompted by the oil crisis of the 1970s. This led to fuel alternatives being sought and in Brazil the Proalcool program begun, where ethanol produced from the fermentation of sugar cane, replaced the need for the importation of oil as a fuel.

1.7.2 Solvent production in clostridia

In batch cultures, solventogenic clostridia display a biphasic growth pattern. Firstly, acids and gases are generated by fermentation of the carbon source, causing a decrease in pH of the culture medium. In the second stage, the acids are reassimilated concomitantly with carbon source uptake, solvents are produced and the pH rises. These two phases are known as acidogenesis and solventogenesis, respectively. During acidogenesis, acetate, butyrate, CO_2 and H_2 are produced. The conversion of pyruvate from glycolysis to acetyl CoA; and electron transfer from pyruvate to protons via hydrogenase (see Figure 1-3 reactions 1 and 2, 3, 4 respectively) generates the CO_2 and H_2, respectively. During solventogenesis products are strain dependent, in the case of *C. acetobutylicum* and *C. saccharobutylicum* the major products are butanol, acetone and ethanol in the ratio of 6:3:1. With *C. beijerinckii* the same solvents are obtained, but acetone can be further reduced to isopropanol.

The pH at which solventogenesis is initiated is also strain dependent, but is generally below pH 5. The role of the acids is as a final e^- acceptor, however this comes at a cost to the cell because in this reduced form acids are toxic to the cell (Jones and Woods, 1986). The acids at high enough concentrations have a deleterious effect on the cell and therefore a shift to solventogenesis is required as a detoxification method (Mitchell, 1998). The acids behave as uncouplers, partitioning in the cell membrane and this allows the entrance of H^+ from the cell's exterior environment. In the presence of undissociated acids at high concentrations the pH gradient across the cell membrane is collapsed resulting in inhibition of metabolic function due to a decrease of the Adenosine triphosphate/ adenosine diphosphate (ATP/ADP) ratio (Bowles and Ellefson, 1985, Jones and Woods, 1986). The concentration of acids in the medium has been shown to have a concentration-dependent effect on the internal pH (Scotcher *et al.*, 2003). Membrane bound ATPase is responsible for controlling the pH of the cell by expending ATP (Bowles and Ellefson, 1985, Durre and Hollergschwandner, 2004,

Jones and Woods, 1986). Treatment of cells to inhibit ATPase is associated with early onset of solventogenesis, this is caused when ATPase is unable to serve its function in pH control. The resulting fall in the internal pH makes conditions favourable for the acids to be in the undissociated form and an increase in undissociated acids levels triggers solventogenesis (Huang *et al.*, 1985, Mitchell, 1998). However, in some situations it should be noted that if an excess of acids is produced too rapidly during the fermentation, this has a toxic effect on the cells and no solventogenesis takes place. This phenomenon is called 'acid crash' (Maddox *et al.*, 2000).

The pSOL1 megaplasmid carries the main genes for solventogenesis in *C. acetobutylicum* and *C. saccharobutylicum* and loss of this plasmid through continuous subculturing removes the ability of the organism to produce solvents; this is termed degeneration (Cornillot *et al.*, 1997, Scotcher *et al.*, 2003). Conversely, genes for solventogenesis in *C. beijerinckii* are carried on the chromosome, but even so degeneration has still been observed (Chen and Blaschek, 1999). Sporulation is concomitant with the onset of solventogenesis and the Spo0A protein is termed the 'master regulator' of endospore formation (Durre and Hollergschwandner, 2004, Mitchell, 1998). Studies have linked the effect Spo0A with the onset of sporulation, granulose formation and a metabolic shift to solventogenesis (Bahl *et al.*, 1995). The inactivation of the *spo0A* gene in *C. acetobutylicum* is associated with a severe deficiency in solventogenesis; no cell swelling and no endospore formation, associated with granulose formation and sporulation. Whereas the overexpression of *spo0A* enhances the butanol concentration and sporulation by accelerating time taken for the expression of genes involved in solvent formation and sporulation, in comparison to the wild type (Harris *et al.*, 2002).

Spore formation in clostridia occurs when the parent cell undergoes asymmetric septation. The mother cell produces as many as five spores per cell and is lysed when the forespore(s) reach maturity. The released spores are covered in a slime coat and have appendages. It is this process of sporulation that allows the organism to survive the impending unfavourable conditions for life (Durre and Hollergschwandner, 2004). Produced during solventogenesis, butanol is considered to be the main toxicant. Butanol is hydrophobic and causes disruption to the phospholipids found in the cell membrane, increasing the fluidity. This disrupts functioning of ATPase and nutrient

uptake, inhibiting cell growth and hence the amount of butanol yielded (Huang *et al.*, 1986, Jones and Woods, 1986).

1.7.3 The future of clostridial fermentations

The replacement of oil as a fuel is of current importance as within the next fifty years, this finite fuel resource is estimated to be fully depleted (Demirbas, 2007). Also, global warming has prompted a greater response into alternative fuels in recent years, in an attempt to cut the carbon (C) emissions produced by fossil fuel combustion. However the success of an alternative fuel would be subject to a number of criteria. The major issues affecting the competitiveness of biobutanol with petrochemicals is the cost of the substrates and the product recovery and are therefore both important targets for research. Substrates that are plentiful and of low value are necessary, lignocellulosic material presents an attractive resource in this respect (Ni and Sun, 2009). The low product yield due to the inhibitory effect butanol has on the growth of the cells could be improved by metabolic engineering to increase the tolerance and improve the ratio of butanol and hence improve the butanol production (Ni and Sun, 2009, Claassen *et al.*, 2000). Additionally the method of product recovery could also allow for improved solvent recovery and reduce the solvent inhibition effect on the cells, by the way of methods such as gas stripping, which remove solvents continuously throughout the duration of the fermentation (Ni and Sun, 2009).

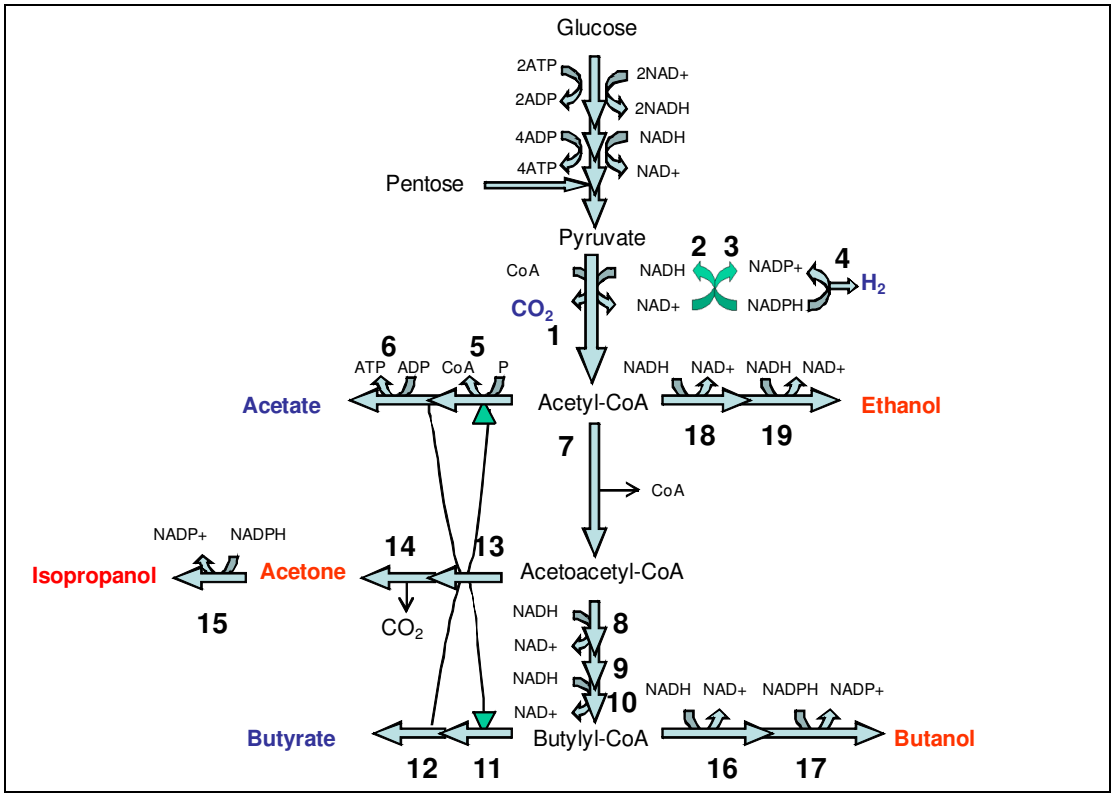


Figure 1-3: Biochemical pathway of solvent production in *C. beijerinckii*

The biochemical pathway shows how solvents are produced in *C. beijerinckii*. Blue signifies the major the products obtained from acidogenesis phase and the red the major products of solventogenesis. The same steps occur in *C. acetobutylicum* with the exception of step 15. **1** pyruvate ferredoxin oxidoreductase; **2** Nicotinamide adenine dinucleotide (NADH) ferredoxin oxidoreductase; **3** Nicotinamide adenine dinucleotide phosphate (NADPH) ferredoxin oxidoreductase; **4** hydrogenase; **5** phosphotransacetylase; **6** acetate kinase; **7** thiolase; **8** β-hydroxybutyryl CoA dehydrogenase; **9** crotonase; **10** butyryl CoA dehydrogenase; **11** phosphotransbutyrylase; **12** butyrate kinase; **13** CoA transferase; **14** acetoacetate decarboxylase; **15** isopropanol dehydrogenase; **16** butyraldehyde dehydrogenase; **17** butanol dehydrogenase; **18** acetaldehyde dehydrogenase; **19** ethanol dehydrogenase.

1.8 Carbohydrate Utilisation in Clostridia

Clostridia are able to metabolise a wide diversity of carbohydrates, making these organisms very versatile when it comes to the range of substrates which can be used for solvent production. This potentially allows for a number of cheap resources in the form of waste streams to be utilised.

1.8.1 The use of polymers

Cellulose, a linear glucose polymer, is the most abundant polymer in nature, however, not all clostridia are able to directly utilise it as a carbon source. *C. acetobutylicum* possesses a cellulosome with components necessary for cellulose degradation (Sabathe *et al.*, 2002, Sabathe and Soucaille, 2003). The cellulases are expressed and have high similarity to those of clostridia known for their ability to use cellulose, such as *C. cellulolyticum* and *C. thermocellum.* However, *C. acetobutylicum* is unable to utilise cellulose directly (Lopez-Contreras *et al.*, 2003, Sabathe *et al.*, 2002). It has been noted that cellulosomes have elevated activity to crystalline cellulose, however despite a cellulosome complex being secreted from *C. acetobutylicum* it is inactive towards cellulose. Reasons for this may lie in the absence of exoglucanases, which are required to work synergistically alongside endoglucanases for efficient cellulose degradation (Mitchell, 1998); or the regulation of the cellulases. However, *C. acetobutylicum* is reported to exhibit endoglucanase, cellobiosidase and cellobiase activities when grown in the presence of a number of carbon sources, such as glucose, xylose, cellobiose and mannose (Lee *et al.*, 1985a). Endoglucanase, *celG*, is expressed in the presence of lichenan, but not in the presence of cellobiose, in contrast to *C. cellulovorans,* where the orthologous gene is expressed in the presence of cellulose and cellobiose (Lopez-Contreras *et al.*, 2003). A similar finding was apparent with another endoglucanase, CelF, which again possessed higher activity in cells grown on lichenan and xylose in comparison to glucose and cellobiose (Lopez-Contreras *et al.*, 2004). Recent research revealed that specific chaperones involved in the secretion of Cel48 and probably Cel9G and Cel9E endoglucanases are missing or insufficiently synthesised because their interaction with the secretory system has a deleterious effect to the cell (Mingardon *et al.*, 2011). Another reason for these results may lie with the lack of catabolite responsive element (CRE) sites in the coding or promoter regions surrounding the putative cellulosome genes, and thus CcpA (catabolite control protein A) cannot bind

and negatively regulate these genes (Lopez-Contreras *et al.*, 2004). Carbon catabolite repression (CCR) of most genes is regulated by a carbon response element. A CRE is a palindromic sequence which either overlaps with a gene promoter or is present in the open reading frame of the gene, where transcription is either unable to initiate or elongation is not able to take place, respectively (Hueck and Hillen, 1995, Seidel *et al.*, 2005). It is CcpA which binds to the CRE in response to the presence of a catabolite, for example glucose and dependent upon the regulatory mechanism, the downstream genes or operon are expressed or repressed (Kim *et al.*, 2005).

Another glucose polymer, starch is also fermentable by clostridia. In both *C. acetobutylicum* and *C. beijerinckii* starch degrading enzymes are present (Mitchell, 1998). α-amylase an endo-acting enzyme cleaves α-1,4 glycosidic links of starch, with the major products being glucose, maltose and maltotriose (Paquet *et al.*, 1991). The other enzyme is glucoamylase which is an exo-acting enzyme cleaving glucose units from the non-reducing ends of the starch molecule.

Xylan the polymer of xylose, has been found to be incompletely used by *C. acetobutylicum* (Mitchell, 1998). An excess of xylose in the culture medium improves the activity of xylanases. Xylanases have an optimal activity at pH6 and thus as the fermentation proceeds, the pH will become unfavourable. Based on this finding the use of a chemostat to regulate the pH at the optimum was found to aid xylan consumption (Lee *et al.*, 1985b). Xylanase genes (Xyn) found on the megaplasmid, endo 1,4-xylanase (*xyn10A*) and β-xylosidase (*xyn10B*) have been cloned from *C. acetobutylicum* and the encoded proteins were shown to convert xylan and xylo-oligosaccharides to xylobiose and xylotriose or xylose, respectively. Xyn10A was found to be moderately active towards a number of substrates including lichenan, carboxymethyl cellulose (CMC), β-glucan, p-nitrophenyl monosaccharides and avicel. Xyn10B was found to only be slightly active towards CMC, p-nitrophenyl monosaccharides and arabinogalacturonic acid (Ali *et al.*, 2004, 2005).

1.8.2 The use of hexose sugars

Much work has been conducted with regards to the utilisation of di- and monosaccharides of hexose sugars. Hexose sugar uptake and metabolism is mediated by the phosphoenolpyruvate dependent phosphotransferase system (PTS). The PTS allows the uptake and phosphorylation of the carbon source via transferring phosphate

from phosphoenolpyruvate along a chain of proteins, enzyme I, HPr and enzyme II domains A and B. Phosphorylation of the substrate occurs concomitant with entry into the cell (Mitchell, 1998, Tangney *et al.*, 2001). In *C. acetobutylicum*, the PTS is the mechanism for glucose, maltose, sucrose and lactose uptake (Tangney and Mitchell, 2000, 2007, Tangney *et al.*, 2001, Yu *et al.*, 2007) and in *C. beijerinckii* glucose, glucitol and sucrose uptake have all been characterised (Mitchell *et al.*, 1991, Reid *et al.*, 1999, Tangney *et al.*, 1998). These characterised PTS are subject to catabolite repression in the presence of glucose, resulting in the sugars not getting utilised until the glucose is exhausted from the culture medium.

1.8.3 The use of pentose sugars

No PTS has been found in bacteria for the uptake of pentose sugars. With the exception of *C. acetobutylicum*, little research has been carried out to identify the genes involved in the uptake, metabolism and regulation of pentoses in solventogenic clostridia. Several studies have determined that *C. acetobutylicum* is able to utilise pentose sugars such as arabinose and xylose as sole carbon sources (Ounine *et al.*, 1983, Ounine *et al.*, 1985). However, if a cheap and abundant resource such as lignocellulosic biomass is to be utilised efficiently, to maximise the productivity of solvents, it is necessary to gain a fundamental understanding of the pathways involved and their regulation. Glucose is the preferred carbon source of many bacteria including clostridia. Indeed, diauxie is observed when *C. acetobutylicum* cultures are grown on media containing both xylose and glucose. It is not until the glucose had been fully depleted from the medium that xylose is utilised (Ounine *et al.*, 1985). Cultures growing on glucose produce a higher butanol yield compared to those growing on xylose. A 28% conversion of xylose to ABE solvents (29% for arabinose), in comparison to 32% on glucose has been observed (Ounine *et al.*, 1983). The growth of *C. acetobutylicum* on arabinose is comparable to glucose, and as a result yields of ethanol and acetone at the end of fermentation are similar. Cultures on xylose are slower to grow and butanol has a more pronounced inhibitory effect on cells growing on xylose (Ounine *et al.*, 1985), accounting for the difference in the yield obtained on xylose in comparison to the other sugars.

1.9 Aims of Research

Lignocellulose is an attractive resource for biofuel production in the future, as it is the most abundant polymer on the planet and is sustainable. To utilise sustainable or waste substrates efficiently as possible it is necessary to convert all constituents to product. One component of lignocellulose, hemicellulose, is usually pentose-rich and organisms currently used to produce biofuel yeasts, are often unable to utilise these pentose sugars and additionally do not produce biobutanol. Conversely, clostridia are able to utilise a wide variety of substrates. In the light of glucose catabolite repression in most bacteria, whereby the presence of glucose affects the utilisation of other sugars at the transcription level, it is necessary to gain a fundamental understanding of the utilisation of other sugars, including pentose sugars in such a biofuel-producing organism. Very little research has been conducted with regards to pentose sugar use in solventogenic clostridia, with the exception of some work in *C. acetobutylicum* (Gu *et al.*, 2010). Knowledge in this area could benefit future research, revealing targets for genetic manipulation, more efficient carbon source utilisation and improved substrate to product yields. In addition, such knowledge could provide information to biofuel producers on potential substrates that could be utilised, how they can be efficiently and wholly utilised and the potential product yields. *C. beijerinckii* is capable of utilising a wide range of sugars. It has been observed from a genomic analysis of phosphotransferases for sugar uptake that there are 42 putative systems in *C. beijerinckii* in comparison to 13 putative PTS in *C. acetobutylicum* (Mitchell, personal communication). In terms of non-PTS for sugar uptake, such as ATP binding cassette (ABC) transporters, symporters etc., there are 30 putative transporters in *C. beijerinckii* in comparison to 6 putative transporters in *C. acetobutylicum*. Thus, *C. beijerinckii* may confer superior carbohydrate utilisation capabilities for biofuel production. *C. beijerinckii* is the preferred solventogenic clostridial strain by the industrial sponsors of this project, Green Biologics limited (LTD). Little research has been carried out into the ability of *C. beijerinckii* to grow on pentose sugars or the effect mixed glucose-pentose sugar media has on sugar use preference, as well as the organism's ability to utilise the xylose polymer xylan. So it stands to reason that research on this organisms' ability to utilise constituents of lignocellulosic waste for future industrial biobutanol production is highly beneficial.

The hypothesis

The purpose of this study is to therefore investigate the ability of *C. beijerinckii* to utilise the pentose sugars xylose and arabinose.

The hypothesis was investigated using a two-tier approach. Firstly to assess the competency of *C. beijerinckii* to utilise the pentose sugars D-xylose and L-arabinose. The effect of a well known carbon catabolite repressor, glucose on the utilisation of these sugars was also investigated, as it is likely any future process using waste streams will be based on a mixture of sugars. Finally the ABE yields from pentose sugar fermentations were monitored. Then following on from this to assess if *C. beijerinckii* is capable of using pentose sugars in the natural state, as the xylose polymer, xylan and a pentose-rich waste stream. The second objective was to propose gene systems involved in the uptake and utilisation of pentose sugars via bioinformatic analysis and to characterise key genes involved in the utilisation of xylose and arabinose by a two tier approach, using proteomic analysis of whole cell protein extracts of *C. beijerinckii* grown on xylose or arabinose then gene cloning and characterisation *in vivo,* employing mutant strains of *Escherichia coli* lacking the key genes essential for pentose sugar utilisation.

Chapter 2:
Materials and Methods

2 Materials and Methods

2.1 Bacterial Strains

Clostridium beijerinckii NCIMB 8052 was obtained from The National Collection of Industrial, Marine and Food Bacteria (NCIMB, Aberdeen, Scotland). *Escherichia coli* UP1089 was purchased from the *E. coli* genetic stock centre (Yale). *E. coli* XK100 was obtained from Dr. W. Mitchell (Heriot-Watt University, Scotland). *E. coli* DS941 was obtained from Prof. I. S. Hunter (University of Strathclyde, Scotland) and *E. coli* TOP10 from Invitrogen (UK). Genotypes of the *E. coli* strains used in this study are given in Table 2-1.

Table 2-1: Details of *E. coli* strains used in this study

***E. coli* strain**	**Genotype**	**Phenotype of relevance**	**Reference**
UP1091	*araA204*	L-arabinose isomerase negative (*araA*$^-$)	(Englesberg, 1961)
DS941	*thr1 leu6 hisG4 thi1 ara14 proA2 argE3 galK2 sup37 xyl15 mtl1 tsx33 str31 recF143 supE44 lacl*^{q}*ZΔM15*	D-xylose isomerase negative (*xylA*$^-$) L-ribulokinase negative (*araB*$^-$)	(Toivonen and Jacobs, 1999) (Englesberg, 1961)
XK100	*pro xxk*	D- xylulokinase negative (*xylB*$^-$)	(Scangos and Reiner, 1978)
TOP10	F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*Δ139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G		

2.2 Growth of *C. beijerinckii*

2.2.1 Media

2.2.1.1 Reinforced clostridial media (RCM)

RCM (Oxoid) was made by adding 38 gram (g) in 1 litre (L) of distilled water (dH_2O).

2.2.1.2 Clostridial basal media (CBM)

CBM consisted of: 10 g/L carbon source (Arabinose, Sigma-Aldrich; Glucose, Fisher Scientific; Xylose, Merck), 0.2 g/L $MgSO_4.7H_2O$ (Sigma), 0.01 g/L $MnSO_4.4H_2O$ (BDH Chemicals), 0.01 g/L NaCl (Fisher Scientific), 0.01 g/L $FeSO_4.7H_2O$ (Sigma-Aldrich), 1 milligram (mg)/L *p*-aminobenzoic acid (Sigma), 2 microgram (µg)/L biotin (Sigma), 1 mg/L thiamine HCl (BDH Chemicals), 4 g/L casein hydrosylate (Oxoid), 0.5 g/L K_2HPO_4 (BDH Chemicals) and 0.5 g/L KH_2PO_4 (BDH Chemicals) (O'Brien and Morris, 1971). All of the ingredients were autoclaved together with the exception of the iron and vitamins which were made as stock solutions and filter sterilised. The carbon sources were also filter sterilised. The phosphate solutions were autoclaved separately and all ingredients added together aseptically after autoclaving.

2.2.1.3 Tryptone yeast extract acetate medium (TYA)

TYA consisted of per litre dH_2O: 60 g sugar (Sigma-Aldrich), 6 g tryptone (Oxoid), 2 g yeast extract (Oxoid), 3 g ammonium acetate (Sigma-Aldrich), 0.5 g KH_2PO_4 (BDH Chemicals), 0.3 g $MgSO_4.7H_2O$ (Sigma-Aldrich), and 10 mg $FeSO_4.7H_2O$ (Sigma-Aldrich) (Batycka *et al.*, 2006). Sugar solutions were autoclaved separately and added aseptically to TYA medium after autoclaving.

2.2.2 Revival of spores

C. beijerinckii spores (1 ml), kept at 4°C in suspension, were heat shocked (10 min, 80°C), inoculated into reinforced clostridial medium (RCM, Oxoid), and grown in an anaerobic cabinet (Modular Atmosphere Controlled System, Don Whitely) overnight under an atmosphere of 80:10:10 (N_2:H_2:CO_2) at 34°C. Cultures of 5% volume per volume (v/v) were used as starter cultures for subculturing and for experiments.

2.2.3 Growth of *C. beijerinckii*

Starter cultures 5% (v/v) were inoculated into 20 ml volumes of either CBM or TYA supplemented with 1% weight per volume (w/v) of the appropriate sugar (glucose, xylose or arabinose), were grown overnight and 5% (v/v) was then used to inoculate 100 millilitre (ml) volumes of TYA or CBM (100 ml) supplemented with 1% (w/v) of the appropriate sugar. Cultures were grown overnight and 5% (v/v) was used to

inoculate media for experiments. All experiments were carried out in an anaerobic cabinet under the conditions previously mentioned and all media was made anaerobic by incubating overnight in an anaerobic cabinet prior to use.

2.2.4 Optical density and pH measurement

Optical density (OD) was measured at 650 nanometres (nm) (Nicolet Evolution 300 spectrophotometer, Thermoelectron Corporation) and pH was measured (Jenway 3100 bench top pH meter, calibrated with buffers of known pH). Samples with an OD > 0.6 were diluted with dH_2O in order to maintain a reading of <0.6.

2.3 Fermentation Conditions

Sartorius Biostat® A plus vessels (1 L) and the BioPAT® MFCS SCADA system software were used to control and record conditions during fermentations. The TYA medium was autoclaved in the 1-L fermenter vessel. Sugar solutions (either glucose, arabinose or xylose) were autoclaved separately and added to the vessels aseptically prior to the fermentation. Prior to inoculation each vessel was sparged with N_2 gas for 1 hour (hr) to achieve anaerobic conditions. Starter cultures 5% (v/v) were grown in an anaerobic cabinet using the conditions previously mentioned, were inoculated into 20 ml volumes of either TYA supplemented with 1% weight per volume (w/v) of glucose, were grown overnight and 5% (v/v) was then used to inoculate 150 ml volumes of TYA supplemented with 1% (w/v) of glucose. Cultures were grown overnight and 5% (v/v) was aseptically injected into each fermenter vessel and grown at 34°C and agitated at 200 revolutions per minute (rpm).

2.4 Waste stream Degradation

2.4.1 Determination of total carbohydrates in biomass

Following the method of Sluiter *et al*. (2008), triplicate samples of weighed 0.3 g dried and milled (Kenwood 'Smoothie' blender) dried distillers' grains (DDGS) were placed in 100 ml bottles and 3 ml of H_2SO_4 (Fisher Scientific; 72% weight per weight (w/w)) was added. Samples were incubated for 1 hr at 30°C at 200 rpm on a horizontal shaking incubator (Unitron, HT Infors). dH_2O (84 ml) was added to each bottle to form a 4% (w/w) H_2SO_4 solution (final concentration). Recovery standards were prepared (1 $g.L^{-1}$ glucose, 1 $g.L^{-1}$ xylose, 0.5 $g.L^{-1}$ arabinose, 0.1 $g.L^{-1}$ galactose, and 0.05 $g.L^{-1}$ mannose). H_2SO_4 (95% w/w, 236 microlitres (µl)) was added to achieve a 4% (w/w,

final conc.). The bottles were weighed, autoclaved (121°C, 1 hr) and re-weighed. Triplicate 10 ml samples from each bottle were dispensed into 50 ml sized bottles and solid Ca_2CO_3 (Sigma) was added slowly to achieve pH 5–6. Each sample supernatant was transferred to 2 ml microcentrifuge tube and centrifuged (10 minutes (min), 12000 rpm, Eppendorf 5810R, Helena Biosciences, Sunderland). The supernatants were filtered through 0.2 µM cellulose acetate syringe filters and sugar content analysed by HPLC (section 2.5.2).

2.4.2 Acid and Enzymatic treatment of dried distillers' grains

H_2SO_4 was used to reduce the pH of each bottle (0.08 molar (M) final concentration). DDGS samples were autoclaved and 10 M NaOH (Sigma) was used to increase the pH to pH 5.0 after cooling. CTec (cellulase, 15% (w/w), Novazymes) and HTec (0.5% w/w, hemicellulase, Novazymes) were added and the reaction mixture was incubated at 50°C on a horizontal shaking incubator (200 rpm, Unitron, HT Infors) for 24 hours.

2.5 Carbon and Solvent-Content Determination

Culture samples (1–2 ml) were centrifuged (10 min, 13000 rpm, MSE microcentaur bench top centrifuge, Sanyo), supernatants were removed, filtered with 0.2-micrometers (µm) cellulose acetate syringe filters and frozen (–20°C). Samples were thawed prior to gas chromatography and high pressure liquid chromatography analysis and transferred to glass vials.

2.5.1 Gas Chromatography (GC)

Solvent analysis was carried out on a Chrompack 9001 gas chromatograph with flame ionisation detector using a 10 metre length, 0.32 millimetres (mm) diameter CP SIL 5CB column (Chrompack). A calibration curve was generated with ethanol, acetone and butanol (all obtained from Sigma) standards of known concentration (0.25%, 0.5%, 0.75%, 1% and 1.5% v/v).

2.5.2 High Performance Liquid Chromatography (HPLC)

Samples were filtered and acidified with 0.25% (v/v) 1 M H_2SO_4. A Varian 920 LC system fitted with an integrated refractive index detector and ultraviolet-visual (UV-VIS) dual wavelength detector (Varian Ltd) was used to detect sugars and acids, respectively. Sugars and acids were separated on a Rezex (rezex organic acid) ROA Organic acid ion (H^+) 8% 300 × 7.8 mm column (Phenomenex) with 0.005 normality

(N) H_2SO_4 as the mobile phase at a flow rate of 0.5 ml/min. Sugars were detected by the refractive index detector and acids by UV detection at 210 nm. A calibration curve was produced with acetic, butyric acid, glucose, arabinose and xylose solutions of known concentrations (0.2%, 0.4%, 0.6%, 0.8%, 1% w/v for sugars and v/v for acids).

2.6 Xylan Degradation

2.6.1 Xylanase kinetics

A xylanase stock solution (0.25%, w/w Xylanase NS22036, Novozymes) was prepared in sterile dH_2O and stored at 4°C. The stock solution of xylanase (100 µl) was added to a Tryptone Yeast Extract (TYA, 20 ml) solution supplemented with beech wood or birch wood xylan (Sigma-Aldrich; 1% (w/v)), incubated at 37°C and shaken at 200 rpm in a horizontal shaking incubator (Unitron, HT infors). At various intervals (before the addition of xylanase and 30 min, 1 hr, 3 hr, 6 hr and 24 hr after the addition of xylanase) a 1 ml sample was taken and heated (100°C, 15 min) to inactivate the enzyme.

2.6.2 Growth of *C. beijerinckii* on xylan

An overnight starter culture (5% v/v) was used to inoculate TYA broths (20 ml) supplemented with 1% (w/v) xylose and grown overnight in an anaerobic cabinet (as previously described). Overnight cultures 5% (v/v) were inoculated into TYA (60 ml) with 1% (w/v) xylose solution and was grown overnight and the following day, a 5% (v/v) was inoculated into TYA (20 ml) supplemented with either 1% or 5% (w/v) of xylose, xylan, xylan pre-treated with xylanase for 24 hr (as in section 1.6.1), xylan supplemented with 0.1% (w/v) xylose. In buffered media experiments 1% (w/v) Ca_2CO_3 (Sigma) was added.

2.7 Thin layer chromatography (TLC)

The samples (2µl) were loaded onto a TLC plate (Polygram Sil G pre-coated plastic sheets, Macherey-Nagel) and placed into a glass tank containing ethyl acetate:pyridine:acetic acid:propanol:dH_2O 5:2:2:1:1 (v/v). The tank was sealed with a glass lid and the plate was removed when the solvent front was approximately 2 cm from the top of the plate. The TLC plate was stained with thymol sulphuric acid reagent (thymol (0.53% w/v) in ethanol with 5.3% (v/v) H_2SO_4) and heated at 120°C for 5 min.

2.8 Bioinformatics

2.8.1 Genomic DNA sequence database

Putative gene systems for xylose and arabinose utilisation were identified by the Kyoto Encyclopaedia of Genes and Genomes (KEGG) http://www.genome.jp/kegg/catalog/orglist.html.

2.8.2 Basic local alignment search tool (BLAST) analysis

BLAST (blastp program) was used to compare the potential genes with those of other organisms, using the national centre for biotechnological information (NCBI) open reading frame (ORF) finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html).

2.8.3 Radial trees

The Amino Acid (AA) sequences were input into ClustalW2 http://www.ebi.ac.uk/Tools/clustalw2.html, which produced a .nxs file. The .nxs file was then used to carry out the phylogenetic analysis using MrBayes http://mrbayes.sourceforge.net/ (using amino acid model set to pr=mixed and lset rates = gamma. The (Markov chain Monte Carlo) MCMC analysis was ran with nchains=1 and ngen=300000 producing standard deviation of the split frequencies of <0.05 and SumT Potential Scale Reduction Factor (PRSF) =~1.000) producing a .tre file. TreeView (Version 1.6.1) http://taxonomy.zoology.gla.ac.uk/rod/treeview.html was then used to produce radial tree using the .tre file (The input amino acid sequences taken from either KEGG or obtained via the accession numbers from published data, which were input into the European molecular biology laboratory (EMBL) nucleotide sequence database http://www.ebi.ac.uk/embl/ to obtain the AA sequence .respectively from files).

2.8.4 Multiple alignments

The Amino Acid (AA) sequences were input into ClustalW2, which produced an .aln file and then GeneDoc (version 2.6.002) program (downloaded from: http://www.nrbsc.org/gfx/genedoc/) was then used to produce the alignment.

2.8.5 Prediction of transmembrane helices in proteins

The transmembrane helices Hidden Markov model (TMHMM) Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) was used. Amino acid sequences were taken from KEGG and input into the server, which gave a data output of the predicted

transmembrane structure of the queried sequences, which was then input into TMRPres2D version 0.93 (http://bioinformatics.biol.uoa.gr/TMRPres2D/).

2.9 Cloning of candidate pentose sugar utilisation genes

2.9.1 Reagents

2.9.1.1 Ampicillin

Ampicillin sodium salt (Sigma, 50 µg/ml) stock solutions of (50 mg/ml) were made, filter sterilised and stored in 1 ml aliquots at -20 °C until use. These were thawed and added to LB agar after autoclaving, when the agar had cooled sufficiently before pouring plates.

2.9.1.2 Agarose gels

1% (w/v) gels were made by adding 1 g of Agarose multipurpose (Bioline) to a solution of 10 ml Tris Acetate Ethylenediaminetetraacetic acid (EDTA) (TAE) buffer x 10 concentration, consisting of 890mM Tris borate, 20mM EDTA at pH8.3(Sigma); with dH_2O (90ml) and either 3 µl Ethidium Bromide (10 mg/ ml) from Sigma or 10 µl Safe View from NBS biologicals (consisting of 96.9% H_2O, <0.1% $C_{21}H_{28}N_4$, <1% tris base, <1% boric acid, <1% EDTA).

2.9.1.3 Luria-Bertoni (LB) media

Agar plates consisted of: 10 g/L tryptone (Oxoid), 5 g/L yeast extract (Oxoid) and 10 g/L NaCl (Fisher Scientific). For LB agar (LBA) 12 g/L technical agar number 3 was added (Oxoid) and for broths the same constituents were added together but in the absence of agar.

2.9.1.4 MacConkey agar

MacConkey agar consisted of: 20 g/L bacteriological peptone (Oxoid), 10 g/L carbon source, 5 g/L bile salts (Oxoid), 5 g/L NaCl, 1 ml/L from a 3% (w/v) solution of neutral red, 1 ml/L from a 0.1% (w/v) solution of crystal violet and 12 g/L agar technical number 3 (Oxoid). The sugar solution was autoclaved separately and added aseptically afterwards. Neutral red stock solution was made by 0.3 g added to 10 ml dH_2O and Crystal violet 0.01 g added to 10 ml dH_2O, these solutions were stored at room temperature and added to media prior to autoclaving.

2.9.2 Deoxyribose nucleic acid (DNA) purification

DNA was prepared from overnight cultures of *C. beijerinckii* grown in 20 ml volumes of RCM (as section 1.1.1). This purification was carried out with a Wizard® Genomic DNA purification kit (Promega). A 1 ml sample of the overnight culture was centrifuged (13200 relative centrifugal force (xg), Eppendorf bench top centrifuge 5415D, Helena Biosciences, Sunderland) in a 1.5 ml microcentrifuge tube for 2 min and the supernatant was removed, leaving behind a pellet of cells, which was resuspended in another 1 ml of overnight culture and centrifuged and the supernatant was removed as before. The pellet of cells was then resuspended in 480 µl of 50 millimolar (mM) EDTA (BDH chemicals) and then 120 µl of a 10 mg/ml Lysozyme solution (Sigma-Aldrich) added and gently mixed. The solution was incubated (45 min, 37°C), centrifuged (13200 xg, 2 min) and the supernatant removed. Nuclei lysis solution (600 µl) was used to gently mix and resuspend the pellet. The mixture was incubated (80°C, 5 min), cooled to room temperature Ribonuclease (RNase) solution (3 µl) was added and the sample was agitated by tube inversion (5-fold). The sample was incubated (37°C, 40 min), and cooled to room temperature. Protein solution (200 µl) was added and the sample was vortexed for 20 seconds (s) followed by 5 min incubation on ice. The solution was then centrifuged (3 min, 13200 xg), the supernatant was transferred into a sterile 1.5 ml microcentrifuge tube and isopropanol (600 µl) was added. The tube was repeatedly inverted until DNA strands were visualised. The samples were centrifuged (2 min, 13200 xg) and the supernatant was discarded. Ethanol (600 µl, 70% v/v) was added; the tube was inverted gently and centrifuged (13200 xg, 2 min). The ethanol was removed and the pellet was left to dry (15 min). The pellet was resuspended in DNA rehydration solution (100 µl, 1 hr, 65°C) and gently agitated. The DNA was then stored at 4°C. This DNA was then used for PCR.

2.9.3 Polymerase chain reaction (PCR) conditions

Genomic DNA gDNA of *C. beijerinckii*, purified as previously stated was used to amplify CBEI4457, 4452 and 4455, putative L- arabinose isomerase, L-ribulokinase and L-ribulose 5-phosphate 4-epimerase (*araABD)* genes, and CBEI2383 and 2384, putative D- xylose isomerase and D-xylulokinase (*xylAB* genes), respectively. Primers, conditions and reagents used are shown in Table 2-2, Table 2-3 and Table 2-4. Primers were designed by eye by looking for GC rich regions up- and downstream of the ORF, these were then synthesised by Eurofins-MWG.

Table 2-2: Primer sequences used to amplify *araABD* and *xylAB* candidates

Putative gene	Locus tag	Primer	Sequence
araA	CBEI4457	4457F	5'-ACCGCAAGGCTGTCTTTACC-3'
		4457R	5'-GGGGAGTGGCAAATATTAGG-3'
araB	CBEI4452	4452F	5'-GGACGGTAACGGTATGAGGGCT-3'
		4452R	5'-AGCTGATCCTTAGCCGCTTCAATGA-3'
araD	CBEI4455	4455F	5'-AGCCTCAATACACATACAGTGCTGAAC-3'
		4455R	5'-TGAATCGCTGCAATTTCTCTTCCTT-3'
xylA	CBEI2383	2383F	5'- GGGCTAGTTATTACTAACGTTGGC -3'
		2383R	5'- CCCGTACAAGTCGGAATAAAAGC -3'
xylB	CBEI2384	2384F	5'- CATGCTTTTATCCCGACTTGTAC -3'
		2384R	5'- AAATCCGATACTACTACACACCC -3'

Table 2-3: PCR conditions used to amplify *araABD* candidates

Stage	Locus Tag	Conditions	
Hot start		95°C	5 min
Denaturation		95°C	1 min
Annealing	CBEI4457	58.5°C	1 min (25 cycles)
	CBEI4452	56.3°C	
	CBEI4455	53.7°C	
	CBEI2383	60.5°C	
	CBEI2384	56.8°C	
Extension		72°C	1 min
Final extension		72°C	10 min
Soak		4°C	∞

Table 2-4: Reagents used for PCR amplification of *araABD* candidates

(A) Biotaq (Bioline)

Component	Volume (µl)
10x NH_4 buffer	5
$dNTP^+$ (100 mM)	1.25
$MgCl_2$ (50 mM)	2
Biotaq (5 units/µl)	2.5
Forward primer (100 pmol)	1
Reverse primer (100 pmol)	1
Template* (190 µg/ml)	2
Molecular grade dH_2O*	35.25
Final reaction volume	50

(B) Biomix (Bioline)

Component	Volume (µl)
Biomix	25
Forward primer (100 pmol)	1
Reverse primer (100 pmol)	1
Template*	2
Molecular grade dH_2O*	21
Final reaction volume	50

*For negative controls mol. grade dH_2O was increased in place of the template

araA and *xylAB* candidate was amplified using the reagents shown in A, *araBD* candidates were amplified using the reagents shown in B. + includes ATP, cytosine triphosphate (CTP), guanine triphosphate (GTP) and thyamine triphosphate (TTP).

2.9.4 Plasmid restriction analysis

Webcutter 2.0 http://rna.lundberg.gu.se/cutter2/ was used to find a suitable restriction enzyme. The restriction enzymes and the reagents used are shown in Table 2-5 A & B. All reaction mixtures were then incubated for 2 hr at 37°C and then 15 min at 65°C to inactivate the enzymes.

Table 2-5: Restriction enzymes and reagents used

(A)

pDNA	Enzyme	Excision sequence
pJW6 & 7 (CBEI^{+} 4457)	*Hind*III	A/AGCTT
pJW1 & 3(CBEI2383)		
pJW19 & pJW20(CBEI2384)		
pJW31 & 33(CBEI4452)	*Spe*I	A/CTAGT
pJW21 & 22(CBEI4455	*Xba*I	T/CTAGA

(B)

Reagent	Volume (µl)
Buffer	2.5
Plasmid DNA (pDNA)	5
Molecular grade dH_2O	16.5
Enzyme	1*
Total	25

* Extra Mol. grade dH_2O was added in place of enzyme for controls, restriction enzyme & buffer were supplied by Roche. $^{+}$ CBEI -*Clostridium beijerinckii* locus tag number

2.9.5 Gel Electrophoresis

PCR products (4 µl) were mixed with 1 µl of 5x DNA loading buffer (Bioline); the solution was loaded and run on a 1% (w/v) agarose gel (Bioline) with size standards shown in Appendix section Figure 7-1 . Gels were viewed and photographed with Biorad molecular imager® FX and Quantity One® software or Biorad Chemidoc XRS^{+} molecular imager and Image lab software.

2.9.6 DNA clean up

Promega Wizard SV gel and a PCR clean-up system was used to extract the PCR products from agarose gels. The products were visualised with a UV lamp and the relevant gel section was cut and dissolved. Membrane binding solution (10 µl) was added per 10 mg of gel slice. The mixture was vortexed and incubated (57°C, 10 min, periodically vortexed). The dissolved mixture was incubated (room temperature, 1 min) and added into a SV minicolumn and centrifuged (14000 rpm, 1 min). The eluent was discarded, membrane wash solution (700 µl) was added to the minicolumn and the column was centrifuged (1 min, 14000 rpm). The eluent was discarded, membrane wash solution (500 µl) was applied and the column was centrifuged (5 min, 14000 rpm).

The eluent was discarded and the column was centrifuged (14000 rpm, 1 min). The spin column was dispensed into a sterile 1.5 ml microcentrifuge tube and nuclease-free water (50 µl) was applied to the centre of the column. This was then incubated (1 min, room temperature) and centrifuged (1 min, 14000 rpm). The fraction that contained the DNA was stored (4°C).

2.9.7 TOPO-TA cloning

E. coli Top10 cells (Invitrogen) stored at –80°C were thawed slowly. DNA solution (4 µl), was missed with 1 µl Salt solution (1.2 M NaCl and 0.06 M $MgCl_2$) was mixed with a TOPO-TA® vector (1 µl) was added, the reaction was gently mixed and incubated (5 min) at room temperature. An aliquot (2 µl) was added to the thawed Top10 cells; the solution was gently mixed and returned to ice (15 min). The cells were heat shocked (42°C, 30 s) and returned to ice. A volume of 250 µl of SOC medium (consisting of 2% Tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose from Sigma) was added to the cells and the solution was incubated (1 hr) on a horizontal shaking incubator (200 rpm). LB plates that contained 50 µg/ml ampicillin were incubated in a WTB Binder incubator (Jencons Scientific LTD; 30 min, 37°C), 40 µl of a 40 mg/ml 5-bromo 4-chloro 3-indolyl β-D-galactopyranoside (X-gal) solution (Bioline) was spread on the surface of the agar and the plates were returned to the incubator. The transformation mixture (50, 100 or 150 µl) was spread on a plate and incubated overnight (37°C). Blue and white colonies were picked and inoculated into LB broth with 50 µg/ml ampicillin and grown overnight (37°C) in a horizontal shaking incubator (200 rpm).

2.9.8 Plasmid purification

Overnight cultures (that contained the pJW plasmid DNA) were used for plasmid extraction with a Purelink quick plasmid miniprep kit (Invitrogen). Overnight culture (1.5 ml) was harvested and centrifuged (10 min, 13000 rpm, Sanyo MSE microcentaur bench top centrifuge). The supernatant was discarded and the cell pellet was completely resuspended in 250 µl of resuspension buffer (20 mg/ ml RNase, 50 mM Tris-HCl pH8.0 and 10 mM EDTA). A volume of 250 µl Lysis buffer (200 mM NaOH and 1% w/v SDS) was added and the reaction mixture was incubated (5 min, room temperature). Precipitation buffer (350 µl) was added; the sample was inverted several times and centrifuged (13000 rpm, 10 min). The supernatant was loaded onto a spin column and centrifuged (13000 rpm, 1 min). The eluent was discarded, wash buffer W10 (500 µl) was added to the spin column, centrifuged (1 min, 13000 rpm) and the eluent discarded.

Wash buffer W9 (700 µl) was added to the spin column, centrifuged as previous, the eluent was discarded and the column was centrifuged once more to remove any residual wash buffer. The spin column was removed, placed into a sterile 1.5 ml microcentrifuge tube and Tris-EDTA (TE) buffer (75 µl, 67°C) was added to the centre of the spin column and centrifuged (13000 rpm, 1 min). The eluted DNA was stored at –20 °C.

2.9.9 The preparation of competent *E. coli* UP1091, XK100 and DS941 cells

A 5 ml volume from a 20 ml overnight culture was used to inoculate a fresh 200 ml LB broth in a 250 ml conical flask and grown at 37°C and 200 rpm on a horizontal incubator to mid-log phase (OD_{600} of 0.5–0.9). The culture was placed in water chilled in an ice bath for 30 min. Aliquots (45 ml) of the culture were added to 50 ml pre-chilled screw-capped vials and centrifuged. (4°C, 4000 rpm, 15 min, pre-chilled Heraeus multifuge 3L-R centrifuge). The pellet was resuspended in 45 ml dH_2O at 4°C and re-centrifuged. This procedure was repeated three times and the final pellet was resuspended in 10% (v/v) glycerol (5 ml) at 4°C and vortexed. Each of the pellets were fully resuspended in 10% (v/v) glycerol and combined in one tube. Glycerol (Sigma) 10% (v/v) was added (final vol. 45 ml). This was centrifuged as before. The supernatant was discarded and the pellet was resuspended in 10% (v/v) glycerol (400 µl), from which 40 µl aliquots were transferred to 1.5 ml cryovials and stored in a freezer at –80°C until use.

2.9.10 Transformation *E. coli* UP1091, XK100 and DS941

Competent cells (40 µl) were used for electroporation. Plasmid DNA (2 µl) was added, and incubated (1 min, on ice) and transferred into a cold 2 mm electroporation cuvette (the mixture completely covered the bottom of the cuvette). The cuvette was placed in the shockpod of a Biorad gene pulser Xcell™ electroporator and was pulsed once (5 milliseconds at 2481 V). The cuvette was immediately removed from the shockpod and 250 µl SOC (Sigma-Aldrich) was added and gently mixed. The cell suspension was then transferred to a 1.5 ml microcentrifuge tube and incubated (1 hr, 37°C, 200 rpm on a horizontal shaking incubator). A volume of 50 – 100 µl transformed cells was spread on LB plates with 50 µg/ µl ampicillin and incubated (overnight at 37°C).

2.10 Proteomics

2.10.1 Reagents

2.10.1.1 Proteomics wash buffer

The wash buffer contained 50 mM KH_2PO_4/K_2HPO_4, 5 mM $MgSO_4$ and 1 mM 1, 4-dithioerythritol (Fluka Biochemika). The phosphate solutions were made up as separate 50 mM stock solutions and added in different ratios to the rest of the ingredients in order to obtain a pH of 7.0. The solution was then filter-sterilised with a 0.2 μM nitrocellulose filter.

2.10.1.2 Bicinchoninic acid solution (BCA)

A 20 ml BCA solution was made by adding 0.4 ml of 4% (w/v) copper (II) sulphate solution (Sigma) to 19.6 ml bicinchoninic acid (Sigma).

2.10.1.3 Standard protein solution

A 1 mg/ml solution was made by gently dissolving Bovine Serum albumin (BSA) from Sigma into sterile molecular grade dH_2O to make a stock solution and serial dilutions were performed to obtain the desired BSA concentration.

2.10.1.4 Wash solution

Methanol (Sigma; 10 ml) was added to 5 ml of nanopure dH_2O, acetic acid (1 ml) was then added and the volume was adjusted to 20 ml with dH_2O.

2.10.1.5 Extraction buffer

ACN (Sigma; 10 ml) was added to 5 ml dH_2O, followed by 1 ml formic acid and the volume was adjusted to 20 ml with dH_2O.

2.10.2 Growth and harvesting of clostridia

Starter culture (1 ml) was inoculated into CBM (20 ml) with 1% (w/v) of either arabinose or xylose. The cultures were grown overnight and 5 ml was added to CBM (100 ml) supplemented with either 1% (w/v) arabinose or xylose. The cultures were grown overnight and 25 ml was used to inoculate CBM (500 ml) supplemented with 1% arabinose or xylose and grown for 14 hr.

The cultures were transferred into 50 ml centrifuge tubes and centrifuged (10 min, 12000 rpm) in a bench-top centrifuge (Eppendorf 5810R, Helena Biosciences). The supernatants were discarded and the pellets were combined, resuspended with wash

buffer and centrifuged (12000 rpm, 10 min). This was repeated three times. The supernatant was discarded and the pellets were frozen and stored at –20 °C.

2.10.3 Protein extraction

Pellets were slowly thawed on ice and fully resuspended in wash buffer with a glass rod. Wash buffer was added gradually in-between stirring to homogenise the cells to a final concentration of 4 ml/g of cell pellet.

Cells were broken in a French pressure cell press (GLM instruments LTD, AMINCO) with a pre-chilled pressure cell at 20000 pounds per inch squared (lb/ in^2). The samples were pressed twice. The cells were kept on ice when out of the press.

The samples were centrifuged (12000 rpm, 15 min, 4°C), the supernatant was extracted, 1 ml aliquots were frozen in liquid N_2 in 1.5 ml microcentrifuge tubes and stored at -80°C.

2.10.4 Bicinchoninic acid protein concentration assay

The assay was conducted in triplicate. An aliquot (10 µl) of a 0 – 1 mg/ml BSA solution was added to each well of 96-well TPP flat bottom test plate (Sigma Aldrich). Three dilutions of the unknown samples from *C. beijerinckii* (unknown sample: water, 1:1, 1:10, 1:100; 10 µl final vol.) were assayed as previously conducted with BSA. Bicinchoninic acid (BCA) solution (200 µl) was added and the plate was incubated (1 hr, 37°C; WTB Binder). The plate was read at 570 nm (Dynex MRX Revelation, Magellan Biosciences) and a standard curve of known concentrations (0, 0.1, 0.2, 0.4, 0.6, 0.8, 0.9, 1 mg/ml) of the standard protein (BSA) was constructed and an equation of the standard curve slope produced, so that the OD_{570} of the whole cell proteins of *C. beijerinckii* could be used to determine a protein concentration.

2.10.5 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Pre-cast Novex tris-glycine gels (gradient 4 – 20%; Invitrogen) were used. Each gel was rinsed with sterile dH_2O, the comb was removed and the wells were washed three times with 1x Tris-glycine SDS running buffer (Invitrogen).

Protein (5 µl of 1, 2, 5, 10 or 20 µg) was added to 5 µl Tris-glycine SDS (2x) sample buffer (Invitrogen), heated (85°C, 2 min) and 10 µl was loaded into each well. The loaded samples were separated by electrophoresis (125 volts (V), 90 min) alongside

protein markers (Colourburst electrophoresis marker 8–220 Dalton (Da), Sigma, Appendix section Figure 7-2).

2.10.6 Gel staining and destaining

The gels were rinsed three times with sterile dH_2O, placed in a plastic container with 20 ml SimplyBlue safe stain (Invitrogen) and left at room temperature for 1 hr on a rotating platform. The stain was discarded and the gel was rinsed with sterile dH_2O. Sterile dH_2O (100 ml) was added to the gel and agitated (1 hr). The gel was imaged on a Biorad Chemidoc XRS^+ molecular imager with Image lab software. The gels were stored in sterile dH_2O at 4°C.

2.10.7 Liquid chromatography electrospray ionisation tandem mass spectrometry (LC-ESI-MS/MS) analysis

All the steps of LC-ESI-MS/MS were carried out by staff at the Moredun Research Institute, with the exception of the data analysis.

2.10.7.1 Gel lane extraction and trypsin digestion

The 20-µg lane (section 2.9.4) was excised and was sliced horizontally into 28 slices (approximately 2.5 mm) and each slice was washed, reduced, alkylated and digested with trypsin. This procedure was carried out as part of the proteomic services at the Moredun Research Institute. The samples were transferred to a sealed HPLC vial and stored at 4°C.

2.10.7.2 LC-ESI-MS/MS

An Ultimate 3000 nano-HPLC system (Dionex) fitted with a WPS-3000 well-plate micro autosampler, FLM-3000 flow manager and column compartment, UVD-3000 UV detector, LPG-3600 dual-gradient micropump and SRD-3600 solvent rack was employed to perform liquid chromatography on the samples. The samples were separated on an i.d monolithic reverse-phase 5 cm x 200 µm column (Dionex-LC Packings) using solvent B (80% ACN, 0.1% (v/v) formic acid) mobile phase with a 8-45% linear gradient applied over 15 min at a final flow rate of 3 µl/min, achieved by combining a micro-pump flow rate of 246 µl/min and cap-flow splitter cartridge with 1/82 split. Peptides were eluted through a 3 nanolitre (nl) UV detector flow cell and fed into a stainless steel nebuliser (Agilent) with a maximum volume of 50 µl/min, and fed into a 3-D high capacity ion trap MS (Esquire HCTplus™, Bruker Daltonics). ESI-MS/MS analysis was applied upon the receiving of a contact closure signal from Chromeleon software that controlled the nano-HPLC system.

2.10.7.3 Data analysis

Onsite DataAnalysis™ software (Bruker Daltonics) was used to process the raw chromatography data and create MASCOT™ compatible files. MASCOT™ software, Matrix Science, (Perkins *et al.*, 1999) was then used to present and interpret the data with published guidelines (Taylor and Goodlett, 2005). A peptide and fragment mass tolerance of 1.5 and 0.5, were used for each respectively. Proteins were identified on the basis of at least two peptides being present. Molecular weight search (MOWSE) scores of 25–40 were manually inspected for two peptides with four continuous y or b ions. A MOWSE score above 25 is equivalent to $p<0.05$ significance. Scores below 25 were insignificant and disregarded.

Chapter 3

Pentose Sugar Utilisation by *C. beijerinckii*

3 Pentose Sugar Utilisation by *C. beijerinckii*

3.1 Introduction

Carbohydrates are a major source of carbon (a source of energy) and are essential to the survival of all heterotrophic organisms. Saccharolytic bacteria including clostridia are able to utilise a wide variety of sugars. A number of bacteria show a preference for glucose over other carbon sources. A classic example is glucose preference over lactose in *E.coli* (Inada *et al.*, 1996). In clostridia glucose is used in preference to maltose, sucrose, lactose and xylose in *C. acetobutylicum*, and over glucitol and sucrose in *C. beijerinckii* (Tangney *et al.*, 2001, Tangney and Mitchell, 2000, Yu *et al.*, 2007, Tangney *et al.*, 1998, Reid *et al.*, 1999, Mitchell *et al.*, 1995). *C. acetobutylicum* is also capable of utilising both arabinose and xylose (Ounine *et al.*, 1983). A preference for glucose over xylose in mixed sugar media containing 0.6% (w/v) is shown regardless of which sugar the cultures are established in (Ounine *et al.*, 1985). However, in mixed xylose-glucose sugar media containing 7.2% (w/v), xylose-established cultures of *C. acetobutylicum* utilise xylose simultaneously with glucose, as long as xylose is in excess in the culture medium (Fond *et al.*, 1985a). Little research has been carried out into the ability of *C. beijerinckii* to grow on pentose sugars or the effect mixed glucose-pentose sugar media has on sugar use preference. One study, carried out on a 6% (w/v) total sugar medium containing a ratio of 5:4:2:1 of glucose, xylose, arabinose and mannose, observed the simultaneous use of all sugars by cultures of *C. acetobutylicum*, *C. beijerinckii*, *C. butylicum* and *C. saccharolyticum* that were established on glucose. Although simultaneous utilisation was observed, the sugars were used at different rates depending on the strain, but glucose was used more rapidly than the other sugars across all the species investigated (Ezeji *et al.*, 2007). Carbon catabolite repression (CCR) exerted by glucose functions to prevent unnecessary expenditure of cellular resources for the metabolism of other carbon sources, which may not be present in the organism's environment and allows expenditure only when glucose, a ubiquitous carbon source, is not available. Sometimes diauxie or 'double growth' is evident on growth curves with mixed carbon source media, whereby an organism will grow on the preferred carbon source until it is exhausted. The growth then arrests, as the organism takes time to synthesise necessary components for the metabolism of the second carbon source. Then once equipped, the cultures will continue to grow again. CCR has potential impacts on the utilisation of all carbon sources in mixed sugar waste streams for biofuel production,

since least preferred sources will be ignored or not wholly utilised and therefore wasted as the fermentation reaches the end point.

Therefore, the aims were to establish if *C. beijerinckii* can utilise pentose sugars arabinose and xylose for solvent production. To assess the effect of glucose on the utilisation of these pentose sugars and then finally to assess the competence of *C. beijerinckii* to use xylan and a pentose sugar-rich waste stream for biofuel production.

3.2 The Utilisation of Pentose Sugars on Minimal Media

3.2.1 Arabinose utilisation

In order to investigate the ability of *C. beijerinckii* to utilise arabinose as a sole carbon source and whether utilisation is subject to CCR in the presence of glucose, spores of *C. beijerinckii* revived and cultures of *C. beijerinckii* were established on glucose or arabinose (by subculturing on CBM with 1% (w/v) glucose or arabinose for two consecutive days) and were inoculated in 100ml of CBM containing 1% (w/v) glucose (positive control), 1% (w/v) arabinose or 0.5% (w/v) of both sugars, or no sugar (negative control). Aliquots (5 ml) were taken from each culture and OD_{650} and the pH was measured at several time-points over 24 hours (Figure 3-1A-C for the glucose-established cultures and Figure 3-2A-C for the arabinose-established cultures of *C. beijerinckii*). This was monitored for 24 hours. Samples were taken at regular intervals up to 12 hours as this period show any CCR, by way of the preferred sugar being utilised and the other remaining unused. A further sample was taken at 24 hours to ascertain if the other carbon source was still utilised after the preferred one has been exhausted. Samples were also taken, stored and analysed as described in Section 2.5 for sugar concentration analysis by HPLC.

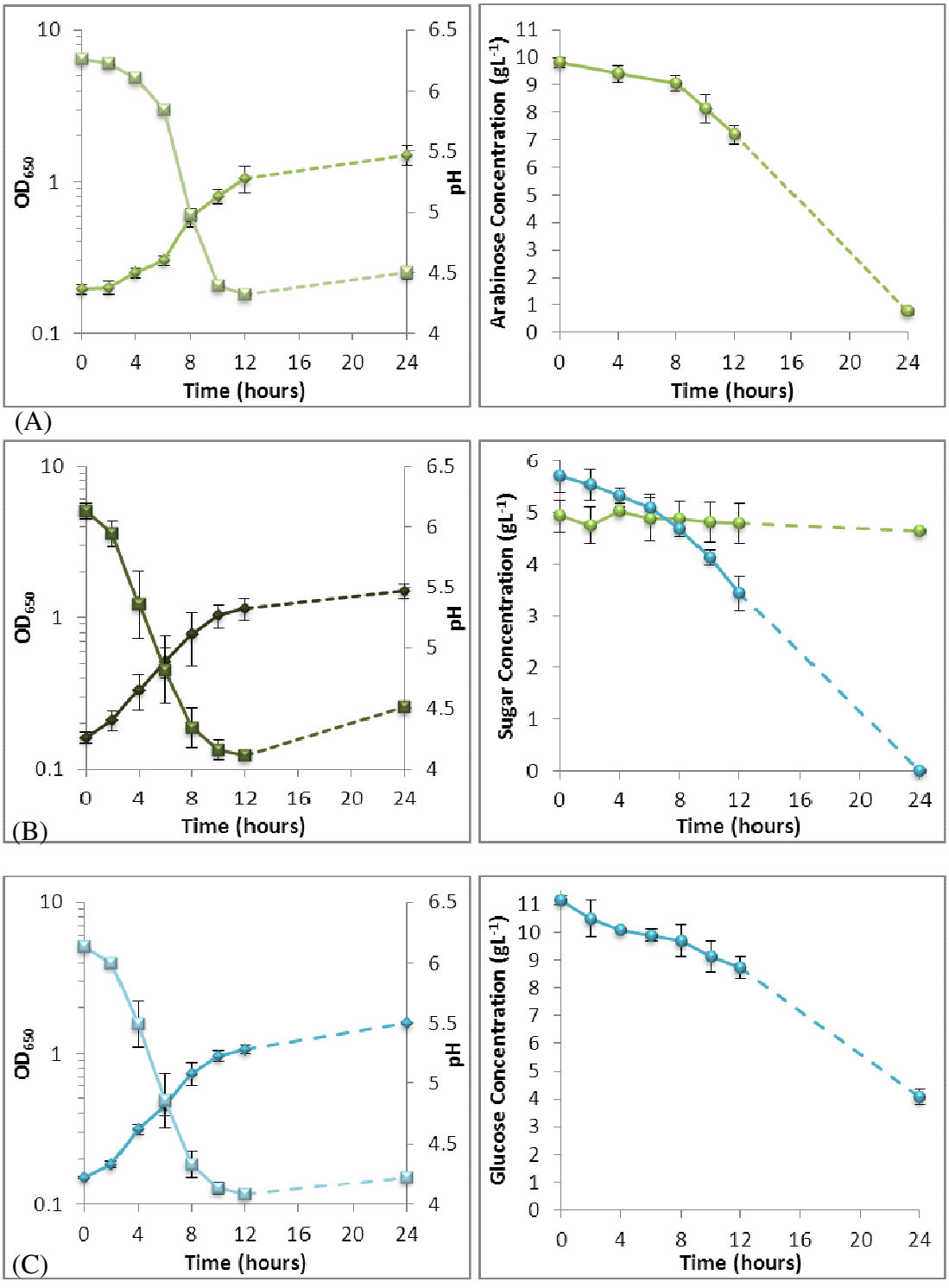


Figure 3-1: Utilisation of glucose and arabinose by glucose-established cultures of *C. beijerinckii*

C. beijerinckii cultures were established on glucose and inoculated into 100 ml CBM supplemented with 1% total concentration of either (A) arabinose; (B) arabinose and glucose; or (C) glucose. OD_{650} on log scale (♦); pH(■); glucose concentration (•); arabinose concentration (•). The average of triplicates was plotted with the standard deviation bars denoting the range between the triplicates.

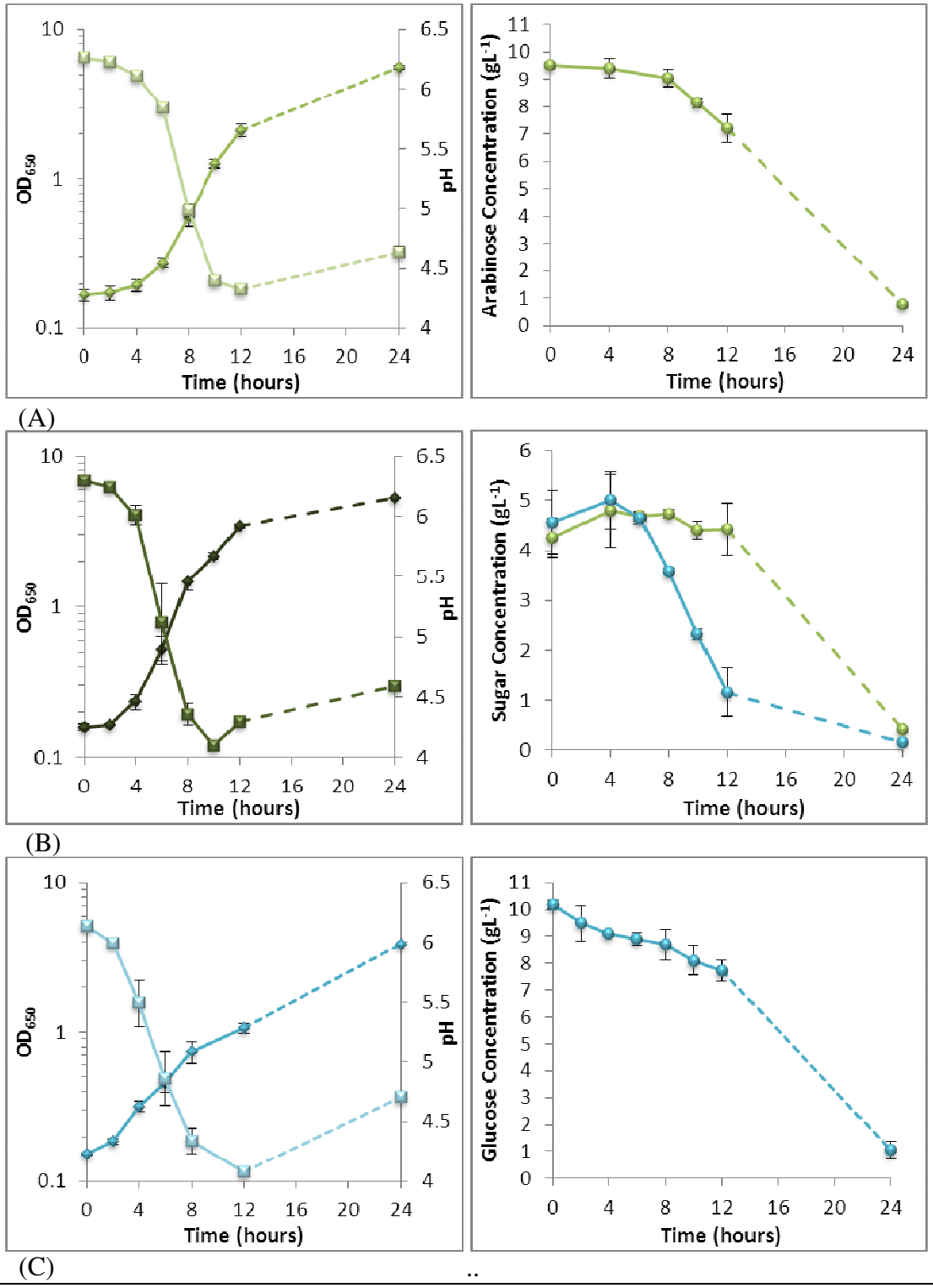


Figure 3-2: Utilisation of glucose and arabinose by arabinose-established cultures of *C. beijerinckii*

C. beijerinckii cultures were established on arabinose and inoculated into 100 ml CBM containing a total concentration of 1% of either (A) arabinose; (B) arabinose and glucose; or (C) glucose. OD_{650} on log scale (♦); pH(■); glucose concentration(●); arabinose concentration(●). The average of triplicates was plotted with the standard deviation bars denoting the range between the triplicates.

C. beijerinckii was unable to grow on CBM without the addition of a sugar (data not shown). *C. beijerinckii* was able to grow well on arabinose and in a comparable fashion to glucose (Figure 3-1A and C). Arabinose was consumed steadily and almost to exhaustion in 24 hours. In the presence of both sugars glucose was used preferentially over arabinose, the concentration of arabinose remaining stable over the course of the fermentation (Figure 3-1B). It is not known whether arabinose was consumed after 24 hours from this particular experiment. To establish if the same effect is a consequence of cultures being established on glucose prior to the experiment, the same experiment was repeated but cultures were established on arabinose prior to the experiment. It was evident in mixed sugar cultures that the concentration of arabinose remained constant over the first 12 hours of the fermentation as glucose was preferentially consumed (Figure 3-2 A-C). At some point during 12–24 hours of the fermentation, arabinose was consumed to almost exhaustion. Such a finding suggests a mechanism of repression of arabinose use in the presence of glucose.

3.2.2 Xylose utilisation

In order to investigate the ability of *C. beijerinckii* to grow on and utilise xylose as a sole carbon source, and whether this utilisation is subject to any CCR in the presence of glucose such as that observed with arabinose, cultures of *C. beijerinckii* were established on either glucose or xylose and experiments conducted in the same manner as in section 3.2.1. *C. beijerinckii* growing on xylose were observed as growing well and in a comparable fashion to glucose, however the utilisation of xylose was poor with cultures established on glucose (Figure 3-3A compared to 3-3C and 3-4A and 3-4C). However, xylose appeared to be poorly utilised with more than half the sugar remaining after 24 hours (Figure 3-3A). In the presence of both sugars, glucose was used preferentially by *C. beijerinckii* cultures and the concentration of xylose remained stable over the course of the first half of the fermentation, decreasing in the latter half to near exhaustion (Figure 3-3B). To establish if the finding of preferential use of glucose over xylose was a consequence of cultures being established on glucose prior to the experiment, another experiment was conducted with cultures established on xylose. As observed with arabinose-established cultures growing in the presence of both glucose and a pentose sugar, a preferential use of glucose was evident on mixed glucose-xylose sugar media (Figure 3-4B). This finding therefore suggests there is also a mechanism of repression of xylose utilisation.

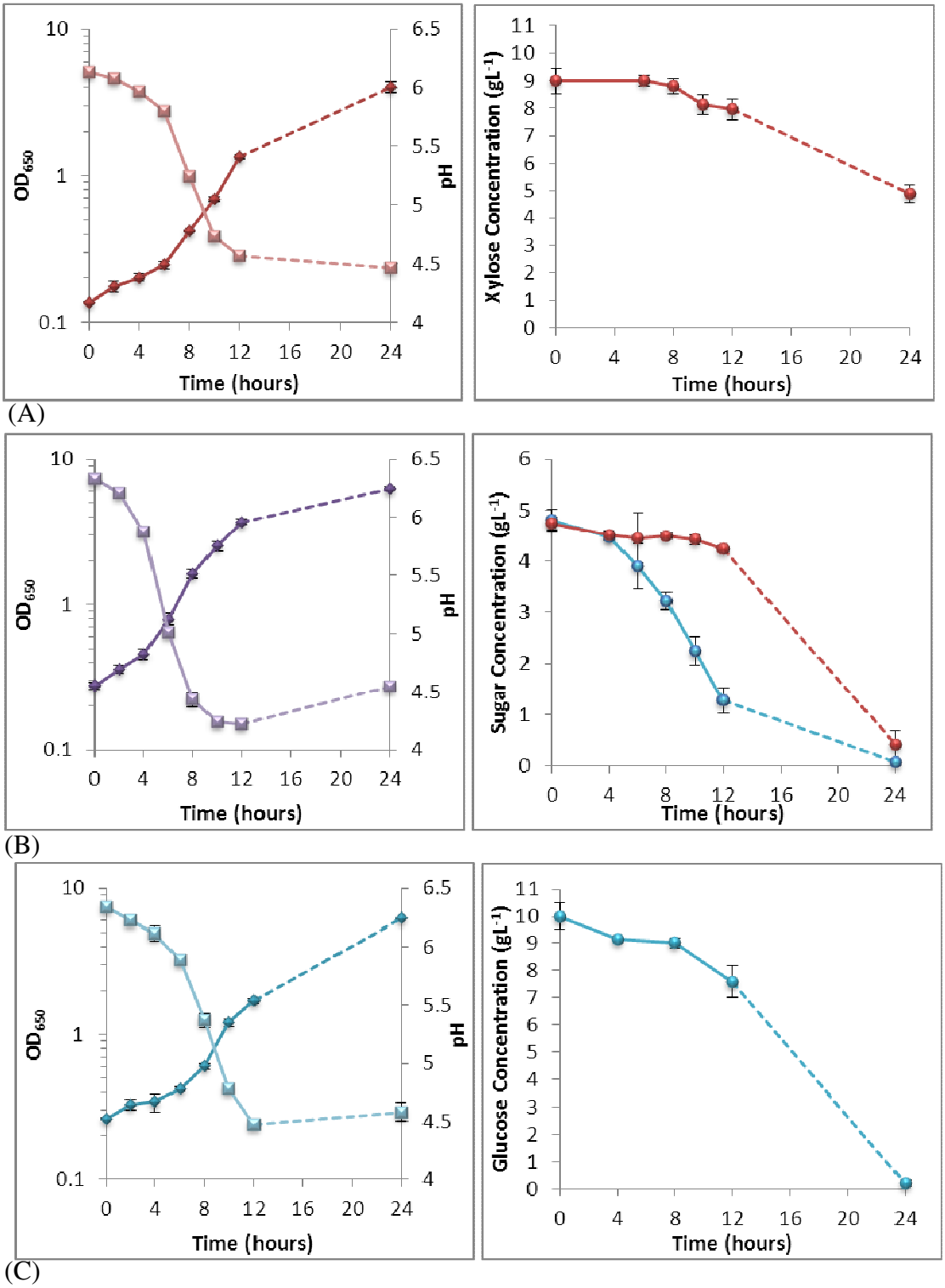


Figure 3-3: Utilisation of glucose and xylose by glucose-established cultures of *C. beijerinckii*

C. beijerinckii cultures were established on glucose and inoculated into 100 ml CBM containing 1% of either (A) xylose; (B) xylose and glucose; or (C) glucose. OD_{650} on log scale (♦); pH (■); glucose concentration (•); xylose concentration (•) The average of triplicates was plotted with the standard deviation bars denoting the range between the triplicates.

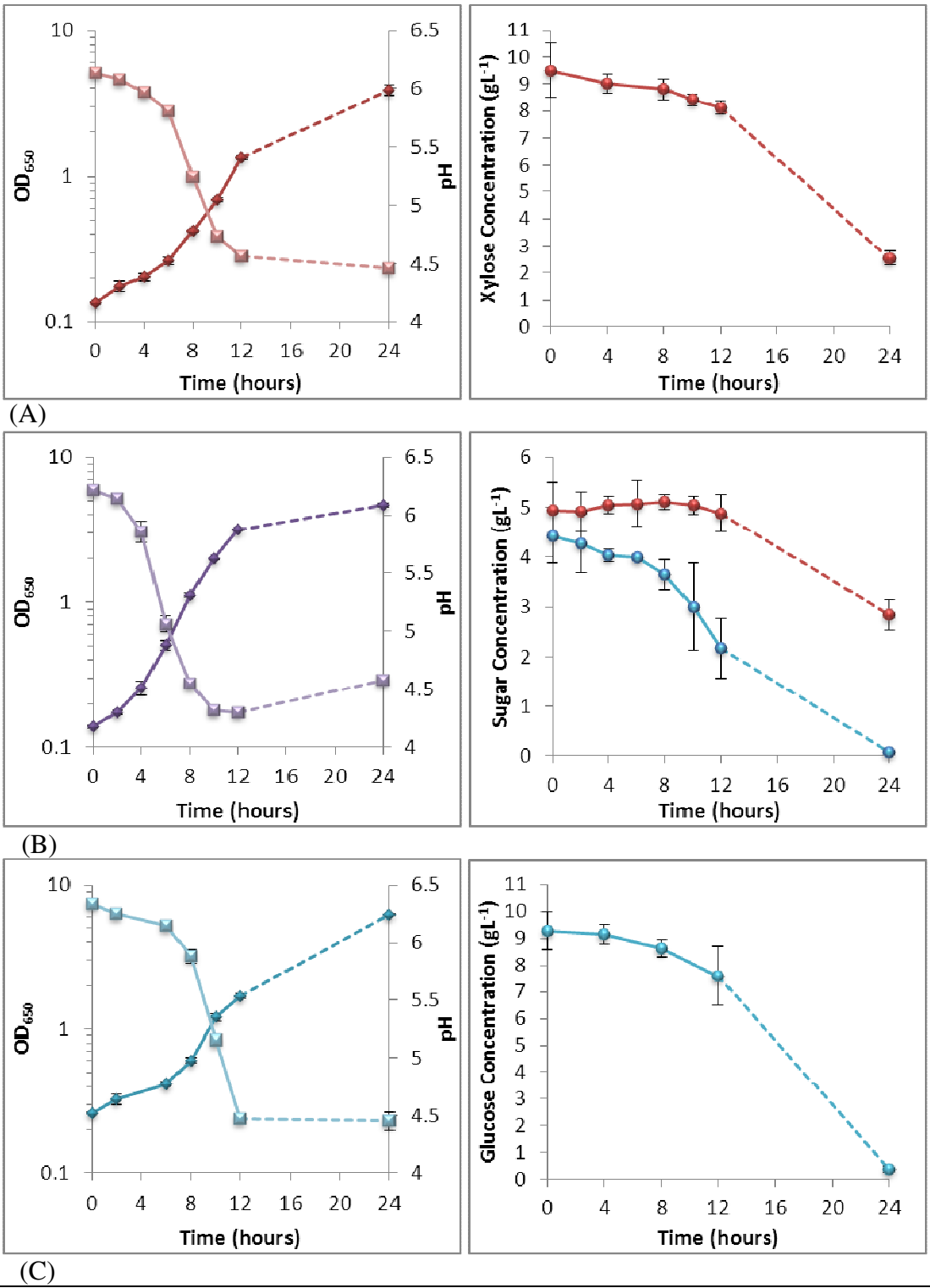


Figure 3-4: Utilisation of glucose and xylose by xylose-established cultures of *C. beijerinckii*

C. beijerinckii cultures were established on xylose and inoculated into 100 ml CBM containing 1% of either (A) xylose; (B) xylose and glucose; or (C) glucose. OD_{650} on log scale (♦); pH (■); glucose concentration (•); xylose concentration (•). The average of triplicates was plotted with the standard deviation bars denoting the range between the triplicates.

3.3 Effect of High Sugar Concentrations on Pentose Sugar Utilisation

In the previous section, regardless of the carbon source cultures of *C. beijerinckii* are established on, glucose is used preferentially over either pentose sugar on a 1% (w/v) total sugar concentration. It is not until glucose has been exhausted or approaching exhaustion that the pentose sugars are used. This has also been observed in *C. acetobutylicum* (Ounine *et al.*, 1985). On high sugar concentrations of 7.2% (w/v) *C. acetobutylicum* has been observed using xylose and glucose simultaneously, if xylose is in excess in the culture medium (Fond *et al.*, 1985a, b). To establish whether the same result would be evident for *C. beijerinckii* and if the inoculum preparation and the ratio of pentose to glucose sugar on the use of pentose sugars has an effect in mixed sugar fermentations, a series of experiments were employed. A longer fermentation time and longer intervals between samples was used to accommodate the time taken for cultures to utilise a higher sugar load, in comparison to the previous experiments.

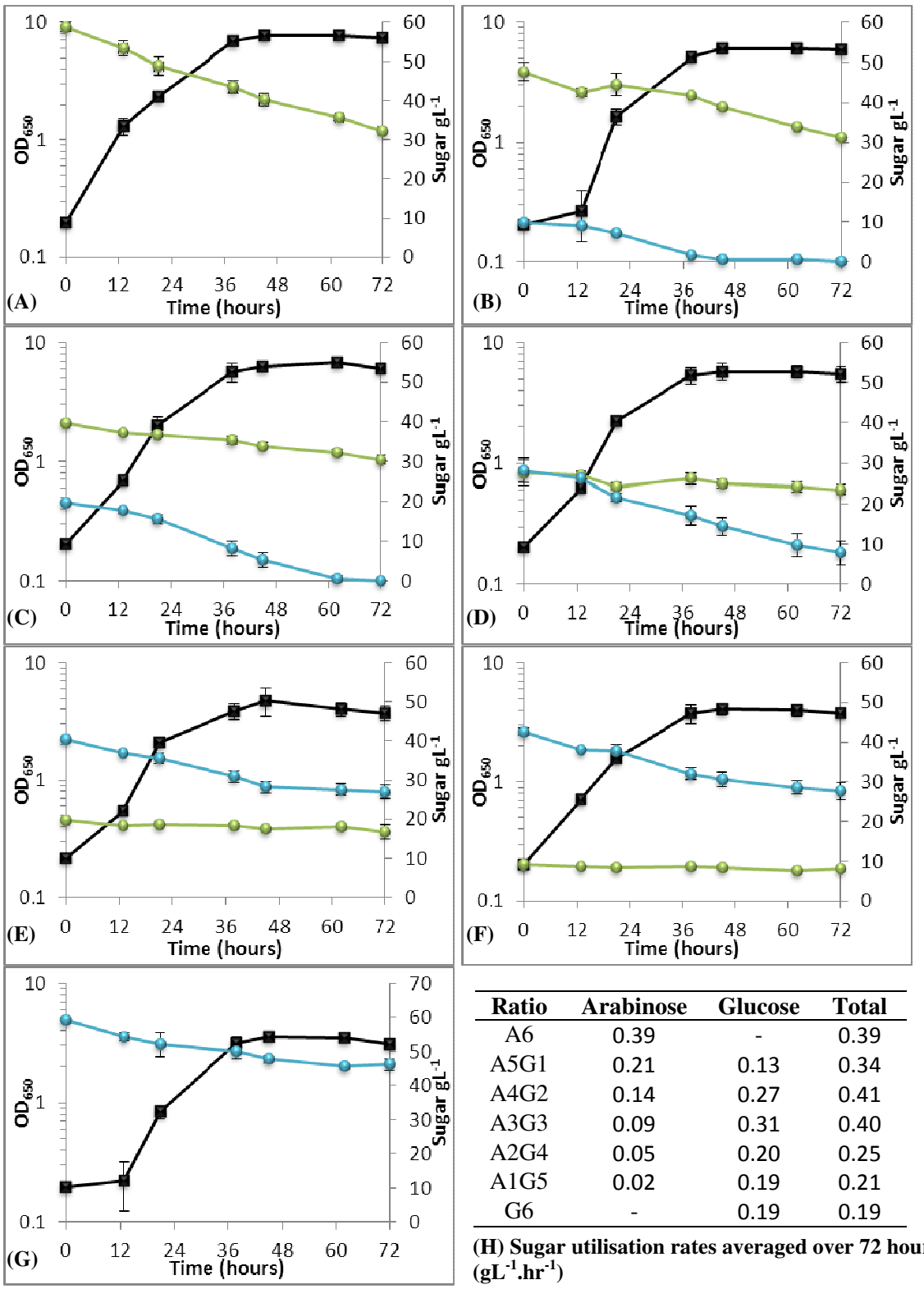
3.3.1 Arabinose

Cultures of *C. beijerinckii* established on either arabinose or glucose prior to the experiments. Despite cultures growing well on 6% (w/v) of arabinose or glucose, arabinose-established cultures (Figure 3-5G) utilised less glucose over the course of the fermentation, than the glucose-established cultures (Figure 3-5H and Figure 3-6H). In contrast the utilisation rates of arabinose were similar (Figure 3-5A & 3-6A) regardless of the carbon source the cultures were established on, though there was a slightly faster utilisation rate for cultures established on arabinose (Figure 3-5H and Figure 3-6H).

On media containing equal ratios of glucose to arabinose, more than double the amount of glucose over arabinose was utilised over 72 hours, regardless of the carbon source that cultures were established on (Figure 3-5D and Figure 3-6D). The rate of utilisation of arabinose was considerably less than that of glucose (Figures 3-5H and 3-6H).

When cultures were grown on media with an excess of glucose, the arabinose was barely used over the course of the fermentation (Figure 3-5 EF and Figure 3-6EF) and the utilisation rates reflect this (Figure 3-5H and Figure 3-6H). In cultures where arabinose was in excess at a ratio of 2:1 (A4G2), a larger amount of arabinose was depleted from the culture medium in comparison to the same ratio where glucose was in excess (Figure 3-5C in comparison to 3-5E; Figure 3-6C in comparison to 3-6E). On the A4G2 ratio glucose utilisation was twice as fast as that of arabinose despite the amount of arabinose present being double that of glucose (Figure 3-5H and Figure 3-6H).

In cultures where arabinose was in excess to glucose by a ratio of 5:1, the rate of utilisation of arabinose was faster than that of glucose (Figure 3-5H and Figure 3-6H).

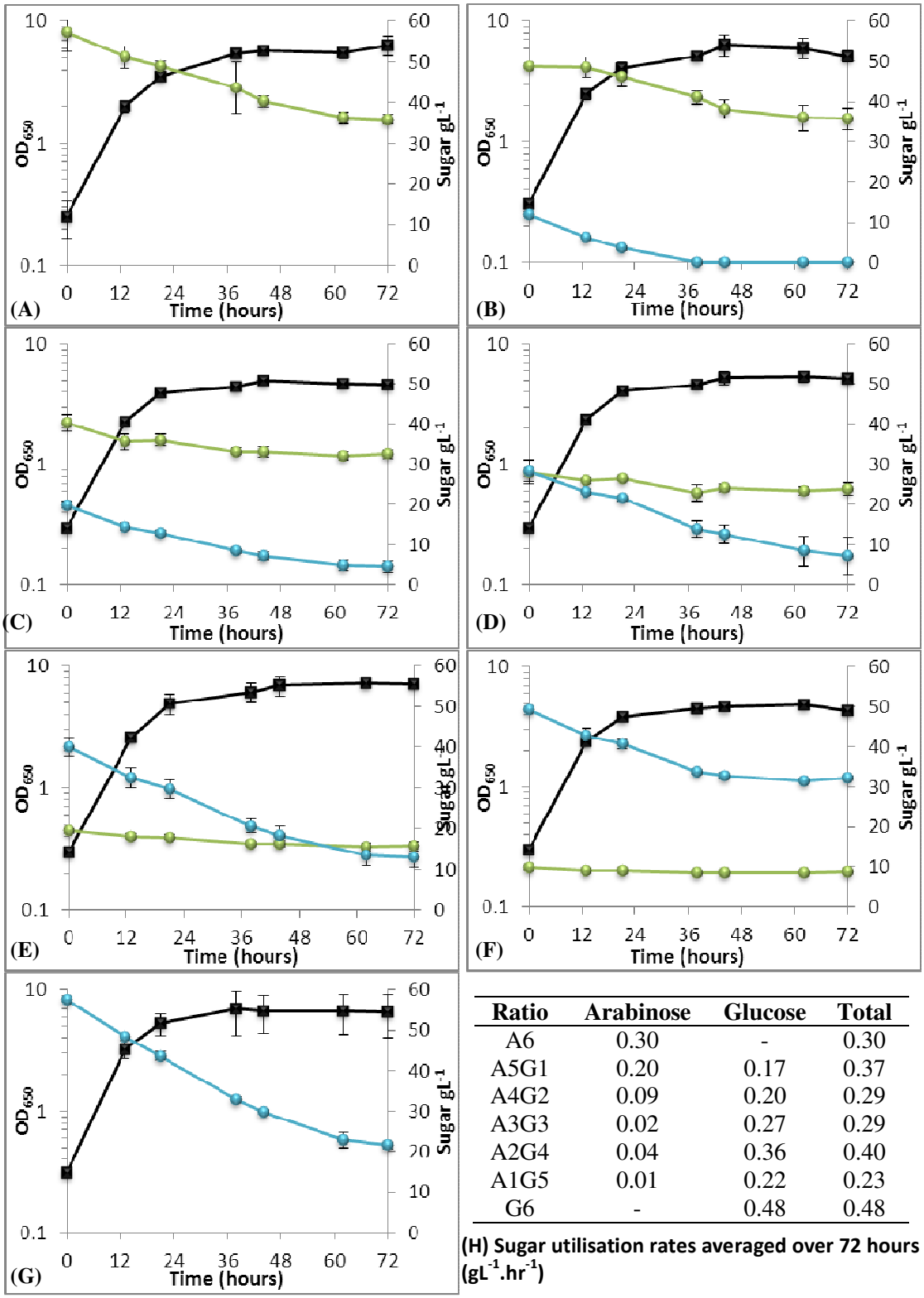


Ratio	Arabinose	Glucose	Total
A6	0.39	-	0.39
A5G1	0.21	0.13	0.34
A4G2	0.14	0.27	0.41
A3G3	0.09	0.31	0.40
A2G4	0.05	0.20	0.25
A1G5	0.02	0.19	0.21
G6	-	0.19	0.19

(H) Sugar utilisation rates averaged over 72 hours ($gL^{-1}.hr^{-1}$)

Figure 3-5: Effect of high sugar concentrations on glucose-arabinose utilisation

Cultures of *C. beijerinckii* were established on TYA supplemented with arabinose and inoculated into triplicate 100 ml volumes of TYA medium with either **(A)** 6% arabinose **(B)** 5% arabinose: 1% glucose **(C)** 4% arabinose: 2% glucose **(D)** 3% arabinose: 3% glucose **(E)** 2% arabinose: 4% glucose **(F)** 1% arabinose: 5% glucose **(G)** 6% glucose **(H)** sugar utilisation rates averaged over 72 hrs ($gL^{-1}.hr^{-1}$), **A** – Arabinose, **G**- glucose, number is the % of sugar in the medium. (•) arabinose (•) glucose concentrations, (■) OD_{650} on log scale. The average of triplicates was plotted with the standard deviation bars denoting the range between the triplicates. The OD_{650} is plotted on a log scale.



Ratio	Arabinose	Glucose	Total
A6	0.30	-	0.30
A5G1	0.20	0.17	0.37
A4G2	0.09	0.20	0.29
A3G3	0.02	0.27	0.29
A2G4	0.04	0.36	0.40
A1G5	0.01	0.22	0.23
G6	-	0.48	0.48

(H) Sugar utilisation rates averaged over 72 hours ($gL^{-1}.hr^{-1}$)

Figure 3-6: Effect of high sugar concentrations on glucose-arabinose utilisation
Cultures of *C. beijerinckii* were established on TYA supplemented with glucose and inoculated into triplicate 100 ml volumes of TYA medium with either **(A)** 6% arabinose **(B)** 5% arabinose: 1% glucose **(C)** 4% arabinose: 2% glucose **(D)** 3% arabinose: 3% glucose **(E)** 2% arabinose: 4% glucose **(F)** 1% arabinose: 5% glucose **(G)** 6% glucose **(H)** sugar utilisation rates averaged over 72 hrs ($gL^{-1}.hr^{-1}$), **A** – Arabinose, **G**- glucose, number is the % of sugar in the medium. (•) arabinose (•) glucose concentrations, (■) OD_{650} on log scale. The average of triplicates was plotted with the standard deviation bars denoting the range between the triplicates. The OD_{650} is plotted on a log scale.

3.3.2 Xylose

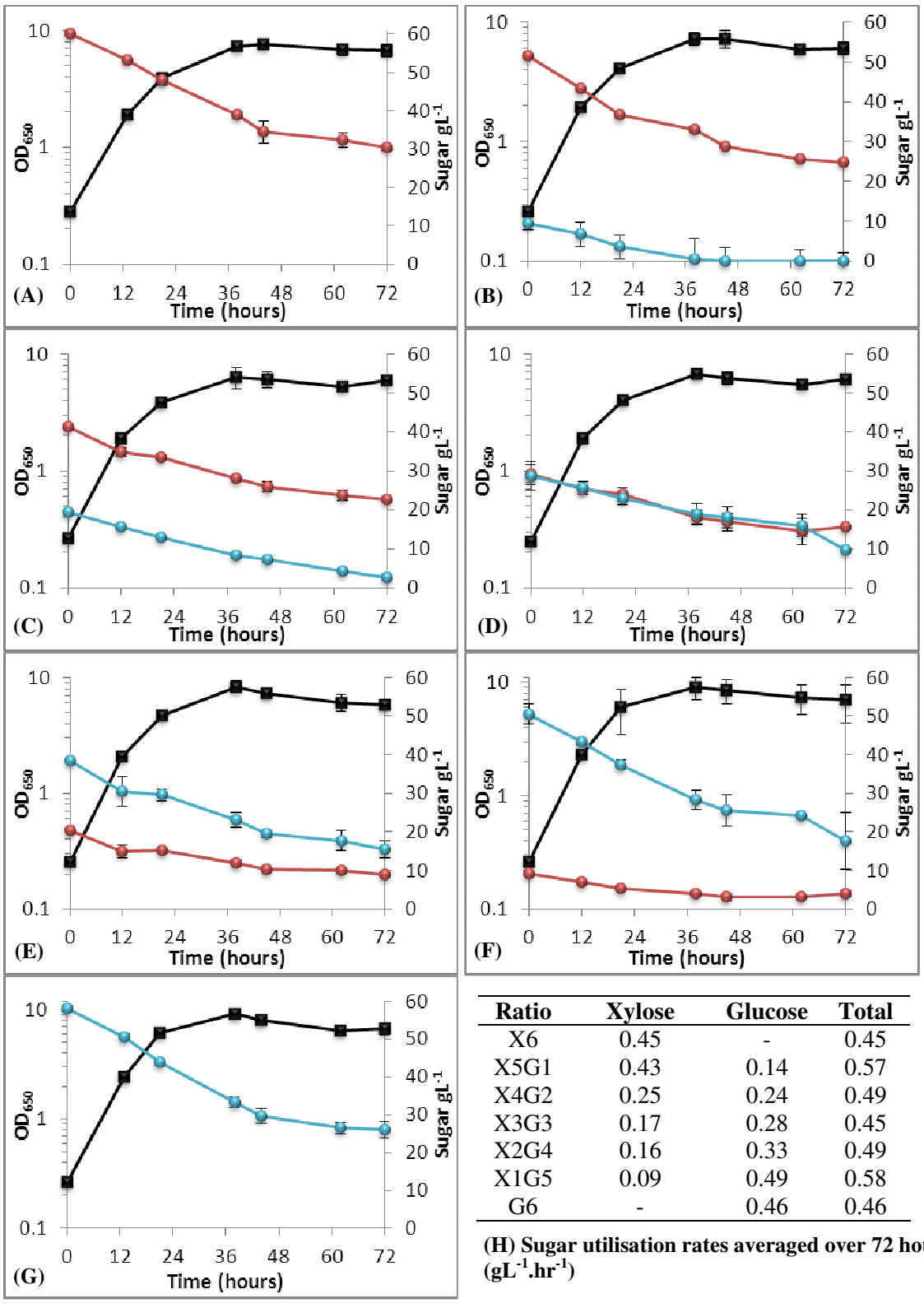
Cultures of *C. beijerinckii* established on either xylose or glucose grew well on 6% (w/v) of either sugar. More xylose was depleted during the course of the fermentation by cultures established on glucose (Figure 3-7A and Figure 3-8A) and a higher rate of xylose utilisation was also evident (Figure 3-7H and Figure 3-8H).

Cultures growing on equal ratios of glucose to xylose, regardless of the carbon source established on, xylose was utilised faster and the difference between the rate of glucose use and pentose sugar use was considerably less than that of glucose and arabinose (Figure 3-7H, Figure 3-8H, Figure 3-5H and Figure 3-6H). The depletion of glucose and xylose is very comparable over the duration of the fermentation whichever carbon source the cultures were established on (Figures 3-7D and 3-8D), especially with xylose-established cultures (Figure 3-7D).

The cultures grown on media with an excess of glucose utilised xylose slower than glucose (Figure 3-7EF and Figure 3-8EF). At a ratio of 2:1 (X2G4), the utilisation rate of glucose was twice as fast as that of xylose (Figure 3-7H and Figure 3-8H), which is in contrast to the arabinose-glucose experiments where the rate difference was considerably larger. A similar observation was noted where the ratio was 5:1 (X1G5), however the rate of xylose utilisation was faster than the equivalent on arabinose (Figure 3-7H, Figure 3-8H, Figure 3-5H and Figure 3-6H).

A larger amount of xylose was depleted from the culture medium where xylose was in excess at a ratio of 2:1 (X4G2), in comparison to cultures established on glucose (Figure 3-7C and Figure 3-8C). The rate of xylose utilisation was very similar between the two (Figure 3-7H and Figure 3-8H). However the glucose was nearly fully depleted at the fermentation end point for xylose established cultures (Figure 3-7C).

In cultures where xylose was in excess to glucose by a ratio of 5:1 (X5G1), the rate of utilisation of xylose was faster than that of glucose (Figure 3-7H and Figure 3-8H), in the arabinose-glucose experiments a much smaller difference in comparison.

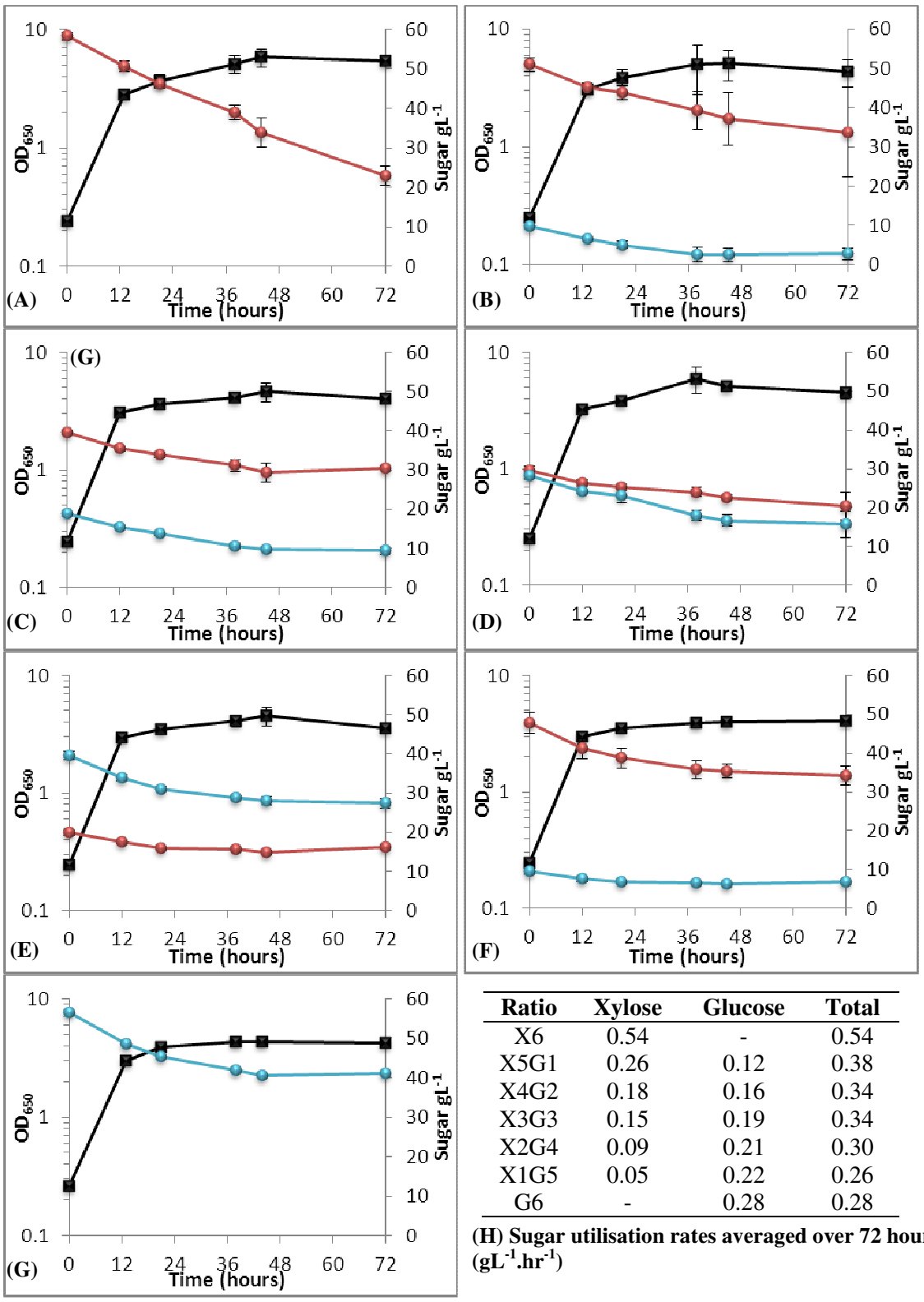


Ratio	Xylose	Glucose	Total
X6	0.45	-	0.45
X5G1	0.43	0.14	0.57
X4G2	0.25	0.24	0.49
X3G3	0.17	0.28	0.45
X2G4	0.16	0.33	0.49
X1G5	0.09	0.49	0.58
G6	-	0.46	0.46

(H) Sugar utilisation rates averaged over 72 hours ($gL^{-1}.hr^{-1}$)

Figure 3-7: Effect of high sugar concentrations on glucose-xylose utilisation

Cultures of *C. beijerinckii* were established on TYA supplemented with xylose and inoculated into triplicate 100 ml volumes of TYA medium with either **(A)** 6% xylose **(B)** 5% xylose: 1% glucose **(C)** 4% xylose: 2% glucose **(D)** 3% xylose: 3% glucose **(E)** 2% xylose: 4% glucose **(F)** 1% xylose: 5% glucose **(G)** 6% glucose **(H)** sugar utilisation rates averaged over 72 hours ($gL^{-1}.hr^{-1}$), **X** – xylose, **G**- glucose, number is the % of sugar in the medium. (•) xylose (•) glucose concentrations, (■) OD_{650} on log scale. The average of triplicates was plotted with the standard deviation bars denoting the range between the triplicates. The OD_{650} is plotted on a log scale.



Ratio	Xylose	Glucose	Total
X6	0.54	-	0.54
X5G1	0.26	0.12	0.38
X4G2	0.18	0.16	0.34
X3G3	0.15	0.19	0.34
X2G4	0.09	0.21	0.30
X1G5	0.05	0.22	0.26
G6	-	0.28	0.28

(H) Sugar utilisation rates averaged over 72 hours ($gL^{-1}.hr^{-1}$)

Figure 3-8: Effect of high sugar concentrations on glucose-xylose utilisation

Cultures of *C. beijerinckii* were established on TYA supplemented with glucose and inoculated into triplicate 100 ml volumes of TYA medium with either **(A)** 6% xylose **(B)** 5% xylose: 1% glucose **(C)** 4% xylose: 2% glucose **(D)** 3% xylose: 3% glucose **(E)** 2% xylose: 4% glucose **(F)** 1% xylose: 5% glucose **(G)** 6% glucose **(H)** sugar utilisation rates averaged over 72 hrs ($gL^{-1}.hr^{-1}$), **X** – xylose, **G**- glucose, number is the % of sugar in the medium. (•) xylose (•) glucose concentrations, (■) OD_{650} on log scale. The average of triplicates was plotted with the standard deviation bars denoting the range between the triplicates. The OD_{650} is plotted on a log scale.

3.4 Solvent Production by *C. beijerinckii* from Pentose Sugars

Next, in order to ascertain what level of solvents could be produced by *C. beijerinckii* cultures grown on pentose sugars, cultures were established on glucose and these were used to inoculate two 1 Litre fermenter vessels, which were ran in parallel, with TYA supplemented with either glucose and arabinose or glucose and xylose (6% w/v). The cultures were monitored for 72 hours and samples taken periodically. Fermenters were used as the agitation they provide allows temperature and nutrient circulation and preventing toxic metabolites pooling. This mixing effect provides the best conditions for culture growth and therefore solvent production.

There was a notable difference in the degree of sugar utilisation by cultures of *C. beijerinckii* depending on the carbon source. Cultures growing on glucose had half the sugar remaining compared to either pentose sugar by the end of the 72 hour fermentation (Figure 3-9 and Table 3-1). The rate of sugar utilisation was more rapid, with a quicker butanol and total solvent production rate, in the first 62 hours of the fermentation on glucose. In the latter 10 hours, the utilisation rate of arabinose was slightly faster in comparison to glucose and xylose, hence the butanol and total solvent production rates were also increased (Figure 3-9A and Table 3-1).

Approximately three-fold more acids were present at the fermentation end-point on both pentose sugars than on glucose (Table 3-1), although it cannot be ascertained from this experiment whether or not the extra acids at the fermentation endpoint would have eventually been converted to solvents if a longer fermentation time was employed.

The solvent levels produced by cultures on arabinose were akin to those grown on xylose (Table 3-1). The ratio of acetone to butanol produced varied depending on the carbon source that the cultures were grown on. Glucose-grown cultures produced a higher acetone to butanol ratio in line with expected ratios, whereas cultures grown on the pentose sugars produced a lower acetone to butanol ratio. This contributed to the higher total solvent production observed by cultures grown on glucose was due to the higher acetone levels produced (Table 3-1). The butanol yield was similar between all three sugars. Cultures grown on xylose had a slightly higher butanol yield of 0.27 g.g in comparison to 0.25 g.g on arabinose and 0.26 g.g on glucose (Table 3-1). Although less xylose was utilised, a greater proportion was converted to butanol.

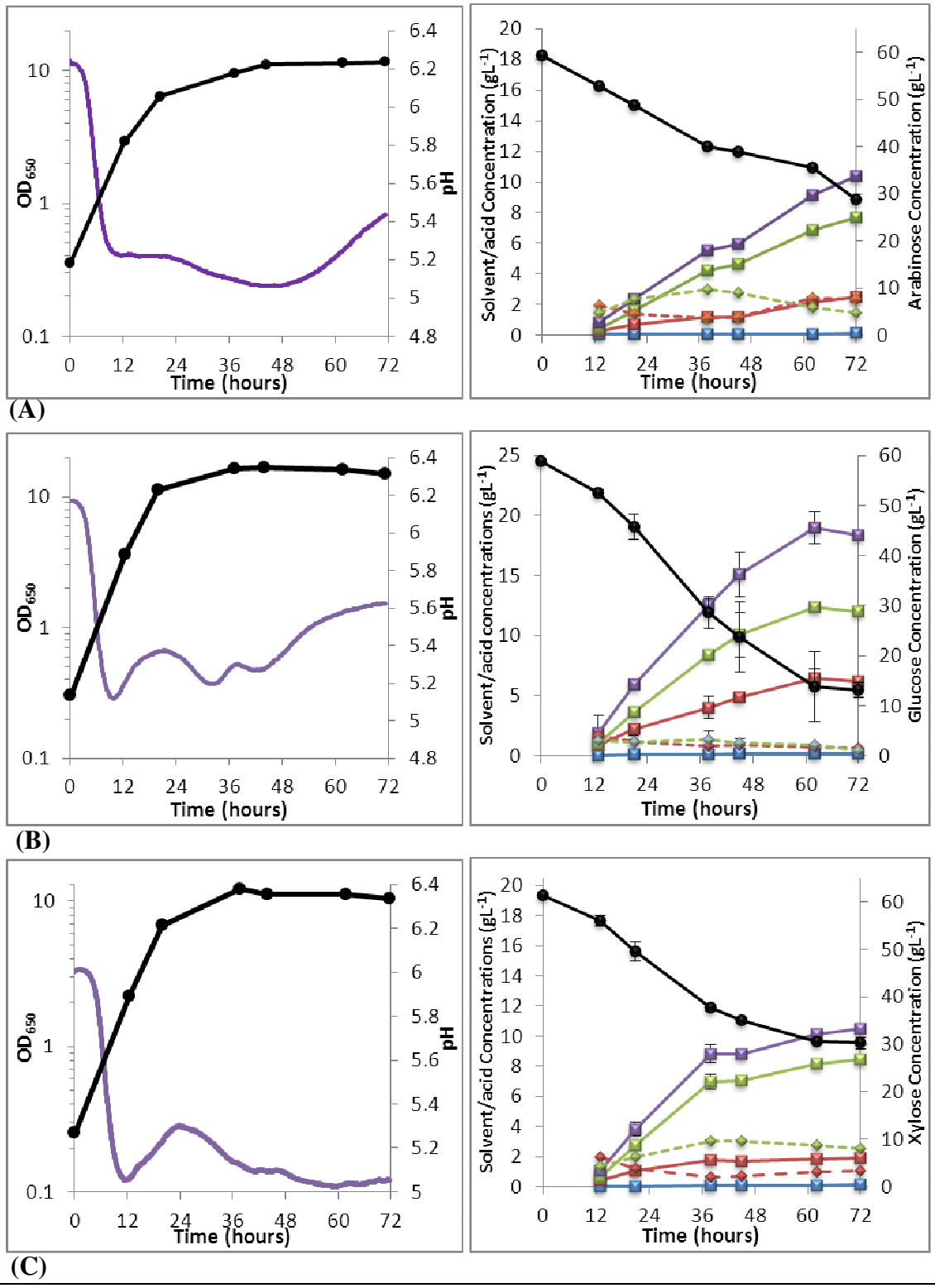


Figure 3-9: Fermentation of glucose and pentose sugars

C. beijerinckii was established on TYA supplemented with glucose and inoculated in 1L fermenters with TYA medium supplemented with 6% of either (A) Arabinose (B) Glucose (C) Xylose. Over a 72 hour fermentation Graph on the left: (•) OD_{650}, (—) pH, Graph on the right: (•) Sugar; Acids (dashed line): (♦) acetate, (♦) butyrate; (■) acetone, (■) butanol, (■) ethanol and (■) total solvents concentrations were monitored periodically over 72 hr. The average of triplicates was plotted with the standard deviation bars denoting the range between the triplicates.

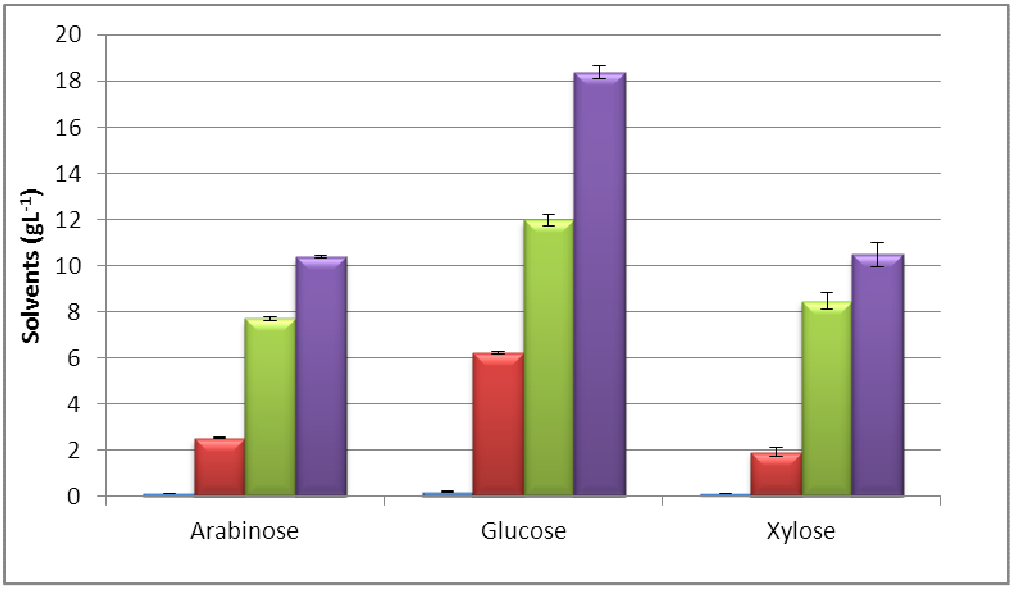


Figure 3-10: Solvent production on glucose and pentose sugars

Solvent end products after 72 hr fermentation (■) ethanol, (■) acetone, (■) butanol, (■) total ABE on TYA supplemented with 6% sugar. The average of triplicates was plotted with the standard deviation bars denoting the range between the triplicates.

Table 3-1: Sugar consumption and solvent production on glucose and pentose sugars

		Arabinose	Glucose	Xylose
Sugar start	(gL^{-1})	59.27	58.41	61.40
Sugar end	(gL^{-1})	28.70	13.17	30.23
Sugar used	**(%)**	**52**	**77**	**51**
Acetate	(gL^{-1})	2.38	0.63	1.03
Butyrate	(gL^{-1})	1.44	0.49	2.57
Total Acids	**(gL^{-1})**	**3.82**	**1.12**	**3.60**
Acetone	(gL^{-1})	2.54	6.21	1.92
Butanol	(gL^{-1})	7.71	11.97	8.46
Ethanol	(gL^{-1})	0.14	0.20	0.11
Total Solvents	**(gL^{-1})**	**10.38**	**18.39**	**10.49**
Acetone:Butanol ratio		1:3	1:2	1:4
Butanol yield	g.g	0.25	0.26	0.27
Solvent yield	**g.g**	**0.34**	**0.41**	**0.34**
Sugar utilisation rate	($gL^{-1}.hr^{-1}$)*	0.42	0.63	0.43
Butanol production rate	($gL^{-1}.hr^{-1}$)*	0.11	0.17	0.12
Solvent production rate	($gL^{-1}.hr^{-1}$)*	0.14	0.26	0.15

* Averaged over 72 hours

3.5 The Utilisation of Xylan and Xylan Hydrolysates by *C. beijerinckii*

It has been established *C. beijerinckii* can utilise xylose as a carbon source for growth. However, the majority of xylose found naturally is in the form of the polysaccharide xylan. Xylans are the second most abundant polymer in nature after cellulose, making up a third of all renewable organic carbon (Collins *et al.*, 2005), it is of interest to determine whether *C. beijerinckii* can utilise xylans for biofuel production.

Xylan is a highly branched heteropolymer. The exact constituents and arrangement of its building blocks are source-specific, but the basic structure is a xylose polymer backbone with side-chains of xylose and other sugars such as galactose, mannose, fucose, glucuronate, rhamnose and arabinose. Xylan is found in most abundance in hardwoods, where it constitutes up to 30% of the cell wall contents. It may also be present in softwoods, grasses and annual plants.

Using commercially available xylan from two different sources (beechwood and birchwood), small-scale fermentations were carried to ascertain the ability of *C. beijerinckii* to utilise xylan and xylan-hydrolysates in parallel to xylose utilisation.

3.5.1 Kinetic Investigation of Activity of Commercial Xylanase on Xylans

In order to ascertain the time required for commercial xylanase (Novozymes) to breakdown commercial xylans from beechwood and birchwood (Sigma), an investigation of the kinetic activity of a commercial xylanase on the two types of xylans was conducted. The hydrolysis of both types of xylans by xylanase was evident within 30 minutes with little difference in the profile of hydrolysates including xylose after six hours of enzymatic treatment (Figure 3-11). Fewer larger xylan hydrolysates, found closest to the origin, after 24 hours of treatment. The effect of autoclaving xylan was also assessed ahead of any fermentations with *C. beijerinckii*, to ensure any breakdown of xylan was not an effect of heat and only attributable to enzymes. Autoclaving did not cause any degradation of either of the xylans used, as can be seen by the lack of visible products above the origin on the TLC plate (Figure 3-11 A & B).

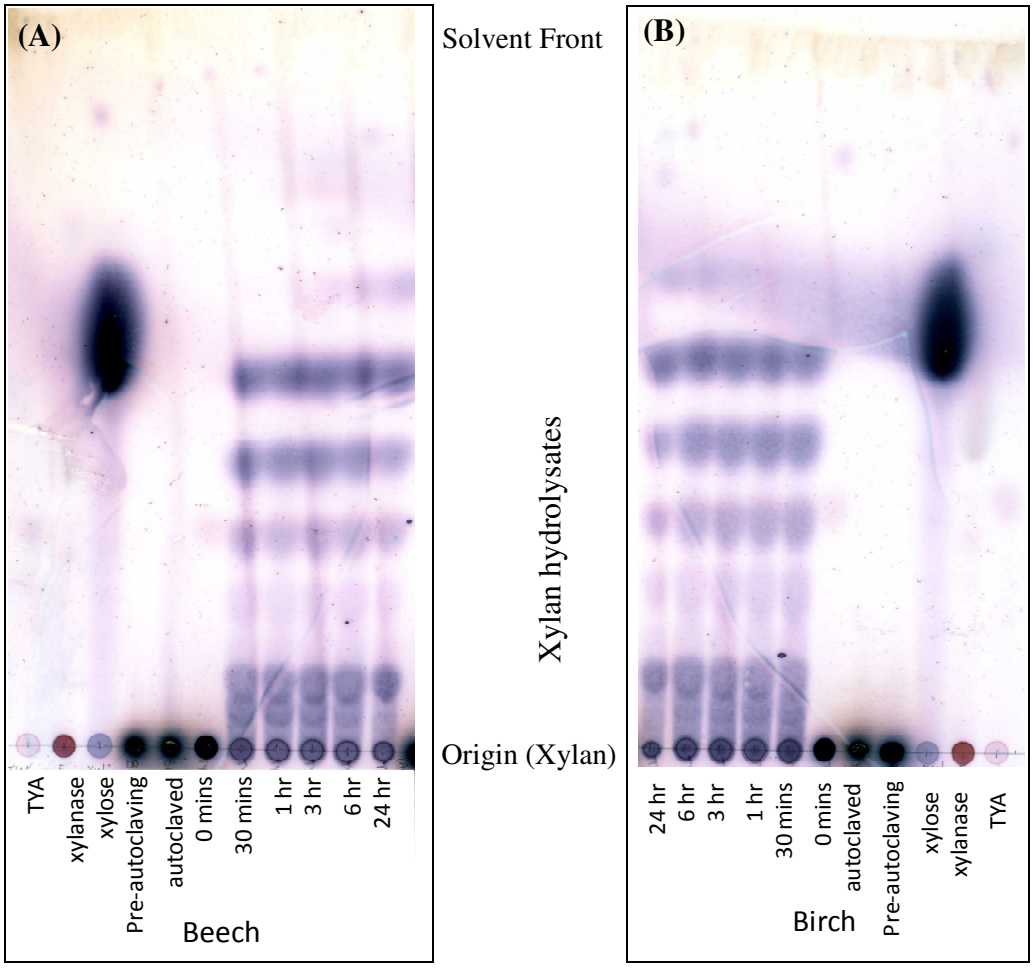


Figure 3-11: Xylanase time treatment of xylan

TYA medium containing 1% (w/v) xylan was autoclaved to assess its effect on xylan degradation then treated with xylanase over a period of 24 hours and samples were taken at various time intervals and boiled to inactivate the xylanase. Samples were then subject to TLC after loading on to TLC plate. Xylan appears at the line of origin, smaller constituents, down to the monomer units (xylose) and a xylose control are found further up the TLC plate **(A)** Beechwood xylan **(B)** Birchwood xylan

3.5.2 The ability of *C. beijerinckii* to utilise xylan and xylan hydrolysates

Different solutions were employed to assess the ability of *C. beijerinckii* to utilise xylans, and xylan hydrolysates; TYA media containing 1 % (w/v) xylan, xylan hydrolysates (xylanase-treated xylan for 24 hours prior to inoculation), xylan supplemented with 0.1% (w/v) xylose, as this was found to boost the utilisation of xylan in *C. acetobutylicum* (Lee *et al.*, 1985b); and xylose. TLC was carried out as previously described on samples from the culture media prior to inoculation and at the end of fermentation (96 hours). *C. beijerinckii* was found to be able to grow on and utilise both xylans and the xylan hydrolysates, as indicated by the pH profiles consistent with

acidogenesis, where the pH is observed as falling steadily over the first 24 hours of the fermentation on all substrates (Figure 3-12A and Figure 3-12B). Figure 3-13 A and B shows the breakdown of beech and birch wood xylan by the appearance of smaller carbohydrate units at 96 hours. Xylan hydrolysates were used during the course of the fermentation as indicated by the disappearance of 'spots' from the profile at 96 hours. *C. beijerinckii* was unable to grow on TYA without an added carbon source (data not shown). The addition of 0.1% (w/v) xylose had no observable effect on the utilisation of either xylans, as indicated by the same breakdown profile as that of xylan after 96 hours (Figure 3-13). The pH reached a plateau on all of the substrates after acidogenesis, at 24 hours and remained stable, failing to rise, indicative of a poor level of solventogenesis occurring (Figure 3-14A). This result was further reflected in the high levels of acid and solvent at the fermentation endpoint (Figure 3-14B). The best solvent production occurred with the xylose control, followed by the pre-enzymatic treatment of xylans. A difference in the solvent profiles of cultures grown on xylans and those grown on xylan hydrolysates and xylose was observed. The ratios of acetone to butanol were almost equal to those found from cultures grown on xylans, whereas the other substrates produced a more commonly observed solvent profile, with a 2:1 ratio of acetone: butanol, or higher.

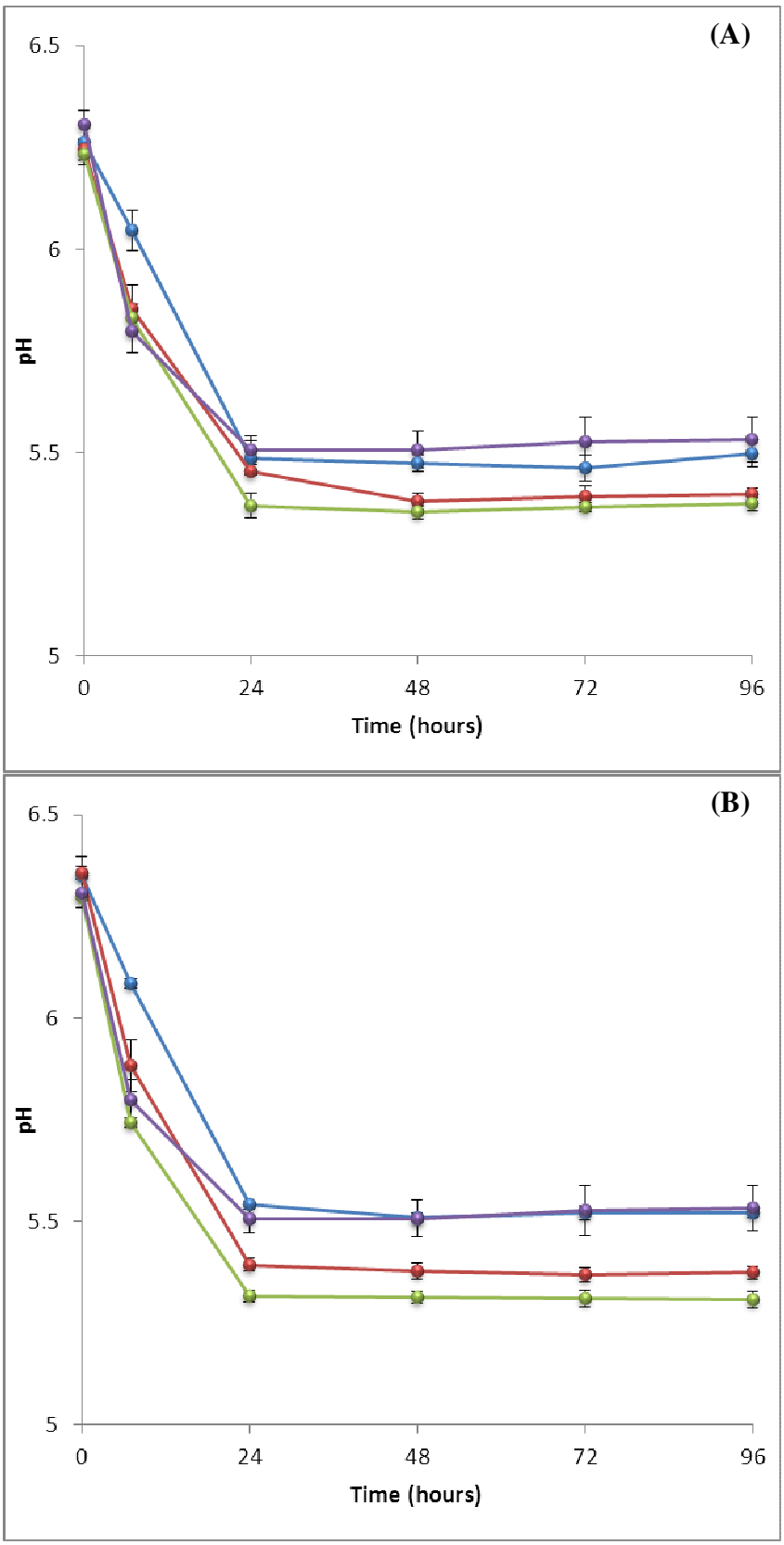


Figure 3-12: pH profile of *C. beijerinckii* cultures grown on 1% (w/v) xylan

Cultures were established on xylose and inoculated into TYA with 1% of **(A)** Beechwood xylan **(B)** Birchwood xylan (•) Untreated xylan (•) Xylan supplemented with 0.1% xylose (•) Pre xylanase-treated xylan (•) Xylose. The error bars denote the differences between triplicates in standard deviation.

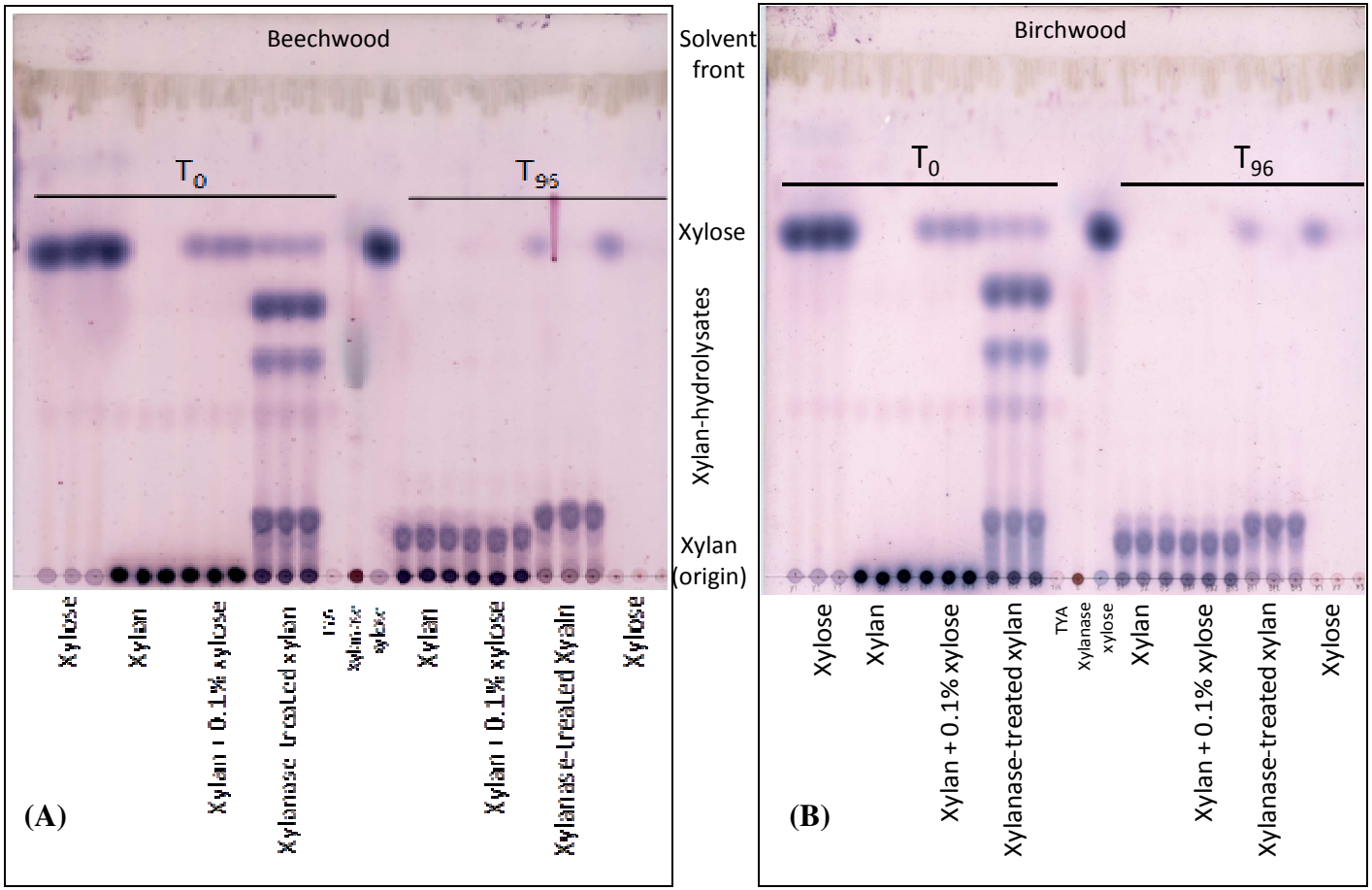


Figure 3-13: Xylan utilisation by *C. beijerinckii*

TLC analysis of the utilisation of **(A)** Beechwood xylan **(B)** Birchwood xylan, by *C. beijerinckii*. T_0 indicates time zero prior to inoculation and T_{96} after 96 hour fermentation. Cultures were inoculated into triplicates of TYA with 1% xylose, 1% xylan, 1% xylan supplemented with 0.1% (w/v) xylose, or 1% xylan pre-treated with xylanase.

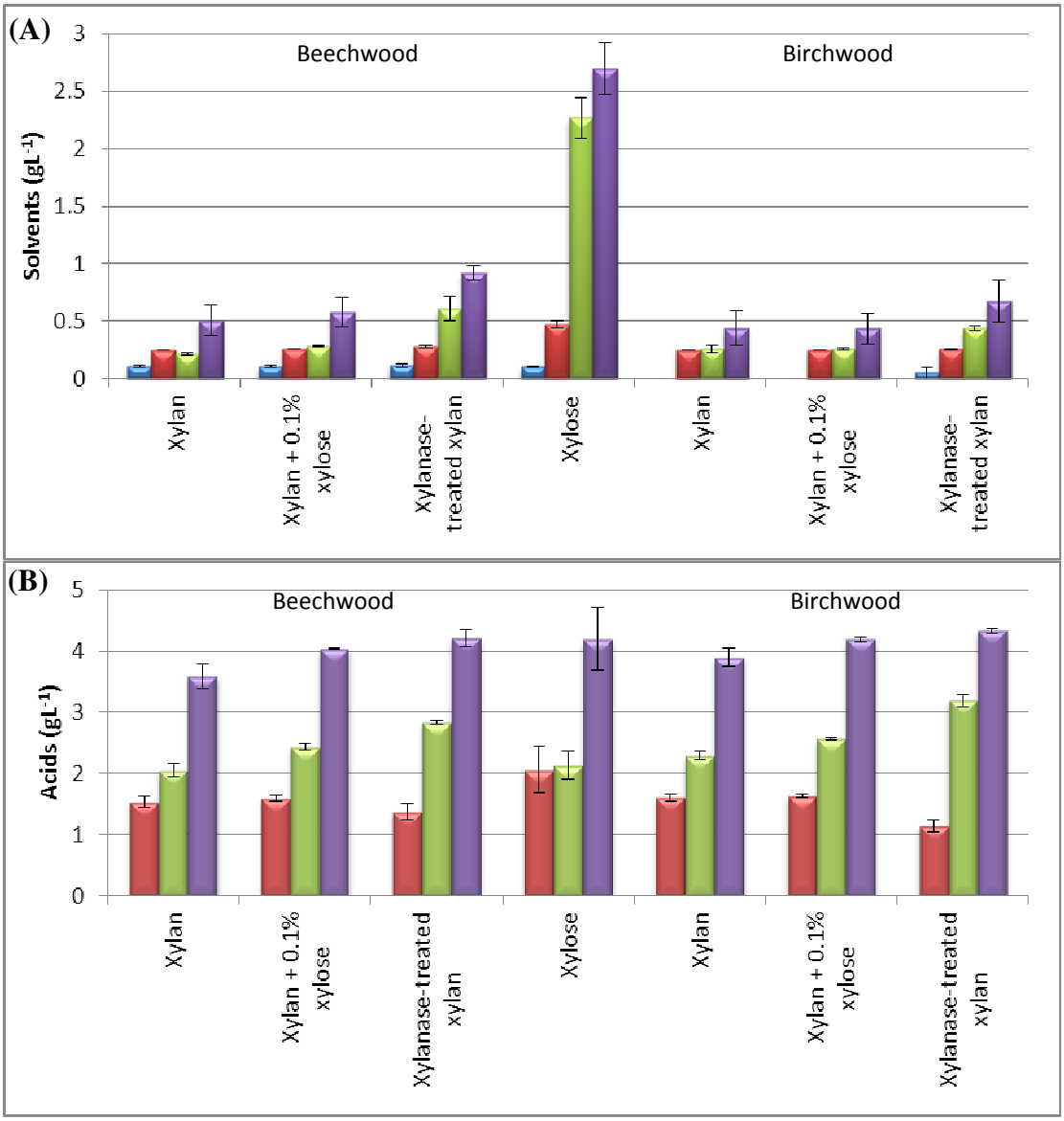


Figure 3-14: Solvent and acid production of *C. beijerinckii* grown on 1% xylan

Cultures were established on xylose and inoculated into 1% of xylan or xylose **(A)** Solvents at 96 hours: (■) Ethanol (■) Acetone (■) Butanol (■) Total ABE **(B)** Acids at 96 hours: (■)Acetate (■)Butyrate (■) Total acids. The error bars denote the differences between triplicates in standard deviation.

3.5.3 Effect of Higher Xylan Concentrations on Solvent Production

One possible reason for the poor solvent production observed in the previous experiment could be that there was an insufficient level of utilisable carbon source remaining in the culture medium to allow any level of sustained solventogenesis. Therefore, the experiment was repeated with a 5% (w/v) concentration of carbon sources, so a plentiful supply was available to allow solventogenesis. Despite a five-fold increase in carbon source, very little difference to the pH profiles of *C. beijerinckii* was observed. A rise in pH, consistent with solventogenesis, was evident only in cultures grown on xylose (Figure 3-15). The pH of the cultures on all the other carbon sources used fell lower towards the fermentation endpoint in comparison to 1% (w/v) concentration and no plateau in the pH level after 24 hours was evident. Consistent with more acid production, in comparison to the 1% (w/v) concentration experiment, thus poor solvent production was still observed (Figure 3-16). Although the concentration of xylan or xylan hydrolysates used was five-fold greater than that used in the previous experiment, only twice the amount of solvent was produced for cultures grown on xylan and around 1.5 times the amount for those grown on xylan supplemented with xylose and the pre-enzymatically treated xylan, whereas the cultures grown on xylose produced ten-fold more solvents compared to cultures grown on xylan (Figure 3-14A). The level of acids at the fermentation endpoint was very similar across all the substrates used (Figure 3-16B) and were only marginally more than in the previous experiments.

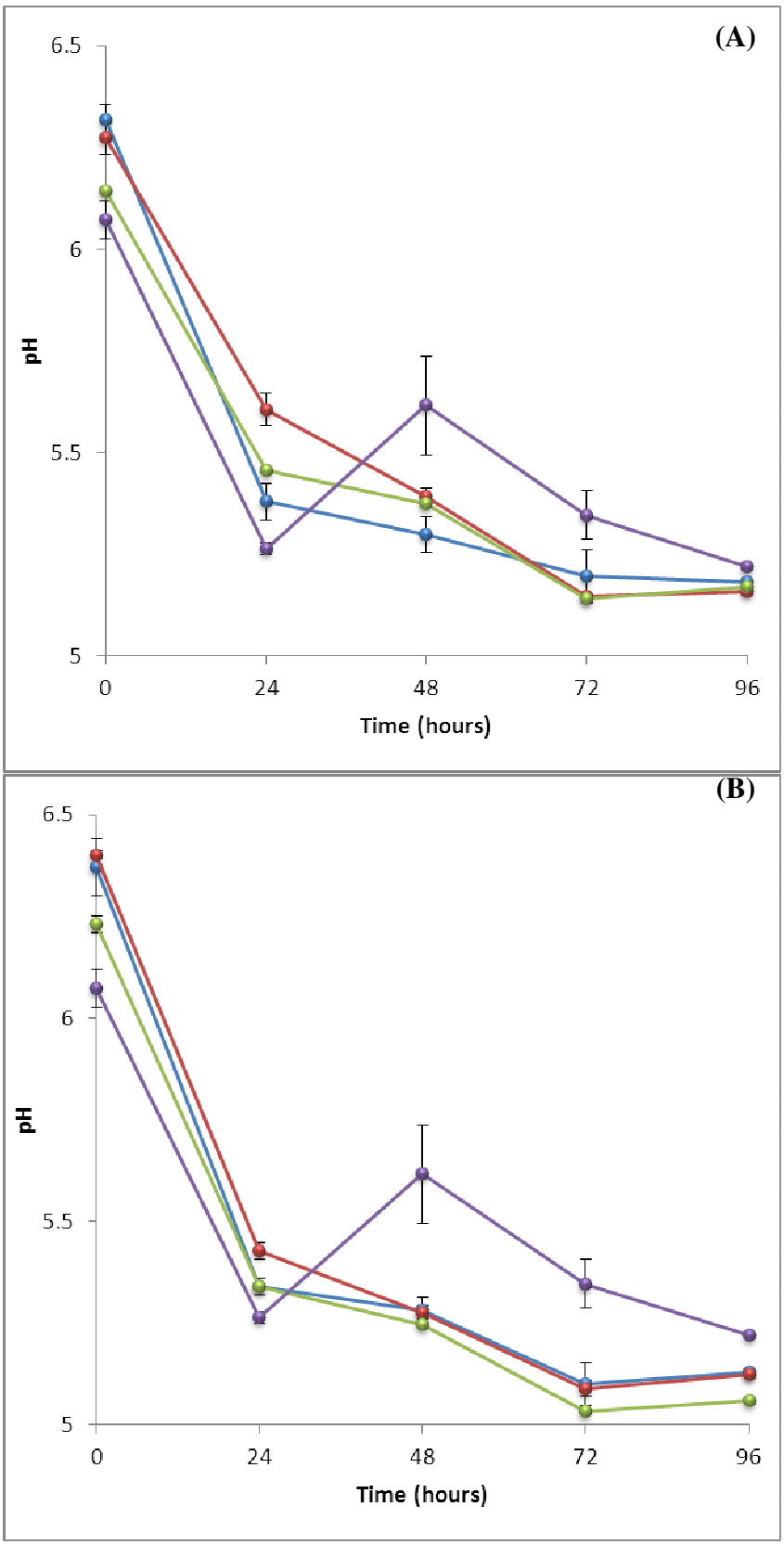


Figure 3-15: pH profile of *C. beijerinckii* grown on 5% xylan

Cultures were established on TYA supplemented with xylose and inoculated into TYA with 5% (w/v) **(A)** Beechwood xylan or **(B)** Birchwood xylan (•) Untreated xylan (•) Xylan supplemented with 0.1% (w/v) xylose (•) Pre xylanase-treated xylan (•) Xylose. The error bars denote the differences between triplicates in standard deviation.

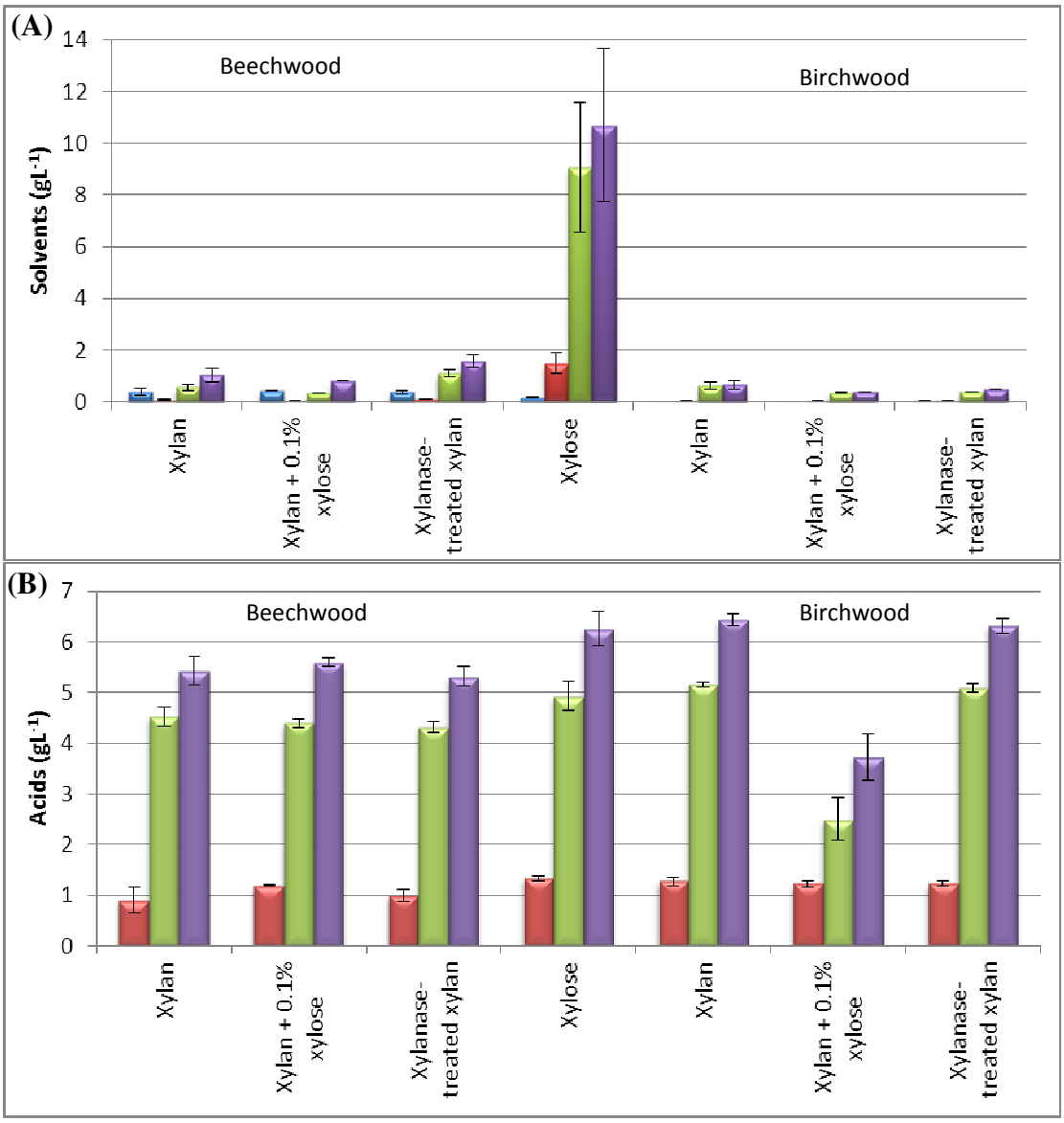


Figure 3-16: Solvent and acid production of *C. beijerinckii* grown on 5% xylan

Cultures were established on xylose and inoculated into 5% of xylan or xylose **(A)** Solvents after 96 hours: (■) Ethanol (■) Acetone (■) Butanol (■) Total ABE **(B)** Acids after 96 hours: (■)Acetate (■)Butyrate (■) Total acids. The error bars denote the differences between triplicates in standard deviation.

3.5.4 Effect of Buffering on Solvent Production from Xylans

It is possible the previous results may be explained by findings of research in *C. acetobutylicum,* where xylanase activity is found to be at an optimum at pH 5.2 in a chemostat (Lee *et al.*, 1985b). It may be that xylanases of *C. beijerinckii* are only active within a narrow pH range and are therefore susceptible to the changes in pH during the course of the fermentation. In order for solventogenesis to occur, acids are reassimilated and used alongside the carbon source cultures are grown on. Therefore, an insufficient or unavailable carbon source, due to unfavourable conditions for xylanase activity, could explain the poor solvent yields observed. It has also been demonstrated in *C. acetobutylicum* that calcium carbonate increases the utilisation of xylose and the solvent yield of cultures grown on xylose (El Kanouni *et al.*, 1998), though this effect was put down to possibly the Ca^{2+} ions providing stability to membrane proteins involved in xylose use and therefore providing an increased butanol tolerance. So to investigate if buffering could help solventogenesis the previous experiment was repeated with the addition of 1% (w/v) calcium carbonate in to the medium as a buffer. A rise in pH was observed between 24–48 hours of fermentation in the pH profile on all the substrates (Figure 3-17) and a solvent yield increase of 1.5–4 fold was observed compared to the previous experiment (Figure 3-18B).

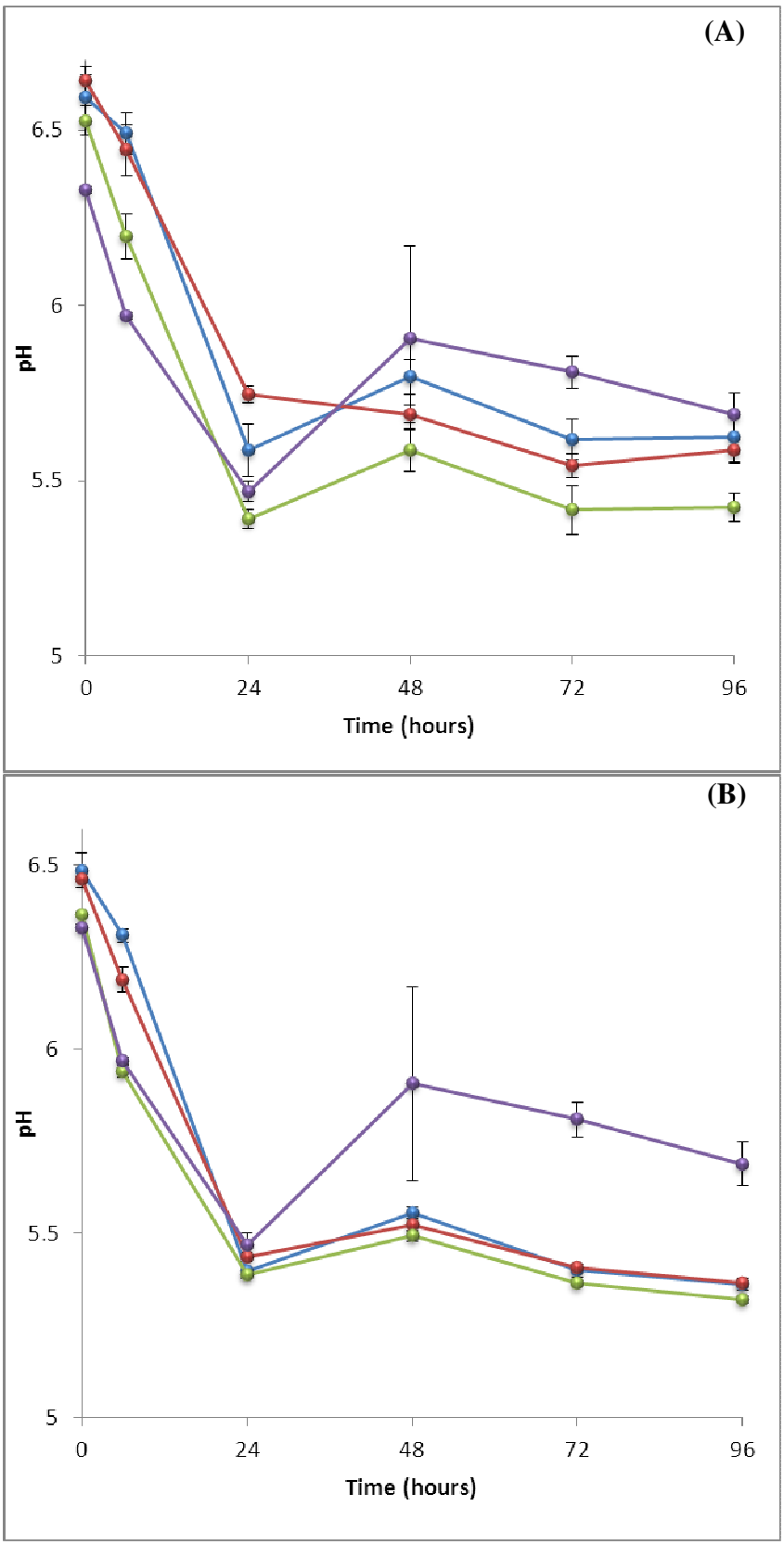


Figure 3-17: pH profile of *C. beijerinckii* grown on 5% xylan supplemented with Ca_2CO_3

Cultures were established on TYA supplemented with xylose and inoculated into TYA with 5% (w/v) **(A)** Beechwood xylan or **(B)** Birchwood xylan (•) Untreated xylan (•) Xylan supplemented with 0.1% xylose (•) Pre xylanase-treated xylan (•) 5% Xylose. The error bars denote the differences between triplicates in standard deviation.

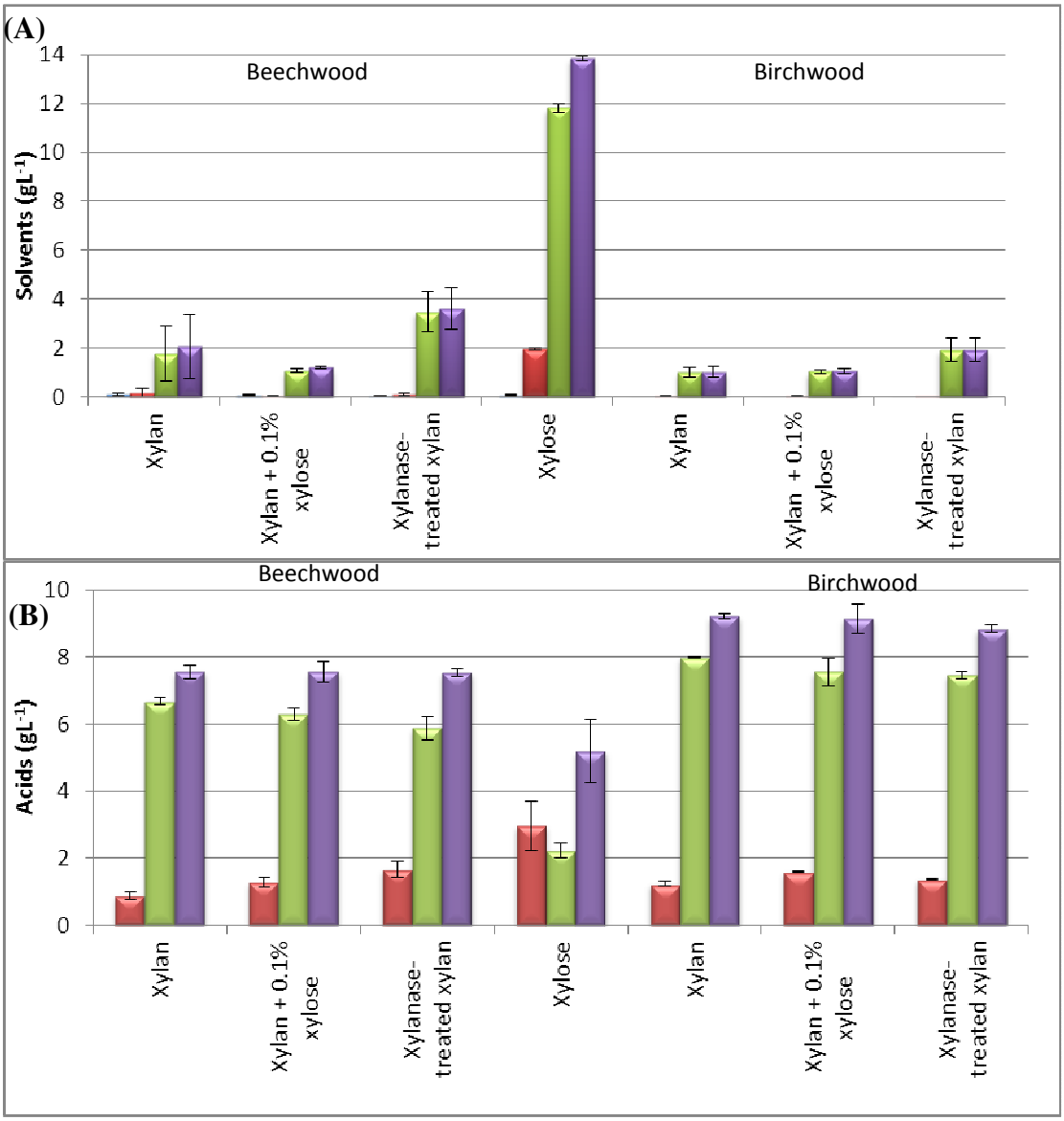


Figure 3-18: Solvent and acid production of *C. beijerinckii* grown on 5% xylan supplemented with Ca_2CO_3

Cultures were established on xylose and inoculated into 5% of xylan or 5% xylose buffered using Ca_2CO_3**(A)** Solvents at 96 hours: (■) Ethanol (■) Acetone (■) Butanol (■) Total ABE **(B)** Acids at 96 hours: (■)Acetate (■)Butyrate (■) Total acids. The error bars denote the differences between triplicates in standard deviation.

3.6 The fermentation of pentose-sugar-rich waste streams by *C. beijerinckii*

It has been established that *C. beijerinckii* is capable of utilising pentose sugars to as a carbon source for solvent production. The next step was to establish the ability of *C. beijerinckii* to utilise pentose sugars in waste streams for future biofuel production. Because waste streams also contain potential inhibitory substrates, such as furfurals, which could interfere with fermentations and any level of interference needs to be assessed, for future biofuel production on these types of resources.

3.6.1 Total carbohydrate determination

Maize-based dried distillers' grains (DDGS, North British Distillery) and were further dried overnight in an oven at 80–90°C, to remove as much of the residual moisture content as possible. More than 7% extra moisture was removed in this way (as calculated by the percentage weight difference before and after the overnight drying). Then samples were taken and total carbohydrate content was determined by acid hydrolysis (Table 3-2).

Table 3-2: Carbohydrate content of maize draff

Carbohydrate*	g per 100g biomass
Glucose (glucan)	19.6 (17.7)
Xylose (xylan)	17.2 (15.2)
Arabinose (arabinan)	6.4 (5.7)

* Galactose was also detected but the concentration was very low and could not be determined accurately by HPLC. No mannose was detected. The polymer values are shown in parentheses

Once the sugar content of DDGS was determined experiments were conducted to assess the ability of *C. beijerinckii* to utilise the sugar constituents of DDGS using different treatment conditions (Table 3-3). It was not possible to measure the OD, because of the appearance of the DDGS in suspension. So only the pH profile and solvent production is presented here.

No appreciative level of solvent was produced by *C. beijerinckii* cultures growing in untreated DDGS (Table 3-4), suggesting that *C. beijerinckii* was unable to breakdown DDGS, or unable to breakdown sufficient amounts to release sugars and sustain growth. There was a difference in the solvent production depending upon when enzymatic treatment was carried out. Cultures of *C. beijerinckii* grown under SSF conditions

produced nearly twice as much solvents as the pre-treated (Figure 3-19). More sugar was released and used by cultures under these conditions (Table 3-4). However, the largest sugar use, and hence the largest solvent production and yield, was observed with cultures growing in soluble sugars. Differences between DDGS and soluble sugar equivalent of DDGS can be explained by potential inhibitors such as ferulic acids or *p*-coumaric acid being present as a product of the acid hydrolysis, which could affect the growth and hence the sugar utilisation and solvent production by cultures. The total solvent and butanol yields of cultures growing in pre-treated DDGS were 60% of that of SSF cultures and 40% of that produced by cultures growing in soluble sugars. SSF cultures total solvent and butanol yields were about 70% of the total solvent and butanol yields of cultures growing in soluble sugars (Table 3-4). Based on all these observations, SSF seems to be the best option for the fermentation of DDGS for by *C. beijerinckii*.

Table 3-3: Treatment conditions of DDGS

Treatment	Acid	Enzyme treatment	inoculated
None	None	none	yes
Pre	Added	24 hours prior (enzymes optimum conditions[#])	yes
SSF	Added	during[+]	yes
SSF control	Added	during[+]	no
Soluble sugars*	none	None	yes

*Sugars autoclaved in separate solutions and added to TYA aseptically after autoclaving. [#] pH 5.0, 50°C, 200 rpm in horizontal shaking incubator. [+]"During" refers to the enzymatic treatment throughout the duration of the fermentation. After 24 hours the bottles were heated for 15 min >100°C in order to inactivate the enzymes. Enzymes used were CTec and HTec (Novazymes). SSF = Simultaneous saccharification and fermentation

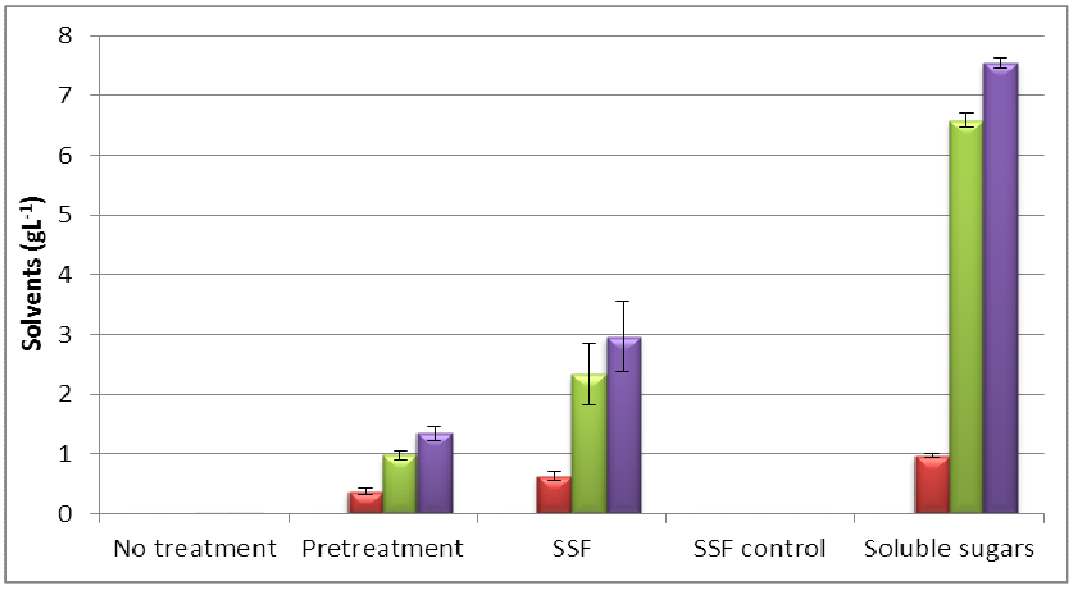


Figure 3-19: The solvent production of *C. beijerinckii* grown in DDGS

C. beijerinckii was established on TYA supplemented with glucose, 72 hour fermentation was carried out on 40g DDGS in 200 ml TYA which was either subjected to acid hydrolysis accompanied with heat treatment (autoclaving) and either treated with cellulase and hemicellulase before the fermentation (pre-treated) or during fermentation (SSF – Simultaneous saccharification and fermentation), or no treatment at all (no treatment and soluble sugars). SSF control was not inoculated with *C. beijerinckii* cultures. (■) Ethanol (■) Acetone (■) Butanol (■) Total ABE The average of triplicates was plotted with the standard deviation bars denoting the range between the triplicates.

Table 3-4: Sugar utilisation, acid and solvent production by *C. beijerinckii* on DDGS

		No treatment	Enzymatic Pre-treatment	SSF	Soluble sugars
Total released sugar * **(gL^{-1})**		0.22	15.52	18.12	25.38
Total used sugar **(gL^{-1})**		0.22	8.47	11.96	21.99
Acids **(gL^{-1})**	Acetic	1.95	2.87	2.21	2.91
	Butyric	1.84	2.39	2.96	1.58
	Total	3.79	5.26	5.17	4.49
Solvents **(gL^{-1})**	Acetone	0	0.37	0.63	0.96
	Butanol	0.003	0.98	2.33	6.58
	Ethanol	0	0	0	0
	Total	0.003	1.34	2.96	7.54
Yields **(g.g)**	Total ABE	0.01	0.16	0.25	0.34
	Butanol	0.01	0.12	0.2	0.3

* this is the total sugar available present at the beginning of the fermentation for pre-enzymatically treated samples or the total sugar released during the course of the fermentation, calculated based on the control SSF.

3.7 Discussion

It has been established that *C. beijerinckii* is able to use both pentose sugars, arabinose and xylose, as sole carbon sources. Regardless of the carbon source (glucose or pentose) used to establish cultures in mixed sugar media, on 1% (w/v) total sugar, glucose is preferentially utilised first. This is in line with studies with a number of organisms, including *B. megaterium*, *T. thermosaccharolyticum*, *C. acetobutylicum*, (Schmiedel and Hillen, 1996, Aduse-Opoku and Mitchell, 1988, Ounine *et al.*, 1985). In *C. acetobutylicum*, xylose permease activity has been observed as being induced by the presence of xylose but inhibited by the presence of glucose on mixed sugar media (Ounine *et al.*, 1985). In *B. subtilis* the utilisation of arabinose is subject to catabolite repression by glucose (Sa-noguiera et al., 1988; Inacio *et al*., 2003). However, *C. thermoaceticum* and *Sulfolobus acidocaldarius* have been observed to utilise xylose and glucose simultaneously (Andreesen *et al.*, 1973, Joshua *et al.*, 2011) and *Corynebacterium glutamicum* utilised arabinose and glucose simultaneously (Kawaguchi *et al.*, 2009).

Despite the presence of glucose affecting the use of xylose, simultaneous use in *C. acetobutylicum* of glucose and xylose can be manipulated by high sugar concentrations (72 gL^{-1}) when xylose is in excess of glucose and cultures are established on xylose (Fond *et al.*, 1985a). In slight contrast to *C. acetobutylicum, C. beijerinckii* used both sugars simultaneously, regardless of the carbon source established on or whether the pentose sugar was in excess or not. The utilisation rate of xylose was greater than that of glucose when the ratio of pentose to glucose is 5:1 or 4:2. On glucose-arabinose containing media, the utilisation of arabinose was greater if the ratio was 5:1. Other strains of *C. beijerinckii*, *C. beijerinckii* NCP260 and *C. beijerinckii* BA101 as well as *C. saccharolyticum* 262, *C. butylicum* NRRL592 and *C. acetobutylicum* ATCC824, have all been observed utilising arabinose and xylose simultaneously with glucose, cellobiose, galactose and mannose in high sugar concentrations of 60 gL^{-1} of mixed sugars (Ezeji and Blaschek, 2008).

The differences in the sugar utilisation behaviour of *C. beijerinckii* cultures in media containing low sugar and high sugar concentrations suggests the mechanism of repression exerted by glucose can be affected by modifying concentrations. Furthermore, xylose appears to exert a greater influence than arabinose in this case.

Solvent production by *C. beijerinckii* cultures was much lower on pentose sugars compared to glucose. This has also been observed in *C. acetobutylicum*, where cultures are observed entering the stationary phase later, at 70 hours, on xylose-containing media in comparison to glucose (33 hours). Additionally, the specific growth rate per hour over the course of the fermentation was also much lower. On the other hand cultures grown in arabinose entered stationary phase after 30 hours (Ounine *et al.*, 1983). In contrast, *C. beijerinckii* cultures grown in glucose or xylose entered stationary phase at the same time (36 hours), whereas arabinose-grown cultures entered after 43 hours. Despite this finding, less sugar was utilised by cultures growing on both pentose sugars and more acids were present at the fermentation end-point. Additionally the acetone: butanol production ratio was higher with pentose sugars than with glucose, a finding which has also been observed in *C. acetobutylicum* (Ounine *et al.*, 1983, Mes-Hartree, 1982)

A number of factors could be responsible for less sugar being utilised and hence less solvents produced from pentose sugars. A lag in the growth rate at the beginning of the fermentation, because all cultures were established on glucose, would cause a delay to the assimilation of the necessary cellular components for pentose sugar use. Another reason may lie in the observation that *C. acetobutylicum* cells grown on pentose sugars may be more sensitive to the effects of butanol levels (Ounine *et al.*, 1985). Therefore growth and solvent production would be reduced.

A number of studies have looked at the addition of chemicals to boost the solvent production of cultures grown on xylose. The addition of acetate or butyrate increases the consumption of xylose and thus the yield of acetone and butanol by xylose-grown cultures of *C. acetobutylicum*, as long as it is added to the culture medium prior to inoculation. Although this suggests that it is a direct conversion of the acids added, the extent at which butanol and acetone levels increased could not be accounted for because upon the addition of only one of them caused both acetone and butanol levels to be boosted (Yu and Saddler, 1983). The addition of 10 gL^{-1} Ca_2CO_3 to the culture medium has also been observed to increase the utilisation of xylose and lessen the inhibitory effect of butanol to *C. acetobutylicum* cells (El Kanouni *et al.*, 1998).

Other factors such as iron limitation have been shown to be detrimental to xylose utilisation, or cysteine addition for utilisation of xylose, mannose, cellobiose and arabinose by *C. acetobutylicum* (El Kanouni *et al.*, 1998, Yu and Saddler, 1983).

Based on the previous finding that *C. beijerinckii* was capable of using xylose as a main carbon source for biofuel production, it was of interest to establish if the polymer xylan, which forms the majority of xylose is found in nature, was also of potential use for biofuel production. It is more economical if xylan does not require to broken down into its constituent units in order to be utilised. An initial experiment indicated that *C. beijerinckii* was indeed capable of utilising xylan and xylan-hydrolysates. However, the solvent profile produced after 96 hours of fermentation was poor, suggesting a condition in the fermentation affecting the solventogenesis. Increasing the amount of carbon source had very little effect on solvent yield. Growth of *C. acetobutylicum* in batch cultures was poor with rapid acid production thought to be a consequence of the lack of buffering capacity in the medium used (Lee *et al.*, 1985b). Although the growth of *C. beijerinckii* was not directly measured it could be inferred from the pH profile, which was consistent with acidogenesis, and the presence of acids at the fermentation endpoint, that poor growth had occurred, which possibly mirrored what is observed with *C. acetobutylicum*.

Lee *et al.*, (1985b) also discovered that xylanase enzyme of *C. acetobutylicum* had a pH range optimum between pH 5.8–6.0 and was stable down to pH 5.2. Below pH 5.2, xylanase activity dropped five-fold, with the utilisation of xylan decreasing from 50% to 30%. Calcium carbonate was therefore added to provide buffering capacity and maintain a higher pH profile, based on the assumption that the xylanase of *C. beijerinckii* had a similar optimal pH range. This boosted the amount of solvent produced by up to two-fold, suggesting the pH has a profound effect on xylanase activity as acidogenesis proceeds. This could be investigated further by using chemostat or by using a series of media buffered at a range of different pH. The putative xylanase from *C. beijerinckii* could also be isolated and activity determined under a range of conditions.

Another explanation to the poor xylan utilisation observed may have come from accumulation of xylose produced from the breakdown of xylan negatively regulated the xylanase genes, the expression of xylanase in the fungus *Hypocrea jecori* has been

observed to be regulated by the concentration of xylose. Higher xylose concentrations resulted in a decrease in xylanase expression (Mach-Aigner *et al.*, 2010). However, such an effect can be discounted, as the TLC analysis showed no evidence of xylose accumulation. A more likely explanation comes from upon onset of solventogenesis, the pH may have fallen below the optimum for the xylanase to function, resulting in insufficient carbon source was available for solventogenesis.

The presence of 0.1% xylose in xylan in the experiments conducted here showed no indication of a delay in the production of acids, indicated by the pH profile or of a reduced solvent production, which would be expected to be associated with this.

The effect of other carbon sources on the utilisation of xylan could also be of importance and is an avenue for future research. Glucose has been observed to affect the expression of xylosidase in *B. subtilis* had no effect on the expression of xylanase (Linder *et al*., 1994) whereas the xylanase promoter in *Trichoderma reesei* was silenced in the presence of glucose (Zeilinger *et al.*, 1996).

Lignocellulose presents an attractive resource in terms of biofuel production for a number of reasons. It is widely available from many sources, such as agriculture, forestry, industry and municipals. It is the most abundant polymer on the planet and is a renewable, sustainable and cheap resource. All qualities needed for a substrate for biofuel production. A number of resources have been investigated and found suitable for biofuel production in solventogenic clostridia. Wheat straw, corn fibre, corn cobs, soy bean, domestic organic waste, soft and hardwoods and dried distillers' grains have all been found to be fermentable substrates.

Agricultural resources can be by-products of grain, oil-seed, fruit and vegetable harvesting, such as stalks, seeds, shells, husks, straw, sludge, wastewater and juice (Howard *et al.*, 2003). Resources are dependent upon the area of the World. Favourable substrates for biofuel production in Asia are rice straw, wheat straw and corn stover, in Europe wheat straw and in North America Corn stover (Kim and Dale, 2006). There is an estimated 60-80 million dry tonnes of corn stover available in the USA for fermentations and a number of investigations into the ability of solventogenic clostridia to utilise products of corn waste have been carried out. *C. beijerinckii* BA101 was able to produce 0.39 grams of ABE solvents per gram of corn fibre, if the acid hydrolysed

corn fibre was treated with XAD-4 resin, as the latter removed the cell growth inhibition attributed to the products arising from acid hydrolysis (Qureshi *et al.*, 2008). *C. beijerinckii* P260 was able to produce 0.41g of butanol per gram of corn fibre xylan (Qureshi *et al.*, 2006). Additionally, *C. beijerinckii* strains, NCIMB8052 and BA101 when grown on corn-steep water (a by-product of the wet-milling industry), were found to produce 16 gL^{-1} and 8.5 gL^{-1} butanol, respectively (Parekh *et al.*, 1999). In Russia, post second World War corn cobs and agricultural waste such as hemp waste and sunflower seeds were a waste problem in some areas (Nakhmanovich and Shcheblykina, 1959, Zverlov *et al.*, 2006). In 1962, at Dokshukino, one of the major plants in Russia, 925 tonnes of corn cob waste was used per month in clostridial fermentations (Zverlov *et al.*, 2006). Reported yields of solvents at this plant averaged 32 g per 100g of sugars. Other agricultural residues, derived from wheat, such as bran and straw have also been used successfully as feedstocks for clostridial fermentations. The fermentation of wheat straw hydrolysate by *C. beijerinckii* P260 was found to produce 0.42 g of solvents per gram of sugars (Qureshi *et al.*, 2007). *C. beijerinckii* ATCC55025 was found to use hexose and pentose sugars simultaneously in wheat bran hydrolysate with a yield of 0.32 g ABE per gram of sugar (Liu *et al.*, 2010).

Other sources for feedstocks come from forestry and paper-pulp mills, including wood, bark, leaves, saw dust, fibre and black liquor (Howard *et al.*, 2003). Steam exploded aspen wood has been shown to be a usable resource by *C. acetobutylicum* ATCC824 for the production of biofuel (Saddler *et al.*, 1983, Yu and Saddler, 1983). Steam exploded woodchips were shown to produce 0.26g of butanol per gram of sugar consumed.

Municipal wastes such as food waste, paper, card and wood (Howard *et al.*, 2003). Waste paper, plant residues, sawdust and fruit and vegetable waste encompasses a considerable amount of the solid-based wastes in Tanzania (Mtui and Nakamura, 2005). Such wastes as these could be dealt with using solventogenic clostridia. Indeed, fruit, vegetable and garden wastes have been found to be fermentable without the addition of extra nutrients to the culture medium in *C. acetobutylicum* strains ATCC824 and DSM1731, *C. beijerinckii* B-592 and Clostridium LMD84.48 (Lopez-Contreras *et al.*, 2000, Claassen *et al.*, 2000). Domestic organic waste hydrolysate fermented by *C. acetobutylicum* DSM1731 produced 28g of ABE per 100g of the sugar available in the (Claassen *et al.*, 2000). A study by (Kobayashi *et al.*, 2005) sludge from a Japanese waste water treatment was not useable by *C. saccharoperbutylacetonicum* without the

addition of a carbon source. Using glucose as a carbon source the more sludge added to the medium resulted in more glucose to be used, which in turn increased the amount of butanol produced. Such a resource would have to be mixed with a carbon-rich waste stream, which maybe poor in other nutrients and thus would complement each other and to be a viable feedstock for biofuel production.

Dried distillers' grains are a waste product of the production of ethanol for beverage and transport industries. DDGS are rich in glucans, arabinans and xylans, so are an attractive resource. In this study the ability of *C. beijerinckii* cultures to ferment DDGS was investigated. A very similar study was carried out with several solventogenic clostridia strains, *C. beijerinckii* BA101 and 260, *C. acetobutylicum*, *C. butylicum* and *C. saccharolyticum*, using hydrolysates of DDGS produced with either dilute acid, hot water or ammonium fibre explosion pre-treatment. Depending on the strain and the production method of the hydrolysate, yields ranged from 0.3–0.35 g.g, in comparison to 0.33–0.39 g.g obtained on the DDGS equivalent soluble sugars (Ezeji and Blaschek, 2008). In this study *C. beijerinckii* had a total solvent yield of 0.34 g.g on the equivalent DDGS soluble sugars, which is similar to the yields obtained of the solventogenic clostridia. However there was a big difference between the yield of the pre-treated DDGS and that of the equivalent soluble sugars. A total solvent yield of 0.16 g.g was obtained, for acid and enzyme pre-treated DDGS.

A number of differences in Materials and Methods exist between this study and that of Ezeji *et al*. (2008) and therefore direct comparisons cannot be made. However, the yields between DDGS and the equivalent sugars were notably different in this study whereas Ezeji *et al*. (2008) found similar yields. These results could be due to inhibitory compounds released from the hydrolysis of DDGS which may have interfered with the enzymatic breakdown and/or the fermentation. Strains of *C. acetobutylicum* have been observed producing total solvent yields of 0.15 g.g and 0.17 g.g for hardwood and softwood hydrolysates, respectively, which is notably lower than the highest theoretical yield of 0.32 for cultures on a glucose soluble sugar-based medium (Saddler *et al.*, 1983, Maddox and Murray, 1983).

Although pre-treatment of substrates allows for optimal conditions for sugar release and fermentation, it also has drawbacks, such as the accumulation of sugar products in the culture medium. This can inhibit the breakdown of other sugars if the expression of a

particular gene associated with their use is inhibited by the presence of another sugar (Galbe and Zacchi, 2002). For the simultaneous saccharification and fermentation cultures the yield was higher at 0.25 g.g, suggesting these conditions are the best to derive solvent from the DDGS for *C. beijerinckii*, despite the compromise on the conditions for enzyme activity and fermentation.

Chapter 4

in silico Analysis for Genes Involved in Pentose Sugar Utilisation

4 *in silico* Analysis for Genes Involved in Pentose Sugar Utilisation

4.1 Introduction

It has been established that *C. beijerinckii* is able to utilise arabinose, xylose, xylan and xylan hydrolysates as sole carbon sources and produce solvents (Chapter 3). These findings, taken along with a preferential use of glucose over both pentose sugars on 1% (w/v) total sugar, suggests there are genes present within the genome for the transport and utilisation of xylan, xylose and arabinose. In a number of organisms gene systems involved in pentose sugar have been firmly established (Sa-Nogueira *et al.*, 1997, Schmiedel *et al.*, 1997, Shamanna and Sanderson, 1979, Kawaguchi *et al.*, 2009, Takeda *et al.*, 1998), as have genes involved in the use of xylan (Ko *et al.*, 1992, Gasparic *et al.*, 1995, Luthi *et al.*, 1990, Wolf *et al.*, 1995). Based on this information, candidates for key genes associated with the use of pentose sugars and xylan were searched for within the annotated genome of *C. beijerinckii*. The genome of *C. beijerinckii* consists of 6Mb circular chromosome and was sequenced and made available 2007 by DOE joint genome institute (Project id 3634512). The candidates identified using KEGG were then used for further analysis using various bioinformatics programs detailed here and in Chapter 2. This was carried out to provide more evidence for the candidate's likely function ahead of any characterisation studies, which are necessary to confirm the definite function.

Bioinformatics is described broadly as the application of computer technology to allow the user to analyse and make sense of large sets of biological data (Attwood and Parry-Smith, 1999). The label applied to a particular ORF within a genome may not necessarily reflect the true function, so it is necessary to employ further analyses. Analyses performed and presented here took the approach of assessing the level of similarity of the candidates with published characterised genes of other bacteria. This was based on the principal proteins which perform the same or similar function, harbour particular structural or functional domains necessary for a particular function. So in order to preserve this function these domains possess identical or very similar AA sequences, even across different species.

For each sequence of an ORF putatively involved in pentose sugar utilisation was identified from a genomic search and input into BLAST. BLAST performs local alignments between a query sequence and those of other species within the database in a

pair-wise fashion. It does so by analysing domains of sequence similarity, which are often involved in function (Madden, 2003). Such a tool displays results in table format providing a statistical significance, to infer whether the two sequences are likely to be orthologues. Additionally, the level of AA sequence identity (identical AAs) in terms of percentage and AA sequence similarity (identical and substituted AAs of similar properties) between the query and the database sequences, are also displayed.

Following on from this the AA sequences were taken either from BLAST searches or from the EMBL accession database, the latter using accession numbers from publications. These were then used to construct multiple alignments and phylograms, or for transmembrane prediction.

Multiple alignments allow analysis of a number of sequences at once for conserved domains across protein families or groups of species. The alignments vertically align the input sequences in such a way as to preserve an individual sequence's residues, as well as taking into account differences accumulated through evolution, by the use of gap insertions. These can reveal possible functional and structural motifs indicative of a particular type of protein (Attwood and Parry-Smith, 1999). The AA sequence of the candidate ORFs or the published characterised proteins were input into ClustalW2, producing an .aln file which was viewed as a multiple alignment in GeneDoc. The results from ClustalW2 also provided the values of percentage identity between candidates that is identical AAs and known pentose sugar utilisation AA sequences.

Radial trees allow analysis of selected sequences in comparison to one another, providing inference of evolutionary history. Such ancestral relationships can be inferred by the 'clustering' in the tree and the number of changes in AAs can be inferred by the branch length (Attwood and Parry-Smith, 1999). The AA sequence of the candidate proteins or the published characterised proteins were input into ClustalW2 producing a .nxs file, which was then run in MrBayes (Huelsenbeck and Ronquist, 2001), in order to obtain a Bayesian inference of phylogeny, providing a confidence level to the topology and accuracy of how the radial tree was built. A .tre file was produced from MrBayes which was viewed TreeView program.

The Hidden Markov Model (HMM) can be used to predict and model membrane α helices in terms of location and orientation. When tested on datasets of proteins with

known topology HMM has an accuracy of 97-98% (Krogh *et al.*, 2001). The model works on a number of known parameters: transmembrane alpha helices have long stretches of hydrophobic AAs, the orientation can be assumed because it is known that arginine and lysine, positive charged amino acids, are mainly found on the cytoplasmic side of a membrane and thus in this way the AA loops found on either side of a membrane can be differentiated (Sonnhammer *et al.*, 1998). All sections of the transmembrane protein are predicted based on the probability score. Some sections show potential qualities of one or more of the sections making up a transmembrane protein, but the final prediction is based on which has the highest probability score. AA sequences were input into the TMHMM server which produced a data output which was used to produce a two dimensional structure of putative transmembrane proteins using TMRPres2D (Spyropoulos *et al.*, 2004).

Bioinformatic analyses were carried out with the aim of proposing gene systems involved in arabinose, xylose and xylan utilisation in *C. beijerinckii*.

4.2 Pentose sugar utilisation in bacteria

The known pathway of pentose sugar metabolism is shown in Figure 4-1 (Bettiga *et al.*, 2008). The breakdown of xylan is achieved naturally by two groups of enzymes of the glycosidase family. These are endo 1,4-xylanase (European Community (EC) 3.2.1.8), which randomly cleave the backbone into short oligomers and β-xylosidase (EC 3.2.1.37), which are exo-acting, hydrolysing the small oligomers into single xylose units, as shown in Figure 4-2 (Shallom and Shoham, 2003, Saha, 2003, Perez *et al.*, 2002, Wong *et al.*, 1988).

Based on this a basic genome search was conducted to find candidate genes for arabinose, xylose and xylan utilisation in *C. beijerinckii* and a candidate gene system was identified for each pentose sugars utilisation (Figure 4-3 & Figure 4-4) and xylan utilisation (Figure 4-5).

Various bioinformatic analyses were performed on each individual ORF within the candidate gene systems to assess the potential identities further and to provide more evidence of the likely function. The results of those analyses are presented here.

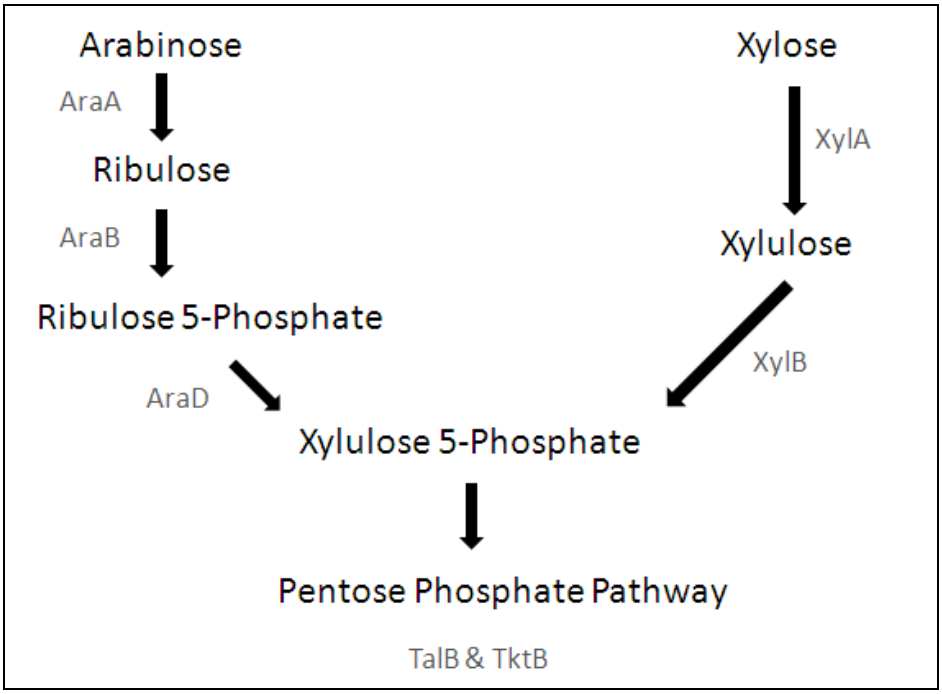


Figure 4-1: The metabolism of Pentose sugars by bacteria

Pentose sugars are transported within the cell then metabolised by this pathway in bacteria. **AraA**- L-arabinose isomerase, **AraB**- L-ribulokinase, **AraD**-L-Ribulose 5-phosphate 4-epimerase, **XylA**- D-xylose isomerase, **XylB**- D-xylulokinase

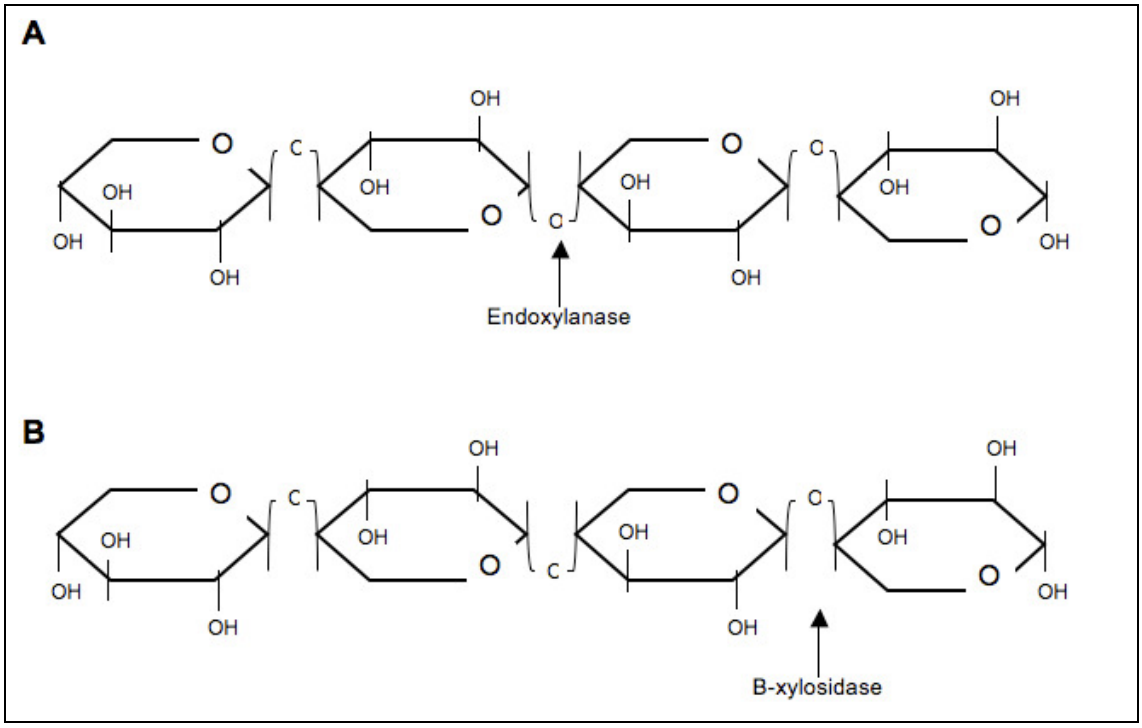


Figure 4-2: Enzymes involved in xylan hydrolysis

(A) shows the action of endoxylanases on xylan backbone and **(B)** shows the action of exo-acting β xylosidases. Diagram taken from Goldman, (2009).

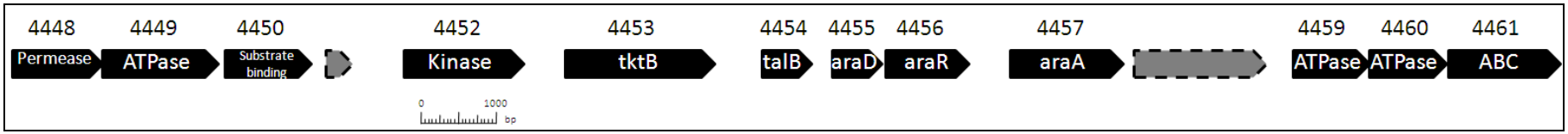


Figure 4-3: The candidate arabinose utilisation gene system in *C. beijerinckii*

This is a candidate gene system for arabinose utilisation proposed after a genomic search, see text for details. Numbers above candidate genes denote the locus tag. The labels given to each ORF remain putative and are the labels applied in the genomic database. ***tktB*** – Transketolase, ***talB*** – Transaldolase, ***araD*** – Ribulose 5-phosphate 4-epimerase, ***araR*** – Arabinose transcriptional repressor, ***araA*** – Arabinose isomerase, **ABC** – ABC transporter related.

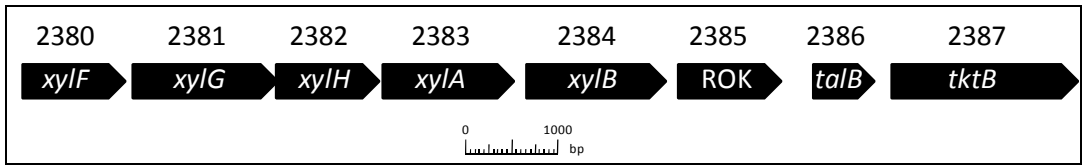


Figure 4-4: The candidate xylose utilisation gene system of *C. beijerinckii*

xylF – Xylose binding protein, ***xylG*** - ATPase, ***xylH*** – Permease, ***xylA*** – Xylose isomerase, ***xylB*** – Xylulokinase, **ROK** – ROK family, ***talB*** – Transaldolase, ***tktB*** – Transketolase. Numbers above denote locus tag.

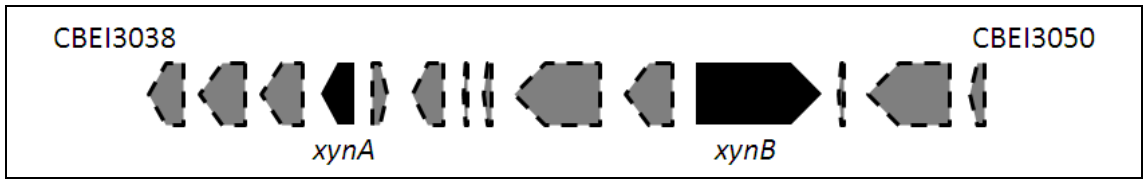


Figure 4-5: The candidate xylan gene system of *C. beijerinckii*

CBEI3038 – Membrane zinc metallopeptidase, **CBEI3038** – Fructose- 1,6 bisphosphate aldolase, **CBEI3040** – Hypothetical protein, **CBEI3041(*xynA*)** – Endo 1,4-β-xylanase, **CBEI3042** – CBEI3044 Hypothetical protein, **CBEI3045** – Methyl-accepting chemotaxis protein, **CBEI3046**- NADPH- dependent FMN reductase, **CBEI3047 (*xynB*)** – Xylan 1,4- β-xylosidase, **CBEI3048** – Pseudogene, **CBEI3049** – ABC transporter, **CBEI3050** – Response regulator receiver protein. The genes are annotated with their putative identity and are not drawn to scale.

4.3 Arabinose Utilisation Genes

araABD are key genes associated with the use of arabinose in a number of prokaryotes (Sa-Nogueira and de Lencastre, 1989, Kawaguchi *et al.*, 2009, Lee *et al.*, 1986, Lin *et al.*, 1985b) and function sequentially to catabolise arabinose into ribulose, ribulose 5-phosphate and xylulose 5-phosphate, respectively (Figure 4-1). The latter is then fed into the pentose phosphate pathway. Candidate genes for arabinose utilisation were identified from a genome search, ORF CBEI4452, 4455 and 4457 were identified as putative *araBDA* genes, respectively. Further bioinformatics analyses as already detailed were applied to provide more evidence for the likely function.

4.3.1 L- arabinose isomerase (*araA*)

The AA sequence of the candidate AraA of *C. beijerinckii* and of published characterised AraA sequences, were used to assess the percentage of sequence identity and to construct a radial tree. Analysis of published characterised AraA sequences and the AA sequence of CBEI4457 (Table 4-1) revealed the most sequence similarity to AraA of *Alicyclobacillus acidocaldarius* (66% identity) and the least identity with *Thermoanaerobacter mathranii* (26%). The radial tree reflected the results of Table 4-1. CBEI4457 was clustered amongst the AraA sequences with 60% or more sequence identity. The clustering and distribution of sequences did not seem to be subject to any particular feature, Gram identity, a high or low genome GC content or the AA sequence length (Figure 4-6).

A multiple alignment of characterised AraA AA sequences of both Gram negative and positive species and CBEI4457, was then constructed (Figure 4-7), large blocks of conserved residues were evident between all the sequences analysed (black shading), suggestive of structurally and functionally important regions, which have been conserved across both the Gram negative and positive species. Indeed, AA residues associated with AraA function in *E. coli* consisting of two histidine (His) and two glutamine (Glu) residues, thought to be a part of the AraA active site were also present in the other the AraA sequences and CBEI4457 (Manjasetty and Chance, 2006). In CBEI4457 they were present at 440 and 342 (His) and 300 and 325 (Glu). In comparison to positions 450, 350, 306 and 333, respectively in *E. coli*.

Table 4-1: The sequence identity of CBEI4457 with published characterised AraA proteins

Organism	Identity (%)
Alicyclobacillus acidocaldarius	66
Bacillus stearothermophilus	65
Geobacillus thermodenitrificans	65
*Thermus sp.*IM6501	65
Bacillus halodurans	62
Bacillus licheniformis	62
Bacillus subtilis	60
Thermotoga maritima	56
Thermotoga neapolitana	56
Escherichia coli	52
Lactobacillus plantarum	52
Salmonella typhimurium	52
Mycobacterium smegmatis	51
Thermoanaerobacter mathranii	26

Identity is the percentage of identical shared AA between CBEI4457 and the AraA sequences in the table

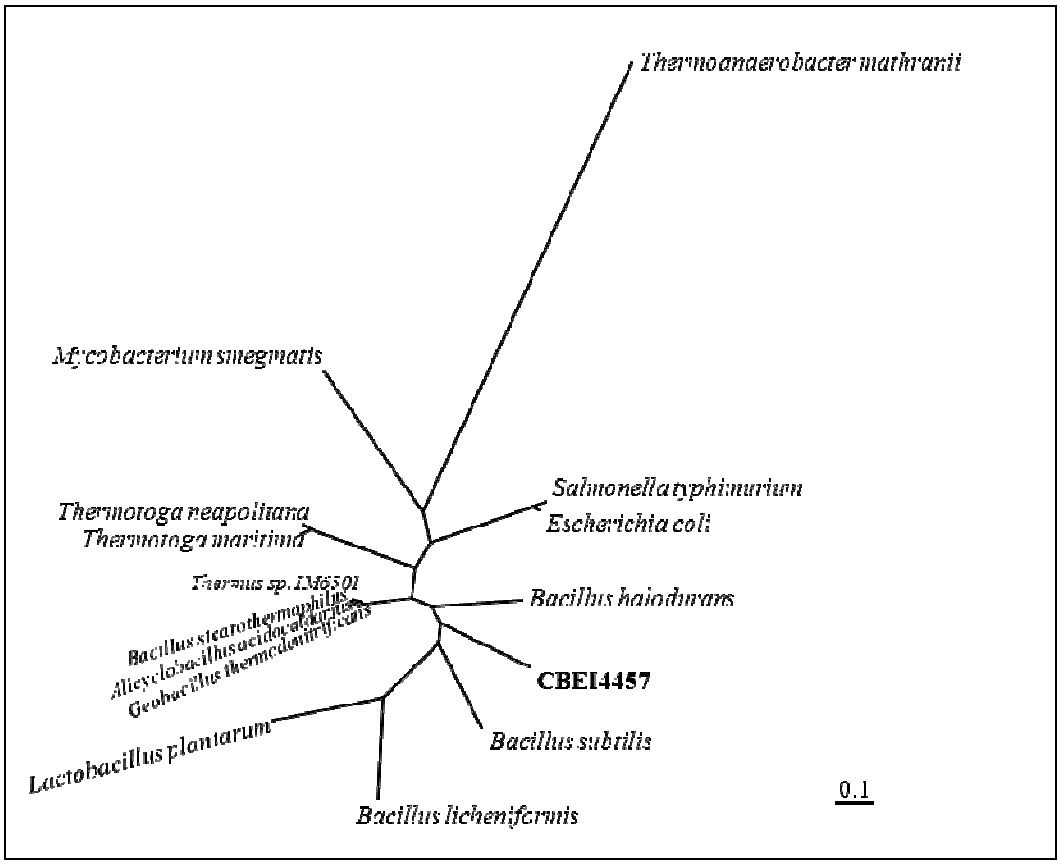


Figure 4-6: A radial tree of published characterised AraA sequences and CBEI4457, the candidate *araA* of *C. beijerinckii*

The AA sequences of AraA proteins and CBEI4457, the AraA candidate of *C. beijerinckii* were used to construct a phylogram, see text for details. Scale denotes 0.1 amino acid substitutions per site.
A. acidocaldarius (Lee *et al.*, 2005), ***B. licheniformis*** (Prabhu *et al.*, 2008), ***B. halodurans*** (Rhimi and Bejar, 2006), ***B stearothermophilus*** (Rhimi and Bejar, 2006), ***B. subtilis*** (Sa-Nogueira and de Lencastre, 1989), ***E. coli*** (Lee *et al.*, 1986), ***G. thermodentrificans*** (Kim and Oh, 2005), ***L. plantarum*** (Chouayekh *et al.*, 2007), ***M. smegmatis*** (Takata *et al.*, 2007), ***S. typhimurium*** *(Lin et al., 1985c)*, ***T. mathranii*** (Jorgensen *et al.*, 2004), ***T. maritima*** (Lee *et al.*, 2004), ***T. neapolitana*** (Kim *et al.*, 2002), ***Thermus sp.* IM6501**(Kim *et al.*, 2003).

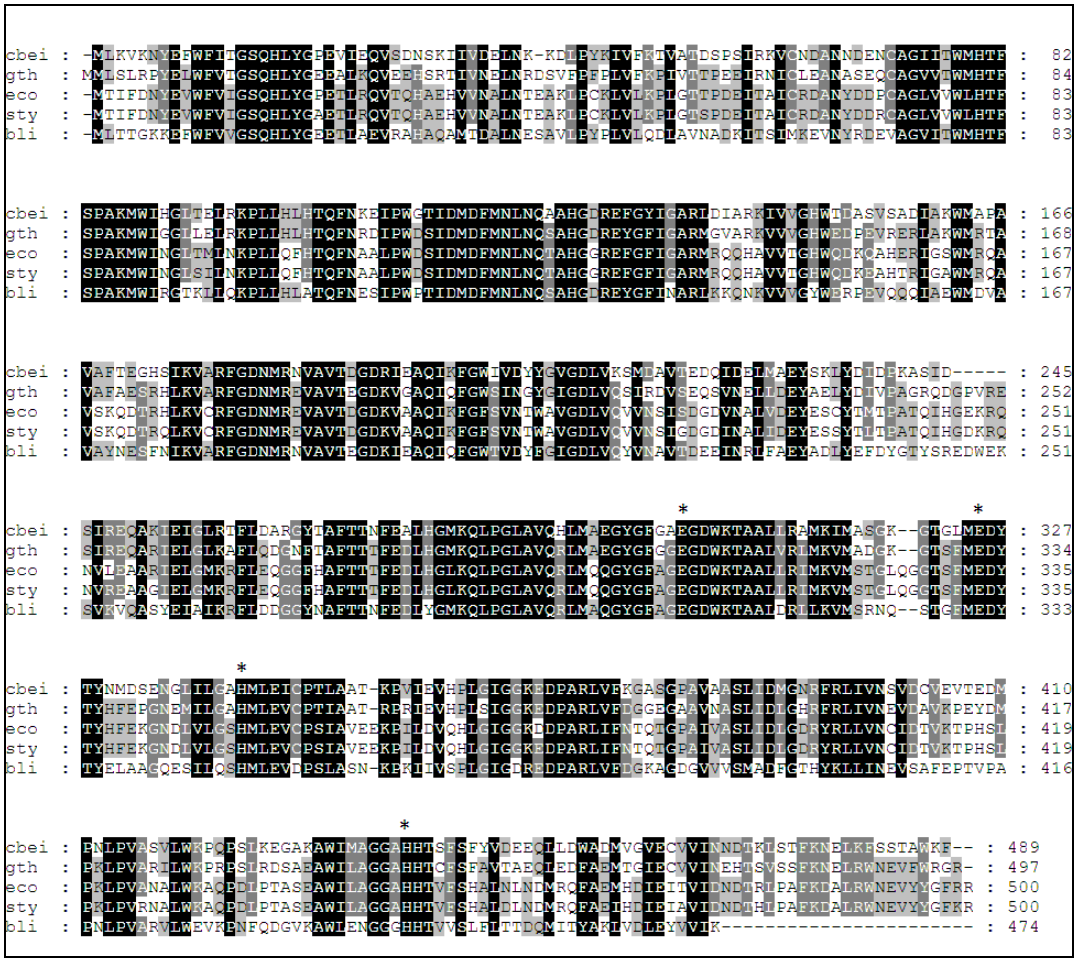


Figure 4-7: Multiple alignment of a selection of published AraA proteins and CBEI4457

The AA sequences of AraA proteins and *C. beijerinckii* candidate were used to construct a multiple alignment, see text for details. **Black** shading denotes 100% sequence similarity (similarity shared by all sequences in the multiple alignment at the shaded point), **Grey shading with white text**- 80%, **grey shading with black text** – 60% sequence similarity. Asterisks indicate the conserved His and Glu residues thought to be associated with the active site. In *E.coli* the positions are 450, 350, 306 and 333, respectively. In CBEI4457 they were present at 440 and 342 and 300 and 325 AAs, respectively (Manjasetty & Chance, 2006).
bli- *B. licheniformis* (Prabhu *et al.*, 2008), **cbei**- *C. beijerinckii* (locus tag CBEI4457) **eco**- *E. coli* (Lee *et al.*, 1986), **gth**- *G. thermodentrificans* (Kim and Oh, 2005), **sty**- *S. typhimurium* (Lin *et al.*, 1985b).

4.3.2 L-ribulokinase (*araB*)

The second protein involved in arabinose metabolism, AraB was then investigated. CBEI4452 shared most sequence identity with AraB of *Cornynebacterium glutamicum* (Table 4-2). A radial tree of characterised AraB sequences and the candidate AraB of *C. beijerinckii* was constructed using the AA sequences (Figure 4-8). This reflected the sequence identity shared between CBEI4452 and that of the AraB of *C. glutamicum* and the differences between this pair and the other AraB sequences in Table 4-2. The extent of sequence similarity, between this pair of sequences is highly likely based on the even distribution of similarity throughout the sequences, as shown in an alignment (Figure 4-9).

Table 4-2: The sequence identity of CBEI4452 with published characterised AraB proteins

Organism	Identity (%)
Corynebacterium glutamican	48
Bacillus stearothermophilus	15
Bacillus subtilis	13
Escherichia coli	11
Salmonella enterica	8

Identity is the percentage of identical shared AA between CBEI4452 and the AraB sequences in the table

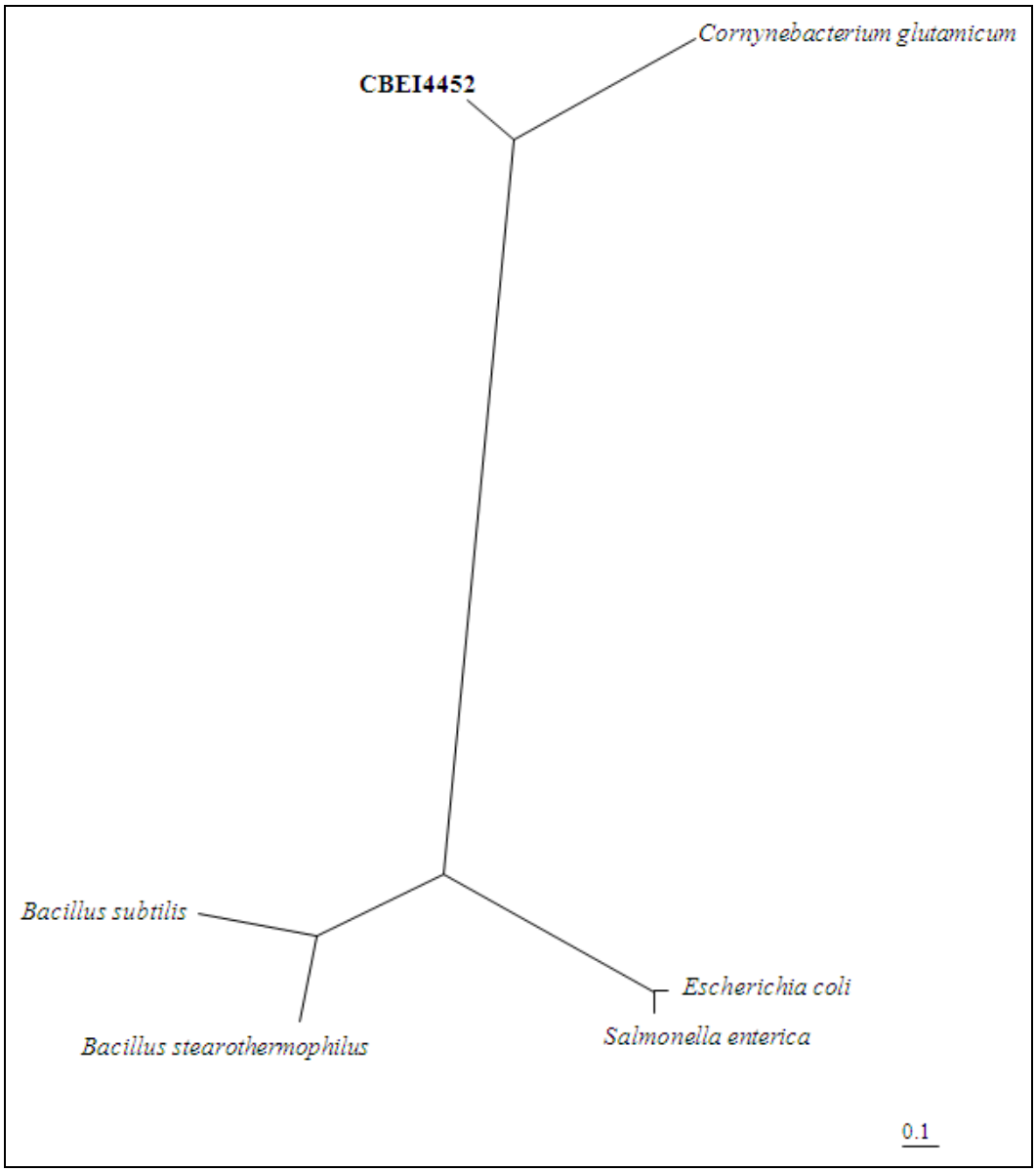


Figure 4-8: A radial tree of characterised AraB proteins and CBEI4452, the AraB candidate of *C. beijerinckii*

The AA sequences of characterised AraB proteins and the putative AraB of *C. beijerinckii* were used to construct a phylogram, see text for details. Scale denotes 0.1 amino acid substitutions per site
B. subtilis (Sa-Nogueira and de Lencastre, 1989), ***C. beijerinckii*** (locus tag CBEI4452), ***E. coli*** (Lee *et al.*, 1986), ***S. enterica*** (Lin *et al.*, 1985b), ***C. glutamicum*** (Kawaguchi *et al.*, 2009)

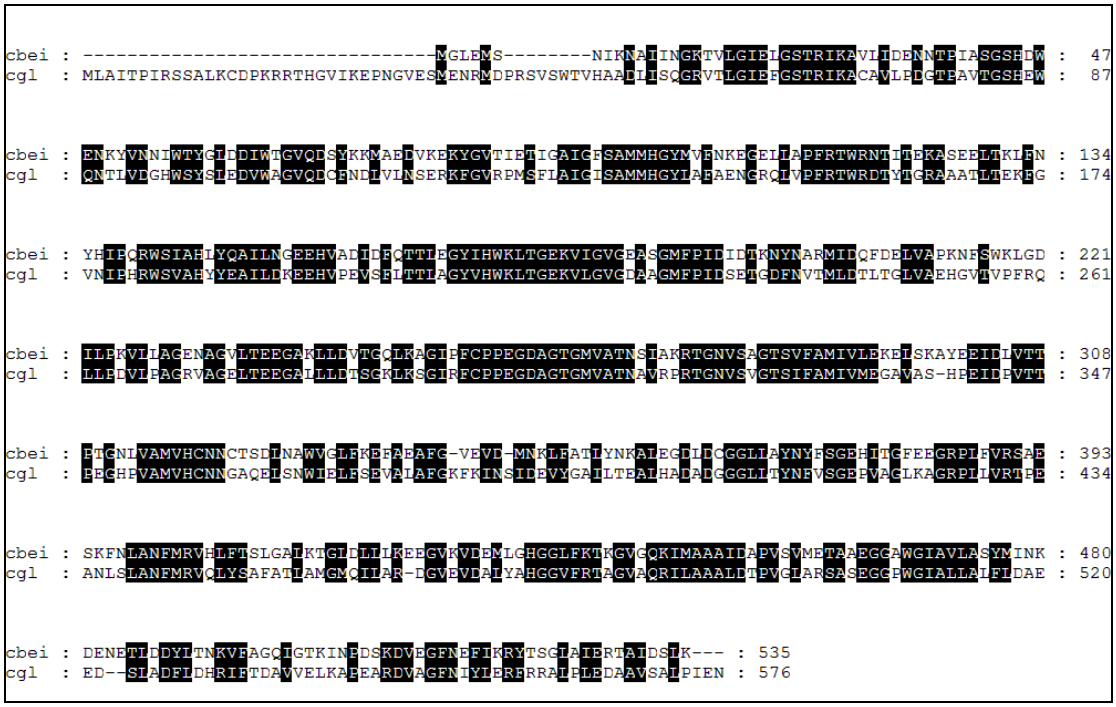


Figure 4-9: Alignment AraB of *C. glutamicum* with the putative AraB, CBEI4452

An alignment was constructed using the AA sequence of a confirmed AraB protein from *C. glutamicum* and the putative AraB of *C. beijerinckii*, see text for details. Sequence similarity between the two sequences is shaded in black. **cbei**- *C. beijerinckii* (locus tag CBEI4452), **cgl**- *C. glutamicum* (Kawaguchi *et al.*, 2009).

4.3.3 L-ribulose 5-phosphate 4-epimerase (*araD*)

The final protein involved in arabinose utilisation prior to the pentose phosphate pathway, AraD (putatively CBEI4455) was then compared to characterised AraD sequences (Table 4-3). The greatest sequence similarity was apparent between AraD of *E. coli* and CBEI4455. The radial tree highlighted Gram identity specific similarities between the sequences, with sequences of Gram negative organisms clustered together and Gram positive sequences, with the exception of AraD of *C. glutamican*, suggesting less sequence similarity to the others. The extent and presence of similarity between AraD of *E. coli* and CBEI4455, was investigated by an alignment (Figure 4-11). An even distribution of similarity was evident throughout the two sequences, suggestive of common or related functions between the two sequences being highly likely.

Table 4-3: The sequence identity of CBEI4455 with published characterised AraD proteins

Organism	Identity (%)
Escherichia coli	62
Bacillus subtilis	56
Salmonella enterica	53
Corynebacterium glutamican	33

Identity is the percentage of identical shared AA between CBEI4455 and the AraD sequences in the table

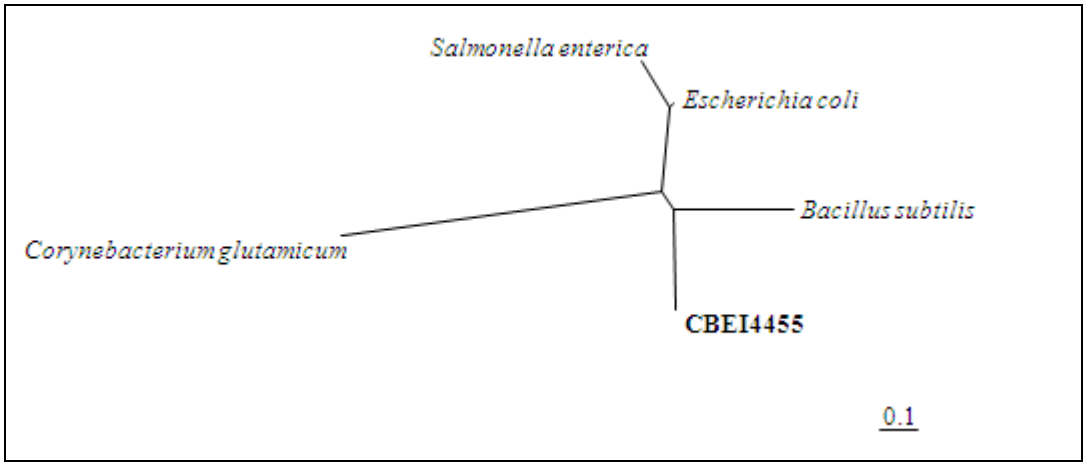


Figure 4-10: A radial tree of characterised AraD proteins and the putative AraD of *C. beijerinckii*, CBEI4455

The AA sequences of characterised AraD proteins were used alongside the putative AraD of *C. beijerinckii* to construct a phylogram, see text for details. Scale denotes 0.1 amino acid substitutions per site.
B. subtilis (Sa-Nogueira and de Lencastre, 1989), ***C. beijerinckii*** (locus tag CBEI4455), ***C. glutamicum*** (Kawaguchi *et al.*, 2009), ***E. coli*** (Lee *et al.*, 1986), ***S. enterica*** (Lin *et al.*, 1985a)

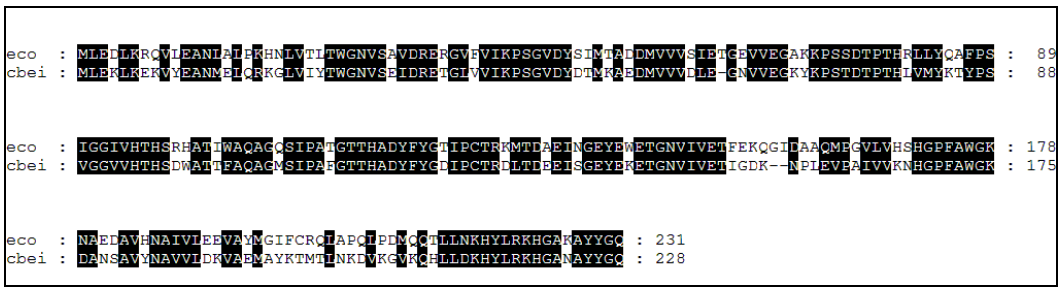


Figure 4-11: A multiple alignment of AraD of *E. coli* with CBEI4455, the putative AraD candidate of *C. beijerinckii*

A multiple alignment was constructed using the AA sequences of a confirmed AraD protein of *E. coli* and the putative AraD of *C. beijerinckii*, see text for details. The sequence similarity between the two is shaded in black. **cbei**- *C. beijerinckii* (locus tag CBEI4455), **eco**- *E. coli* (Lee *et al.*, 1986).

4.4 Xylose Utilisation Genes

Candidates for the genes involved xylose utilisation were identified from a genome search as previously described (Chapter 2). ORFs CBEI2383 and CBEI2384 were identified as putative *xylAB* genes, respectively. *xylAB* are key genes associated with the use of xylose in a number of other bacteria (Scheler *et al.*, 1991, Schmiedel *et al.*, 1997, Wilhelm and Hollenberg, 1984, Lawlis *et al.*, 1984, Feldmann *et al.*, 1992, Lokman *et al.*, 1991, Erlandson *et al.*, 2000, Wong *et al.*, 1991, Takeda *et al.*, 1998, Erbeznik *et al.*, 1998). XylAB work sequentially to convert xylose into xylulose and xylulose 5-phosphate, respectively and xylulose 5-phosphate then enters the pentose phosphate pathway.

4.4.1 Xylose isomerase (*xylA*)

The candidate XylA for *C. beijerinckii*, CBEI2383, was firstly compared to published characterised XylA sequences to investigate the level of sequence identity. CBEI2383 shared 71% identity with three species of *Thermoanaerobacterium* (Table 4-4). A radial tree constructed using the AA sequences of characterised XylA proteins and the CBEI2383 (Figure 4-12) revealed two groups of XylA sequences. A reason for the two distinct clusters of groups these was then sought. What was evident was there were differences in the sequence length and the genomic G/C content of the organism. One group consisted of organisms with a lower G/C content and generally longer AA sequence length, this included CBEI2383 and XylA of bacilli, thermoanaerobacteriae, lactobacilli and others (AA length of 435-465). Though *E. coli* was an exception in this group (333AA). These sequences shared 48-72% identity with CBEI2383. The second group contained XylA sequences from G/C rich organisms, including streptomycetes, *Actinoplanes*, *Arthrobacter* and *Thermus* species and had an AA length of 387-395AA. The members of this group shared 15-22% sequence identity with CBEI2383.

Residues involved in the activity of XylA, have been identified in *L. brevis* (Bor *et al.*, 1992). These consist of two His residues involved in catalytic activity and metal binding, respectively, and a 'WGGREG' motif is involved in xylose binding. This revealed the presence of the two residues and motif residues present in all the XylA sequences and CBEI2383, as shown on a multiple alignment constructed of a selection of XylA sequences (Figure 4-13). The two residues and the motif in the *C. beijerinckii* protein were present at the same or at residue positions of close proximity as that of *L. brevis*, (at 103AA, 273AA and 190-195AA, for each respectively). In *C. beijerinckii* the two his residues were at 101AA, 271AA and the WGGREG motif at 188-193AA.

Table 4-4: The sequence identity of CBEI2383 with published characterised XylA proteins

Organism	Identity (%)
Thermoanaerobacter pseudethanolicus	71
Thermoanaerobacterium saccharolyticum	71
Thermoanaerobacterium thermosaccharolyticum	71
Bacillus sp. LW2	68
Thermoanaerobacterium thermosulfurogenes	68
Bacillus megaterium	67
Bacillus subtilis	66
Thermotoga neapolitana	66
Bacillus licheniformis	63
Staphylococcus xylosus	63
Tetrageneococcus halophilus	58
Lactococcus lactis	54
Lactobacillus brevis	53
Lactobacillus pentosus	52
Salmonella enterica	50
Klebsiella pneumoniae	49
Escherichia coli	48
Thermus thermophilus	22
Actinoplanes sp ATTC 31351	20
Actinoplanes missouriensis	19
Arthrobacter sp. NRRL B3728	18
Streptomyces diastaticus	18
Streptomyces olivaceoviridis	18
Streptomyces rubiginosus	18
Streptomyces chorchorusii	17
Streptomyces murinus	17
Streptomyces violaceusniger	17
Streptomyces lividens	15

Identity is the percentage of identical shared AA between CBEI2383 and the XylA sequences in the table

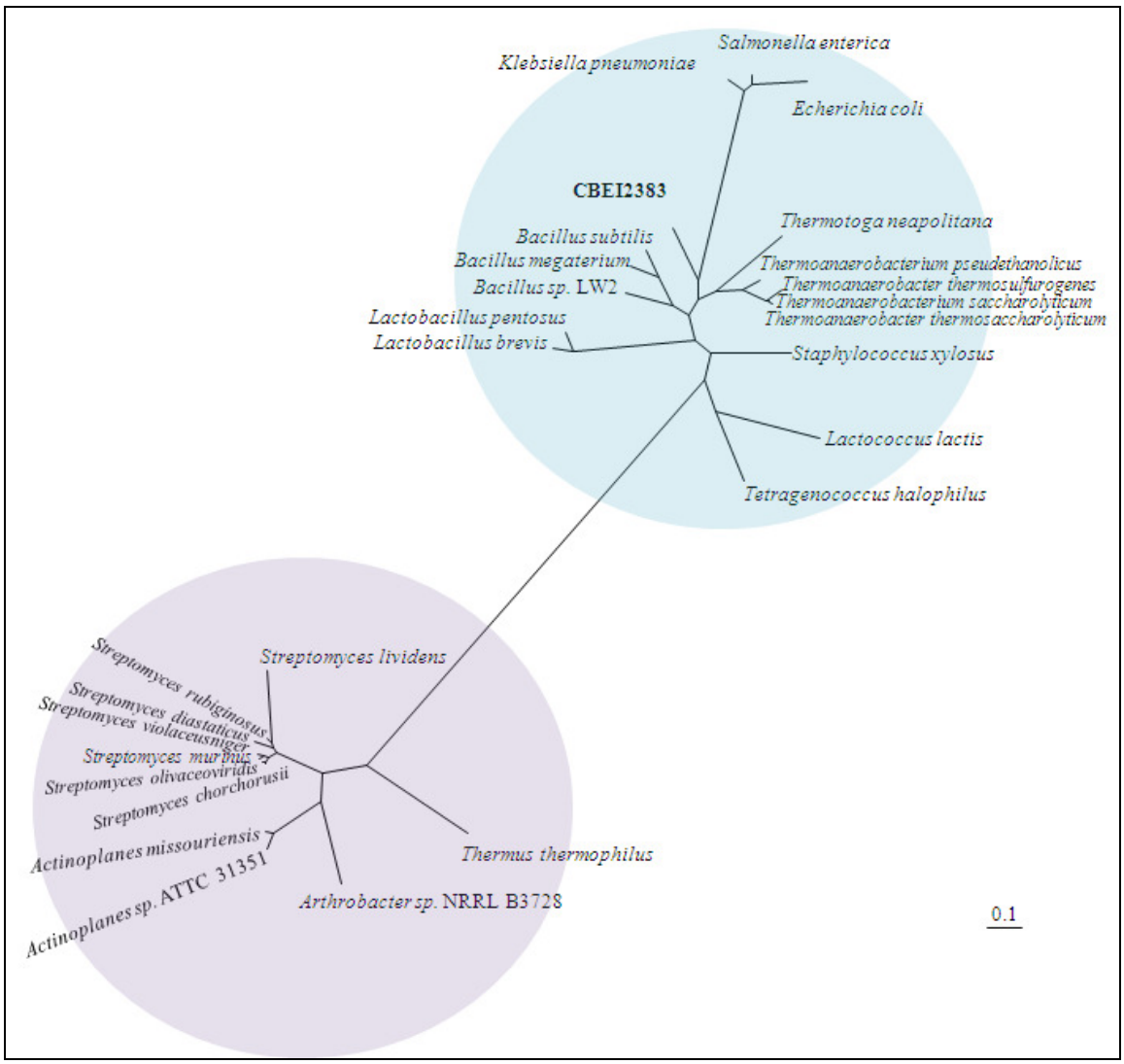


Figure 4-12: A radial tree of characterised XylA proteins and candidate XylA of *C. beijerinckii*

The AA sequences of confirmed XylA proteins and that of the *C. beijerinckii* XylA candidate were used to construct a radial tree, see text for details. Scale denotes 0.1 amino acid substitutions per site.

Actinoplanes sp.* ATCC 31351** (Saari *et al.*, 1987), ***A. missouriensis (Amore and Hollenberg, 1989), ***Arthrobacter sp* NRRL B3728** (Loviny-Anderton *et al.*, 1991), ***Bacillus sp.* LW2** (Liao *et al.*, 1995), ***B. licheniformis*** (Scheler *et al.*, 1991), ***B. megaterium*** (Schmiedel *et al.*, 1997), ***B. subtilis*** (Wilhelm and Hollenberg, 1984), ***C. beijerinckii*** CBEI2383, ***E. coli*** (Lawlis *et al.*, 1984), ***K. pneumoniae*** (Feldmann *et al.*, 1992), ***L. brevis*** (Bor *et al.*, 1992), ***L. pentosus*** (Lokman *et al.*, 1991), ***L. lactis*** (Erlandson *et al.*, 2000), ***S. typhimurium*** (Shamanna and Sanderson, 1979), ***S. xylosus*** (Sizemore *et al.*, 1992), ***S. albus*** (griseus) (Sanchez and Smiley, 1975), ***S. chibaensis* J-59** *(chorchorusii)* (Joo *et al.*, 2005), ***S. diastaticus*** (Wang *et al.*, 1994), ***S. rochei* S-41***(griseofuscus)* (Kikuchi *et al.*, 1990), ***S. lividens*** (Heo *et al.*, 2008), ***S. murinus*** (Rasmussen *et al.*, 1994), ***S. olivaceoviridis* E-86** (Kaneko *et al.*, 2001), ***S. rubiginosus*** (Wong *et al.*, 1991), ***S. violaceusniger*** (Drocourt *et al.*, 1988), ***T. halophilus*** (Takeda *et al.*, 1998), ***T. pseudethanolicus*** (Erbeznik *et al.*, 1998), ***T. thermosaccharolyticum*** (Meaden *et al.*, 1994), ***T. thermosulfurogenes*** (Haldrup *et al.*, 1998), ***T. saccharolyticum*** (Lee *et al.*, 1993), ***T. neapolitana*** (Vieille *et al.*, 1995), ***T. thermophilus*** (Dekker *et al.*, 1991).

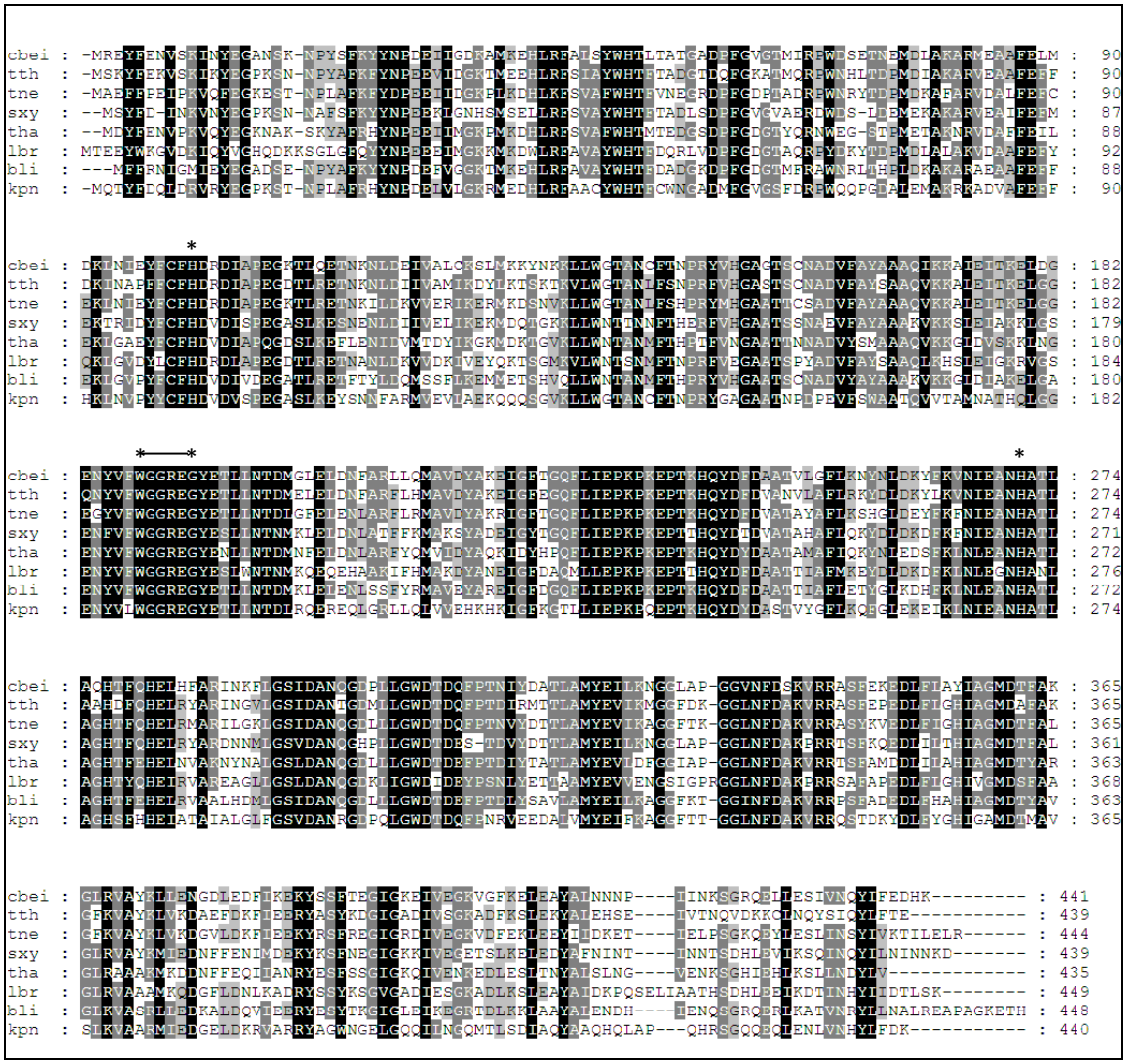


Figure 4-13: A Multiple alignment of XylA proteins and the XylA candidate of *C. beijerinckii*

A multiple alignment was constructed using AA sequences of a selection of confirmed XylA proteins, most similar in terms of sequence similarity from BLAST analysis to the candidate XylA of *C. beijerinckii* (CBEI2383), see text for details. **Black** shading denotes 100% sequence similarity (similarity shared by all sequences in the multiple alignment at the shaded point), **Grey shading with white text**- 80%, **grey shading with black text** – 60% sequence similarity. The presence of conserved residues indicative of XylA proteins, which have been identified in XylA of *L. brevis* (Bor *et al.*, 1992). These are indicated by the asterisks above the sequences and include an H residue thought to be involved in catalytic activity (103AA), 'WGGREG' motif involved in xylose binding (190-195AA), and an H residue involved in metal binding at the active site (273AA).

bli- *B. licheniformis* (Scheler *et al.*, 1991), **cbei**- *C. beijerinckii* (CBEI2383), **kpn**- *K. pneumoniae* (Feldmann *et al.*, 1992), **lbr**- *L. brevis* (Bor *et al.*, 1992), **sxy**- *S. xylosus* (Sizemore *et al.*, 1992), **tha**- *T. halophilus* (Takeda *et al.*, 1998), **tne**- *T. neapolitana* (Vieille *et al.*, 1995), **tth**- *T. thermophilus* (Dekker *et al.*, 1991)

4.4.2 Xylulokinase (*xylB*)

The final protein involved in xylose utilisation prior to the pentose phosphate pathway, XylB (putatively CBEI2384) was then compared to characterised XylB sequences (Table 4-5). *T. pseudethanolicus* had the highest sequence identity with CBEI2384 of 50%. A radial tree was then constructed to compare the all the AA sequences collectively with one and another. The tree reflected the results of Table 4-5 as CBEI2384 was clustered with XylB of *T. pseudethanolicus*. Distinct clusters of XylB were evident dependent upon the Gram identity of the organism the XylB belonged to (Figure 4-14). One cluster contained XylB sequences of Gram negative organisms and another of XylB sequences of Gram positive organisms (with the exception of the XylB of *L. brevis*, *T. pseudethanolicus* and CBEI2384). An alignment of the most similar sequence to CBEI2384 and the XylB of *T. pseudethanolicus* (Figure 4-15) highlighted the level of sequence similarity between the two sequences.

Table 4-5: The sequence identity of CBEI2384 with published characterised XylB proteins

Organism	Identity (%)
Thermoanaerobacter pseudethanolicus	50
Lactobacillus pentosus	39
Bacillus subtilis	38
Tetrageneococcus halophilus	38
Lactococcus lactis	37
Staphylococcus xylosus	37
Escherichia coli	35
Klebsiella pneumoniae	33
Bacillus megaterium	32
Bacillus licheniformis	29
Streptomyces rubiginosus	24
Streptomyces lividens	21
Lactobacillus brevis	6

Identity is the percentage of identical shared AA between CBEI2384 and the XylB sequences in the table

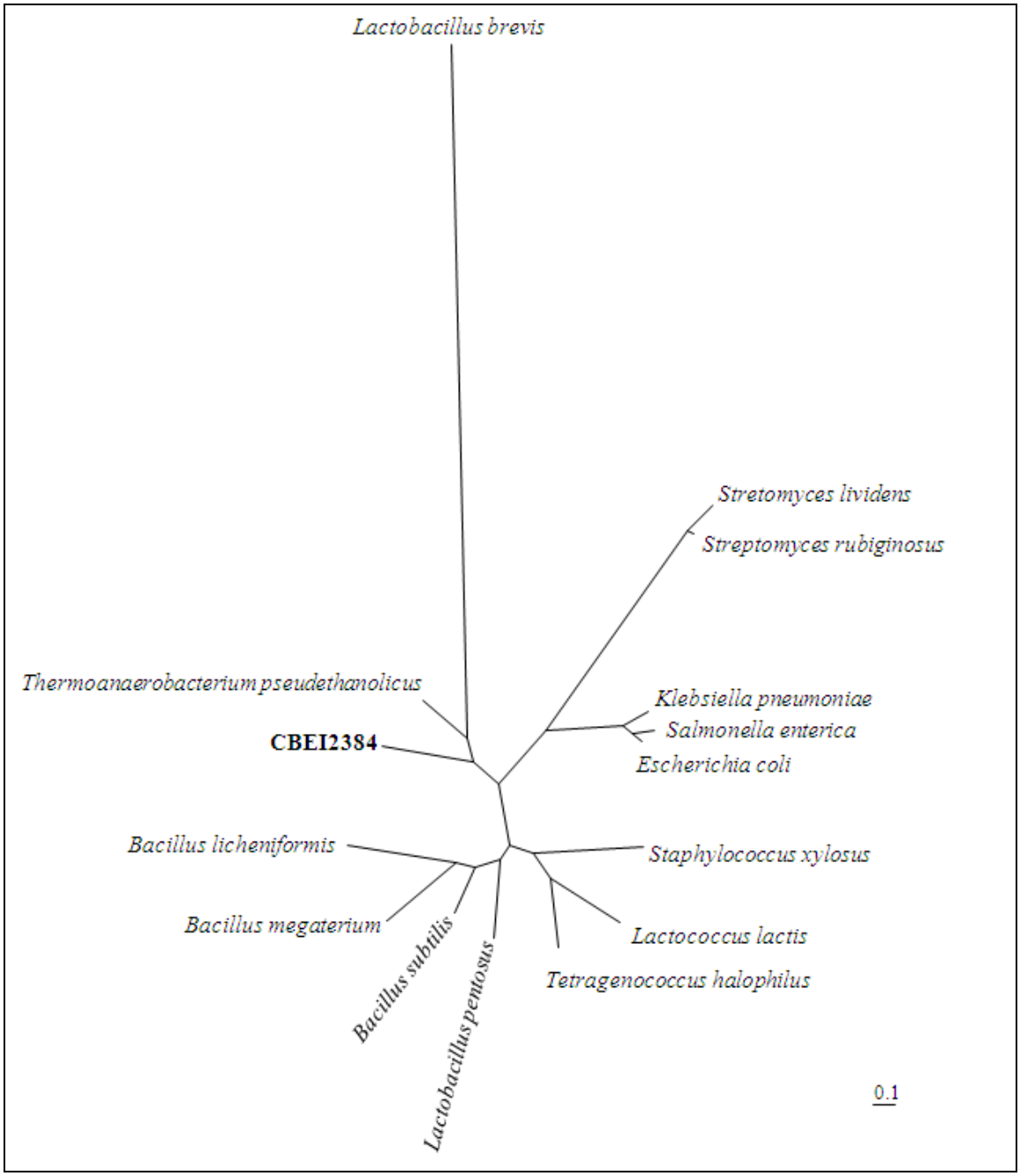


Figure 4-14: A radial tree of characterised XylB proteins with the XylB candidate of *C. beijerinckii*

A radial tree was constructed using the AA sequences of confirmed XylB proteins and the XylB candidate of *C. beijerinckii*, see text for details. Scale denotes 0.1 amino acid substitutions per site.
B. licheniformis (Scheler *et al.*, 1991), ***B. megaterium*** (Schmiedel *et al.*, 1997), ***B. subtilis*** (Wilhelm and Hollenberg, 1984), ***C. beijerinckii*** CBEI2384, ***E. coli*** (Lawlis *et al.*, 1984), ***K. pneumoniae*** (Feldmann *et al.*, 1992), ***L. pentosus*** (Lokman *et al.*, 1991), ***L. lactis*** (Erlandson *et al.*, 2000), ***S. typhimurium*** (Shamanna and Sanderson, 1979), ***S. xylosus*** (Sizemore *et al.*, 1992), ***S. rubiginosus*** (Wong *et al.*, 1991), ***T. halophilus*** (Takeda *et al.*, 1998).

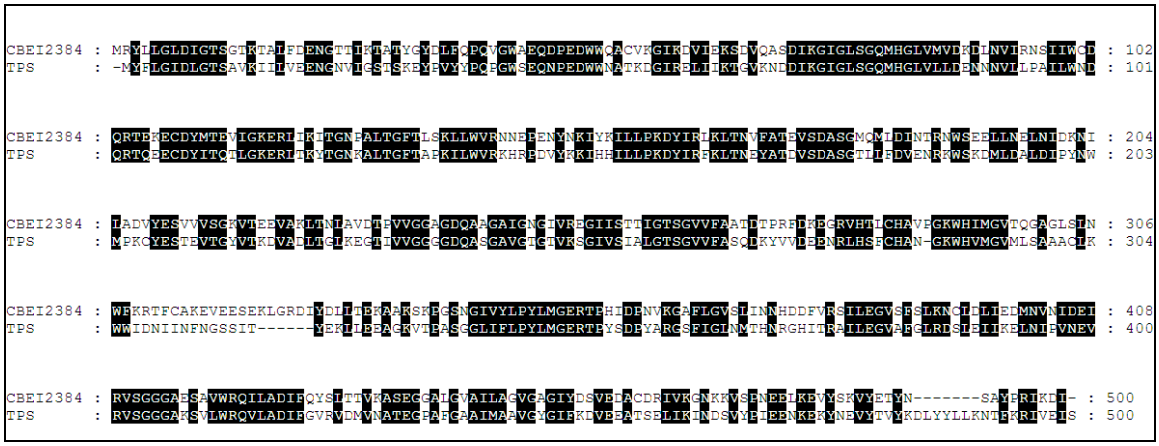


Figure 4-15: An alignment of the XylB of *T. pseudethanolicus* with the *C. beijerinckii* XylB candidate

A multiple alignment was constructed using the AA sequences of the putative XylB of *C. beijerinckii* and the XylB *T. pseudethanolicus* proteins. **Black** shading denotes 100% sequence similarity between the two sequences. **cbei**- *C. beijerinckii* CBEI2384, **tps**- *Thermoanaerobacter pseudethanolicus* (Erbeznik *et al.,* 1998)

4.5 Candidate Xylanases and Xylosidases in *C. beijerinckii*

As previously described there are two types of enzymes which can hydrolyse xylans, xylanase and xylosidase. Xylanase and xylosidase candidates were identified in the genome and are separated by five putatively unrelated genes (Figure 4-5). The deduced AA sequences of the encoded enzymes were used in further analyses. First of all a radial tree was created of the AA sequences of the putative enzymes in *C. beijerinckii* and known xylanase (XynA) and xylosidase (XynB) sequences of various organisms (Figure 4-16). The tree shows four clusters of sequences; two groups for XynB enzymes and two groups of XynA enzymes. CBEI3041, the putative xylanase of *C. beijerinckii*, is found clustered with XynA sequences with species of bacilli and had the most sequence similarity to the xylanase of *B. subtilis* sharing 72% identity (Table 4-6). Next, to observe sequence similarity further, a multiple alignment of sequences was constructed from the AA sequences of CBEI3041 and a very high level of sequence identity was observed between these three sequences (Figure 4-17). For CBEI3047, the putative xylosidase, was clustered with the XynB sequence of *C. saccholyticus* sharing 26% identity (Table 4-7). There is a substantial difference in the length of CBEI3047 and other XynB sequences. All of the characterised xylosidases are around 500 AA in length with the exception of *Prievotella ruminicola* which contains 319 AA in length, while CBEI3047 is much larger at 841 AA (Figure 4-18). A considerable number of gap insertions were required to accommodate the length differences between CBEI3047 and xylosidase of *C. saccholyticus*. The CBEI3047 appears to have extra N-terminal domain. Upon BLAST analysis there appeared to be an AraC-like ligand binding domain between 11-146 AAs and a XynB domain from 351-836, within the latter was a family 39 glycosyl hydrolase catalytic core which is associated with β-xylosidases (Czjzek *et al.*, 2004). There were small regions of identity shared between all three sequences, but the extent of gap insertions makes it difficult to assess if these enzymes show any meaningful homology.

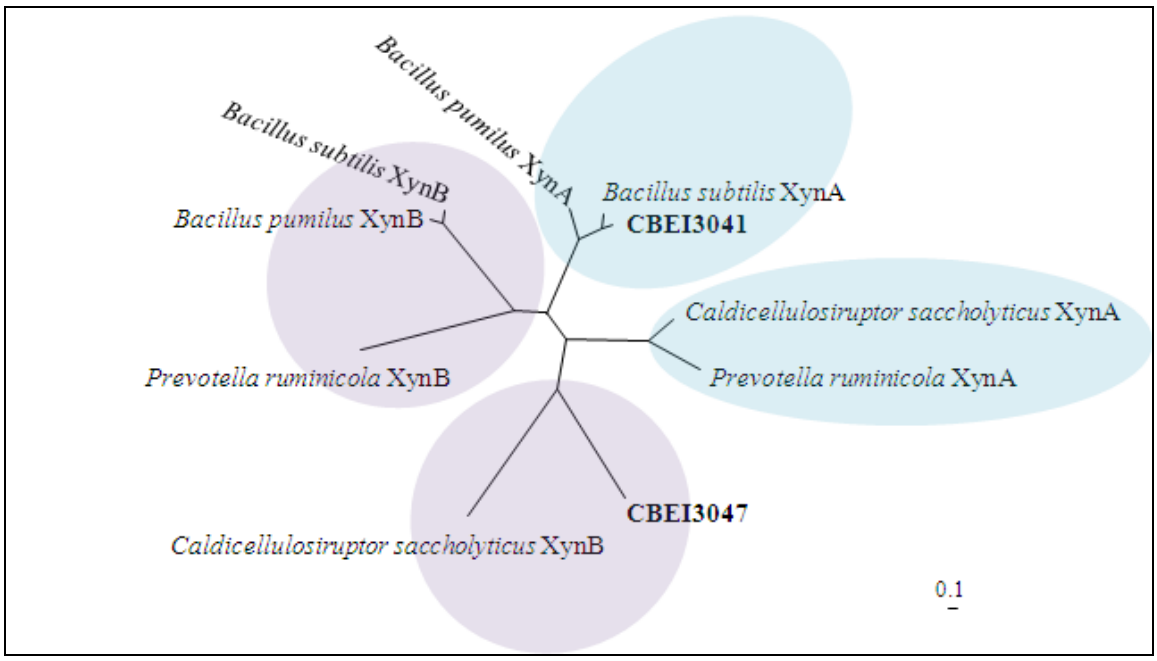


Figure 4-16: A radial tree of characterised XynAB proteins with the XynAB candidates of *C. beijerinckii*

A radial tree was constructed using the AA sequences of confirmed XynAB proteins and the XynAB candidates of *C. beijerinckii*, see text for details. Scale denotes 0.1 amino acid substitutions per site.
Bacillus pumilus (X00660/ X05793); ***Bacillus subtilis*** (Z34519/ U66480); ***Caldicellulosiruptor saccholyticus*** (AF005383); **CBEI**- ***Clostridium beijerinckii*** (number denotes the locus tag); ***Prevotella ruminicola*** (Z49241). Accession numbers are quoted in parenthesis.

Table 4-6: The sequence identity of CBEI3041 with published characterised XynA proteins

Organism	Identity (%)
Bacillus subtilis	72
Bacillus pumilus	43
Prevotella ruminicola	9
Caldicellulosiruptor saccholyticus	6

Identity is the percentage of identical shared AA between CBEI3041 and the XynA sequences in the table

Table 4-7: The sequence identity of CBEI3047 with published characterised XynB proteins

Organism	Identity (%)
Caldicellulosiruptor saccholyticus	26
Bacillus subtilis	6
Bacillus pumilus	6
Prevotella ruminicola	3

Identity is the percentage of identical shared AA between CBEI3047 and the XynB sequences in the table

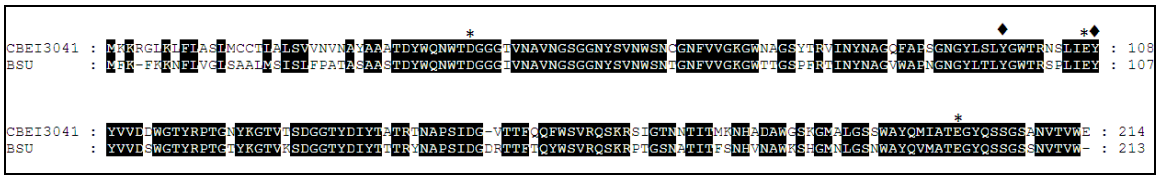


Figure 4-17: An alignment of a characterised xylanase with the *C. beijerinckii* candidate

The AA sequences of characterised xylanases with the greatest sequence identity with the candidate XynA gene of *C. beijerinckii* (CBEI3041) and the xylanase of *B. subtilis* were used to construct a multiple alignment using ClustalW2 and GeneDoc. **Black** shading indicates AA identity shared between two of the sequences. As can be observed, a substantial level of identity exists between all three sequences indicative of the same or similar function. * denotes positions of residues associated with the active site function. ♦ denotes positions likely to be involved in substrate binding (Wakarchuk *et al.*, 1994). **BSU**- *B. subtilis* (accession number: Z34519), **CBEI3041**- *C. beijerinckii* CBEI3041

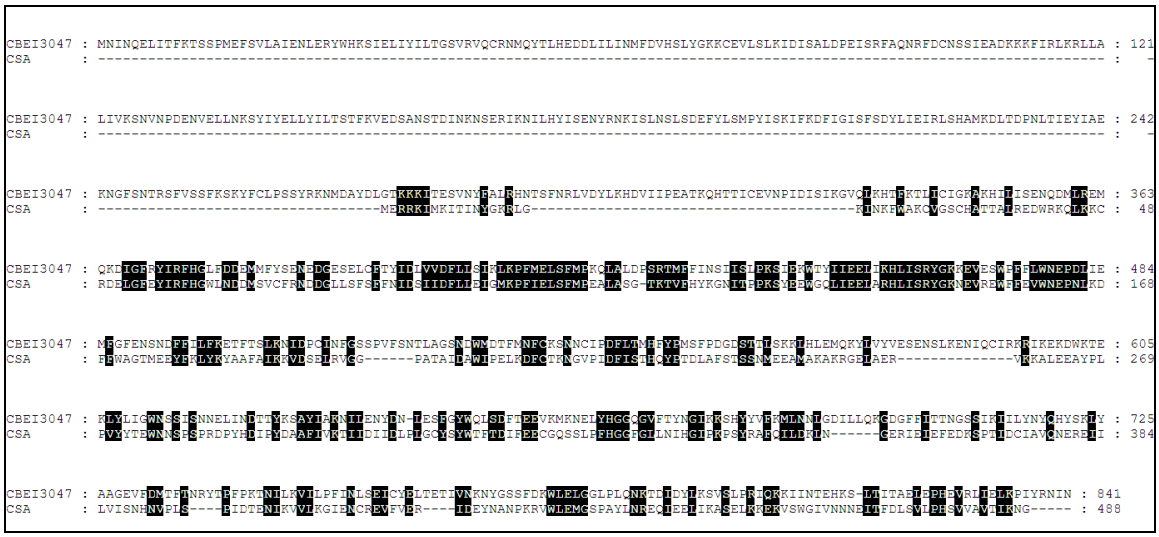


Figure 4-18: An alignment of a characterised xylosidase with the *C. beijerinckii* candidate

The AA sequences of a characterised xylosidase with the greatest sequence identity with the candidate XynB gene of *C. beijerinckii* (CBEI3047) was used to construct an alignment using ClustalW2 and GeneDoc. **Black** shading indicates AA identity shared across the sequences. **CBEI3047**- *C. beijerinckii* CBEI3047, **CSA**- *Caldicellulosiruptor saccharolyticus* (accession number: AF005383)

4.6 Transporters

A genome search identified two candidate ABC transporter ORFs for arabinose (CBEI4448-50 and CBEI4459-61) and one for xylose (CBEI2380-82).

ABC transporters are composed of three main units which are coded for on separate genes (Schneider, 2001). There is a substrate-binding protein, found on the outside of the membrane which has a role in binding and presenting the substrate to a transmembrane protein, or permease, which is coupled to the third unit, the ATP-binding protein. When the substrate is presented to the permease, ATP is hydrolysed by the ATP-binding protein and this facilitates the translocation of the substrate across the membrane to the inside of the cell.

The AA sequences of candidate ABC transporter components of pentose sugars in *C. beijerinckii* and published characterised ABC transporter proteins were used to construct a radial tree (Figure 4-19). Three distinct clusters of similarity were apparent for each ABC transporter component, one for solute-binding proteins, one for ATPases and one for permeases. For the putative arabinose ABC transporter components, there were two candidate groups, CBEI4448-4450 and CBEI4459-61. CBEI4448, 4459 and 4460 showed most similarity with permeases, CBEI4448 had the most percentage similarity to xylose binding protein (XylF) of *T. pseudethanolicus* and the latter two with *B. subtilis* ribose binding protein (RbsC) (Table 4-8). CBEI4449 and 4461 with ATPases, having most similarity to xylose ATPase (XylG) of *T. pseudethanolicus* (Table 4-8); and CBEI4450 with substrate binding proteins (which had most sequence similarity to XylF of *E. coli*). The results suggest that only of one set of genes, encode all the components necessary for an arabinose ABC transporter, from ORF CBEI4448-50; and the other has an incomplete set, CBEI4459-61 encoding two candidate permeases but lacking a candidate substrate binding protein.

A candidate for each component a xylose ABC transporter was identified. CBEI2380 has most similarity to solute-binding proteins, CBEI2381 to ATPases and CBEI2382 to permeases, all three had most sequence identity with the AA sequences of XylF, XylG and xylose permease (XylH) of *T. pseudethanolicus* respectively (Table 4-8).

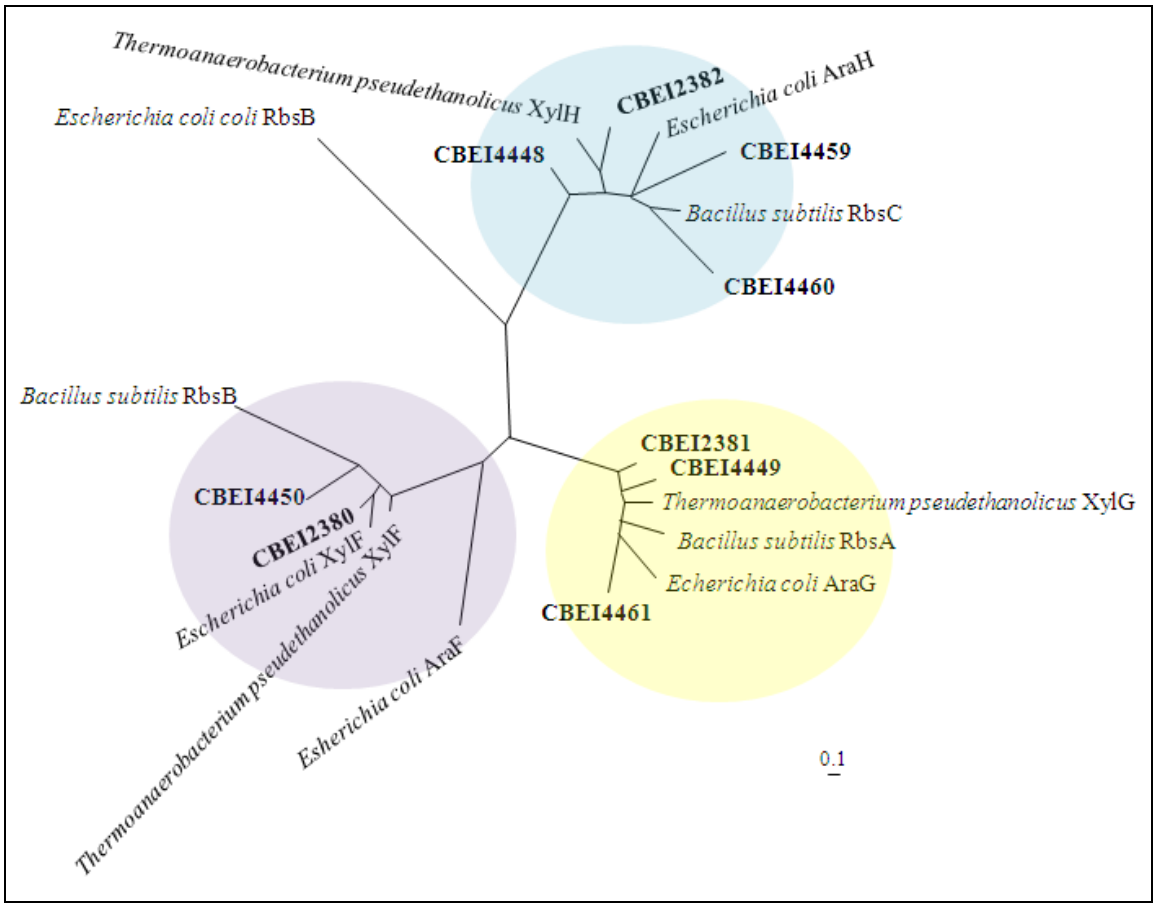


Figure 4-19: A radial tree of characterised pentose sugar ABC transporter proteins and candidates of *C. beijerinckii* putatively involved in pentose sugar transport

The AA sequences of candidate and confirmed ABC transporters were used to construct a radial tree. Scale denotes 0.1 amino acid substitutions per site.
Candidate AA sequences of *C. beijerinckii* are labelled with locus tag CBEIXXXX, published characterised sequences are labelled with the organism and the protein.
AraF- arabinose-binding protein, **RbsB**- Ribose-binding protein, **XylF**- xylose-binding protein, **AraG/XylG**- ATPase, **AraH/XylH**- permease. ***E. coli*** K12 (Horazdovsky and Hogg, 1989, Sumiya and Henderson, 1989, Groarke *et al.*, 1983), ***T. ethanolicus*** ATCC 33223 (Erbeznik *et al.*, 2004).

Table 4-8: Sequence identities of pentose sugar ABC transporters with the putative ABC transporters of *C. beijerinckii*

Sequence	CBEI2380	CBEI2381	CBEI2382	CBEI4448	CBEI4449	CBEI4450	CBEI4459	CBEI4460	CBEI4461
Escherichia coli (AraF)	10	13	2	6	4	12	4	7	8
Escherichia coli (AraG)	7	39	8	3	39	4	7	7	37
Escherichia coli (AraH)	2	4	32	32	4	3	25	25	7
Thermoanaerobacter pseudethanolicus (XylF)	**62**	7	2	8	12	34	2	2	7
Thermoanaerobacter pseudethanolicus (XylG)	3	**53**	10	4	**54**	7	7	9	**39**
Thermoanaerobacter pseudethanolicus (XylH)	5	9	**47**	**40**	4	4	24	24	7
Bacillus subtilis (RbsA)	7	43	8	6	44	9	4	3	37
Bacillus subtilis (RbsB)	25	7	4	3	9	18	2	0	11
Bacillus subtilis (RbsC)	5	11	37	35	8	4	**31**	**34**	6
Escherichia coli (RbsB)	3	1	2	2	1	2	5	2	1
Escherichia coli (XylF)	61	6	2	7	11	**37**	3	4	8
CBEI2380	-	6	2	5	6	**35**	2	2	8
CBEI2381	6	-	5	4	**52**	6	1	6	38
CBEI2382	2	5	-	**36**	3	3	**27**	**25**	5
CBEI4448	5	4	**36**	-	6	3	**27**	23	8
CBEI4449	6	**52**	3	6	-	7	1	10	**40**
CBEI4450	**35**	6	3	3	7	-	1	2	5
CBEI4459	2	1	27	27	1	1	-	22	5
CBEI4460	2	6	25	23	10	2	22	-	2
CBEI4461	8	38	5	8	40	5	5	2	-

Identity is the percentage of identical shared AA between putative ABC transporters and published characterised ABC transporter sequences. **Red bold** – denotes the highest sequence identity of the putative ABC transporter proteins of *C. beijerinckii* with published characterised ABC transporter proteins. **Black bold** – denotes the highest identity share amongst the putative ABC transporters of *C. beijerinckii*. **Ara** – Arabinose, **Xyl** – Xylose **Rbs** – Ribose. **Suffix: B/F**- Sugar binding protein; **A/G**-ATPase,**C/H**- Permease. **CBEI**- *C. beijerinckii* (suffix number denotes the locus tag)

The binding site of ATP on ATP-binding proteins is highly conserved across both prokaryotes and eukaryotes. Known as the Walker motif, it provides a cleft in which the ATP is positioned and is hydrolysed. The cleft is made by two domains A and B which are linked by a helix of approximately one hundred amino acids containing an ABC signature (Fath and Kolter, 1993, Schneider, 2001). The Walker motifs and ABC signatures of AraG of *E. coli* and XylG of *T. ethanolicus* share a considerable level of sequence similarity at similar AA sequence positions with putative AraG and XylG sequences of *C. beijerinckii* (Figure 4-20). However the motifs are not perfectly conserved. Some AAs are substituted with AAs of similar properties and few similar properties.

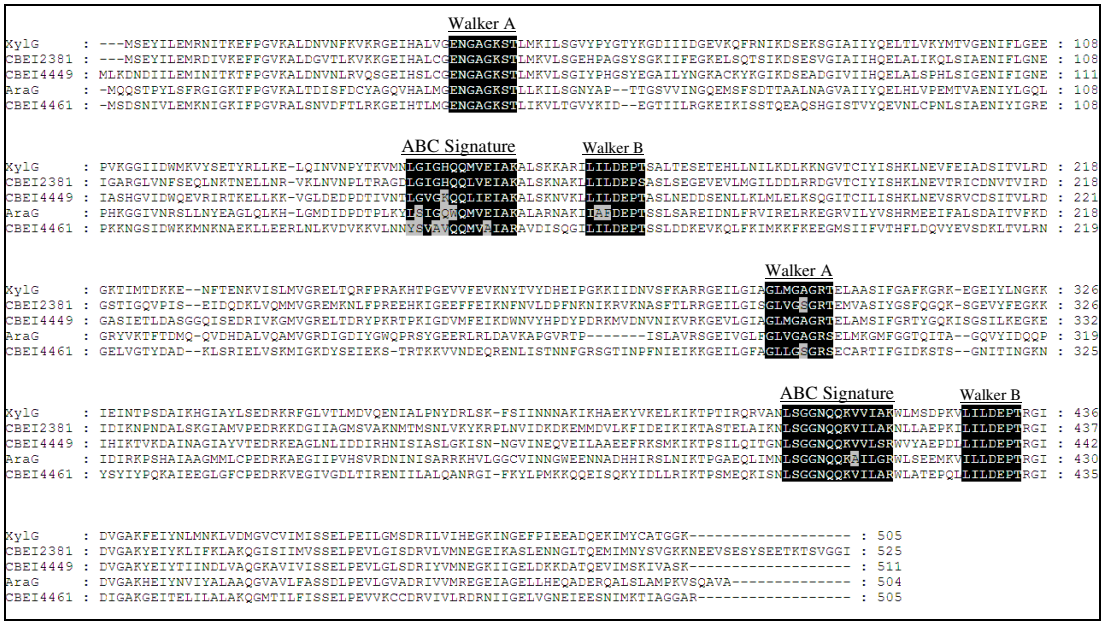


Figure 4-20: Proposed Walker motifs and ABC signatures of *C. beijerinckii* candidates in comparison with characterised AraG of *E.coli*

The AA sequences of candidates and AraG proteins were used to construct a multiple alignment. The predicted Walker motifs and ABC signatures are marked. Residues which are the same or have the same AA properties are highlighted in black and residue differences with different AA properties are highlighted in grey. **AraG**- *E. coli* K12 AraG sequence (Horazdovsky and Hogg, 1989), **CBEI4449/4461**- Locus tag numbers of AraG candidates of *C. beijerinckii*, **CBEI2381**- XylG candidate of *C. beijerinckii,* **XylG** – XylG of *T. ethanolicus* (Erbeznik *et al.*, 1998).

As described earlier the Hidden Markov Model (HMM) can be used to predict the structure of membrane-bound proteins, based on the AA sequences. Within an ABC transporter group there was only one transmembrane protein, for arabinose, AraH and xylose, XylH. TMHMM was used to analyse for and predict the helical structure of the putative AraH and XylH of *C. beijerinckii*. Using the results of the TMHMM were presented on sequence alignments, to compare the sequence positions of the transmembrane regions, then TMRPres2D program was used to produce a two-dimensional model of the structures. The other putative ABC transporter proteins, AraFG and XylFG, were also analysed but no helical structure was observed using the model in these sequences (results not shown). What was apparent was there appears to be an additional >50 AAs on the sequence of CBEI4448 (Figure 4-21) in comparison to the other sequences analysed. When comparing the sequences of the alignment with the two dimensional model (Figure 4-22), there was a noticeably larger extracellular region from 124-180 AA and therefore it would appear there is one less transmembrane

spanning domain in this region in comparison to the AraH of *E. coli*. CBEI4459 also had one less transmembrane spanning helix. Of the three putative AraH sequences of *C. beijerinckii*, CBEI4460, had the same number of transmembrane spanning domains as the AraH of *E. coli* (Figure 4-22).

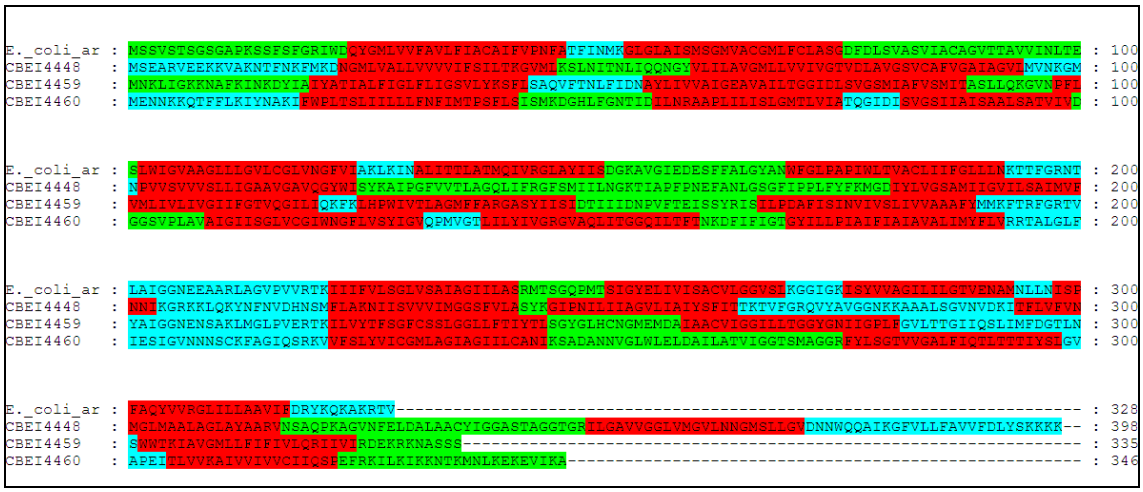


Figure 4-21: A multiple alignment of a characterised AraH sequence and the candidate AraH sequences of *C. beijerinckii*

A multiple alignment was produced based on the results of TMHMM to compare the transmembrane structure of the AraH of *E. coli* and putative AraH proteins of *C. beijerinckii*. The putative transmembrane regions (■) the cytoplasmic regions (■); extracellular regions (■). **CBEIXXXX** denotes the *C. beijerinckii* candidate sequence locus tag; **E.coli** – *E. coli* AraH (Horazdovsky and Hogg, 1989)

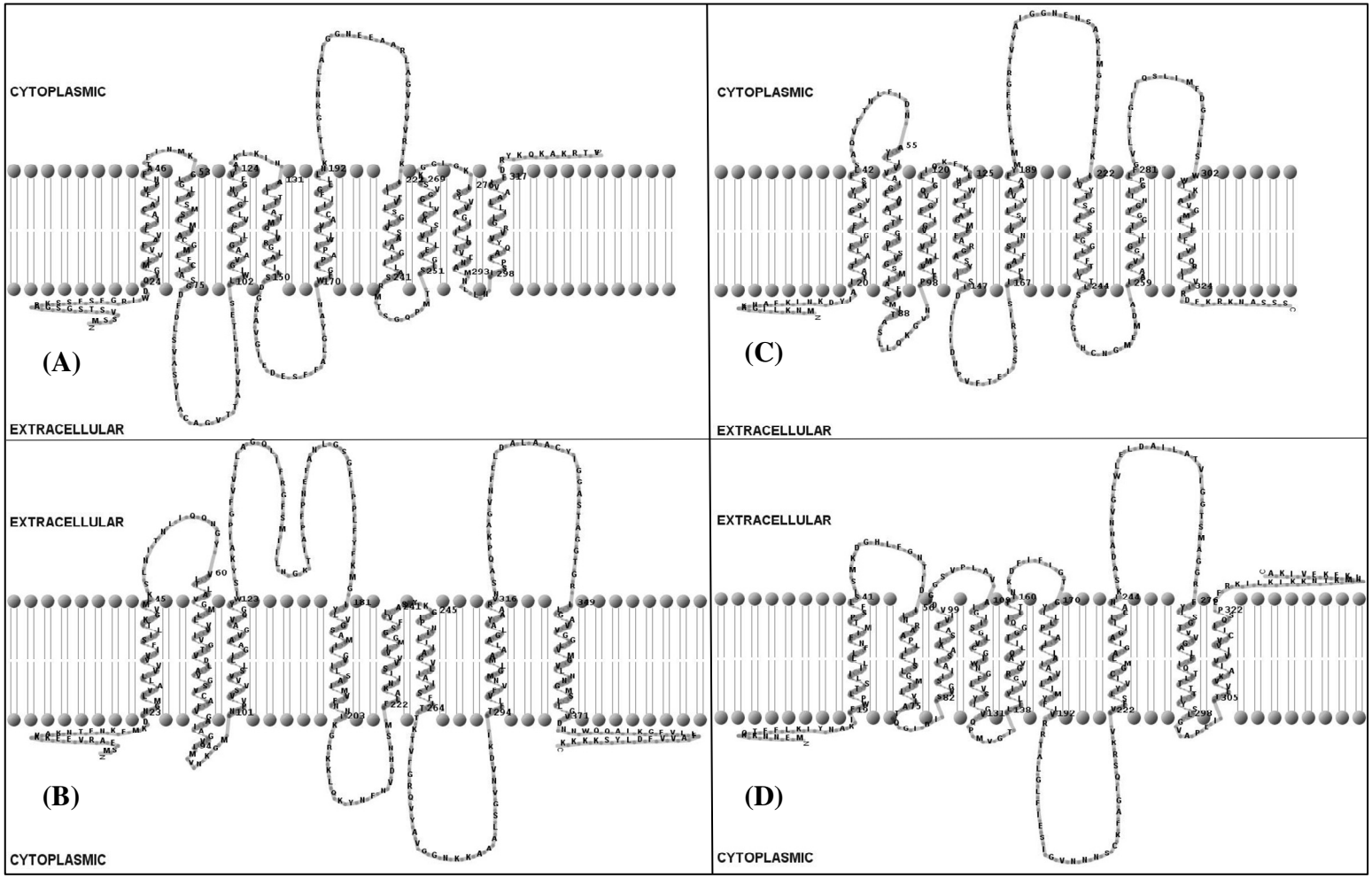


Figure 4-22: Modelling of transmembrane protein AraH proteins and *C. beijerinckii* candidates
The hidden Markov model was applied to all candidate ABC transporter associated AA sequences of *C. beijerinckii* and compared to the published characterised AraH of *E. coli*. TMRPres2D was used to produce a 2D model of the structure.
(A) *E. coli* K12 AraH (Horazdovsky and Hogg, 1989); **(B)** CBEI4448; **(C)** CBEI4459; **(D)** CBEI4460

Out of the three xylose ABC transporter candidates, CBEI2380-2382, only CBEI2382 was predicted as being transmembrane protein. The predicted number of transmembrane helices of CBEI2382 was eleven in comparison to ten in XylH of *T. pseudethanolicus* (Figure 4-23& Figure 4-24), there appeared to be an extra transmembrane spanning region between the 328-348 AA region for CBEI2382 and an extra-long extracellular region for *T. pseudethanolicus* in the same area (Figure 4-23). Both AA sequences were of a similar length (Figure 4-24).

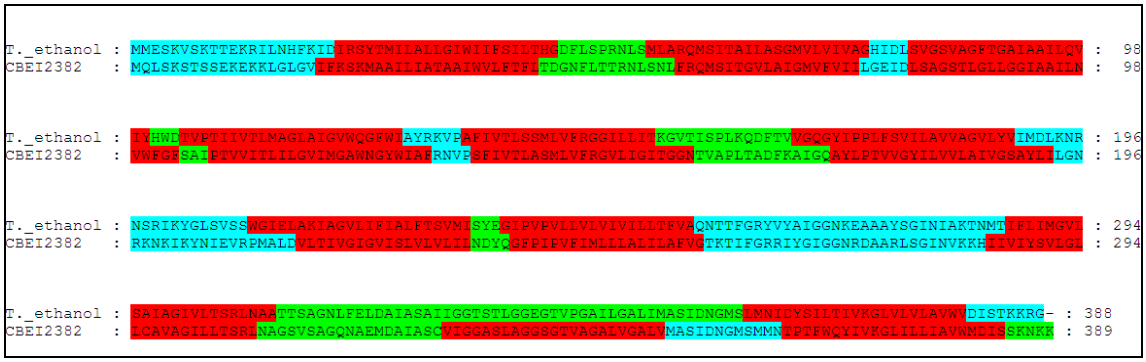


Figure 4-23: An alignment of a characterised XylH sequence and the candidate XylH sequence of *C. beijerinckii*

An alignment was produced based on the results of TMHMM to compare the transmembrane structure of the XylH of *T. pseudethanolicus* and putative XylH protein of *C. beijerinckii*. The putative transmembrane regions (■) the cytoplasmic regions (■); extracellular regions (■). **CBEIXXXX** denotes the *C. beijerinckii* candidate sequence locus tag; **T. ethanol** – *T. pseudethanolicus* XylH (Erbeznik *et al.*, 2004)

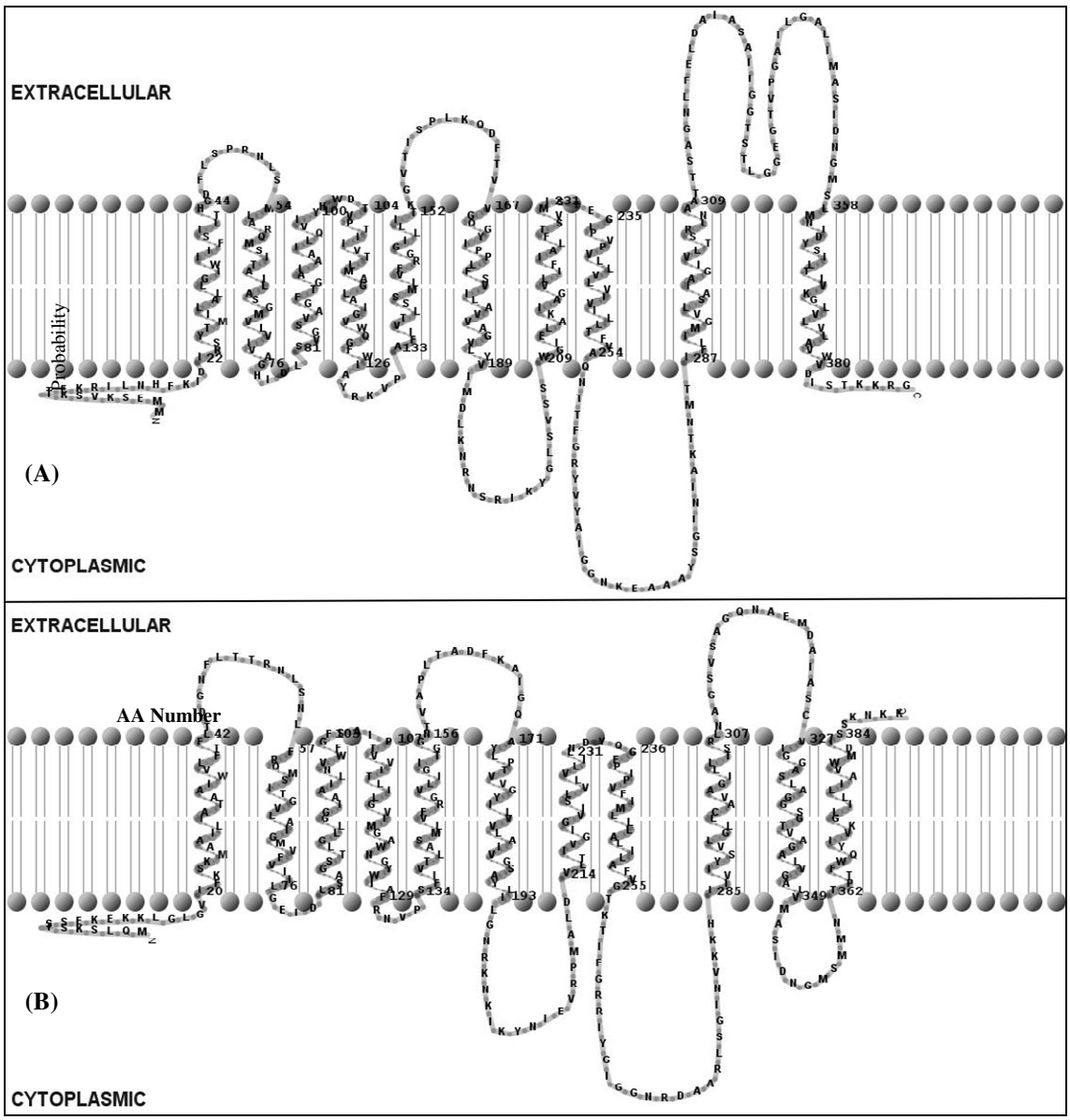


Figure 4-24: Modelling and comparison of a known transmembrane protein and a *C. beijerinckii* candidate

The hidden Markov model was used to predict the presence of transmembrane helices and then TMRPres2D was used to produce a 2D model of the structure of a confirmed XylH protein of *T. ethanolicus* and the XylH candidate of *C. beijerinckii*, CBEI2382, using the AA sequences of each. **(A)** *T. pseudethanolicus* XylH (Erbeznik *et al.*, 2004), **(B)** *C. beijerinckii* locus tag CBEI2382.

4.7 Regulation

Regulation of pentose utilisation is distinct between Gram negative and Gram positive organisms, studied thus far. The arabinose operon in Gram positive organisms is negatively regulated by the repressor AraR, binding to operator regions of the arabinose operon in the absence of arabinose, forming a loop in the DNA. In the presence of arabinose, arabinose is thought to bind to araR causing a conformational change and thus preventing its binding to DNA (Franco *et al.*, 2006). In Gram negative organisms arabinose transcription regulator (AraC) functions as a activator/positive regulator and a repressor, in the absence of arabinose, blocking transcription of the arabinose operon by forming a DNA loop between operator and initiator sites. The presence of arabinose stimulates transcription by binding to two initiator sites, therefore AraC behaves as a positive regulator (Kawaguchi *et al.*, 2009).

To predict a mechanism of regulation for arabinose and xylose utilisation in *C. beijerinckii*, the AA sequences of characterised pentose sugar utilisation gene regulators that act positively and/or negatively, were used to compare to the candidates of *C. beijerinckii*. The radial tree produced revealed three main clusters of similarity between the different pentose sugar regulatory genes, one for xylose repressor (XylR), another for AraC and XylS (xylose transcription regulator) and finally one for both arabinose repressor (AraR) and ribose regulator (RbsR) (Figure 4-25). Both CBEI4456 and CBEI2385 showed the least similarity with AraC and XylS, a longer branch length was evident between the two. The shorter branch length and clustering with AraR and RbsR demonstrated CBEI4456, is more likely to be AraR than AraC. CBEI4456 shared the highest sequence identity of 40% (Table 4-9), with AraR of *B. subtilis*. An alignment highlighted an extensive amount of sequence similarity between the two, therefore suggestive of a common or related function between these two proteins (Figure 4-25). All DNA binding proteins carry a helix-turn-helix motif. This has two helices, one for the interaction with DNA molecules and the other for stabilisation (Religa *et al.*, 2007). Indeed, a putative DNA binding site (helix-turn helix motif) was identified from the BLAST analysis of CBEI4456 and is indicated in Figure 4-26.

CBEI2385 had the greatest similarity to XylR sequences (Figure 4-25), it shared 28% sequence identity with the XylR sequences of *B. subtilis, B. megaterium* and *T. halophilus* (Table 4-10). The greatest similarity was observed with the XylR sequence of *B. subtilis* (46% sequence similarity), the sequence similarity was evenly distributed

throughout the sequences with some long regions of consecutive identical AAs or AAs with similar chemical properties being conserved (Figure 4-27).

These results suggest a negative regulatory mechanism for the use of arabinose and xylose by *C. beijerinckii*.

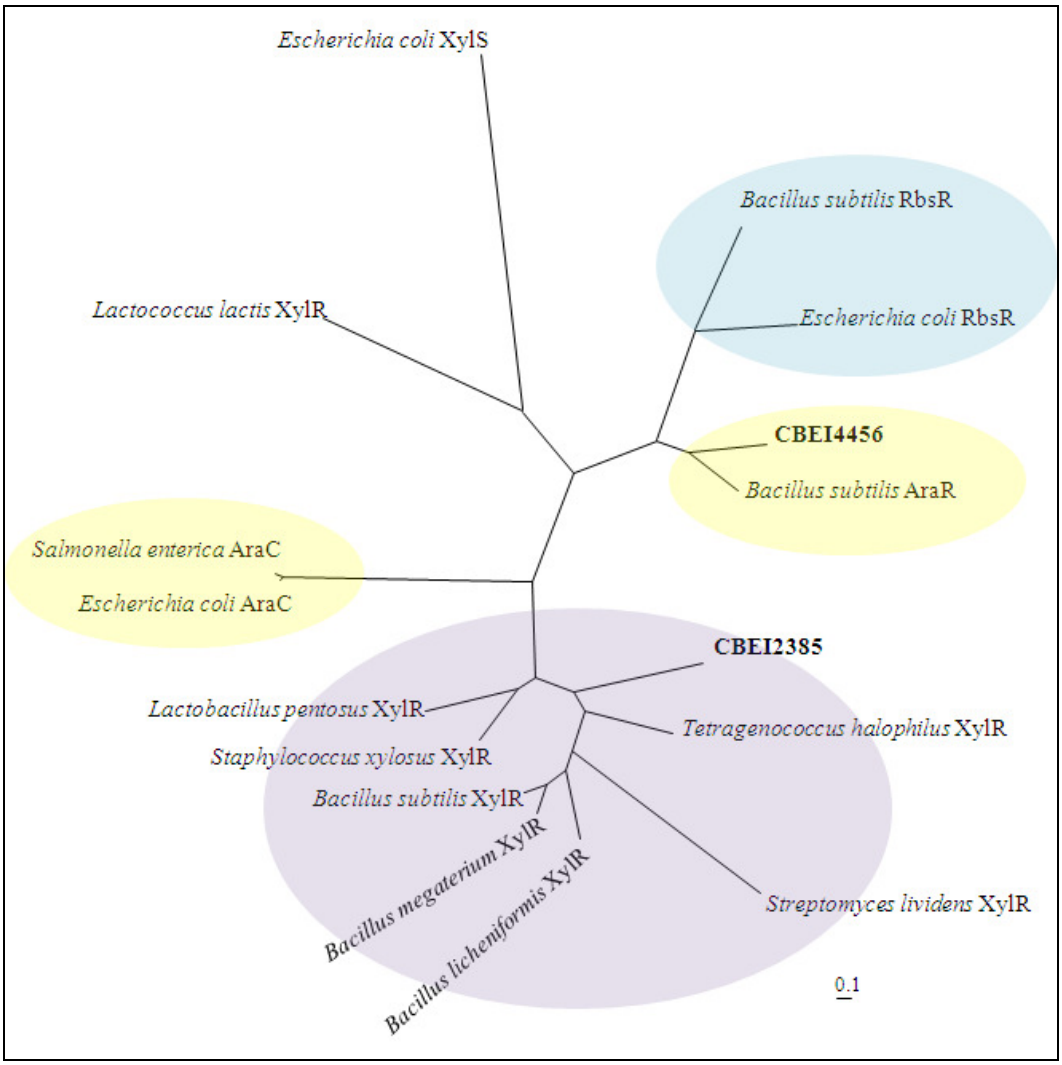


Figure 4-25: Radial Tree of characterised pentose sugar repressors and the putative pentose repressors of *C. beijerinckii*

The AA sequences of a number of proteins with functions confirmed in pentose sugar regulation were used along with candidates putatively involved in pentose sugar utilisation in *C. beijerinckii*, see text for details. Scale denotes 0.1 amino acid substitutions per site.

AraC – arabinose regulator/activator, **AraR**- arabinose repressor, **RbsR**- ribose repressor, **XylR**- xylose repressor, **XylS**- xylose regulator/activator. **CBEIXXXX** denotes the locus tag of *C. beijerinckii*.

B. licheniformis (Scheler *et al.*, 1991), ***B. megaterium*** (Rygus *et al.*, 1991), ***B. subtilis*** **AraR** (Sa-Nogueira and Mota, 1997) **RbsR** (Woodson & Devine, 1994), **XylR** (Kreuzer *et al.*, 1989), ***C. glutamicum*** (Kawaguchi *et al.*, 2009), ***E. coli*** **AraC** (Wallace *et al.*, 1980) **XylS** (Song and Park, 1997), **RbsR** (Mauzy and Hermodson, 1992), ***L. pentosus*** (Lokman *et al.*, 1991), ***S. enterica*** (Clarke *et al.*, 1982), ***S. xylosus*** (Sizemore *et al.*, 1992), ***S. lividens*** (Heo *et al.*, 2008), ***T. halophilus*** (Takeda *et al.*, 1998).

Table 4-9: The sequence identity of CBEI4456 with published characterised arabinose regulation proteins

Organism	Identity (%)
Bacillus subtilis	55
Escherichia coli (AraC)	48
Salmonella enterica (AraC)	48

Identity is the percentage of identical shared AA between CBEI3047 and the XynB sequences in the table

Table 4-10: The sequence identity of CBEI2385 with published characterised xylose regulation proteins

Organism	Identity (%)
Bacillus megaterium	59
Tetragenococcus halophilus	59
Staphylococcus xylosus	55
Bacillus licheniformis	52
Lactobacillus pentosus	52
Escherichia coli (XylS)	46
Bacillus subtilis	26

Identity is the percentage of identical shared AA between CBEI3047 and the XynB sequences in the table

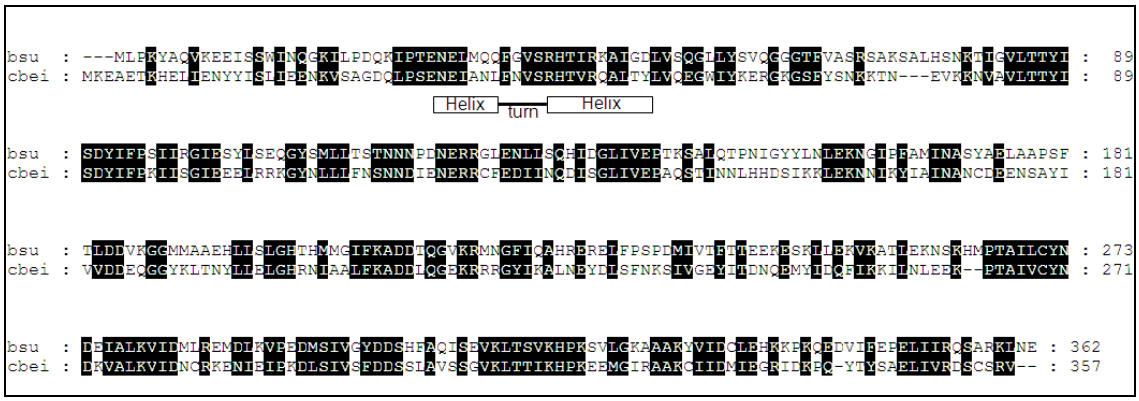


Figure 4-26: An alignment of characterised AraR of *B. subtilis* and CBEI4456

A multiple alignment was constructed using the AA sequences of a confirmed AraR protein of *B. subtilis* and the putative AraR of *C. beijerinckii*, see text for details. Sequence similarity between the two is shaded black. The helix-turn-helix motif associated with DNA binding in *B. subtilis* (Sa-Nogueira and Ramos, 1997) is indicated in the alignment. A BLAST analysis revealed a putative DNA binding site in CBEI4456. **bsu**- *B. subtilis* (Sa-Nogueira and de Lencastre, 1989) **cbei** – *C. beijerinckii* (CBEI4456)

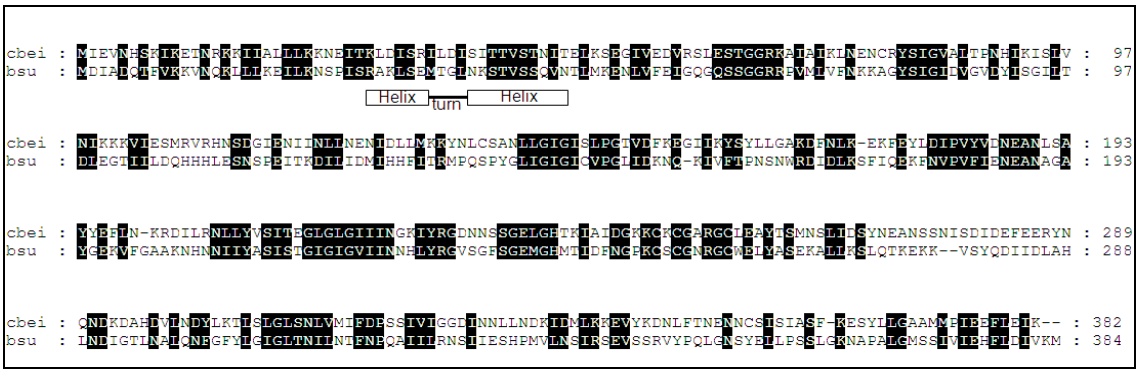


Figure 4-27: An alignment of XylR of *B. subtilis* and CBEI2385

A multiple alignment was constructed using the AA sequences of a confirmed XylR protein of *B. subtilis* and the putative XylR of *C. beijerinckii*, see text for details. Homologous sequences between the two are shaded in black. The helix-turn-helix motif associated with DNA binding in *B. subtilis* (Titgemeyer *et al.*, 1994) is indicated in the alignment. A BLAST analysis revealed a putative DNA binding site in CBEI2385. **bsu**- *B. subtilis* (Kraus *et al.*, 1994), **cbei**- *C. beijerinckii* (CBEI2385)

4.8 Pentose Phosphate Pathway

After arabinose and xylose are catabolised to form xylulose 5-phosphate, transketolase and transaldolase act together to produce glyceraldehyde 3-phosphate, fructose 6-phosphate and NADPH in a series of reactions that form part of the pentose phosphate pathway (Figure 4-28).

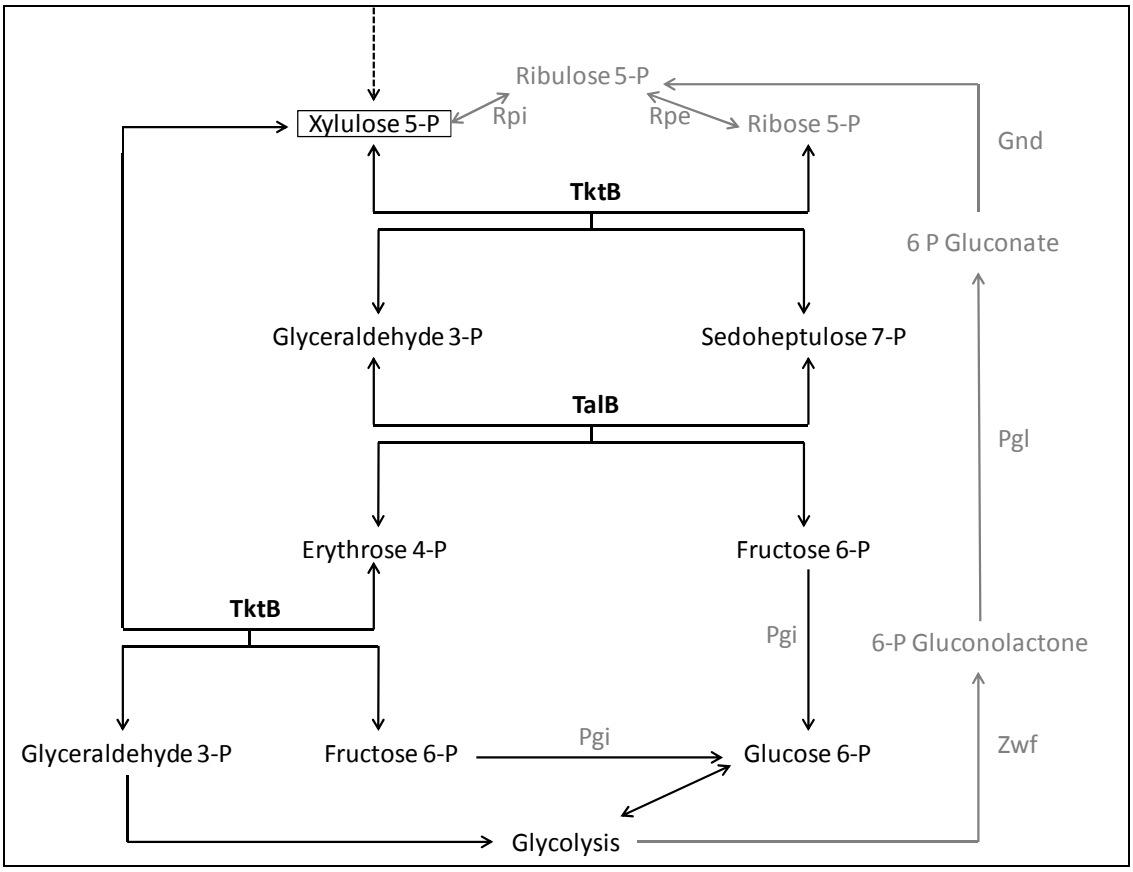


Figure 4-28: Pentose phosphate pathway

After arabinose and xylose are metabolised to xylulose 5-phosphate (boxed text) - the entry point is shown by the dashed arrow; transketolase and transaldolase metabolise these intermediates via the pentose phosphate pathway. Note 'P' is shorthand for phosphate. **Gnd** – Gluconate-6-phosphate dehydrogenase, **Pgi** – Glucose 6-phosphate isomerase, **Pgl**- Phosphogluconolactonase, **Rpe** – Ribulose 5-phosphate 3-epimerase, **Rpi** – Ribulose 5-phosphate isomerase, **TalB** – Transaldolase, **TktB** – Transketolase, **Zwf** – Glucose 6-phosphate dehydrogenase

Several candidates for transaldolases and transketolases were identified by a basic genome search of *C. beijerinckii*. A candidate for each gene was observed within the putative arabinose and xylose utilisation gene systems, these were then compared to

characterised transketolases and transaldolases of *E. coli* and a transaldolase of *C. acetobutylicum*.

4.8.1 Transketolase

An extensive sequence similarity between TktB AA sequence of *E. coli* (Iida *et al.*, 1993) and the two TktB candidates AA sequences of *C. beijerinckii* was evident (Figure 4-29). Both of the TktB candidates had 60% sequence identity with the TktB of *E. coli* and the two *C. beijerinckii* protein sequences had 99% sequence similarity with each other.

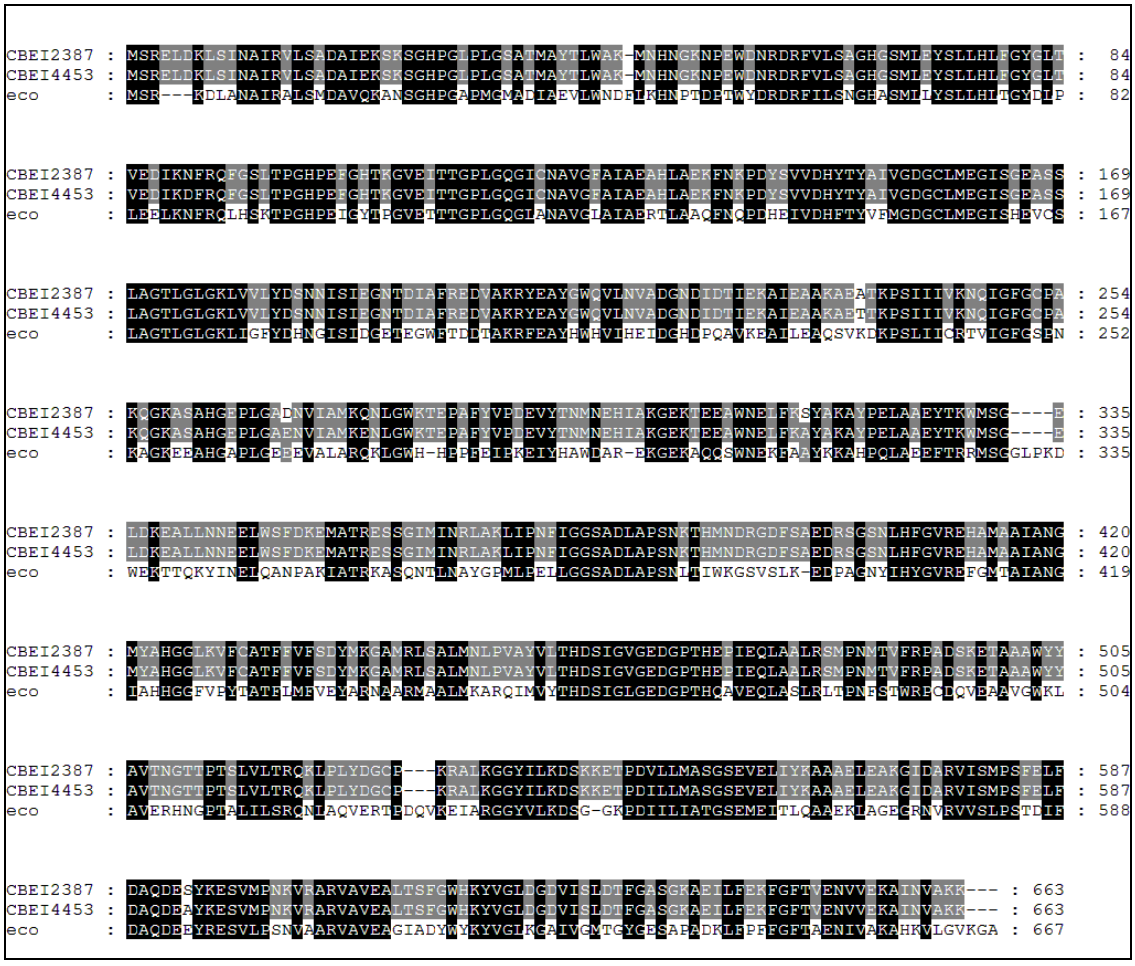


Figure 4-29: A multiple alignment of characterised TktB of *E. coli* and two *C. beijerinckii* candidates

A multiple alignment was constructed using the AA sequences of a confirmed TktB protein of *E. coli* and the putative TktB sequences of *C. beijerinckii* (CBEI2387 & CBEI4453). 100% sequence similarity is shaded black, sequence similarity between two of the sequences is shaded grey. **cbei**- *C. beijerinckii*, **eco**- *E. coli* (Iida *et al.*, 1993)

4.8.2 Transaldolase

Upon the comparison of TalB AA sequence of *E. coli* and *C. acetobutylicum* (Sprenger *et al.*, 1995) and the TalB candidates of *C. beijerinckii* (Figure 4-30), an approximate 100 AA difference in length between TalB of *E. coli* and the other sequences was evident and as a result there were a lot of gap insertions applied to the alignment. Despite this there were sizeable regions of similarity shared by all four sequences (shaded in black). However, the extent of gap insertions in the alignment makes it difficult to assess if there is any meaningful similarity between the *E. coli* sequence and the other three. Indeed both *C. beijerinckii* TalB candidate sequences had 22% sequence identity and *C. acetobutylicum* 20%. Conversely there was 73% sequence identity between the TalB sequence of *C. acetobutylicum* and the two candidates of *C. beijerinckii.* The extent of similarity is evident between these three sequences, highly suggestive of a common function and common ancestry. The two *C. beijerinckii* sequences shared 99% sequence identity and thus are highly likely to have the same function.

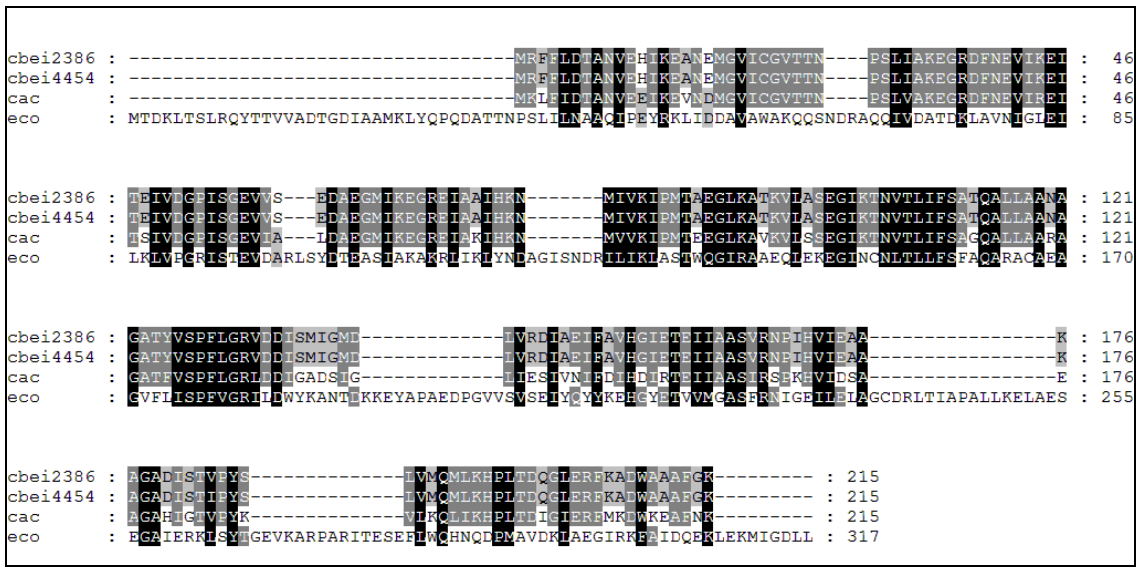


Figure 4-30: A multiple alignment of the TalB *E. coli* and *C. acetobutylicum* and two candidates of *C. beijerinckii*

A multiple alignment was constructed using the AA sequences of a confirmed TalB protein of *E. coli* and *C. acetobutylicum* the putative TalB sequences of *C. beijerinckii* (CBEI2386 & CBEI4454) Black shading denotes 100%, dark grey (white text) 75% and light grey (black text) 50% sequence similarity. **cac** – *C. acetobutylicum* (Gu *et al.*, 2009), **cbei**- *C. beijerinckii* (CBEI2386/4454) **eco**- *E. coli* (Iida *et al.*, 1993).

As mentioned there were several candidates for transaldolase and transketolases discovered from a genome search. CBEI2386 and CBEI4454 shared 99% identity with one and another and they shared 100% and 99% respectively with CBEI4645, the other candidates had 27-34% identity with the other candidates (Table 4-11A). For the putative transketolases, CBEI2387 and CBEI4453, had 98% identity with each other and 23-35% identity with other candidates (Table 4-11B).

Table 4-11: The identity shared between all putative TalB and TktB in *C. beijerinckii*

(A)

	CBEI2386	**CBEI4454**
CBEI0317	34%	34%
CBEI0338	35%	36%
CBEI2742	27%	27%
CBEI4454	99%	-
CBEI4645	100%	99%

The percentage of identical shared AA between all putative TalB of *C. beijerinckii*

(B)

	CBEI2387	**CBEI4453**
CBEI0224	31%	31%
CBEI0225	23%	23%
CBEI0545	33%	33%
CBEI0546	23%	23%
CBEI4453	98%	-
CBEI4870	26%	26%
CBEI4871	35%	35%

The percentage of identical shared AA between all putative TktB of *C. beijerinckii*

4.9 Discussion

Based on the findings of the previous chapter which revealed that *C. beijerinckii* could utilise arabinose and xylose as carbon sources and produce solvents when grown on them, this indicated the presence of genes involved in the uptake and metabolism of pentose sugars. Therefore the next step was to search for candidate gene systems within the genome responsible for this. This revealed two potential separate gene systems for the use of the pentose sugars. Each candidate AA sequence was assessed further using various modes of bioinformatics.

In this study, AA sequences of proteins putatively involved in xylose, arabinose and xylan utilisation were subjected to analysis. ClustalW2 and BLAST programs were used to perform pairwise alignments comparing ORF sequence with characterised sequences to produce a percentage identity. The percentage similarity between the unknown and characterised sequences gives a level of reliability in inferring supposed relationships. Indeed, with a sequence identity score of <50%, the inference of a relationship becomes increasingly less reliable (Attwood and Parry-Smith, 1999). Pairwise alignments can be deceptive, in that it is easy to find weak similarity, where no true relationships lie (Fasman and Salzberg, 1998). It is therefore important to use other means of analysis, as well as assigning a threshold indicating the reliability of the relationship.

The pairwise and multiple alignments were based on the AA sequences rather than nucleotide sequences. This is because the comparison of the AA sequences is much more sensitive, due to the redundancy of the genetic code, whereby often the same AA will have a number of differing codons and therefore a base substitution between two sequences may not affect the amino acid coded for (Attwood and Parry-Smith, 1999, Fasman and Salzberg, 1998). Additionally, a change in the AA sequence at a particular position does not always impact on the structure and function of the coded protein. Alignments vertically align the enquiries in such a way as to preserve an individual sequence's residues, but also take into account differences accumulated through evolution, by the use of gap insertions. This can reveal possible functional and structural motifs indicative of a particular protein (Attwood and Parry-Smith, 1999). However the use of gap insertions must be taken as a penalty as too many could be inserted to deceptively infer a relationship where none exist.

Bioinformatic analyses conducted here have implicated a potential mechanism of pentose sugar transport used by *C. beijerinckii*. There are several modes of transport of solutes observed in bacteria and these may be passive or active. Passive transport equates to facilitated diffusion, whereby the substrate moves from a high concentration to a low one, but this method does not allow the accumulation of a substrate in the cell. Active transport on the other hand allows this through the expense of energy. Such mechanisms include proton or ion symporters, whereby a substrate is taken up the same time as a hydrogen, potassium or sodium ion (Dills *et al.*, 1980). *E. coli, L. brevis* and *B. subtilis* use such a mechanism to transport pentose sugars (Henderson, 1990, Chaillou *et al.*, 1998, Sa-Nogueira and Mota, 1997); also group translocation for example the PEP dependent phosphotransferase system (PTS), whereby the transport and accumulation of a substrate is coupled to a chain of reactions which phosphorylates the substrate, upon entry into the cell. Such a mechanism has been firmly established in clostridia for a number of substrates including glucose, maltose, sucrose and glucitol, but never with pentose sugars (Tangney and Mitchell, 2000, 2007, Tangney *et al.*, 2001, Yu *et al.*, 2007, Mitchell *et al.*, 1991, Reid *et al.*, 1999, Tangney *et al.*, 1998). No PTS has ever been discovered for pentose sugar transport in bacteria studied thus far.

Based on the analyses done here pentose sugar transport in *C. beijerinckii* is predicted to be via ATP binding cassette (ABC) transporters. ABC transporters are made up of three constituents. First, two hydrophobic units, permeases, found within the cell membrane forming a channel through which the substrate can pass to the inside of the cell. Second there are two further units, ATP-binding proteins, which function to hydrolyse ATP providing energy for facilitating the transport of the substrate across the membrane. Such a mechanism confers substrate accumulation (Matuschek *et al.*, 1997). These units are found in the cytoplasm but are associated with the permease units. The final unit is a periplasmic binding protein. The substrate binds to this and is transported to the permeases and interaction between the substrate binding protein complex and the permeases occurs allows the substrate to be released and transported across the cell membrane (Fath and Kolter, 1993, Greller *et al.*, 1999). In prokaryotes including *Sulfolobus solfatorious*, *Thermoccous litoralis*, *E. coli*, *G. stearothermophilus* and *T. ethanolicus* (Elferink *et al.*, 2001, Greller *et al.*, 2001, Horazdovsky and Hogg, 1989, Sumiya and Henderson, 1989, Groarke *et al.*, 1983, Erbeznik *et al.*, 2004) a variety of carbohydrates have been found to be associated with transportation in this manner. Carbohydrates accumulated by this mechanism include monosaccharides, such as

arabinose, ribose and xylose and disaccharides such as lactose, maltose and cellobiose ((Saier, 2000).

Though one putative group of genes encoding each of the ABC transporter components was identified for xylose transport, there appeared to be two candidate groups of ABC transporter components for arabinose transport. These were positioned adjacent either up- or downstream to other genes putatively involved in the metabolism of arabinose (CBEI4448-4450 and CBEI4459-4461, respectively). An initial analysis in the form of a radial tree revealed that only one of the candidate groups possessed a putative arabinose binding protein, AraF. Upon further analysis each of the groups was shown possess a putative AraG, while the upstream group possessed a putative AraH and the downstream group appeared to possess two AraH candidates. Such a finding is not unusual, in *T. ethanolicus* there is a duplication of both *xylF* and *xylH* genes (Erbeznik *et al.*, 2004). It might be that the downstream group is a result of an event such as a horizontal gene transfer.

As previously described, bacteria use isomerases to metabolise pentose sugars, distinct from fungi which use redox reactions. In *S. typhimurium*, *E. coli*, *B. subtilis* and *C. glutamicum* AraABD, are key proteins involved sequentially in arabinose utilisation, in *E. coli*, *B. subtilis*, *L. pentosus*, *B. licheniformis*, *B. megaterium*, *T. halophila*, and *L. lactis* XylAB, are key proteins involved sequentially in xylose utilisation, prior to being fed into the PPP. Candidate *araABD* and *xylAB* were identified within the genome of *C. beijerinckii*. A high level of evidence was accumulated for the function of AraA and XylA, due to the presence of motifs associated with the proteins' functioning, as well as a high level of sequence identity shared with characterised proteins. What was also evident about these two proteins is that there appeared to be two classes or groups. Such a finding has been highlighted previously for XylA sequences (Meaden *et al.*, 1994). This appears to be dictated by the G/C content of the genome the species the XylA is from. XylA sequences of high (≥65%) G/C content species, were found in one group and XylA sequences of lower G/C content (≤52%) species, were found in the other group. With regards AraA, differences between the groups did not correlate with the G/C content of the species, the length of the sequence, or the Gram reaction of the organism.

Evidence for the next proteins involved in pentose sugar metabolism, AraB and XylB was less persuasive. The percentage identities with characterised proteins were lower than the recommended threshold for relationship reliability. The putative XylB shared a higher level of similarity with Gram negative species XylB sequences rather than Gram positive sequences while the opposite was true for AraB, hinting at distinct evolutionary paths. Evidence for the presence of the next protein involved in arabinose metabolism, AraD, CBEI4455, was more convincing, due to a high level of sequence similarity between the candidate and characterised AraD proteins. In both putative pentose sugar utilisation gene systems there were genes encoding a transaldolase and transketolase enzymes involved in the PPP, a known pathway involved in pentose sugar use in both prokaryotes and eukaryotes. Comparison with a Gram positive TalB revealed a more convincing line of evidence of the similarity than with a Gram negative. Conversely, putative transketolase sequences showed a reliable sequence relationship with the TktB sequence from *E. coli*.

Having established that *C. beijerinckii* could utilise xylan, candidate genes within the genome were sought. Two candidate genes, one a putative xylanase and the other a putative xylosidase, were identified and bioinformatic analyses revealed a notable sequence identity existed between the candidate XynA with two bacilli XynA AA sequences. However, little demonstrable homology could be established for the candidate XynB with characterised XynB sequences and therefore no inference of the identity of this candidate sequence could be made.

In prokaryotes regulation of pentose sugar utilisation occurs via one of two mechanisms, depending on the Gram identity of the organism. Gram positive organisms use negative regulation and Gram negatives use positive and negative regulation. Putative pentose sugar repressors of *C. beijerinckii* appear to follow this trend as highlighted by their relationship with repressors of other Gram positive organisms. In *C. beijerinckii* the induction of gene expression would presumably occur due to the pentose sugar binding to the repressor protein, causing a conformational change that prevents the repressor from binding to the promoter/operator regions and allowing the expression of genes associated with pentose sugar metabolism, a mechanism firmly established in a number of Gram positive organisms (Sizemore *et al.*, 1992, Scheler and Hillen, 1994, Schmiedel *et al.*, 1997, Takeda *et al.*, 1998, Mota *et al.*, 1999). In the absence of the pentose sugar the regulator would bind to the operator/promoter region

preventing transcription (Mota *et al.*, 1999, Scheler and Hillen, 1994). This mechanism is very distinct from that of Gram negative organisms. In *E. coli* and *S. enterica* the presence of the pentose sugar the repressor, to the operator site and in the absence, is thought to bind to the initiator site and operator sites, forming a DNA loop. Therefore activating or blocking expression of genes associated with pentose sugar use, respectively (Gallegos *et al.*, 1997, Song and Park, 1997, Kawaguchi *et al.*, 2009).

Based on the bioinformatic analyses performed distinct gene systems have been identified as candidates for xylose and arabinose utilisation and therefore set a base for further studies of *in vivo* gene function characterisation, to ascertain the true function of the genes. A proposed pathway of arabinose and xylose utilisation in *C. beijerinckii* based on the results accumulated here is shown in Figure 4-31.

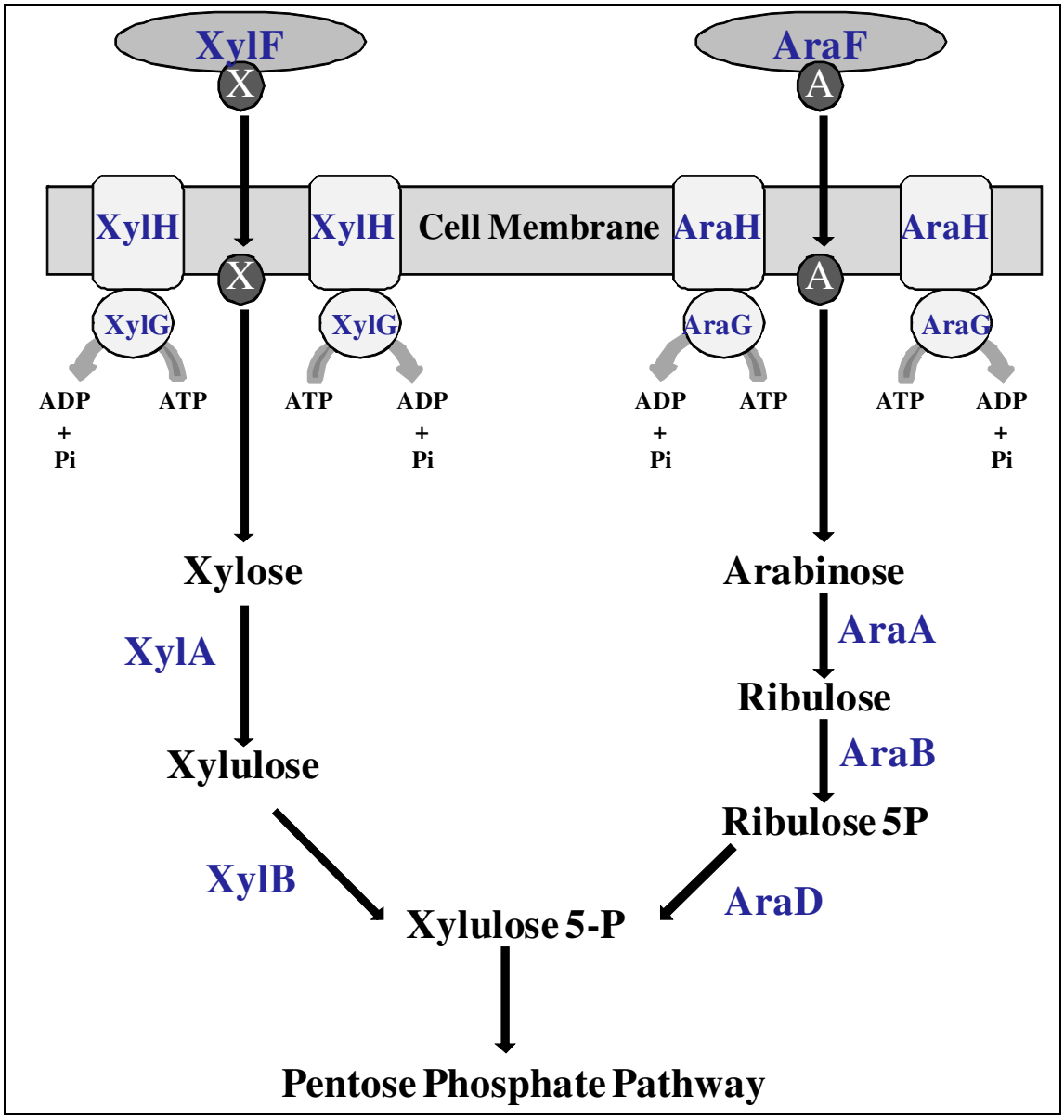


Figure 4-31: The proposed pentose sugar utilisation pathway in *C. beijerinckii*

Based on bioinformatic analyses performed, gene systems and hence a pathway of pentose sugar utilisation in *C. beijerinckii* could be proposed. **AraA**- L-arabinose isomerase, **AraB**- L-ribulokinase, **AraD**-L-Ribulose 5-phosphate 4-epimerase, **AraF**-arabinose binding protein, **AraG**-ATP-binding protein, **AraH**-permease, **XylA**- D-xylose isomerase, **XylB**- xylulokinase, **XylF**- xylose binding protein, **XylG**- ATP-binding protein, **XylH**- permease.

Chapter 5

Characterisation and Expression of Genes Involved in Pentose Sugar Utilisation in *C. beijerinckii*

5 Characterisation and Expression of Pentose Utilisation Genes

5.1 Introduction

It has been established that *C. beijerinckii* is capable of both arabinose and xylose utilisation (Chapter 3). This indicates that *C. beijerinckii* must possess genes for the utilisation of both pentose sugars and potential candidates within the genome were sought and identified using bioinformatics (Chapter 4). The next step was to ascertain if these candidate genes were indeed involved in the use of pentose sugars. To do this whole cell protein extracts from cultures of *C. beijerinckii* grown on minimal media supplemented with either arabinose or xylose as the sole carbon sources were analysed using LC-ESI-MS/MS, to provide a qualitative picture of the proteins necessary for arabinose or xylose utilisation and compare them to the in order to ascertain if the genes bioinformatic analyses (Chapter 4) to provide targets for gene cloning and characterisation.

Proteomics provides whole cell analysis of gene expression under a particular condition, such as a stimulus or stress; at a particular time, for example during a cellular process; and in a particular location for example a specific organelle. As such it is therefore described as an alternative to cDNA microarrays, the current preferred method of gene expression analysis. Proteomics is more expensive, time consuming and technically demanding. However, these two methods are on a par in terms of analysis capability (Cutillas and Timms, 2010) and proteomic analysis is becoming more and more popular. Proteomic analysis can be quantitative but the analysis conducted here was qualitative to provide a snapshot of the genes expressed and hence the proteins required for pentose sugar utilisation in *C. beijerinckii*.

Proteomic analysis was carried out as shown in the schematic diagram (Figure 5-1). As already described whole-cell protein extracts of *C. beijerinckii* grown on pentose sugar were prepared and were fractionated using SDS-PAGE. The lane was then excised and sliced horizontally into 28 slices of ~2.5mm size. Each gel slice was then subject to in-gel trypsinolysis, which included the gel being washed, the proteins reduced and alkylated and then trypsin digestion was conducted. The peptides produced were then subjected to LC-ESI MS/MS analysis.

The first stage of LC-ESI MS/MS analysis is reverse phase HPLC, the most commonly used LC method for this purpose. This separates sample peptides based on hydrophobicity. Each peptide will have a different hydrophobicity which dictates how strongly the peptide adsorbs to the hydrophobic solid phase of the column. A gradient of organic solvent, ACN, the mobile phase, causes the desorption and therefore elution of peptides from the solid phase. Peptides with lower hydrophobicity elute first and peptides with higher hydrophobicity are eluted last, as a higher organic solvent concentration is required to provide decreasing polarity to allow these peptides to desorb. Chromatography is advantageous step because it allows the detection of more ions, because the peptides elute separately on a gradient based on hydrophobicity rather than all at once, therefore providing a more complete ionisation of a protein and therefore a better sample coverage.

In the next stage of LC-ESI MS/MS the peptides elute from the HPLC column into a nebuliser into which N_2 gas is pumped. The purpose is to volatilise or create aerosol droplets containing the peptides, the so called electro-spray ionisation (ESI). Once droplets are produced the next step is to decrease the droplet size and increase the charge prior to MS analysis. This was done when the eluted peptides pass through an electrospray needle with a high voltage applied to it under atmospheric pressure. Droplets with the same charge as the needle are repelled and therefore sprayed from the needle. The droplets then come into contact with N_2 gas, allowing evaporation of solvent from the droplets and size reduction occurs, until the surface tension reaches the Rayleigh limit causing a Coulombic explosion where the droplet is fragmented into smaller ones and this process continues until single analyte ions are produced.

The final stage is MS analysis. Ions are separated by MS via the presence of a magnetic or electric field affecting the movement depending on the mass-to-charge ratio. In tandem MS (MS/MS) the first MS separates according to a single mass of the precursor ion then further fragmentation via collision with an inert gas produces fragment ions termed as daughter ions. These daughter ions are then separated by a second MS according to the mass. This is ideal for sequencing peptides because it provides higher sensitivity, by reducing interference from the large volume of peptides.

At the end of LC-ESI MS/MS a peptide mass fingerprint is detected and this was compared to sequences within the *C. beijerinckii* proteome database, using MASCOT.

MASCOT produced a score, reflecting a statistical significance. In line with published guidelines, protein identification was subject to the presence of two or more peptides, complete with four consecutive b or y ions (peptide fragments have a charge on the N-terminal or the C terminal, respectively), with a Mowse score of >25 and therefore significant to $p<0.05$.

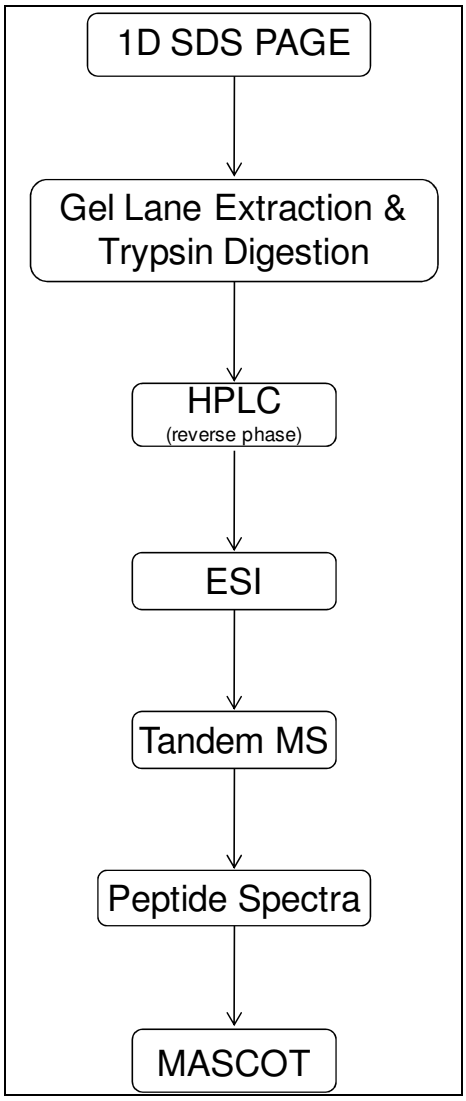


Figure 5-1: Proteomic analysis via LC-ESI MS/MS of whole cell protein extracts of *C. beijerinckii* grown on pentose sugars

C. beijerinckii was grown overnight into log phase and proteins were extracted, separated by SDS PAGE. The whole-cell protein sample was then cut from gel lane, washed, digested with trypsin, and then LC-ESI-MS/MS was conducted on the digested sample.

The next step after establishing what genes were expressed by *C. beijerinckii* cells grown on xylose and arabinose, candidates of , key genes associated with pentose sugar utilisation, *xylAB* and *araABD*, were amplified by PCR and cloned separately on pCR2.1-TOPO cloning vector which was then used to transform *E. coli* TOP10 cells (Invitrogen). Aliquots of the *E. coli* TOP10 used in the transformation reaction were spread on screening plates. The screening involved media containing ampicillin, because ampicillin resistance is carried on the plasmid, so only successful transformants can grow. Secondly, X-GAL was added to the plates for blue-white colony screening. pCR2.1-TOPO vector carries the *lacZα* gene, a truncated form of the *lacZ* gene encoding β-galactosidase (this gene is made fully functional by the host cell providing the other part of the gene lacZΩ). β-galactosidase catabolises lactose into its constituents, glucose and galactose. X-GAL is a chromogenic analogue of lactose which can be catabolised by β-galactosidase to produce galactose and a blue precipitate, 4-chloro-3-brom-indigo and hence colonies containing this gene will appear blue (Fermentas, 2010). However if the *lacZ* gene sequence is interrupted by a cloned gene (by an inserted sequence within the *lacZα* gene of the cloning vector) then no fully functional β-galactosidase can be produced and therefore the colonies will appear a white colour. Thus recombinants and non-recombinants can be distinguished colormetrically.

Recombinants were then harvested and the pCR2.1-TOPO vector carrying the cloned gene was then extracted and inserted into *E. coli* mutants that lacked the functional gene being investigated (they are unable to metabolise either xylose or arabinose depending on the deficiency). Recombinants were then screened *in vivo* for functional complementation, using MacConkey medium. MacConkey medium has three main components that made it a suitable screening method for this purpose; pentose sugar as a carbon source, peptone as an alternative carbon source and neutral red pH indicator. Neutral red changes colour dependent upon the pH of the medium, at pH<6.8 it is a red colour and at pH>6.8 a yellow colour. So in the natural state a mutant unable to utilise the pentose sugar has to utilise the alternative carbon source, peptone, in order to grow. The metabolism of peptone produces ammonia and thus the pH of the medium rises and the pH indicator and the medium appears a yellow colour. If the mutant acquires the lacked gene for pentose sugar utilisation, then the pentose sugar is utilised, producing acids therefore the pH of the medium falls below pH6.8 and neutral red turns the

medium red. If the gene acquired by the mutant does not encode the missing function, then the appearance of the medium will be of that of the mutant in the natural state, still unable to metabolise the pentose sugar. Furthermore, as the plasmid vector carried genes for antibiotic resistance, the addition of ampicillin to the MacConkey medium assured that only recombinants could grow on the medium, as each of the mutants used were sensitive and hence unable to grow on media containing ampicillin. Such a technique has been used successfully previously (Bird, 2011).

The plasmid vector used in this study, pCR2.1-TOPO, is a bidirectional vector and therefore has two different promoters for the purpose of expressing the cloned gene whatever the orientation the gene is cloned. pCR2.1-TOPO possesses a lactose (*lac)* promoter, subject to being cloned under this promoter the expression of the gene required the addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) to the MacConkey medium. IPTG is an analogue of lactose and induces the expression of the genes downstream of the *lac* promoter (the cloned gene). The other promoter T7 functions in the opposite direction and requires the host *E. coli* mutant to possess T7 RNA polymerase, for the expression of the cloned gene

5.2 Proteomics

C. beijerinckii cultures were grown on minimal medium (CBM) supplemented with 1% (w/v) either arabinose or xylose. Cells were harvested, washed and broken using a French pressure cell, to recover the whole-cell proteins. The concentration of the recovered proteins was determined by a BCA assay standard curve (Appendix section 7.2). Then SDS-PAGE was then carried out using five different amounts of the extract from *C. beijerinckii* (Figure 5-2). The 20µg lane was then excised and used for LC-ESI-MS/MS analysis.

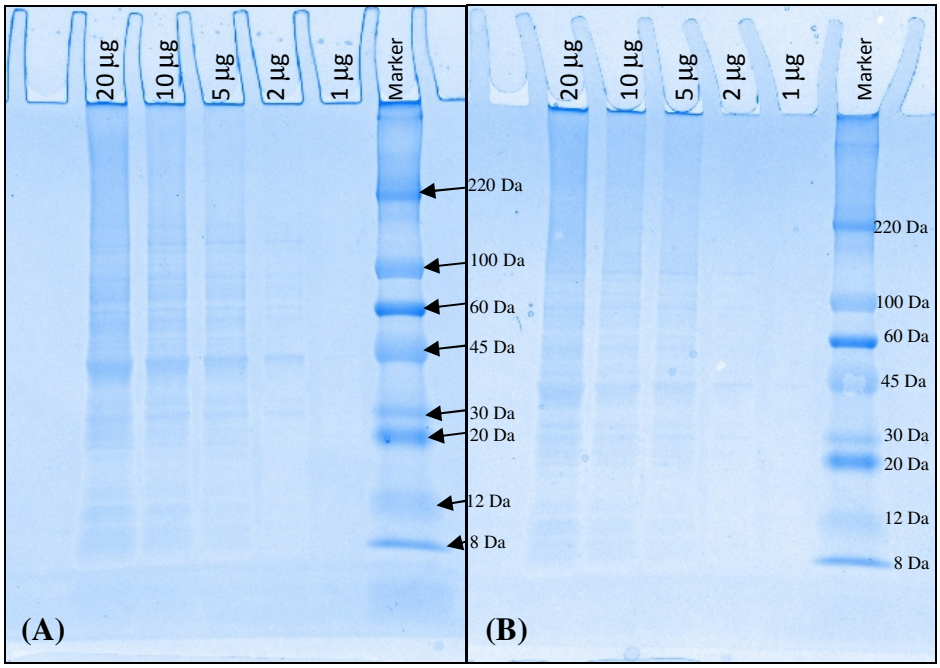


Figure 5-2: SDS-PAGE of *C. beijerinckii* crude extract

C. beijerinckii cultures grown on **(A)** arabinose **(B)** xylose. Cells were harvested, washed and broken with a French pressure cell and whole-cell proteins recovered. Five quantities of the extract were then separated by SDS-PAGE. 20µg lane was then excised and used for LC-ESI-MS/MS analysis.

After LC-EMI-MS/MS analysis was carried out the peptide spectra were analysed using MASCOT. This gives qualitative rather than quantitative results of the whole-cell protein composition of *C. beijerinckii* grown on arabinose or xylose a as single carbon sources. The identification of proteins from peptide spectra presented are significant to $p<0.05$ (with a Mowse score of >25). That is there is a 5% chance that a false-positive match will occur between the MASCOT database with *C. beijerinckii* proteins and the peptides from the LC-ESI-MS/MS results. A full list of the proteins identified and the MOWSE scores can be found in Appendix sections 7.4 and 7.5.

A number putative genes were identified via bioinformatic analyses as having roles in arabinose and xylose utilisation (as presented in Chapter 4) and these were found to be expressed in cultures of *C. beijerinckii* grown on arabinose (

Table 5-1) and xylose (Table 5-2). Gene products of two candidates for arabinose utilisation, the putative AraH (CBEI4448) and AraR (CBEI4456) and three potential proteins involved in xylose utilisation, two putative ABC transporter components

(CBEI2381 and CBEI2382) and the putative XylR (CBEI2385) a identified in Chapter 4 however were not observed. Of the 5020 proteins *C. beijerinckii* possesses, 267 proteins (5.3 %) and 243 proteins (4.8 %) were present in the proteomics analysis of arabinose and xylose grown cultures.

Table 5-1: Expression of the putative arabinose utilisation gene system

ORF	Putative function	Mass (M_r)	Mowse Score	Sequence coverage (%)*
CBEI4449	AraG (ABC transporter)	56239	223	32
CBEI4450	AraF (ABC transporter)	41113	626	61
CBEI4452	L-ribulokinase	58827	684	43
CBEI4453	Transketolase	72572	1365	69
CBEI4454	Transaldolase	23178	1490	74
CBEI4455	L-ribulose-5-phosphate 4-epimerase	25411	371	69
CBEI4457	L-arabinose isomerase	54563	306	49

This shows the presence of protein encoded by the proposed arabinose utilisation gene system in *C. beijerinckii* identified in Chapter 4. The sequence coverage shows the percentage of the amino acid sequence of the protein observed from LC-ESI MS/MS analysis of whole cell proteins of *C. beijerinckii* grown on arabinose. All Mowse scores are the equivalent of a significance of $p<0.05$. * this is the percentage of peptide fragments covering the protein sequence observed in the spectral results.

Table 5-2: Expression of putative xylose utilisation gene system

ORF	Putative function	Mass (M_r)	Mowse Score	Sequence coverage (%)*
CBEI2380	XylF (ABC transporter)	38942	165	19
CBEI2383	D-xylose isomerase	50455	177	47
CBEI2384	D-xylulokinase	55260	889	78
CBEI2386	Transaldolase	23164	1651	68
CBEI2387	Transketolase	72544	1044	29

This shows the presence of proteins encoded by the proposed xylose utilisation gene system in *C. beijerinckii* identified in Chapter 4. The sequence coverage shows the percentage of the amino acid sequence of the protein observed from LC-ESI MS/MS analysis of whole cell proteins of *C. beijerinckii* grown on xylose. All Mowse scores are the equivalent of a significance of $p<0.05$. * this is the percentage of peptide fragments covering the protein sequence observed in the spectral results.

5.3 Genes involved in xylose utilisation

5.3.1 Xylose isomerase

5.3.1.1 PCR amplification and pDNA purification

Genomic DNA (gDNA) was isolated from *C. beijerinckii* as described in Materials and Methods and used as a template for the PCR amplification of the putative *xylA*. Primers were designed up- and downstream of the target ORF CBEI2383 and are listed in Materials and Methods. A ~1500 base pair (bp) product was obtained (Figure 5-3), the expected size for CBEI2383 amplification was 1509bp. The band was then excised from the agarose gel and purified as described in Materials and Methods.

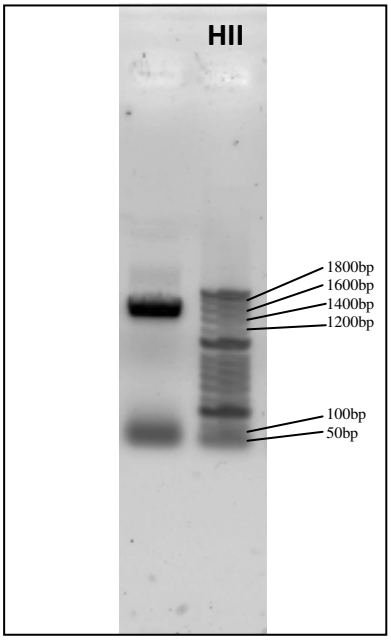


Figure 5-3: PCR amplification of CBEI2383

ORF CBEI2383 was amplified by PCR and ran on a 1% agarose gel alongside Hyperladder II (**HII**) (Bioline) was used as a size standard

5.3.2 Cloning and restriction analysis

The purified CBEI2383 DNA was cloned into the pCR2.1 TOPO cloning vector and the recombinant plasmid was then used to transform *E. coli* TOP10 cells. Colonies of the transformed *E. coli* TOP10 cells were screened using X-GAL to identify colonies carrying CBEI2383. Nine transformants were harvested and grown overnight and two were randomly selected for pDNA extraction. The pDNA (pJW1 and pJW3) was then subjected to digestion by the restriction enzyme *Hin*dIII, to determine the orientation of CBEI2383. The two possible orientations the gene could have been cloned in are shown in Figure 5-4. Cloning under control of the *lac* promoter would result in two fragments of 1338 and 4102bp after digestion with *Hin*dIII and under control of the T7 promoter would result in fragments of 287 and 5153bp.

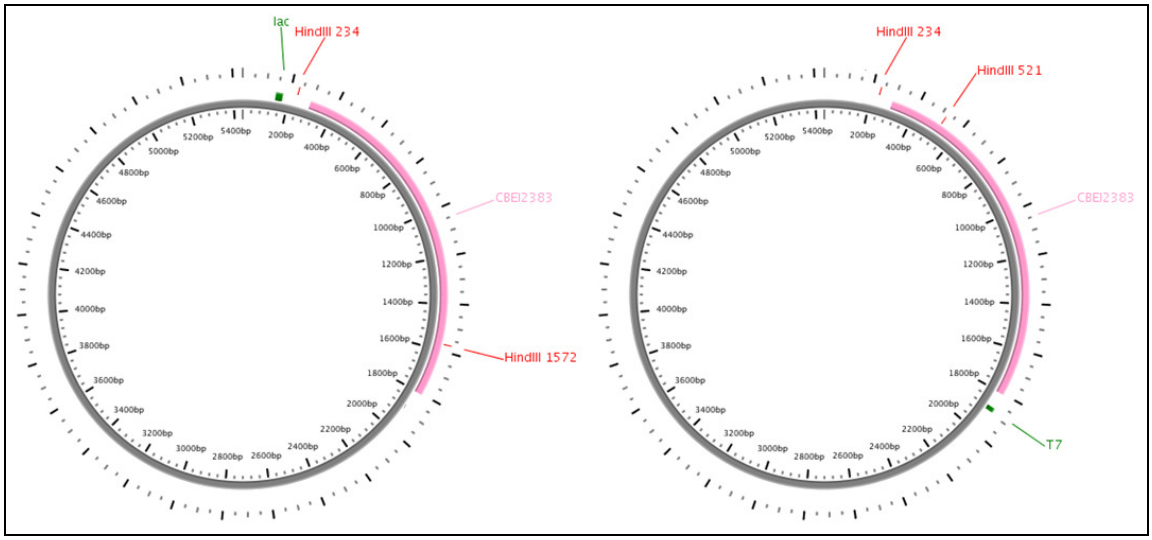


Figure 5-4: Possible orientations of the cloned CBEI2383

This diagram shows the possible orientations of CBEI2383(—) cloned into pCR2.1-TOPO vector (—). On the left is the orientation with expression under the control of the *lac* promoter (—), *Hin*dIII (—) restriction digest would produce 2 fragments of 1338bp and 4102bp. On the right is the orientation with expression under the control of the T7 promoter (—), *Hin*dIII (—) restriction digest would produce 2 fragments of 287bp and 5153bp

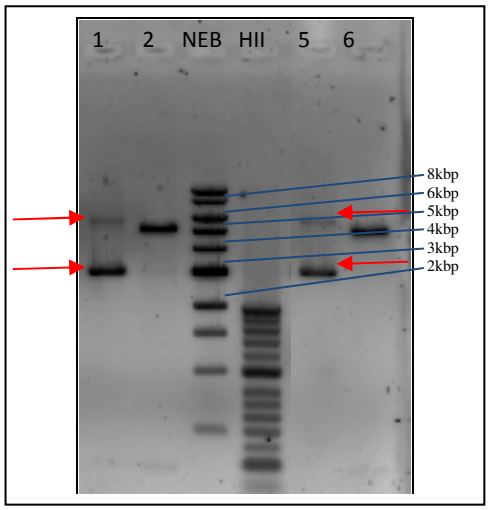


Figure 5-5: *Hin*dIII restriction digest of pJW1 and pJW3
Lane **1**- pJW1 undigested, **2**- pJW1 *Hin*dIII treated, **NEB**- NEB 1kb ladder (NEB), **HII**- Hyperladder II (Bioline), **5**- pJW3 undigested, **6**- pJW3 *Hin*dIII treated. *Hin*dIII digestion has produced two fragments for both pJW1 and 3 of sizes <300bp and ~5000bp as indicated by the red arrows.

Restriction digestion of pJW1 and pJW3 with *Hin*dIII resulted in 2 fragments of <300bp and ~5000bp (Figure 5-5) consistent with CBEI2383 being cloned in the direction of expression under control of the T7 promoter (Figure 5-4).

5.3.3 Complementation screening in vivo

pJW1 and pJW3 were then used to transform a *xylA*$^-$ *E. coli* host (also an *araB*$^-$ mutant), via electroporation. For the expression of CBEI2383 it was required that the host strain produces T7 polymerase for T7 expression. The host strain used, *E. coli* DS941, lacked T7 polymerase. However a sizeable amount of upstream sequence (114bp) was also cloned within the PCR product and therefore a potential native promoter may have also been cloned, allowing the expression of CBEI2383. A putative Pribnow box and a -35 sequence were identified (Figure 5-6). Indeed, there must have been such a sequence present as the insertion of pJW1 and pJW3 into *E. coli* DS941 yielded a positive fermentation phenotype on xylose like the positive control, the cloned *xylA* gene of *T. thermosaccharolyticum* (Meaden *et al.*, 1994). The negative control (pUC19) produced a negative phenotype on xylose (Figure 5-7A). All strains gave a positive fermentation phenotype on glucose as expected (Figure 5-7B).

so in conclusion the *in vivo* activity of the gene encoding ORF CBEI2383 has been demonstrated here as a xylose isomerase.

GGGCTAGTTATTACTAACGTTGGCTTTTATAAATAGATAAATGGCTTAATAAGTTATTACGCTATTTACT
AGAAATAATTTATATCCAATATAGAAAAATGAGGAGGGTATCATATGAGAGAATACTTTGAAAATGTTTC

Figure 5-6: Putative Pribnow box, -35 and Shine-Dalgarno sequences upstream of CBEI2383

CBEI2383 is shown in black text, the intergenic region is shown in blue text, a putative Pribnow box and -35 sequence are highlighted in green, putative Shine-Dalgarno sequence is highlighted in blue.

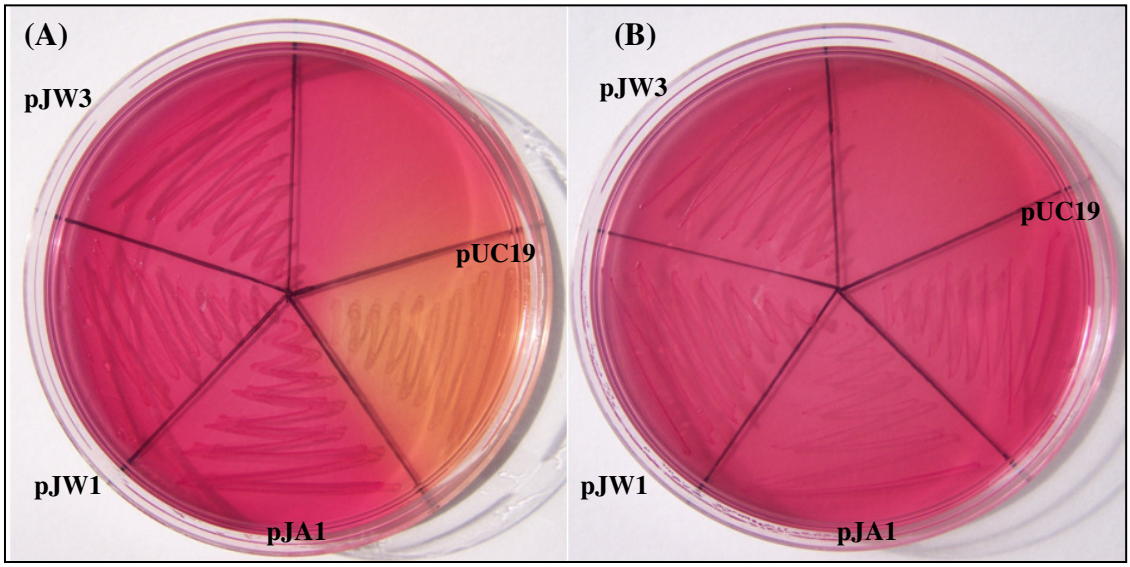


Figure 5-7: Screening of pJW1 and pJW3 on MacConkey medium

(A) MacConkey medium supplemented with xylose and ampicillin was used to screen for any complementation of the *E. coli* DS941 *xylA*$^{-}$ mutant strain, naturally unable to utilise xylose, when transformed with either pJW1, pJW3 (containing CBEI2383), pJA1 (positive control carrying the *xylA* gene of *T. thermosaccharolyticum*) and the negative control pUC19 (carrying no genes associated with xylose use). **(B)** MacConkey medium supplemented with glucose (as a screening control) and ampicillin Transformation of *E. coli* DS941 with pJW1, pJW3 and pJA1 is observed as giving a positive fermentation phenotype on both sugars and with pUC19 only a positive phenotype on glucose but still a negative fermentation phenotype on xylose.

5.3.4 Xylulokinase

5.3.4.1 PCR amplification and pDNA purification

The putative *xylB* gene was amplified and purified in the same manner as the *xylA* gene using the conditions and primers described in Materials and Methods. The PCR produced a ~2000bp product, the expected size was 1998bp (Figure 5-8).

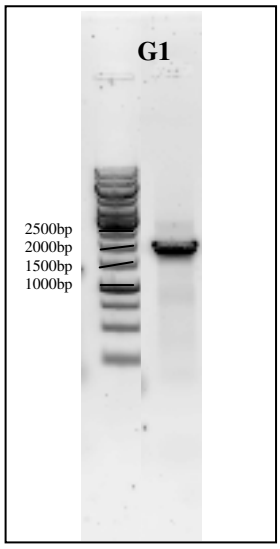


Figure 5-8: PCR amplification of CBEI2384

ORF CBEI2384 was amplified by PCR and ran on a1% agarose gel alongside Generuler 1Kb ladder (**G1**) (Fermentas) was used as a size standard

5.3.4.2 Cloning and restriction analysis

CBEI2384 DNA was purified and cloned into the pCR2.1-TOPO cloning vector and used to transform *E. coli* TOP10 cells Screening and harvesting of the transformed *E. coli* TOP10 cells was performed as previously described. Four transformants were harvested, grown overnight and two transformants were randomly selected for pDNA extraction. The orientation of the cloned gene (CBEI2384) was ascertained using the restriction enzyme *Hin*dIII. The possible orientations that CBEI2384 could have been cloned in are shown in Figure 5-9. Cloning under control of the *lac* promoter would result in two fragments of 680 and 5249bp after digestion with *Hin*dIII, and cloning under control of the T7 promoter would result in fragments of 1434 and 4495bp.

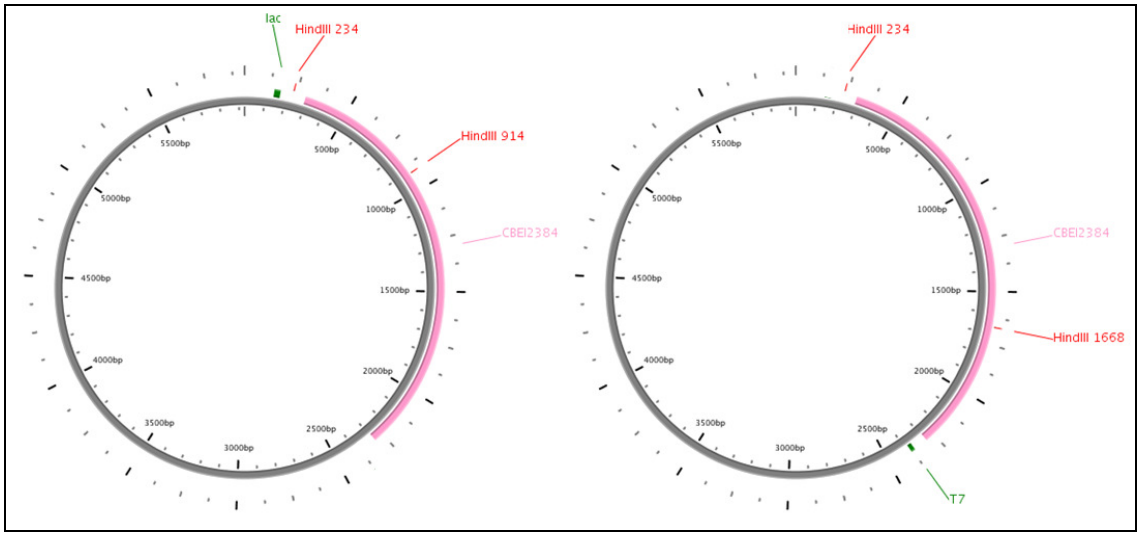


Figure 5-9: Possible orientations of the cloned CBEI2384

This diagram shows the possible orientations of CBEI2384 (—) cloned into pCR2.1-TOPO vector (—). On the left is the orientation with expression under the control of the *lac* promoter (—), *Hin*dIII (—) restriction digest would produce 2 fragments of 680bp and 5249bp. On the right is the orientation with expression under the control of the T7 promoter (—), *Hin*dIII (—) restriction digest would produce 2 fragments of 1434bp and 4495bp

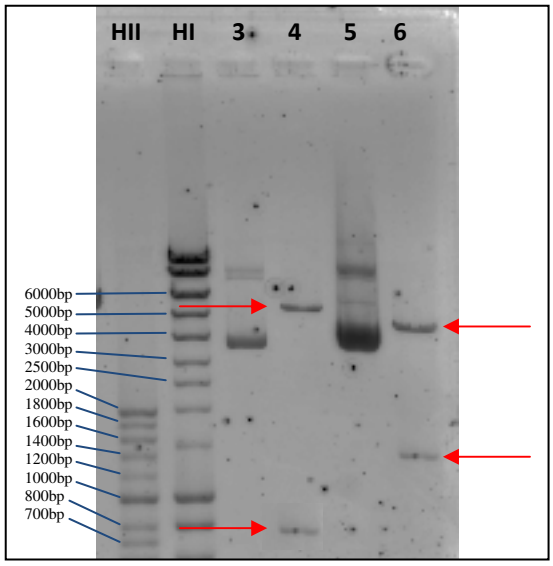


Figure 5-10: *Hin*dIII restriction digest of pJW19 and 20

Lane **HII**- Hyperladder II (Bioline), **HI**- Hyperladder I (Bioline), **3**-pJW19 undigested, **4**- pJW19 *Hin*dIII treated, **5**- pJW20 undigested**, 6**- pJW20 *Hin*dIII treated. *Hin*dIII digestion has produced two fragments for both pJW19 and 20 of sizes ~1400bp, ~4500bp and ~5000bp, ~700bp, respectively. As indicated by the red arrows.

Restriction digestion of pJW19 with *Hin*dIII resulted in 2 fragments of ~1400bp and ~4500bp while pJW20 gave fragments of >5000bp and ~700bp (Figure 5-10), consistent with CBEI2384 being cloned in the direction of expression under control of the T7 promoter and *lac* promoter, respectively (Figure 5-9).

5.3.4.3 Complementation screening in vivo

The *xylB*$^-$ *E. coli* XK100 host was transformed with pJW19 and pJW20 using electroporation. Like the *xylA*$^-$ host, *E. coli* XK100 lacks the T7 polymerase gene required for expression under control of the T7 promoter. However, following previous success with CBEI2383 the host strain was transformed with pJW19 and pJW20. Putative sequences of a Pribnow box and -35 and Shine-Dalgarno sequences were also identified upstream of CBEI2384, like that of CBEI2383 (Figure 5-11). Therefore the possibility of a native promoter being present within the 417bp upstream of CBEI2384 also amplified, was strong. The insertion of pJW19 and pJW20 into *E. coli* XK100 yielded a positive fermentation phenotype on xylose. The negative control transformed with pUC19 gene gave a negative phenotype on xylose (Figure 5-12A). All produced a positive fermentation phenotype on glucose as expected (Figure 5-12B).

So in conclusion the *in vivo* activity of the gene encoding ORF CBEI2384 has been demonstrated here as a xylulose kinase.

CATGCTTTTATCCCGACTTGTACGGGATTTTATGGATGATCCTAAGATTAGAAGTTGTACCCAAAGGCTA
GCCTTAAAGGTAAGATCTAAGTAAATACATTTGAAAAAACTTCTAACTTGTAAATTTATAGTTAAAAGGA
CTTCTTGTAATATTTTTCAGTATTATTATTGCAAGAAGTTCCTTTTTATTTAGGACTCTATTTTGGTAGA
TATTTTATAATAAGTGCAACATCCTAAATACAAAAAATAACTAGTAAAATAAAAAATGAAAATGTATCCA
AAAAACTATTGACCGCAAAGAAAAAGGGAAGTATAATAAACTTATTAAAATACATTTAAAAAGTTATTTA
AAAAATTAAGTAAAAAGTTTGTAATTATGTTAATTTAGAAAAGATATATAACTTAGGAGGTTTATTAATG

Figure 5-11: Putative Pribnow box, -35 and Shine-Dalgarno sequences upstream of CBEI2384

CBEI2384 is shown in black text, the intergenic region is shown in blue text, a putative Pribnow box and -35 sequence are highlighted in green, putative Shine-Dalgarno sequence is highlighted in blue.

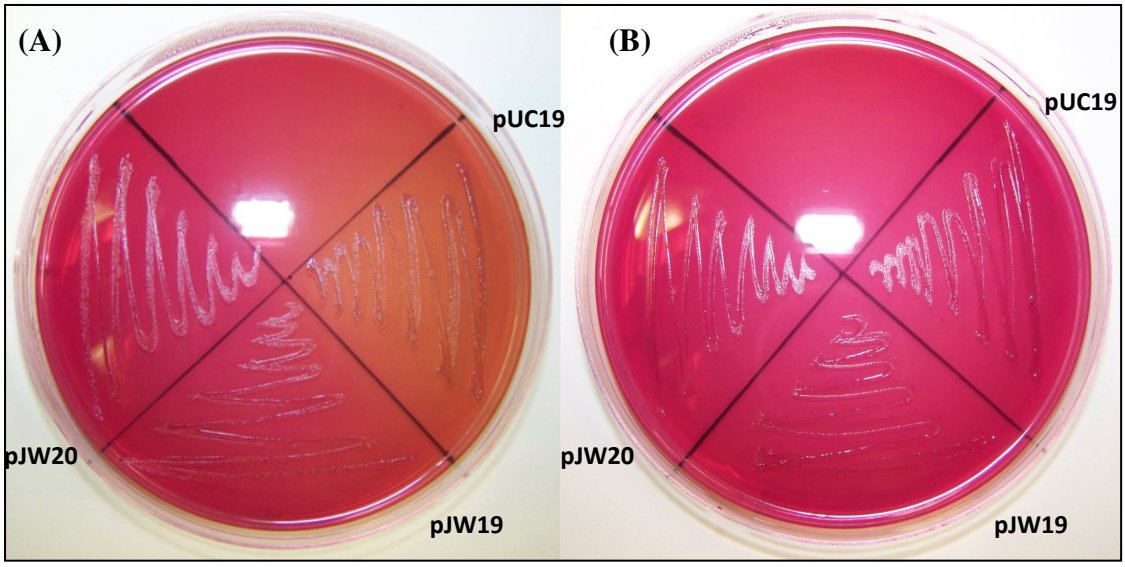


Figure 5-12: Screening of pJW19 and pJW20 on MacConkey medium

(A) MacConkey medium supplemented with xylose, ampicillin and 40mM IPTG was used to screen for any complementation in *E. coli* XK100 *xylB*$^{-}$ mutant strain, naturally unable to utilise xylose, when transformed with either pJW19, pJW20 or the negative control pUC19 (carrying no genes associated with xylose use). **(B)** MacConkey medium supplemented with glucose (as a screening control), ampicillin and 40mM IPTG Transformation of *E. coli* XK100 with pJW19 and pJW20 resulted in a positive fermentation phenotype on both sugars and with pUC19 only a positive phenotype on glucose but still a negative fermentation phenotype on xylose.

5.4 Genes involved in arabinose utilisation

5.4.1 Arabinose isomerase

5.4.1.1 PCR amplification and pDNA purification

gDNA was isolated as previously described and used as a template for the PCR amplification of the putative *araA*, CBEI4457. Primers were designed up- and downstream of the target ORF CBEI4457 and are listed in Materials and Methods. A ~2000bp product was obtained, the expected size was 1937bp (Figure 5-13). The band was excised from the agarose gel and purified as previously described.

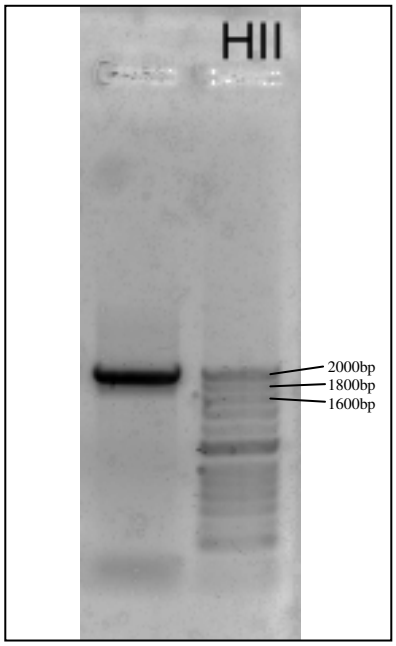


Figure 5-13: PCR amplification of CBEI4457

ORF CBEI4457 was amplified by PCR and ran on a 1% agarose gel alongside Hyperladder II (**HII**) (Bioline) was used as a size standard

5.4.1.2 Cloning and restriction analysis

The purified CBEI4457 DNA was cloned into the pCR2.1-TOPO cloning vector and *E. coli* TOP10 cells were then transformed with the ligated plasmid. As previously carried out two randomly selected colonies from over 300 transformed *E. coli* Top10 cells were screened for the successful cloning of CBEI4457 using X-GAL, then harvested, grown overnight and pDNA extracted. As previously described TOPO-TA cloning is

non-directional and therefore to ascertain the orientation of the cloned gene, digestion with H*ind*III restriction enzyme was carried out. The two possible orientations of the cloned gene are shown in Figure 5-14. Cloning under control of the *lac* promoter would result in three fragments of 4746, 947 or 174bp. Cloning in the opposite orientation under control of the T7 promoter would result in three fragments of 3989, 931 and 947bp, which on a gel would appear as two fragments.

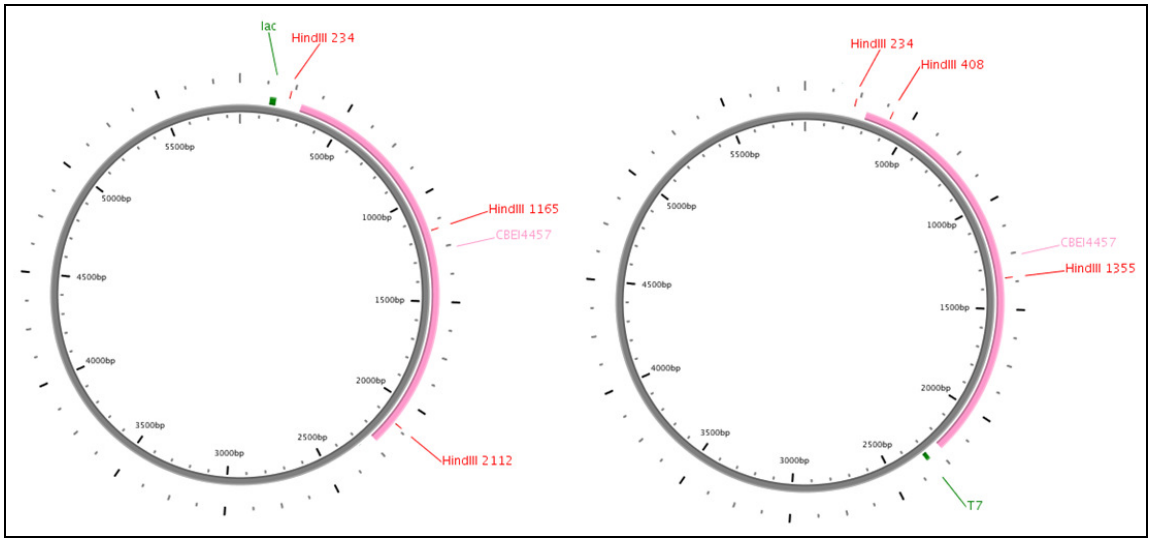


Figure 5-14: Possible orientations of the cloned CBEI4457

This diagram shows the possible orientations of CBEI4457 (—) cloned into pCR2.1-TOPO vector (—). On the left is the orientation with expression under lac promoter (—), *Hin*dIII (—) restriction digest would produce 3 fragments 174bp, 947bp, 4746bp. On the right is the orientation with expression under T7 promoter (—), *Hin*dIII (—) restriction digest would produce 3 fragments 931bp, 947bp, 3989bp, and would probably appear as two fragments of ~900bp and 3989bp.

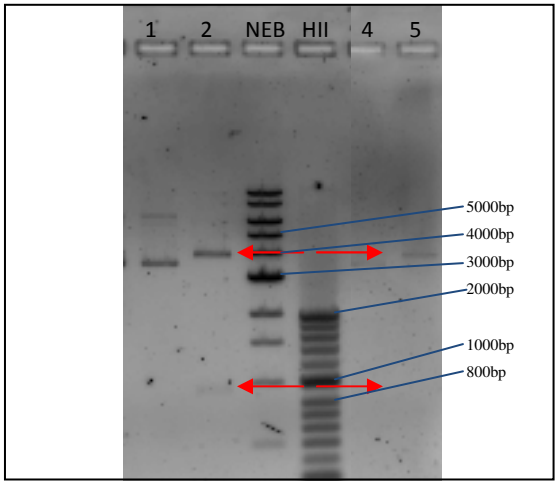


Figure 5-15: *Hind*III restriction digest of pJW6 and pJW7

Lane 1- pJW6 undigested, **2**- pJW6 *Hin*dIII treated, **NEB**-NEB 1Kb ladder (NEB), **HII**-Hyperladder II (Bioline), **4**- pJW7 undigested, **5**- pJW7 *Hin*dIII treated. *Hin*dIII digestion has produced two fragments for both pJW6 and 7 of sizes <1000bp and ~4000bp as indicated by the red arrows.

Restriction digestion using *Hin*dIII produced two fragments of <1000bp and ~4000bp, from both pJW6 and pJW7 (Figure 5-15). This suggested with reference to Figure 5-14, that in both plasmids the CBEI4457 gene was cloned under the control of the T7 promoter.

5.4.2 Complementation screening *in vivo*

pJW6 and pJW7 were then used to transform an *araA*$^-$ *E. coli* host using electroporation. The orientation of the cloned gene was such that it is required the host strain produces T7 polymerase for expression. The host strain used, *E. coli* UP1091, lacks T7 polymerase, however the CBEI4457 gene was amplified together with 340 bp upstream of the open reading frame. This meant that potentially a native promoter may have also been cloned, allowing the expression of CBEI4457. A putative Pribnow box and -35 sequence have been identified in this upstream region (Figure 5-16). Based on the results of the complementation screening (Figure 5-17) it is likely that such a native sequence was present allowing the expression of CBEI4457. The insertion of pJW6 and pJW7 into *E. coli* UP1091 produced a positive phenotype on arabinose-containing medium (Figure 5-17A) whereas the insertion of pUC19, carrying no genes associated with arabinose utilisation had a negative phenotype on arabinose. All pDNA produced a positive phenotype on glucose in the control screen (Figure 5-17B).

Therefore the *in vivo* activity of the gene encoding ORF CBEI4457 has been demonstrated here as an arabinose isomerase.

ACCGCAAGGCTGTCTTTACCTAATTTAAACCATAATAATGTTTATCTGGAAAGTTAATATTTTTTATTT
TTTTGATTTTACATCTTGTATGTACAACATTATTGTATTATAATAAGCTTGTACATACATCAAAAACTT
GTTTGTACTCATGAAAAGGTATACTAAATTTAGATAAGCATTTTTATTTATCCAATAGCTAGCATCCGT
AGTACTCCATCTTCTTAAGTGAGAGATAACGGCCGCAAGCTCCTGGACTAGGTGCCCCTTAAGGGTATA
AATCCAAGGTGGAGTTAAAATCCCACTTAAAAAGTTTCACTTTATATATTAGGAGGAATGTTAAATGTT

Figure 5-16: Putative Pribnow box, -35 and Shine-Dalgarno sequences upstream of CBEI4457

CBEI4457 is shown in black text, the intergenic region is shown in blue text, a putative Pribnow box and -35 sequence are highlighted in green, putative Shine-Dalgarno sequence is highlighted in blue.

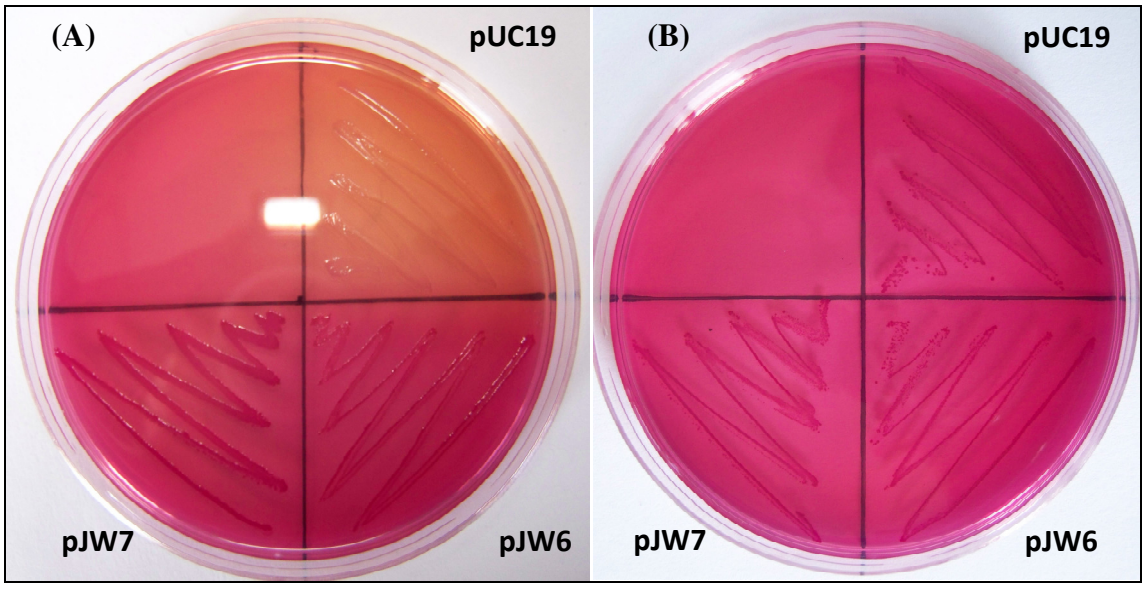


Figure 5-17: screening of pJW6 and pJW7 on MacConkey medium

(A) MacConkey medium supplemented with arabinose and ampicillin was used to screen for any complementation of the *E. coli* UP1091 *araA*$^-$ mutant strain (naturally unable to utilise arabinose), when transformed with either pJW6, pJW7 and the negative control pUC19 (carrying no genes associated with arabinose use) **(B)** MacConkey medium supplemented with glucose (as a screening control) and ampicillin. Transformation of *E. coli* UP1091 with pJW6 and pJW7 was observed as having a positive fermentation phenotype on both sugars. pUC19 had a positive phenotype on glucose and a negative fermentation phenotype on arabinose.

5.4.3 L-ribulokinase (Putative)

5.4.3.1 PCR amplification and pDNA purification

The putative *araB* gene (CBEI4452) was amplified and purified in the same manner as the *araA* gene, producing a ~2000bp product shown in Figure 5-18 (the expected size was 1995bp).

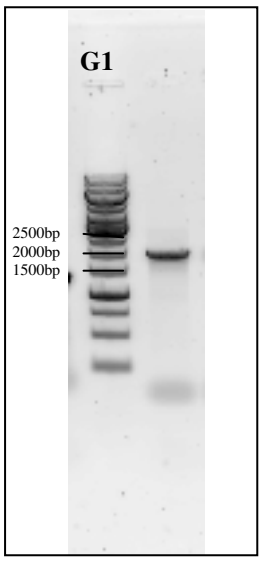


Figure 5-18: PCR amplification of CBEI4452

ORF CBEI4452 was amplified by PCR and ran on a1% agarose gel alongside Generuler 1kb DNA ladder (**G1**) (Fermentas) was used as a size standard. This shows a PCR product of ~2000bp.

5.4.4 Cloning and restriction analysis

E. coli TOP10 cells were transformed with CBEI4452 DNA which was purified and cloned into the pCR2.1-TOPO plasmid vector as previously described. Screening and harvesting of two randomly selected transformed *E. coli* TOP10 cells out of eight was performed as previously described and pDNA extracted. The orientation of the gene was ascertained using the restriction enzyme *Spe*I. The possible orientations are shown in Figure 5-19. Cloning under control of the *lac* promoter would result in two fragments of 1801 and 4122bp following digestion with *Spe*I and under control of the T7 promoter would result in fragments of 262 and 5661bp.

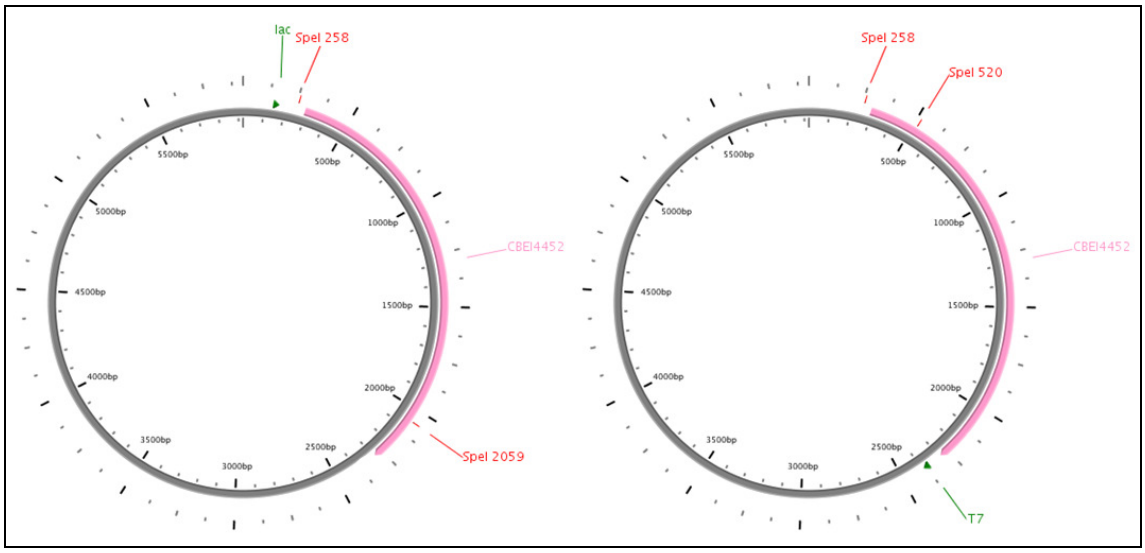


Figure 5-19: Possible orientations of the cloned CBEI4452

This diagram shows the possible orientations of CBEI4452(—) cloned into pCR2.1-TOPO vector (—). On the left is the orientation with expression under control of the *lac* promoter (—), *Spe*I (—) restriction digest would produce 2 fragments of 1801bp and 4122bp. On the right is the orientation with expression under T7 promoter (—), *Spe*I (—) restriction digest would produce 2 fragments of 262bp and 5661bp

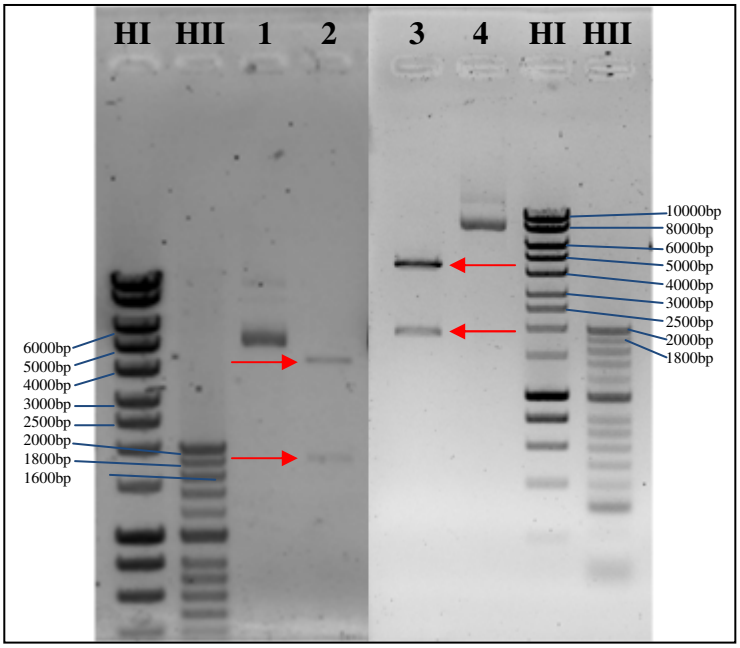


Figure 5-20: *Spe*I restriction digest of pJW24 and pJW25

HII-Hyperladder II (Bioline), **HI-**Hyperladder I (Bioline), **1**-pJW24 undigested, **2**-pJW24 *Spe*I treated, **3**-pJW25 *Spe*I treated, **4**-pJW25 undigested

Restriction digestion of pJW24 and pJW25 with *Spe*I resulted in 2 fragments of ~1800bp and >4000bp (Figure 5-20). With reference to Figure 5-19 is consistent with CBEI4452 being cloned in the orientation with its expression under control of the *lac* promoter.

5.4.5 Complementation screening *in vivo*

The *araB*$^-$ *E. coli* DS941 host was transformed with pJW24 and pJW25 using electroporation. The insertion of pJW24 and pJW25 into *E. coli* DS941 yielded a negative fermentation phenotype on arabinose as did the negative control (Figure 5-21A), despite the addition of IPTG. But all strains produced a positive fermentation phenotype on glucose as expected (Figure 5-21B). The other six *E. coli* TOP10 transformants had their pDNA extracted in order to ascertain their orientation in the hope that the orientation was under the expression of the T7 promoter, as previous success with the expression of the gene under control of the T7 promoter has been observed. Unfortunately these were all cloned in the orientation under control of the *lac* promoter.

Therefore in conclusion, no evidence has been obtained here to show the *in vivo* activity of the gene encoding ORF CBEI4452 demonstrates a ribulokinase function. Reasons were then sought, given the high level of similarity CBEI4452 AA sequence had with AraB of *C. glutamicum* (Chapter 4) and its presence in whole-cell protein extracts from *C. beijerinckii* cells grown on arabinose (proteomic analysis).

E. coli DS941 has an *ara14* genotype, carries a point mutation in *araB* gene and is shown to accumulate large quantities of ribulose sugar and is inhibited by arabinose. It has been observed that *araB* mutations have a dual effect in that different *araB* mutants producing varying levels of AraA activity. The *ara14* mutant used in this study has been found to have lowest level of AraA activity in comparison to other *araB* mutants and the wild type (Englesberg, 1960). This might explain why there was no complementation observed. Had a positive control been available this would have highlighted this. Therefore the experiment would need to be repeated with a more appropriate *araB*- strain which does not affect the AraA activity and a positive control would need to be obtained.

pJW24 and pJW25 was also inserted into *E. coli* XK100 (*xylB*- mutant) to check for *xylB* complementation alongside a positive control pJW19 the *xylB* gene of *C. beijerinckii* (Figure 5-22). Like that of the insertion of pUC19, a negative fermentation phenotype was observed with insertion of pJW6 and pJW7 into *E. coli* XK100 on xylose-containing medium. Whereas the insertion of pJW19 into *E. coli* XK100 yielded a positive phenotype.

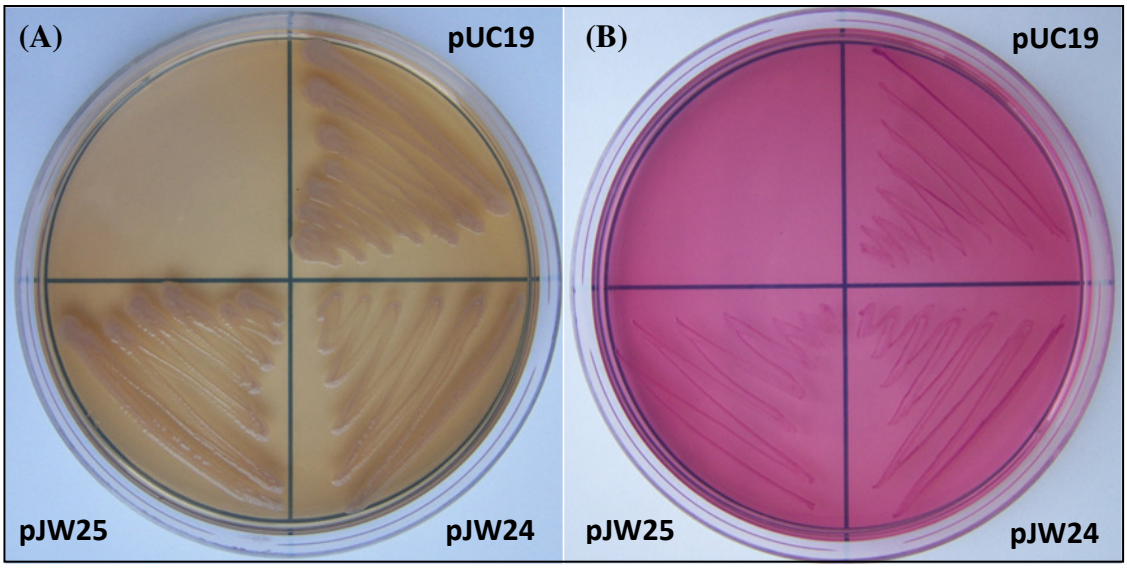


Figure 5-21: Screening of pJW24 and pJW25 MacConkey medium

(A) MacConkey media supplemented with arabinose, ampicillin and 40mM IPTG was used to screen for any complementation in *E. coli* DS941 *araB*$^-$ mutant strain, naturally unable to utilise arabinose, when transformed with either pJW24, pJW25 and the negative control pUC19 (carrying no genes associated with arabinose use). **(B)** MacConkey media supplemented with glucose (as a screening control), ampicillin and 40mM IPTG. Transformation of *E. coli* DS941 with pJW24, pJW25 and pUC19 are all observed as having a negative fermentation phenotype on arabinose and positive on glucose

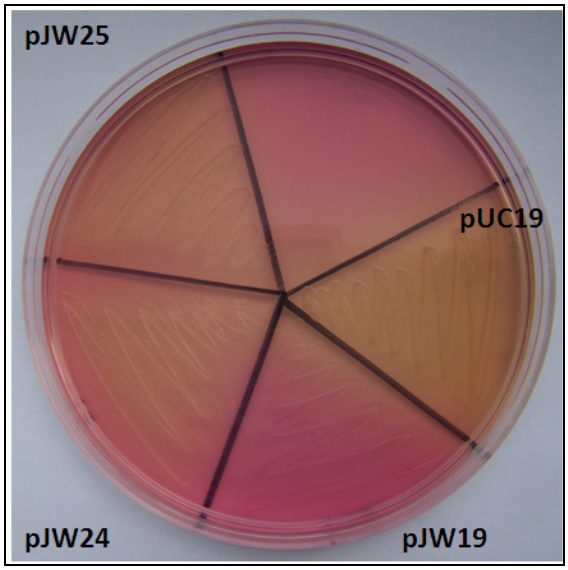


Figure 5-22: Screening of pJW24 and pJW25 MacConkey medium supplemented with xylose

MacConkey media supplemented with xylose, ampicillin and 40mg/ml IPTG was used to screen for complementation in *E.coli* XK100 *xylB*$^{-}$ mutant strain, naturally unable to utilise xylose, when transformed with pJW24, pJW25 and pJW19 a positive control (*xylB* of *C. beijerinckii*) the negative control pUC19 (carrying no genes associated with xylose use) a negative fermentation phenotype was observed with pJW24, pJW25 and pUC19 and a positive phenotype with pJW19.

5.4.6 L-ribulose 5-Phosphate 4-Epimerase (Putative)

5.4.6.1 PCR amplification and pDNA purification

The putative *araD* gene was amplified and purified in the same manner as the *araA* and the putative *araB* genes. This produced a <1500bp product, (the expected size was 1349bp) as shown in Figure 5-23.

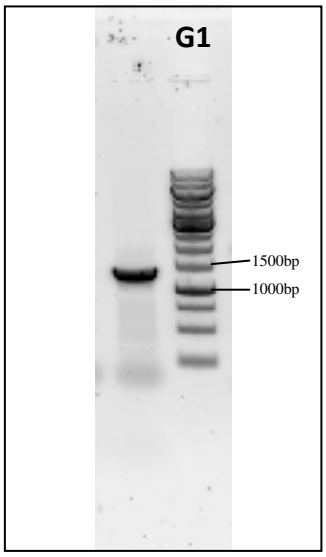


Figure 5-23: PCR amplification of CBEI4455

ORF CBEI4455 was amplified by PCR and ran on a 1% agarose gel alongside Generuler 1kb DNA ladder (**G1**) (Fermentas) was used as a size standard. This showed a PCR product between 1000-1500bp in size.

5.4.7 Cloning and restriction analysis

E. coli TOP10 cells were transformed with CBEI4455 DNA which was purified and cloned into the pCR2.1-TOPO cloning vector. Screening and harvesting of the transformed *E. coli* TOP10 cells was performed as previously described and pDNA was extracted. The orientation of the cloned gene was ascertained using the restriction enzyme *Xba*I. The possible orientations of the gene are shown in Figure 5-24. Cloning under control of the *lac* promoter would result in two fragments of 1353 or 3927bp following digestion with *Xba*I and under control of the T7 promoter would result in fragments of 104 and 5176bp.

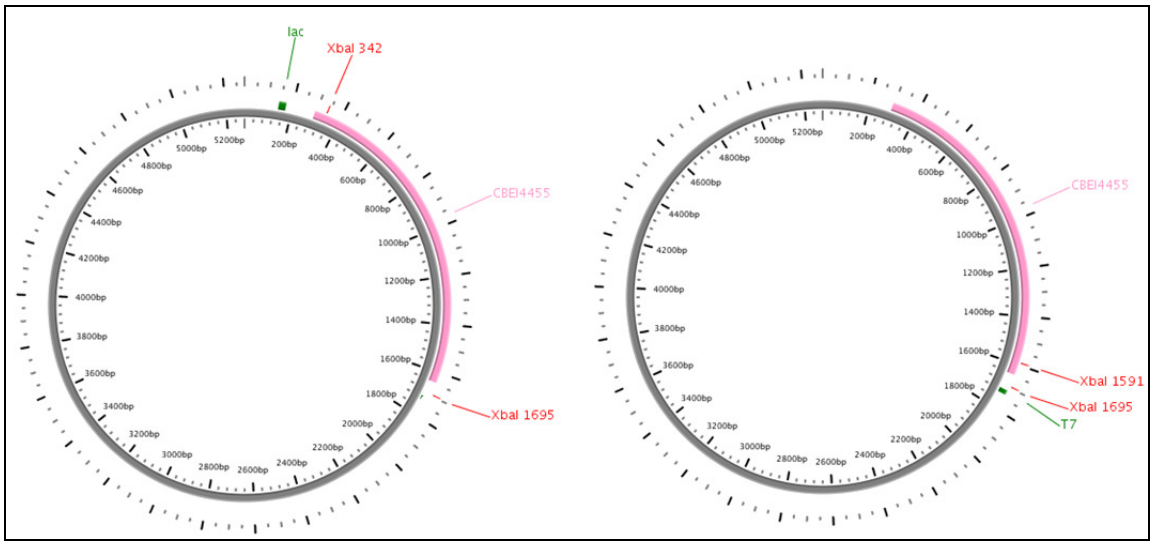


Figure 5-24: Possible orientations of the cloned CBEI4455

This diagram shows the possible orientations of CBEI4455(—) cloned into pCR2.1-TOPO vector (—). On the left is the orientation with expression under control of the *lac* promoter (—), *Xba*I (—) restriction digest would produce 2 fragments of 1353bp and 3927bp. On the right is the orientation with expression under control of the T7 promoter (—), *Xba*I (—) restriction digest would produce 2 fragments of 104bp and 5176bp

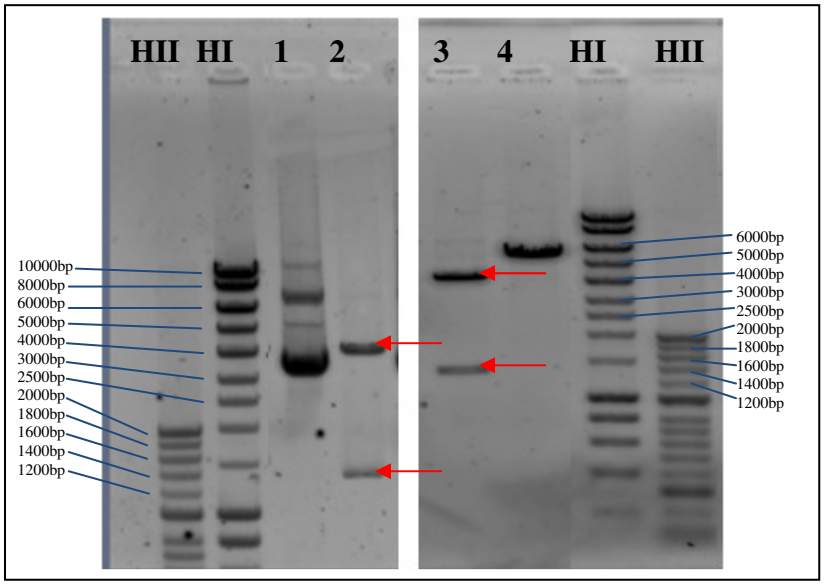


Figure 5-25: *Xba*I restriction digest of pJW21 and pJW22

HII-Hyperladder II (Bioline), **HI**-Hyperladder I (Bioline), **1**-pJW22 undigested, **2**-pJW23 *Xba*I treated, **3**- pJW23 *Xba*I treated, **4**-pJW23 untreated

Restriction digestion of pJW22 and pJW23 with *Xba*I resulted in 2 fragments of ~4000bp and ~1500bp (Figure 5-25) consistent with CBEI4455 being cloned in the direction of expression under control of the *lac* promoter (Figure 5-24).

5.4.8 Complementation screening *in vivo*

Unfortunately the strain procured thought to contain an *araD*$^{-}$ mutation was found to also have mutations in *araABR* genes as well. Unfortunately another *araD* mutant could not be obtained within the time remit of this research and thus the clones are still to undergo complementation screening to characterise them.

5.5 Discussion

LC-ESI-MS/MS was used to qualitatively observe the genes expressed by *C. beijerinckii* when grown on the pentose sugar xylose. *In silico* studies (Chapter 4) had identified potential candidates involved in xylose utilisation and most of these candidates were present in the analyses. Although LC-ESI-MS/MS is a powerful technique for providing a picture of proteins required for all aspects of pentose sugar use, there are a number of drawbacks which could explain the low (~5 %) coverage of the proteome being present in the analysis. Proteins may have been subject to degradation on a number of fronts, such as from the freeze-thawing of samples and natural proteolytic compounds in the environment. The analysis itself produces large volumes of peptide spectra, and it may be that some of these spectra 'mask' other peptide spectra, additionally some of the data may not have been significant at $p<0.05$. It is beneficial to carry out such analyses out in triplicate. The data presented here are from a single analysis, and therefore further analyses may reveal more proteins. Additionally, it would have been of interest to do a proteomic analysis of whole-cell protein extracts of *C. beijerinckii* cultures grown on glucose to investigate if the proteins characterised and putatively involved in arabinose and xylose use are present. However, unfortunately this is an expensive technique and thus out of the remit of this project. The technique does not provide quantitative data so does not provide a clear cut picture that the genes putatively involved in xylose utilisation are definitely involved in xylose utilisation, but rather the genes are expressed on this sugar. However, if taken together with the two genes characterised here and the *in silico* studies, it seems very likely that the xylose utilisation gene system proposed in Chapter 4 is indeed involved in xylose utilisation in *C. beijerinckii*.

Candidates for *xylAB* in *C. beijerinckii* were previously identified by *in silico* studies (Chapter 4) and observed here in the proteomic analysis. These candidates were cloned and screened for complementation *in vivo* using MacConkey agar as a screening tool and both complemented the absent *xylA* or *xylB* activity in the *E. coli* hosts used.

Candidates for *araABD* in *C. beijerinckii* were previously identified by *in silico* studies (Chapter 4) and from the proteomic analysis (this chapter). *araAB* candidates were also cloned and screened for complementation *in vivo* using MacConkey agar as a screening tool. However, only the insertion of CBEI4457 complemented the absence of *araA* activity in the *E. coli* host used. It is not possible to say whether the *araB* candidate,

CBEI4452, does or does not encode AraB, here because of the nature of the *araB*$^{-}$ mutant used and therefore an appropriate *araB*- mutant would need to be obtained which has a fully-functional *araA* gene. Based on the very high level of sequence homology of the AA sequence of CBEI4452 had with *C. glutamicum* AraB (Figure 4-10) and the characterisation of an adjacent gene *araA*, as well as the results from the proteomic analysis of *C. beijerinckii* grown on arabinose, it is quite likely *C. glutamicum* AraB shares a similar function with CBEI4452. Work began on the putative *araD* gene, CBEI4455, however the mutant strain that was to be used for screening was discovered to carry other deletions of other *ara* utilisation genes and time restraints meant this could not be completed. In the future a more suitable *araD*$^{-}$ mutant could be obtained and the experiment completed.

Chapter 6:
Summary and Future Research

6 Summary and Future Research

6.1 Summary

In the light of oil reserves reaching exhaustion in the near future coupled with the greenhouse gas emissions associated with its combustion it is necessary to seek alternative and sustainable fuel sources in order to secure future fuel sources as well as meeting legislative targets of greenhouse gas reduction and mitigation of nations around the World (United Nations, 1998). One way of addressing this is through the use of alcohols as fuels. In particular butanol, which can be produced from the ABE fermentation process, has superior fuel qualities to one of the most current widely used alcohol fuels ethanol.

Lignocellulose is found in plant-based material and is a constituent of plant cell walls. It is a cheap, abundant and sustainable as a substrate for ABE fermentation. Lignocellulosic waste can come from a plethora of sources, such as, municipal, forestry, agricultural and industrial (Zaldivar *et al.*, 2001). Recently an industrial waste stream in the form of spent grains and pot ale from the Scottish whisky industry has been employed and demonstrated as a viable substrate for the ABE production here at the Biofuel Research Centre. This is a waste stream which has low value, is a disposal nuisance and cost to the industry can made valuable by conversion to biofuel. The reason that spent grains are useable after yeast fermentation is because yeast is unable to utilise pentose sugars and as such leaves a glucose depleted and pentose-rich substrate which can be fermented further (Hahn-Hagerdal *et al.*, 2006).

The purpose of this study was to investigate if *C. beijerinckii* could utilise the pentose sugars arabinose and xylose. Strains of clostridia are known to be capable of pentose sugar fermentation and *C. beijerinckii* has been observed here as being able to grow on both pentose sugars as a sole carbon source and to produce similar yields of total solvents to each other but lower than on glucose (Chapter 3). However what was evident was not much difference in the butanol yields between glucose, xylose and arabinose, despite more pentose sugar being present at the end of 72 hour fermentation. Differences in the ratios of acetone to butanol were apparent dependent upon the sugar *C. beijerinckii* was grown on, a finding also observed in *C. acetobutylicum* (Mes-Hartree, 1982, Ounine *et al.*, 1983), whereby the acetone: butanol ratio produced was higher on pentose sugars in comparison to glucose. The levels of both butyric and

acetic acids at the fermentation end-point were higher for cultures grown on pentose sugars in comparison to cultures grown on glucose.

When grown on mixed sugar media (glucose with a pentose sugar) at 1% (w/v) sugar concentration, *C. beijerinckii* was observed as utilising glucose in preference to both pentose sugars in media, regardless of what sugars the cultures were established on prior to the experiment. Conversely, on high sugar concentrations (6% w/v), regardless of the sugar cultures were established on, pentose sugars were used at a quicker rate than glucose, in media containing a higher concentration of pentose sugars. This finding suggests although there is a repression in the presence of glucose (based on the findings of utilisation preferences on low sugar fermentations), it may be that the mechanism of repression is influenced in a concentration dependent manner. This knowledge is invaluable to future biofuel production using lignocellulose-based waste streams, which are made up of multiple carbohydrates. If the pentose sugars are in excess they will be utilised alongside glucose. This is of benefit as it avoids the preferred carbon source being depleted and pentose sugar wastage, when cultures maybe reaching the fermentation end-point.

It would be valuable to gain insight into whether any other sugars prevalent in lignocellulosic wastes also have an effect on the utilisation. It has been noted in several solventogenic strains of clostridia although sugars are used simultaneously, when in high concentration, some more rapidly or preferentially to others and this is strain dependent. *C. beijerinckii* BA101 and *C. beijerinckii* NCP260 have both been observed as using cellobiose and glucose in preference to pentose sugars. *C. beijerinckii* BA101 used pentose sugars in preference over mannose and galactose, whereas *C. beijerinckii* NCP260 uses mannose in preference over arabinose, arabinose is used in preference over galactose (Ezeji *et al.*, 2007, Ezeji and Blaschek, 2008). In previous studies of *C. beijerinckii* NCIMB8052, glucose has exerted a strong inhibition on the utilisation or expression of gene systems involved in glucitol, galactose and mannitol (Mitchell *et al.*, 1995, Behrens *et al.*, 2001). However fructose was observed as being simultaneously utilised alongside glucose. Additionally in the presence of excess mannose, the phosphorylation of glucose has been observed as prevented by the glucose PTS (Mitchell *et al.*, 1995, Mitchell *et al.*, 1991).

For the future of the biofuel industry it is necessary to increase the efficiency and therefore make the process more economically viable, therefore it is necessary to use a cheap and plentiful substrate. To obtain the most value out of the substrate, it needs to be as wholly used as possible, therefore it is of an advantage for the solventogenic organism to be able to use the polymer (cellulose and hemicellulose) directly to reduce costs incurred by resources such as enzymes required to breakdown the substrate into a form where it is directly useable by the organism. *C. beijerinckii* cannot utilise cellulose directly without the aid of the addition of cellulases to the culture medium (Lopez-Contreras *et al.*, 2001). Therefore enzymes, co-cultures (with *C. thermocellum* or *C. cellulolyticum*) or genetic modification would be necessary for *C. beijerinckii* to be able to utilise the cellulose fraction. In this study *C. beijerinckii* was capable of utilising xylan and xylan hydrolysates directly and was aided by the addition of Ca_2CO_3 as a buffer (Chapter 3). However the solvent production by *C. beijerinckii* cultures grown on xylan was observed as poor.

Investigating further into xylanase activity of *C. beijerinckii* could provide the necessary information to improve utilisation of xylan. It was observed in *C. acetobutylicum* that both XynA and XynB had optimum activities at pH 5.0 and 60°C for XynA and 70°C for XynB. Also of interest XynB was found to be partly inhibited by 1mM $FeCl_2$, $CuSO_4$ and $CdCl_2$ and strongly inhibited by $HgCl_2$ (Ali *et al.*, 2004, 2005). Such information could be highly useful in order to achieve the best conditions for xylan breakdown in order to achieve as much product from the substrate as possible. However, it is probably expected that on mixed sugar fermentations *C. beijerinckii* will utilise the smaller carbohydrate units in preference, especially if there is an excess of glucose thus xylan is likely to be wasted before the fermentation endpoint. Therefore a more fruitful line of research would be to look at ways to breakdown xylan into smaller constituent units in the most economically viable way, perhaps by using co-cultures or by genetic modification.

Based on the results of Chapter 3, work had begun to provide evidence of the possession of xylanases in *C. beijerinckii*. A zymogram using whole cell protein extracts of *C. beijerinckii* grown on minimal medium supplemented with xylan were made in view of using the supernatant and cell debris being used for native PAGE. Using gels supplemented with xylan and looking for the presence of clearing zones via employing congo red staining and destaining using a NaCl solution (Geib *et al.*, 2010).

Unfortunately due to time constraints this work could not be completed. An alternative route would be to clone the putative xylanase genes and over-express the putative genes and purify the protein for measuring the activity of the protein and its activity under different conditions such as temperature and pH, using the methodology of (Jalal *et al.*, 2009).

The extent of resources on the planet that can be used for biofuel production is vast. Kim and Dale (2006) estimated there is 73.9 teragrams annually of waste crops that could be used to make 49.1 gigalitres of fuel. The specific resource that could be utilised by a particular region of the World would depend upon the native species grown or consumed. Resources for fermentations can come from forestry, agricultural, municipal and industrial sectors.

Agricultural resources can be by-products of grain, oil-seed, fruit and vegetable harvesting, such as stalks, seeds, shells, husks, straw, sludge, wastewater and juice. Forestry and paper-pulp mills, including wood, bark, leaves, sawdust, fibre and black liquor. Municipal and industrial wastes such as food waste, paper, paper-pulp wastes, card and wood (Howard *et al.*, 2003). Resources abundant for solventogenesis are dependent upon the area of the World. For example in North America an estimated 60-80 million dry tonnes of corn stover available in the USA for fermentations. Whereas waste paper, plant residues, sawdust and fruit and vegetable waste encompasses a considerable amount of the solid-based wastes in Tanzania (Mtui and Nakamura, 2005). Therefore it would be necessary for nations to identify and test the suitability of a potential feedstock.

Although there are countless waste streams that have been investigated and could be used for clostridial fermentations, research is required into the best methods to release the carbon source, if required, in terms of the economics, the effect on the fermentation equipment and the organism and limiting the loss of sugars, by the chemical or physical methods used (Galbe and Zacchi, 2002). The more the costs associated with this can be reduced and the more wholly a resource can be used, the more biobutanol and economic benefit can be derived. Indeed in this study the use of DDGS as a fermentation substrate was investigated (Chapter 3). The total solvent yields obtained on acid and enzymatically treated DDGS was less in comparison to soluble sugars, suggesting that inhibitor compounds may be present from the treatment of DDGS. Therefore ways of

reducing or preventing the effect of inhibitory compounds would have been highly beneficial here. XAD-4 resin has been used successfully to remove cell growth inhibition of *C. beijerinckii* BA101 attributed to acid-hydrolysed corn fibre (Qureshi *et al*., 2008).

Lignocelluloses are not exhaustive of the resources that could be used and have been investigated for biofuel production in clostridia. Industrial wastes such as glycerol, a product of the transesterification process of oil (Andrade and Vasconcelos, 2003), dairy wastes, including caseins and whey which a protein-rich, lactose and fat (Audic *et al.*, 2003) could also be used as a feedstock for biofuel production. Indeed, *C. acetobutylicum* strains ATCC824 and ATCC 4259 when grown on glucose-glycerol mixtures had an ABE yield of 0.32 g/g and 0.41 g/g, despite the glycerol being of low-grade. Continuous cultures were also stable for 70 days grown on this media, which was thought to provide a selective pressure for plasmid maintenance, thus preventing strain degeneration which often comes with continuous fermentations (Andrade and Vasconcelos, 2003). Such resources and others should be investigated as biofuel production in the future.

Based on the finding *C. beijerinckii* can utilise both arabinose and xylose (Chapter 3) it was then expected that a gene system would be present responsible for the use of both of the pentose sugars. Using a number of bioinformatics analyses (using programmes, BLAST, MrBayes, CLUSTALW2, THMHH, TMRPres2D, TreeView and GeneDoc) and knowledge of the enzymes involved in pentose sugar utilisation in other bacteria, two separate candidate gene systems, one for arabinose (CBEI4448-4457) and one for xylose (CBEI2380-2387) were identified within the genome of *C. beijerinckii* (Chapter 4). In order to firmly establish the candidate systems roles in xylose and arabinose use it was necessary to carry out further studies to confirm their role. Several candidates were targeted for cloning (Chapter 5). The ensuing *in vivo* studies confirmed *xylAB* (CBEI2383 and CBEI2384) and *araA* (CBEI4457) by a positive fermentation phenotype on MacConkey screening medium. The candidate *araBD* genes were also cloned however unsuitable mutants were only available and so these experiments would have to be repeated with host cell mutants carrying a fully-functional arabinose utilisation gene system with only a single gene mutation, in *araB* or *araD*. Both *araB*

and *araD* candidates had a high level of similarity with published characterised *araBD* genes. Because of the level of similarity with other characterised *araBD* genes, and the *araA* gene being in close proximity, it would be expected that these two are likely candidate *araBD* genes. In *E. coli*, *B. subtilis*, *S. enterica* and *C. glutamicum* all arabinose utilisation genes are adjacent within the arabinose gene system (Lee *et al.*, 1986, Sa-Nogueira and de Lencastre, 1989, Lin *et al.*, 1985b, Kawaguchi *et al.*, 2009) Whereas the candidate arabinose gene system of *C. beijerinckii* there are genes between *araA* and the putative *araB* and *araD* genes (Figure 6-1). The order of the putative arabinose utilisation genes in *C. beijerinckii* is the same as *C. glutamicum araBDA* but different to *E. coli*, *S. enterica* (*araBAD*) and *B. subtilis araDBA*.

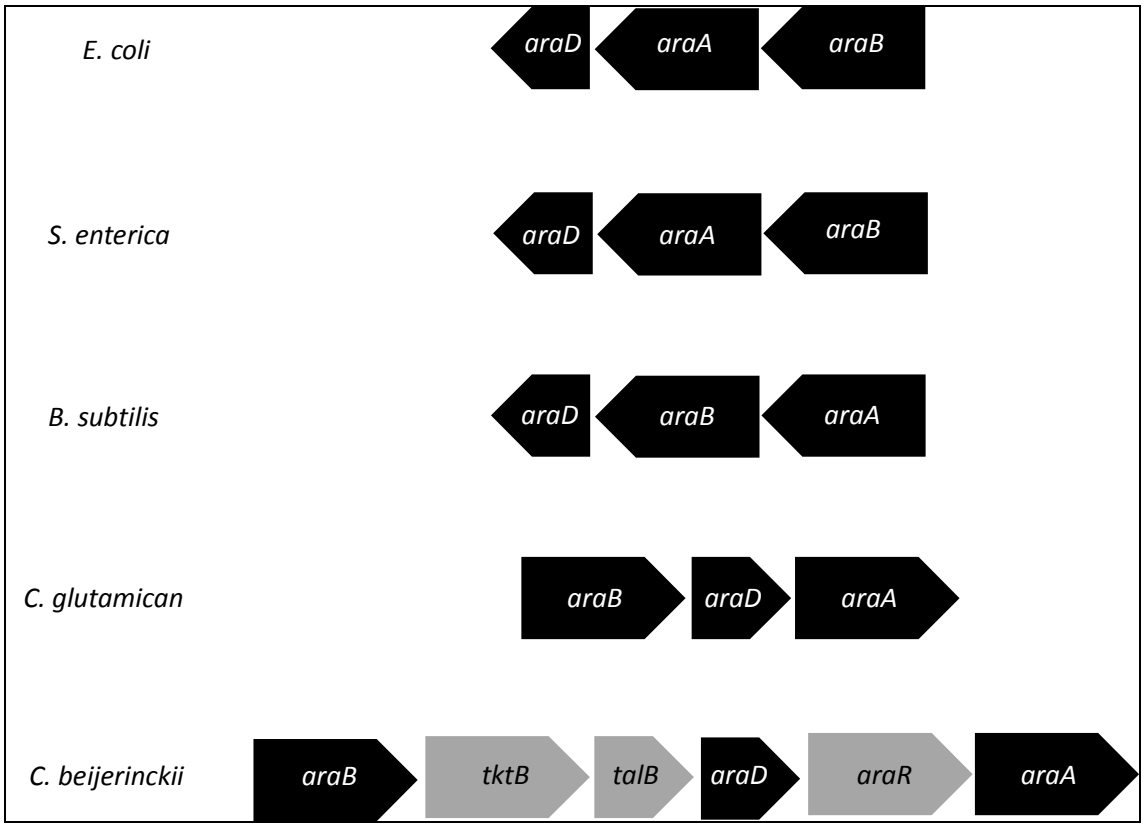


Figure 6-1: Comparison of the arabinose operons

Published characterised arabinose operons in comparison to the putative arabinose gene system of *C. beijerinckii*. The putative arabinose utilisation genes in *C. beijerinckii* are interrupted by other genes. Whereas other organisms arabinose utilisation genes are adjacent to one and another. ***B. subtilis*** (Sa-Nogueira and de Lencastre, 1989); ***C. glutamican*** (Kawaguchi *et al.*, 2009) ***E. coli*** (Lee *et al.*, 1986); ***S. enterica*** (Lin *et al.*, 1985b)

Cloning experiments do not always work and this can be due to the differences between the host and the nature of the organism that the cloned gene comes from. *C. beijerinckii*

is a Gram positive organism with a low G/C content and therefore it is possible this produces a defective protein in a Gram negative host with a high G/C content, such as *E. coli*. The expression of non-native proteins can be a stress to the host cell (Lesley *et al.*, 2002). Although *E. coli* is often a popular host choice for gene cloning, genes of Gram positive organisms can produce unstable clones (Graves and Rabinowitz, 1986) and AT-rich inserts can be unstable in circular vectors (Godiska *et al.*, 2009). Additionally some genes cannot be expressed efficiently in *E. coli* because of particular or unique features of the gene being cloned. Codons that are used more rarely in *E. coli* have been demonstrated to produce a low copy number and thus a low gene expression (Konigsberg and Godson, 1983), so particular genes highly expressed in one organism may not be expressed to the same extent in another simply because of the codon usage differences between species (Sharp *et al.*, 1988). It maybe that mRNA stability or the translational efficiency is different, or the folding of the protein differs and thus is prone to degradation by the host cell's proteases (Makrides, 1996). Proteins which are not properly folded have a tendency to aggregate which may cause inclusion body formation (Fink, 1999) and therefore can be toxic to the host cell. Alternative methods could be employed to characterise the candidate *araBD* genes. Producing knockouts of these genes in *C. beijerinckii* and analysing the ability of the strain to grow on arabinose in minimal medium in comparison to the parent strain. A knockout could be produced by transposon mutagenesis. Tn1545 and Tn916 conjugative transposons have been successfully transferred into *C. beijerinckii* previously (Woolley *et al.*, 1989). However, such a method is very time consuming and laborious. Commercially available knockout kits are available, which are simple to use and rapid. TargeTron (Sigma) has been used successfully in *Clostridium perfringens* and works by inserting a permanent, stable intron to disrupt a specific targeted gene (Sigma-Aldrich, 2008). Obtaining a Gram positive host with *araB* and *araD* mutations and then cloning *C. beijerinckii araB* and *araD* candidates may have success.

Additionally positive controls would be beneficial to check that the screening medium here was not the reason for the negative phenotype observed. A positive control could be made by cloning characterised *araB* or *araD* genes from either *E. coli*, *B. subtilis*, *S. enterica* and *C. glutamicum* ((Sa-Nogueira and de Lencastre, 1989) (Lee *et al.*, 1986) (Lin *et al.*, 1985b); (Kawaguchi *et al.*, 2009) into the $araB^-$ and $araD^-$ mutant host cells used in this study.

Proteomic analysis (Chapter 5) revealed the expression of the candidate *araBD* and *araA* genes when *C. beijerinckii* was grown on arabinose, adding further evidence that these genes have a role in arabinose utilisation. *xylAB* was expressed when *C. beijerinckii* was grown on medium containing xylose. There was also expression of the other candidate genes in both of the gene systems. The candidates *araFG*, *talB*, *tktB* (CBEI4449, 4450, 4453 and 4454) and *xylF*, *talB*, *tktB* (CBEI2380, 2386, 2387) and were also expressed. However evidence of candidate *araH*, *araR* and *xylGH* and *xylR* expression was not observed. It is recommended to get proteomic analyses carried out in triplicate. The analysis itself produces large volumes of peptide spectra, and it may be that some of these spectra 'mask' other peptide spectra (Higdon and Kolker, 2007) or some of the data is not significant at $p<0.05$. Although this method provides a whole-cell protein snap shot of *C. beijerinckii* when grown on arabinose or xylose it does not provide a quantitative result, which could have provided further evidence implicating the role of the candidate genes. Real-time PCR is another tool for analysing gene expression of genes associated with the utilisation of arabinose and xylose. It is a very sensitive method of quantifying mRNA. mRNA could be extracted from cells of *C. beijerinckii* grown on minimal medium containing the pentose sugar as the sole carbon source. Probes could be designed for each individual gene involved in arabinose utilisation. Using a different reporter dye, each gene could be quantified in one reaction (Nolan *et al.*, 2006). The effect of the presence of glucose on the expression of the candidate genes associated with both arabinose and xylose utilisation could be investigated. Looking at the transcripts on both high and low sugar concentrations would provide evidence at the gene level to the effect observed in Chapter 3. Real-time PCR is reliant on a number of factors such as the quality of the mRNA template, which is very vulnerable to degradation by environmental RNases; prevention of contamination to prevent false-positive amplification and degradation; careful design of primers to prevent mis-priming and primer dimers (Udvardi *et al.*, 2008, Nolan *et al.*, 2006, Bustin, 2002).

6.2 Future Research

A three-pronged approach for future research is thus proposed and summarised: firstly, further characterisation of xylose, arabinose and xylan utilisation genes; secondly, further fermentations in relation to the use of these carbohydrates; and finally strain development.

1. Further Characterisation of Genes Involved in Pentose Sugar Utilisation

There are many other avenues which could be explored with regards to pentose sugar utilisation. Of particular value would be to explore the regulation mechanism of arabinose and xylose use, for fermentations with mixed sugar feedstocks. Although simultaneous utilisation of both arabinose and xylose was observed alongside glucose in this study, it would be of benefit to characterise the mechanism for improving pentose sugar use, to provide knowledge for strain development, for the benefit of potential feedstocks where the ratios of pentose sugars are not in excess to glucose.

The expression of genes involved in pentose sugar utilisation in the presence of glucose on low and high concentrations of sugar could be investigated via the use of real-time PCR using mRNA extracted from cultures at different time points on a growth curve. The expression of the putative transporter, repressor as well as utilisation genes could be investigated. Further characterisation of genes that have not been characterised in this study could be carried out using similar methods. Target genes could be amplified, cloned and characterised *in vivo* using strains with single gene mutations. The use of Gram positive hosts, such as bacilli, if available could reduce the likelihood of problems that can be observed when using *E. coli*.

As well as the use of pentose sugars by *C. beijerinckii*, the use of the pentose-polymer xylan, and the use arabinan could also be investigated. Zymogram experiments could be completed to establish the ability of *C. beijerinckii* to utilise xylan before any candidate genes were cloned and characterised using the methodology for genes involved in pentose sugar utilisation. Alternatively, genes involved in xylan utilisation could be cloned and the protein over-expressed then purified and characterised for activity under a range of different conditions to establish the optimum conditions for xylan utilisation by *C. beijerinckii*. These conditions could then be used to manipulate culture conditions and improve xylan utilisation.

2. Utilisation of Carbohydrates

This would encompass the effect other sugars present in lignocellulosic waste have on the utilisation of pentose sugars, by conducting fermentations on the pure-form of the sugars at high and low sugar concentrations. Additionally, the effect other sugars have on the utilisation of pentose polymers, such as xylan could also be investigated in a similar manner. Further lignocellulosic waste streams which are abundant on the planet could be investigated as a feedstock for biobutanol production, perhaps waste streams indigenous to different regions of the World. This could incorporate studies into best ways to degrade the feedstock prior to fermentation, this may use of co-cultures, for example using xylanolytic and cellulolytic clostridia, such as *C. cellulolyticum* or *C. cellulovorans* with *C. beijerinckii*, as the former two do not produce biobutanol.

3. Strain Development

Ultimately a strain which is not subject to any repression of genes associated with sugar utilisation would be of high value so as to negate the wastage of any sugars whatever the nature of the constituents making up a feedstock are. Such a strain was produced in *B. subtilis*. A mutation in the 'global regulator' CcpA prevented catabolite repression exerted by several sugars, including glucose, fructose and mannitol (Hueck and Hillen, 1995). CcpA has been identified in several in Gram positive bacteria with a low G/C content, including *B. megaterium* and *Lactobacillus casei* (Hueck *et al.*, 1995, Monedero *et al.*, 1997) and is responsible for carbon catabolite repression, whereby CcpA binds to catabolite-responsive elements found in the operons involved in the utilisation of the less preferred carbon sources and thus prevents the expression of the genes comprising the operon (Muscariello *et al.*, 2001). Indeed, CcpA homologues have been identified in *C. beijerinckii* (Reid *et al.*, 1999).

6.3 Final conclusions

This research has provided fundamental and important knowledge of the ability of *C. beijerinckii* NCIMB8052 ability to utilise pentose sugars. This is of importance in the light of utilising lignocellulose, a highly abundant and ubiquitous feedstock found on the planet. The ability of the organism to utilise pentose sugars, which are often prevalent in the hemicellulose fraction of lignocelluloses is of importance because the uptake and utilisation of a carbon source is directly related to solvent production and economics of biofuel production. Having a fundamental knowledge of how a carbon source is utilised, opens doors to strain improvement, what sorts of feedstocks are usable and the effect of CCR sugar (such as glucose) has on the utilisation of the feedstock. The techniques used in this study could be applied to other sugars and carbon sources as well as other solventogenic strains of clostridia. Further avenues of research into pentose sugars have been eluded to here as well as highlighting an important a requirement for research for the best and economically viable release of carbon sources in terms of minimising the destruction of the carbon source and the release of inhibitors of high value to biofuel industry in the future.

Chapter 7

Appendices

7 Appendices

7.1 Size standards for gel electrophoresis used in this study

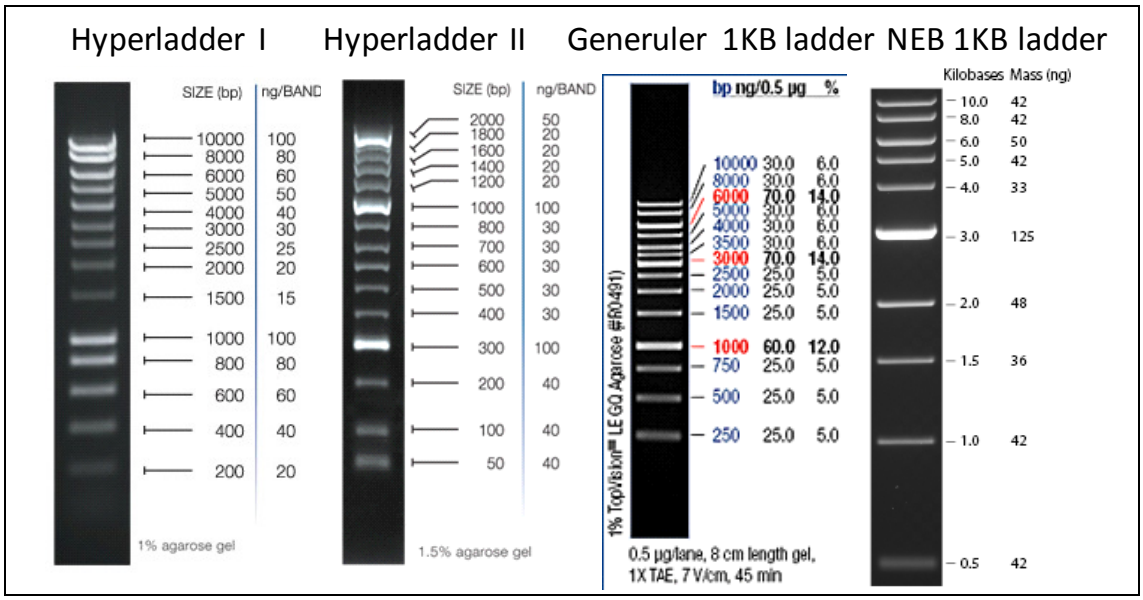


Figure 7-1: Hyperladder I and II and Generuler size standards

Hyperladder I and **II** and **Generuler** were used as DNA size standards during electrophoresis

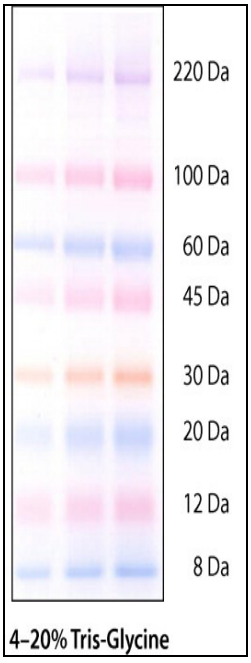


Figure 7-2: Protein Marker

Colourburst electrophoresis marker (Sigma) was ran alongside samples during SDS-PAGE

7.2 Bicinchoninic Acid Protein Concentration Assay Standard Curve

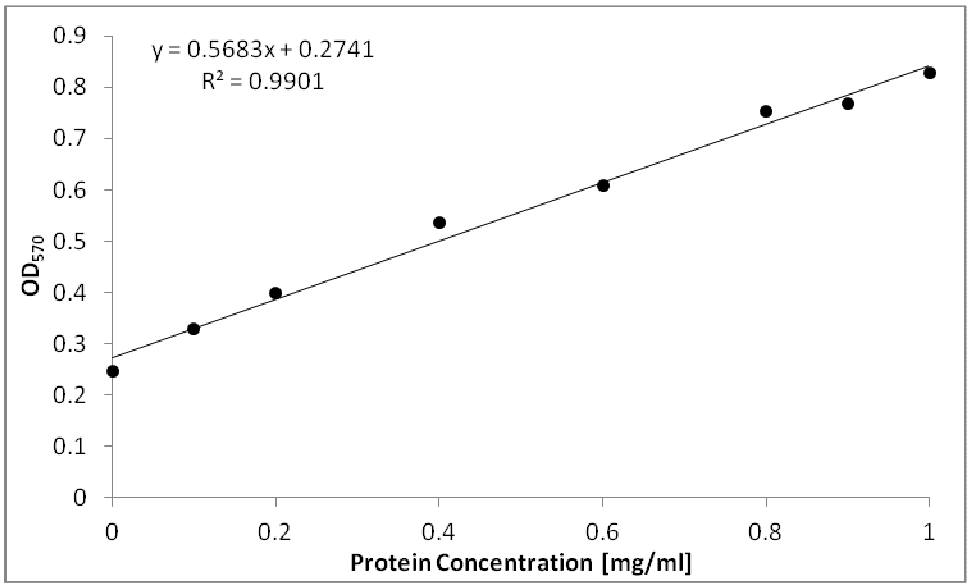
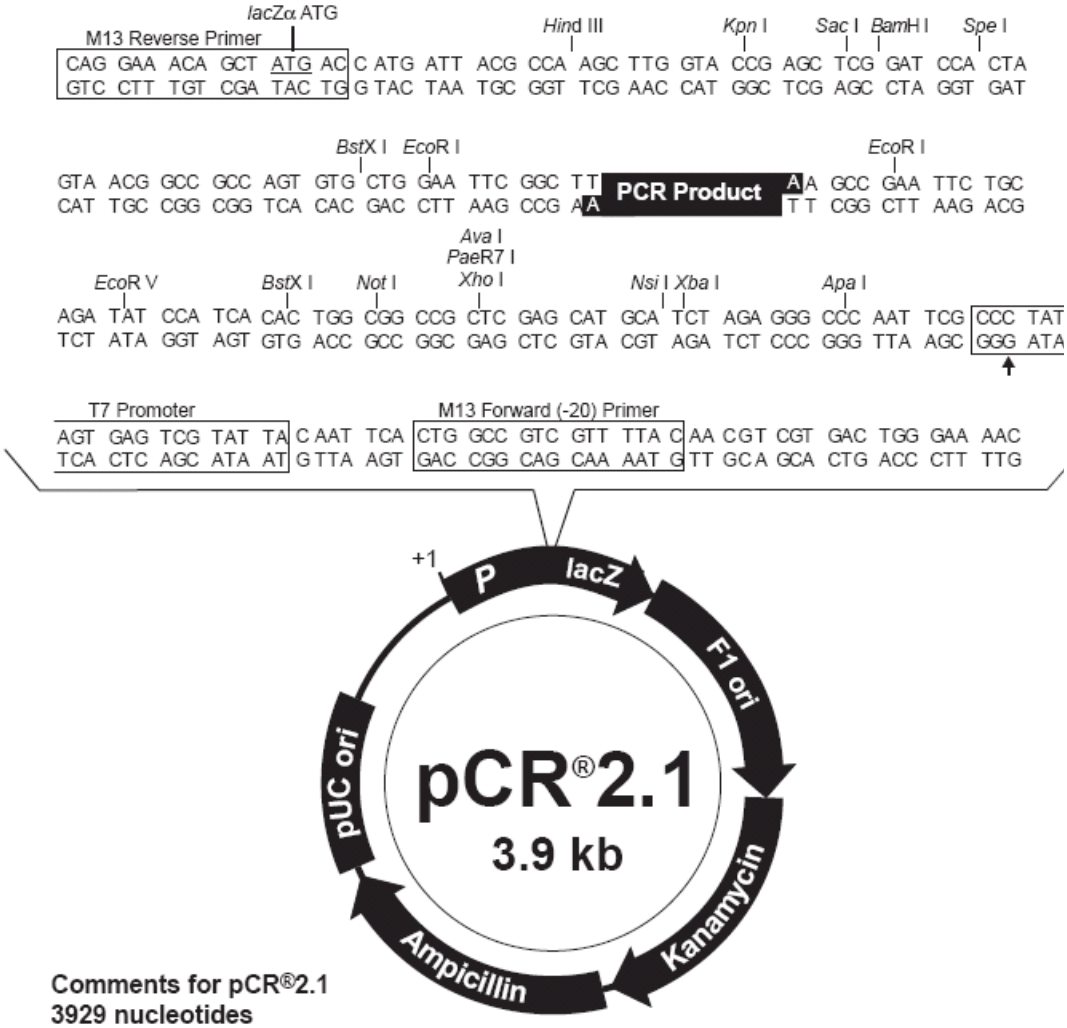


Figure 7-3: Bicinchoninic Acid Protein Concentration Assay Standard Curve

BCA assay standard curve was carried out using BSA of known concentrations

7.3 pCR2.1 vector

Map



Comments for pCR®2.1
3929 nucleotides

LacZα gene: bases 1-545
M13 Reverse priming site: bases 205-221
T7 promoter: bases 362-381
M13 (-20) Forward priming site: bases 389-404
f1 origin: bases 546-983
Kanamycin resistance ORF: bases 1317-2111
Ampicillin resistance ORF: bases 2129-2989
pUC origin: bases 3134-3807

pCR2.1 Sequence

AGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATCCATTAATGCAGCTGGCACGACAGGTT
TCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAG
GCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAA
ACAGCTATGACCATGATTACGCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGC
TGGAATCCGGCTTAAGCCGAATCCTGCAGATATCCATCACACTGGCGGCCGCTCGAGCATGCATCTAGAG
GGCCCAATCCGCCCTATAGTGAGTCGTATTACAATCCACTGGCCGTCGTTTTACAACGTCGTGACTGGGA
AAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAA
GAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGG
CGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCC
GCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGG
GGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATAGGGTGATGGT
TCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATA
GTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATCCTTTTGATTTATAAGGGAT
TTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAA
ATCCAGGGCGCAAGGGCTGCTAAAGGAAGCGGAACACGTAGAAAGCCAGTCCGCGAAACGGTGCTGACCC
CGGATGAATGTCAGCTACTGGGCTATCTGGACAAGGGAAAACGCAAGCGCAAAGAGAAAGCAGGTAGCTT
GCAGTGGGCTTACATGGCGATAGCTAGACTGGGCGGTTTTATGGACAGCAAGCGAACCGGAATTGCCAGC
TGGGGCGCCCTCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATC
TGATGGCGCAGGGGATCAAGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATG
GATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATCCGGCTATGACTGGGCACAACAGACAAT
CGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGAC
CTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTC
CTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGG
GCAGGATCTCCTGTCATCCCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGG
CTGCATACGCTTGATCCGGCTACCTGCCCATCCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTA
CTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGA
ACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGC
TTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATCCATCGACTGTGGCCGGCTGGGTGTGGCGG
ACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCG
CTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATCCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAG
TTCTTCTGAATTGAAAAAGGAAGAGTATGAGTATCCAACATTTCCGTGTCGCCCTTATCCCCTTTTTTGC
GGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTG
GGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAG
AACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGG
GCAAGAGCAACTCGGTCGCCGCATACACTATCCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAA
AAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTG
CGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGA
TCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACC
ACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCC
GGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGC
TGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGG
CCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAA
ATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATA
TATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAAT
CTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAG
GATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGC
GGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAG
ATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTA
CATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTT
GGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCC
AGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTC
CCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCT
TCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTT
TTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGG
CCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATCCTGTGGATAACCGTATTAC
CGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAA
GCGGAAG

Cloning site

7.4 Proteomic Analysis of *C. beijerinckii* Grown on Arabinose

GI #	Gene label	Mass (Mr)	Mowse score
gil150014899	DNA gyrase subunit A	92723	182
gil150014913	ABC transporter related	59727	296
gil150014925	peptidylprolyl isomerase	19103	180
gil150014967	phosphoserine aminotransferase	40121	491
gil150014980	AbrB family transcriptional regulator	20286	114
gil150014983	histone family protein DNA-binding protein	10127	170
gil150014991	hypoxanthine phosphoribosyltransferase	20166	236
gil150014993	formate--tetrahydrofolate ligase	60745	367
gil150014996	transcription elongation factor GreA	17769	89
gil150014997	lysyl-tRNA synthetase	57333	105
gil150015006	hypothetical protein Cbei_0114	12071	100
gil150015028	elongation factor Tu	43653	745
gil150015032	ribosomal protein L11	14988	549
gil150015033	50S ribosomal protein L1	24623	712
gil150015034	50S ribosomal protein L10	18275	657
gil150015035	ribosomal protein L7/L12	12592	143
gil150015036	DNA-directed RNA polymerase subunit beta	139257	591
gil150015037	DNA-directed RNA polymerase subunit beta'	132490	175
gil150015038	30S ribosomal protein S12	13758	88
gil150015039	30S ribosomal protein S7	17661	1313
gil150015040	elongation factor G	76160	1130
gil150015042	30S ribosomal protein S10	11502	58
gil150015043	50S ribosomal protein L3	22768	811
gil150015044	50S ribosomal protein L4	22863	742
gil150015046	50S ribosomal protein L2	30397	490
gil150015049	30S ribosomal protein S3	24581	541
gil150015050	50S ribosomal protein L16	16342	176
gil150015051	50S ribosomal protein L29	8116	113
gil150015053	50S ribosomal protein L14	13212	353
gil150015055	50S ribosomal protein L5	20424	1017
gil150015057	30S ribosomal protein S8	14756	419
gil150015058	50S ribosomal protein L6	19874	570
gil150015059	50S ribosomal protein L18	13196	109
gil150015060	30S ribosomal protein S5	17312	947
gil150015061	50S ribosomal protein L30	6610	396
gil150015062	50S ribosomal protein L15	15648	458
gil150015069	30S ribosomal protein S13	13847	667
gil150015070	30S ribosomal protein S11	14080	258
gil150015071	30S ribosomal protein S4	24032	403
gil150015072	DNA-directed RNA polymerase subunit alpha	35215	178
gil150015073	50S ribosomal protein L17	12814	291
gil150015078	50S ribosomal protein L13	16340	603
gil150015079	30S ribosomal protein S9	14476	412
gil150015088	phosphoenolpyruvate-protein phosphotransferase	59644	571
gil150015095	phosphate butyryltransferase	32592	1196
gil150015096	butyrate kinase	38524	546
gil150015137	mannitol dehydrogenase domain-containing protein	42461	43
gil150015138	glucosamine--fructose-6-phosphate aminotransferase, isomerising	66893	611
gil150015152	RND family efflux transporter MFP subunit	52841	187
gil150015178	rubrerythrin	22497	123
gil150015211	3-hydroxybutyryl-CoA dehydratase	28449	1316
gil150015212	acyl-CoA dehydrogenase domain-containing protein	41307	2121
gil150015213	electron transfer flavoprotein, alpha/beta-subunit-like protein	27918	1242
gil150015214	electron transfer flavoprotein, alpha/beta-subunit-like protein	35994	1372
gil150015215	3-hydroxybutyryl-CoA dehydrogenase	30618	2641
gil150015218	co-chaperonin GroES	10149	340

gil150015219	chaperonin GroEL	57575	4717
gil150015221	inosine 5'-monophosphate dehydrogenase	52586	145
gil150015222	GMP synthase	57592	87
gil150015279	glutathione peroxidase	20334	104
gil150015281	peptidoglycan-binding LysM	59124	974
gil150015297	RpiB/LacA/LacB family sugar-phosphate isomerase	18060	128
gil150015301	acetyl-CoA acetyltransferase	41327	3761
gil150015304	F0F1 ATP synthase subunit B	18305	129
gil150015305	F0F1 ATP synthase subunit delta	21314	132
gil150015306	F0F1 ATP synthase subunit alpha	55366	486
gil150015313	peptidase M23B	27546	233
gil150015318	S-adenosylmethionine synthetase	43543	122
gil150015323	pre-protein translocase subunit SecA	96025	220
gil150015334	glutamine synthetase, catalytic region	77489	292
gil150015376	aminoacyl-histidine dipeptidase	53996	294
gil150015398	50S ribosomal protein L21	11301	598
gil150015422	transketolase	72429	748
gil150015455	putative serine protein kinase, PrkA	73669	1019
gil150015466	cysteine synthase	31988	532
gil150015486	glyceraldehyde-3-phosphate dehydrogenase, type I	35775	1659
gil150015487	phosphoglycerate kinase	42136	482
gil150015488	triosephosphate isomerase	27162	1089
gil150015489	phosphoglyceromutase	56514	320
gil150015491	phosphopyruvate hydratase	47201	1578
gil150015524	5'-nucleotidase domain-containing protein	119048	104
gil150015527	YaeC family lipoprotein	29814	114
gil150015546	methyltransferase type 11	21713	220
gil150015568	O-methyltransferase family protein	23751	126
gil150015570	hypothetical protein Cbei_0682	10093	135
gil150015571	radical SAM domain-containing protein	50434	644
gil150015577	hypothetical protein Cbei_0689	11485	123
gil150015644	phosphotransferase system, lactose/cellobiose-specific IIB subunit	11293	465
gil150015651	alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen	17828	89
gil150015652	O-acetylhomoserine/O-acetylserine sulfhydrylase	46363	128
gil150015659	pepF/M3 family oligoendopeptidase	68210	137
gil150015682	UspA domain-containing protein	16476	270
gil150015712	hypothetical protein Cbei_0826	16769	377
gil150015716	molecular chaperone DnaK	65416	1087
gil150015725	GatB/Yqey domain-containing protein	16935	119
gil150015735	pyruvate phosphate dikinase	97049	786
gil150015751	glycogen/starch/alpha-glucan phosphorylase	91707	84
gil150015868	ribosomal 5S rRNA E-loop binding protein Ctc/L25/TL5	20762	139
gil150015869	phosphoglucomutase/phosphomannomutase alpha/beta/subunit	64725	424
gil150015885	beta-lactamase, putative	41307	114
gil150015895	formate acetyltransferase	83631	1492
gil150015897	formate acetyltransferase	83456	2316
gil150015921	hypothetical protein Cbei_1035	11843	95
gil150015929	branched-chain amino acid aminotransferase	38271	232
gil150015932	band 7 protein	34759	192
gil150015939	phosphoribosylformylglycinamidine synthase	137920	198
gil150015945	bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase	56438	218
gil150016048	phosphotransacetylase	36053	71
gil150016060	30S ribosomal protein S16	9229	303
gil150016064	50S ribosomal protein L19	13359	139
gil150016072	30S ribosomal protein S2	26356	198
gil150016073	elongation factor Ts	34123	252
gil150016083	transcription elongation factor NusA	44957	103

gil150016086	translation initiation factor IF-2	76480	271
gil150016091	30S ribosomal protein S15	10189	203
gil150016092	polynucleotide phosphorylase/polyadenylase	77276	216
gil150016111	6,7-dimethyl-8-ribityllumazine synthase	16396	326
gil150016121	S-ribosylhomocysteinase	18301	95
gil150016208	trigger factor	48472	747
gil150016209	ATP-dependent Clp protease, proteolytic subunit ClpP	22056	300
gil150016219	hypothetical protein Cbei_1337	12591	573
gil150016286	hypothetical protein Cbei_1406	15725	211
gil150016296	rubrerythrin	20007	348
gil150016307	glutamyl-tRNA synthetase	63942	119
gil150016309	glutaminyl-tRNA synthetase	64207	87
gil150016338	pyruvate flavodoxin/ferredoxin oxidoreductase domain-containing protein	130496	358
gil150016362	hypothetical protein Cbei_1482	35086	162
gil150016373	hypothetical protein Cbei_1494	24923	777
gil150016374	hypothetical protein Cbei_1495	35782	567
gil150016385	superoxide dismutase	23152	378
gil150016417	adenine phosphoribosyltransferase	18747	200
gil150016423	aspartyl-tRNA synthetase	68424	136
gil150016445	translation initiation factor IF-3	21563	200
gil150016447	50S ribosomal protein L20	13621	272
gil150016456	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase B	55243	112
gil150016464	alanine racemase domain-containing protein	25848	223
gil150016564	elongation factor P	22075	274
gil150016573	stage III sporulation protein AH	19180	233
gil150016574	hypothetical protein Cbei_1700	13826	154
gil150016586	response regulator receiver protein	30721	405
gil150016596	iron-containing alcohol dehydrogenase	43043	552
gil150016611	NAD(P)H dehydrogenase (quinone)	22331	199
gil150016623	anthranilate synthase component I	53456	258
gil150016624	glutamine amidotransferase of anthranilate synthase	21769	243
gil150016625	anthranilate phosphoribosyltransferase	36466	178
gil150016626	indole-3-glycerol-phosphate synthase	31355	77
gil150016628	tryptophan synthase subunit beta	43014	1367
gil150016629	tryptophan synthase, alpha subunit	28702	180
gil150016663	single-stranded DNA-binding protein	24874	112
gil150016668	dihydrodipicolinate reductase	27720	80
gil150016673	Hsp33-like chaperonin	32229	60
gil150016698	heat shock protein DnaJ domain-containing protein	18975	64
gil150016741	serine hydroxymethyltransferase	45243	52
gil150016743	hypothetical protein Cbei_1870	9608	103
gil150016747	Serine-type D-Ala-D-Ala carboxypeptidase	44299	163
gil150016776	fructose-1,6-bisphosphate aldolase, class II	30931	769
gil150016781	membrane -like protein	21839	86
gil150016830	hypothetical protein Cbei_1958	19522	216
gil150016840	aspartate/ornithine carbamoyltransferase family protein	45081	137
gil150016904	3-hydroxybutyryl-CoA dehydratase	28556	108
gil150016905	acyl-CoA dehydrogenase domain-containing protein	41391	1758
gil150016907	electron transfer flavoprotein, alpha/beta-subunit-like protein	36021	130
gil150016983	inosine 5-monophosphate dehydrogenase	55868	197
gil150017050	iron-containing alcohol dehydrogenase	42947	123
gil150017058	nitroreductase	20612	138
gil150017147	hydroxylamine reductase	60892	554
gil150017192	methionine synthase, vitamin-B12 independent	43641	231
gil150017213	hypothetical protein Cbei_2351	10018	185
gil150017248	putative transaldolase	23164	1401
gil150017249	transketolase	72544	1005
gil150017281	iron-containing alcohol dehydrogenase	43050	113

gil150017287	xanthine phosphoribosyltransferase	21248	160
gil150017311	electron transport complex, RnfABCDGE type, G subunit	20254	62
gil150017349	rubrerythrin	20009	349
gil150017371	extracellular solute-binding protein	31572	401
gil150017403	4-hydroxy-3-methylbut-2-enyl diphosphate reductase/S1 RNA-binding domain protein	71710	230
gil150017407	Ferritin, Dps family protein	20036	136
gil150017452	hypothetical protein Cbei_2594	25673	260
gil150017453	hypothetical protein Cbei_2595	29611	95
gil150017459	dTDP-4-dehydrorhamnose 3,5-epimerase	22223	119
gil150017467	pantoate--beta-alanine ligase	31367	222
gil150017468	3-methyl-2-oxobutanoate hydroxymethyltransferase	29773	735
gil150017508	3-oxoacid CoA-transferase, B subunit	22977	416
gil150017509	3-oxoacid CoA-transferase, A subunit	25671	321
gil150017530	coagulation factor 5/8 type domain-containing protein	80810	471
gil150017535	redoxin domain-containing protein	18083	307
gil150017593	phosphotransferase system, lactose/cellobiose-specific IIB subunit	10742	88
gil150017595	putative translaldolase	26780	65
gil150017661	hypothetical protein Cbei_2809	19517	120
gil150017731	acyl-CoA dehydrogenase domain-containing protein	41968	110
gil150017734	electron transfer flavoprotein, alpha subunit-like protein	32922	70
gil150017861	carbon-monoxide dehydrogenase, catalytic subunit	70862	73
gil150017880	fructose-1,6-bisphosphate aldolase	33328	277
gil150017884	hypothetical protein Cbei_3043	22683	185
gil150018108	class II aldolase/adducin family protein	24899	370
gil150018109	coenzyme A transferase	56285	1244
gil150018155	hypothetical protein Cbei_3326	23887	62
gil150018543	2,5-didehydrogluconate reductase	31986	191
gil150018622	molybdenum cofactor synthesis domain-containing protein	17547	99
gil150018649	aldehyde dehydrogenase	51867	149
gil150018652	acetoacetate decarboxylase	27532	216
gil150018793	extracellular solute-binding protein	62390	223
gil150018806	hypothetical protein Cbei_3994	17665	67
gil150018853	2-oxoglutarate ferredoxin oxidoreductase subunit beta	31494	86
gil150018854	pyruvate flavodoxin/ferredoxin oxidoreductase domain-containing protein	62543	481
gil150018878	hypothetical protein Cbei_4066	24056	95
gil150018959	basic membrane lipoprotein	37308	167
gil150018972	heat shock protein 90	75068	131
gil150019016	glutamate synthase (ferredoxin)	171247	63
gil150019040	hypoxanthine phosphoribosyltransferase	19883	158
gil150019044	nucleoside-diphosphate kinase	15405	156
gil150019054	hypothetical protein Cbei_4242	16701	82
gil150019101	flagellin domain-containing protein	29021	178
gil150019116	putative CheW protein	14865	226
gil150019119	CheA signal transduction histidine kinase	76215	385
gil150019128	aminotransferase, class I and II	43945	233
gil150019130	pyruvate flavodoxin/ferredoxin oxidoreductase domain-containing protein	128281	3447
gil150019139	oligoendopeptidase F	68788	371
gil150019140	hypothetical protein Cbei_4328	15671	377
gil150019149	thiamine pyrophosphate binding domain-containing protein	42357	46
gil150019168	cysteine synthase A	32640	1931
gil150019169	nitroreductase	19702	381
gil150019216	hypothetical protein Cbei_4405	15493	1820
gil150019217	hypothetical protein Cbei_4406	15519	1871
gil150019218	hypothetical protein Cbei_4407	83020	58
gil150019220	saccharopine dehydrogenase	45459	57
gil150019260	ABC transporter related	56239	223

gil150019261	monosaccharide-transporting ATPase	41113	626
gil150019263	carbohydrate kinase, FGGY	58827	684
gil150019264	transketolase	72572	1365
gil150019265	putative transaldolase	23178	1490
gil150019266	L-ribulose-5-phosphate 4-epimerase	25411	371
gil150019268	L-arabinose isomerase	54563	306
gil150019273	periplasmic binding protein/LacI transcriptional regulator	36287	413
gil150019276	aldose 1-epimerase	39131	292
gil150019414	methyl-accepting chemotaxis sensory transducer	33969	92
gil150019445	phosphotransferase system, lactose/cellobiose-specific IIB subunit	11135	85
gil150019454	transketolase	72473	856
gil150019505	choline/ethanolamine kinase	71972	176
gil150019527	cell wall binding repeat-containing protein	69478	1094
gil150019569	cell wall binding repeat-containing protein	63700	2099
gil150019585	dihydrodipicolinate synthase	32461	182
gil150019597	enzyme with TIM-barrel fold	24119	74
gil150019643	catalase	26132	102
gil150019659	pyruvate kinase	50903	551
gil150019660	6-phosphofructokinase	34271	610
gil150019662	stress responsive alpha-beta barrel domain-containing protein	10959	92
gil150019666	UDP-N-acetylenolpyruvoylglucosamine reductase	27428	99
gil150019713	glucose-1-phosphate adenylyltransferase	43189	67
gil150019725	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase	25537	127
gil150019735	D-isomer specific 2-hydroxyacid dehydrogenase, NAD-binding	32469	368
gil150019736	extracellular solute-binding protein	61171	137
gil150019767	response regulator receiver protein	34454	100
gil150019848	pyridoxal biosynthesis lyase PdxS	31410	290
gil150019863	beta-lactamase domain-containing protein	46047	311
gil150019864	acyl-CoA dehydrogenase domain-containing protein	69341	401
gil150019868	hypothetical protein Cbei_5060	19627	264
gil150019882	adenylosuccinate synthetase	47519	75
gil150019894	30S ribosomal protein S6	10989	335
gil150015036	DNA-directed RNA polymerase subunit beta	139257	220
gil150015308	F0F1 ATP synthase subunit beta	50258	123
gil150015375	pyruvate kinase	52122	141
gil150016020	sporulation stage IV, protein A	56351	131
gil150016296	rubrerythrin	20007	56
gil150016423	aspartyl-tRNA synthetase	68424	49
gil150016625	anthranilate phosphoribosyltransferase	36466	94
gil150016726	pyruvate flavodoxin/ferredoxin oxidoreductase domain-containing protein	131577	171

7.5 Proteomic Analysis of *C. beijerinckii* Grown on Xylose

GI #	Gene label	Mass	score
gil150014899	DNA gyrase subunit A	92723	92
gil150014913	ABC transporter related	59727	151
gil150014925	peptidylprolyl isomerase	19103	76
gil150014945	peptidase M24	47840	123
gil150014967	phosphoserine aminotransferase	40121	338
gil150014972	SpoVG family protein	10169	135
gil150014974	ribose-phosphate pyrophosphokinase	34994	97
gil150014983	histone family protein DNA-binding protein	10127	216
gil150014993	formate--tetrahydrofolate ligase	60745	545
gil150014996	transcription elongation factor GreA	17769	43
gil150015028	elongation factor Tu	43653	1288
gil150015032	ribosomal protein L11	14988	518
gil150015033	50S ribosomal protein L1	24623	873
gil150015034	50S ribosomal protein L10	18275	710
gil150015035	ribosomal protein L7/L12	12592	788
gil150015036	DNA-directed RNA polymerase subunit beta	139257	800
gil150015037	DNA-directed RNA polymerase subunit beta'	132490	149
gil150015038	30S ribosomal protein S12	13758	101
gil150015039	30S ribosomal protein S7	17661	1123
gil150015040	elongation factor G	76160	1476
gil150015042	30S ribosomal protein S10	11502	332
gil150015043	50S ribosomal protein L3	22768	689
gil150015044	50S ribosomal protein L4	22863	857
gil150015045	50S ribosomal protein L23	11108	691
gil150015046	50S ribosomal protein L2	30397	494
gil150015047	30S ribosomal protein S19	10278	141
gil150015048	50S ribosomal protein L22	12337	321
gil150015049	30S ribosomal protein S3	24581	578
gil150015053	50S ribosomal protein L14	13212	432
gil150015054	50S ribosomal protein L24	11651	240
gil150015055	50S ribosomal protein L5	20424	980
gil150015057	30S ribosomal protein S8	14756	502
gil150015058	50S ribosomal protein L6	19874	903
gil150015059	50S ribosomal protein L18	13196	331
gil150015060	30S ribosomal protein S5	17312	808
gil150015061	50S ribosomal protein L30	6610	508
gil150015062	50S ribosomal protein L15	15648	482
gil150015064	adenylate kinase	24232	59
gil150015069	30S ribosomal protein S13	13847	564
gil150015070	30S ribosomal protein S11	14080	217
gil150015071	30S ribosomal protein S4	24032	343
gil150015072	DNA-directed RNA polymerase subunit alpha	35215	121
gil150015073	50S ribosomal protein L17	12814	318
gil150015078	50S ribosomal protein L13	16340	633
gil150015079	30S ribosomal protein S9	14476	458
gil150015088	phosphoenolpyruvate-protein phosphotransferase	59644	408
gil150015095	phosphate butyryltransferase	32592	1387
gil150015096	butyrate kinase	38524	541
gil150015136	phosphoenolpyruvate-dependent sugar phosphotransferase system, EIIA 2	16181	128
gil150015138	glucosamine--fructose-6-phosphate aminotransferase, isomerising	66893	245
gil150015152	RND family efflux transporter MFP subunit	52841	290
gil150015207	transaldolase	24704	209
gil150015211	3-hydroxybutyryl-CoA dehydratase	28449	702
gil150015212	acyl-CoA dehydrogenase domain-containing protein	41307	1851
gil150015213	electron transfer flavoprotein, alpha/beta-subunit-like	27918	1106

	protein		
gil150015214	electron transfer flavoprotein, alpha/beta-subunit-like protein	35994	1140
gil150015215	3-hydroxybutyryl-CoA dehydrogenase	30618	2777
gil150015218	co-chaperonin GroES	10149	265
gil150015219	chaperonin GroEL	57575	62
gil150015221	inosine 5'-monophosphate dehydrogenase	52586	574
gil150015222	GMP synthase	35775	174
gil150015231	glucose-6-phosphate isomerase	49829	731
gil150015297	RpiB/LacA/LacB family sugar-phosphate isomerase	18060	124
gil150015301	acetyl-CoA acetyltransferase	41327	4628
gil150015304	F0F1 ATP synthase subunit B	18305	90
gil150015305	F0F1 ATP synthase subunit delta	21314	148
gil150015306	F0F1 ATP synthase subunit alpha	55366	1248
gil150015308	F0F1 ATP synthase subunit beta	50258	861
gil150015318	S-adenosylmethionine synthetase	43543	479
gil150015323	preprotein translocase subunit SecA	96025	176
gil150015334	glutamine synthetase, catalytic region	77489	128
gil150015353	hemerythrin-like metal-binding protein	15797	77
gil150015375	pyruvate kinase	52122	106
gil150015376	aminoacyl-histidine dipeptidase	53996	338
gil150015398	50S ribosomal protein L21	11301	780
gil150015455	putative serine protein kinase, PrkA	73669	47
gil150015466	cysteine synthase	31988	374
gil150015486	glyceraldehyde-3-phosphate dehydrogenase, type I	35775	1349
gil150015487	phosphoglycerate kinase	42136	549
gil150015488	triosephosphate isomerase	27162	978
gil150015489	phosphoglyceromutase	56514	141
gil150015491	phosphopyruvate hydratase	47201	296
gil150015527	YaeC family lipoprotein	29814	84
gil150015546	methyltransferase type 11	21713	273
gil150015551	putative oxidoreductase	50621	109
gil150015568	O-methyltransferase family protein	23751	136
gil150015570	hypothetical protein Cbei_0682	10093	88
gil150015571	radical SAM domain-containing protein	50434	554
gil150015577	hypothetical protein Cbei_0689	11485	177
gil150015641	putative CheW protein	33591	182
gil150015644	phosphotransferase system, lactose/cellobiose-specific IIB subunit	11293	166
gil150015645	Serine--pyruvate transaminase	42965	134
gil150015652	O-acetylhomoserine/O-acetylserine sulfhydrylase	46363	488
gil150015682	UspA domain-containing protein	16476	186
gil150015690	methyl-accepting chemotaxis sensory transducer	62657	77
gil150015712	hypothetical protein Cbei_0826	16769	281
gil150015716	molecular chaperone DnaK	65416	769
gil150015868	ribosomal 5S rRNA E-loop binding protein Ctc/L25/TL5	20762	220
gil150015869	phosphoglucomutase/phosphomannomutase alpha/beta/subunit	64725	350
gil150015892	orotate phosphoribosyltransferase	25225	105
gil150015895	formate acetyltransferase	83631	952
gil150015897	formate acetyltransferase	83456	3018
gil150015914	AsnC family transcriptional regulator	17964	105
gil150015921	hypothetical protein Cbei_1035	11843	189
gil150015929	branched-chain amino acid aminotransferase	38271	277
gil150015932	band 7 protein	34759	58
gil150015939	phosphoribosylformylglycinamidine synthase	137920	643
gil150015940	phosphoribosylaminoimidazole carboxylase, catalytic subunit	17102	210
gil150015941	phosphoribosylaminoimidazole-succinocarboxamide synthase	26763	486

gil150015945	bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase	56438	536
gil150015994	GTP-binding protein TypA	68060	69
gil150016002	cell division protein FtsZ	40622	417
gil150016040	hypothetical protein Cbei_1156	12883	265
gil150016049	acetate kinase	43774	286
gil150016054	ribonuclease III	26714	80
gil150016064	50S ribosomal protein L19	13359	304
gil150016072	30S ribosomal protein S2	26356	104
gil150016073	elongation factor Ts	34123	340
gil150016082	hypothetical protein Cbei_1198	14563	119
gil150016083	transcription elongation factor NusA	44957	96
gil150016086	translation initiation factor IF-2	76480	351
gil150016087	ribosome-binding factor A	13620	95
gil150016091	30S ribosomal protein S15	10189	294
gil150016092	polynucleotide phosphorylase/polyadenylase	77276	200
gil150016103	phosphotransferase system, phosphocarrier protein HPr	9077	545
gil150016105	adenylosuccinate lyase	54084	491
gil150016121	S-ribosylhomocysteinase	18301	209
gil150016175	threonine synthase	54816	121
gil150016208	trigger factor	48472	735
gil150016209	ATP-dependent Clp protease, proteolytic subunit ClpP	22056	201
gil150016219	hypothetical protein Cbei_1337	12591	515
gil150016220	nicotinate phosphoribosyltransferase	56957	86
gil150016260	hypothetical protein Cbei_1378	21832	182
gil150016296	rubrerythrin	20007	328
gil150016307	glutamyl-tRNA synthetase	63942	195
gil150016309	glutaminyl-tRNA synthetase	64207	67
gil150016311	seryl-tRNA synthetase	48555	266
gil150016338	pyruvate flavodoxin/ferredoxin oxidoreductase domain-containing protein	130496	167
gil150016373	hypothetical protein Cbei_1494	24923	848
gil150016374	hypothetical protein Cbei_1495	35782	791
gil150016423	aspartyl-tRNA synthetase	68424	52
gil150016445	translation initiation factor IF-3	21563	84
gil150016446	50S ribosomal protein L35	7506	51
gil150016447	50S ribosomal protein L20	13621	289
gil150016564	elongation factor P	22075	205
gil150016586	response regulator receiver protein	30721	312
gil150016596	iron-containing alcohol dehydrogenase	43043	472
gil150016623	anthranilate synthase component I	53456	215
gil150016624	glutamine amidotransferase of anthranilate synthase	21769	198
gil150016625	anthranilate phosphoribosyltransferase	36466	117
gil150016626	indole-3-glycerol-phosphate synthase	31355	144
gil150016628	tryptophan synthase subunit beta	43014	731
gil150016629	tryptophan synthase, alpha subunit	28702	204
gil150016648	hypothetical protein Cbei_1774	24133	174
gil150016655	valyl-tRNA synthetase	101354	55
gil150016661	hypothetical protein Cbei_1788	20099	76
gil150016664	tetrahydrodipicolinate succinyltransferase domain-containing protein	25184	190
gil150016726	pyruvate flavodoxin/ferredoxin oxidoreductase domain-containing protein	131577	171
gil150016741	serine hydroxymethyltransferase	45243	428
gil150016776	fructose-1,6-bisphosphate aldolase, class II	30931	716
gil150016830	hypothetical protein Cbei_1958	19522	293
gil150016905	acyl-CoA dehydrogenase domain-containing protein	41391	1473
gil150016995	aspartyl-tRNA synthetase	50162	121
gil150017050	iron-containing alcohol dehydrogenase	42947	122
gil150017051	ferredoxin-NADP(+) reductase subunit alpha	32747	105

gil150017147	hydroxylamine reductase	60892	625
gil150017213	hypothetical protein Cbei_2351	10018	336
gil150017242	D-xylose ABC transporter, periplasmic substrate-binding protein	38942	165
gil150017245	xylose isomerase	50455	177
gil150017246	xylulokinase	55260	889
gil150017248	putative transaldolase	23164	1651
gil150017249	transketolase	72544	1044
gil150017250	hypothetical protein Cbei_2388	23555	80
gil150017251	L-fucose isomerase related protein	54503	904
gil150017287	xanthine phosphoribosyltransferase	21248	218
gil150017349	rubrerythrin	20009	304
gil150017371	extracellular solute-binding protein	31572	337
gil150017403	4-hydroxy-3-methylbut-2-enyl diphosphate reductase/S1 RNA-binding domain protein	71710	200
gil150017407	Ferritin, Dps family protein	20036	295
gil150017466	aspartate alpha-decarboxylase	13792	349
gil150017467	pantoate--beta-alanine ligase	31367	661
gil150017468	3-methyl-2-oxobutanoate hydroxymethyltransferase	29773	1525
gil150017606	glycerol dehydrogenase	39532	77
gil150017652	M protein-like MukB domain-containing protein	126228	37
gil150017661	hypothetical protein Cbei_2809	19517	173
gil150017675	NERD domain-containing protein	50337	61
gil150017732	FAD linked oxidase domain-containing protein	51693	79
gil150017855	hydroxylamine reductase	61608	253
gil150017874	hypothetical protein Cbei_3033	26571	109
gil150017884	hypothetical protein Cbei_3043	22683	114
gil150018108	class II aldolase/adducin family protein	24899	118
gil150018109	coenzyme A transferase	56285	533
gil150018161	aliphatic sulfonate ABC transporter periplasmic ligand-binding protein	36930	223
gil150018387	flavodoxin	16717	68
gil150018543	2,5-didehydrogluconate reductase	31986	53
gil150018652	acetoacetate decarboxylase	27532	116
gil150018959	basic membrane lipoprotein	37308	216
gil150018972	heat shock protein 90	75068	212
gil150018998	sulphate adenylyltransferase, large subunit	63575	103
gil150019101	flagellin domain-containing protein	29021	148
gil150019114	flagellar motor switch protein	43264	245
gil150019116	putative CheW protein	14865	137
gil150019119	CheA signal transduction histidine kinase	76215	286
gil150019123	putative CheW protein	17077	114
gil150019128	aminotransferase, class I and II	43945	199
gil150019130	pyruvate flavodoxin/ferredoxin oxidoreductase domain-containing protein	128281	3650
gil150019139	oligoendopeptidase F	68788	150
gil150019140	hypothetical protein Cbei_4328	15671	369
gil150019149	thiamine pyrophosphate binding domain-containing protein	42357	45
gil150019168	cysteine synthase A	32640	1749
gil150019169	nitroreductase	19702	257
gil150019216	hypothetical protein Cbei_4405	15493	1545
gil150019217	hypothetical protein Cbei_4406	15519	1492
gil150019218	hypothetical protein Cbei_4407	83020	95
gil150019264	transketolase	72572	760
gil150019273	periplasmic binding protein/LacI transcriptional regulator	36287	77
gil150019278	Cof-like hydrolase	32693	86
gil150019407	HesB-related (seleno)protein	12035	124
gil150019454	transketolase	72473	678
gil150019491	L-fucose isomerase related protein	54604	887
gil150019499	hypothetical protein Cbei_4689	9915	108

gil150019500	membrane spanning protein	38915	34
gil150019505	choline/ethanolamine kinase	71972	253
gil150019527	cell wall binding repeat-containing protein	69478	261
gil150019569	cell wall binding repeat-containing protein	63700	1365
gil150019583	putative aminopeptidase 2	49843	46
gil150019585	dihydrodipicolinate synthase	32461	174
gil150019621	hypothetical protein Cbei_4812	13089	73
gil150019659	pyruvate kinase	50903	913
gil150019660	6-phosphofructokinase	34271	518
gil150019725	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase	25537	136
gil150019734	hypothetical protein Cbei_4926	48142	75
gil150019735	D-isomer specific 2-hydroxyacid dehydrogenase, NAD-binding	32469	413
gil150019736	extracellular solute-binding protein	61171	87
gil150019758	hypothetical protein Cbei_4950	13298	82
gil150019767	response regulator receiver protein	34454	143
gil150019848	pyridoxal biosynthesis lyase PdxS	31410	104
gil150019864	acyl-CoA dehydrogenase domain-containing protein	69341	115
gil150019868	hypothetical protein Cbei_5060	19627	250
gil150019882	adenylosuccinate synthetase	47519	398
gil150019889	50S ribosomal protein L9	16731	102
gil150019892	30S ribosomal protein S18	10148	87
gil150019894	30S ribosomal protein S6	10989	194

References

References

Aduse-Opoku, J. & Mitchell, W.J. (1988) Diauxic growth of Clostridium thermosaccharolyticum on glucose and xylose. *FEMS Microbiology Letters.* **50** (1) 45-49.

Ali, M.K., Rudolph, F.B. & Bennett, G.N. (2004) Thermostable xylanase10B from *Clostridium acetobutylicum* ATCC824. *Journal of Industrial Microbiology and Biotechnology.* **31** (5) 229-234.

Ali, M.K., Rudolph, F.B. & Bennett, G.N. (2005) Characterization of thermostable Xyn10A enzyme from mesophilic *Clostridium acetobutylicum* ATCC 824. *Journal of Industrial Microbiology and Biotechnology.* **32** (1) 12-18.

Amore, R. & Hollenberg, C.P. (1989) Xylose isomerase from Actinoplanes missouriensis: primary structure of the gene and the protein. *Nucleic Acids Research.* **17** (18) 7515.

Andrade, J.C. & Vasconcelos, I. (2003) Continuous cultures of *Clostridium acetobutylicum*: culture stability and low-grade glycerol utilisation. *Biotechnology Letters.* **25** (2) 121-125.

Andreesen, J.R., Schaupp, A., Neurauter, C., Brown, A. & Ljungdahl, L.G. (1973) Fermentation of glucose, fructose, and xylose by Clostridium thermoaceticum: effect of metals on growth yield, enzymes, and the synthesis of acetate from CO 2. *Journal of Bacteriology.* **114** (2) 743-751.

Attwood, T.K. & Parry-Smith, D.J. (1999) *Introduction to bioinformatics,* Harlow, Prentice-Hall.

Audic, J.-L., Chaufer, B. & Daufin, G. (2003) Non-food applications of milk components and dairy co-products: A review. *Lait.* **83** 417-438.

Bahl, H., Muller, H., Behrens, S., Joseph, H. & Narberhaus, F. (1995) Expression of heat shock genes in Clostridium acetobutylicum. *FEMS Microbiology Reviews.* **17** (3) 341-348.

Batycka, M., Inglis, N.F., Cook, K., Adam, A., Fraser-Pitt, D., Smith, D.G., Main, L., Lubben, A. & Kessler, B.M. (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., Mitchell, W. & Bahl, H. (2001) Molecular analysis of the mannitol operon of *Clostridium acetobutylicum* encoding a phosphotransferase system and a putative PTS-modulated regulator. *Microbiology.* **147** (Pt 1) 75-86.

Bettiga, M., Hahn-Hagerdal, B. & Gorwa-Grauslund, M.F. (2008) Comparing the xylose reductase/xylitol dehydrogenase and xylose isomerase pathways in arabinose and xylose fermenting Saccharomyces cerevisiae strains. *Biotechnology Biofuels.* **1** (1) 16.

References

Bird, E.J. (2011) Characterisation of Carbohydrate Transport in the Solventogenic organism Clostridium acetobutylicum ATCC 824. Edinburgh Napier University.

Bor, Y.C., Moraes, C., Lee, S.P., Crosby, W.L., Sinskey, A.J. & Batt, C.A. (1992) Cloning and sequencing the Lactobacillus brevis gene encoding xylose isomerase. *Gene.* **114** (1) 127-132.

Bowles, L.K. & Ellefson, W.L. (1985) Effects of butanol on *Clostridium acetobutylicum*. *Applied and Environmental Microbiology.* **50** (5) 1165-1170.

Bustin, S.A. (2002) Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *Journal of Molecular Endocrinology.* **29** (1) 23-39.

Chaillou, S., Bor, Y.C., Batt, C.A., Postma, P.W. & Pouwels, P.H. (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** (12) 4720-4728.

Chen, C.K. & Blaschek, H.P. (1999) Effect of acetate on molecular and physiological aspects of *Clostridium beijerinckii* NCIMB 8052 solvent production and strain degeneration. *Applied and Environmental Microbiology.* **65** (2) 499-505.

Chouayekh, H., Bejar, W., Rhimi, M., Jelleli, K., Mseddi, M. & Bejar, S. (2007) Characterization of an l-arabinose isomerase from the *Lactobacillus plantarum* NC8 strain showing pronounced stability at acidic pH. *FEMS Microbiology Letters.* **277** (2) 260-267.

Claassen, P.A., Budde, M.A. & 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.

Clarke, P., Lin, H.C. & Wilcox, G. (1982) The nucleotide sequence of the araC regulatory gene in Salmonella typhimurium LT2. *Gene.* **18** (2) 157-163.

Collins, T., Gerday, C. & Feller, G. (2005) Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiology Reviews.* **29** (1) 3-23.

Copeland, A., Lucas, S., Lapidus, A., Barry, K., Detter, J.C., Glavina del Rio, T., Hammon, N., Israni, S., Dalin, E., Tice, H., Pitluck, S., Chertkov, O., Brettin, T., Bruce, D., Han, C., Tapia, R., Gilna, P., Schmutz, J., Larimer, F., Land, M., Hauser, L., Kyrpides, N., Mikhailova, N., Wu, J.H.D., Newcomb, M. & Richardson, P. (2007a) Complete sequence of Clostridium thermocellum ATCC 27405. NCBI. Available at: http://www.ncbi.nlm.nih.gov/nuccore/NC_009012. Last accessed: 18/01/11.

Copeland, A., Lucas, S., Lapidus, A., Barry, K., Detter, J.C., Glavinadel Rio, T., Hammon, N., Israni, S., Dalin, E., Tice, H., Pitluck, S., Sims, D., Brettin, T., Bruce, D., Tapia, R., Brainard, J., Schmutz, J., Larimer, F., Land, M., Hauser, L., Kyrpides, N., Mikhailova, N., Bennet, G., Cann, I., Chen, J.-S., Contreras, A.L., Jones, D., Kashket, E., Mitchell, W., Stoddard, S., Schwarz, W., Qureshi, N., Young, M., Shi, Z., Ezeji, T., White, B., Blaschek, H.P. & Richardson, P. (2007b) Clostridium beijerinckii NCIMB 8052, complete genome. *NCBI.* Available at: http://www.ncbi.nlm.nih.gov/nuccore/NC_009617. Last accessed: 09/09/08.

Cornillot, E., Nair, R.V., Papoutsakis, E.T. & Soucaille, P. (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** (17) 5442-5447.

Cutillas, P.R. & Timms, J.F. (2010) Approaches and applications of quantitative LC-MS for proteomics and activitomics. *Methods in Molecular Biology.* **658** 3-17.

Czjzek, M., Bravman, T., Henrissat, B. & Shoham, Y. (2004) Crystallization and preliminary X-ray analysis of family 39 beta-D-xylosidase from Geobacillus stearothermophilus T-6. *Acta Crystallographica Section D Biological Crystallography.* **60** (Pt 3) 583-585.

de Vos, W.M., Kleerebezem, M. & Kuipers, O.P. (1997) Expression systems for industrial Gram-positive bacteria with low guanine and cytosine content. *Current Opinion in Biotechnology.* **8** (5) 547-553.

Dekker, K., Yamagata, H., Sakaguchi, K. & Udaka, S. (1991) Xylose (glucose) isomerase gene from the thermophile *Thermus thermophilus*: cloning, sequencing, and comparison with other thermostable xylose isomerases. *Journal of Bacteriology.* **173** (10) 3078-3083.

Demain, A.L., Newcomb, M. & Wu, J.H. (2005) Cellulase, clostridia, and ethanol. *Microbiology and Molecular Biology Reviews.* **69** (1) 124-154.

Demirbas, A. (2007) Progress and recent trends in biofuels. *Progress in Energy and Combustion Science.* **33** 1-18.

Dills, S.S., Apperson, A., Schmidt, M.R. & Saier, M.H., Jr. (1980) Carbohydrate transport in bacteria. *Microbiology Reviews.* **44** (3) 385-418.

Drocourt, D., Bejar, S., Calmels, T., Reynes, J.P. & Tiraby, G. (1988) Nucleotide sequence of the xylose isomerase gene from Streptomyces violaceoniger. *Nucleic Acids Research.* **16** (19) 9337.

Durre, P. (1998) New insights and novel developments in clostridial acetone/butanol/isopropanol fermentation. *Applied Microbiology and Biotechnology.* **49** 639-648.

Durre, P. (2007) Biobutanol: an attractive biofuel. *Biotechnology Journal.* **2** (12) 1525-1534.

Durre, P., Fischer, R.J., Kuhn, A., Lorenz, K., Schreiber, W., Sturzenhofecker, B., Ullmann, S., Winzer, K. & Sauer, U. (1995) Solventogenic enzymes of *Clostridium acetobutylicum*: catalytic properties, genetic organization, and transcriptional regulation. *FEMS Microbiology Reviews.* **17** (3) 251-262.

Durre, P. & Hollergschwandner, C. (2004) Initiation of endospore formation in *Clostridium acetobutylicum*. *Anaerobe.* **10** (2) 69-74.

El Kanouni, A., Zerdani, I., Zaafa, S., Znassni, M., Boudouma, L. & Boudouma, M. (1998) The improvement of glucose/xylose fermentation by *Clostridium acetobutylicum* using calcium carbonate. *World Journal of Microbiology and Biotechnology.* **14** 431-435.

Elferink, M.G., Albers, S.V., Konings, W.N. & Driessen, A.J. (2001) Sugar transport in *Sulfolobus solfataricus* is mediated by two families of binding protein-dependent ABC transporters. *Molecular Microbiology.* **39** (6) 1494-1503.

Englesberg, E. (1961) Enzymatic characterization of 17 L-arabinose negative mutants of *Escherichia coli*. *Journal of Bacteriology.* **81** 996-1006.

Erbeznik, M., Dawson, K.A. & Strobel, H.J. (1998) Cloning and characterization of transcription of the xylAB operon in *Thermoanaerobacter ethanolicus*. *Journal of Bacteriology.* **180** (5) 1103-1109.

Erbeznik, M., Hudson, S.E., Herrman, A.B. & Strobel, H.J. (2004) Molecular analysis of the xylFGH operon, coding for xylose ABC transport, in Thermoanaerobacter ethanolicus. *Current Opinion in Microbiology.* **48** (4) 295-299.

Erlandson, K.A., Park, J.H., Wissam, El, K., Kao, H.H., Basaran, P., Brydges, S. & Batt, C.A. (2000) Dissolution of xylose metabolism in *Lactococcus lactis*. *Applied and Environmental Microbiology.* **66** (9) 3974-3980.

European Parliament and the council of the European Union (2003) Directive 2003/30/EC of the European parliament and of the council of 8 May 2003 on the promotion of the use of biofuels or other renewable fuels for transport. *Official Journal of the European Union.* **L123** 42-46.

Ezeji, T. & Blaschek, H.P. (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.

Ezeji, T., Qureshi, N. & Blaschek, H.P. (2007) Butanol production from agricultural residues: Impact of degradation products on Clostridium beijerinckii growth and butanol fermentation. *Biotechnology Bioengineering.* **97** (6) 1460-1469.

Fasman, K.H. & Salzberg, S.L. (1998) An introduction to biological sequence analysis. IN Salzberg, S.L., Searls, D.B. & Kasif, S. (Eds.) *New Comprehensive Biochemistry: Computational methods in molecular biology.* Amsterdam, Elsevier.

Fath, M.J. & Kolter, R. (1993) ABC transporters: bacterial exporters. *Microbiological Reviews.* **57** (4) 995-1017.

Feldmann, S.D., Sahm, H. & Sprenger, G.A. (1992) Cloning and expression of the genes for xylose isomerase and xylulokinase from Klebsiella pneumoniae 1033 in Escherichia coli K12. *Molecular Genetics and Genomics.* **234** (2) 201-210.

Fermentas (2010) X-Gal. Product Information. Available at: http://www.fermentas.de/product_info.php?info=p871. Last accessed: 14/07/2011.

Fink, A.L. (1999) Chaperone-mediated protein folding. *Physiological Reviews.* **79** (2) 425-449.

Fond, O., Engasser, J.-M., Matta-El-Amouri, G. & Petitdemange, H. (1985a) The acetone butanol fermentation on glucose and xylose. I. Regulation and kinetics in batch cultures. *Biotechnology and Bioengineering.* **28** (2) 160-166.

Fond, O., Engasser, J.-M., Matta-El-Amouri, G. & Petitdemange, H. (1985b) The acetone butanol fermentation on glucose and xylose. II. Regulation and kinetics in fed-batch cultures. *Biotechnology and Bioengineering.* **28** (2) 167-175.

Franco, I.S., Mota, L.J., Soares, C.M. & de Sa-Nogueira, I. (2006) Functional domains of the *Bacillus subtilis* transcription factor AraR and identification of amino acids important for nucleoprotein complex assembly and effector binding. *Journal of Bacteriology.* **188** (8) 3024-3036.

Frondel, M. & Peters, J. (2007) Biodiesel: A new oildorado? *Energy Policy.* **35** 1675-1684.

Gabriel, C.L. (1928) Butanol fermentation process. *Industrial and Engineering Chemistry.* **20** (10) 1063-1067.

Galbe, M. & Zacchi, G. (2002) A review of the production of ethanol from softwood. *Applied Microbiology and Biotechnology.* **59** (6) 618-628.

Gallegos, M.T., Schleif, R., Bairoch, A., Hofmann, K. & Ramos, J.L. (1997) Arac/XylS family of transcriptional regulators. *Microbiology and Molecular Biology Reviews.* **61** (4) 393-410.

Gallert, C. & Winter, J. (2002) Solid and liquid residues as raw materials for biotechnology. *Naturwissenschaften.* **89** (11) 483-496.

Gasparic, A., Martin, J., Daniel, A.S. & Flint, H.J. (1995) A xylan hydrolase gene cluster in Prevotella ruminicola B(1)4: sequence relationships, synergistic interactions, and oxygen sensitivity of a novel enzyme with exoxylanase and beta-(1,4)-xylosidase activities. *Applied and Environmental Microbiology.* **61** (8) 2958-2964.

Geib, S.M., Tien, M. & Hoover, K. (2010) Identification of proteins involved in lignocellulose degradation using in gel zymogram analysis combined with mass

spectroscopy-based peptide analysis of gut proteins from larval Asian longhorned beetles, *Anoplophora glabripennis*. *Insect Science.* **17** 253-264.

Godiska, R., Mead, D., Dhodda, V., Wu, C., Hochstein, R., Karsi, A., Usdin, K., Entezam, A. & Ravin, N. (2009) Linear plasmid vector for cloning of repetitive or unstable sequences in Escherichia coli. *Nucleic Acids Research.* **38** (6) 1-9.

Goldman, N.R. (2009) Methods for optimizing enzymatic hydrolysis of xylan to improve xylooligosaccharide yield. *MMG 445: Basic biochemistry.* **5** (1) 31-36.

Graves, M.C. & Rabinowitz, J.C. (1986) In vivo and in vitro transcription of the Clostridium pasteurianum ferredoxin gene. Evidence for "extended" promoter elements in gram-positive organisms. *Journal of Biological Chemistry.* **261** (24) 11409-11415.

Gray, K.A., Zhao, L. & Emptage, M. (2006) Bioethanol. *Current Opinon in Chemical Biology.* **10** (2) 141-146.

Greller, G., Horlacher, R., DiRuggiero, J. & Boos, W. (1999) Molecular and biochemical analysis of MalK, the ATP-hydrolyzing subunit of the trehalose/maltose transport system of the hyperthermophilic archaeon *Thermococcus litoralis*. *Journal of Biological Chemistry.* **274** (29) 20259-20264.

Greller, G., Riek, R. & Boos, W. (2001) Purification and characterization of the heterologously expressed trehalose/maltose ABC transporter complex of the hyperthermophilic archaeon *Thermococcus litoralis*. *European Journal of Biochemistry.* **268** (14) 4011-4018.

Groarke, J.M., Mahoney, W.C., Hope, J.N., Furlong, C.E., Robb, F.T., Zalkin, H. & Hermodson, M.A. (1983) The amino acid sequence of D-ribose-binding protein from Escherichia coli K12. *Journal of Biological Chemistry.* **258** (21) 12952-12956.

Gu, Y., Ding, Y., Ren, C., Sun, Z., Rodionov, D.A., Zhang, W., Yang, S., Yang, C. & Jiang, W. (2010) Reconstruction of xylose utilization pathway and regulons in Firmicutes. *BMC Genomics.* **11** 255.

Gu, Y., Li, J., Zhang, L., Chen, J., Niu, L., Yang, Y., Yang, S. & Jiang, W. (2009) Improvement of xylose utilization in Clostridium acetobutylicum via expression of the talA gene encoding transaldolase from Escherichia coli. *Journal of Biotechnology.* **143** (4) 284-287.

Hahn-Hagerdal, B., Galbe, M., Gorwa-Grauslund, M.F., Liden, G. & Zacchi, G. (2006) Bio-ethanol--the fuel of tomorrow from the residues of today. *Trends in Biotechnology.* **24** (12) 549-556.

Haldrup, A., Petersen, S.G. & Okkels, F.T. (1998) The xylose isomerase gene from *Thermoanaerobacterium thermosulfurogenes* allows effective selection of transgenic plant cells using D-xylose as the selection agent. *Plant Molecular Biology.* **37** (2) 287-296.

Hansen, A.C., Zhang, Q. & Lyne, P.W. (2005) Ethanol-diesel fuel blends -- a review. *Bioresource Technology.* **96** (3) 277-285.

Harris, L.M., Welker, N.E. & Papoutsakis, E.T. (2002) Northern, morphological, and fermentation analysis of spo0A inactivation and overexpression in Clostridium acetobutylicum ATCC 824. *Journal of Bacteriology.* **184** (13) 3586-3597.

Henderson, P.J. (1990) Proton-linked sugar transport systems in bacteria. *Journal of Bioenergetics and Biomembranes.* **22** (4) 525-569.

Heo, G.Y., Kim, W.C., Joo, G.J., Kwak, Y.Y., Shin, J.H., Roh, D.H., Park, H.D. & Rhee, I.K. (2008) Deletion of xylR gene enhances expression of xylose isomerase in Streptomyces lividans TK24. *Journal of Microbiology and Biotechnology.* **18** (5) 837-844.

Higdon, R. & Kolker, E. (2007) A predictive model for identifying proteins by a single peptide match. *Bioinformatics.* **23** (3) 277-280.

Horazdovsky, B.F. & Hogg, R.W. (1989) Genetic reconstitution of the high-affinity L-arabinose transport system. *Journal of Bacteriology.* **171** (6) 3053-3059.

Howard, R.L., Abotsi, E., Jansen van Rensburg, E.L. & Howard, S. (2003) Lignocellulose biotechnology: issues of bioconversion and enzyme production. *African Journal of Biotechnology.* **2** (12) 602-619.

Huang, L., Forsberg, C.W. & Gibbins, L.N. (1986) Influence of External pH and Fermentation Products on *Clostridium acetobutylicum* Intracellular pH and Cellular Distribution of Fermentation Products. *Applied and Environmental Microbiology.* **51** (6) 1230-1234.

Huang, L., Gibbins, L.N. & Forsberg, C.W. (1985) Transmembrane pH gradient and membrane potential in *Clostridium acetobutylicum* during growth under acetogenic and solventogenic conditions. *Applied and Environmental Microbiology.* **50** (4) 1043-1047.

Hueck, C.J. & Hillen, W. (1995) Catabolite repression in Bacillus subtilis: a global regulatory mechanism for the gram-positive bacteria? *Molecular Microbiology.* **15** (3) 395-401.

Hueck, C.J., Kraus, A., Schmiedel, D. & Hillen, W. (1995) Cloning, expression and functional analyses of the catabolite control protein CcpA from Bacillus megaterium. *Molecular Microbiology.* **16** (5) 855-864.

Huelsenbeck, J.P. & Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics.* **17** (8) 754-755.

Iida, A., Teshiba, S. & Mizobuchi, K. (1993) Identification and characterization of the tktB gene encoding a second transketolase in Escherichia coli K-12. *Journal of Bacteriology.* **175** (17) 5375-5383.

References

Inada, T., Kimata, K. & Aiba, H. (1996) Mechanism responsible for glucose-lactose diauxie in Escherichia coli: challenge to the cAMP model. *Genes Cells.* **1** (3) 293-301.

Jalal, A., Rashid, N., Rasool, N. & Akhtar, M. (2009) Gene cloning and characterization of a xylanase from a newly isolated Bacillus subtilis strain R5. *Journal of Bioscience and Bioengineering.* **107** (4) 360-365.

Jones, D.T., van der Westhuizen, A., Long, S., Allcock, E.R., Reid, S.J. & Woods, D.R. (1982) Solvent Production and Morphological Changes in *Clostridium acetobutylicum*. *Applied and Environmental Microbiology.* **43** (6) 1434-1439.

Jones, D.T. & Woods, D.R. (1986) Acetone-butanol fermentation revisited. *Microbiological Reviews.* **50** (4) 484-524.

Joo, G.J., Shin, J.H., Heo, G.Y., Kim, Y.M. & Rhee, I.K. (2005) Molecular cloning and expression of a thermostable xylose (glucose) isomerase gene, xylA, from Streptomyces chibaensis J-59. *The Journal of Microbiology.* **43** (1) 34-37.

Jorgensen, F., Hansen, O.C. & Stougaard, P. (2004) Enzymatic conversion of D-galactose to D-tagatose: heterologous expression and characterisation of a thermostable L-arabinose isomerase from *Thermoanaerobacter mathranii*. *Applied Microbiology and Biotechnology.* **64** (6) 816-822.

Joshua, C.J., Dahl, R., Benke, P.I. & Keasling, J.D. (2011) Absence of diauxie during simultaneous utilization of glucose and Xylose by Sulfolobus acidocaldarius. *Journal of Bacteriology.* **193** (6) 1293-1301.

Kaneko, T., Saito, K., Kawamura, Y. & Takahashi, S. (2001) Molecular cloning of acid-stable glucose isomerase gene from Streptomyces olivaceoviridis E-86 by a simple two-step PCR method, and its expression in Escherichia coli. *Bioscience Biotechnology and Biochemistry.* **65** (5) 1054-1062.

Kawaguchi, H., Sasaki, M., Vertes, A.A., Inui, M. & Yukawa, H. (2009) Identification and functional analysis of the gene cluster for L-arabinose utilization in Corynebacterium glutamicum. *Applied and Environmental Microbiology.* **75** (11) 3419-3429.

Kikuchi, T., Itoh, Y., Kasumi, T. & Fukazawa, C. (1990) Molecular cloning of the xylA gene encoding xylose isomerase from Streptomyces griseofuscus S-41: primary structure of the gene and its product. *Agricultural and Biological Chemistry.* **54** 2469-2472.

Kim, B.C., Lee, Y.H., Lee, H.S., Lee, D.W., Choe, E.A. & Pyun, Y.R. (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.

Kim, H.J. & Oh, D.K. (2005) Purification and characterization of an L-arabinose isomerase from an isolated strain of *Geobacillus thermodenitrificans* producing D-tagatose. *Journal of Biotechnology.* **120** (2) 162-173.

Kim, J.H., Yang, Y.K. & Chambliss, G.H. (2005) Evidence that Bacillus catabolite control protein CcpA interacts with RNA polymerase to inhibit transcription. *Molecular Microbiology.* **56** (1) 155-162.

Kim, J.W., Kim, Y.W., Roh, H.J., Kim, H.Y., Cha, J.H., Park, K.H. & Park, C.S. (2003) Production of tagatose by a recombinant thermostable L-arabinose isomerase from *Thermus sp*. IM6501. *Biotechnology Letters.* **25** (12) 963-967.

Kim, S. & Dale, B.E. (2006) Life cycle assessment of various cropping systems utilized for producing biofuels: bioethanol and biodiesel. *Biomass and Bioenergy.* **29** 426-439.

Ko, E.P., Akatsuka, H., Moriyama, H., Shinmyo, A., Hata, Y., Katsube, Y., Urabe, I. & Okada, H. (1992) Site-directed mutagenesis at aspartate and glutamate residues of xylanase from Bacillus pumilus. *Biochemical Journal.* **288 (Pt 1)** 117-121.

Kobayashi, G., Eto, K., Tashiro, Y., Okubo, K., Sonomoto, K. & Ishizaki, A. (2005) Utilization of excess sludge by acetone-butanol-ethanol fermentation employing *Clostridium saccharoperbutylacetonicum* N1-4 (ATCC 13564). *Journal of Bioscience and Bioengineering.* **99** (5) 517-519.

Konigsberg, W. & Godson, G.N. (1983) Evidence for use of rare codons in the dnaG gene and other regulatory genes of Escherichia coli. *Proceedings of the National Academy of Science U S A.* **80** (3) 687-691.

Kraus, A., Hueck, C., Gartner, D. & Hillen, W. (1994) Catabolite repression of the Bacillus subtilis xyl operon involves a cis element functional in the context of an unrelated sequence, and glucose exerts additional xylR-dependent repression. *Journal of Bacteriology.* **176** (6) 1738-1745.

Kreuzer, P., Gartner, D., Allmansberger, R. & Hillen, W. (1989) Identification and sequence analysis of the *Bacillus subtilis* W23 xylR gene and xyl operator. *Journal of Bacteriology.* **171** (7) 3840-3845.

Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology.* **305** (3) 567-580.

Lawlis, V.B., Dennis, M.S., Chen, E.Y., Smith, D.H. & Henner, D.J. (1984) Cloning and sequencing of the xylose isomerase and xylulose kinase genes of *Escherichia coli*. *Applied and Environmental Microbiology.* **47** (1) 15-21.

Lee, C., Shallenberger, R. & Bittum, M. (1970) Free sugars in fruits and vegetables. *New York Food and Life Science Bulletin.* **1** 1-12.

Lee, D.W., Jang, H.J., Choe, E.A., Kim, B.C., Lee, S.J., Kim, S.B., Hong, Y.H. & Pyun, Y.R. (2004) Characterization of a thermostable L-arabinose (D-galactose) isomerase from the hyperthermophilic eubacterium *Thermotoga maritima*. *Applied and Environmental Microbiology.* **70** (3) 1397-1404.

Lee, N., Gielow, W., Martin, R., Hamilton, E. & Fowler, A. (1986) The organization of the araBAD operon of Escherichia coli. *Gene.* **47** (2-3) 231-244.

Lee, S.F., Forsberg, C.W. & Gibbins, L.N. (1985a) Cellulolytic Activity of *Clostridium acetobutylicum*. *Applied and Environmental Microbiology.* **50** (2) 220-228.

Lee, S.F., Forsberg, C.W. & Gibbins, L.N. (1985b) Xylanolytic Activity of *Clostridium acetobutylicum*. *Applied and Environmental Microbiology.* **50** (4) 1068-1076.

Lee, S.J., Lee, D.W., Choe, E.A., Hong, Y.H., Kim, S.B., Kim, B.C. & Pyun, Y.R. (2005) Characterization of a thermoacidophilic L-arabinose isomerase from *Alicyclobacillus acidocaldarius*: role of Lys-269 in pH optimum. *Applied and Environmental Microbiology.* **71** (12) 7888-7896.

Lee, Y.E., Ramesh, M.V. & Zeikus, J.G. (1993) Cloning, sequencing and biochemical characterization of xylose isomerase from Thermoanaerobacterium saccharolyticum strain B6A-RI. *Journal of General Microbiology.* **139 Pt 6** 1227-1234.

Lenz, T.G. & Moreira, A.R. (1980) Economic evaluation of the acetone-butanol fermentation. *Industrial and Engineering Chemistry Product Research and Development.* **19** 478-483.

Lesley, S.A., Graziano, J., Cho, C.Y., Knuth, M.W. & Klock, H.E. (2002) Gene expression response to misfolded protein as a screen for soluble recombinant protein. *Protein Engineering.* **15** (2) 153-160.

Liao, W.X., Earnest, L., Kok, S.L. & Jeyaseelan, K. (1995) Molecular cloning and characterization of the xylose isomerase gene from a thermophilic Bacillus species. *Biochemistry and Molecular Biology International.* **36** (2) 401-410.

Lin, H.C., Lei, S.P., Studnicka, G. & Wilcox, G. (1985a) The araBAD operon of Salmonella typhimurium LT2. III. Nucleotide sequence of araD and its flanking regions, and primary structure of its product, L-ribulose-5-phosphate 4-epimerase. *Gene.* **34** (1) 129-134.

Lin, H.C., Lei, S.P. & Wilcox, G. (1985b) The araBAD operon of Salmonella typhimurium LT2. I. Nucleotide sequence of araB and primary structure of its product, ribulokinase. *Gene.* **34** (1) 111-122.

Lin, H.C., Lei, S.P. & Wilcox, G. (1985c) The araBAD operon of Salmonella typhimurium LT2. II. Nucleotide sequence of araA and primary structure of its product, L-arabinose isomerase. *Gene.* **34** (1) 123-128.

Lin, Y. & Tanaka, S. (2006) Ethanol fermentation from biomass resources: current state and prospects. *Applied Microbiology and Biotechnology.* **69** (6) 627-642.

Liu, Z., Ying, Y., Li, F., Ma, C. & Xu, P. (2010) Butanol production by Clostridium beijerinckii ATCC 55025 from wheat bran. *Journal of Industrial Microbiology and Biotechnology.* **37** (5) 495-501.

Lokman, B.C., van Santen, P., Verdoes, J.C., Kruse, J., Leer, R.J., Posno, M. & Pouwels, P.H. (1991) Organization and characterization of three genes involved in D-xylose catabolism in Lactobacillus pentosus. *Molecular and General Genetics.* **230** (1-2) 161-169.

Lopez-Contreras, A.M., Claassen, P.A., Mooibroek, H. & De Vos, W.M. (2000) Utilisation of saccharides in extruded domestic organic waste by Clostridium acetobutylicum ATCC 824 for production of acetone, butanol and ethanol. *Applied Microbiology and Biotechnology.* **54** (2) 162-167.

Lopez-Contreras, A.M., Gabor, K., Martens, A.A., Renckens, B.A., Claassen, P.A., Van Der Oost, J. & De Vos, W.M. (2004) Substrate-induced production and secretion of cellulases by *Clostridium acetobutylicum*. *Applied and Environmental Microbiology.* **70** (9) 5238-5243.

Lopez-Contreras, A.M., Martens, A.A., Szijarto, N., Mooibroek, H., Claassen, P.A., van der Oost, J. & de Vos, W.M. (2003) Production by *Clostridium acetobutylicum* ATCC 824 of CelG, a cellulosomal glycoside hydrolase belonging to family 9. *Applied and Environmental Microbiology.* **69** (2) 869-877.

Lopez-Contreras, A.M., Smidt, H., van der Oost, J., Claassen, P.A., Mooibroek, H. & de Vos, W.M. (2001) Clostridium beijerinckii cells expressing Neocallimastix patriciarum glycoside hydrolases show enhanced lichenan utilization and solvent production. *Applied and Environmental Microbiology.* **67** (11) 5127-5133.

Loviny-Anderton, T., Shaw, P.C., Shin, M.K. & Hartley, B.S. (1991) D-Xylose (D-glucose) isomerase from Arthrobacter strain N.R.R.L. B3728. Gene cloning, sequence and expression. *Biochemical Journal.* **277 (Pt 1)** 263-271.

Luthi, E., Love, D.R., McAnulty, J., Wallace, C., Caughey, P.A., Saul, D. & Bergquist, P.L. (1990) Cloning, sequence analysis, and expression of genes encoding xylan-degrading enzymes from the thermophile "Caldocellum saccharolyticum". *Applied and Environmental Microbiology.* **56** (4) 1017-1024.

Mach-Aigner, A.R., Pucher, M.E. & Mach, R.L. (2010) D-Xylose as a repressor or inducer of xylanase expression in Hypocrea jecorina (Trichoderma reesei). *Applied and Environmental Microbiology.* **76** (6) 1770-1776.

Madden, T. (2003) The BLAST Sequence Analysis Tool. *The NCBI handbook.* The National Library of Medicine. Available at: http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.chapter.ch16. Last accessed: 01/04/2009.

Maddox, I.-S. & Murray, A.E. (1983) Production of *n*-butanol by fermentation of wood hydrolysate. *Biotechnology Letters.* **5** (3) 175-178.

Maddox, I.S., Steiner, E., Hirsch, S., Wessner, S., Gutierrez, N.A., Gapes, J.R. & Schuster, K.C. (2000) The cause of "acid-crash" and "acidogenic fermentations" during the batch acetone-butanol-ethanol (ABE-) fermentation process. *Journal of Molecular Microbiology and Biotechnology.* **2** (1) 95-100.

Makrides, S.C. (1996) Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiological Reviews.* **60** (3) 512-538.

Manjasetty, B.A. & Chance, M.R. (2006) Crystal structure of Escherichia coli L-arabinose isomerase (ECAI), the putative target of biological tagatose production. *Journal of Molecular Biology.* **360** (2) 297-309.

Matuschek, M., Sahm, K. & Bahl, H. (1997) Molecular characterization of genes encoding a novel ABC transporter in *Thermoanaerobacterium thermosulfurigenes* EM1. *Current Microbiology.* **35** (4) 237-239.

Mauzy, C.A. & Hermodson, M.A. (1992) Structural and functional analyses of the repressor, RbsR, of the ribose operon of Escherichia coli. *Protein Science.* **1** 831-842.

Meaden, P.G., Aduse-Opoku, J., Reizer, J., Reizer, A., Lanceman, Y.A., Martin, M.F. & Mitchell, W.J. (1994) The xylose isomerase-encoding gene (xylA) of *Clostridium thermosaccharolyticum*: cloning, sequencing and phylogeny of XylA enzymes. *Gene.* **141** (1) 97-101.

Mes-Hartree, M.a.S., J.N. (1982) Butanol production of *Clostridium acetobutylicum* grown on sugars found in hemicellulose hydrolysates. *Biotechnology Letters.* **4** (4) 247-252.

Mingardon, F., Chanal, A., Tardif, C. & Fierobe, H.P. (2011) The issue of secretion in heterologous expression of Clostridium cellulolyticum cellulase-encoding genes in Clostridium acetobutylicum ATCC 824. *Applied and Environmental Microbiology.* **77** (9) 2831-2838.

Mitchell, W.J. (1998) Physiology of carbohydrate to solvent conversion by clostridia. *Advances in Microbial Physiology.* **39** 31-130.

Mitchell, W.J., Albasheri, K.A. & Yazdanian, M. (1995) Factors affecting utilization of carbohydrates by clostridia. *FEMS: Microbiology Reviews.* **7** 317-329.

Mitchell, W.J., Shaw, J.E. & Andrews, L. (1991) Properties of the glucose phosphotransferase system of *Clostridium acetobutylicum* NCIB 8052. *Applied and Environmental Microbiology.* **57** (9) 2534-2539.

Monedero, V., Gosalbes, M.J. & Perez-Martinez, G. (1997) Catabolite repression in Lactobacillus casei ATCC 393 is mediated by CcpA. *Journal of Bacteriology.* **179** (21) 6657-6664.

Mota, L.J., Tavares, P. & Sa-Nogueira, I. (1999) Mode of action of AraR, the key regulator of L-arabinose metabolism in *Bacillus subtilis*. *Molecular Microbiology.* **33** (3) 476-489.

Mtui, G. & Nakamura, Y. (2005) Bioconversion of lignocellulosic waste from selected dumping sites in Dar es Salaam, Tanzania. *Biodegradation.* **16** (6) 493-499.

Muscariello, L., Marasco, R., De Felice, M. & Sacco, M. (2001) The functional ccpA gene is required for carbon catabolite repression in Lactobacillus plantarum. *Applied and Environmental Microbiology.* **67** (7) 2903-2907.

References

Nakhmanovich, B.M. & Shcheblykina, N.A. (1959) Fermentation of pentoses of corn cob hydrolyzates by Clostridium acetobutylicum. *Mikrobiologiya.* **28** (1) 99-104 (Article in Russian).

Ni, Y. & Sun, Z. (2009) Recent progress on industrial fermentative production of acetone-butanol-ethanol by Clostridium acetobutylicum in China. *Applied Microbiology and Biotechnology.* **83** (3) 415-423.

Nimcevic, D. & Gapes, J.R. (2000) The acetone-butanol fermentation in pilot plant and pre-industrial scale. *Journal of Molecular Microbiology and Biotechnology.* **2** (1) 15-20.

Nolan, T., Hands, R.E. & Bustin, S.A. (2006) Quantification of mRNA using real-time RT-PCR. *Nature Protocols.* **1** (3) 1559-1582.

Nolling, J., Breton, G., Omelchenko, M.V., Makarova, K.S., Zeng, Q., Gibson, R., Lee, H.M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y.I., Tatusov, R.L., Sabathe, F., Doucette-Stamm, L., Soucaille, P., Daly, M.J., Bennett, G.N., Koonin, E.V. & Smith, D.R. (2001) Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *Journal of Bacteriology.* **183** (16) 4823-4838.

O'Brien, R.W. & Morris, J.G. (1971) Oxygen and the growth and metabolism of *Clostridium acetobutylicum*. *Journal of General Microbiology.* **68** (3) 307-318.

Ounine, K., Petitdemange, H., Raval, G. & Gay, R. (1983) Acetone-butanol production from pentoses by *Clostridium acetobutylicum Biotechnology Letters.* **5** (9) 605-610.

Ounine, K., Petitdemange, H., Raval, G. & Gay, R. (1985) Regulation and butanol inhibition of D-xylose and D-glucose uptake in *Clostridium acetobutylicium*. *Applied and Environmental Microbiology.* **49** (4) 874-878.

Paquet, V., Croux, C., Goma, G. & Soucaille, P. (1991) Purification and characterization of the extracellular α-amylase from *Clostridium acetobutylicum* ATCC 824. *Applied and Environmental Microbiology.* **57** (1) 212-218.

Parekh, M., Formanek, J. & Blaschek, H.P. (1999) Pilot-scale production of butanol by *Clostridium beijerinckii* BA101 using a low-cost fermentation medium based on corn steep water. *Applied Microbiology and Biotechnology.* **51** 152-157.

Perez, J., Munoz-Dorado, J., de la Rubia, T. & Martinez, J. (2002) Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview. *International Microbiology.* **5** (2) 53-63.

Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis.* **20** (18) 3551-3567.

Prabhu, P., Tiwari, M.K., Jeya, M., Gunasekaran, P., Kim, I.W. & Lee, J.K. (2008) Cloning and characterization of a novel L-arabinose isomerase from Bacillus licheniformis. *Applied Microbiology and Biotechnology.* **81** (2) 283-290.

Qureshi, N., Ezeji, T.C., Ebener, J., Dien, B.S., Cotta, M.A. & Blaschek, H.P. (2008) Butanol production by *Clostridium beijerinckii*. Part I: Use of acid and enzyme hydrolyzed corn fiber. *Bioresource Technology.* **99** (13) 5915-5922.

Qureshi, N., Li, X.L., Hughes, S., Saha, B.C. & Cotta, M.A. (2006) Butanol production from corn fiber xylan using *Clostridium acetobutylicum*. *Biotechnology Progress.* **22** (3) 673-680.

Qureshi, N., Saha, B.C. & Cotta, M.A. (2007) Butanol production from wheat straw hydrolysate using *Clostridium beijerinckii*. *Bioprocess and Biosystems Engineering.* **30** (6) 419-427.

Rasmussen, H., la Cour, T., Nyborg, J. & Schulein, M. (1994) Structure determination of glucose isomerase from Streptomyces murinus at 2.6 A resolution. *Acta Crystallographica Section D Biological Crystallography.* **50** (Pt 2) 124-131.

Reid, S.J., Rafudeen, M.S. & Leat, N.G. (1999) The genes controlling sucrose utilization in *Clostridium beijerinckii* NCIMB 8052 constitute an operon. *Microbiology.* **145** (6) 1461-1472.

Religa, T.L., Johnson, C.M., Vu, D.M., Brewer, S.H., Dyer, R.B. & Fersht, A.R. (2007) The helix-turn-helix motif as an ultrafast independently folding domain: the pathway of folding of Engrailed homeodomain. *Proceedings of the National Academy of Science U S A.* **104** (22) 9272-9277.

Rhimi, M. & Bejar, S. (2006) Cloning, purification and biochemical characterization of metallic-ions independent and thermoactive l-arabinose isomerase from the *Bacillus stearothermophilus* US100 strain. *Biochimica et Biophysica Acta.* **1760** (2) 191-199.

Rygus, T., Scheler, A., Allmansberger, R. & Hillen, W. (1991) Molecular cloning, structure, promoters and regulatory elements for transcription of the Bacillus megaterium encoded regulon for xylose utilization. *Archives of Microbiology.* **155** (6) 535-542.

Sa-Nogueira, I. & de Lencastre, H. (1989) Cloning and characterization of araA, araB, and araD, the structural genes for L-arabinose utilization in *Bacillus subtilis*. *Journal of Bacteriology.* **171** (7) 4088-4091.

Sa-Nogueira, I. & Mota, L.J. (1997) Negative regulation of L-arabinose metabolism in *Bacillus subtilis*: characterization of the araR (araC) gene. *Journal of Bacteriology.* **179** (5) 1598-1608.

Sa-Nogueira, I., Nogueira, T.V., Soares, S. & de Lencastre, H. (1997) The *Bacillus subtilis* L-arabinose (ara) operon: nucleotide sequence, genetic organization and expression. *Microbiology.* **143** (3) 957-969.

Sa-Nogueira, I. & Ramos, S.S. (1997) Cloning, functional analysis, and transcriptional regulation of the Bacillus subtilis araE gene involved in L-arabinose utilization. *Journal of Bacteriology.* **179** (24) 7705-7711.

Saari, G.C., Kumar, A.A., Kawasaki, G.H., Insley, M.Y. & O'Hara, P.J. (1987) Sequence of the Ampullariella sp. strain 3876 gene coding for xylose isomerase. *Journal of Bacteriology.* **169** (2) 612-618.

Sabathe, F., Belaich, A. & Soucaille, P. (2002) Characterization of the cellulolytic complex (cellulosome) of *Clostridium acetobutylicum*. *FEMS Microbiology Letters.* **217** 15-22.

Sabathe, F. & Soucaille, P. (2003) Characterization of the CipA scaffolding protein and in vivo production of a minicellulosome in *Clostridium acetobutylicum*. *Journal of Bacteriology.* **185** (3) 1092-1096.

Saddler, J.N., Yu, E.K.C., Mes-Hartree, M., Levitin, N. & Brownell, H.H. (1983) Utilization of enzymatically hydrolyzed wood hemicelluloses by microorganisms for production of liquid fuels. *Applied and Environmental Microbiology.* **45** (1) 153-160.

Saha, B.C. (2003) Hemicellulose bioconversion. *Journal of Industrial Microbiology and Biotechnology.* **30** (5) 279-291.

Saier, M.H., Jr. (2000) Families of proteins forming transmembrane channels. *Journal of Membrane Biology* **175** (3) 165-180.

Sanchez, S. & Smiley, K.L. (1975) Properties of D-xylose isomerase from Streptomyces albus. *Applied Microbiology.* **29** (6) 745-750.

Scangos, G.A. & Reiner, A.M. (1978) Ribitol and D-arabitol catabolism in Escherichia coli. *Journal of Bacteriology.* **134** (2) 492-500.

Scheler, A. & Hillen, W. (1994) Regulation of xylose utilization in Bacillus licheniformis: Xyl repressor-xyl-operator interaction studied by DNA modification protection and interference. *Molecular Microbiology.* **13** (3) 505-512.

Scheler, A., Rygus, T., Allmansberger, R. & Hillen, W. (1991) Molecular cloning, structure, promoters and regulatory elements for transcription of the Bacillus licheniformis encoded regulon for xylose utilization. *Archives of Microbiology.* **155** (6) 526-534.

Schmiedel, D. & Hillen, W. (1996) Contributions of XylR CcpA and cre to diauxic growth of Bacillus megaterium and to xylose isomerase expression in the presence of glucose and xylose. *Molecular and General Genetics.* **250** (3) 259-266.

Schmiedel, D., Kintrup, M., Kuster, E. & Hillen, W. (1997) Regulation of expression, genetic organization and substrate specificity of xylose uptake in *Bacillus megaterium*. *Molecular Microbiology.* **23** (5) 1053-1062.

Schneider, E. (2001) ABC transporters catalyzing carbohydrate uptake. *Research in Microbiology.* **152** (3-4) 303-310.

Scotcher, M.C., Huang, K.X., Harrison, M.L., Rudolph, F.B. & Bennett, G.N. (2003) Sequences affecting the regulation of solvent production in *Clostridium acetobutylicum*. *Journal of Industrial Microbiology and Biotechnology.* **30** (7) 414-420.

Seidel, G., Diel, M., Fuchsbauer, N. & Hillen, W. (2005) Quantitative interdependence of coeffectors, CcpA and cre in carbon catabolite regulation of Bacillus subtilis. *FEBS Journal.* **272** (10) 2566-2577.

Shallom, D. & Shoham, Y. (2003) Microbial hemicellulases. *Current Opinion in Microbiology.* **6** (3) 219-228.

Shamanna, D.K. & Sanderson, K.E. (1979) Genetics and regulation of D-xylose utilization in *Salmonella typhimurium* LT2. *Journal of Bacteriology.* **139** (1) 71-79.

Sharp, P.M., Cowe, E., Higgins, D.G., Shields, D.C., Wolfe, K.H. & Wright, F. (1988) Codon usage patterns in Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster and Homo sapiens; a review of the considerable within-species diversity. *Nucleic Acids Research.* **16** (17) 8207-8211.

Sigma-Aldrich (2008) TargeTron® Gene Knockout System. User Guide. Available at: http://www.sigmaaldrich.com/etc/medialib/docs/Sigma/General_Information/targetron-user-guide.Par.0001.File.tmp/targetron-user-guide.pdf. Last accessed: 10/10/2011.

Sizemore, C., Wieland, B., Gotz, F. & Hillen, W. (1992) Regulation of Staphylococcus xylosus xylose utilization genes at the molecular level. *Journal of Bacteriology.* **174** (9) 3042-3048.

Song, S. & Park, C. (1997) Organization and regulation of the D-xylose operons in *Escherichia coli* K-12: XylR acts as a transcriptional activator. *Journal of Bacteriology.* **179** (22) 7025-7032.

Sonnhammer, E.L., von Heijne, G. & Krogh, A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings of the International Conference on Intellegent Systems for Molecular Biology.* **6** 175-182.

Sprenger, G.A., Schorken, U., Sprenger, G. & Sahm, H. (1995) Transaldolase B of Escherichia coli K-12: cloning of its gene, talB, and characterization of the enzyme from recombinant strains. *Journal of Bacteriology.* **177** (20) 5930-5936.

Spyropoulos, I.C., Liakopoulos, T.D., Bagos, P.G. & Hamodrakas, S.J. (2004) TMRPres2D: high quality visual representation of transmembrane protein models. *Bioinformatics.* **20** (17) 3258-3260.

Sumiya, M. & Henderson, P.J. (1989) The D-xylose binding protein of Escherichia coli. *Biochemical Society Transactions.* **17** 553-554.

Takata, G., Poonperm, W., Rao, D., Souda, A., Nishizaki, T., Morimoto, K. & Izumori, K. (2007) Cloning, expression, and transcription analysis of L-arabinose isomerase gene from Mycobacterium smegmatis SMDU. *Bioscience Biotechnology and Biochemistry.* **71** (12) 2876-2885.

Takeda, Y., Takase, K., Yamato, I. & Abe, K. (1998) Sequencing and characterization of the xyl operon of a gram-positive bacterium, *Tetragenococcus halophila*. *Applied and Environmental Microbiology.* **64** (7) 2513-2519.

Tangney, M., Brehm, J.K., Minton, N.P. & Mitchell, W.J. (1998) A gene system for glucitol transport and metabolism in *Clostridium beijerinckii* NCIMB 8052. *Applied and Environmental Microbiology.* **64** (5) 1612-1619.

Tangney, M. & Mitchell, W.J. (2000) Analysis of a catabolic operon for sucrose transport and metabolism in *Clostridium acetobutylicum* ATCC 824. *Journal of Molecular Microbiology and Biotechnology.* **2** (1) 71-80.

Tangney, M. & Mitchell, W.J. (2007) Characterisation of a glucose phosphotransferase system in *Clostridium acetobutylicum* ATCC 824. *Applied Microbiology and Biotechnology.* **74** (2) 398-405.

Tangney, M., Winters, G.T. & Mitchell, W.J. (2001) Characterization of a maltose transport system in *Clostridium acetobutylicum* ATCC 824. *Journal of Industrial Microbiology and Biotechnology.* **27** (5) 298-306.

Taylor, G.K. & Goodlett, D.R. (2005) Rules governing protein identification by mass spectrometry. *Rapid Communications in Mass Spectrometry.* **19** (23) 3420.

Thostrup, L. (2006) Marken - en stor solfanger. *Forskning det biovidenskabelige Fakultet Kobenhavns Universitet.* Available at: http://www.life.ku.dk/forskning/online_artikler/artikler/marken_en_stor_solfanger.aspx. Last accessed: 02/04/2009.

Titgemeyer, F., Reizer, J., Reizer, A. & Saier, M.H., Jr. (1994) Evolutionary relationships between sugar kinases and transcriptional repressors in bacteria. *Microbiology.* **140 (Pt 9)** 2349-2354.

Toivonen, J.M., Boocock, M.R. and & Jacobs, H.T. (1999) Modelling in *Escherichia coli* of mutations in mitoribosomal protein S12: novel mutant phenotypes of rpsL. *Molecular Microbiology.* **31** (6) 1735-1746.

Ture, S., Uzun, D. & Ture, E. (1997) The potential use of sweet sorghum as a non-polluting source of energy. *Energy.* **22** (1) 17-19.

Udvardi, M.K., Czechowski, T. & Scheible, W.R. (2008) Eleven golden rules of quantitative RT-PCR. *Plant Cell.* **20** (7) 1736-1737.

United Nations (1998) Kyoto protocol to the United Nations framework convention on climate change. Available at: http://unfccc.int/resource/docs/convkp/kpeng.pdf. Last accessed: 14/02/2012.

Vieille, C., Hess, J.M., Kelly, R.M. & Zeikus, J.G. (1995) xylA cloning and sequencing and biochemical characterization of xylose isomerase from *Thermotoga neapolitana*. *Applied and Environmental Microbiology.* **61** (5) 1867-1875.

Wakarchuk, W.W., Campbell, R.L., Sung, W.L., Davoodi, J. & Yaguchi, M. (1994) Mutational and crystallographic analyses of the active site residues of the Bacillus circulans xylanase. *Protein Science.* **3** (3) 467-475.

Wallace, R.G., Lee, N. & Fowler, A.V. (1980) The araC gene of Escherichia coli: transcriptional and translational start-points and complete nucleotide sequence. *Gene.* **12** (3-4) 179-190.

Wang, Y., Huang, Z., Dai, X., Liu, J., Cui, T., Niu, L., Wang, C. & Xu, X. (1994) The sequence of xylose isomerase gene from Streptomyces diastaticus No. 7 M1033. *Chinese Journal of Biotechnology.* **10** (2) 97-103.

Wheals, A.E., Basso, L.C., Alves, D.M. & Amorim, H.V. (1999) Fuel ethanol after 25 years. *Trends in Biotechnology.* **17** (12) 482-487.

Wilhelm, M. & Hollenberg, C.P. (1984) Selective cloning of *Bacillus subtilis* xylose isomerase and xylulokinase in *Escherichia coli* genes by IS5-mediated expression. *EMBO Journal.* **3** (11) 2555-2560.

Wolf, M., Geczi, A., Simon, O. & Borriss, R. (1995) Genes encoding xylan and beta-glucan hydrolysing enzymes in Bacillus subtilis: characterization, mapping and construction of strains deficient in lichenase, cellulase and xylanase. *Microbiology.* **141 (Pt 2)** 281-290.

Wong, H.C., Ting, Y., Lin, H.C., Reichert, F., Myambo, K., Watt, K.W., Toy, P.L. & Drummond, R.J. (1991) Genetic organization and regulation of the xylose degradation genes in Streptomyces rubiginosus. *Journal of Bacteriology.* **173** (21) 6849-6858.

Wong, K.K., Tan, L.U. & Saddler, J.N. (1988) Multiplicity of beta-1,4-xylanase in microorganisms: functions and applications. *Microbiological Reviews.* **52** (3) 305-317.

Woolley, R.C., Pennock, A., Ashton, R.J., Davies, A. & Young, M. (1989) Transfer of Tn1545 and Tn916 to Clostridium acetobutylicum. *Plasmid.* **22** (2) 169-174.

Yu, E.K.C. & Saddler, J.N. (1983) Enhanced acetone-butanol fermentation by Clostridium acetobutylicum grown on D-xylose in the presence of acetic or butyric acid. *FEMS Microbiology Letters.* **18** 103-107.

Yu, Y., Tangney, M., Aass, H.C. & Mitchell, W.J. (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., Nielsen, J. & Olsson, L. (2001) Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration. *Applied Microbiology and Biotechnology.* **56** (1-2) 17-34.

Zeilinger, S., Mach, R.L., Schindler, M., Herzog, P. & Kubicek, C.P. (1996) Different inducibility of expression of the two xylanase genes xyn1 and xyn2 in Trichoderma reesei. *Journal of Biological Chemistry.* **271** (41) 25624-25629.

Zverlov, V.V., Berezina, O., Velikodvorskaya, G.A. & Schwarz, W.H. (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** (5) 587-597.