INVESTIGATION OF LACTOSE UTILIZATION GENES IN *CLOSTRIDIUM ACETOBUTYLICUM*

by

KERRIE RUTH HANCOCK

A thesis presented in partial fulfilment of the requirements
for the degree of Doctor of Philosophy in Biotechnology
at Massey University, Palmerston North
New Zealand

1988

ABSTRACT

Preparatory to constructing a genomic library of Clostridium acetobutylicum, cell lysis and DNA preparation conditions leading to undegraded DNA were examined. Early to mid-exponential phase cells lysed more readily than cells at late-exponential or stationary phase. Lysis was facilitated by 0.3% w/v glycine in the growth medium. Achromopeptidase was a more effective lytic enzyme than either mutanolysin or lysozyme.

All strains of C. acetobutylicum produced high levels of DNase activity, coinciding with the late-exponential and stationary growth phases. Chromosomal DNA isolated from all strains of C. acetobutylicum was rapidly degraded. This degradation was not prevented by the use of various protein inactivating agents. The adverse effect of oxygen and related radicals on the DNA of this strict anaerobe was considered to be responsible. Undegraded DNA was isolated by protecting the cells to be lysed from oxygen.

A genomic library of C. acetobutylicum NCIB2951 in the cosmid vector pLAFR1, constituting 3,500 recombinant clones, was prepared. Clones from this library complemented various Escherichia coli auxotrophic mutations, showing that C. acetobutylicum genes are expressible in E. coli.

Recombinant clones coding for the ß-galactosidase of NCIB2951 were isolated from the genomic library using the chromogenic substrate X-gal. The lacY mutation of HB101 could not be complemented by these clones, suggesting that a classical lac operon system does not exist in C. acetobutylicum.

The ß-galactosidase (cbg) gene was further subcloned on a 5.2 kb EcoRI fragment, and was expressed when the fragment was cloned in either orientation. Cbg was thus expressed from its own promoter. The cbg gene is inducible by lactose in C. acetobutylicum. When cloned into E. coli, however, this gene was expressed constitutively, the level being unaffected by the presence of the inducer,

IPTG or glucose.

Six strains of C. acetobutylicum possessed a sequence highly homologous to the cloned ß-galactosidase fragment. The ß-galactosidase gene region of NCIB2951 showed only low homology to the DNA from other Gram-positive bacteria (Streptococcus lactis ATCC7962, Streptococcus thermophilus DRI1424, Lactobacillus bulgaricus DRI20056, Lactobacillus helviticus DRI20064), and no detectable homology to DNA from Gram-negative bacteria (E. coli DC272 or Rhizobium loti PN2231).

The ß-galactosidase activity of the 5.2 kb fragment was inactivated by Tn5 insertion at either of two loci. Locus I (400 bp) was approximately 500 bp from locus II (approximately 3.2 kb). Maxi-cell analysis identified a 100 kDa protein as the ß-galactosidase gene product.

The 5.2 kb fragment was sequenced and analyzed. Three ORF's were identified. ORF1 (cbgA) coded for the structural ß-galactosidase gene. Significant amino acid homology was detected with the amino acid sequences of the lacZ, ebgA (E. coli) and lacZ (Klebsiella pneumoniae). ORF2 (cbgR) coded for a small regulatory protein which shared homology with the amino acid sequence of the "0.3 kb gene" from Bacillus subtilis. ORF3 coded for a truncated protein which shared significant homology with the N-terminal amino acid regions of spo0A and spo0F (B. subtilis), two regulatory proteins of the two-component system. Hence, no lac operon exists in C. acetobutylicum.

ACKNOWLEDGEMENTS

I am indebted to my supervisors, Dr Ian Maddox, Dr Lindsay Pearce and Professor Barry Scott for their collective inspiration, encouragement, guidance, advice, and patience during this PhD, and for providing the facilities which enabled this study. Special thanks goes to the following people for their encouragement, advice and friendship which was greatly appreciated.

At New Zealand Dairy Research Institute (NZDRI): Audrey, Graham and Guy for your constant interest and encouragement, thanks. Also Sam and Charlotte.

At Massey: Noemi, Feng Feng, and Sirinda (fellow PhDs at Biotechnology) and Mike, for your friendship.

Everyone at 'Scott Base' and the surrounding area: Caroline, Michelle, Grant, Robert, Gopher, Chris, Eva, Sharon, Trish, Mike, Lawrence, Sally, George, and Leslie for their friendship, sense of humour and for sharing the lab with me (i.e. tolerance). To those involved in the 'conspiracy' concerning my computer and I - the joke was much appreciated!!!

To Eva, thanks for your assistance and tutoring concerning the SDS-PAGE.

Also to 'Garfield' and 'Footrot Flats' for keeping me sane.

I gratefully acknowledge the following grants, without which this work could not have been done: New Zealand Energy Research Development Council, and the New Zealand Dairy Research Institute.

I would also like to thank Dan Legg (NZDRI) and Robert Hickson (Scott Base) for their great patience and help in using the computers to create this thesis. Also thanks to Paul Le Ceve for his expertise with the photos, and Gail Haydock for help with the typing.

To my family, I would like to show my greatest gratitude, for their unending love and support.

Dedicated to my mother

and especially to my late father

LIST OF CONTENTS

Page

Page

LIST OF FIGURES

Page

Page

Page

LIST OF TABLES

ABBREVIATIONS

kb	kilobases
pfu	plaque-forming units
Ap	ampicillin
Cm	chloramphenicol
Km	kanamycin
Tc	tetracycline
SDS	sodium dodecyl sulphate
PEG	polyethylene glycol
kDa	kilodaltons
MDa	megadaltons
RF	replicative form
TTC	2,3,5 triphenyl tetrazolium chloride
X-gal	5-bromo-4-chloro-3-indoyl galactopyranoside
SS	single stranded
DS	double stranded
IPTG	isopropyl thiogalactopyranoside
ORF	open reading frame
SD	Shine-Dalgarno sequence

CHAPTER 1: LITERATURE REVIEW

PART I: ABE FERMENTATION

1.1 INTRODUCTION

There is an extensive literature on the acetone/butanol/ethanol (ABE) fermentation, including numerous reviews. They include those by Killifer (1927), Gabriel (1928), Beesch (1952, 1953), Steel (1958), Prescott and Dunn (1959), Rose (1961), Ross (1961), Spivey (1978), Hastings (1978), Walton and Martin (1979), Zeikus (1980), Volesky et al. (1981), Robson and Jones (1982), Volesky and Szeczesny (1983), Linden and Moreira (1982), Moreira (1983), Gibbs (1983), Jones and Woods (1986b), and Ennis et al. (1986b). These authors have adequately covered the topic, and therefore no attempt will be made to give a detailed review of all the aspects of the literature. Instead, this chapter will give a basic outline of the ABE fermentation and present those factors which relate to the present study.

1.1.1 History

The ABE fermentation has a long and well documented history as a successful industrial fermentation process (Johnson, 1947; Prescott and Dunn, 1959; Ross, 1961; Hastings, 1978; Compere and Griffith, 1979; Walton and Martin, 1979; Volesky et al., 1981; Gibbs, 1983).

The first investigation was conducted by Pasteur in 1862, who observed the microbial production of butanol from lactic acid and calcium lactate. In 1909, interest in commercializing butanol production was initiated as a means of obtaining butadiene, used as a raw material for the manufacture of a synthetic rubber (Gabriel, 1928; Gabriel and Crawford, 1930), and acetone, which was being used in the dyestuff industry. In 1911 Strange and Graham Ltd employed

two teams of microbiologists (Fernbach and Schoen at the Pasteur Institute; Perkin and Weizmann at the University of Manchester) to find fermentation methods for meeting synthetic rubber feedstock needs (Gabriel, 1928). A bacterium capable of producing isoamyl alcohol or butanol was required, which provided the best route for the production of isoprene or butadiene. Both groups isolated bacteria capable of producing a mixture of amyl alcohol, butanol, ethanol and acetone. Weizmann left the employment of Strange and Graham Ltd in 1912 and, continuing with his own culture isolation work, eventually isolated a strain capable of producing four times the yield of acetone and butanol as the original bacterium as well as being able to utilize a variety of carbohydrates. The bacterium was originally named Bacillus granulobacter pectinovorum but was later renamed Clostridium acetobutylicum Weizmann by McCoy et al. (1926). From 1913 onwards it was used to produce acetone and butanol from potatoes in two commercial plants.

The outbreak of World War I in 1914 caused a dramatic increase in the industrial demand for acetone. It was now required as the gelatinizing agent for nitrocellulose in the naval warfare explosive cordite, as a propellant for heavy artillery shells, and as the solvent for cellulose acetate in aircraft wing dope production. To meet this need, the British government introduced the Weizmann strain into Strange and Graham plants, and adapted distilleries to take advantage of the higher production of acetone and the bacterium's ability to utilise a greater variety of the readily available starches as feedstocks. In May 1917, upon entry of the USA into the war, a plant was opened in Indiana which supplemented the acetone output from the existing plants. The butanol produced was stockpiled.

After the end of World War I there was an immediate decline in the requirement for acetone, causing the closure of some plants, but the ABE fermentation continued for some years due to its product use in manufacturing products such as varnishes, and in textile and isoprene production. Later,

DuPont de Nemours and Co. developed low viscosity nitrocellulose lacquers for the automobile industry and found butanol and butyl acetate to be the solvents of choice for this coating system. Weizmann (1919) obtained a US patent and large factories began supplying the butanol for conversion to butyl acetate. On expiration of the patent, new plants were opened in USA, Puerto Rico, and Japan, coinciding with development of new bacterial strains (Walton and Martin, 1979).

The onset of World War II again saw demand for the manufacture of cordite, and the ABE fermentation was given top priority, which led to the development of continuous distillation processes. Subsequent to 1936, plants were also built in Japan, India, Australia, South Africa, USSR and eastern bloc countries, with the plant in South Africa continuing production until 1983. At the end of World War II in 1945, 66% of butanol and 10% of acetone in the USA were produced by the ABE fermentation.

After World War II, the substantial use of the primary substrate, molasses, in cattle fodder, and the commercial production of acetone, butanol and ethanol by chemical means, led to a rapid decline in the ABE fermentation process and industrial production ceased in both the U.S.A. and England during the 1960's due to the unfavourable economics. Adverse factors included the increasing cost both of raw materials such as molasses and of disposing of the fermentation effluent. Further, the petrochemical industry was by now capable of producing vast quantities of chemicals at relatively cheap prices during the cracking of petroleum. Today, acetone is produced by the cumene hydroperoxide process or by catalytic dehydrogenation of isopropanol. Butanol is produced by the oxo process or the aldol process (Moreira, 1983).

1.1.2 Revival of the fermentation process

Renewed interest in the C. acetobutylicum fermentation process has arisen due to society being confronted with the dwindling resources of fossil fuels and chemical feedstocks, plus the proliferation of wastes generated by municipalities, agriculture and industries. These problems have been coupled with large increases in the price of petroleum and mounting environmental pressures and expenditures resulting from the required treatment of waste products. The above problems have had a profound effect on the development of alternative technology based on potentially attractive resources, that is, the conversion of reusable waste products into more desirable commodities such as alcohols.

In addition, new uses of both butanol and acetone are being found. Butanol is used extensively in the manufacture of hydraulic fluids, detergents, butylated melamine resins, urea-formaldehyde resins, amines for gasoline additives, esters in the protective coatings industries, plasticizers, dyeing agents and as a coagulant in the micellular flooding process in tertiary oil recovery systems from existing wells (Casida, 1968; Compere and Griffith, 1979). Acetone is used for nitrocellulose solvents, vinyl resins, and derivatives such as methyl isobutyl ketone, methyl isobutyl carbinol, methyl methacrylate and as the spinning solvent in the production of cellulose acetate.

Although the ABE fermentation is not at present competitive with the petrochemical route, in light of the above, a number of groups worldwide are reexamining the fermentation to increase our understanding of the process. Much of the work is on developments in fermentation technology to improve yield and recovery (see review by Ennis et al., 1986b). More recently, studies on the biochemical, physiological and genetic aspects of C. acetobutylicum have been initiated (Woods, 1982; Truffaut and Sebald, 1983; Hartmanis et al., 1984). If a combination of these two approaches leads to sufficiently improved economics, the ABE fermentation is likely to be revived as the preferred method

for butanol and acetone production.

1.2 THE ORGANISM

1.2.1 Classification

The ABE fermentation is carried out by butyric anaerobic bacteria belonging to the genus Clostridium. Two clostridial species have been historically used in, or developed for, solvent production. These are Clostridium acetobutylicum and Clostridium beijerinckii, [syn C. butylicum] (George et al., 1983). They are commonly known as butyric organisms and produce significant amounts of neutral products such as acetone, butanol and ethanol from carbohydrate fermentation by way of organic acid intermediates (Prescott and Dunn, 1959; Cummins and Johnson, 1971; George et al., 1983).

The morphological, cultural and biochemical characteristics of C. acetobutylicum are well documented (McCoy et al., 1926; Cato et al. (1986); Volesky et al., 1981). Cato et al. (1986) used classical morphological, biochemical and physiological criteria to divide the genus into four groups based on the position of the spores, gelatin liquefication and toxin-antitoxin identification. C. acetobutylicum was placed in group II based on the subterminal position of the spores and positive gelatin hydrolysis.

1.2.2 Morphological characteristics

C. acetobutylicum can be described as a saccharolytic chemoorganotroph. Vegetative cells are straight rods with rounded ends and exist singly, paired or in chains, depending on the growth phase and medium used. Cell dimensions vary with a length of 2.4 to 4.7 μm and a width of 0.6 to 0.9 μm. Subterminal oval spores without appendages or exosporia are approximately 1 μm by 1.5 μm in size. Spores are usually produced when growth conditions become

unfavourable for continued vegetative growth, such as toxic product accumulation (Ross, 1961; Casida, 1968; Hastings, 1978). C. acetobutylicum is a strict anaerobe with an optimal growth temperature of 34°C (Cato et al., 1986). The optimal pH value for growth and acid production is 6.5, while a pH of less than 5.3 is often required for solvent production (Spivey, 1978; Gottschal and Morris, 1981; Andersch et al., 1982; Bahl et al., 1982a; George et al., 1983; Holt et al., 1984, Monot et al., 1984). Cells are Gram-positive, but tend to become gram-variable with age. The vegetative cells are motile, by means of peritrichous flagella. Actively motile cells are regarded as a prerequisite for good solvent production (Gutierrez and Maddox, 1987). Division by transverse fission results in long chains of bacteria on solid media. These break apart into single cells in liquid medium during vigorous fermentation.

1.2.3 Cultural characteristics

C. acetobutylicum grows anaerobically on agar as raised, circular surface colonies (3-5 mm in diameter). The colonies appear greyish white in colour, with a translucent glossy surface and an irregular margin (Cato et al., 1986). Growth spreads on the agar surface, a feature common to many clostridia. Gas bubbles may be apparent in the colonies (Kutzenok and Aschner, 1952).

C. acetobutylicum can utilize a wide variety of carbohydrates, owing to its unusual ability to ferment not only hexoses but also pentoses, disaccharides and oligosaccharides (Compere and Griffith, 1979; Mes-Hartree and Saddler, 1982; Maddox, 1980). Those fermented are listed in Table 1.1. Metabolites produced include acetate, butyrate, acetone, butanol, ethanol, CO_2, H_2, acetoin and acetylmethylcarbinol. C. acetobutylicum fixes atmospheric nitrogen (Rosenblum and Wilson, 1949), hydrolyses gelatin and grows in milk. In the latter medium, it produces a stormy fermentation with abundant gas. Some strains are caseinase positive, produce H_2S or are DNase positive. C. acetobutylicum, however, is

Table 1.1: Carbohydrates fermented by C. acetobutylicum

Pentoses:	arabinose	Disaccharides:	sucrose
	xylose		maltose
			trehalose
			lactose
Hexoses:	glucose	Polysaccharides:	starch
	galactose		dextrin
	mannose		glycogen
	salicin		amygdalin
	mannitol		inulin
	X methyl-glucoside		melezitose

unable to produce urease, lecithinase, indole, lipase, or catalase, and cannot reduce nitrate or sulphate to nitrite or sulphite, respectively. Most strains produce granulose and capsules.

1.2.4 Substrate range and nutritional requirements

The ability of C. acetobutylicum to utilise a large variety of carbohydrates enables it to convert many different waste biomass and agricultural residues into solvents. Substrates that have been examined include: wood hydrolysates, (Mes-Hartree and Saddler, 1982; Maddox and Murray, 1983): whey filtrate (Maddox, 1980; Maddox et al., 1981; Linden and Moriera, 1982; Welsh and Veliky, 1984): bagasse, wheat and rice straws (Marchal et al., 1984a,b; Soni et al., 1982): lignocellulosic residues (Yu et al., 1984; Saddler et al., 1983): molasses (Maddox 1982; Qadeer et al., 1980): agricultural residues (Soni et al., 1982; Ounine et al., 1983): and corncobs (Langlykke, 1948; Nakhmanovich and Shcheblykina, 1959).

Oxford et al. (1940) identified biotin and a factor present in yeast as necessary for the growth of some strains of C. acetobutylicum. The yeast factor was later identified as p-aminobenzoic acid by Rubbo et al., (1941). These results have

been confirmed by Lampen and Peterson (1943), and Reyes and Mickleson (1945). Monot et al. (1982) found that cell growth was also dependent on the presence of magnesium, iron and potassium, but that in excess, magnesium and manganese had deleterious effects on growth and solvent production.

The organism thus has considerable versatility and with its ability to ferment a wide range of carbohydrates, and limited nutritional requirements, can be readily grown on many substrates.

1.2.5 Life cycle

The morphology of the culture varies depending on the strain of organism used and the formulation of the medium. There is also variation during culture growth which can be used to assess the age and progress of the fermentation process, as there is a distinct and relatively constant variation with time (Spivey, 1978). The morphological changes in C. acetobutylicum P262, the commercial strain used at National Chemical Products Ltd., Germiston, South Africa, have been studied and correlated with growth and solvent production (Jones et al., 1982; Long et al., 1983, 1984b).

In a typical batch fermentation process, there are several distinct stages. When the medium is first inoculated, the rods are elongated, with developed septa, and form long chains of phase dark, strongly Gram-positive cells with sporadic and sluggish motility. After 6 h the chains begin to break up, releasing highly motile individual rods. The rods are actively dividing and the growth rate is at a maximum. By 14 h growth decreases to zero and motility steadily decreases. There is a large population of non- motile cells by 18 h. The pH breakpoint is reached and granulose accumulation begins after 19 to 20 h. At the end of fermentation (30 h) the majority of the cells are typical swollen, phase-bright Gram-positive clostridial forms with a clearly defined extracellular capsule. A small percentage are Gram-negative that proceed to degenerate.

1.2.6 Sporulation

Bacterial sporulation has been extensively studied in various Bacillus species as well as a few Clostridium species (Kay and Warren, 1968; Dawes et al., 1969; Walker,1970; Hsu and Ordal, 1970; Mackey and Morris, 1971; Aronson and Fitz-James, 1976) resulting in several excellent reviews (Tipper and Gauthier, 1972; Fitz-James and Young, 1969; Young and Mandelstam, 1979). Woods and Jones (1986) review the events associated with sporulation in Clostridium species.

The stages of sporulation by anaerobic Clostridium are basically similar to those for Bacillus species (Fitz-James and Young, 1969). However, clostridial sporulation is often accompanied by characteristic changes in cell shape, such as considerable increase in cell length, a clubbing to accommodate the forespore, and later the spore, as in C. pectinovorum (Fitz-James and Young, 1969). Formation of the axial filament is also rare in Clostridium (Johnstone and Holland, 1977; Woods and Jones, 1986).

Prescott and Dunn (1959) noted that good solvent yield is associated with an ability to sporulate well. Gottschal and Morris (1981) suggested a relationship between sporulation and solvent production, showing that a loss of sporeforming capacity during continuous culture is associated with loss of ability to produce solvents. A recent study (Long et al., 1984a, b) on sporulation of C. acetobutylicum P262 revealed that the initiation of solvent production and clostridial stage formation were essential for sporulation. They concluded that the shift to the solventogenic phase during the fermentation process was linked to both the inhibition of cell division and the triggering of various physiological and morphological changes. Under nutrient limitation there was inhibition of growth, no solvent production and no differentiation into spores. A requirement for a minimum concentration of glucose and ammonium ions during endospore formation was shown, the yield of endospores at low levels of glucose and ammonium being proportional to the concentration of these substrates. This is

in agreement with data published by Gottschal and Morris (1981), Bahl et al. (1982a), Martin et al. (1983) and Yu and Saddler (1983). This situation is in contrast to that existing with aerobic spore-formers such as B. subtilis, where sporulation is initiated by nutrient starvation (Young and Mandelstam, 1979).

C. acetobutylicum sporulation mutants blocked both before and after forespore septum formation are still able to accumulate granulose, produce a clostridial stage and form solvents (Jones et al., 1982). Long et al. (1984a) reported that C. acetobutylicum sporulation mutants which were unable to form clostridial stages could not produce solvents, store granulose or sporulate. Granulose mutants could still produce solvents and mature spores. Meinecke et al. (1984) reported that C. acetobutylicum mutants which are unable to store granulose were also unable to initiate sporulation but still able to produce solvents. Thus, at least the final stages of sporulation are not a pre-requisite for solvent production.

1.3 BIOCHEMISTRY OF C. ACETOBUTYLICUM

1.3.1 Course of fermentation

In a normal batch culture, the fermentation process using C. acetobutylicum consists of two distinct phases which correspond to the two stage mechanism of product formation. A suitable example of a typical outline of an ABE fermentation profile can be seen in Fig. 1.1 (Prescott and Dunn, 1959; Spivey, 1978; Jones et al., 1982). The initial acidogenic phase consists of active logarithmic growth of the bacterium over approximately 12 h. At a certain pH value known as the pH breakpoint, there is a switching from the acidogenic phase to the solventogenic phase (Davies and Stephenson, 1941; Barber et al., 1979). The reassimilation of the acids occurs simultaneously with continued consumption of the carbohydrate source. This uptake of acids is seen as a

detoxification process in response to an unfavourable environment as the resulting neutral solvents that are produced are less toxic to the cells than the previously synthesised organic acids (Woods and Jones, 1986).

1.3.2 Metabolic pathways

The main biochemical pathways by which various carbohydrate substrates are converted to acids, solvents, hydrogen and CO_2 by C. acetobutylicum and other butyric acid bacteria have been investigated and reviewed (Woods et al., 1945; Doelle, 1981; Thauer et al., 1977; Lenz and Moreira, 1980; Gottschalk and Bahl, 1981; Volesky et al., 1981; Gottschalk, 1986; Volesky and Szczesny, 1983; Monot et al., 1983; Kim et al., 1984; Papoutsakis, 1984; Hartmanis and Gatenbeck, 1984; Haggstrom, 1985; Matta-El-Amouri et al., 1985; Peptitdemange and Gay, 1986; Rogers, 1986; Jones and Woods, 1986b). A general diagram of the biochemical pathways involved in solvent production by C. acetobutylicum is given in Fig. 1.2 (Jones and Woods, 1986b).

1.3.2.1 Acid producing pathways

Hexose sugars are metabolised via the Embden-Meyerhof (EMP) pathway. When glucose is converted to pyruvate two molecules of ATP and two molecules of reduced NADH are produced for each molecule of glucose metabolised. These are subsequently used in biosynthesis reactions (energy) and as a hydrogen donor, respectively.

Pentose sugars, such as arabinose and xylose are utilized by the pentose phosphate pathway, leading to formation of pentose-5-phosphate or other intermediate compounds (Volesky and Szczesny, 1983). Conversion of the pentose-5-phosphate occurs via the transketolase and transaldolase enzymes resulting in the production of fructose-6-phosphate and glyceraldehyde-3

Fig.1.1: Typical outline of an ABE fermentation as carried out by C. acetobutylicum (Volesky et al., 1981).

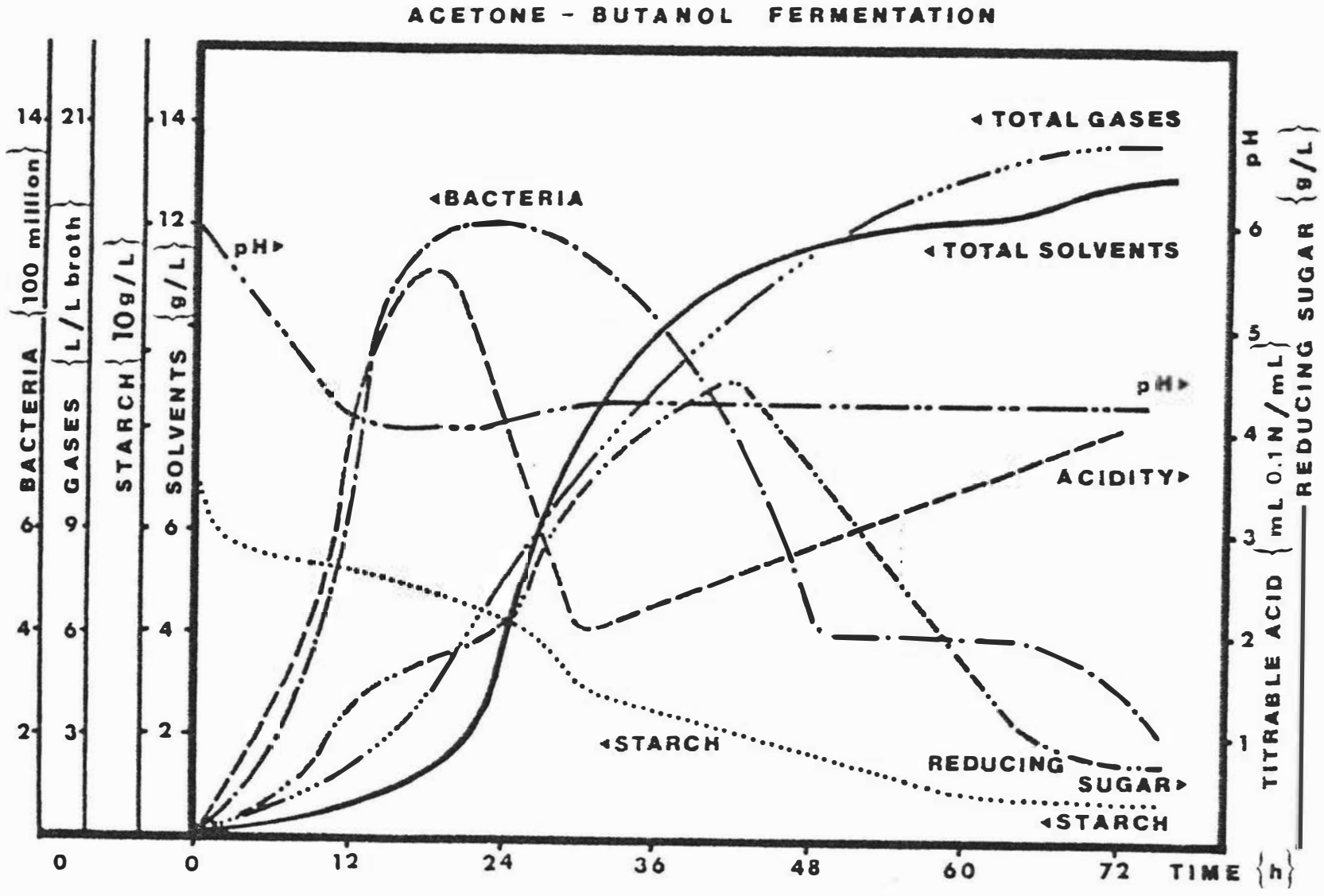

ACETONE - BUTANOL FERMENTATION
BACTERIA (100 million)
GASES (L/L broth)
STARCH (10g/L)
SOLVENTS (g/L)
TOTAL GASES
BACTERIA
TOTAL SOLVENTS
pH
pH
ACIDITY
STARCH
REDUCING SUGAR
STARCH
TITRABLE ACID (mL 0.1N/mL)
REDUCING SUGAR (g/L)
TIME (h)
0
12
24
36
48
60
72

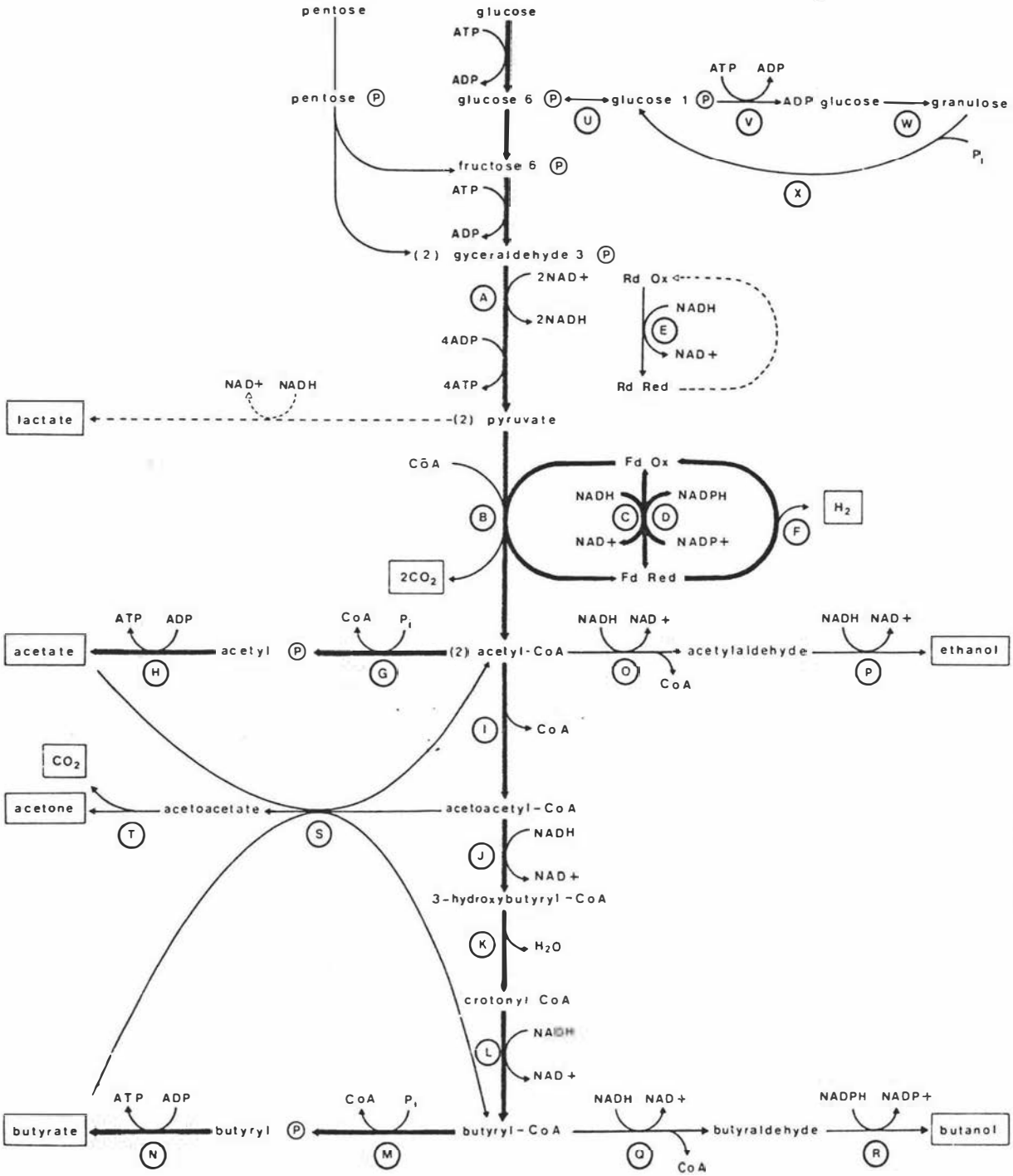
Fig.1.2: Biochemical pathways for the synthesis of acetone, butanol, and ethanol in C. acetobutylicum (Jones and Woods,1986).

Reactions which predominate during the acidogenic phase (a) and the solventogenic phase (b) of the fermentation are shown in thick arrows. Extracellular products are shown in boxes.

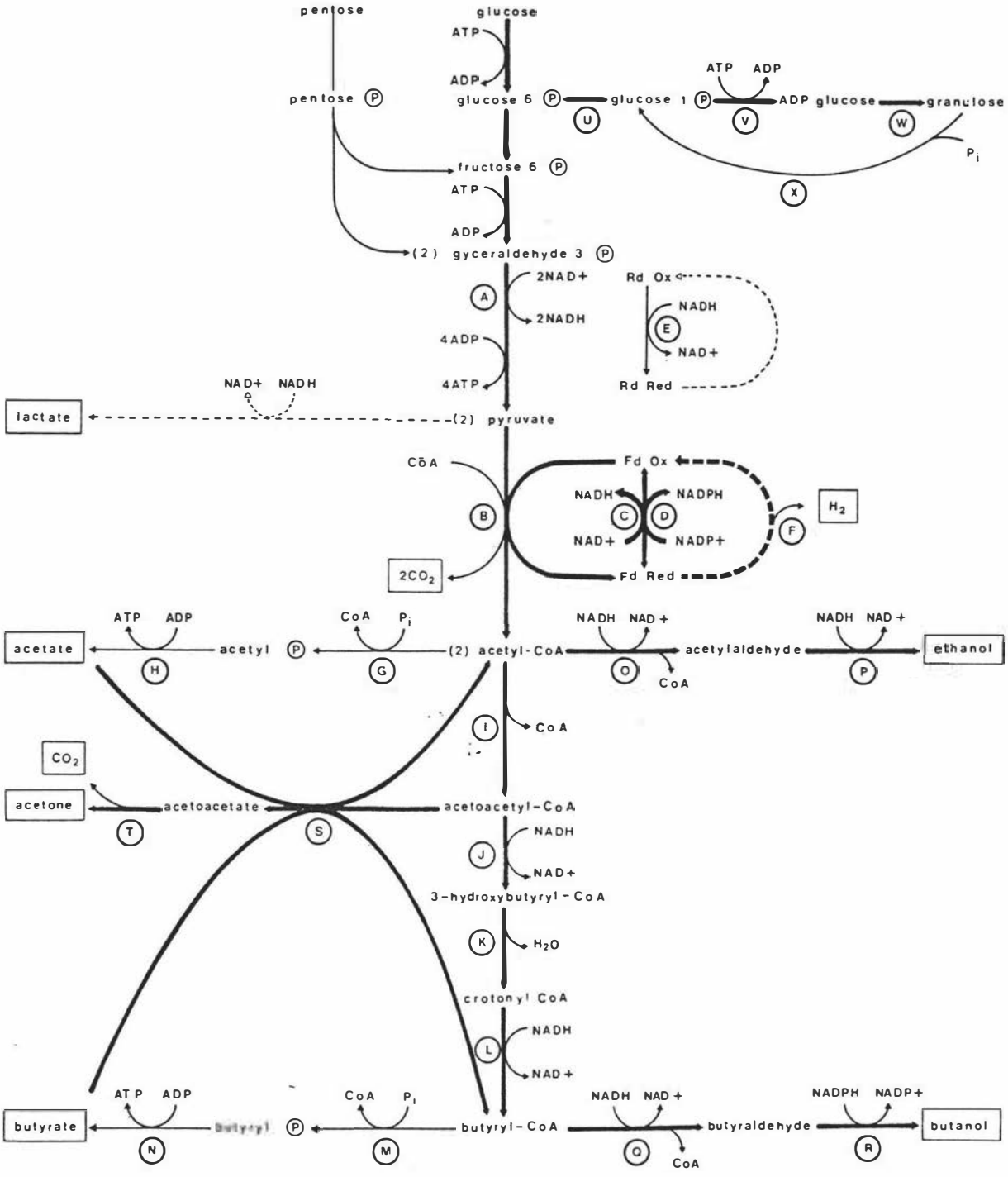
Enzymes are indicated by letters as follows:

(A) glyceraldehyde-3-phosphate dehydrogenase
(B) pyruvate-ferredoxin oxidoreductase
(C) NADH-ferredoxin oxidoreductase
(D) NADPH-ferredoxin oxidoreductase
(E) NADH-rubredoxin oxidoreductase
(F) hydrogenase
(G) phosphate acetyltransferase
(H) acetate kinase
(I) thiolase (acetyl-CoA acetyltransferse)
(J) 3-hydroxybutyryl-CoA dehydrogenase
(K) crotonase
(L) butyryl-CoA dehydrogenase
(M) phosphate butyryltransferase
(N) butyrate kinase
(O) acetaldehyde dehydrogenase
(P) ethanol dehydrogenase
(Q) butyraldehyde dehydrogenase
(R) butanol dehydrogenase
(S) acetoacetyl-CoA:acetate/butyrate-CoA transferase
(T) acetoacetate decarboxylase
(U) phosphoglucomutase
(V) ADP-glucose pyrophosphorylase
(W) granulose (glycogen) synthase
(X) granulose phosphorylase

(a)


pentose
glucose
ATP
ADP
pentose (P)
glucose 6 (P)
U
glucose 1 (P)
ATP ADP
V
ADP glucose
granulose
W
P_i
X
fructose 6 (P)
ATP
ADP
(2) gyceraldehyde 3 (P)
2NAD+
2NADH
A
Rd Ox
NADH
E
NAD+
Rd Red
4ADP
4ATP
NAD+ NADH
lactate
(2) pyruvate
CoA
Fd Ox
NADH
NADPH
C
D
NAD+
NADP+
Fd Red
B
F
H_2
$2CO_2$
ATP ADP
acetate
acetyl (P)
H
CoA P_i
G
(2) acetyl-CoA
NADH NAD+
O
acetylaldehyde
CoA
NADH NAD+
P
ethanol
I
CoA
CO_2
acetone
acetoacetate
T
S
acetoacetyl-CoA
NADH
J
NAD+
3-hydroxybutyryl-CoA
K
H_2O
crotonyl CoA
NADH
L
NAD+
ATP ADP
butyrate
butyryl (P)
N
CoA P_i
M
butyryl-CoA
NADH NAD+
Q
butyraldehyde
CoA
NADPH NADP+
R
butanol

(b)


pentose
glucose
ATP
ADP
pentose (P)
glucose 6 (P)
U
glucose 1 (P)
ATP ADP
V
ADP glucose
granulose
W
Pi
X
fructose 6 (P)
ATP
ADP
(2) gyceraldehyde 3 (P)
A
2NAD+
2NADH
Rd Ox
NADH
E
NAD+
Rd Red
4ADP
4ATP
NAD+ NADH
lactate
(2) pyruvate
CoA
B
Fd Ox
NADH
NADPH
C
D
NAD+
NADP+
Fd Red
F
H2
2CO2
ATP ADP
acetate
acetyl (P)
H
CoA Pi
G
(2) acetyl-CoA
NADH NAD+
O
acetylaldehyde
CoA
NADH NAD+
P
ethanol
I
CoA
CO2
acetone
acetoacetate
T
S
acetoacetyl-CoA
NADH
J
NAD+
3-hydroxybutyryl-CoA
K
H2O
crotonyl CoA
NADH
L
NAD+
ATP ADP
butyrate
butyryl (P)
N
CoA Pi
M
butyryl-CoA
NADH NAD+
Q
butyraldehyde
CoA
NADPH NADP+
R
butanol

-phosphate which are then further metabolised to pyruvate by entering the EMP pathway (Cynkin and Delwiche, 1958; Cynkin and Gibbs, 1958). The fermentation of 3 molecules of pentose yields 5 molecules of ATP and 5 molecules of reduced NADH.

Pyruvate resulting from glycolysis is oxidised by pyruvate-ferredoxin oxidoreductase in the presence of CoA, to yield acetyl-CoA, H_2 and CO_2, and is coupled with the production of NAD+ and the reduction of ferredoxin. The acetyl-CoA produced by this phosphoroclastic cleavage is the central intermediate in the branched fermentation pathways of C. acetobutylicum, from where it leads to the production of two other key intermediates, acetoacetyl-CoA and butyryl-CoA. These intermediates serve as precursors to all the products of both the acid- and solvent-producing pathways.

The production of butyrate occurs by the following steps. Two molecules of acetyl-CoA are converted to acetoacetyl-CoA, liberating CoA by acetyl-CoA acetyl transferase. Acetoacetyl-CoA is further converted, via the 3-hydroxybutryl-CoA and crotonyl-CoA intermediates to butyryl-CoA. Butyryl-CoA is converted to butyrate via butyryl phosphate resulting in the production of ATP and the release of CoA. Acetyl-CoA is converted to acetate via acetyl phosphate, enabling the production of ATP, and release of CoA. Both acetate and butyrate are excreted outside the cell.

1.3.2.2 Solvent producing pathways

The onset of solvent production involves a switch in the carbon flow from the acidogenic to the solventogenic pathways. During solvent production acetyl-CoA and butyryl-CoA function as key intermediates in the production of ethanol and butanol. Acetoacetyl-CoA is diverted from the production of butyrate and acetate, and is utilized via the transferase system, for the production of acetoacetate and thence acetone. This diversion of the original

system to production of acetone, prevents acid production, and eliminates the two steps which regenerated NAD+. By reducing the acids to alcohols, this regeneration process is replaced.

Reassimilation of the acids does not occur through a reversal of the acid-forming pathways (Hartmanis et al., 1984). Reassimilation of acetate and butyrate occurs via the acetoacetyl-CoA:acetate/butyrate-CoA transferase (Doelle, 1975, 1981; Andersch et al., 1983). The acetoacetate produced is decarboxylated by acetoacetate decarboxylase to produce acetone. This step is irreversible. The enzyme involved has broad carboxylic activity, catalyzing CoA transfer to either acetate or butyrate.

The majority of butyrate reassimilated is rapidly converted to butanol. Also, over 55% of the reassimilated acetate is converted to butanol, the remainder being converted to acetone via decarboxylation. Ethanol is also produced via acetyl CoA enabling regeneration of NAD+. The acetate reassimilated as acetyl CoA is converted to acetaldehyde by aldehyde dehydrogenase, and then to ethanol by ethanol dehydrogenase. This results in the release of CoA and the production of NAD+. The butyrate is reassimilated as butyryl-CoA which in turn is converted to butanol via butyraldehyde dehydrogenase and butanol dehydrogenase. This results in the release of CoA and the production of NAD+ and NADP+. The produced alcohols are then excreted from the cell.

1.3.3 Factors triggering solventogenesis

Knowledge of the factors involved in the switch from acidogenesis to solventogenesis is important to allow us to understand the way in which the production of solvents is initiated and maintained. Several studies have been performed to investigate these factors (Abou-Zeid et al., 1978; Baghlaf et al.,

1980; Bu'Lock and Bu'Lock, 1983; Long et al., 1984b). Experiments have shown that certain growth factors must be present at a minimum concentration, which allows sufficient growth and substrate consumption to provide the initial threshold of acids required for initiation of solvent production. However, it appears that there is no single growth-limiting nutrient whose exhaustion specifically induces solvent production (Bahl and Gottschalk, 1985). Solvent production is therefore believed to be due to a number of factors including culture pH value, cessation of growth, excess carbohydrate substrate and threshold concentrations of acids (Bahl et al., 1982a,b; Andersch et al., 1982; Martin et al., 1983; Monot et al., 1984; Hartmanis et al., 1984; Fond et al., 1985).

PART II: GENETICS OF C. ACETOBUTYLICUM

1.4 GENETICS OF C. ACETOBUTYLICUM

1.4.1 Taxonomy of the genus Clostridium

The genus Clostridium contains some of the most primitive eubacteria (Fox et al., 1980). By definition members of this genus are anaerobic, spore-forming rods, but beyond these common features there is a wide diversity in size, metabolic products, nutritional requirements and DNA G+C composition. Traditionally, Clostridium species have been classified by phenotypic characteristics such as substrate utilisation, optimum growth temperature, spore position, nutritional requirements, gelatin hydrolysis, fermentation products, motility and toxin production (Cato et al., 1986).

Table 1.2: Homology studies between Clostridium species (Cummins and Johnson, 1971)

Group	Spp.	Wall Sugars	Nutrient Requirements	G+C (%)	DNA Homology (%) I	II	III
I	18	glucose DAP	simple	27-28	72-100	17-47	-
II	20	glucose galactose DAP	complex	26-28	15-30	76-100	-
III	36	glucose glucose rhamnose mannose +/-DAP	complex	22-43	0-29	1-18	-

The results of early studies on the taxonomy of 'butyric acid' clostridial species by Cummins and Johnson (1971) were based on comparison of the cell wall composition, nutritional requirements, DNA homologies, and %G+C between various species. This resulted in the division of the species into three groups (Table 1.2). Cummins and Johnson (1971) reported great variation in cell wall composition, %G+C, and low DNA homologies among phenotypically similar butyric acid-producing clostridia, including bacteria of the same species. Investigation of four C. acetobutylicum strains identified similar % G+C ratios (28-29%), with similar cell wall structure consisting of meso-DAP, glutamic acid, rhamnose and mannose. DNA homology between the four strains was high (90-96%), but there was low homology to other clostridial species (7-10%). They were significantly different from other Clostridia such that they could not be placed in either of the two major homology groups and were designated to group III. This third group, differed widely among its own members in all categories. Cell wall structure and the type of isomeric form of linking between the diamino acids and sugars was found to correlate well with DNA homology (Cummins and Johnson, 1971).

Recently, more sensitive methods such as ribosomal RNA (rRNA) homology and oligonucleotide sequencing have been used for examining the relatedness of bacteria (Ludwig et al., 1985). Ribosomes are universally distributed, being present in all groups of organisms examined. They exhibit constancy of function and as their genes change in sequence very slowly, are conserved to a greater extent than other genes. This enables a more rigorous measure of genetic relatedness (Fox et al., 1980; Ludwig et al., 1985). Investigation of rRNA sequences has led to a clearer taxonomy in various genera, and has identified rRNA sequence similarities between organisms normally considered distantly related and representatives of diverse bacterial groups (Moore and McCarthy, 1967; Johnson et al., 1970; Pace and Campbell, 1971).

Table 1.3: Classification of Clostridium species (Johnson and Frances, 1975)

Group	%G+C	No. of Species	Comments	DNA Homology (%) I	II	III	IV
I	28	30	Well defined 10 subgroups	50+	40	0	0
II	28	11		40	69+	0	0
III	27-29	6		0-31	13-17	0	0
IV	41-45	5	similar to Bacillus spp.	0	0 (0-14%)	0	0

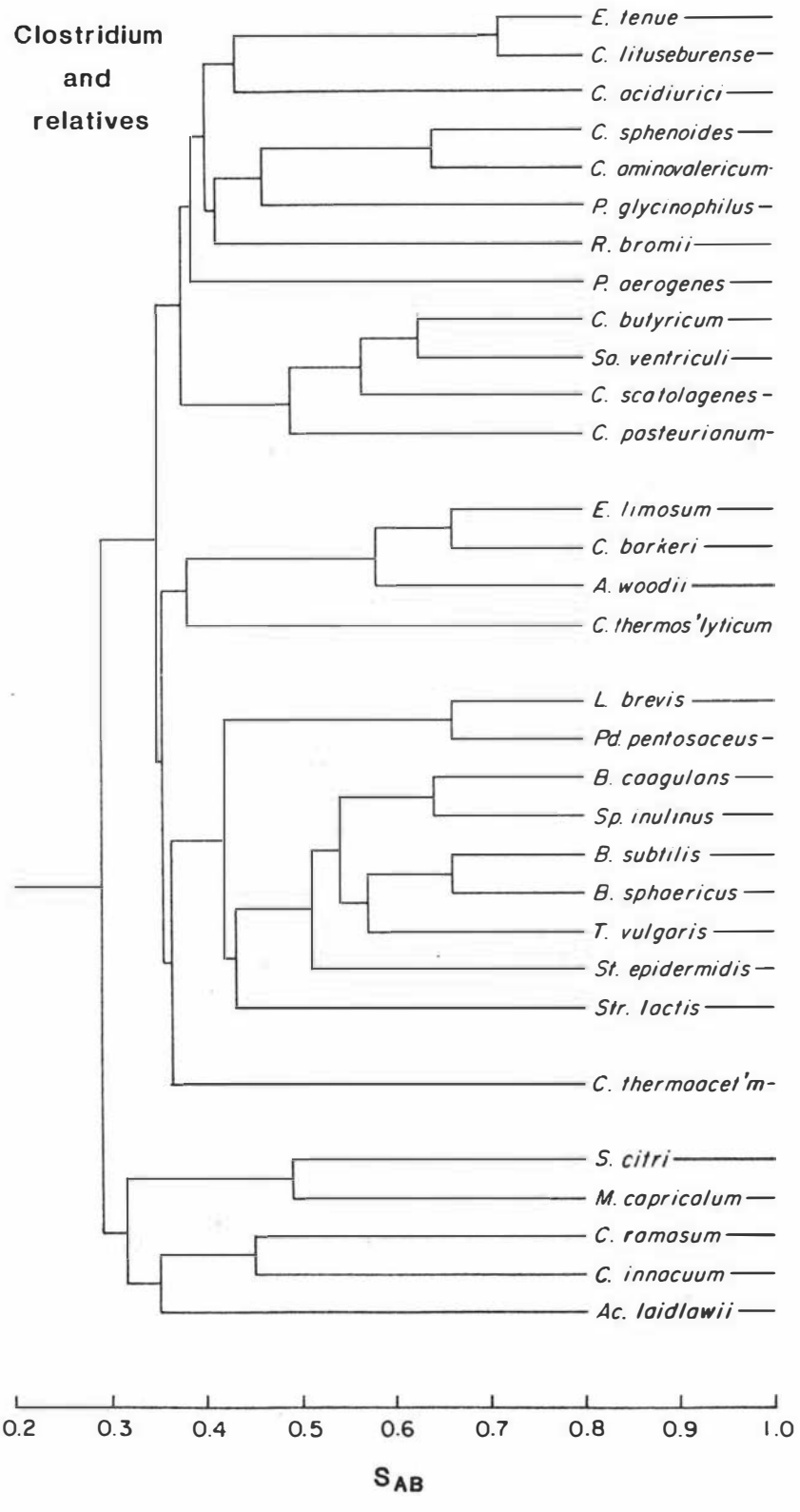
Johnson and Frances (1975) used 23S rRNA homology among 56 Clostridium species to establish more detailed groupings within this genus. The species were assigned to four groups (Table 1.3). The authors concluded that there were three main homology groups (%G+C 27-29%), plus one minor group (%G+C 41-45%). Group I was the largest of the groups, and species were subdivided into 10 related subgroups. Only one strain of C. acetobutylicum (ATCC 824) was investigated and this was placed in Group I. Species which fell into rRNA homology groups I and II shared a moderate degree of homology among themselves, but no homology to Group III or IV. Group III members shared little or no homology to reference strains of Groups I and II, and shared little homology between themselves. The species in Group IV had a high %G+C, similar to that of Bacillus species. They shared more homology to Bacillus species than any species of Groups I, II, III or among themselves.

Cell shape and size correlated very well with these rRNA homology results. Motility and nutritional requirements correlated reasonably well with rRNA homology values, although diversity existed in some groups. In contrast, characteristics such as spore position, utilization of individual carbohydrates,

Fig.1.3: Phylogenetic relationship between Clostridium and other Gram-positive bacteria (Fox et al., 1980).

Generic Abbreviations

A	Acetobacterium
Ac	Acholeplasma
B	Bacillus
C	Clostridium
E	Eubacterium
L	Lactobacillus
M	Mycoplasma
P	Peptococcus
Pd	Pediococcus
R	Ruminococcus
S	Spiroplasma
Sa	Sarcina
Sp	Sporolactobacillus
St	Staphylococcus
Str	Streptococcus
T	Thermactinomyces


Clostridium
and
relatives
E. tenue
C. lituseburense
C. acidiurici
C. sphenoides
C. aminovolericum
P. glycinophilus
R. bromii
P. aerogenes
C. butyricum
Sa. ventriculi
C. scatologenes
C. posteurionum
E. limosum
C. barkeri
A. woodii
C. thermos'lyticum
L. brevis
Pd. pentosaceus
B. coagulans
Sp. inulinus
B. subtilis
B. sphaericus
T. vulgaris
St. epidermidis
Str. lactis
C. thermoacet'm
S. citri
M. capricolum
C. ramosum
C. innocuum
Ac. laidlawii
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
SAB

gelatin liquefaction and organic acid fermentation correlated very poorly with rRNA homology (Mead, 1971).

Fox et al. (1980) investigated the homology of some prokaryotes, including Clostridium spp, using 16S rRNA (Fig. 1.3). These results support the above work by Johnson and Frances (1975), that many Clostridium species are more closely related to other Gram-positive bacteria than to each other. However, the groupings given do not always agree with those given by Johnson and Frances (1975). Results show that the genus Clostridium is an ancient and very diverse group of bacteria. A lot of work has still to be done to place Clostridium into a better classification. Cato et al. (1986) suggested that the genus may eventually be divided into at least two separate genera, possessing 22-34% G+C and 40-55% G+C, respectively.

1.4.2 Genetics and strain improvement

Fundamental studies on the molecular genetics of anaerobes is an essential prerequisite for enhancing the understanding of the biochemistry, physiology and genetics of the anaerobic bacteria. Recent interest in the genetics of anaerobes stems from the medical or industrial viewpoint, due to the emergence of antibiotic strains involved in disease, and strains which can utilize diverse substrates eventually producing useful products (Woods, 1982; Woods and Jones, 1984).

It has long been recognised that the performance of strains effects the efficiency of industrial fermentations. The use of genetic techniques to produce strains of C. acetobutylicum with higher butanol production, increased butanol tolerance or more diverse substrate potential has often been suggested by various authors (Walton and Martin, 1979; Volesky et al., 1981; Blaschek, 1986). However, there is a need to establish a working system for the genetic manipulation of C. acetobutylicum and other anaerobic bacteria before such improvements are

possible. Technical difficulties in working with anaerobes, as well as the need to adapt available genetic techniques to these genera have hampered progress (Woods, 1982; Woods and Jones, 1984).

1.4.2.1 Mutagenesis

A useful tool in the study of the genetics and regulation of bacteria is mutagenesis. Mutations in bacteria can be induced by mutagens, which can be divided into two classes. Indirect mutagens act by inducing a postreplication repair system (SOS processing) which is prone to errors. In E. coli this system is dependant upon products of such genes as recA, lexA and umuC. Examples of such mutagens are UV irradiation and mitomycin C. The second class, called direct mutagens, cause mutations by mispairing mechanisms. This results in misreplication of template DNA or nucleotide precursors. They operate mostly independently of the complex system of SOS processing. Examples are N-methyl-N'-nitro-N -nitrosoguanadine (MNNG) and ethyl methanesulphonate (EMS) (Miller, 1983; Walker, 1984).

Droffner and Yamamoto (1983) reported that chromosomal auxotrophic mutants of Clostridium species were difficult to obtain. However, mutants have been produced and employed in physiological studies from such species as C. perfringens (Mendez and Gomez, 1982; Sebald, 1982); C. pasteurianum (Robson et al., 1974); C. histolyticum (Sebald, 1982); C. thermoaceticum (Schwartz and Keller, 1982), C. saccharolyticum, (Murray et al., 1983).

1.4.2.2 Mutagenesis of C. acetobutylicum

Mutants of C. acetobutylicum have been obtained both spontaneously and by the use of various mutagens. These mutants effect a wide variety of morphological, biochemical and physiological characteristics and have proved useful in studying these characteristics (Table 1.4).

Table 1.4: Mutants of C. acetobutylicum

D/I	Mutagen	Efficiency	Characteristics	Reference
D	EMS	+	autolysin -	Allcock et al. (1981)
I	UV	-		
D	EMS	+	Rif^R, spo - His-, Glut-, Pro-, Trp- granulose - solvent-	Jones et al. (1982)
spontaneous		+	granulose - spo -	Meinecke et al. (1984)
D	EMS	+	Ab^R, spo - capsule - granulose- solvent -	Long et al. (1984)
D	EMS	+	Arg-, Tyr-, His-, Met-	Jones et al. (1985)
D	MNNG	+	Ade-, Uracil- riboflavin -	Bowring and Morris (1985)
D	EMS	+	Rif^R	
I	UV	-		
I	mitomycin C	-		
I	nalidixic acid	-		
I	Metroni- dazole	-		
I	H_20_2	-		
D	EMS	+	Ap^R, Em^R Butanol R	Lemmel (1985)
D	MNNG	-		
I	UV	-		
I	acridine half mustard	-		

Table 1.4 cont'd:

D/I	Mutagen	Efficiency	Characteristics	Reference
D	MNNG	+/-	Butanol ^{R} spo -	Hermann et al. (1985)
D	EMS	-		
D	EMS	+	Xylose ++	Lee et al. (1985b)
D	EMS + spontaneous +		Sm^{R}, Rif^{R} Allyl alcohol^{R}	Rogers (1986)
D	EMS + spontaneous +		Acid - spo-, solvent- Allyl alcohol^{R}	Rogers and Palosaari (1987)
D	EMS	+	autolysin decrease	Reysset et al. (1987)

D - direct mutagens I - indirect mutagens
(+) - efficient (-) - inefficient

Direct mutagens were efficient in the production of various phenotypic mutants of C. acetobutylicum strains. It was noted by Bowring and Morris (1985) that the use of MNNG resulted in the isolation of certain auxotrophic lesions at a greater frequency than others. The direct mutagen, EMS, appears to be the mutagen of choice, producing a reproducible array of mutants. However, under some conditions, EMS and MNNG were ineffective in producing certain mutants (Lemmel, 1985; Hermann et al., 1985).

In contrast, indirect mutagens such as UV and mitomycin C, whose efficiency relies on the misrepair of damaged DNA via the error-prone pathway, were ineffective in producing mutants. They also had no significant effect on the viability of the cells (Allcock et al., 1981; Bowring and Morris, 1985). This

apparent inefficiency of indirect mutagens on C. acetobutylicum is commonly observed in the clostridia, (Mendez and Gomez, 1982; Walker, 1983). This may be due to C. acetobutylicum having either a very efficient repair mechanism, or no error-prone system at all.

1.4.3 Plasmids

Several species of the genus Clostridium have been shown to contain plasmids. The plasmids best characterised are mainly those of the important pathogenic species such as C. perfringens, C. difficile and C. tetani. In addition, there have been investigations over a range of other Clostridium species (Ionesco et al., 1975; Brefort et al., 1977; Mihelc et al., 1978; Duncan et al., 1978; Rood et al., 1978; Lee et al., 1987; Blaschek and Solberg, 1981). Some of these plasmids have been identified as being responsible for a variety of phenotypic characteristics including α-toxin, bacteriocin and caseinase production, antibiotic resistance and degradation of organomercurial compounds (Duncan et al., 1978; Rood et al., 1978; Mihelc et al., 1978; Lee et al., 1987; Laird et al., 1980; Pan Hou et al., 1980; Blaschek and Solberg, 1981). Most plasmids however, are still cryptic, as no known function can be ascribed to their presence.

1.4.3.1 Plasmids in C. acetobutylicum

Plasmids have been detected in several strains of C. acetobutylicum. However, their size and number vary and no function has been correlated with their presence in the various strains. Truffaut and Sebald (1983) investigated a wide range of acetone- and butanol-producing strains of C. acetobutylicum and related strains. Investigation of the 17 strains of C. acetobutylicum identified five strains which contained plasmids (Table 2.5). In the remaining 12 strains, no plasmid DNA could be detected as bands on an agarose gel. Lee et al. (1987) investigated plasmid content in six strains of C. acetobutylicum. Of these, four contained plasmids (Table 1.5).

Table 1.5: Plasmids detected in C. acetobutylicum

Strain	No.	Size	Reference
NRRL598	1	5.2 Md	Truffaut and Sebald
NCIB2951	2	6.8, 45 Md	(1983)
NRRL6442	2	6.8, 45 Md	
NCIB6443	1	8 Md	
NCIB6444	3	4, 5.5, 8-10 Md	
DSM1732	1	12.5 kb	Lee et al., (1987)
DSM1733	2	11.5, 12.5 kb	
DSM1737	1	12.5 kb	
DSM1738	2	11.5, 12.5 kb	

1.4.3.2 Construction of shuttle vectors using plasmids from Clostridium species

The essential requirement and obvious lack of suitable vectors for gene transfer and transformation is a major obstacle in the development of genetics of C. acetobutylicum. Clostridium plasmid cloning vectors should replicate in and contain selective markers for E. coli, as at present, genetic and biochemical experiments of higher sophistication can be performed in E. coli compared to Clostridium.

Attempts have been made to construct shuttle vectors from native plasmids of C. butyricum (Collins et al., 1985; Luczak et al., 1985) C. perfringens, (Squires et al., 1984) C. acetobutylicum and C. butyricum (Truffaut and Sebald, 1988). These are summarized in Table 1.6.

Squires et al. (1984) constructed a series of chimeric shuttle vectors from cryptic plasmids of C. perfringens. Each of the vectors was stably maintained in E. coli conferring Ap^R.

Table 1.6: Summary of Clostridium based shuttle vectors that have been transferred into Gram-positive bacteria

Plasmids	Vector	Shuttle vector	Maintenance/Stability	
(Squires et al., 1984): C. perfringens			E. coli	C. perfringens
pJU121	pBR322	pJU3	+	ND[a]
pJU122	pBR322	pJU5	+	ND
pJU281	pBR322	pJU1	+	ND
pJU1	(pCW3/	pJU7	+	-
	pJU124)	pJU10	+	-
pJU3	Tc^R	pJU12	+	+
JU5		pJU13	+	+
		pJU16	+	+
(Collins et al., 1985): C. butyricum			E. coli	B. subtilis
pCB101	pJAB1	pRB1	+	+
		pRB5	+	+ (chromosomal integration)
pCB102	pJAB1	pRB7	-	+
(Truffaut and Sebald, 1988): C. acetobutylicum			E. coli	B. subtilis
pIP413	pVH33	pIP4131	+	+/unstable
C. butyricum				
pIP414	pHV33	pIP4141	+	+/+
		pIP4144	+	+/unstable
pIP4142	pHV33	pIP4142	+	-

[a] ND= Not done

Introduction of the Tc^R gene from two native C. perfringens plasmids into these vectors, and subsequent transformation into E. coli, resulted in expression of the Tc^R in E. coli. When introduced into C. perfringens, three of the five examined were stably maintained, while the others were not.

Collins et al. (1985) constructed a series of shuttle vectors from two cryptic plasmids of C. butyricum. Of the 30 formed, three were maintained in E. coli and B. subtilis strains. Plasmid pRB1 could be successfully transferred back from B. subtilis to E. coli. Truffaut and Sebald (1988) constructed three shuttle vectors using cryptic plasmids of C. acetobutylicum and C. butyricum. Only one was stable in B. subtilis. These results show that shuttle vectors are often unstable when placed in bacterial strains other than their native strain. Since none of the shuttle vectors were transferred to C. acetobutylicum, their stability in, or usefulness as, shuttle vectors for a genetic system in C. acetobutylicum is unknown.

1.4.4 Bacteriophages

The use of bacteriophages in E. coli and B. subtilis cloning systems are well established and useful tools. The presence of bacteriophages has been reported in many clostridial species, including C. acetobutylicum (Eklund et al., 1971; Ogata and Hongo, 1979). They are the causative agents of sluggish, abnormal fermentations in industrial ABE processes.

The feasibility of developing bacteriophage DNA as a possible vector for transfer of recombinant DNA plasmids into C. acetobutylicum P262 was investigated by Reid et al. (1983). Protoplasts of C. acetobutylicum were transfected by bacteriophage CA1 DNA. They reported a requirement for a long (2 hr) incubation of protoplast and DNA at 37°C for successful transfection, and observed that the rate of adsorption of bacteriophage CA1 to protoplasts was slower than the rate of adsorption to cells, although the

maximum overall amount (80- 86%) was similar.

However, protoplasts infected with bacteriophage DNA did not release mature bacteriophage particles spontaneously. Instead the protoplasts had to be regenerated on regeneration medium for 24-26 hr before phage particles were observed. The results indicated that the DNA of bacteriophage CA1 is taken up and expressed by C. acetobutylicum protoplasts that are allowed to regenerate. This successful transfection of C. acetobutylicum with phage DNA means that the phage could be developed as a potential vector for the transfer of recombinant plasmids into C. acetobutylicum.

Despite these developments, the use of these bacteriophages as cloning vectors is probably not feasible, mainly because of the evolution of copious quantities of gases, and the fact that C. acetobutylicum bacteriophages have been shown to have only a narrow host range and to infect only some strains of the same species with the same fermentation ability (Ogata and Hongo, 1979). Transduction has not as yet been reported in C. acetobutylicum or other solvent-producing strains.

1.4.5 Genetic transfer systems in C. acetobutylicum

1.4.5.1 Protoplast regeneration

Transformation of protoplasts is dependant upon the development of methods for the production and regeneration of protoplasts. These have been reported for several Gram-positive bacteria such as C. pasteurianum (Minton and Morris, 1983), C. saccharroperbutylacetonicum (Yoshino et al., 1984) and C. acetobutylicum (Allcock et al., 1982; Jones et al., 1985; Reysset et al., 1987).

In an attempt to develop a transformation system using protoplasts, Allcock et al. (1982) investigated protoplast formation and regeneration in C. acetobutylicum. The induction of autolysis and protoplast formation was due to the induction of a cell-bound autolysin. They showed that high concentrations of sucrose alone induced the production of the autolysin, and that when sucrose-induced cells were maintained in an osmotically stabilized medium, between 70 and 85% stable protoplasts could be obtained. A number of other chemical agents were also identified that were able to induce autolysins with varying degrees of effectiveness. The addition of lysozyme increased both the rate and the final percentage of protoplasts obtained. As the presence of autolysin was found to affect the regeneration of protoplasts, autolysin-deficient mutants were isolated from C. acetobutylicum (Allcock et al., 1981). One of these mutants, lyt-1 was utilized in the regeneration studies. This mutant produced less autolysin than the parent strain. Jones et al. (1985) obtained an optimum protoplast production using mid-exponential phase cells, with high regeneration frequencies (in excess of 80%). This frequency of regeneration is considerably higher than that obtained with the other clostridial strain investigated (C. pasteurianum) where only 10% was achieved (Minton and Morris, 1983).

Reysset et al. (1987) investigated protoplast production and regeneration in C. acetobutylicum NI-4080. They reported that with this strain, autoplast production using sucrose and lysozyme was achieved at a rate of 99%. This efficiency of autoplast production was related to the growth stage, being maximum at the end of the exponential phase. Protoplast production using sucrose, lysozyme and penicillin was achieved at a rate close to 100%, being more efficient at the early exponential phase. However, irreproducible regeneration of protoplasts was obtained, at rates of lower than 0.01%, and a large number of L-form colonies were produced. Various treatments to increase regeneration were tried, such

as the presence of N-acetyl glucosamine, D-alanine, D-glutamate and gelatin, but these proved unsuccessful. This low level of regeneration was due to the presence of autolysin from the L-forming cells in the population which hampered the regeneration of the protoplasts. When an autolysin-decreased mutant was produced (NI-4081), protoplast regeneration frequencies increased to about 1%. The use of xylose instead of sucrose in the regeneration medium reduced the regeneration time required, and also inhibited the formation of the L-form colonies, although the final number of regenerants obtained was the same.

1.4.5.2 Protoplast fusion

PEG-induced fusion of bacterial protoplasts has been utilized for studies into fundamental and applied genetics of other bacteria (Hopwood, 1981). Demonstration of protoplast fusion and isolation of recombinants in C. acetobutylicum P262 is a potential help for fundamental genetic mapping studies and in obtaining hybrid strains with suitable industrial traits. It may also facilitate interspecific gene and plasmid transfer in C. acetobutylicum by protecting plasmids from the high levels of extracellular DNase produced by strains of C. acetobutylicum.

Jones et al. (1985) investigated the production of recombinants using protoplast fusion in C. acetobutylicum P262. Analysis of the regenerated colonies after protoplast fusion showed the presence of three groups: stable parental colonies, stable recombinant colonies and segregating biparentals. Parental colonies represented the majority, with one of the parental types (Arg-, Tyr-) having a persistently higher regeneration frequency than the other parental type (His-, Met-). Stable recombinant colonies accounted for 0.3-2% of the colonies and possessed a mixture of the auxotrophic markers.

The diploid biparentals represented 1.4-8.5% of the colonies. The stability of the biparental state was varying, 30-40% retained the biparental trait after subcloning. Continued segregation of both parental genomes during growth was therefore observed with the concurrent production of recombinants. The biparental trait was not retained after sporulation. The segregating biparentals gave progeny of both parental types. These were subdivided into prototrophic complementating biparentals, partially complementating biparentals and non-complementating biparentals (either one parent or neither). Non-complementating biparentals had a phenotype of one of the parental strains (single parent non-complementating biparentals) or expressed neither of the parental phenotypes (zero non--complementating biparentals). These partially complementating and zero non-complementating biparentals had not been reported earlier. All classes segregated to produce parentals, stable recombinants and occasionally biparentals with other phenotypes. C. acetobutylicum is therefore capable of undergoing homologous recombination at frequencies similar to those of aerobic species such as B. subtilis. This may facilitate interspecific gene and plasmid transfer in C. acetobutylicum (Jones et al., 1985).

1.4.5.3 Protoplast transformation

The transformation of C. acetobutylicum cells by chromosomal or plasmid DNA via natural competence or treatments to facilitate DNA uptake by whole cells has not been successful in C. acetobutylicum (Jones and Woods, 1986a). In contrast, protoplast transformation has been successfully used for the transfer of plasmid and bacteriophage DNA into C. acetobutylicum protoplasts (Reid et al., 1983; Lin and Blaschek, 1984).

Lin and Blaschek (1984) used methods which decreased the extracellular DNase activity associated with protoplasts to obtain successful

transformation. Heat treatment (55°C, 15 min) of C. acetobutylicum SA1 allowed significant inactivation of the extracellular DNase which markedly interferes with transformation. However, the heat treatment impaired the cell wall regeneration capability, decreasing regeneration 1000-fold. They reported the successful transformation of the plasmid pUB110 from B. subtilis into C. acetobutylicum resulting in isolation of Km^R transformants. The frequency achieved was 10^{-4} transformants per viable regenerated protoplast (28 transformants/μg of pUB110 DNA). Without the inactivation of the DNase, no transformants were isolated. Prolonged incubation of the transformation mixture did not improve the transformation frequency. The efficiency per μg of DNA, however, was extremely low. Detection of pUB110 DNA was difficult, requiring the use of diethyl pyrocarbonate (DEP) in an attempt to inactivate the DNase activity. One can conclude that transformation of C. acetobutylicum is a complex and difficult process, and that no simple routine procedure is available for getting DNA into C. acetobutylicum cells at high efficiency.

1.4.5.4 Conjugation in C. acetobutylicum

Conjugation between bacteria is dependent upon the presence of a transfer (Tra) gene region on plasmids, enabling self-transfer. Some self-mobilisable plasmids can also mobilise other plasmids which lack the Tra gene region but which have compatible mob, and oriT sites. Plasmids in this category which have been used to investigate conjugation in C. acetobutylicum are pAMB1 and pVA797. The plasmid pAMB1 (26.5 kb) confers resistance to macrolides, lincosamides and streptogamin B upon the host cell (Clewell et al., 1974; LeBlanc and Lee, 1984). It is a broad host range, conjugal self-transfer plasmid derived from the Gram-positive anaerobe S. faecalis, and has the ability to undergo conjugal transfer to a number of other species such as Streptococcus spp., Lactobacillus spp., S. aureus, and Bacillus species.

The plasmid pVA797 is also a conjugative plasmid isolated from Streptococcus agalactiae. It can replicate in S. sanguis and E. coli (Macrina et al., 1982; Evans and Macrina, 1983). It codes for erythromycin resistance and has been used to mobilise non-conjugative plasmids into other Streptococcus strains.

Oultram and Young (1985) investigated the conjugal transfer of pAMß1 into C. acetobutylicum. They reported that pAMB1 could be transferred from S. lactis to C. acetobutylicum at high efficiency (10^{-4} to 10^{-5}). The plasmid was stably maintained in C. acetobutylicum and was transferable back to S. lactis at lower frequencies (10^{-5} to 10^{-7}) than that observed from S. lactis to C. acetobutylicum. Transfer of pAMB1 between various strains of C. acetobutylicum was also observed at frequencies comparable to that for transfer between C. acetobutylicum and S. lactis. Transfer of pAMB1 between B. subtilis and C. acetobutylicum was also possible (10^{-6} to 10^{-7}).

Reysset and Sebald (1985) reported the transfer and stable maintenance of the broad host range plasmids pAMB1, pIP501 and pJH41 from S. faecalis to C. acetobutylicum. Frequencies obtained ranged from 10^{-5} (pAMB1, pIP501) to 10^{-7} (pJH4). Sm^R, Km^R, and Er^R from all three plasmids were fully expressed at high levels in C. acetobutylicum. The Cm determinant of pIP501 however, was poorly expressed, giving exconjugants with only a slight, though reproducible resistance to chloramphenicol. C. acetobutylicum could retransfer the plasmid to Rif^R strains of C. acetobutylicum at frequencies of 10^{-5} to 10^{-6}.

Yu and Pearce (1986) further demonstrated that conjugal transfer was possible using pAMB1 and pVA797. Generally, all three donor species (S. faecalis, S. sanguis and S. lactis) transferred pAMB1 and pVA797 with comparable efficiency into C. acetobutylicum (10^{-4} to 10^{-7}). C. acetobutylicum carrying pAMB1 or pVA797 was also able to retransfer the

plasmids to S. lactis at comparable frequencies (10^{-4} to 10^{-5}) to the original transfer. Transfer of pVA79::Tn917 from S. faecalis occurred at essentially the same frequency (10^{-4}). Tn917, which codes for kanamycin resistance, conferred this resistance upon C. acetobutylicum. The expression of this transposon has the potential to be used for the study of structural and regulator genes of C. acetobutylicum, as reported in the study of the developmentally regulated genes of B. subtilis (Youngman et al., 1985).

The ability of self-conjugable plasmids pAMB1 and pVA797 to co-mobilise the non-conjugative plasmid pAM610 into C. acetobutylicum was also investigated by Yu and Pearce (1986). Plasmid pAM610 is a 9.5 kb plasmid encoding for resistance to tetracycline and kanamycin. It has a single unique cloning site and has potential as a vector for the construction of a shuttle vector system for gene cloning and transfer into C. acetobutylicum. However, although pAM610 co-transferred, the frequency was low (10^{-6} to 10^{-7}). Hence, the efficiency of conjugational transfer is too low to enable cloning of genes en masse into C. acetobutylicum.

For this reason, the cloning of C. acetobutylicum genes into well known hosts such as E. coli and B. subtilis is an alternative method which has been utilised by various groups (Efstathiou & Truffaut, 1986; Zappe et al., 1986, 1987; Usdin et al., 1986; Youngleson et al., 1988).

1.5 CHARACTERIZATION OF VARIOUS C. ACETOBUTYLICUM GENES

1.5.1 Introduction

Cloning and characterization of C. acetobutylicum genes in C. acetobutylicum is not possible at present due to difficulties in transferring DNA into C. acetobutylicum. The alternative approach of cloning and study of genes from C.

acetobutylicum in other bacteria which are well characterised, such as E. coli and B. subtilis, has many advantages. There is the availability of many genetically identified mutants which will enable the analysis and characterization of important structural and regulatory genes by standard methods such as complementation. Study of gene functions by site-directed mutagenesis is also possible. The ease of transfer of DNA into these well characterized bacteria is a great advantage.

Using other bacteria relies on the expression and stability of C. acetobutylicum genes (Gram-positive anaerobe) in a different host system, such as in the Gram-negative aerobe (E. coli) and Gram-positive aerobe (B. subtilis). This possibility is supported by evidence of cloned genes from other clostridial species expressing in E. coli. These include the α-isopropylmalate dehydrogenase and hydrogenase genes of C. butyricum (Karube et al., 1983; Ishii et al., 1983), the ferredoxin gene from C. pasteurianum (Graves et al., 1985, Daldal and Applebaum, 1985), and the endoglucanase, cellobiase and cellobiohydrolase genes from C. thermocellum (Cornet et al., 1983a, b; Millet et al., 1985).

1.5.2 Cloning of genes from C. acetobutylicum

It is only in the past few years that reports have appeared on the cloning of genes from C. acetobutylicum (Table 1.7). The first reported cloning of genes from C. acetobutylicum was by Efstathiou and Truffaut (1986). A genomic library to strain ABKn8 was constructed using the vector pBR322, and transformed into host E. coli HB101. The recombinant plasmids were able to complement the HB101 proA2, and leuB6 auxotrophic mutations. Complementation of the leuB6 mutation of B. subtilis was also obtained. The results showed that genes from C. acetobutylicum were able to express and complement the defects present in the genetic system of two different hosts, E. coli and B. subtilis.

Table 1.7: Current list of genes cloned from C. acetobutylicum

Gene	Host	Reference
proA2	E. coli	Efstathiou and Truffaut (1986)
leuB6	E. coli B. subtilis	
arg-	E. coli	Zappe et al., (1986)
endoglucanase		
cellobiase		
his-		
xylanase	E. coli	Zappe et al., (1987)
glutamine synthetase	E. coli	Usdin et al., (1987)
alcohol dehydrogenase	E. coli	Youngleson et al. (1988)

As the first step to understanding and improving cellulolytic activity of C. acetobutylicum, Zappe et al. (1986) cloned and characterized two cellulase genes from C. acetobutylicum P262. They constructed a genomic library to strain P262 in E. coli, using the vector pEcoR251. Complementation of E. coli Arg- and His- mutants supported the results of Efstathiou and Truffaut (1986) that expression of C. acetobutylicum genes in a Gram-negative host such as E. coli was possible. A plasmid (designated pHZ100) was isolated which possessed endoglucanase and cellobiase activity, carried on a 4.9 kb insert.

The conditions for optimum activity of the endoglucanase from C. acetobutylicum P270 and E. coli HB101 (pHZ100) differed. The optimum pH and temperature for P270 endoglucanase activity were pH 4.6 and 37°C respectively. These differed from the enzyme extracts of E. coli (pHZ100) which had an optimum of 50°C at pH 5-7. The endoglucanase activity from C. acetobutylicum cloned in

E. coli pHZ100 did not require molasses for induction and was expressed constitutively. This is in contrast to that observed in C. acetobutylicum P270, where the enzyme was inducible, requiring the presence of a small molecule in molasses for induction (Allcock and Woods, 1981). This apparent loss of regulation observed here has been noted in many Gram-positive genes cloned into E. coli. The endoglucanase activity occurred mainly in the periplasmic fraction in E. coli cells containing the cloned gene. This is in contrast to the endoglucanase activity in C. acetobutylicum P270, which is an extracellular enzyme (Lee et al., 1985a). Zappe et al.,(1988) subsequently reported the sequencing of the fragment containing the endo-glucanase gene.

Zappe et al. (1987) cloned the xylanase gene from C. acetobutylicum P262, utilizing the previously made pEcoR251 genomic library. A recombinant plasmid containing a 5.4 kb fragment from C. acetobutylicum genomic DNA (pHZ300) was shown to contain xylanase activity by a zone of hydrolysis on xylan plates. Experiments verified that 98% of the xylanase activity was present in the cytoplasm of E. coli cells, with no appreciable activity detectable in the extracellular culture medium. This is in contrast to C. acetobutylicum xylanase activity, which is an extracellular enzyme (Lee et al., 1985).

The effects of pH and temperature on the xylanase purified from E. coli HB101 (pHZ300) and C. acetobutylicum were essentially similar. Results showed maximum activity between pH 5.5 - 6.5, at 37-43°C. This is in contrast to the results obtained with the endoglucanase gene (Zappe et al., 1986). The cloned insert was shown to code for the production of a 28 kDa protein, representing the xylanase protein. A similar sized protein was observed from a partially purified preparation of C. acetobutylicum culture medium.

Usdin et al. (1986) reported the cloning and characterization of glutamine synthetase (glnA) gene from C. acetobutylicum P262, utilizing the previously made pEcoR251 genomic library. The recombinant plasmid clone in the

genomic library coding for the glnA gene was identified by complementation of E. coli ET8051 deleted for glnA-, ntrB- and ntrC-. The plasmid (pHZ200), containing the cloned C. acetobutylicum glnA gene, was capable of directing the synthesis of an active glutamine synthetase. However, the C. acetobutylicum glnA gene had no detectable DNA homology to the E. coli glnA gene.

The plasmid pHZ200 coded for the production of a major protein of 59 kDa, expressed from the glnA gene. The C. acetobutylicum GlnA was not subject to regulation by an adenylylation-deadenylylation system as is the case in most Gram-negative bacteria (Magasanik, 1982), but was similar to that observed in the Gram-positive B. subtilis where the gene product is autoregulatory (Schreier et al., 1985; Gardner and Aronson, 1984). The C. acetobutylicum glnA gene in E. coli expressed high levels of GlnA in the presence of excess nitrogen, whereas the E. coli gene is repressed under identical conditions. The glnA gene of C. acetobutylicum was subject by excess glutamine to nitrogen regulation in E. coli. Under conditions of nitrogen limitation the activity of the C. acetobutylicum glutamine synthetase in E. coli increased five fold yielding extremely high levels of activity. However, the cloned C. acetobutylicum glnA gene was unable to complement certain nitrogen regulatory gene functions in E. coli ntrB and ntrC deletion strains. Janssen et al. (1988) investigated further the structure and regulation of the cloned glnA gene of C. acetobutylicum by sequencing.

Youngleson et al. (1988) cloned and characterized an alcohol dehydrogenase (adh) gene from C. acetobutylicum P262. Recombinant plasmids from the previously constructed genomic library in pEcoR251 were screened for alcohol dehydrogenase (ADH) activity, enabling identification of one clone (pCADH100) which exhibited increased sensitivity to the lethal suicide substrate allyl alcohol when compared to wild type E. coli. Allyl alcohol is converted by ADH to a lethal alkylating aldehyde. This recombinant plasmid (containing a 3.2 kb C. acetobutylicum insert DNA) was transformed into E. coli HB101-adh1, a mutant of HB101 resistant to 200mM allyl alcohol. The presence of pCADH100 in this

strain caused inhibition by only 10mM allyl alcohol. The C. acetobutylicum adh gene showed no significant DNA homology to the E. coli adh gene.

E. coli HB101/pCADH100 showed a 64% increase in ethanol production and a proportionate decrease in acetate production compared to wild type E. coli which usually produces equimolar amounts of ethanol and acetate in anaerobic conditions. E. coli HB101-adh1/pCADH100 showed an increase of 116% in ethanol production, whilst there was no change in the level of acetate production, which remained at the same level as in E. coli HB101-adh1. HB101-adh1/ pCADH100 showed a 2-4 fold increase in alcohol dehydrogenase activity when compared to HB101/pCADH100. E. coli HB101/pCADH100 produced a polypeptide of Mr 33,000 identified as the adh gene product based on comigration with a band produced by C. acetobutylicum cell extracts. Cell extracts from E. coli HB101/pCAD100 grown both anaerobically and aerobically exhibited increased levels of NADP-dependant ADH activity, when alcohol was used as substrate. In contrast, there was no significant difference in the level of NAD- dependent ADH activity. The enzyme showed a broad substrate specificity to primary alcohols and aldehydes but not to secondary alcohols or ketones. These results showed that the Adh required NADP+ as a cofactor for function.

The adh gene of C. acetobutylicum is the first NADP-dependent alcohol dehydrogenase which has been cloned, although a number of NAD-dependent alcohol dehydrogenases have been cloned and sequenced from Zygomonas mobilis (Conway et al., 1987) and Saccharomyces cerevisiae (Williamson et al., 1980).

PART III: LACTOSE UTILIZATION IN BACTERIA

1.6 LACTOSE UTILIZATION

1.6.1 The use of whey for the ABE fermentation

One of the primary concerns in investigating the economic feasibility of the ABE fermentation process is the availability and cost of a suitable substrate for the process. Starch and molasses are the traditional substrates for commercial production, but neither is readily available in NZ, at a sufficiently low cost.

However, two potential substrates are readily available, one being wood hydrolysates and waste products from the forestry and paper industries. Investigations on the potential use of this substrate for production of solvents via the ABE fermentation have been reported (Compere and Griffith, 1979; Maddox and Murray, 1983; Saddler et al., 1983; Yu et al., 1984a,b).

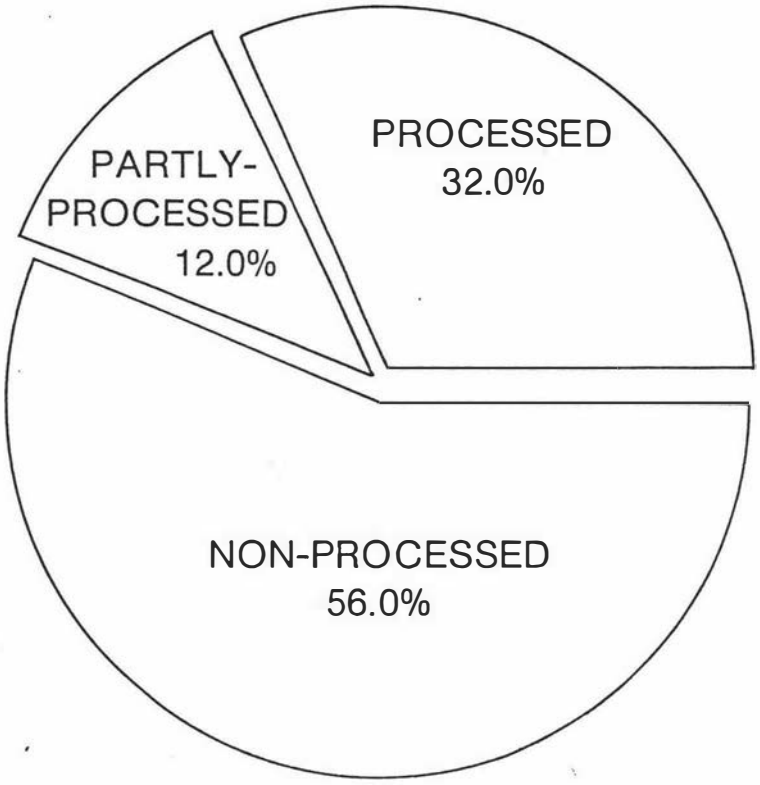
The second potentially useful substrate available at a low cost is whey, which is produced in vast quantities by the dairy industry. Although whey has a relatively low sugar content (lactose, 40-50 g/l) (and is therefore unsuitable for many fermentation processes without prior concentration) it is a satisfactory substrate for production of solvents via the ABE fermentation by C. acetobutylicum as product inhibition at approximately 20 g/l greatly restricts the amount of sugar utilized (Maddox, 1980; Maddox et al., 1981; Gapes et al., 1982; Welsh and Veliky, 1984; Ennis and Maddox, 1985; Ennis et al., 1986; Ennis and Maddox, 1987a,b).

Generally, while whey is a suitable substrate for the growth of C. acetobutylicum, solvent concentration, productivity, and the extent of lactose utilization are lower than reported for industrial substrates (Beesch, 1952; Spivey, 1978). Thus it appears that lactose is generally a poorer substrate, compared to glucose, for

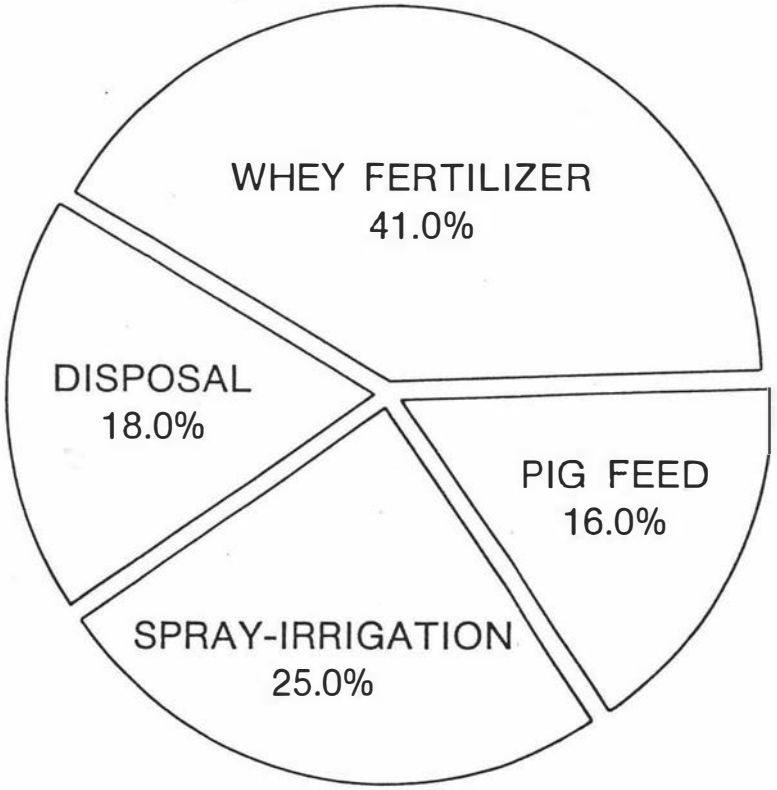
Facing page 43

Fig.1.4: Utilization of waste whey within New Zealand (1985/86 season)

WHEY UTILIZATION 1985/86 SEASON
Total volume 2732000 cubic metres



NON-PROCESSED WHEY 1985/86 SEASON
Total volume 1530000 cubic metres



solvent production (Abou-Zeid et al., 1978; Compere and Griffith, 1979). Studies performed using whey (or permeate) as the fermentation substrate include those reported by Maddox (1980), Gapes et al.(1982), Moreira (1983), Welsh and Veliky (1984), Ennis and Maddox (1987a,b), Qureshi and Maddox (1987).

1.6.2 Whey production in New Zealand

During the 1985/86 season 2,732,000 m^3 of whey was produced from the manufacture of cheese and casein, leading to large disposal problems. Some of this whey is treated by ultrafiltration to remove the protein, resulting in deproteinated whey (permeate) containing approximately 6% total solids. Of the remaining total solids the majority (over 80%) is lactose with smaller quantities of protein, non-protein nitrogen and mineral salts (Short, 1978; Hobman, 1984). However, despite the deproteination step, there is a negligible decrease in the BOD value of the whey, which approximates 100,000ppm. This presents a serious waste disposal problem to the dairy industries in many countries, including NZ. At present, less than 50% of the whey is partly or fully processed (Fig. 1.4). The 56% of non-processed whey is disposed of as fertilizer, pig feed, spray irrigation or diluted into water.

There is, however, the possibility that fermentation to a useful end-product, such as butanol, may provide a greater return to the dairy industry than that from present disposal methods. Feasibility studies have been undertaken by NZERDC (1982) to investigate the economics of solvent production from whey. Investigation by Hobman (1984) on the current domestic market for butanol and acetone suggested that production of these solvents via the ABE fermentation utilizing whey as the substrate, would require the development of further large markets, such as in liquid fuels, for cost competitiveness and economic feasibility. Nevertheless, research continues into the use of whey as the substrate for the ABE fermentation, because of its potential, rather than actual, importance.

Fig.1.5: Pathways for lactose utilization in bacteria.

Abbreviations:

PEP:PTS: -phosphoenolpyruvate-dependant phosphotransferase system
Lac-6P: lactose-6-phosphate
Gal-6P: galactose-6-phosphate
P-ß-gal: phospho-ß-galactosidase
Tag-6P: tagatose-6-phosphate
TDP: tagatose-1,6-diphosphate
DHAP: dihydroxyacetone-phosphate
Gly-3P: glyceraldehyde-3-phosphate
Glu-6P: glucose-6-phosphate
Glu-1P: glucose-1-phosphate
Gal-1P: galactose-1-phosphate
ß-gal: ß-galactosidase
FDP: Fructose-1,6-diphosphate

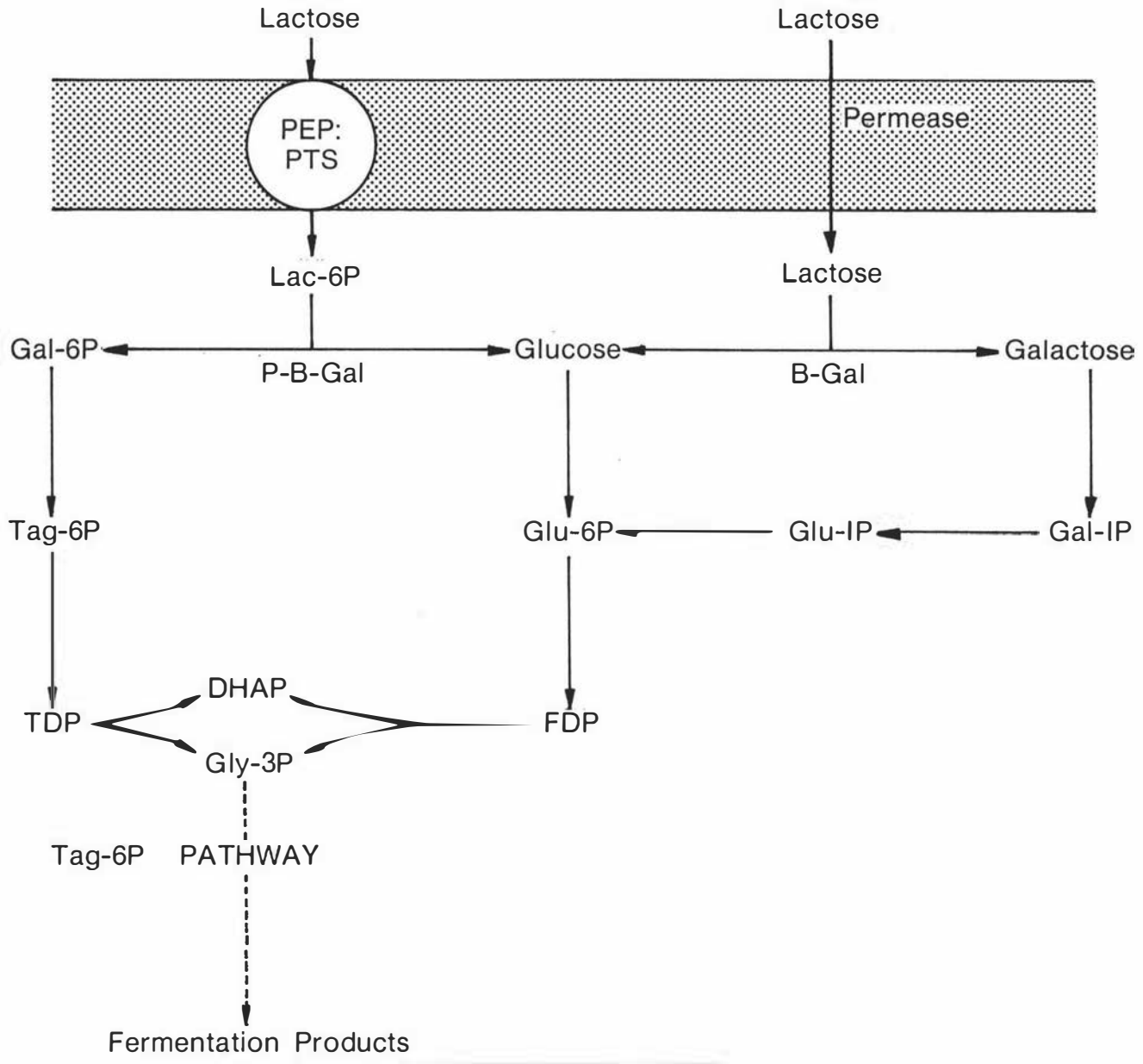
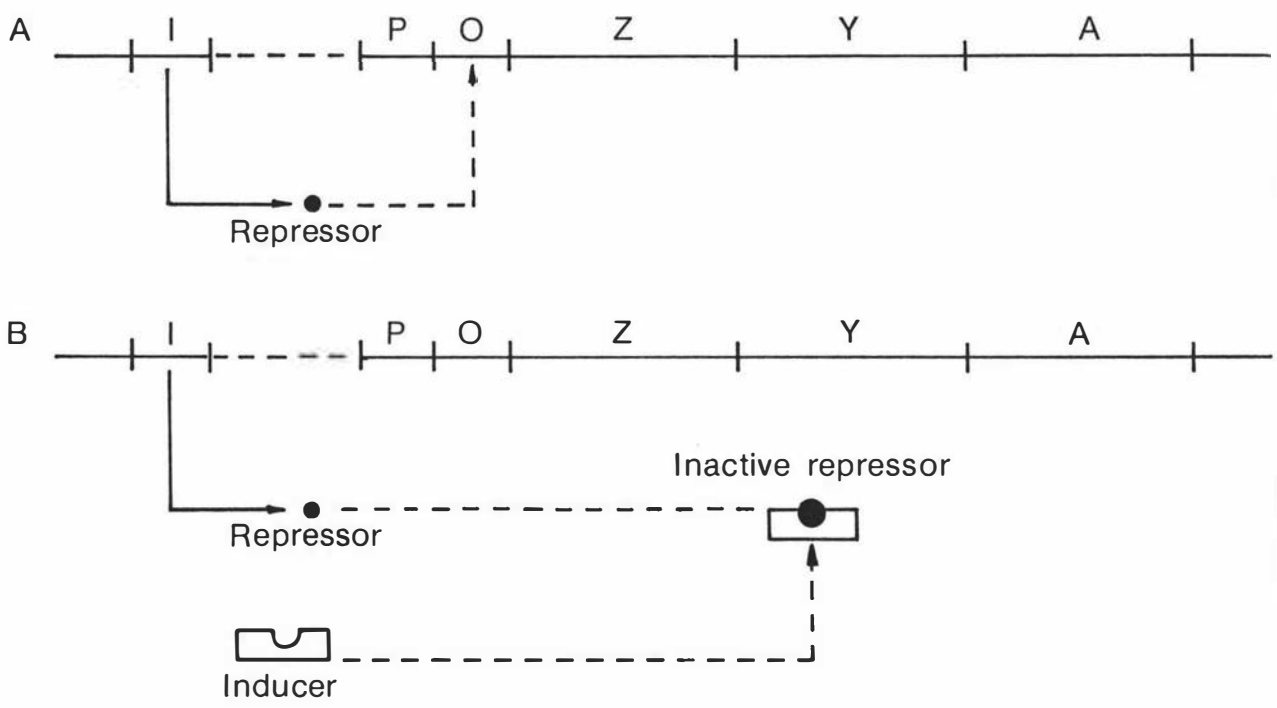

Lactose
Lactose
PEP:
PTS
Permease
Lac-6P
Lactose
Gal-6P
P-B-Gal
Glucose
B-Gal
Galactose
Tag-6P
Glu-6P
Glu-IP
Gal-IP
DHAP
TDP
FDP
Gly-3P
Tag-6P PATHWAY
Fermentation Products

Fig.1.6: The Lactose operon in E. coli.

A The inactive form of the operon. The repressor protein binds to the operator region (O) and prevents synthesis of the mRNA which codes for the structural genes lacZYA.

B The active form of the operon. Maximum rate is achieved in the presence of positive-activating CAP-cAMP complex (bound to the CAP site) and with the inducer present. The inducer binds to the repressor, converting it to an inactive form. Synthesis of the mRNA which codes for the structural genes lacZYA proceeds. Low levels of lac expression occur in the presence of the inducer without cAMP.

Z - ß-galactosidase structural gene
Y - lactose permease structural gene
A - lactose transacetylase structural gene
O - operator site
I - repressor structural gene
P - promoter site


A
I
P
O
Z
Y
A
Repressor
B
I
P
O
Z
Y
A
Inactive repressor
Repressor
Inducer

1.7 PATHWAYS OF LACTOSE UTILIZATION IN MICROORGANISMS

Microorganisms can utilize lactose by either or both of two metabolic pathways (Fig. 1.5). The first pathway involves the ß-galactosidase and is common in Gram-negative bacteria. The second system is known as the phosphoenol- pyruvate--dependant phosphotransferase (PEP:PTS) lactose system and is more commonly found in Gram-positive bacteria (McKay et al., 1970; Dills et al., 1980; Saier, 1985).

1.7.1 Metabolism of lactose in E. coli (The Lac Operon)

The four genes that encode gene products responsible for lactose utilization by E. coli are located at 9 min on the E. coli bacterial map. Three of these structural genes (lacZ, lacY, and lacA) constitute a polycistronic operon transcribed in the direction of lacZ to lacA (Fig. 1.6). The first gene, lacZ, codes for the enzyme ß-galactosidase which hydrolyses the disaccharide lactose to glucose and galactose. The lacY gene directs the synthesis of the hydrophobic lactose permease which is required for active transport of the lactose into the cell (Rickenberg et al., 1956). The lacA gene codes for the ß-galactoside transacetylase. The biological function of the ß-galactoside transacetylase, which acetylates lactose and other ß-galactosides with acetyl CoA, is probably that of a detoxification enzyme. Non-metabolizable structural analogues of lactose are acetylated and excreted from the cell (Lederberg, 1950; Andrews and Lin, 1976; Gottschalk, 1986; Miller and Reznikoff, 1980). However, lacA- mutants have no detectable impairment of lactose utilization, and LacA does not appear to be essential for cell growth on lactose. The lacI gene encodes for the repressor, which is active in inhibiting lac transcription (Fox et al., 1966). Synthesis of the lac proteins other than the repressor is not constitutive. The lacI repressor protein is produced constitutively from a separate promoter, at a low rate.

Expression of the lacZYA operon is negatively controlled by the repressor lacI and positively regulated by the catabolite-activating protein:cyclic adenosine-3-5-

-monophosphate (CAP-cAMP) complex (Loomis and Magasanik, 1964; Perlman and Pastan, 1968; Goeddel et al., 1978). This lac regulatory region represents the majority of the lacI- lacZ intercistronic region. The upstream promoter of the lacZYA genes directs the synthesis of mRNA, transcribing a polycistronic operon. Glucose represses the expression of the lac operon in part, by reducing the concentration of cAMP. The production of lac mRNA requires the CAP-promoted RNA polymerase binding and a vacant operator site (Dickson et al., 1975; Miller and Reznikoff, 1980). This synthesis is controlled by the repressor which is able to bind to the operator O region of the operon and block the production of the lac mRNA. Under these conditions, the synthesis of the lac-specific enzymes is repressed (Fox et al., 1966; Goeddel et al., 1978).

The promoter (lacP) is required for initiation of the lac operon (Dickson et al., 1975; Miller and Reznikoff, 1980). It can be divided into two main regions; RNA polymerase binding site, and the CAP-cAMP site. The first is adjacent to the operator (lacO) and carries information for the initiation of transcription. The efficiency of RNA polymerase initiation is strongly dependant upon the binding of the CAP-cAMP complex to the CAP site (Beckwith et al., 1972; Majors, 1975; Miller, 1978).

Control of lac expression is mediated by the operator lacO site. Binding of the lac repressor prevents activation of lacZYA, as although RNA polymerase binding can occur, the presence of the repressor prevents productive transcription initiation. Binding of the lacI product does not totally inhibit lacZYA transcription, allowing the production of basal levels of the ß-galactosidase and permease enzymes, which are synthesised in the absence of a suitable inducer (Goeddel et al., 1978). In the presence of a suitable substrate, such as lactose, entering the cell, allolactose (the true inducer) is formed by the action of the ß-galactosidase. Allolactose binds to the repressor, whether the repressor is bound to the O site or free (Muller-Hill et al., 1964). The repressor is modified in such a way so that it can no longer bind to the operator region and

transcription can proceed. This allows for the synthesis of the lac mRNA and the production of the lac gene products (Fig. 1.6) (Dickson et al., 1975).

This system has served as a model system which led to the formulation and study of the operon model for gene regulation (Jacob et al., 1960; Jacob and Monod, 1961). Therefore, the structural and regulatory features of the lac operon in prokaryotic cells is well understood. Most Gram-negative enteric bacteria have the three lac genes organized as a polycistronic operon. ß-galactosidase activity is also present in many Gram-positive bacteria, such as Lactobacillus helviticus, L. bulgaricus, L. acidophilus, (Premi et al., 1972; Hickey et al., 1986), S. lactis, (Citti et al., 1965; McKay et al., 1970; Farrow, 1980; Fisher et al., 1985; Crow and Thomas, 1984) S. thermophilus (Tinson et al., 1985) Streptococcus salivarius, S. mutans, (Hamilton and Lo, 1978) and Bacillus megaterium (Landman, 1957).

1.7.2 Metabolism of lactose by the PEP:PTS system

Certain Gram-positive bacteria have an alternative pathway of lactose metabolism, which involves transport of lactose into the cell as lactose phosphate, via the phosphoenol pyruvate-dependant phosphotransferase (PEP:PTS) system (see review by Postma and Lengeler, 1985). This system was first described in S. aureus (Hengstenberg et al., 1970; Simoni et al., 1973), and uses energy derived from the high energy phosphate bond in phosphoenol pyruvate (PEP) to translocate lactose through the cell membrane (Fig. 1.5). Four proteins are required to catalyse this transfer of the phosphoryl group from PEP to lactose (Simoni et al., 1968). Enzyme I and HPR (histidine-containing phosphocarrier protein) are soluble cytoplasmic components. They are synthesised constitutively and are necessary for phosphorylation of many sugar substrates of the PTS system (Hays et al., 1973; McKay et al., 1969). A lactose-specific membrane-associated protein, enzyme $II^{lactose}$, which has a catalytic function, is responsible for recognising and binding lactose (Schafer et al., 1981). The lactose-specific soluble cytoplasmic protein, enzyme $III^{lactose}$ serves as an

immediate phosphate donor to lactose in the presence of enzyme $II^{lactose}$ (Nakazawa et al., 1971; Deutscher et al., 1982). Enzyme $II^{lactose}$, enzyme $III^{lactose}$, and phospho-ß-galactosidase are formed after induction of the S. aureus lac system.

The lactose phosphate is hydrolysed by phospho-ß-galactosidase to glucose and galactose-6-phosphate (Hengstenberg et al., 1970). The glucose is subsequently metabolised via the glycolytic pathway to lactic acid (Bissett and Anderson, 1973). The galactose-6-phosphate is metabolised to triose phosphates via the three enzymes of the tagatose bisphosphate pathway (Dills et al., 1980; Saier, 1985). The triose phosphates are then further metabolised by the glycolytic pathway (Bissett and Anderson, 1973). Phospho-ß-galactosidase has been isolated from a number of different bacteria, and shown to bear little resemblance to ß-galactosidase.

The lactose metabolic pathway in S. lactis is similar to that observed in S. aureus. Enzymes involved in the PEP:PTS pathway are often plasmid-encoded, and the plasmids carry genes enabling conjugal transfer (Morse et al., 1968; Chassy et al., 1978; Gasson and Davies, 1980; Lee et al., 1982; Inamine et al., 1986).

The regulation and expression of lactose genes in these bacteria (unlike the other pathway) appears to be quite complex, and little is known. In S. aureus, expression appears to be subject to catabolic repression (McClatchy and Rosenblum, 1963), the preferred inducer being galactose-6-phosphate (Morse et al., 1968; Hengstenberg et al., 1969). In contrast, expression in S. lactis is induced by the presence of lactose or galactose, and repressed by glucose (McKay et al., 1970) with the exception of the phospho-ß-galactosidase of S. lactis L13 which is not repressed by glucose (Boizet et al., 1988). In L. casei, a ß-galactoside is required for induction of the lac PEP:PTS and lac genes, with neither galactose nor galactose-6-phosphate being able to induce expression (Lee et al., 1982). In all species, glucose at least partially represses expression of the

lactose genes, even though the bacteria are reported to lack cAMP. Involvement of cAMP in catabolite repression is unlikely as it is not present in physiologically significant levels in Gram-positive bacteria (Blumenthal, 1972; Botsford 1981).

In S. aureus, the chromosomally located phospho-ß-galactosidase and determinants for PEP:PTS enzyme $II^{lactose}$ and enzyme $III^{lactose}$ are linked (Hengstenberg et al., 1970). Constitutive mutants have shown coordinated increased levels of the three proteins (Morse et al., 1968). The repressor protein, long postulated, has an uncertain map location (McClatchy and Rosenblum, 1963; Morse et al., 1968). In contrast, in L. casei the lac genes are not in a single transcriptional unit and there is an observed lack of coordinate regulation of phospho-ß-galactosidase and the structural genes of the PEP:PTS system (Lee et al., 1982). This system of lactose utilization is known to be used by S. aureus (Hengstenberg et al., 1969), S. cremoris (Okamoto and Morichi, 1979; McKay et al., 1970), S. faecalis (Okamato and Morichi, 1979), S. lactis (Farrow, 1980), S. mutans (Hamilton and Lo, 1978), S. salivarius (Hamilton and Lo, 1978), S. thermophilus (Hemme et al., 1980) L. casei (Premi et al., 1972) and L. acidophilus, (Fisher et al., 1985). Some of these bacteria possess both pathways, though the majority lack the better-known pathway of lactose metabolism involving a ß-galactosidase.

1.7.3 Lactose utilization in C. acetobutylicum

Yu et al. (1987) studied the potential mechanisms for utilization of lactose in various strains of C. acetobutylicum. When grown in the presence of lactose, all strains tested, except ATCC824, showed both ß-galactosidase and phospho-ß-galactosidase activities. Only phospho-ß-galactosidase activity was detected in strain ATCC824. Neither strain P262 nor ATCC824 showed either enzyme activity when grown on glucose. When strain P262 was grown on whey permeate, differential induction of the two enzymes was demonstrated, showing early induction of phospho-ß-galactosidase associated with the acidogenic phase, and

the ß-galactosidase activity peaking at a later stage of fermentation (22h) coinciding with the solventogenic production phase. This differential induction, both with carbon supply and time during growth, clearly demonstrated that phospho-ß-galactosidase is distinct from ß-galactosidase. This raised the possibility that the two lactose-hydrolysing enzymes serve quite different functions during fermentation. In contrast, the regulation of phospho-ß-galactosidase and ß-galactosidase in S. salivarius and S. lactis is such that there is co-induction of both enzymes in the presence of a suitable inducer (Hamilton and Lo, 1978; Crow and Thomas, 1984).

The relative importance of these enzyme systems in solvent production during growth on a lactose substrate remains unknown. Nevertheless, these results suggested that a PEP-dependant- phosphotransferase system, a tagatose pathway and a ß-galactosidase and lactose permease existed in C. acetobutylicum strains.

1.8 CLONING OF LACTOSE UTILIZATION GENES

1.8.1 Lac operon system

Very few lac genes have been cloned and characterized, apart from the E. coli system which is very well understood (Kalnins et al. 1983). Dickson and Markin (1978) cloned the ß-galactosidase from the yeast Kluyveromyces lactis in E. coli. The DNA structure and regulation of the yeast ß-galactosidase was markedly different to that of E. coli. Herman and McKay (1986, 1987) cloned the ß-galactosidase gene from S. thermophilus 19258. This gene was also significantly different at the DNA level from the ß-galactosidase of E. coli.

Gilbert and Hall (1987) cloned the lactose genes from S. bovis H3. This organism possessed a lac operon system similar that of the E. coli system. The ß-galactosidase genes of both bacteria shared strong DNA homology. However, the other genes were dissimilar to each other in DNA homology. In S. bovis,

the production of ß-galactosidase is repressed by glucose. However, the genes cloned in E. coli were not, due to lack of a CAP binding site. Interspecies regulation existed between the systems of both bacteria. Hirata et al. (1984) cloned two ß-galactosidase genes from Bacillus stearothermophilus. This bacterium possessed three ß-galactosidase enzymes (ß-galI, ß-galII, ß-galIII), which were coded for by two ß-galactosidase genes, bgaA (ß-galII, ß-galIII) and bgaB (ß-galI). These two genes were located very close to each other on the chromosome. Hybridization and immunological techniques revealed that the genes and protein products were different to each other and to the ß-galactosidase gene of E. coli.

1.8.2 PEP:PTS system

Several genes have been cloned from the PEP:PTS systems of various Gram-positive bacteria. Lee et al. (1982) cloned the phospho-ß-galactosidase gene for L. casei 64H. The gene was expressible in E. coli, encoding for a protein of Mr 43,000. Breidt and Stewart (1986) cloned the phospho-ß-galactosidase from S. aureus KUS74, which possessed a constitutive phospho-ß-galactosidase due to insertion of Tn551. Inamine et al. (1986) cloned the phospho-ß-galactosidase gene of S. thermophilus H2 and identified that the lac-PTS region was organised as an operon, with a putative phospho-ß-galactosidase regulatory region, lac-PTS gene, followed by the phospho-ß-galactosidase structural gene. Boizet et al. (1988) cloned the phospho-ß-galactosidase of S. lactis L13. The gene was expressed constitutively in E. coli and was not regulated by glucose, galactose or lactose. This is in contrast to the function and regulation of the phospho-ß-galactosidase in S. lactis itself. Maeda and Gasson (1986) cloned and characterized the phospho-ß-galactosidase gene from S. lactis 712, which produced a 58,000 Mr protein. The gene was not expressed fully in E. coli, nor could it be induced. This is again in contrast to the gene when in S. lactis.

PART IV: AIM OF THE THESIS

1.9 AIM OF THE THESIS

A readily available substrate for the ABE fermentation in NZ is whey, obtained from the manufacture of cheese in the dairy industry. Unfortunately, lactose has been shown to be a relatively poor substrate for solvent production when compared to glucose (Maddox, 1980; Abou-Zeid et al., 1978; Compere and Griffith, 1979).

Study of the enzymes involved in lactose utilization pathways is important since the uptake and metabolism of the carbohydrate source helps determine the efficiency of whey as a substrate for the ABE fermentation. Although the pathways for lactose utilization and the lac gene regulation in E. coli are very well understood, the equivalent pathways and their regulation systems in other organisms are less well known. The presence of the two lactose-metabolising systems in C. acetobutylicum has only recently been established (Yu et al., 1987). Apart from this study, however, there are no reports on the physiology, biochemistry or genetics of lactose-utilization enzymes in clostridia.

In addition to its direct application to the ABE fermentation, a study of the genetics of the enzymes involved in lactose uptake and utilization in C. acetobutylicum will provide further information with respect to the presence and role of lactose metabolism in bacteria in general. Hence the work described in this thesis was undertaken to provide information on the genetic systems involved in lactose utilization in C. acetobutylicum. The procedure was to, first, create a genomic library of total DNA using the genetic systems available in E. coli, namely the bacteriophage packaging system and cosmid vectors. Secondly, the genomic library would be used to identify cosmids containing genes involved in the utilization of lactose by C. acetobutylicum, followed by characterization of these genes and comparison with those of other microorganisms.

CHAPTER 2: METHODS AND MATERIALS

2.1 BACTERIAL STRAINS AND PLASMIDS

These are described in Table 2.1.

Table 2.1:

Strain	Relevant characteristics	Reference
Clostridium acetobutylicum:		
ATCC824	Wild type	ATCC[b]
NCIB2951	Wild type	Maddox (1980)
NRRL594	Wild type	NRRL Culture collection[a]
NRRL598	Wild type	NRRL Culture collection
NRRL2490	Wild type	NRRL Culture collection
P262	Wild type	Allcock *et al.* (1981)
Clostridium perfringens:		
CW92	Wild type	Rood *et al.* (1978)
Clostridium butylicum:		
NRRL593	Wild type	Truffaut and Sebald (1983)
Streptococcus lactis:		
ATCC7962	Wild type	NZDRI[c] Culture collection
Streptococcus thermophilus:		
DRI1242	Wild type	NZRCC[d]
Lactobacillus bulgaricus:		
20056	Wild type	NZRCC
Lactobacillus helviticus:		
20064	Wild type	NZRCC
Staphylococcus aureus:		
ATCC9144	DNase+	MMCC[e]

Table 2.1 cont'd:

Strain	Relevant characteristics	Reference
Escherichia coli:		
BHB2688	N205 recA [imm^{434} cIts b2 red Dam Sam]	Hohn (1979)
BHB2690	N205 recA [imm^{434} cIts b2 red Eam Sam]	Hohn (1979)
DC272	adh80 aceF10 lac+	Clark & Cronan (1980)
DRI4477	HB101/pLZ600	Lee et al. (1982)
DRI4988-4989	HBB101/pKH90-91	This study
DRI4990-4992	HB101/pKH93-95	This study
DRI4993-4996	HB101/pKH93-95	This study
DRI4988-4989	HBB101/pKH90-91	This study
DRI4990-4992	HB101/pKH93-95	This study
DRI4993-4996	HB101/pKH96-99	This study
DRI4997-4999	HB101/pKH100-102	This study
DRI5013	HB101/pKH106	This study
DRI5014	HB101/pKH107	This study
DRI5015	HB101/pKH179	This study
DRI5019-DRI5059	HB101/pKH108-pKH172	This study
DRI8519	PB2959/pKH107	This study
DRI8520	PB2959/pKH179	This study
DRI8521	PB2959/pSUP202	This study
DRI8530	JM101/pSUP202	This study
DRI8531	JM101/pKH107	This study
DRI8532	JM101/pKH179	This study
DRI8533	HB101/pKH184	This study
DRI8534	HB101/pKH185	This study
DRI8535	HB101/pKH186	This study
DRI8536	HB101/pKH187	This study
DRI8537	HB101/pKH188	This study
DRI8538	HB101/pKH189	This study
HB101	F- pro leu thi lacY Str^{R} hsdR hsdM endA recA	Boyer and Roulland -Dussoix(1969)
JM101	(Δlac proAB) supE thi F' proAB lacI^{q} Z M15	Yanisch-Perron et al. (1985)
JM109	(Δlac proAB) supE thi recA relA gyrA hsdR [F' proAB lacI^{q}Z M15]	Yanisch-Perron et al. (1985)

Table 2.1 cont'd:

Strain	Relevant characteristics	Reference
K802	supF gal met supE hsdR+ hsdM+	Wood (1966)
PB2959	F- Δ lac pheC	P. Bergquist (Auckland University)
PN1126	HB101/pAD10	Egelhoff and Long (1985)
Bacteriophage:		
467	b221 rex::Tn5 cI857 Oam 29 Pam 80	(Berg et al., 1982)
Plasmids:		
pBR328	Ap^R Tc^R Cm^R	Boliver et al. (1977)
pSUP202	Ap^R Tc^R Cm^R oriT (from RP4)	Simon et al. (1984)
pLAFR1	cos derivative of pRK290; Tc^R, repRK2	Friedman et al. (1982)
c2RB	double cos derivative Ap^R	Bates and Swift (1982)
pUC18/19	Ap^R	Vieira and Messing (1982)
pRK2013	rep col_{E1} Km^R	Ditta et al. (1980)
pAD10	Ap^R pUC8 S. typhimurium promoter, and E. coli rpoC terminator	Egelhoff & Long (1985)
pLZ600	pBR322 clone containing 7.9kb pbg region of L. casei	Lee et al. (1982)
pKH90-91	pLAFR1 cosmids containing NCIB2951 DNA that complements the pro mutation in HB101	This study
pKH93-95	pLAFR1 cosmids containing NCIB2951 DNA that complements the thi mutation in HB101	This study
pKH96-99	pLAFR1 cosmids containing NCIB2951 DNA that complments the leu mutation in HB101	This study
pKH100-102	pLAFR1 cosmids containing the ß-galactosidase gene region of NCIB2951	This study

Table 2.1 cont'd:

Strain	Relevant characteristics	Reference
pKH106	pSUP202 clone containing 7.2 kb EcoRI fragment from pKH101	This study
pKH107	pSUP202 clone containing 5.2 kb EcoRI fragment from NCIB2951	This study
pKH108-pKH172	pKH107::Tn5 clones containing ß-galactosidase region of NCIB2951	This study
pKH173	pUC18 clone containing 0.9 kb EcoRI/HaeIII fragment of the 5.2 kb fragment	This study
pKH174	pUC18 clone containing 4.3 kb EcoRI/HaeIII fragment of the 5.2 kb fragment	This study
pKH179	pSUP202 clone containing 5.2 kb EcoRI fragment from NCIB2951 in the opposite orientation to pKH107	This study
pKH184	pAD10 clone containing 5.2 kb ß-galactosidase region of NCIB2951	This study
pKH185	pAD10 clone containing 5.2 kb ß-galactosidase region in opposite orientation	This study
pKH186	pAD10 clone containing 10.9 kb ß-galactosidase::Tn5 region of pKH112	This study
pKH187	pAD10 clone containing 10.9 kb ß-galactosidase::Tn5 region of pKH127	This study
pKH188	pAD10 clone containing 10.9 kb ß-galactosidase::Tn5 region of pKH124	This study
pKH189	pAD10 clone containing 10.9 kb ß-galactosidase::Tn5 region of pKH153	This study

[a] NRRL Culture Collection, Illinois, USA
[b] ATCC American Type Culture Collection, Maryland, USA
[c] NZDRI New Zealand Dairy Research Institute, Palmerston North, New Zealand

[d] NZRCC New Zealand Reference Culture Collection (NZDRI), Palmerston North, New Zealand

[e] MMCC Massey University Microbiology Culture Collection, Palmerston North, New Zealand

2.2 PREPARATION OF CULTURE MEDIA

2.2.1 Liquid Media

Clostridial Growth Medium: (CGM; Ennis & Maddox, 1985).

Composition (g/l): lactose, 10; yeast extract, 5; ammonium acetate, 2; NaCl, 1; cysteine hydrochloride, 0.5; solution I, 10 ml (K_2HPO_4, 75; KH_2PO_4, 75); solution II, 2 ml ($MgSO_4.7H_2O$, 50; $MnSO_4.4H_2O$, 2; $FeSO_4.7H_2O$, 2), pH 7.0.

Cooked Meat Medium (Difco). Composition (g/l):125, water to 1l.

Luria broth (LB; Miller, 1972). Composition (g/l):tryptone, 10; yeast extract, 5; NaCl, 5. pH 7.2.

Minimal Medium 56 (Bergquist and Alderberg, 1972). Composition (g/l): KH_2PO_4, 5.36; Na_2HPO_4, 8.69; $(NH_4)_2SO_4$, 2; 10% $MgSO_4.7H_2O$, 2 ml; 1% $Ca(NO_3)_2$, 1 ml; 0.027% $FeSO_4.7H_2O$, 1ml. pH 7.2. For use the medium was diluted with an equal volume of deionized water.

M17 broth (Terzarghi and Sandine, 1975). Composition (g/l):lactose, 5; polypeptone, 5; beef extract, 5; soy peptone, 5; yeast extract, 2.5; ascorbic acid, 0.5; sodium glycerophosphate, 19; 1M $MgCl_2$, 1ml. pH7.2.

MRS medium (de Man et al., 1960). Composition (g/l): peptone, 10; meat extract, 10; yeast extract, 5; glucose, 20; Tween 80, 1ml; K_2HPO_4, 2; sodium acetate, 3; di-ammonium citrate, 2; salt solution ($MgSO_4.7H_2O$, 46; $MnSO_4.4H_2O$, 9.6), 5 ml.

YT broth Composition (g/l): tryptone, 16; yeast extract, 10; NaCl, 5. pH 7.4.

Lambda broth (de Bruijn & Lupski, 1984). Composition (g/l): tryptone, 10; NaCl, 2.5. pH 7.0.
YM broth (de Bruijn & Lupski, 1984). Composition (g/l): tryptone, 10; NaCl, 2.5; maltose, 2; yeast extract, 0.1. pH 7.0.

2.2.2 Solid Media

These were prepared by adding agar (15 g/l) (Difco) to liquid media.
DNase medium: Composition : DNase medium (Difco), 30g/l. pH 7.0. For growth of Clostridium strains the medium was supplemented with the following (g/l): yeast extract, 5; cysteine hydrochloride, 3; lactose, 10; agar, 20.
L top agar: Composition (g/l): tryptone. 10; NaCl, 8; agar, 8.

2.3 MAINTENANCE OF CULTURES

E. coli cultures were maintained on LB agar plates at room temperature or at 4°C. C. acetobutylicum cultures were maintained as a spore suspension in distilled water at 4°C. C. perfringens cultures were maintained on CGM plates. Streptococcus and Lactobacillus strains were maintained on M17 and MRS broth respectively at 4°C. For long term maintenance, all cultures were stored in 50% (v/v) glycerol at -70°C.

2.4 GROWTH OF BACTERIA

Clostridial spores were activated (70°C/3 min) in cooked meat medium and the cells were grown anaerobically overnight at 30°C before subculturing into CGM. C. acetobutylicum cells were grown in CGM with or without glycine (3 g/l). Cells were harvested at early, mid, or late exponential phase or stationary phase as required. Cell stage was checked by monitoring the absorbance at 600nm, in combination with phase contrast microscopy. E. coli were grown at 30°C or 37°C in solid or liquid media supplemented as required with appropriate antibiotics: ampicillin (Ap; 100 μg/ml), chloramphenicol (Cm; 15 μg/ml), kanamycin (Km; 50

μg/ml), tetracycline (Tc; 15 μg/ml). All antibiotics were supplied by Sigma Chemical Co (St Louis, Mo., USA). Streptococcus species were grown at 30°C in M17 media. Lactobacillus species were grown at 37°C in MRS medium.

2.5 METHOD OF LYSIS

Three different bacteriolytic enzymes were investigated for their effect on the lysis of C. acetobutylicum.

2.5.1 Lysis experiments using lysozyme

Lysis experiments using lysozyme were carried out using cells of various culture ages grown either in the presence or absence of glycine (3 g/l). The lysozyme was obtained from Sigma Chemical Co. and was used according to the manufacturer's instructions.

Materials

(1) Lysis buffer; 25mM Tris-HCl, 10mM Na_2EDTA, pH8.0 (2) Lysozyme (50mg/ml) in lysis buffer

Method

As a standard procedure, cells from a 100 ml culture were harvested by centrifugation (3,640 **g**, 10 min, 4°C), washed once with cold sterile lysis buffer, and resuspended in lysis buffer to a standard OD_{600} of approximately 0.6. The lysozyme solution was mixed into 4 ml of the cell suspension to give final concentrations of 0.5, 1, 2, 5, and 10 mg/ml, and incubated at 37°C for digestion. Lysis in the cell/enzyme suspension was monitored at various time intervals by recording the optical density, at 600nm, of the contents of each tube. The extent of cell lysis was expressed as the percentage of unlysed cells by calculating the ratio OD_{600} at time y to OD_{600} at time zero, x 100.

2.5.2 Lysis experiments using mutanolysin

Experiments were carried out using cells of various culture ages grown either in the presence or absence of glycine (3 g/l). The mutanolysin was obtained from Sigma Chemical Co. and was used according to the manufacturer's instructions.

Materials

(1) 100mM Sodium phosphate storage buffer; 18.5mM $K_2HPO_4.7H_2O$, 81.5mM NaH_2PO_4, pH 6.2 (2) Lysis buffer; 20mM Tris-HCl, 1mM $MgCl_2$, pH 7.0 (3) TES buffer; 10mM Tris-HCl, 1mM Na_2EDTA, 50mM NaCl, pH7.0

Method

The mutanolysin (25 mg) was dissolved in storage buffer to a concentration of 5 mg/ml (12650U/ml), and stored in aliquots frozen at -20°C. Cells were centrifuged (3,640 **g**, 10 min, 4°C), washed in TES buffer and resuspended in the lysis buffer to produce a suspension with an absorbance of approximately 0.6 at 600nm. To 4 ml of this cell suspension mutanolysin, in storage buffer, was added to give 0, 5, 12.5, 25, 50, 100, and 200 μg/ml. The mixture was incubated at 55°C, and the decrease in turbidity at OD_{600} was measured at various time intervals. The percentage of cell lysis was calculated as described for lysozyme.

2.5.3 Lysis experiments using achromopeptidase

Experiments using achromopeptidase were carried out using cells of various culture ages grown in the absence of glycine. The achromopeptidase (containing 1,000,000 units/g) was obtained from Wako Pure Chemicals Industries Ltd, Osaka, Japan, and was used according to the manufacturer's instructions.

Materials

(1) Storage buffer; 10mM Tris-HCl, 10mM NaCl pH 8.0 (2) Lysis buffer; 10mM Tris-HCl, 10mM NaCl pH 8.0 (3) TES buffer; 10mM Tris-HCl, 1mM

Na_2EDTA, 50mM NaCl pH 8.0.

Method

The enzyme was dissolved in 20 ml of storage buffer at a concentration of 50U/ml(50 μg/ml). Cells were centrifuged (3,640 **g**, 10 min, 4°C), washed once in TES buffer, and resuspended in lysis buffer to produce a suspension with an absorbance of approximately 0.6 at 600nm. The correct number of units of achromopeptidase was added to 5 ml of the cell suspension to give enzyme concentrations of 0, 25, 50, 100, 200, and 300 units/ml and the mixture was incubated at 40°C . The decrease in turbidity at 600nm was measured at various time intervals, and the percentage of cell lysis was calculated as described for lysozyme.

2.6 BATCH FERMENTATION RUN

The fermenter used was a Microferm Laboratory Fermenter (New Brunswick Scientific Co, New Brunswick, New Jersey, USA) equipped with a 2 l pyrex glass vessel of 1.6 l working volume. The fermenter vessel was removed from the autoclave and attached to the fermenter apparatus while still hot (85-90°C). Cooling was initiated by means of cold water flowing through hollow baffles, controlled by an electronic thermostat and oxygen-free nitrogen gas was swept across the medium surface. Gas sweeping was continued until good gassing due to microbial growth was observed. When the fermentation culture was at 34°C, the medium pH was measured and the fermenter was inoculated with an overnight culture of highly motile, gassing C. acetobutylicum. The pH and OD_{600} were measured by standard techniques. The number of cells/ml was measured using a cytometer counter, after suitable dilution of culture.

2.6.1 DNase assay

DNase activity was measured by the method of Blaschek and Klacik (1984).

Materials

(1) DNase buffer: 20mM Tris-HCl, 50mM NaCl, 10mM $MgCl_2$, 5mM ß-mercaptoethanol (pH 8) (2) DNA stock solution; 1mg/ml DNA type III (Sigma) (3) 5M HCl.

Method

A sample (1 ml) was removed from the fermenter at various time intervals, and the cells were spun down. The supernatant was removed and both the cell fraction and the medium were tested for DNase activity. The cells were lysed in 1 ml of DNase buffer until clear, and the solution was re-centrifuged to remove cell debris. Then 750 μl of DNA stock solution was mixed with 7.25 ml of DNAse buffer, and pre-warmed at 37°C, 15 min. 1 ml of cell fraction was added (extracellular medium, or intracellular), mixed, and incubated at 37°C for 15 min. The polymerised DNA present at time zero and remaining after 15 min at 37°C was precipitated with 1 ml of 5M HCl. Turbidity was developed at 37°C and the OD_{600} was measured. The DNase activity present was based on the amount of substrate DNA depolymerised in 15 min as measured from a standard curve minus the zero control. 1U = the amount of enzyme depolymerising 1 μg of DNA/min at 37°C.

2.7 ISOLATION OF PLASMID DNA

Plasmid DNA was isolated by the three methods described below.

2.7.1 Preparative isolation of plasmid DNA from E. coli by the cleared lysate method

A modification of the cleared lysate procedure of Clewell and Helinski (1969) was used.

Materials

(1) 25% (w/v) sucrose solution in 0.05 M-Tris pH 8.0 (2) Lysozyme (20 mg/ml) freshly prepared in 25% (w/v) sucrose solution (3) TE (10/1) buffer pH 8.0: 10mM Tris-HCl, 1mM Na_2EDTA (4) Triton lysis mixture; 0.25M Na_2EDTA (25 ml), 1M Tris pH 8.0 (50 ml), 20% (v/v) Triton X-100 in 0.01M Tris-HCl pH 8.0 (10 ml), and sterilized water (60 ml) (5) CsCl (6) Ethidium bromide (10 mg/ml).

Method

An overnight culture of E. coli in LB (25 ml) was inoculated into 1 l of LB and incubated at 37°C with shaking (200 rpm). For isolation of amplifiable plasmid DNA, chloramphenicol (Cm, 170 μg/ml) was added when OD_{600} was approximately 0.4. Spectinomycin (399 μg/ml) was used if the plasmid contained the Cm^R gene. Cells were incubated for a further 12-16 h and were harvested by centrifugation (12,100 **g**, 10 min, 4°C). The remaining operations were performed on ice using chilled solutions. The cell pellet was resuspended in 8 ml of sucrose solution, lysozyme (2 ml) and Na_2EDTA (4 ml) were added and the mixture was incubated for 10 min. Triton lysis mixture (12 ml) was added to the mixture and gentle agitation was maintained until lysis was completed. The lysate was centrifuged (48,200 **g**, 45 min, 4°C) and the clear supernatant was collected. Care was taken to avoid the gelatinous chromosomal DNA pellet.

Cesium chloride was added to the supernatant to a final concentration of 1 g/ml, and the mixture was incubated at room temperature for 1 h followed by centrifugation (48,200 **g**, 15 min, 4°C) to remove cell debris. Ethidium bromide was added to a final concentration of 300 μg/ml and the mixture was held at 4°C overnight. The next day, the mixture was centrifuged (48,200 **g**, 15 min, 4°C) to pellet the ribonucleoprotein. The refractive index of the supernatant was adjusted to n = 1.393-1.397 using cesium chloride and the mixture was centrifuged in a Beckman Type 65 rotor at 50,000 rpm for 40 h at 16°C. The plasmid DNA band was collected and the ethidium bromide was removed by sequential mixing

with NaCl-saturated isopropanol until the isopropanol was clear. The cesium chloride was removed by dialysing in several changes of TES (10:1:100) solution for at least 24 h. The DNA solution was concentrated by ethanol precipitation and the purity and concentration of the DNA was determined spectrophotometrically. The DNA solution was stored at 4°C.

2.7.2 Alkaline lysis method

The method of Ish-Horowitz and Burke (1981) was scaled up for the preparation of the plasmids from E. coli.

Materials

(1) Solution I: 5 mM glucose; 25 mM Tris-HCl pH 8.0; 10 mM Na_2EDTA (2) Solution II: 0.2 M NaOH, 1% (w/v) SDS (3) Solution III: 5 M potassium acetate pH 5.8 (4) Lysozyme (5) Isopropanol (6) 70% ethanol (7) TE 10:1 buffer pH 8.0: 10 mM Tris-HCl; 1 mM Na_2EDTA (8) CsCl (9) Ethidium bromide (10 mg/ml).

Method

E. coli cells (1 l), grown to late exponential phase, were harvested by centrifugation (12,100 **g**, 10 min, 4°C) and resuspended in 30 ml of solution I. Lysozyme (3 ml of 50 mg/ml) was added and the mixture was incubated at room temperature for 10 min. Solution II (60 ml) was added, mixed by inversion, and incubated on ice for 10 min. Solution III (45 ml) was added and the mixture was agitated briefly, incubated for 10 min on ice, and then centrifuged (4,080 **g**, 10 min, 4°C). DNA was precipitated by adding 0.6 volume of isopropanol to the supernatant, followed by incubation for 10 min at room temperature and centrifugation (4,080 **g**, 10 min, 4°C). The pellet (DNA + RNA) was washed twice with 70% ethanol at room temperature, dried and resuspended in TE 10:1 buffer. CsCl and ethidium bromide were added, and the DNA-CsCl mixture was centrifuged (vertical rotor, Vi Type 65, 60,000 rpm, 6 h, 15°C). The plasmid

DNA was removed and purified. The purity and concentration of the DNA were determined spectrophotometrically.

2.7.3 Rapid boiling method

The method of Holmes and Quigley (1981) was used for small scale preparation of plasmid DNA from E. coli.

Materials

(1) STET buffer: 8% (w/v) sucrose; 5% (w/v) Triton X-100; 50 mM Na_2EDTA pH 8.0; 50 mM Tris-HCl, pH 8.0 (2) Lysozyme (10 mg/ml) in 10 mM Tris pH 7.6 (3) Isopropanol.

Method

E. coli cells (1.5 ml), grown to late exponential phase, were harvested in a microfuge tube and resuspended in 350 μl of STET buffer. Lysozyme (25 μl) was added, and the cells were boiled for 40 s and immediately centrifuged (microcentrifuge, top speed, 10 min). The supernatant (300-350 μl) was collected and an equal volume of isopropanol was added, incubated at -20°C for 10 min and centrifuged (microcentrifuge, top speed, 5 min). The pellet (DNA + RNA) was washed with 95% ethanol, dried in vacuo and resuspended in 50 μl of sterile water.

2.8 ISOLATION OF GENOMIC DNA

2.8.1 Isolation of E. coli, Streptococcus and Lactobacillus DNA

E. coli DNA was prepared by a modification of the method described by Fisher and Lerman (1979).

Materials

(1) TE 50:20 (50 mM Tris-HCl, 20 mM Na_2EDTA) pH 8.0 (2) Lysozyme (10 mg/ml, stored at -20°C) (3) Proteinase K (10 mg/ml) (4) Sarkosyl (60 mg/ml) (5) Phenol (6) chloroform-isoamyl alcohol (24:1, v/v) (7) CsCl (8) Ethidium bromide (9) Dialysis tubing.

Method

1 l of each bacterial strain was grown in LB broth at 37°C for 16 h on a shaker. Cells were harvested by centrifugation (4,080 **g**, 10 min, 4°C) and resuspended in 50 ml of TE (50:20) buffer. Lysozyme was added to a final concentration of 1 mg/ml and the suspension was incubated at 37°C for 1 h. Proteinase K was added to a final concentration of 300 μg/ml and incubation was continued at 50°C for 30 min. Sarkosyl was added to a final concentration of 1% (w/v) and the mixture was incubated at 50°C overnight. After incubation, an equal volume of phenol: chloroform:isoamyl alcohol (25:24:1,v/v) was added, mixed until homogeneous, and then centrifuged (3,640 **g**, 10 min, 4°C). The upper aqueous phase was carefully removed, and an equal volume of chloroform:isoamyl alcohol was added, mixed until homogeneous, and then centrifuged (3,640 **g**, 10 min, 4°C). Cesium chloride was added to the top aqueous phase to a final concentration of 1 g/ml, allowed to dissolve, then 800 μl of ethidium bromide (10 mg/ml) was added for every 10 ml. The DNA:cesium chloride solution was centrifuged in a Beckman ultracentrifuge in a Ti80 rotor for 16 h at 50,000 rpm. The DNA band was collected, the ethidium bromide was removed and the solution was dialysed extensively against TES (10:1:100).

DNA from Streptococcus and Lactobacillus strains was isolated by the same procedure, except that the concentration of lysozyme was increased to 20 mg/ml.

2.8.2 Isolation of C. acetobutylicum DNA

Option 1

C. acetobutylicum genomic DNA was isolated by a modification of the method of Marmur (1961).

Materials

(1) Sucrose-TE 50:20 (25%(w/v)) sucrose, 50 mM Tris-HCl, 20 mM Na_2EDTA) pH 7.0 (2) Mutanolysin (300 μg/ml) (3) 250mM Na_2EDTA, pH 7.0 (4) 20% (w/v) SDS (5) Proteinase K (50 mg/ml) (6)5M sodium perchlorate (7) Chloroform-isoamyl alcohol (24:1, v/v) (8) 3M Na acetate (9) 95% ethanol.

Method

The clostridial spores were activated and grown anaerobically overnight in CMM. The fresh culture was inoculated into freshly autoclaved CGM at 1:200 (v/v) and incubated anaerobically at 30 or 37°C for 6-8 h until the culture was at mid-log phase. This was determined by microscopic examination looking for very actively moving, rapidly dividing clostridia. The cells were then centrifuged (4,080 **g**, 10 min, 4°C) and resuspended in 10 ml sucrose-TE (25%:50:20). Mutanolysin (3 ml) was added, and the mixture was incubated at 50°C for 30 min. Na_2EDTA (10 ml of 250mM) was added, and the mixture was incubated at 50°C for 10 min followed by addition of 3 ml of proteinase K and 1.5 ml of 20% SDS. The mixture was incubated at 50°C for 30 min or overnight until the solution cleared. Sodium perchlorate (6.6 ml of 5M) was added, mixed until homogeneous, followed by an equal volume of chloroform: isoamyl alcohol, and gently emulsified. The emulsified solution was centrifuged (12,100 **g**, 10 min, 4°C) and the aqueous top layer was removed. Twice the volume of ice-cold ethanol was added, and the DNA was spooled out on a stirring rod, and dissolved in TE 50:20.

Option 2

Cesium chloride was added to the DNA solution to a final concentration of 1 g/ml, 800 μl of ethidium bromide was added per 10 ml of DNA:cesium chloride solution and the mixture was centrifuged on a Beckman rotor (50,000 rpm, 20°C, 16 h). The DNA band was collected, ethidium bromide was removed, and the DNA was dialysed extensively against TES (10:1:100).

Option 3

Materials

(1) Guanidinium isothiocyanate solution: 4M-guanidinium isothiocyanate, 5mM-sodium citrate (pH 7.0), 0.1M ß-mercaptoethanol, 0.5% SDS (2) 5.7M CsCl (3) Lysis solution:Sucrose-TE 50:20 (25% (w/v)) sucrose, 50 mM Tris-HCl, 20 mM Na_2EDTA, pH 7.0.

Method

The cells were lysed in 2 ml of lysis buffer, and guanidinium isothiocyanate solution (10 ml) was added. The resulting DNA: guanidinium isothiocyanate solution was loaded on top of a 3 ml 5.7M-cesium chloride cushion, centrifuged (35,000 rpm, 20°C, 12 h) and the DNA band was removed from the interface and dialysed extensively against TES (10:1:100) as described previously.

2.9 DIGESTION OF DNA WITH RESTRICTION ENZYMES

DNA was routinely digested with restriction enzymes as described below.

Materials

(1) Universal buffer (x10) pH 7.6: 0.06M Tris pH 7.6; 0.1M $MgCl_2.6H_20$; 0.1M ß-mercaptoethanol (2) 1M NaCl (3) SDS dye mixture: 20% (w/v) sucrose, 1% (w/v) SDS, 5mM Na_2EDTA, 0.2% (w/v) bromophenol blue (4) Restriction enzymes (New England Biolabs).

Method

All DNA digestions were carried out in Universal buffer (1x) with the salt concentration adjusted as recommended by the manufacturer, using 1M NaCl. Digestions were performed at 37°C for 1-2 h. An aliquot of the digest was checked on a minigel to ensure the digestion had gone to completion before the reaction was stopped by adding 0.2 volume of an SDS dye mixture. If the DNA digestion was incomplete, more enzyme was added and the mixture was incubated further, or the DNA was further purified and the experiment repeated.

2.10 PREPARATION OF PACKAGING EXTRACTS

Packaging extracts were prepared by a modification of the protocol of Maniatis et al. (1982).

Materials

(1) Wash buffer; 40mM Tris-HCl (pH 8.0), 0.1% (v/v) ß-mercaptoethanol (2) CH buffer: 40mM Tris-HCl (pH 8.0), 1mM spermidine, 1mM putrescine, 0.1%(v/v) ß-mercaptoethanol, 7%(v/v) DMSO (3) Liquid nitrogen.

Method

Stocks of E. coli strains BHB2690 and BHB2688 were checked to verify their genotypes by replica-plating streaks of each strain onto LB plates. One plate of each strain was incubated at 30-32°C and the other at 42°C overnight. Growth occurred only on the plates incubated at 30-32°C, establishing that the genotype was correct.

An overnight culture of each strain in LB broth was diluted 100-fold into fresh 500 ml of the same medium in a 2 l flask, and grown with vigorous shaking at 32°C for 3 h or until the OD_{600} was 0.3. The lysogens were then induced by placing the flasks in a water bath pre-heated to 45°C, and the cultures were continuously swirled for 15 min. The cells were then re-incubated at 39°C for 2-3 h with

vigorous aeration. The lysogens were checked for successful induction by adding a drop of chloroform into a small sample of each culture strain, to detect lysis of cells. When the culture samples cleared within 2 min, the two cultures were mixed in a pre-cooled 4 l flask and chilled in an ice-water bath for 10 min. The bacteria were then collected by centrifugation (3,640 **g**, 10 min, 4°C), washed in 300 ml of ice-cold buffer, and re-centrifuged. The fluid was well drained from the centrifuge tube and the cells were resuspended in 4 ml of freshly prepared CH buffer. The bacterial suspension was transferred to a cold glass tube and 50 μl aliquots were dispensed into pre-cooled 1.5 ml Eppendorf tubes as quickly as possible. The cap was placed on each tube, and the tube was plunged into liquid nitrogen. The tubes were removed from the liquid nitrogen and stored at -70°C.

The efficiency of four of the packaging extract aliquots was tested using L47ab DNA and E. coli strain BNN45. The L47ab DNA (500ng) was added to each packaging extract immediately following thawing and mixed by stirring with a pipette tip, followed by incubation at 37°C for 1 h. After incubation, 500 μl of SM buffer was added and three drops of chloroform were added as a preservative. E. coli BNN45 cells were inoculated with an aliquot of the packaging extract and incubated at 37°C for 45 min without shaking. The cells were then diluted serially, mixed with 3 ml of LB top agar and plated onto LB plates, and incubated overnight at 37°C. The resulting plaques were counted, giving an average efficiency of 1.73 x 10^5 pfu/μg of DNA.

2.11 CONSTRUCTION OF A pLAFR1 GENE BANK TO C. ACETOBUTYLICUM NCIB 2951

Materials

(1) SM buffer pH 7.5. Composition g/l: NaCl, 5.8 g; $MgSO_2.7H_2O$, 2.0 g; 1 M Tris-HCl pH 7.5, 50 ml; 2% gelatin, 5 ml. The remainder of the materials used have been described previously.

Method

Genomic DNA from strain NCIB 2951 was isolated and partially digested with EcoRI to an average size of 25 kb, then separated by electrophoresis (1.5 V/cm for 6 h) in a horizontal 1.0% Seaplaque agarose (Marine Colloids) gel. DNA in the size range of 15-30 kb was electroeluted from the agarose gel, extracted with phenol-chloroform and concentrated by ethanol precipitation. Vector DNA, pLAFR1, (Friedman et al., 1982) was isolated by the alkaline lysis method and digested with EcoRI. The electroeluted clostridial DNA (5 μg) was mixed with EcoRI-digested pLAFR1 (0.5 μg) and ligated overnight at 4°C in a 10 μl reaction mixture containing 6 mM Tris pH 7.6, 10 mM $MgCl_2$, 10 mM ß-mercaptoethanol, 1 mM ATP and 400 units of T4-DNA ligase.

DNA packaging was carried out by the method of Hohn (1979) using the packaging extract prepared from the lysogenic strains BHB2688 and BHB2690. A 3 μl aliquot of ligated mixture was added to the packaging extract (40 μl) immediately following thawing, and mixed by stirring with a pipette tip followed by incubation at 37°C for 1 h. After incubation, 500 μl of SM buffer (pre-warmed to 37°C) was added and 3 drops of chloroform were added as a preservative. E. coli HB101 cells, grown overnight in 20 ml of LB broth were harvested by centrifugation (1,085 **g**, 5 min, 4°C), re-suspended in 10 ml of SM buffer at room temperature, and 4.5 ml of the cell suspension was mixed with the 500 μl of DNA-packaging mixture prepared above. The mixture was incubated at 37°C for 45 min without shaking, centrifuged (1,085 **g**, 5 min, room temperature), resuspended in 5 ml of LB and incubated at 37°C for 60 min with slow shaking. The cells were harvested by centrifugation (1,085 **g**, 5 min, room temperature), resuspended in 1 ml of SM buffer, and 200 μl aliquots were plated on LB plates containing tetracycline (15 μg/ml). After overnight incubation at 37°C the Tc^R colonies were washed off with 50% glycerol and stored at -70°C. To check the quality of the gene bank, recombinant plasmid DNA from 24 randomly selected Tc^R colonies was isolated by the rapid boiling method, restricted with EcoRI enzyme and separated on an agarose gel.

2.12 HORIZONTAL AGAROSE GEL ELECTROPHORESIS OF DNA

The horizontal slab gels (150x200x5mm) used contained 0.7% (w/v) agarose (Sigma, Type 1). The TAE electrophoresis buffer pH 7.8, contained 40mM Tris, 1mM Na_2EDTA and 5mM Na acetate; TBE electrophoresis buffer pH 8.3 contained 89 mM Tris, 1 mM Na_2EDTA and 89 mM Boric acid. Electrophoresis was usually carried out at 1.5V/cm for 16 h. Gels were stained and photographed. For rapid electrophoresis of DNA, small samples were analyzed on a minigel (93 x 68 x 1.5mm) run at 13 V/cm for 1 h.

2.13 DETERMINATION OF MOLECULAR WEIGHTS

A HindIII digest of lambda DNA was used as a standard to determine the molecular weight of DNA fragments. The relative mobilities of DNA fragments were measured and molecular weights determined graphically from a plot of relative mobility versus $\log_{10}$ molecular weight (Sanger et al., 1982) or by using a computer programme designed to transform mobilities to molecular weights. In experiments utilizing Bal31, the BRL 1 kb ladder was used to determine the size of the resulting fragments.

2.14 DETERMINATION OF DNA PURITY

DNA purity was determined from UV absorption at 230 nm, 260 nm and 280 nm. The following spectral ratios were regarded as satisfactory: A260/A230nm 1.8-2.3, and A260/A280nm 1.8-2.0. DNA concentrations were determined at 260 nm using extinction coefficients of 20 for unsheared DNA and 24 for sheared DNA, for 1 mg/ml DNA solution (Brenner and Falkow, 1971).

2.15 PURIFICATION OF DNA PREPARATION

The following protocol was used to purify small quantities of DNA.

A measured volume of DNA was transferred to a microcentrifuge tube and an equal volume of phenol-chloroform (1:1; v/v) was added, mixed thoroughly and centrifuged at top speed for 3 min in a microcentrifuge. The aqueous phase was collected, an equal volume of sterilized water was added to the phenol-chloroform mixture and the extraction was repeated. The DNA was then precipitated using ethanol. E. coli transfer-RNA (20 mg/ml) was added to a final concentration of 20-50 μg/ml) when the concentration of DNA was less than 1-5 μg/ml.

2.16 EXTRACTION OF DNA FROM AGAROSE

Freeze-thaw method

Preparative 1% Seaplaque agarose gels were stained with ethidium bromide and DNA was visualised under long wave (350 nm) UV light. The band from the 1% Seaplaque gel containing the DNA fragments was cut into small pieces and placed in a microcentrifuge tube. Tris equilibrated phenol was added to cover the gel fragments and the tube was placed at -20°C overnight to freeze the gel fragments. The microcentrifuge tube was centrifuged (10 min, top speed, room temperature) and the aqueous phase was collected, extracted using phenol-chloroform, and precipitated using ethanol. The DNA was resuspended in water and used directly for cloning or preparation of a [^{32}P]-labelled probe.

2.17 PREPARATION OF DIALYSIS TUBING

Dialysis tubing was prepared according to the method of Maniatis et al. (1982).

Materials

(1) Dialysis tubing (2) $NaHCO_3$ (3) 250 mM Na_2EDTA, pH 8.0.

Method

The dialysis tubing (10 or 40 mm) was cut into 15 cm lengths and boiled in a large volume of 2% (w/v) $NaHCO_3$ and 1 mM Na_2EDTA for 10 min. The tubing was

rinsed thoroughly in distilled water and boiled for 10 min in distilled water. After cooling, the tubing was stored submerged at 4°C. Before use, the tubing was washed with sterile distilled water.

2.18 PREPARATION OF [^{32}P]-LABELLED DNA PROBES

The procedure described by Taylor et al. (1976) and Whitfeld et al. (1982) was used.

Materials

(1) Deoxycytidine 5'[^{32}P] triphosphate (Amersham PB, 10205 3,000 Ci/mmol) (2) Random primers (prepared by Professor D B Scott (Massey University, NZ) from herring sperm DNA) (3) DNA polymerase I (Klenow fragment from Boehringer Mannheim) (4) Deoxyribonucleoside triphosphates (Sigma): dATP (20 mM), dGTP (20 mM), dTTP (20 mM) (5) HaeIII restriction enzyme (New England Biolabs) (6) Sephadex G-50 (Pharmacia, fine grade) (7) 0.25M Na_2EDTA, pH 8.0 (8) Phenol (equilibrated with TE (50/20) buffer, pH 8.0) (9) Chloroform (10) TES (10/1/100) buffer pH 8.0: 10 mM Tris, 1 mM Na_2EDTA, 100 mM NaCl.

Method

DNA (0.25-1.0 μg) to be labelled, was digested with HaeIII restriction enzyme in a 25 μl reaction mixture for 30 min at 37°C. Random primers (100 μg) were added and the mixture was boiled for 2 min, then chilled rapidly on ice. Reagents were added in the order listed below: Sterilized distilled water, 2.5 μl; Universal (x10) buffer, 1.5 μl; dTTP, 1 μl; dATP, 1 μl; dGTP, 1 μl; [^{32}P] dCTP, 3 μl; DNA polymerase I, 1 μl. The mixture was then incubated at 37°C for 30 min, and the reaction was stopped by adding 2 μl of 0.25 M Na_2EDTA. The reaction mixture was extracted with phenol-chloroform and the aqueous phase was loaded onto a Sephadex G-50 (packed in a 1 ml syringe) column equilibrated with TES buffer pH 8.0. The syringe column was spun (1,085 **g**, 2 min) and the DNA peak was

collected and stored at -20°C. DNA probes were usually labelled to a specific activity of 1-5x10^8 cpm per μg of DNA.

2.19 DNA TRANSFER AND HYBRIDIZATION

The procedure described by Southern (1975) was used.

Materials

(1) 0.25 M HCl (2) 0.5 M NaOH, 0.5 M NaCl (3) 0.5 M Tris pH 7.4, 2.0 M NaCl (4) 20x SSC (3 M NaCl, 0.3 M sodium citrate) (5) 2x SSC
(6) Denhardts solution (10x) (Denhardt, 1966) contained: 1M Hepes buffer pH 7.0, 25 ml; 20 x SSC, 75 ml; herring sperm DNA (28 mg/ml), 0.32ml; E. coli tRNA (10 mg/ml), 1 ml; 20% (w/v) SDS, 2.5 ml; Ficoll (Sigma 70), 1 g; bovine serum albumin (BSA), 1 g; polyvinylpyrrolidone (PVP) 5 g; distilled water, 397 ml
(7) Nitrocellulose filter (Schleicher and Schull BA 85).

Method

Gels were stained and photographed. DNA was subjected to partial depurination by shaking the gels in 0.25 M HCl for 15 min (Wahl et al., 1979), then denatured by shaking for 15 min in 0.5 M NaOH, 0.5 M sodium chloride, followed by shaking for 15 min in 0.5 M Tris pH 7.4, 2.0 M sodium chloride to neutralize the gel. The gel was washed in 2x SSC for 2 min and placed on a blotting stand for DNA transfer to nitrocellulose filter. The nitrocellulose filter was removed after 16h, washed in 2x SSC for 5 min, blotted dry and baked in vacuo at 80°C for 2 h. The filter was sealed in a bag containing 20 ml (x10) Denhardts solution and pre-hybridized at 65°C for at least 2 h. Most of the liquid in the bag was drained off, and the boiled probe (10^7 cpm) was added. The bag was sealed and incubated in a shaking waterbath at 65°C overnight. After incubation the filter was removed from the bag, washed in three changes of 2x SSC, blotted dry, covered in "Gladwrap" and exposed to Ilford Curex X-ray film in the presence of Cronex intensifying screens for 1-5 d at -70°C. When filters were re-used, the probe DNA

was removed by washing for 20 min in 20 mM NaOH, 15 min in 0.5M Tris, pH 7.4, 2.0M NaCl and 15 min in 2x SSC.

2.20 DNA LIGATION

Materials

(1) Universal buffer (x10) pH 7.6: 0.06M Tris pH 7.6; 0.1M $MgCl_2.6H_20$; 0.1M ß-mercaptoethanol (2) 10 mM ATP (3) T4 DNA ligase (New England Biolabs).

Method

Ligation mixtures (20 μl) containing 0.2-1.0 μg of vector DNA, 0.2-0.5 μg of insert DNA, 2 μl of Universal buffer (x10) pH 7.6, 2 μl of 10 mM ATP, 1 μl of ligase (400 units) and an appropriate volume of sterile water to make up the final volume, were incubated at 4°C for 16 h. An aliquot (1 μl) was checked on a 1% minigel before and after the ligation. The ligated DNA was then used directly for transformation.

2.21 TRANSFORMATION OF E. COLI

A modified procedure of Cohen et al. (1972) was used.

Materials

(1) 60 mM $CaCl_2.2H_20$ (2) TEC buffer pH 8.0: 10mM Tris.HCl pH 8.0; 0.25 mM Na_2EDTA pH 8.0; 30 mM $CaCl_2.2H_20$.

Method

A stationary phase culture of E. coli strain HB101 was diluted 1/100 into 25 ml of LB and incubated at 37°C. The cells (OD_{600} = 0.4) were harvested by centrifugation (1,085 **g**, 5 min, 4°C). All subsequent steps were carried out at 4°C. The cells were resuspended in 10 ml of calcium chloride solution (60 mM) and incubated on ice for 20 min. These cells were then harvested and resuspended in 250 μl of

calcium chloride solution (60 mM) and used in transformation experiments. The transformation mixture containing 50 μl competent cells, 5 μl ligated DNA and 45 μl of TEC buffer was incubated at 4°C for 1 h, heat shocked at 42°C for 2.5 min, then diluted 1/10 in LB and incubated at 37°C for 2.5 h. The culture (100 μl) was plated on LB medium supplemented with the appropriate antibiotic and incubated at 37°C overnight.

2.22 CONJUGATION

Triparental crosses (Ditta et al., 1980) were performed by dispensing 100 μl of a late log culture of the recipient as a spot on a LB plate. After allowing to dry, a mixture (100 μl) of the donor plus the helper (50:50 v/v) was dispensed on top of the recipient spot and dried. The plate was incubated overnight at 37°C.

2.23 PHOSPHO-ß-GALACTOSIDASE ENZYME ASSAY

Phospho-ß-galactosidase activity was assayed using the chromogenic substrate o-nitro-phenyl-ß-D-galactoside 6-phosphate (ONPG-6-P) by the protocol of Inamine et al. (1986).

Materials

(1) 50mM sodium phosphate buffer, pH 7.0; 20mM NaH_2P0_4, 30mM$Na_2HP0_4.7H_20$ (2) toluene-acetone (1:9 v/v) (3) 24mM o-nitro-phenyl-ß-D-galactoside 6-phosphate (ONPG-6-P) in sodium phosphate buffer (4) 0.5M Na_2C0_3.

Method

The E. coli gene library was plated out for single colonies onto LB agar plates, supplemented with tetracycline, and incubated overnight at 37°C. One thousand colonies were selected from these plates and were replated onto LB plates prior to testing. Cells from each culture were grown to mid-exponential phase in LB containing lactose (20 g/l) followed by centrifugation. The cells were dispersed

into microtitre tray wells in 25 μl of 50mM sodium phosphate buffer and were then permeabilized by the addition of 1.5 μl of toluene-acetone (1:9 v/v), followed by incubation at 37°C for 15 min. After the addition of 25 μl of 24mM ONPG-6-P, incubation was continued for 15 min before adding 50 μl of 0.5M Na_2C0_3. Wells containing cells with phospho-ß-galactosidase activity were visibly yellow. Two positive controls (Streptococcus lactis H1, and DRI4477 containing pLZ600) and a negative control (E. coli HB101) were used every time.

2.24 ß-GALACTOSIDASE ENZYME ASSAY

ß-galactosidase activity in permeabilized cells was assayed using the chromogenic substrate o-nitro-phenyl-ß-D-galactoside (ONPG) by the protocol of Miller (1972).

Materials

(1) Z buffer; 60mM Na_2HP0_4, 40mM $NaH_2P0_4.2H_20$, 10mM KCl, 1mM $MgS0_4.7H_20$, 50mM ß-mercaptoethanol (2) Triton-X (10% w/v) (3) Sodium dodecyl sulphate (SDS) (0.1% w/v) (4) o-nitro-phenyl-ß-D-galactoside (ONPG), 4 mg/ml stock in Z buffer (5) 1 M Na_2C0_3

Method

Exponential-phase cells (0.5 ml), grown in supplemented M-56 medium, were added to an equal volume of Z buffer; two drops of Triton-X solution and one drop of sodium dodecyl sulphate (SDS) were added. The mixture was vortexed for 10 sec and incubated for 5 min at 28°C. ONPG (200 μl) was added and the assays were timed until a pale yellow colour developed. The reaction was stopped by addition of 450 μl of 1M Na_2CO_3. The cells were removed by centrifugation (12,000 rpm, 2 min, 20°C) and the amount of o-nitro-phenyl (ONP) produced was determined by measuring the absorbance at 420 nm. The activity was measured as Units of ß-galactosidase = 1000 x OD_{420}/time x vol x OD_{600}.

2.25 PREPARATION OF BACTERIOPHAGE LAMBDA 467

In order to prepare stock of lambda 467 bacteriophage the procedure of de Bruijn and Lupski (1984) was used with modifications.

Materials

(1) SM buffer: 100 mM NaCl, 200 mM Tris.HCl (pH 7.5), 100 mM $MgS0_4$

(2) Chloroform.

Method

Serial dilutions of lambda 467 lysate were made and incubated at 37°C for 20 min with mid-exponential E. coli K802 cells. 3 ml of melted top agar was added, mixed well, plated onto YM plates and incubated overnight at 37°C. An overnight culture of E. coli K802 grown in YM was diluted one-hundred fold into the same medium and the culture was grown with vigorous shaking at 37°C until the OD_{600} was approximately 0.8 (cell density of 10^9/ml, 3 h). Four fresh plaques from the prepared plates were picked using a sterile toothpick tip and added to 500 μl of the mid-exponential phase K802 cells, mixed well and incubated at 37°C for 20 min without shaking to allow phage adsorption. To the 500 μl of cell and bacteriophage mixture, 7.5 ml of melted lambda top agar (equilibrated at 45°C) was added, mixed well and plated in 2.5 ml aliquots per YM agar plate. An equivalent amount of K802 cells with no bacteriophage was plated as a control. The plates were incubated at 37°C for 5-7 h, and inspected for lysis by comparison with the control plate. When confluent plaques were visible on the plates containing the lambda-infected cells, the top agar was scraped off into a sterile screw top centrifuge tube; 2.5 ml of SM buffer was added and the mixture was vortexed thoroughly. Then 500 μl of chloroform was added and the mixture was again vortexed thoroughly, before being incubated on ice for 5 min. The lysate was then centrifuged (12,100 **g**, 4°C, 5 min) and the supernatant was decanted into a sterile glass screw top tube. Two drops of chloroform were added and the lysate was stored at 4°C. This lambda 467 bacteriophage lysate was titred using strain E. coli K802 as recipient.

Mid-exponential phase cells of K802 were infected with serial dilutions of the lysate and incubated for 20 min at 37°C, before plating out as described previously. The final titre of the lambda 467 lysate was 1.08 $x10^{10}$ pfu/ml.

2.26 MUTAGENESIS OF PLASMID DNA WITH TRANSPOSON Tn5

Tn5-mediated transposon mutagenesis was performed by a modification of the method of de Bruijn and Lupski (1984).

Materials

(1) Sucrose solution: 25% sucrose (w/v), 50 mM Tris.HCl (pH 8.0).

Methods

The plasmid to be mutagenised was transformed into strain HB101 and the culture was grown overnight in LB containing tetracycline (15 μg/ml). The presence and the correct structure of the plasmid in the strain was physically verified by isolating the plasmid by the method of Holmes and Quigley (1981), preparing restriction enzyme digests and by testing for ß-galactosidase activity on X-gal plates. The strain was then grown to saturation in 5 ml of YM medium containing tetracycline, at 30°C. The culture was diluted one-hundred fold into 5 ml of YM medium containing tetracycline (15 μg/ml) and was grown at 30°C for 3 h until the OD_{600} was 0.8. 1 ml of the culture (1 $x10^9$ cells) was mixed with the Tn5-carrying lambda 467 bacteriophage to give a multiplicity of infection of 1.0 (100 μl of stock lysate) and the mixture was incubated at 30°C for 2 h. Aliquots (150 μl) of the infected cells were then plated onto LB plates containing kanamycin (50 μg/ml) and tetracycline (15 μg/ml) and incubated at 30°C for 48 h. The cells were then washed off the plates using sucrose solution (5 ml) to each plate (the colonies were gently re-suspended using a sterile glass spreading rod), and the cell suspension was decanted into a sterile centrifuge tube. The plates were washed with additional sucrose solution. The cells were collected via centrifugation (5,900 **g**, 4°C, 10 min) and re-suspended in the sucrose solution (8 ml of solution/ml of cells plated). The

plasmids contained within the cells were isolated using the clear lysate method, and were then purified using phenol-chloroform. The plasmid DNA so isolated was used to transform strain HB101 competent cells and grown overnight on LB plates containing tetracycline and kanamycin. The resulting transformant colonies were screened for the presence of the Tn5 in the resident plasmid by the method of Holmes and Quigley (1981).

2.27 MAXI-CELL ANALYSIS

Maxi-cells were prepared by a modification of the method by Sancar et al. (1979).

Materials

(1) cycloserine, 10 mg/ml (2) [^{35}S]-methionine.

Methods

E. coli cells were grown overnight in minimal medium (supplemented with 1% casamino acids, glucose, Ap) at 37°C. The cells were inoculated into fresh medium at 1:20 dilution and grown at 37°C with vigorous shaking until the density was 30 Klett Units with filter$_{54}$. The UV source was pre-warmed for 15 min, and then 10 ml of culture was transferred to a sterile glass petri plate and irradiated with UV to give a total dose of 6 Joules (30 s in the laminar flow hood). The cells were diluted 1:1 with supplemented minimal medium and incubated in the dark for 37°C for 1 h. Cycloserine was added to a final concentration of 200 μg/ml and the cells were incubated overnight at 37°C with shaking. The cells were washed twice in minimal salts, and re-suspended in minimal medium (sulphate-free). After incubation at 37°C for 1 h to allow sulphate depletion, 5-20μCi of ^{35}S-methionine was added to each sample and incubated for 1 h to allow labelling.

2.28 PROTEIN GEL ANALYSIS

Protein gels were run to analyze the labelled protein products from maxi-cell analysis.

Materials

(1) Stock solution A (g/l); Acrylamide, 300; Bis-acrylamide, 8 (2) Stock solution B; 1.5 M Tris-HCl (pH 8.8) (3) Stock solution C; SDS 10% (w/v) (4) Stock solution D (g/l); Tris-HCl 30; Acrylamide, 400, Bis-acrylamide, 10. (pH 6.8) (5) Loading Buffer: 2.5 ml of Solution D; 4 ml of Solution C; Glycerol, 2 ml; ß-mercaptoethanol, 1ml; Bromophenol blue, 500 μl of 0.02% (w/v); 400 μl of water (6) Running buffer: 25 mM Tris-HCl; 190 mM glycine, SDS (0.1%: w/v)
(7) Separation gel (12.5%): 40% acrylamide, 18.75 ml; 1% bis-acrylamide, 6.35 ml; Solution B, 15 ml; Solution C, 600 μl; TEMED, 15μl; 10% ammonium persulphate, 240 μl; 19.14 ml water (8) Stacking gel (3% acrylamide, 0.08% Bisacrylamide pH 6.8). Solution A, 1ml; Solution C, 100 μl; Solution D, 2.5 ml; Temed, 10 μl; 10% ammonium persulphate 100 μl; 6.25 ml water (9) Staining solution: 50% methanol, 7% acetic acid, Coomassie Blue R250 (0.2% w/v) (10) Destaining solution: 5% methanol, 7% acetic acid.

Method

After labelling, the cells were pelleted, washed once with minimal salts and resuspended in 100 μl of gel sample buffer. The separation gel was poured, carefully covered with water and allowed to set. The water was removed and the stacking gel poured on top, the comb inserted, and the gel left to set. The samples were boiled for 2 min, spun briefly, and 20 μl of each sample was loaded onto a gel, and the gel was run until the dye was close to the gel bottom. The gel was stained in the staining solution for 30 min, and then de-stained overnight in the de-staining solution. The gel was then placed on blotting paper and dried (2 h, 80°C) before being exposed to X-ray film.

2.29 PREPARATION OF Bal31 DELETIONS

Bal31 digestion of the 5.2 kb EcoRI fragment was carried out as described below.

Materials

(1) 5x Bal31 buffer: 60 mM $CaCl_2.2H_2O$, 3M NaCl, 5 mM Na_2EDTA, 100 mM Tris-HCl (2) 100 mM $MgCl_2$ (3) 3 M Sodium acetate (4) Ethanol (5) Phenol:chloroform (1:1 v/v).

Method

The plasmid containing the insert was linearized at a site adjacent to the start of the insert using the required enzymes, as described in Section 2.9. The DNA was purified with phenol:chloroform, and precipitated with ethanol overnight at -20°C. The DNA solution was centrifuged (12,100 **g**, 4°C, 30 min) and re-suspended to a concentration of 1 μg/μl. Determination of the concentration of Bal31 required for digestion was identified by the following method. DNA (2 μ1) was added to 50 μl of 1x Bal31 buffer, 12mM $MgCl_2$, and 9 μ1 aliquots were dispensed into five Eppendorf tubes. Addition of 1 μl of enzyme solution at various dilutions of 1/4, 1/8, 1/16, 1/32, and 1/64 were made, the solutions mixed, and the DNA digested at 30°C for 30 min. SDS stop dye was added immediately, and the DNA solution was examined by gel electrophoresis. The Bal31 concentration used was the one which had entirely digested the DNA within the 30 min. The number of bases removed/ end/min was calculated to allow identification of the time at which samples were to be removed. A large digest using 30 μg of DNA was carried out under identical conditions, removing samples at the various appointed time intervals. The samples were immediately purified using phenol:chloroform, and precipitated with ethanol overnight at -20°C. After centrifugation, each DNA sample was re-suspended in 100 μl water, and digested with a second restriction enzyme which released the insert form the vector as described in Section 2.9. A 2 μl aliquot was run on a gel to check for complete digestion and the samples were precipitated overnight at -20°C using ethanol. The samples were then centrifuged

(12,100 **g**, 4°C, 30 min) and re-suspended in 30 μl of water. A ligation was set up using an appropriately prepared M13mp18 or mp19 vector (SmaI/HindIII or SmaI/EcoRI).

2.30 LARGE-SCALE PREPARATION OF M13 RF DNA

M13 RF DNA for cloning of Bal31 deletions was prepared by the following method.

Materials

(1) STET buffer: 8% (w/v) sucrose; 5% (v/v) Triton X-100; 50 mM Na_2EDTA pH 8.0, 50 mM Tris-HCl, pH 8.0 (2) Lysozyme, 10 mg/ml (3) isopropanol (4) CsCl.

Method

A single colony of JM101 was inoculated into 2 ml of YT medium and grown overnight at 37°C. Two 10 ml aliquots of fresh medium were inoculated with 1:100 dilution of the overnight JM101 culture and grown at 37°C for 3 h. One tube was inoculated with a single plaque of mp18 and mp19 and grown overnight at 37°C. Two 500 ml flasks of YT medium were inoculated with 5 ml of a fresh overnight culture of JM101, and grown with vigorous shaking for 3 h. 1 ml aliquots of the overnight JM101/phage broth were centrifuged (12,100 **g**, 5 min, 20°C) and the respective supernatants were inoculated into the flasks and grown up at 37°C for 6 h. The cells were harvested by centrifugation (3,640 **g**, 10 min, 4°C), re-suspended in 34 ml of STET buffer, and transferred to a 250 ml flask. Lysozyme (2 ml of 10 mg/ml) was added and the cell suspension was brought to the boil over a bunsen flame, then transferred to a boiling water bath for 40 s. The viscous solution was transferred into SS34 tubes and centrifuged. (12,100 **g**, 4°C, 25 min). The supernatant was transferred to a fresh tube and an equal volume of isopropanol was added, and held at -20°C for 15 min before centrifugation (12,100 **g**, 4°C, 10 min). The isopropanol was drained off, and the pellet was washed with ethanol and dried. The DNA pellet was resuspended in 9 ml of TE (10:1) and purified through a

cesium chloride gradient. The phage band was removed and dialysed extensively with TES (10:1:100).

2.31 PREPARATION AND TRANSFORMATION OF E. COLI CELLS

Preparation and transformation of E. coli cells with M13 RF DNA was performed by the following method.

Materials

(1) 50 mM $CaCl_2$ (2) X-gal:(5-bromo-4-chloro-3-indoyl- ß-D-galactopyranoside). 20 mg/ml in dimethyl formamide (3) IPTG (isopropyl-ß-D-thiogalacto-pyranoside), 24 mg/ml in water.

Method

An overnight culture of JM101/JM109 was inoculated from a single colony and grown in YT medium with shaking at 37°C overnight. The overnight culture was subcultured (1:100) into 25 ml of fresh YT medium and grown at 37°C with vigorous shaking for 2 h. (The overnight culture was also subcultured (1:100) into 5 ml of fresh medium for use in plating.) The cells were harvested in 30 ml Corex tubes by centrifugation (3,640 **g**, 10 min, 4°C) and the supernatant was poured off. The cells were gently re-suspended in 10 ml of ice-cold 50 mM $CaCl_2$ and left on ice for 20 min. The cells were re-centrifuged, the supernatant poured off, and the competent cells were re-suspended in 2.5 ml of 50 mM $CaCl_2$ and stored on ice. To transform, 200 μl of competent cells were added to the ligation mix in a glass tube, gently mixed, and incubated on ice for 40 min. The mixture was then heat shocked for 3 min at 42°C. To each glass tube, 20 μl of X-gal, 20 μl IPTG, 200 μl of exponentially growing cells and 3 ml of molten top agar were added, vortexed briefly, and poured onto a fresh MM56 (glucose and thiamine supplemented) plates. The top agar was allowed to set for 30 min, the plates inverted and incubated at 37°C overnight. To check the transformation efficiency of the cells, 5ng of M13mp18 RF DNA was also transformed.

2.32 HYBRIDIZATION OF M13 PLAQUES

M13 recombinant plaques containing inserts of the 5.2 kb EcoRI fragment were isolated by the method of Benton and Davis (1977).

Materials

(1) Millipore 0.45 μM filters (Millipore Corporation, Bedford, MA,USA).

Method

A nitrocellulose filter was placed on the plate containing the recombinant and parental M13 plaques and left for 5 min to obtain a good plaque print. The plate and the filter were marked in such a way as to allow identification of any positive hybridised plaques. The filter was carefully removed, air dried, and baked in vacuo at 80°C for 2 h. The pre-hybridization and hybridization were performed as described in Section 2.19.

2.33 PREPARATION OF RECOMBINANT M13 SINGLE STRANDED (SS) DNA TEMPLATES

M13 templates were prepared for sequencing by the following method.

Materials

(1) Phage precipitation solution: 20% PEG 6,000 (w/v), 2.5 M NaCl (2) TE 10:1: 10ml 1 M Tris-HCl, 400 μl 250 mM Na_2EDTA pH 8.0 (3) Phenol, Tris equilibrated (4) Chloroform (5) 3 M-Sodium acetate (6) Ethanol.

Method

2 ml of YT medium was inoculated with a single colony of JM101/JM109 and incubated with shaking overnight at 37°C. The culture was sub-cultured 100-fold into fresh medium and 1.5 ml aliquots were dispensed into test tubes. Individual colourless plaques were picked with a toothpick tip, inoculated into the test tubes, and incubated at 37°C for 5-6 h with vigorous shaking. The cultures were

transferred into a 1.5 ml microfuge tube and the cells were harvested by centrifugation (12,000 **g**, 5 min, 20°C). The supernatant was carefully removed into a fresh tube and 200 μl of the phage precipitation solution was added, mixed and allowed to stand for 15 min at room temperature. The tubes were centrifuged (12,000 **g**, 5 min, 20°C) and the supernatant was removed by aspiration. The tubes were then re-spun for 1 min and all remaining liquid was removed by aspiration. The bacteriophage pellet was re-suspended in 100 μl of TE (10:0.1) and incubated at 65°C for 10 min to aid re-suspension. Phenol (50 μl) was added and the mixture was vortexed for 15 s. The mixture was allowed to stand for 15 min and vortexed again for 10 s before being centrifuged (12,000 **g**, 5 min, 20°C). The aqueous phase was carefully removed and transferred to a fresh microfuge tube. To the supernatant, 10 μl of 3M sodium acetate and 250 μl of ethanol was added, mixed and stored overnight at -20°C.

The tubes were centrifuged (12,000 rpm, 10 min, 4°C) to precipitate the ss template DNA, the ethanol was drained off, and the DNA was dried under vacuum. The DNA pellet was re-suspended in 30 μl of TE (10:0.1) and a 2 μl aliquot was checked by gel electrophoresis to determine the DNA concentration. The template DNA was stored at -20°C until sequenced.

2.34 PREPARATION AND ASSEMBLY OF GLASS PLATES

To prevent formation of bubbles when pouring the gel, the plates were cleaned thoroughly with a scourer, washed well, and wiped over with ethanol. The inside of the top plate was siliconised with Sigmacote, and the plates were wiped with ethanol and polished dry with paper towels. After checking that the plates were spotlessly clean, the spacers were placed on the sides of the back plate and the top plate was placed on top of the back plate. The plates were wrapped with "Sleek" tape ensuring no gaps. The sides were then clamped and a line was the drawn 4 mm from the top of the top plate as a guide for the comb.

2.35 POURING AND RUNNING OF SEQUENCING GELS

Sequencing gels were run as described below.

Materials

(1) Acrylamide mix: 288 g urea, 34.2 g acrylamide, 1.8 g bis-acrylamide (2) TEMED (3) 10% ammonium persulphate (4) Sharkstooth combs (5) 1X TBE buffer: 89 mM Tris-HCl, 8.9 mM boric acid, 12 mM Na_2EDTA, pH 8.6 (6) Formamide dye mix: 100 ml deionized formamide, 0.03 g xylene carbol FF, 0.03 g bromophenol blue, 20 mM Na_2EDTA (7) Fixing solution: 10% glacial acetic acid (v/v), 10% ethanol (v/v).

Method

The urea and acrylamide were dissolved in water, and 5 g of Amberlite (MB-3) resin was added to deionise the solution. The resin was removed by filtration through a sintered glass funnel (#2) with suction. 60 ml of 10X TBE was added and the volume was made up to 600 ml with water. The mixture was de-gassed and stored in a dark glass bottle at 4°C.

To 40 ml of acrylamide mix, 24 μl of TEMED and 240 μl of 10% ammonium persulphate were added, mixed well, and the gel poured, ensuring that no air bubbles were present. The back of the sharkstooth comb was inserted up to the guide line and the gel was allowed to set. Surface acrylamide was washed off and the tape was cut off the bottom of the plates. The comb was flushed with water, removed, the well flushed out, and the comb was re-inserted with the sharksteeth just inserting into the gel. The plate was then clamped to the gel box together with an aluminium plate, and TBE buffer was added to the top and bottom reservoirs. The wells were marked off and flushed out immediately before loading. The labelled templates were thawed, and boiled for 3 min. 3 μl of each sample was loaded into the well using a tip kept hard against the glass plates, avoiding air bubbles. The gel was run at 1500V, 30mA, 45W. The gel containing the short runs

was run for 2.5 h, the long gel was run for 4 h. The power was turned off, the plates unclasped and the tape removed. The plates were carefully prised apart and the gels were fixed in the fixing solution for 10 min. The plates were removed from the fixing solution, drained and blotted dry with handi-towels. The gel was transferred to a piece of blotting paper, and dried for 45 min at 80°C under vacuum on a gel drier. The gel was exposed to X-ray film for 1-3 d, then developed.

2.36 PREPARATION OF THE DEOXYNUCLEOTIDE (dNTP) AND DIDEOXYNUCLEOTIDE (ddNTP) MIXES

(1) 20 mM dNTP stock solutions

20 mM dATP (25 mg in 2.34 ml water)
20 mM dGTP (25 mg in 2.47 ml water)
20 mM dCTP (25 mg in 2.67 ml water)
20 mM dTTP (25 mg in 2.59 ml water)

(2) 0.5 mM dNTP stock solutions

1 in 40 dilutions of above stocks in distilled water

(3) 10 mM stocks of ddNTPs

10 mM ddATP (4 μmol in 400 μl)
10 mM ddGTP (4 μmol in 400 μl)
10 mM ddCTP (4 μmol in 400 μl)
10 mM ddTTP (4 μmol in 400 μl)

(4) Dilutions of 10 mM ddNTP stock solutions

0.02 mM ddATP (2 μl/ml)
0.1 mM ddGTP (10 μl/ml)
0.08 mM ddCTP (8 μl/ml)
0.5 mM ddTTP (50 μl/ml)

(5) Preparation of d/ddNTP solutions

Nucleotide solutions	Mixture (μl) A	G	C	T
0.5 mM dGTP	200	10	200	200
0.5 mM dCTP	200	200	10	200
0.5 mM dTTP	200	200	200	10
TE (50/1)	50	50	50	50
SUBTOTAL	650	460	460	460
0.02 mM ddATP	650	-	-	-
0.1 mM ddGTP	-	376	-	-
0.08 mM ddCTP	-	-	306	-
0.5 mM ddTTP	-	-	-	460
TOTAL	1300	836	766	920

(6) Chase mixes. 0.125 mM of each of the dNTPs. Prepared by adding equal volume of the 0.5 mM dNPT stocks.

2.37 PREPARATION OF [^{35}S]-LABELLED M13 SINGLE STRANDED TEMPLATES

Recombinant M13 ss templates were sequenced using a modification of the method of Sanger et al. (1977), adapted for use with [^{35}S-dATP] (Biggin et al., 1983).

Materials

(1) [^{35}S]-dATP, deoxyadenosine 5'-α-[^{35}S] thiotriphosphate (2) X10 annealing buffer: 100 mM Tris-HCl pH 8.5, 100mM $MgCl_26H_20$ (3) Universal primers: M13 17mer 3' TGACCGGCAGCAAAATG 5' at 2.5ng/μl. Tn5 17mer primer 3' GTTCATCGCAGGACTTG 5'at 2.5 ng/μl (4) DNA polymerase (Klenow fragment) (5) d/ddNTP termination mixes (6) Chase mix.

Method

A reaction mixture cocktail was prepared for four templates as follows: 5 μl 10x annealing buffer, 5 μl 17mer primer and 15 μl of water. An aliquot (5 μl) was

dispensed into four 500 μl Eppendorf tubes containing 5 μl of ss template. The tubes were incubated at 80°C for 5 min in a hot block and then allowed to cool slowly to room temperature (30 min). The d/ddNTP termination mixes were thawed on ice, and 9 μl of each mix was placed into a separate Eppendorf tube containing 1.2 μl of [^{35}S]-dATP. DNA polymerase (0.5 μl) was added to each cooled template mix. A grid of 16 Eppendorf tubes was set up:

A	A1	A2	A3	A4
G	G1	G2	G3	G4
C	C1	C2	C3	C4
T	T1	T2	T3	T4
Template	**1**	**2**	**3**	**4**

and 2.2 μl of annealed template mix was added to the corresponding tubes followed by 2.2 μl of the d/dd termination mixes. The tubes were spun for 30 s to collect the liquid in the bottom of the tubes, and incubated at 37°C for 20 min. The chase mix (1 μl) was added to each tube, centrifuged briefly, and incubated further for 15 min at 37°C. The labelled templates were frozen at -20°C until ready for use on a sequencing gel.

2.38 SEQUENCING WITH THE 'SEQUENASE' KIT

The 'Sequenase' sequencing kit (Tabor and Richardson, 1987) was also used in the latter stages of sequencing exactly as recommended by the manufacturers (USB, Cleveland, Ohio, USA).

2.39 ASSEMBLY AND ANALYSIS OF SEQUENCE DATA

Compilation of the sequencing data, analysis of the compiled sequence to determine ORF's and comparison of sequence data to the known ß-galactosidase genes performed using the programs available through the University of Winsconsin Genetics Computer Group (UWGCG) sequence analysis software package installed on a VAX 750 (Devereux et al., 1984).

CHAPTER 3: EFFECT OF VARIOUS BACTERIOLYTIC ENZYMES ON THE LYSIS OF CLOSTRIDIUM ACETOBUTYLICUM

3.1 INTRODUCTION

Bacterial cell walls are composed of a complex of polymers, the principal ones being peptidoglycan, polysaccharides, teichoic acids, teichuronic acids, lipids and proteins (Cummins, 1973; Strominger and Tipper, 1974; Schleifer and Stackebrandt, 1983).

3.1.1 Bacterial Cell Walls

The Gram-positive cell wall is a thick (30-100nm) and usually homogenous layer (Ghusyen et al., 1968; Shockman and Barrett, 1983). In contrast, Gram-negative bacteria have a thinner though more complex multi-layered cell wall (10nm) with an obvious double-membrane structure (Ghusyen et al., 1968; Nikaido, 1973; Glauert et al., 1976).

Table 3.1: General differences in cell wall components between Gram-positive and Gram-negative bacteria (Cota-Robles & Stein, 1973)

Component	Gram-positive	Gram-negative
Peptidoglycan	50-90%	1-10%
Amino sugar	10-30%	1-10%
Lipids	0- 2%	10-20%
Amino acids (no.)	3- 4	14-18
Polysaccharides	+	-
Teichoic acids	+	-
Teichuronic acids	+	-
Protein	-	+
O-antigen complex	-	+
Lipoprotein	-	+

There are major chemical and structural differences between the cell walls of Gram-positive and Gram-negative bacteria (Table 3.1). The low peptidoglycan content of Gram-negative bacteria usually facilitates lysis. By contrast, it is often difficult to lyse Gram-positive bacteria.

3.1.2 C. acetobutylicum cell envelope

The cell envelope of C. acetobutylicum NCIB8952 has been studied by Cho and Doy (1973). The general morphology shows that the envelope consists of a slightly wavy cell wall which is triple layered, and an asymmetric cell membrane (Fig.3.1). The triple layered cell wall consists of an undulating outer electron dense layer (4-6nm) separated from the inner dense layer (22-24nm) by a light layer (4nm). The outer electron dense layer of the cell wall, shown to be differentiated into subunits separated by electron light materials, is not always visible in stationary phase cells.

The cell membrane is also triple-layered and asymmetric. The outer layer is thick (6-8nm), very electron dense and separated by a light layer (3-4nm) from the inner dense layer (2.5-3nm). The outer electron-dense layer can be clearly separated from the inner in plasmolysed cells, which suggests that the outer layer of the cell membrane may represent a fusion of both the original layer of the cell membrane and the innermost component of the cell wall (Cho and Doy, 1973). The difference in electron density between the various layers suggests that these may differ in composition. Various structures exist between the cell wall and the cell membrane. Often there is a thickening of the outer electron-dense layer of the cell membrane, and the presence of vesicles or tubules. These structures are shown to be connected with both the cell wall and the outer layer of the cell membrane.

Fig.3.1: The ultrastructure of the C. acetobutylicum cell wall and membrane (Cho and Doy, 1973).

Diagram illustrating the various layers of the bacterial cell wall and membrane, including the structures found between the cell wall and the cell membrane.

(1) thickening of the outer layer of the cell membrane
(2) vesicles
(3) tubules

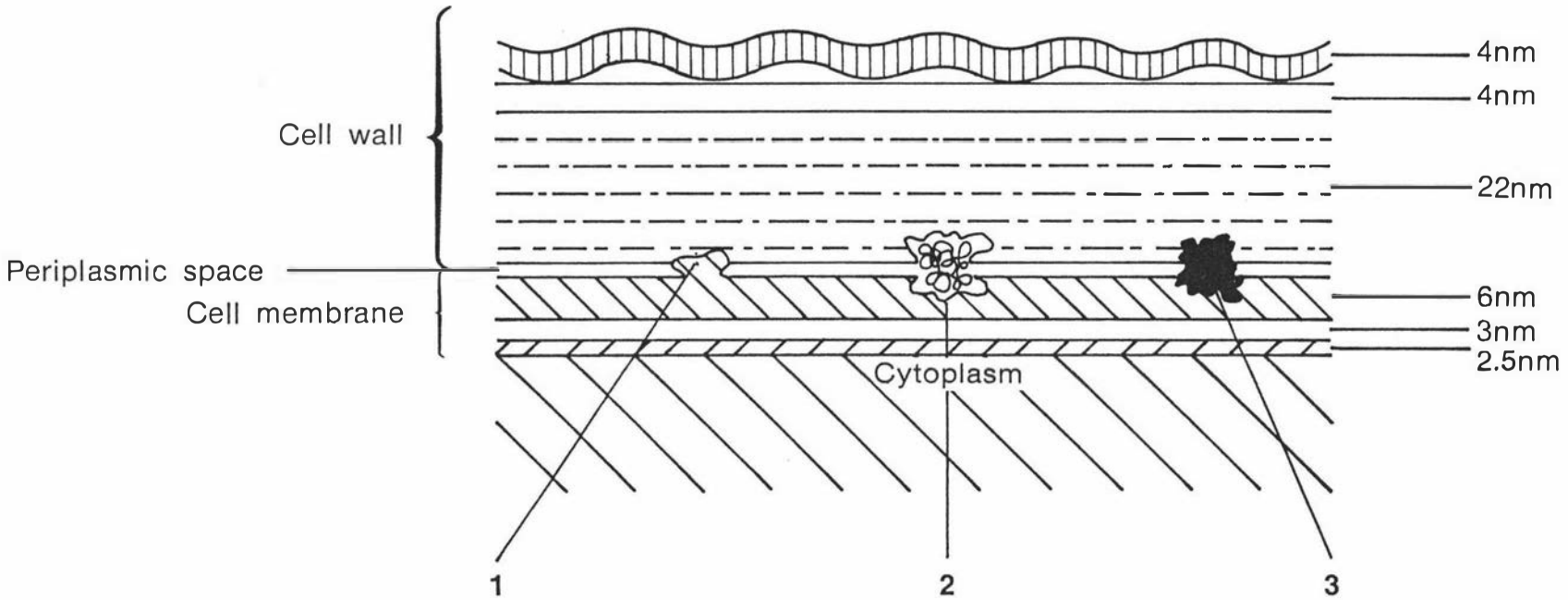
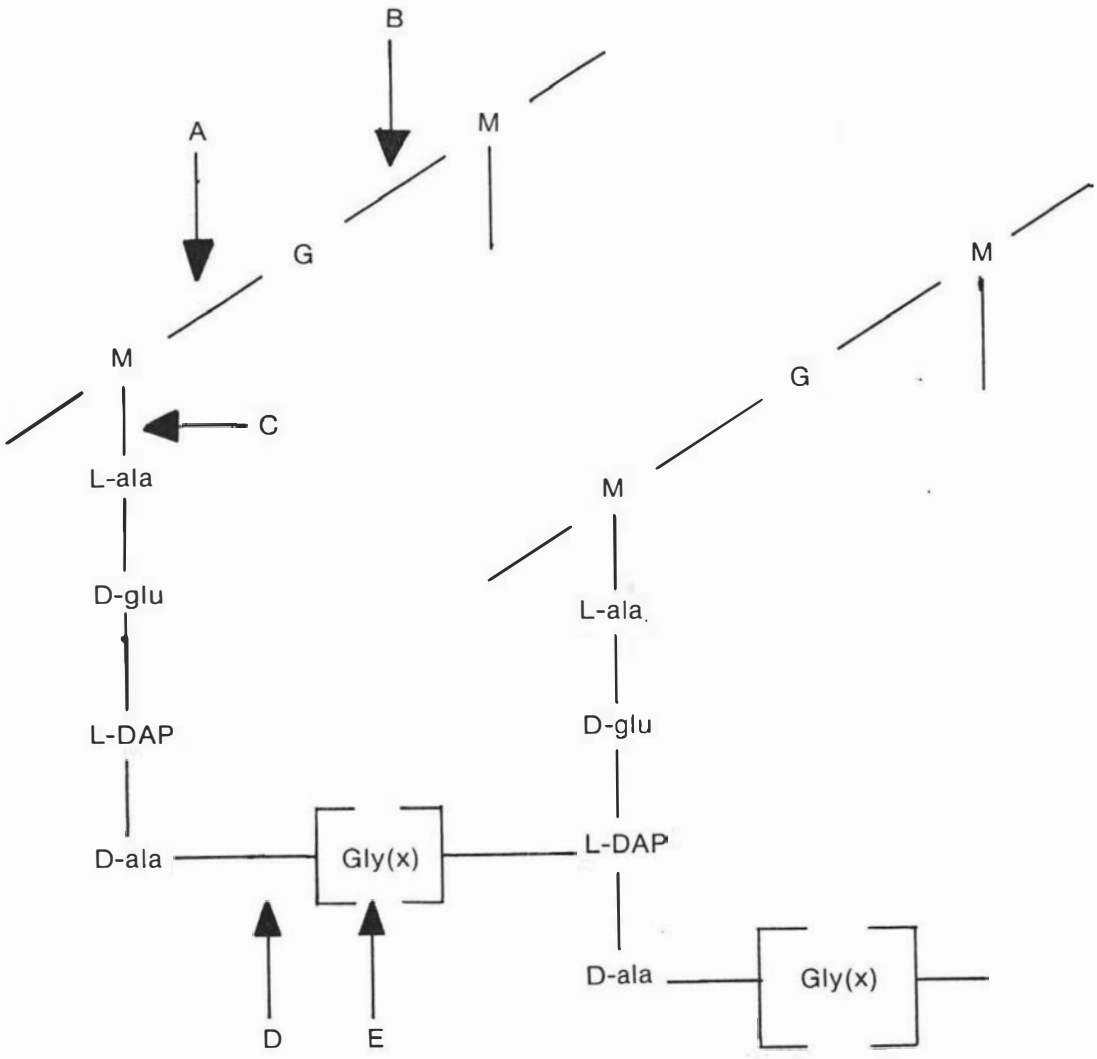

Cell wall
Periplasmic space
Cell membrane
4nm
4nm
22nm
6nm
3nm
2.5nm
Cytoplasm
1
2
3

Fig.3.2: The general structure of cell wall peptidoglycan and enzyme sites.

G (N-acetyl glucosamine) and M (N-acetyl muramic acid) which are linked by ß-1-4 linkages and alternate in the glycan strand, plus the units of the tetrapeptide and the cross-linking bridges between these tetrapeptide units. L-ala (L-alanine), D-glu (D-glutamic acid), L-DAP (L-diaminopimelic acid), gly (glycine) and D-ala (D-alanine). (Cota Robles and Stein, 1973).

Various enzyme sites are represented by the arrows.

Class I glycosidases (A,B)
Class II amidases (C)
Class III endopeptidases (D,E)
(Cummins, 1973)


B
A
M
G
M
M
C
G
L-ala
M
D-glu
L-ala
L-DAP
D-glu
D-ala
Gly(x)
L-DAP
D-ala
Gly(x)
D
E

3.1.3 Peptidoglycan

Peptidoglycan is the most important structural polymer, being responsible for the shape and structural integrity of the cell in face of its high internal osmotic pressure and hypotonic environmental conditions (Ghuysen and Shockman, 1973). It forms a large proportion of the cell wall in Gram-positive bacteria, but only a small proportion in Gram-negative bacteria. The peptidoglycan of Gram-positive bacteria is multi-layered, with great variation in composition and structural arrangement, especially in the peptide moiety. In contrast, the peptidoglycan present in Gram-negative bacteria is remarkably uniform, with only a few variations known.

Peptidoglycan is a polymer with a backbone of amino sugar chains (N-acetylhexosamines) cross linked through tetrapeptide side chains (Fig.3.2). The backbone of amino sugar chains is composed of alternating residues of N-acetylglucosamine and N-acetylmuramic acid linked by ß-1,4 glycosidic bonds. The carboxyl groups of lactic acid on C-3 of N-acetylmuramic acid are usually substituted by tetrapeptide units containing L- and D- amino acids. The peptidoglycan strands are further linked to each other by means of interpeptide bridges between the carboxyl group of the fourth amino acid (D-alanine) of one tetrapeptide to the amino group of the third amino acid in another tetrapeptide. The resulting carbohydrate skeleton of peptidoglycan forms a three dimensional net-like structure resembling that of chitin (Ghuysen and Shockman, 1973; Cummins, 1973; Strominger, 1975).

3.1.4 Bacteriolytic enzymes

Bacteriolytic enzymes are hydrolases that act on the peptidoglycan in bacterial cell walls and cause cell lysis. These have been isolated from both eukaryotic and prokaryotic organisms. Three classes of bacteriolytic enzymes are recognised, differentiated by their mechanisms of action on the basic structure

of the peptidoglycan (Fig. 3.2); glycosidases, amidases, and endopetidases. The first class, the glycosidases (acetyl muramidases), are divided into two further groups; the endo-acetylmuramidases and the endo-acetyl-glucosaminidases which act on the ß-1-4 linkages between N-acetyl muramic acid and N-acetyl glucosamine residues (A) and the alternate bonds between N-acetyl glucosamine and N-acetyl muramic acid residues (B), respectively. The second class, the N-acetylmuramyl-L-alanine amidases, specifically hydrolyse the linkages at the junction between the N-acetyl muramic acid and the first amino acid of the tetrapeptide (C). The third class contains the endopeptidases, which work by splitting the bond within the peptide bridges and their cross linkages (D,E) (Ghuysen, 1974; Ghuysen and Shockman, 1973).

The best characterized bacteriolytic enzyme of the first class is lysozyme, mucopeptide N-acetyl muramoylhydrolase,(E.C. 3.2.1.17). Albumin lysozyme efficiently lyses a wide range of Gram-negative bacteria but only a few species of Gram-positive bacteria from Micrococcus, Bacillus, Sarcina, and Pediococcus. It is not effective against such genera as Staphylococcus, Streptococcus and Clostridium (Ghuysen et al., 1968; Cota-Robles and Stein, 1973; Nakamura et al., 1973; Cords et al., 1974; Yokogawa et al., 1975; Horonouchi et al., 1977; Yu et al., 1984c).

Mutanolysin, an N-acetyl muramidase, belongs to the second class. It is produced by Streptomyces globisporus and has a pronounced lytic activity on cell walls of Streptococcus mutans and a vast array of other Gram-positive bacteria difficult to lyse with lysozyme (Yokogawa et al., 1975; Siegal et al., 1981).

The soil bacterium Achromobacter, has been found to produce a highly active bacteriolytic enzyme, achromopeptidase, with a broad specificity. It belongs to the third class of bacteriolytic enzymes since it hydrolyses peptide linkages of peptidoglycan. Like lysozyme, the enzyme causes lysis of Micrococcus, Bacillus, and Sarcina and in addition has an increased activity spectrum, lysing many other

Fig.3.3: Lysis results of C. acetobutylicum cells of various growth phases grown in the absence of glycine using lysozyme.

Cell lysis was monitored by the reduction in OD_{600} as described in Section 2.5.

A Early exponential phase
B Mid exponential phase
C Late exponential phase
D Stationary phase

Concentration of lysozyme:

● 0 mg/ml
+ 1 mg/ml
○ 2 mg/ml
× 5 mg/ml
□ 10 mg/ml

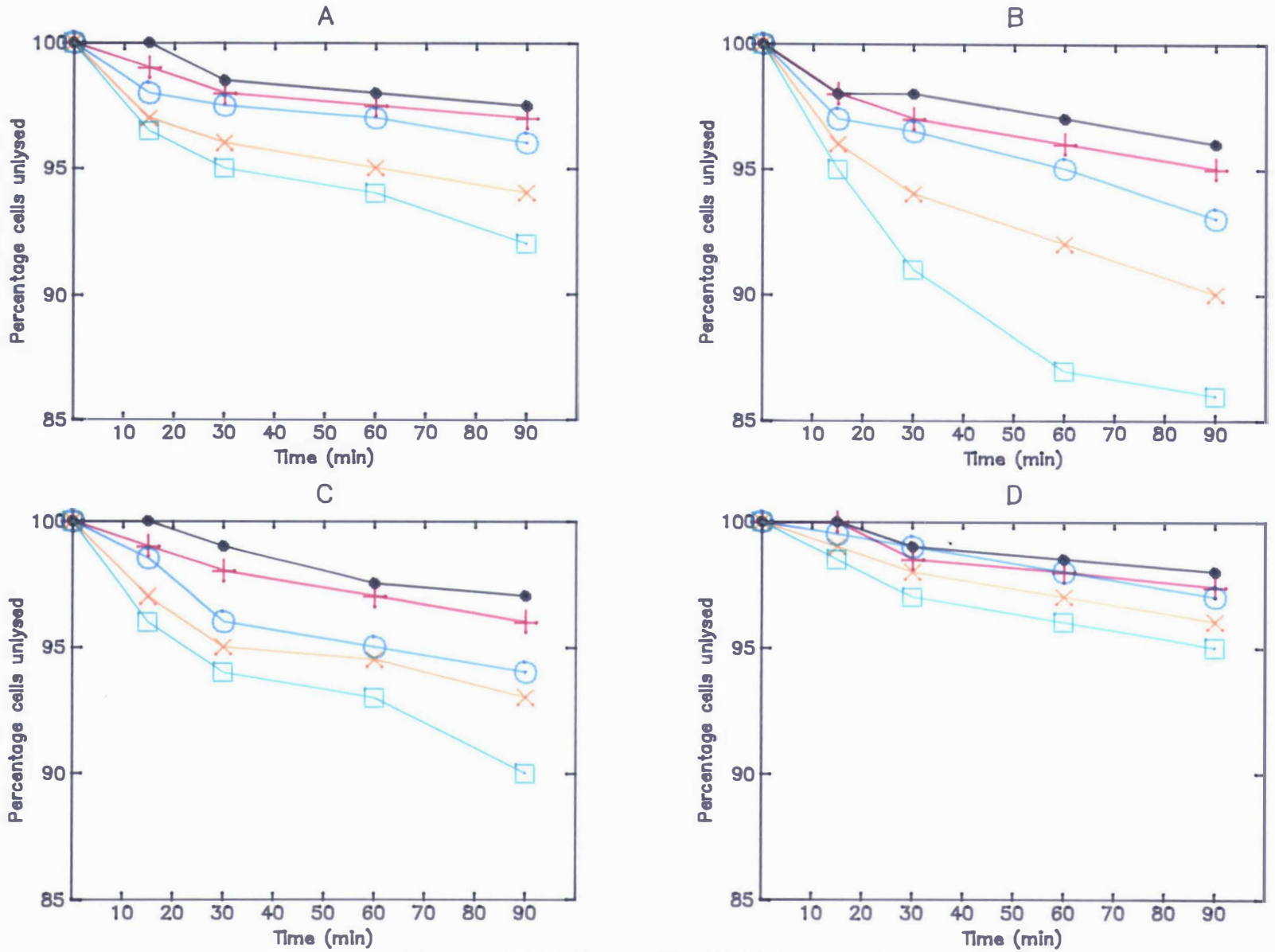

A
B
C
D
Percentage cells unlysed
100
95
90
85
10 20 30 40 50 60 70 80 90
Time (min)

Fig.3.4: Lysis results of C. acetobutylicum cells of various growth phases grown in the presence of glycine using lysozyme.

Cell lysis was monitored by the reduction in OD_{600} as described in Section 2.5.

A Early exponential phase
B Mid exponential phase
C Late exponential phase
D Stationary phase

Concentration of lysozyme:

- ● 0.0 mg/ml
- ┼ 0.5 mg/ml
- ○ 1.0 mg/ml
- × 2.0 mg/ml
- □ 5.0 mg/ml
- △ 10.0 mg/ml

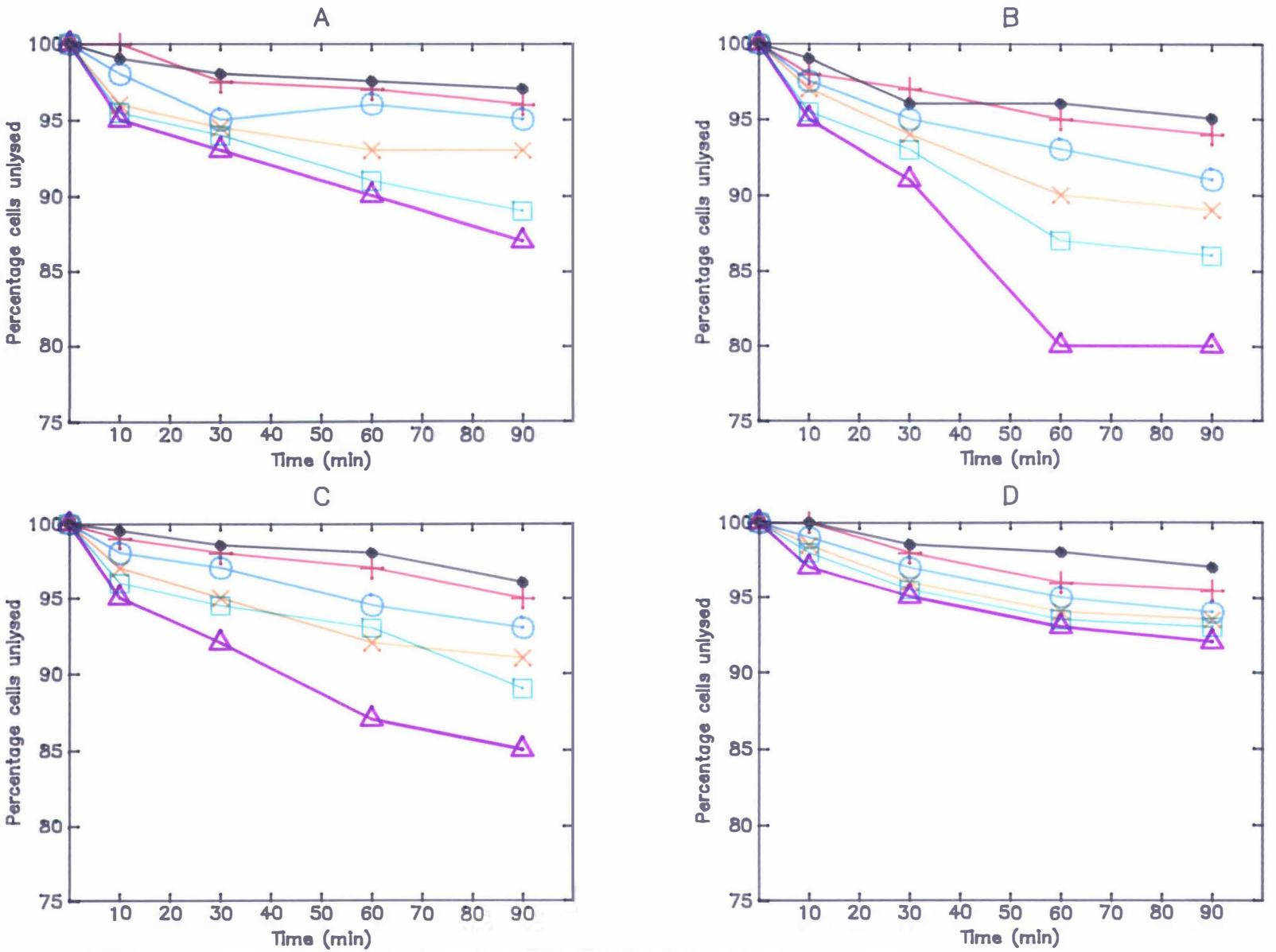
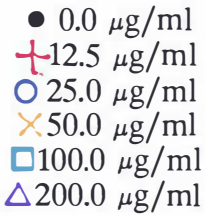

A
B
C
D
Percentage cells unlysed
Time (min)
100
95
90
85
80
75
10
20
30
40
50
60
70
80
90

Fig.3.5: Lysis results of C. acetobutylicum cells of various growth phases grown in the absence of glycine using mutanolysin.

Cell lysis was monitored by the reduction in OD_{600} as described in Section 2.5.

A Early exponential phase
B Mid exponential phase
C Late exponential phase
D Stationary phase

Concentrations of mutanolysin:



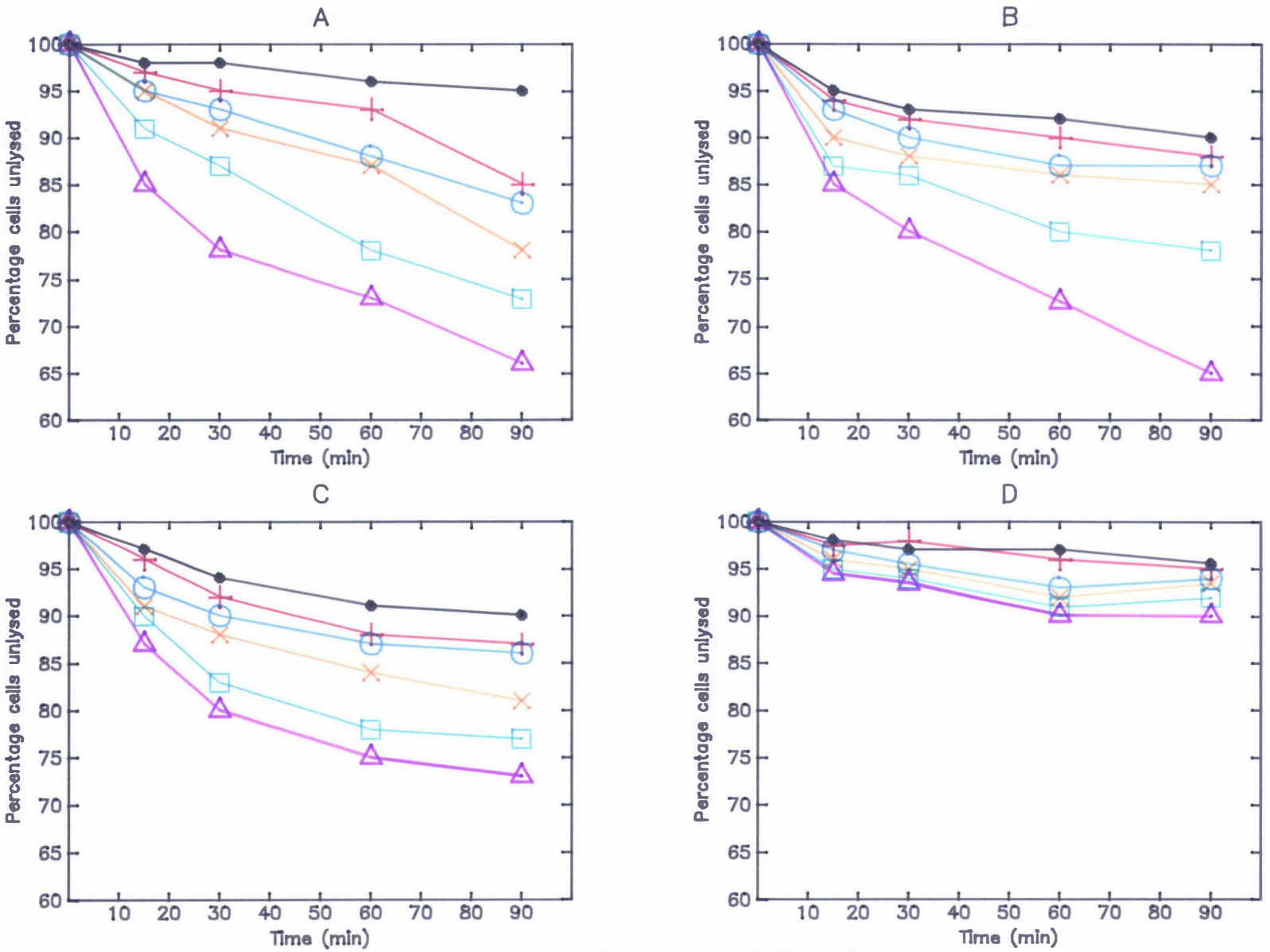

A
B
C
D
Percentage cells unlysed
Time (min)
100
95
90
85
80
75
70
65
60
10
20
30
40
50
60
70
80
90

Fig.3.6: Lysis results of C. acetobutylicum cells of various growth phases grown in the presence of glycine using mutanolysin.
Cell lysis was monitored by the reduction in OD_{600} as described in Section 2.5.

A Early exponential phase
B Mid exponential phase
C Late exponential phase
D Stationary phase

Concentrations of mutanolysin:

● 0.0 μg/ml
+ 5.0 μg/ml
○ 12.5 μg/ml
× 25.0 μg/ml
□ 50.0 μg/ml
△ 100.0 μg/ml
* 200.0 μg/ml

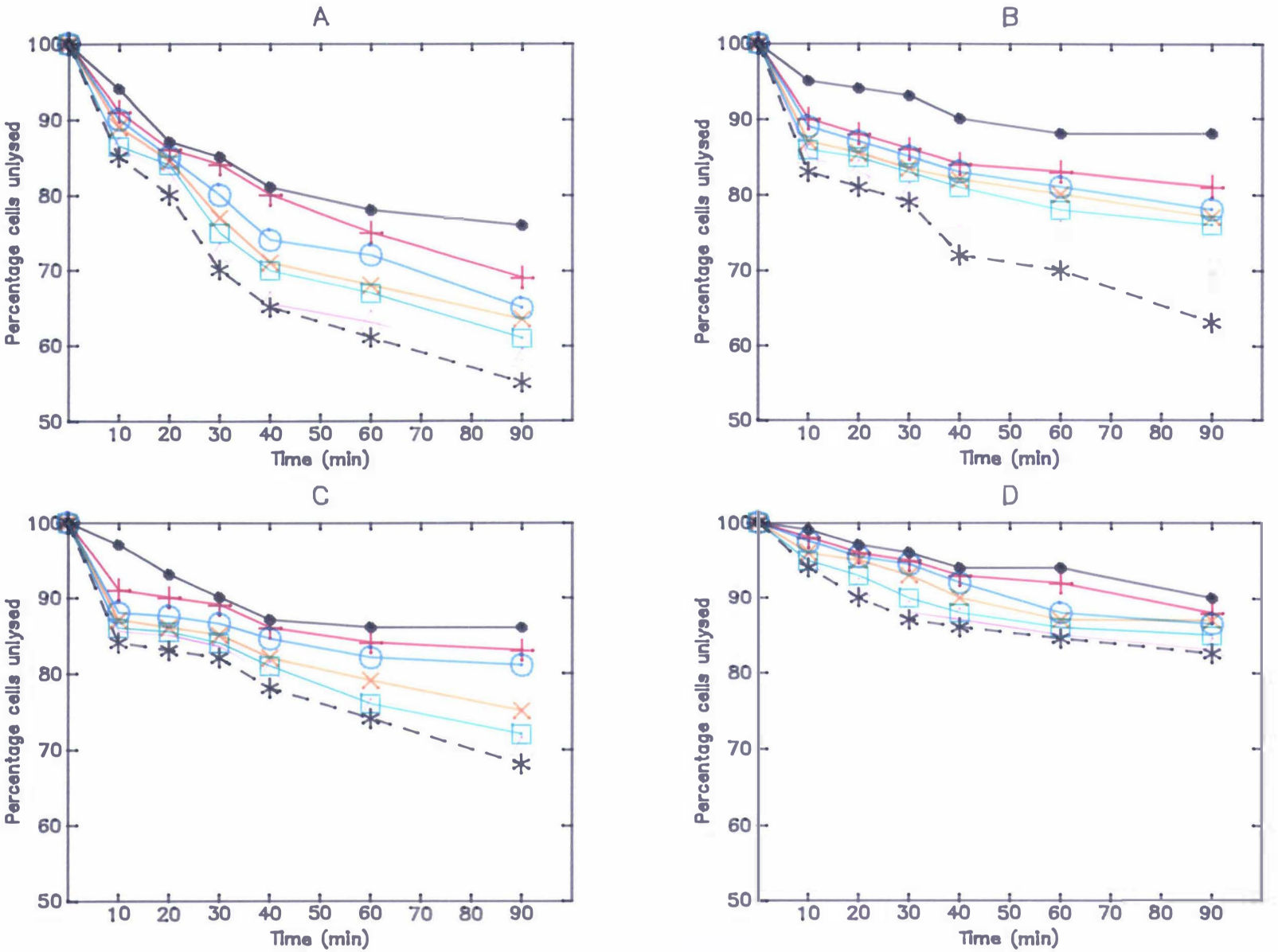
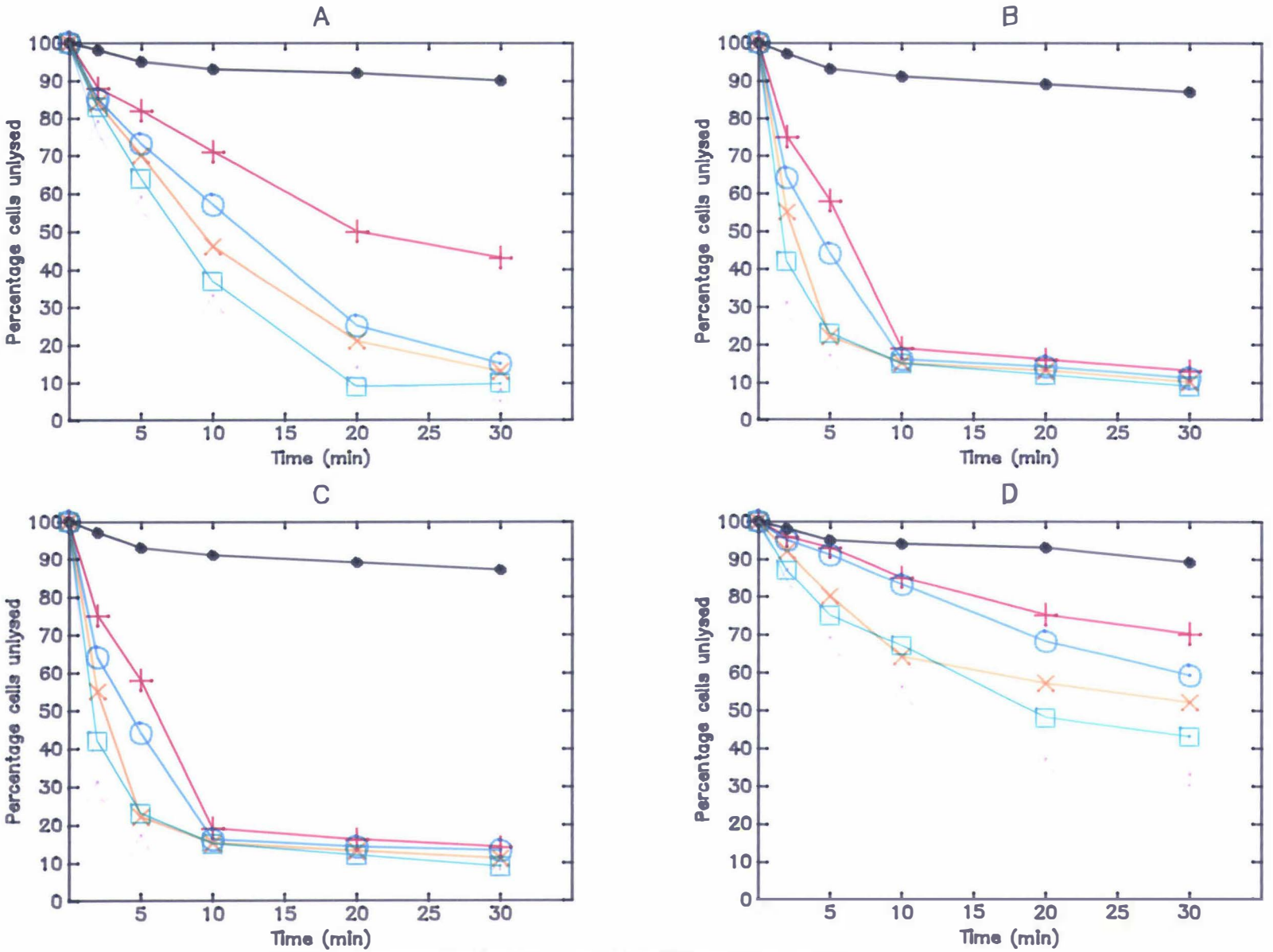

A
B
C
D
Percentage cells unlysed
Time (min)
100
90
80
70
60
50
10
20
30
40
50
60
70
80
90

Fig.3.7: Lysis results of C. acetobutylicum cells of various growth phases grown in the absence of glycine using achromopeptidase.
Cell lysis was monitored by the reduction in OD_{600} as described in Section 2.5.

A Early exponential phase
B Mid exponential phase
C Late exponential phase
D Stationary phase

Concentrations of achromopeptidase:

● 0U/ml
+ 25U/ml
○ 50U/ml
× 100U/ml
□ 200U/ml
△ 300U/ml


A
B
C
D
Percentage cells unlysed
Time (min)
0
10
20
30
40
50
60
70
80
90
100
5
10
15
20
25
30

Gram-positives, such as Staphylococcus, Streptococcus, Clostridium, and Leuconostoc which are normally highly resistant to lysozyme. In addition, it is reported to be active against some Gram-negative microorganisms (Horinouchi et al., 1977).

3.1.5 Aim

The aim of this phase of the project was to determine the effectiveness of different enzymes for use in the rapid and efficient, but controlled, lysis of cells of C. acetobutylicum.

3.2 RESULTS

The rates of cell lysis at different concentrations of lysozyme, mutanolysin and achromopeptidase were examined as described in section 2.5 using C. acetobutylicum P262 as a test organism. The effect of lysozyme on the lysis of early, mid, late exponential and stationary phase cells of C. acetobutylicum grown in CGM medium without glycine is shown in Fig. 3.3. Cells in the mid-exponential phase were the most susceptible to lysis. However, even at the highest lysozyme concentration of 10 mg/ml, lysis did not proceed beyond 15% after a 90 min incubation period. There was insignificant lysis of stationary phase cells. Using cells grown in the presence of glycine a 2-fold increase in lysis was observed (Fig.3.4). Again, cells in the mid-exponential phase were the most susceptible, although the maximum lysis was still only 20% after 90 min incubation at the highest lysozyme concentration of 10 mg/ml. The lysis of stationary phase cells was again insignificant.

The effect of mutanolysin on lysis of early, mid, late exponential and stationary phase cells grown without glycine is shown in Fig.3.5. Cells in the mid-exponential phase were the most susceptible to lysis, with 35% lysis in 90 min at a mutanolysin concentration of 200 μg/ml. There was insignificant lysis of stationary phase cells.

Lysis was, however, significantly higher than that achieved using lysozyme. Using cells grown in the presence of glycine a slight increase in lysis was observed (Fig. 3.6). Cells in the early to mid exponential phase were the most susceptible, the maximum extent of cell lysis being 43% after 90 min incubation using 200 μg/ml.

Achromopeptidase rapidly lysed the cells decreasing the OD_{600} over a relatively short time period (Fig.3.7). With early exponential phase cells, at an achromopeptidase concentration of 300U/ml, almost complete lysis was observed in 20 min, and appreciable lysis (43%) was obtained with 50U/ml in 10 min. The effect was even more pronounced with mid exponential phase cells. Extensive lysis (87%) occurred at an enzyme concentration as low as 25U/ml. With late exponential phase cells, cell lysis was still appreciable. With stationary phase cells, lysis still occurred (70% at an enzyme concentration of 300U/ml; 30% at 25U/ml).

3.3 DISCUSSION

In order to isolate native undegraded nucleic acid, gentle but rapid means of cell disruption is required. Since various groups have reported that for C. acetobutylicum cell lysis in any form was very difficult to achieve, it was decided to investigate the effects of various bacteriolytic enzymes in the hope of finding a reliable one. Lysis was performed using just the enzyme alone, without the addition of substances which could increase lysis efficiency such as ionic and non-ionic detergents or proteinases, and utilizing temperatures and pH values that were optimal for each enzyme. It was assumed that once enough peptidoglycan had been removed from the cell wall that the osmotic shock due to the low hypotonicity of the environment and the high internal atmospheric pressure would cause the cell to lyse.

3.3.1 Effect of glycine in the medium

The addition of amino acids to the growth medium has been utilized to aid the lysis of many different species of bacteria (Hammes et al., 1973). Growth in the presence of glycine apart from causing morphological changes and inhibiting growth, also exerts a destabilizing effect on the cell wall. This occurs by the incorporation of modified glycine-containing precursors into the cell wall, resulting in a high percentage of uncrosslinked mucopeptides. As a result hydrolysis by lysozyme is more efficient (Hammes et al., 1973; Cota-Robles and Stein, 1973). This method has been used by various groups working on various Clostridium species including C. acetobutylicum. For protoplast formation of C. perfringens, Stal and Blaschek (1985) found that growth in the presence of glycine increased the ease of lysis. Allcock et al. (1982) noted an enhancement of C. acetobutylicum P262 protoplast formation with glycine in the growth media, as did Lin and Blaschek (1984), who found that the presence of glycine was very dramatic on lysis of C. acetobutylicum SA-1. Reid et al. (1983) also used glycine to enhance protoplast formation. However, even though the addition of glycine did result in a doubling of the degree of lysis, giving comparable results to those obtained by Allcock et al. (1982) and Reid et al. (1983), the effect was not as dramatic as found by Lin and Blaschek (1984).

3.3.2 Culture ages

The importance of the physiological condition of the culture on cell lysis has been reported by various groups, especially with other Clostridium species (Urano et al., 1983; Strom et al., 1984; Stal and Blaschek, 1985; Weickert et al., 1986). It is noted in the article by Cota-Robles and Stein (1973) that bacteria from stationary phase were harder to lyse with lysozyme than actively growing cells. Urano et al. (1983) reported that isolation of plasmid DNA from various Clostridium species, including C. acetobutylicum, was difficult due to resistance of the cell wall to lytic action of lysozyme, especially in older cells. Lin and

Blaschek (1984) reported that lysis was affected by the growth stage of the bacteria, optimized at exponential phase cells. Strom et al. (1984) found that the physiological condition of C. botulinum cultures was very important. Most strains would not lyse by any method when using lysozyme once they had reached stationary phase. Weickert et al., (1986) also reported difficulty in lysing C. botulinum, and found exponentially growing cells easiest to lyse. Lee et al. (1987) also reported that it was essential to use cells from the exponential growth phase to achieve lysis. Cells harvested later were very resistant to lysis with lysozyme. Cells were therefore collected in various stages of growth as determined by the absorbance reading as well as phase contrast microscopy (Table 3.2).

The results reported here supported these observations; that there was a greater degree of difficulty of lysis with increasing culture age. Early, mid and late exponential phase cells could lyse to a certain degree, depending on the enzyme

Table 3.2: Morphology of C. acetobutylicum cells at various stages of growth.

Culture phase	Growth (h)	OD_{600}	Phase contrast microscopy of cells
Early exp.	3- 6	0.25 - 0.4	Long, thin dividing cells, slightly motile, no spores, slight gassing of medium
Mid exp.	8-10	0.65 - 0.75	Long thin cells in pairs and chains, dividing, highly motile, no spores, high gassing of medium
Late exp.	12-14	0.92 - 1.2	Single short fat cells, slightly motile, no spores, little gassing of medium
Stationary	24 h	2.25 - 3.00	Single short stubby cells, clostridial forms present, phase bright spores often visible, non-motile, little gassing of medium

used, whereas stationary phase cells were virtually unlysable, with the exception of experiments using achromopeptidase where significant lysis occurred at all cell stages.

In mid to late exponential cultures of C. acetobutylicum automatic lysing could occur as C. acetobutylicum cells are known to produce an autolysin (Barber et al., 1979; Allcock et al., 1981; Webster et al., 1981; Soucaille et al., 1987; van der Westhuizen et al., 1982).

However, this autolysin is activated by the presence of high concentrations of sucrose in the protoplast formation medium (Allcock et al., 1982; Reid et al., 1983). Stationary cells, on the production of spores might lyse, freeing the mature spore from the old cell wall. C. acetobutylicum cells stop growing very early upon induction of solvent production and cell differentiation, which causes inhibition of cell growth and cessation of normal cell division (Jones et al., 1982; Long et al., 1984a). This should be manifested by a decrease in obtainable cell lysis as observed by these results.

3.3.3 Peptidoglycan sensitivity and accessibility

There are a number of gross features of cell wall structure which are known to be involved in resistance to lysozyme. It is known that in Gram-positive bacteria, including C. acetobutylicum, the rigid cell wall component is considerably thicker than in Gram-negative bacteria and contains a far higher amount of peptidoglycan (Cho and Doy, 1973; Shockman and Barrett, 1983). Even though the Gram-positive cell wall lacks an outer membrane, it contains an almost endless variety of polysaccharides, frequently negatively charged and teichoic acids covalently linked to the peptidoglycan. The tightness of the net of peptidoglycan may also effect the susceptibility of the cell wall to enzyme action. This is shown by the effect of glycine in the medium which, as mentioned before, increased the degree of cell lysis.

The use of lysozyme was the least effective of the three enzymes tested, supporting the findings of Wako (1986), who reported that C. acetobutylicum showed little decrease in turbidity (10-40%) in 60 min or no appreciable lysis at all with lysozyme, whereas the use of achromopeptidase caused a marked decrease in turbidity of at least 80% in 10 min. Lee et al, (1987) used achromopeptidase to isolate plasmid DNA from various Clostridial species, including C. acetobutylicum and, although noting improved cell lysis, also found that its use resulted in a decreased yield of DNA. Urano et al. (1983) also used a mixture of lysozyme and achromopeptidase to achieve lysis of C. butyricum cells.

If the approach to the peptidoglycan network, which is the substrate, is hindered by other polymers, such as teichoic acids, highly negative charged polymers present in the cell walls of some Gram-positive bacteria can bind a highly basic protein such as lysozyme and prevent its activity. C. acetobutylicum is also known to produce an exopolysaccharide upon cells shifting to solvent production (Jones et al., 1982; Long et al., 1984a). This extracellular capsule is a polysaccharide composed of repeating sugar units thought to contain a high degree of acetylation. Apart from acting as a reserve carbohydrate store, this could also act as a barrier, preventing and binding basic enzymes.

The presence of 6-O-acetylated group on N-acetyl muramic acid is a known basis for lysis resistance. This is known in several Gram-positive bacteria, where some of the muramic acid residues are O-acetylated on C-6 or substituted by phosphodiester groups, which covalently link other wall polymers to the peptidoglycan (Ghuysen, 1974; Ghuysen and Shockman, 1973). Such resistance has been found in Streptococcus faecalis and M. lysodekiticus mutants with substitution by O-acetyl groups or by chemical acetylation of Bacillus megaterium walls (Ghuysen, 1974). There appears to be no report of the presence or absence of O-acetylation in C. acetobutylicum. Elimination of naturally occurring O-acetyl groups off the glycan chains and removal of teichoic acids

however does not render such peptidoglycans significantly susceptible to lysozyme, when compared with native cell walls. This indicates that some intrinsic property of the peptidoglycan must be involved in the resistance to lysozyme action.

Lysozyme performs a transglycosidase activity and its lytic activity is product inhibited (Ghuysen, 1974). In Gram-positive cells with their greater peptidoglycan content complete hydrolysis of all sensitive bonds may be hindered (Ghuysen et al., 1968).

The effect of achromopeptidase on cultures of all ages, when compared with that observed using lysozyme or mutanolysin is dramatic. The results are well within those cited by the manufacturer who found that at a concentration of 300U/ml 80% or more of C. acetobutylicum cells lysed in 10 min. Dramatic activity has also been noticed by Nakamura et al. (1973) on Micrococcus radiodurans, which, while virtually resistant to high concentrations of lysozyme, underwent a rapid decrease in absorbance, from an OD_{600} of 1.5 to less than 0.2 in 8 min using achromopeptidase at 110 μg/ml. However, the lysis was very hard to control and occurred far too quickly in a short period of time at the concentrations of achromopeptidase used to allow adequate control. Horinouchi et al. (1977), using the enzyme on Staphylococcus aureus, achieved total lysis at 37°C in 10 min using 1176U/ml of a highly concentrated bacterial suspension. It was also noted that a dramatic decrease in viscosity occurred far more rapidly than that noticed using lysozyme or mutanolysin, probably due to exogeneous DNase activity.

Because of the lack of control of achromopeptidase action and the noticeable degree of loss of viscosity, it was decided to use mutanolysin as, although its activity was slower than achromopeptidase, it provided better control of the lysis and had the added advantage of causing less noticeable loss of viscosity. Addition of anionic detergents, high EDTA concentrations and proteinase K would be expected to increase the degree of lysis. The fact that mutanolysin

was stable at the higher temperature of 50°C compared with lysozyme and achromopeptidase was also an advantage as temperature is known to inactivate some DNAses. C. acetobutylicum is reported by various groups to produce significant levels of DNAse activity (Efstathiou and Truffaut, 1986; Johnson and Frances, 1975).

3.4 SUMMARY

Early to mid exponential phase cells are easier to lyse than cells at late exponential or stationary phase. The latter are almost unable to be lysed by all enzymes tested except by achromopeptidase which resulted in significant lysis. It was also shown that achromopeptidase is far more effective than mutanolysin or lysozyme on cells whether grown in the presence or absence of glycine. Growth in the presence of 0.3% (w/v) glycine exerts a minor destabilizing effect on the C. acetobutylicum cell wall, and as a result, hydrolysis by lysozyme or mutanolysin is slightly more effective.

CHAPTER 4: PREPARATION OF A GENOMIC LIBRARY OF CLOSTRIDIUM ACETOBUTYLICUM NCIB2951

4.1 INTRODUCTION

Following the development of methods for cloning foreign DNA into plasmids, researchers began to investigate means of preparing gene banks representative of the entire genomic DNA both of prokaryotes and eukaryotes. Plasmid vectors have the theoretical capacity to accommodate fragments of unlimited size, but the transformation efficiency of plasmids containing greater than 10 kb of insert DNA is so low that their use in constructing genomic libraries is not practical.

Lambda vectors were constructed to overcome this problem, being able to incorporate into the non-essential region up to 20 kb of insert DNA. More recently, cosmids were developed by Collins and Hohn (1978) as vectors specifically designed to clone even larger DNA fragments up to 50 kb in length. An additional advantage of cloning larger inserts is the ease of identification of overlapping fragments from the original genome, as well as the isolation of intact genes and operons.

Two examples of cosmids are c2RB and pLAFR1. c2RB is a 6.8 kb cosmid containing an unique restriction enzyme site (BamHI) for the insertion of foreign DNA, flanked by EcoRI sites which facilitates rapid isolation for screening the insert DNA (Bates and Swift, 1982). To enable packaging in bacteriophage heads, this vector requires the insertion of an average fragment size of 44 kb. The vector, pLAFR1 (21.6 kb) carries the lambda cos site cloned into the unique BglII site of the broad host range vector pRK290 (Ditta et al., 1980). It encodes for tetracycline resistance and contains a unique EcoRI site for cloning of insert DNA (Freidman et al., 1982). It also possesses the RK2 replicon (mob+) which enables recombinant plasmids to be mobilised into other species as well as other strains of E. coli. The mobilization requires the presence of the helper plasmid pRK2013

which contains the RK2 transfer genes (tra+), and encodes for kanamycin resistance (Figurski and Helinski, 1979).

4.1.1 Aim

The aim of this phase of the project, therefore was to construct a genomic library of C. acetobutylicum.

4.2 RESULTS

4.2.1 Difficulties in preparation of the genomic library

As the first step in the construction of a genomic library, the isolation of genomic DNA was attempted as described in section 2.8.2. However, the viscosity of the lysate decreased almost instantaneously upon cell lysis, and resulted in the isolation of degraded DNA. In most instances, spooling of DNA from the ethanol precipitation step was virtually impossible. For this reason, the DNase activity of the various strains of C. acetobutylicum was investigated.

The various strains were tested for extracellular DNase activity on DNase agar plates, modified for the growth of Clostridium strains (section 2.2.2). A fermentable carbohydrate source was included (lactose), and cysteine was added to reduce the oxygen-reduction potential. The percentage of agar was also increased (4%) to curb the "blowing" seen due to the production of gas during active growth. C. acetobutylicum strains were activated and grown anaerobically overnight in CMM at 30°C. Cells were then streaked onto the modified DNase plates and incubated overnight at 30°C. A strain of C. perfringens was also plated out as a negative control, while a known extracellular DNase producer, S. aureus (ATCC9144), was used as a positive control on an unmodified DNase agar plate.

Fig.4.1: Detection of extracellular DNase activities in various strains of C. acetobutylicum.
DNase activity was measured as described in Section 2.2.

1 S. aureus ATCC9144
2 C. perfringens CW92
3 C. acetobutylicum P262
4 C. acetobutylicum NCIB2951
5 S. aureus ATCC9144
6 C. perfringens CW92
7 C. acetobutylicum NRRL594
8 C. acetobutylicum NRRL598
9 C. acetobutylicum NRRL2490
10 C. acetobutylicum ATCC824

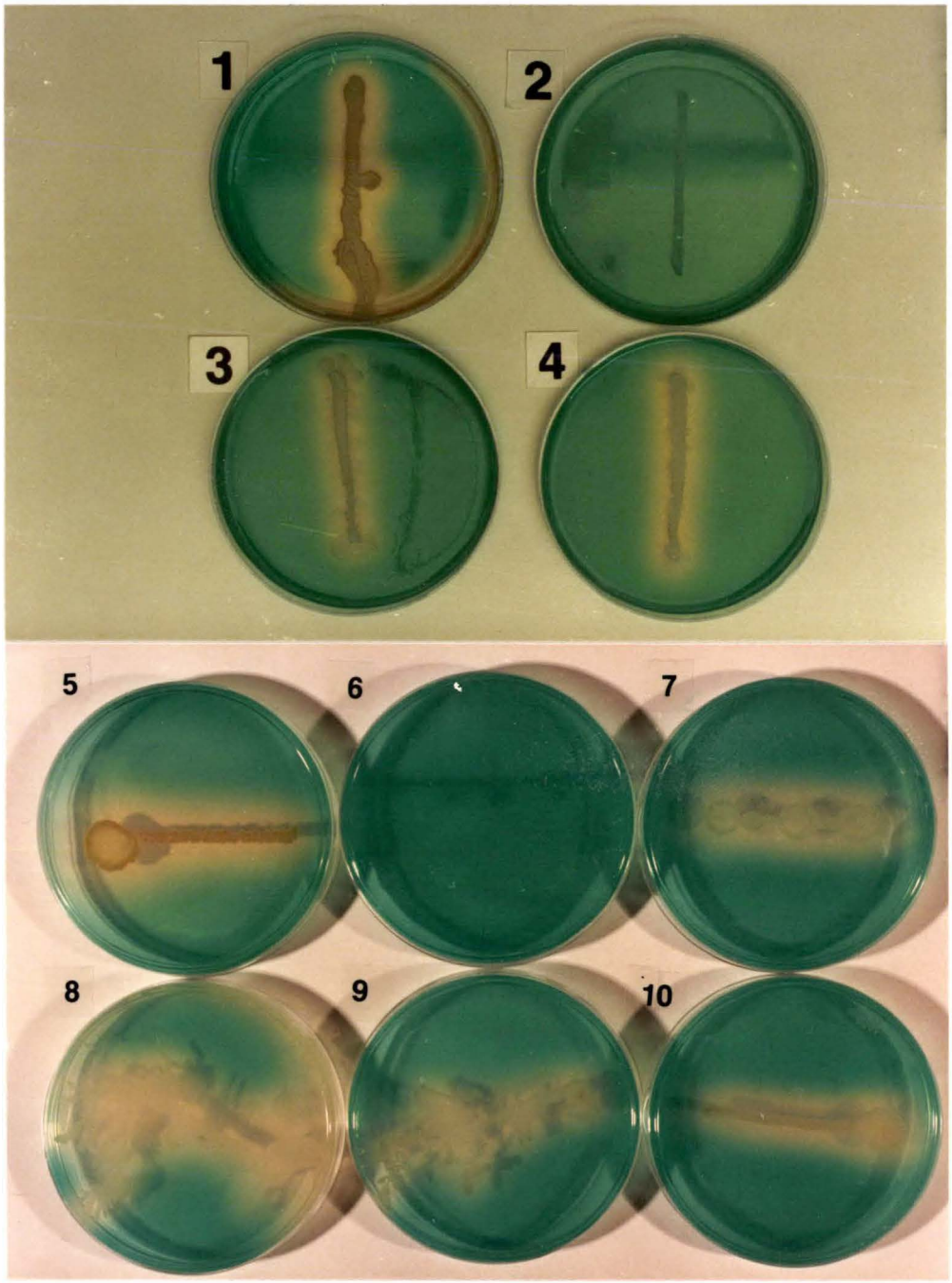
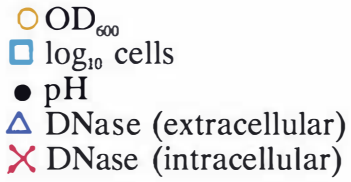
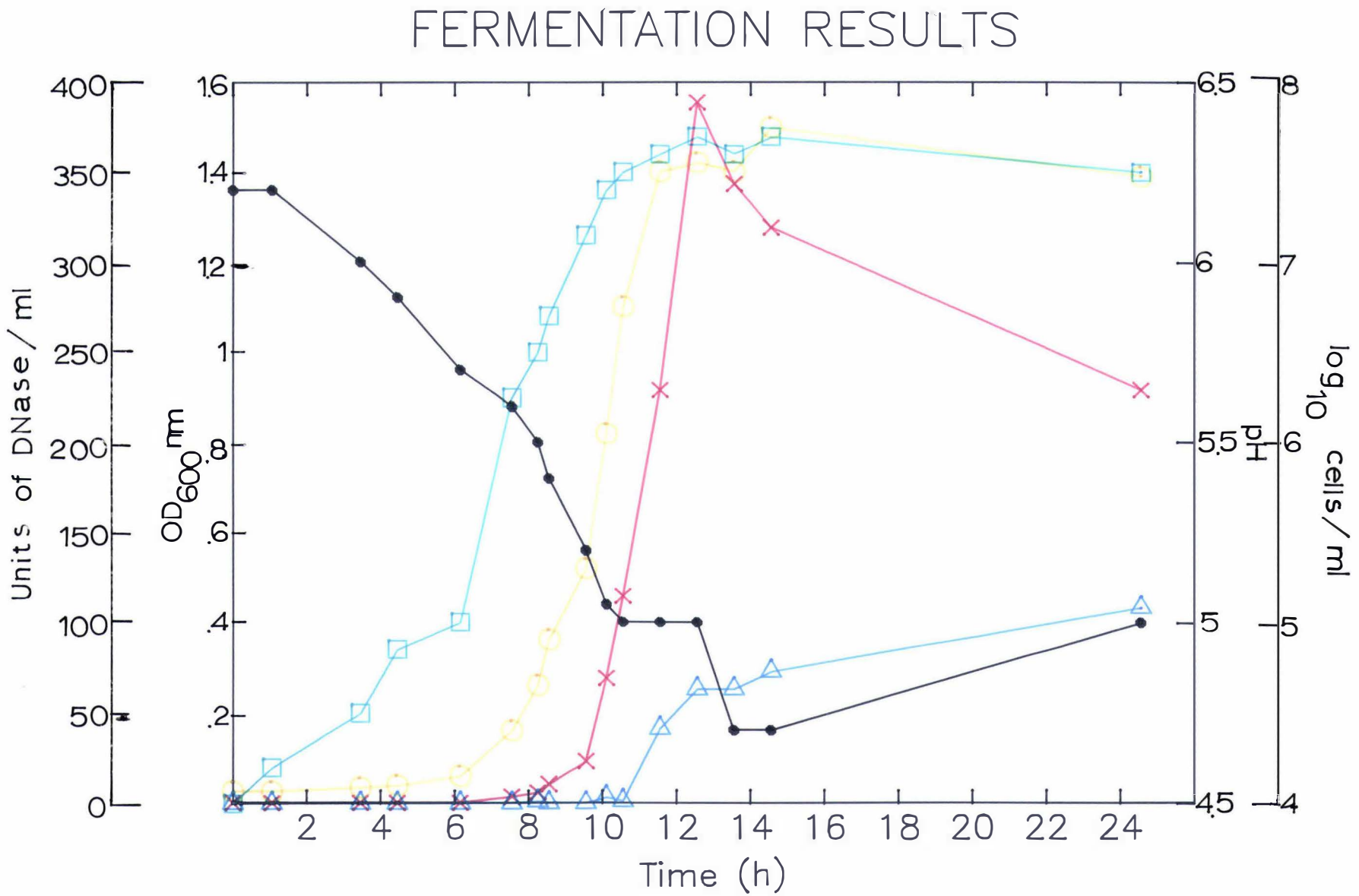

1
2
3
4
5
6
7
8
9
10

Fig.4.2: Levels of intracellular and extracellular DNase activity in C. acetobutylicum NCIB2951 during fermentation.
DNase activity was monitored as described in Section 2.6.1.




FERMENTATION RESULTS
Units of DNase / ml
400
350
300
250
200
150
100
50
0
OD600 nm
1.6
1.4
1.2
1
.8
.6
.4
.2
pH
6.5
6
5.5
5
4.5
log10 cells / ml
8
7
6
5
4
2
4
6
8
10
12
14
16
18
20
22
24
Time (h)

All six strains of C. acetobutylicum tested showed high levels of extracellular DNase activity (Fig.4.1). The levels produced were comparable to that observed in S. aureus ATCC9144 and showed that C. acetobutylicum produced significant levels of extracellular DNase activity at some stage during their growth cycle.

Further experiments were carried out to determine when the extracellular DNase was produced during the fermentation growth cycle, and to determine if any significant levels of intracellular DNase were produced. A batch fermentation was performed, using strain NCIB2951 as described in section 2.6. The pH, cell population and cell morphology were measured at various time intervals during the fermentation run, and an aliquot was removed at these intervals for measurement of the DNase content as described in section 2.6.1. Results showed that extracellular DNase activity was absent during the early and mid-exponential phases (Fig.4.2), but significant activity was detected at late-exponential phase and thereafter. Cells from early to mid-exponential phase also did not have any detectable intracellular DNase activity. However, this was detected in the late-exponential phase and stationary phase cells, and preceded detection of extracellular DNase activity. The intracellular activity was significantly higher than the extracellular activity (Fig. 4.2).

In view of these results (and including the earlier results that early to mid-exponential phase cells were easiest to lyse), the use of early- to mid-exponential phase cells was strictly adhered to for the isolation of chromosomal DNA from strains of C. acetobutylicum.

Theoretically, this protocol should have minimised degradation of DNA. However, further attempts to isolate undegraded chromosomal DNA from early or mid-exponential phase cells still provided evidence of a DNA degradation (stability) problem. Various substances known to inhibit or inactivate DNase activity were included during the cell lysis procedure in an attempt to prevent the DNA degradation. These included diethyl pyrocarbonate, proteinase K,

thermolysin, sodium dodecyl sulphate, guanidinium isothiocyanate, sodium perchlorate and SLS. However, no beneficial effects were observed. Further, performing the lysis at 55°C (a temperature known to inactivate some DNases) did not cause any appreciable difference compared to lysis at a lower temperature (37°C).

Isolation of C. acetobutylicum DNA in the presence of the chaotropic agent guanidinium isothiocyanate also did not increase the ability to isolate undegraded chromosomal DNA. The method, however, produced a large yield of high molecular weight DNA from E. coli. The same effect was seen using sodium perchlorate, another chaotropic agent. These results led to the conclusion that some intrinsic factor other than nuclease was responsible for the observed DNA degradation.

In a further attempt to identify the cause and alleviate the problem, a set of experiments was carried out where different samples of isolated chromosomal DNA and lysate DNA from several strains of C. acetobutylicum were tested for DNAse activity. In some cases, pBR322 DNA was incorporated as a marker. The samples were placed in a sterile microfuge tube and incubated at 4°C for 1-7 d prior to analysis by gel electrophoresis.

The results are summarized in Table 4.2. Chromosomal DNA from E. coli was relatively stable, as was linearized pBR322. However, significant degradation of the chromosomal DNA occurred in lysates of C. acetobutylicum, while the plasmid pBR322 DNA remained stable.

These results confirmed that DNAse was not entirely responsible for the instability of the DNA. DNA sensitivity to oxygen has been reported by Martin and Savage, (1988).

Table 4.1:Investigation of C. acetobutylicum DNA stability

DNA	Stability			
	1 day		7 days	
	total	pBR322	total	pBR322
E. coli	+		+	
pBR322		+		+
E. coli/pBR322	+	+	+	+
Lysates:				
C. acetobutylicum				
P262	-		-	
P262/pBR322	-	+	-	+
NCIB2951	-		-	
NCIB2951/pBR322	-	+	-	+
NRRL594	-		-	
NRRL594/pBR322	-	+	-	+
NRRL598	-		-	
NRRL598/pBR322	-	+	-	+
NRRL2490	-		-	
NRRL2490/pBR322	-	+	-	+
Isolated DNA:				
P262	+		-	
P262/pBR322	+	+	-	+
NCIB2951	+		-	
NCIB2951/pBR322	+	+	-	+
NRRL594	-		-	
NRRL594/pBR322	-	+	-	+
NRRL598	-		-	
NRRL598/pBR322	-	+	-	+
NRRL2490	-		-	
NRRL2490/pBR322	-	+	-	+

\+ DNA stable as measured by gel electrophoresis (section 2.12).

\- DNA unstable as measured by gel electrophoresis (section 2.12).

Isolation of chromosomal DNA was therefore carried out while keeping the cells as anaerobic as possible prior to lysis. This improved the ability to isolate undegraded stable DNA. One strain in particular, NCIB2951, was chosen for further use in the construction of the genomic library, as this strain was the most reliable for isolating high molecular weight DNA.

4.2.2 Preparation of the genomic library

The first cosmid chosen for the construction of a genomic library was c2RB. Chromosomal DNA from NCIB2951 was partially digested with Sau3A (section 2.9), and subjected to electrophoresis on a 1% Seaplaque gel (section 2.11). The band in the region of 40-45 kb was cut out and the DNA was extracted as described in section 2.16. This DNA was ligated into c2RB, and packaged as described in section 2.11. However, no recombinant clones were observed. Since controls using E. coli chromosomal DNA always resulted in the isolation of large numbers of clones constituting the genomic library, it seemed problems still existed with the clostridial DNA. Gel electrophoresis revealed that the size of the selected DNA seemed to decrease with time, and that ligation did not incorporate the insert DNA.

Thus, it appeared that although NCIB2951 chromosomal DNA could be isolated in a stable form, its molecular weight was not always high enough to enable easy isolation of the required fragment sizes. The decrease in the size of the DNA was probably due to damage that had occurred during the cell lysis which, although small, was significant enough to cause the size to decrease upon frequent handling, especially during ethanol precipitation steps.

A different cosmid (pLAFR1) was chosen. pLAFR1 was useful for the construction of this genomic library, due to the fact that less handling was required. Also, it was larger than c2RB, and hence required pieces of insert DNA which were smaller. Insert DNA was prepared by an adapted method

which minimized the handling of the DNA as described in section 2.11. This DNA was partially digested with EcoRI, (heated 65°C, 10 min to inactivate the enzyme), washed through a centricon filter, and ligated into EcoRI digested pLAFR1. The ligation mix was then packaged into bacteriophage heads (section 2.10 and 2.11) and plated overnight onto LB plates containing tetracycline. The resultant recombinant clones were washed off the plates with LB broth, washed and resuspended in glycerol. The genomic library was stored at -70°C. Approximately 3,500 recombinant clones were isolated from the C. acetobutylicum NCIB2951 genomic library. This number was significantly less than that achieved with an E. coli genomic library (9,700), prepared in an identical manner.

Packaging into the bacteriophage heads imposes a size selection on the recombinant inserts. The average insert size, using pLAFR1 is 23 kb. As the actual size of the C. acetobutylicum NCIB2951 chromosome is not known, an arbitrary value of 3,000 kb was set in order to determine the number of pLAFR1 clones required for the desired probability of a given sequence being present in the library. This was based on the known chromosome size of S. lactis (3,000 kb). The desired probability (0.99) of a unique DNA sequence being represented in the genomic library, was calculated using the equation of Clark and Carbon (1979):

$$N = \frac{\ln(1-P)}{\ln(1-1/n)}$$

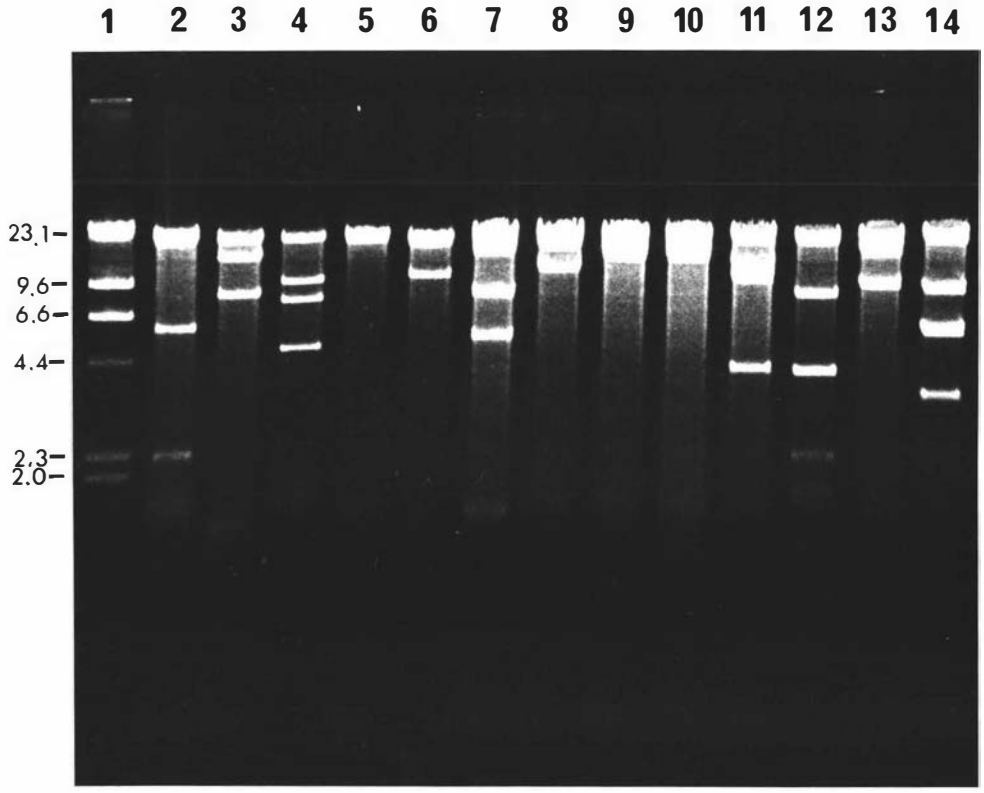
where P = the probability of a given sequence being in a genomic library (0.99); N = the number of clones required in the clone bank; and n = the size of the genome relative to the average size of the cloned fragment.

$$N = \frac{\ln(1-0.99)}{\ln(1-1/130.43)} = \frac{-4.605}{-7.696\times10^{3}} = 598 \text{ clones.}$$

Therefore, based on the above formula, the 3500 recombinant clones achieved should be representative of the entire genome of C. acetobutylicum. If the

Fig.4.3: Analysis by gel electrophoresis of a random selection of cosmid clones from the genomic library of C. acetobutylicum NCIB2951. Sizes are in kilobases.

Lane 1 : Lambda HindIII size standards
Lane 2-14 : EcoRI digests of random pLAFR1 cosmids


1
2
3
4
5
6
7
8
9
10
11
12
13
14
23.1
9.6
6.6
4.4
2.3
2.0

genome is as large as E. coli (4,100 kb), 820 clones would be required to maintain 99% probability of a given sequence being present.

Several of the recombinant clones from the genomic library were chosen at random, and the cosmid DNA was isolated (section 2.7.3), digested with EcoRI (section 2.9), and analyzed by gel electrophoresis (section 2.11) to determine the percentage and size of insert DNA (Fig.4.3). In all cases, a fragment the size of pLAFR1 (21.6 kb) was present, and in 12 out of the 13 cases other fragments, presumably insert DNA, were also produced. One out of the 13 clones (Fig.4.3, lane 5) yielded only a pLAFR1-sized fragment. Since pLAFR1 alone is of insufficient size for packaging, this recombinant clone was probably a cosmid-dimer. The total length of insert DNA in the 12 recombinant clones was measured, and found to vary from 13 to 35 kb, the mean being 22 kb. The total length of DNA being packaged appears to be in the range of 34.4 to 51.6 kb. This corresponds well with the reported values for size selectivity in lambda packaging in vitro (Hohn, 1979).

4.3 DISCUSSION

A number of bacterial genera (e.g. Staphylococcus, Serratia, Pseudomonas, Bacillus, Clostridium) are known to produce intracellular and extracellular DNase (Streitfold et al, 1962; Nestle and Roberts, 1969; Miller and Clark, 1976; Akrigg, 1978; Mihelc et al., 1978; Scurlock and Miller, 1979; Rood et al., 1978; Solberg et al., 1981; Luczak et al., 1985). A significant problem which arises in preparing DNA from these bacteria is that these DNases, if not inhibited, rapidly degrade DNA during preparation, giving variable and inconsistent results (Johnson and Frances, 1975; Urano et al., 1983; Lin and Blaschek, 1984; Blaschek and Solberg, 1981; Blaschek and Klacik, 1984; Strom et al., 1984; Mahony et al., 1986; Efstathiou and Truffaut, 1986; Yu and Pearce, 1986; Zappe et al., 1986; Lee et al., 1987).

4.3.1 DNase activity

The results obtained in the present study showed that all strains of C. acetobutylicum investigated possessed active intracellular and extracellular DNases. The detection of extracellular and intracellular DNase activity towards the end and during stationary phase growth is similar to that reported in B. subtilis (Akrigg, 1978; Reshetnik et al., 1978; Akrigg and Mandelstam, 1977). Initially, it was thought that by harvesting cells during early- to mid-exponential phase, the DNase problem would be avoided. However, problems of DNA degradation still occurred, so attempts were made to inhibit or inactivate DNase.

The addition of detergents did not decrease the DNA degradation. Similar results have been reported for various clostridial species by Strom et al. (1984) and Lee et al. (1987). Proteinase K is an unspecific serine protease which possesses the ability to rapidly inactivate, by proteolytic action, any endogenous nucleases with or without SDS being present. However, the use of proteinase K did not give any appreciable decrease in the DNA degradation. Martin and Savage (1988) used many proteolytic enzymes, including proteinase K and did not detect any inhibition of the DNA degradation. Similar results were reported by Strom et al. (1984) and Lee et al. (1987).

Diethyl pyrocarbonate (DEP) reacts with and denatures proteins, and therefore inactivates nucleases (Ehrengberg et al., 1976). It can also react with single stranded DNA but does not affect double stranded DNA. DEP has been used extensively during the isolation of undegraded nucleic acid (DNA and RNA) from a variety of organisms, such as viruses, bacteria, blue-green algae, protozoa, slime molds, fungi, plants and animals (Ehrenberg et al., 1976). However, its use to aid the isolation of undegraded plasmid and chromosomal DNA from various Clostridium species has resulted in conflicting results as to its effectiveness (Strom et al., 1984; Blaschek and Klacik, 1984; Efstathiou and Truffaut, 1986; Mahony et al., 1986; Lee et al., 1987). The use of DEP in the lysis protocol

during extraction of DNA should have resulted in less degraded DNA. However, the results obtained were negative and a persistent increase in the isolation of undegraded DNA was not obtained. This apparent failure is in agreement with reports from other groups.

Chaotropic agents (often used in the isolation of RNA via inhibition of RNase) have been used successfully in the isolation of high molecular weight DNA from the DNAse-producing S. aureus (Dyer and Tandolo, 1983). Proteins rapidly dissolve in solutions of potent chaotropic agents as the orderly secondary structure is lost (Maniatis et al., 1982). However, the use of the chaotropic agents guanidinium isothiocyanate and sodium perchlorate did not help in the isolation of undegraded DNA in the present work. Hence the results obtained suggested that DNase was not the only factor involved in DNA degradation.

Another possible factor is the presence of oxygen. The ability of an organism to survive and grow in the presence of oxygen is dependent upon factors such as possession of a variety of enzymes which detoxify oxygen reduction products. Organisms which lack these systems generate significantly high amounts of these products (Fridovich, 1978). DNA is one of the sites attacked by such products. Production of hydrogen peroxide via the presence of oxygen is known to produce single stranded breakage in DNA (Ananthaswamy and Eisenstark, 1977; Brawn and Fridovich, 1981; Kreig and Hoffmen, 1986). Hydrogen peroxide is also reported to cause the loss of adenine residues from DNA (Uchida et al., 1965) leading to single stranded breakage of DNA (Ljungquist, 1977).

This is put forward as a possible explanation for the DNA degradation observed in C. acetobutylicum lysates since all detectable levels of DNAse were removed and its presence is no longer a viable explanation. Martin and Savage (1988) reported similar results when attempting to isolate high molecular weight DNA from the oxygen-intolerant Gram-negative Roseburia cecicola. This high molecular weight DNA was isolated only when the bacterial cells were kept

under anaerobic conditions up to lysis. Exposure of cells to oxygen caused chromosomal DNA to degrade. They also reported extensive degradation of linear lambda DNA but only single stranded nicking of plasmid DNA. Lee et al. (1987) reported results which showed that isolated plasmid DNA was relatively unstable at 4°C; after 2-3 days of storage the DNA was completely degraded.

The qualitative results reported in the present study are similar to the above. The lysates of C. acetobutylicum caused no degradation of pBR322. However, many of the samples of C. acetobutylicum genomic DNA decreased in molecular weight. An explanation could be that the presence of numerous single stranded nicks present in the chromosomal DNA caused it to literally fall apart. The evidence on the role of oxygen in the degradation of DNA is also an acceptable explanation as to the ineffectiveness of DEP in the isolation of undegraded plasmid and chromosomal DNA reported by some groups (Blaschek and Klacik, 1984; Strom et al., 1984; Mahony et al., 1986; Lee et al., 1987). DEP, which has the ability to inhibit nuclease, has no ability to protect DNA from the oxygen radicals.

E. coli normally has an efficient mechanism to repair DNA under the control of the recA and lexA gene products. This system is independent of DNA replication, and is usually effected by the repair system I (Miller, 1983). Mutants of E. coli, lacking this system are reported to become highly sensitive to agents which cause chromosomal breaks, including oxygen and hydrogen peroxide (Ananthaswamy and Eisenstark, 1977; Morimyo, 1982). This leads to the hypothesis that C. acetobutylicum (as an anaerobe) may be deficient in a repair system required for protection against oxygen and free radicals. If C. acetobutylicum does not contain any major oxygen-detoxifying enzyme system, it has no known way to prevent the damage caused by the presence of oxygen during lysis. The reports by various groups (see Section 1.4.2), that indirect mutagens were unsuccessful in the production of mutants supports this

hypothesis.

When cells were protected and kept as anaerobic as possible, DNA stability was noticeably improved and allowed a significant increase in the isolation of stable, high molecular weight DNA.

4.3.2 Preparation of genomic library

There is an extensive list of genes from Gram-positive bacteria (Bacillus, Clostridium, Streptococcus, Lactobacillus, Ruminococcus) that have been successfully cloned and expressed in E. coli (Lee et al., 1982; Brenier et al., 1983; Panbangred et al., 1983; Ishii et al., 1983; Karube et al., 1983; Cornet et al., 1983; Gardner and Aronson, 1984; Breidt and Stewart, 1986; Limsowtin et al., 1986; Howard and White, 1988).

The production of a genomic library is the essential first step to allowing investigation and genetic analysis of C. acetobutylicum. An understanding of the genetics will help to enable alteration of biochemical characteristics which will lead to increasing the efficiency of the ABE fermentation.

The trouble experienced in producing the genomic library was decreased by changing some of the steps from the normal method of producing a genomic library. Washing the DNA through a centricon filter removed NaCl and enzymes, as well as concentrating the DNA. It also decreased the degree of handling of the DNA. Although the possibility of getting ligation of two fragments together that do not align within the genome exists, the level would be extremely low, as the DNA sample was collected during the EcoRI partial digestion at a time interval where the DNA dropped only from 50 to 23 kb. The DNA once cloned in E. coli as the library was completely stable.

4.4 SUMMARY

Strains of C. acetobutylicum produced high levels of extracellular and intracellular DNase activities. The DNase activities coincided with the late-exponential and stationary phases of growth. Detectable intracellular DNase activity preceded detectable extracellular activity.

Isolation of chromosomal DNA from strains of C. acetobutylicum was associated with degradation. This degradation was not prevented by the use of various protein inhibiting agents, and was therefore not due entirely to the DNase. This observed degradation is presumed to be due to the adverse affect of oxygen and related radicals on the double-stranded DNA. Isolation of undegraded DNA could be achieved reproducibly by protecting the cells to be lysed from oxygen, and precipitation of DNA as soon as lysis occurred. A genomic library to C. acetobutylicum NCIB2951 was made using the cosmid vector pLAFR1. It constituted 3,500 recombinant clones.

CHAPTER 5: IDENTIFICATION AND CHARACTERIZATION OF LACTOSE UTILIZATION GENES IN C. ACETOBUTYLICUM NCIB2951

5.1 INTRODUCTION

5.1.1 Lactose Utilization in C. acetobutylicum

C. acetobutylicum has been shown to possess both ß-galactosidase and phospho-ß-galactosidase activities (Yu et al., 1987). In P262, induction of ß-galactosidase and phospho-ß-galactosidase occurs during the fermentation of whey. High levels of phospho-ß-galactosidase are detected in early stages of fermentation, coinciding to the acidogenic (growth) phase. The level of phospho-ß-galactosidase then drops sharply to a low level within 5 h of induction. ß-galactosidase activity is observed after 6 h growth and increases slowly, peaking at 22 h (coinciding with maximum solvent production). Both ß-galactosidase and phospho-ß-galactosidase have been detected in NCIB2951 at 16 h after inoculation, with higher levels of ß-galactosidase than that of phospho-ß-galactosidase.

This differential expression in the induction of ß-galactosidase and phospho-ß-galactosidase could indicate that the two lactose utilization pathways serve two different functions during the ABE fermentation (Yu et al., 1987). The presence of these two enzymes suggest that both pathways for lactose utilization (as shown in Section 1.7) operate in C. acetobutylicum. Nothing is known about the regulation of these two pathways in C. acetobutylicum or the regulation of the ß-galactosidase and phospho-ß-galactosidase enzymes themselves. The relative importance in vivo of these two potential pathways for lactose utilization and solvent production in C. acetobutylicum has yet to be determined. However, ATCC824, which possesses only a phospho-ß-galactosidase produces lower amounts of solvents than P262 (ß-galactosidase and phospho-ß-galactosidase) when grown on lactose (Yu et al., 1987). Of these two known lactose utilisation pathways, only the ß-galactosidase is well understood at the genetic level (Dills

et al., 1980). The PEP:PTS system is a complex system, with regulation and induction differing markedly between different bacteria (Postma and Lengeler, 1985).

5.1.2 Aim

The aim of this phase of the project was to isolate and characterize the genes encoding the ß-galactosidase and phospho-ß-galactosidase from the NCIB2951 genomic library.

5.2 RESULTS

5.2.1 Complementation of HB101 Auxotrophic Mutations

In order to establish whether the C. acetobutylicum insert DNA in the recombinant clones was able to express in E. coli, the ability to complement auxotrophic mutations present in the host (E. coli HB101) was investigated. HB101 possesses mutations in the proline, leucine and thiamine biosynthesis pathways.

The genomic library was plated onto minimal medium (supplemented with glucose, Tc, and TTC), and lacking one of the three required amino acids or vitamins (i.e. proline, leucine or thiamine). The plates were incubated at 37°C for two days. Several of the red colonies which grew were picked and replated onto the same medium for isolation of single colonies to establish the stability of their prototrophy. Using this method, clones were isolated from the library that complemented each of the auxotrophic defects of HB101.

Six clones were selected that complemented the proA2 (α-glutamyl phosphate reductase) mutation of HB101. Cosmid DNA was isolated from these clones

Fig.5.1: Analysis by agarose gel electrophoresis of cosmids isolated from the C. acetobutylicum NCIB2951 genomic library which complemented the proA2, leuB6, and thi auxotrophic mutations present in E. coli HB101.
Size markers are in kilobases.

(a) EcoRI restriction enzyme digests of cosmids complementing proA2 mutation.

Lane 1 Lambda HindIII size standards
Lane 2 pKH90
Lane 3 pKH91
Lane 4 pKH92

(b) EcoRI restriction enzyme digests of cosmids complementing leuB6 mutation.

Lane 1 Lambda HindIII size standards
Lane 2 pKH96
Lane 3 pKH97
Lane 4 pKH98

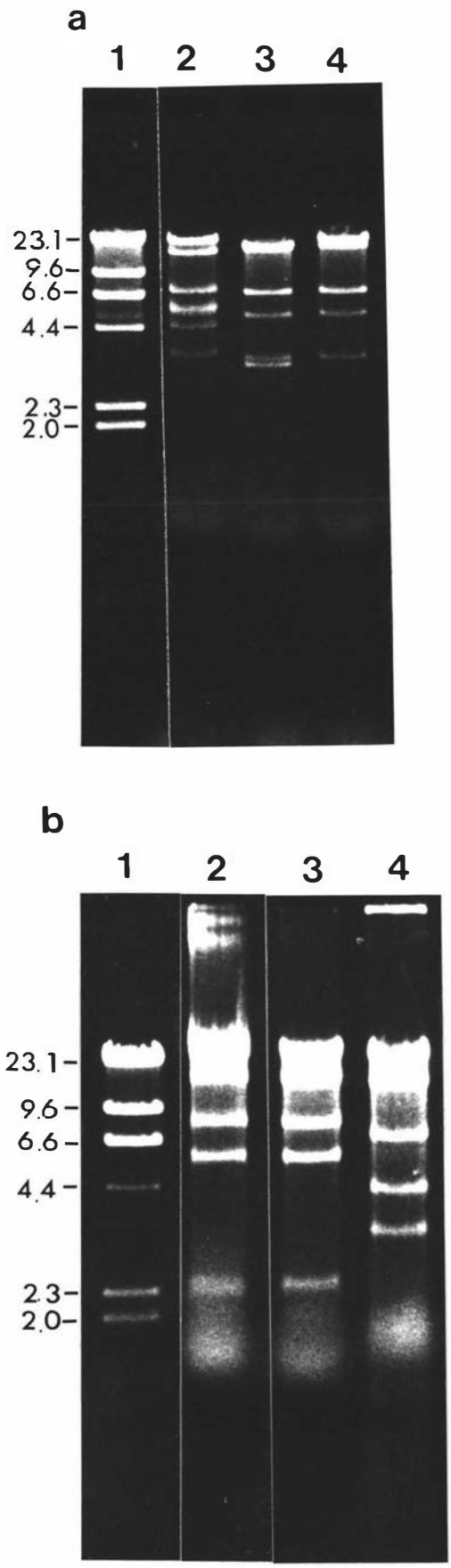
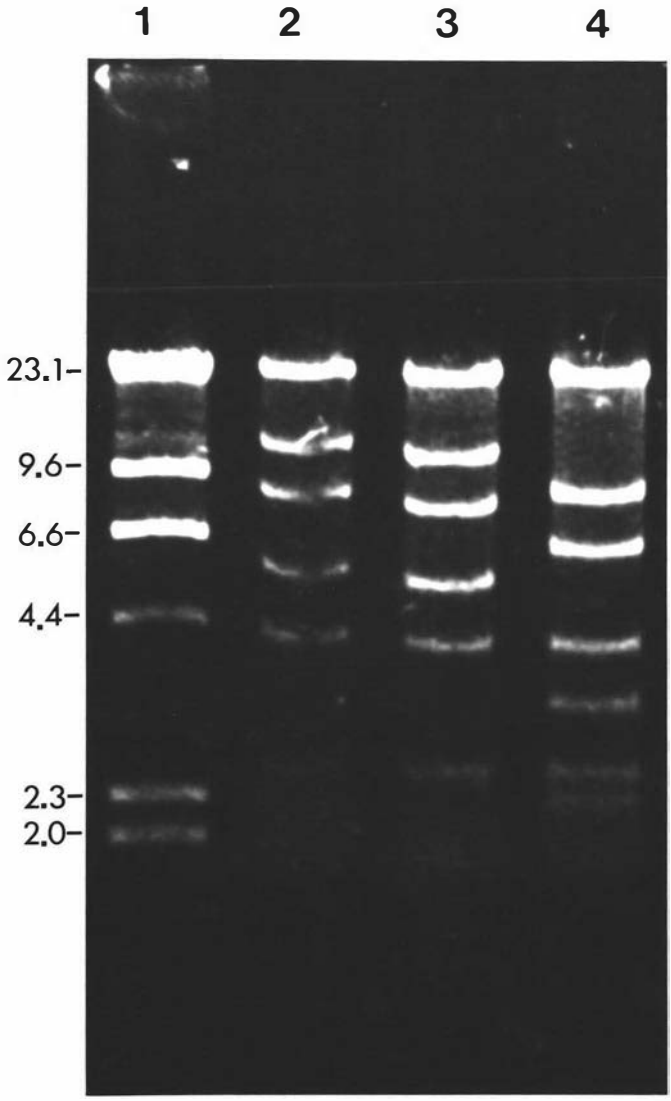

a
1
2
3
4
23.1
9.6
6.6
4.4
2.3
2.0
b
1
2
3
4
23.1
9.6
6.6
4.4
2.3
2.0

Fig.5.1 cont'd:

(c) EcoRI restriction enzyme digests of cosmids complementing thi mutation.

Lane 1 Lambda HindIII size standards
Lane 2 pKH93
Lane 3 pKH94
Lane 4 pKH95

c



(section 2.7.3), digested with EcoRI (section 2.9), and analyzed by agarose gel electrophoresis as described in section 2.12. All clones shared common DNA fragments of 4.6 and 6.8 kb, and the three unique digest patterns observed (designated pKH90-92) are shown in Fig. 5.1a.

Six positive clones were selected, which had acquired the stable leucine prototrophy and therefore must express the α-isopropyl malate dehydrogenase, absent in HB101 because of the leuB6 mutation. These clones were grown up overnight, the plasmid DNA was isolated, digested with EcoRI and analyzed by agarose gel electrophoresis. Each clone was shown to share a common fragment insert DNA (15.4 kb) as well as the 21.6 kb band corresponding to pLAFR1. Three such clones (designated pKH96-98) representing the six analyzed are shown in Fig. 5.1b.

Clones that complemented the thiamine defect were also isolated. Cosmid DNA from 6 of these clones was digested with EcoRI and also analyzed by agarose gel electrophoresis. The clones shared two common DNA fragments (3.8 and 2.5 kb). Three clones which represented the restriction digest patterns observed were chosen and designated pKH93-pKH95 (Fig 5.1c).

These results showed that genes from the Gram-positive anaerobe, C. acetobutylicum, were able to express and complement mutations present in the Gram-negative host strain E. coli HB101 genome.

5.2.2 Attempts to isolate the pbg gene from the C. acetobutylicum library using a probe from Lactobacillus casei 64H

As the phospho-ß-galactosidase (pbg) gene from L. casei 64H had been cloned, initial experiments were carried out to see if this probe would be suitable to select for the C. acetobutylicum pbg gene. The L. casei pbg gene has been cloned on a 7.9 kb PstI fragment in pBR322, designated pLZ600 (Lee et al., 1982). Plasmid pLZ600 was isolated from DRI4477 (section 2.7.3) and was

Fig.5.2: Hybridization of the 1.1 kb SphI fragment containing the pbg gene region of L. casei to EcoRI digested genomic DNA of C. acetobutylicum NCIB2951.
Size markers are in kilobases.

(a) Agarose gel electrophoresis of EcoRI genomic digests of C. acetobutylicum NCIB2951.

Lane 1 Lambda HindIII size standards
Lane 2 C. acetobutylicum NCIB2951

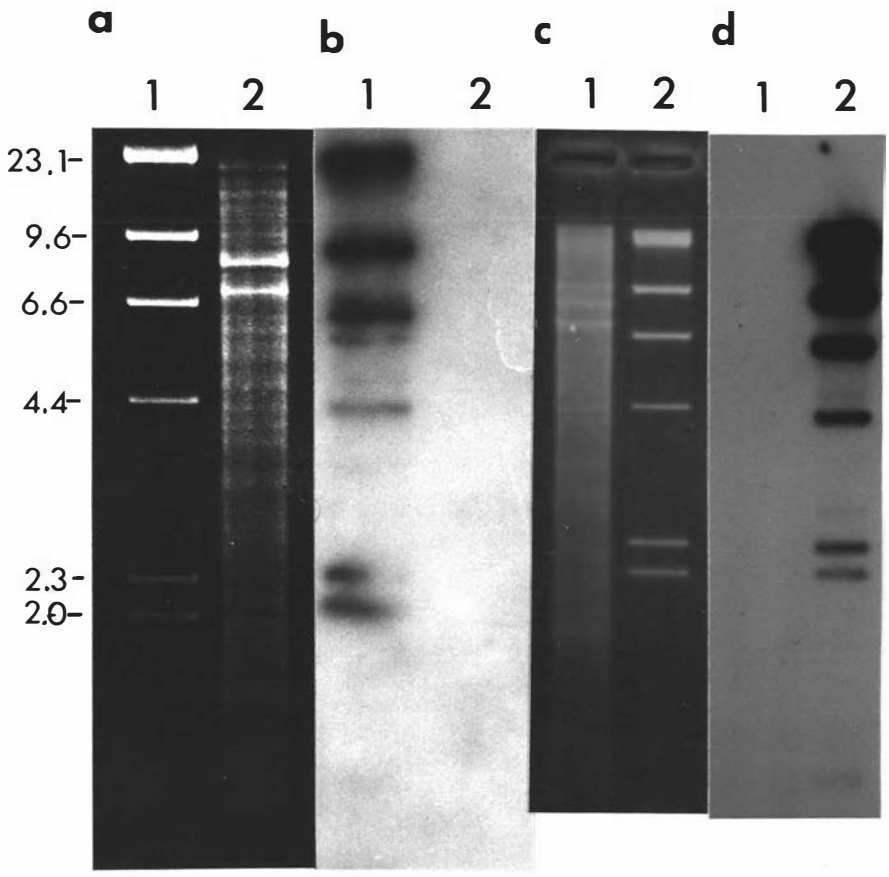
(b) Autoradiograph of the same gel hybridized at 65°C with [^{32}P]-labelled 1.1 kb SphI fragment carrying the L. casei pbg structural gene.

(c) Agarose gel electrophoresis of EcoRI genomic digests of C. acetobutylicum NCIB2951.

Lane 1 C. acetobutylicum NCIB2951
Lane 2 Lambda HindIII size standards

(d) Autoradiograph of the same gel hybridized at 60°C with [^{32}P]-labelled 1.1 kb SphI fragment carrying the L. casei pbg structural gene.

[^{32}P]-labelled lambda DNA was included to identify the positions of the lambda standards.


a
b
c
d
1
2
1
2
1
2
1
2
23.1-
9.6-
6.6-
4.4-
2.3-
2.0-

digested (section 2.9) with the restriction enzyme SphI to yield, amongst other bands, a 1.1 kb fragment known to contain the pbg gene. This fragment was isolated from agarose gel (section 2.16) and used as a probe for the pbg gene. To check whether there was an acceptably high degree of homology between L. casei and C. acetobutylicum pbg genes to allow detection of the clostridial gene in the genomic library, a Southern blot of EcoRI restricted genomic DNA of C. acetobutylicum NCIB2951 (section 2.19) was probed with the [^{32}P]-labelled 1.1 kb SphI pbg fragment of L. casei under conditions of high stringency (65°C and 6xSSC) and at a lower stringency (60°C, and 6xSSC) as described in section 2.18. In both cases, no significant hybridization was detected (Fig. 5.2). These results suggested that it would not be possible to utilize the L. casei pbg gene probe to detect and isolate the equivalent C. acetobutylicum NCIB2951 gene from the genomic library.

5.2.3 Attempts to select for the C. acetobutylicum pbg gene by expression in E. coli

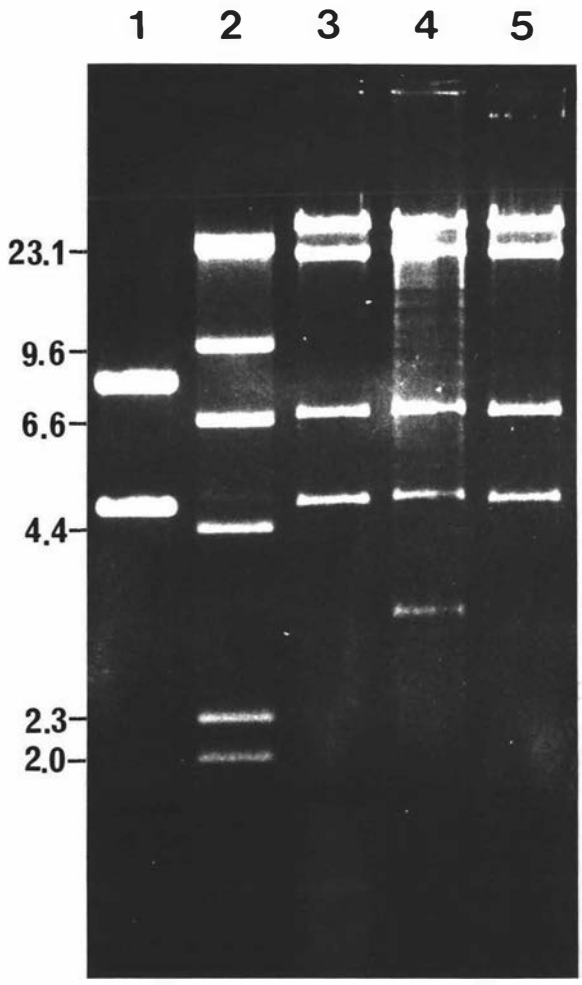
The genomic library was serially diluted and plated out for single colonies on LB agar plates containing tetracycline. One thousand colonies were selected and replated onto numbered sectors on plates.
According to the calculation shown in Section 4.2.2, the number of clones in the genomic library should be representative of the entire C. acetobutylicum genome. These E. coli clones were grown overnight in LB medium, supplemented with galactose, diluted 1/10 in fresh medium and after growing for 4 h, were sequentially screened for phospho-ß-galactosidase activity in microtitre trays using appropriate controls as described in section 2.23. However, after screening 850 colonies, no positive clones were detected.

5.2.4 Isolation of the cbg gene from C. acetobutylicum

To isolate the ß-galactosidase gene (cbg), two techniques were investigated. In the first, the genomic library of C. acetobutylicum NCIB2951 was plated out

Fig.5.3: Restriction enzyme digests of cosmids that expresses ß-galactosidase activity. Agarose gel of EcoRI digests of cosmid DNA. Size markers are in kilobases.

Lane 1 pKH107
Lane 2 Lambda HindIII size standards
Lane 3 pKH100
Lane 4 pKH101
Lane 5 pKH102


1
2
3
4
5
23.1
9.6
6.6
4.4
2.3
2.0

onto minimal medium supplemented with Tc, TTC, casamino acids, and with lactose as the carbon source. Colonies capable of catabolizing the lactose would, in the presence of TTC, produce a deep red formazin dye due to TTC reduction. Colonies unable to catabolize the test substrate would remain colourless. When this was attempted, no colonies were obtained on the plate. However, this selection system relies on complementation of the lacY mutation of HB101 to ensure lactose is taken up by the cell.

Therefore, an alternative method using X-gal was tried. X-gal (5-bromo-4--chloro-3-indoyl galacto-pyranoside) is a chromogenic substrate which is cleaved by ß-galactosidase to galactose plus the insoluble blue indoyl derivative. The genomic library was crossed en masse, using the tri-parental mating system (see section 2.22), into E. coli strain PB2959 (△lac, which lacks the entire lac operon) and screened for active ß-galactosidase activity by streaking a loopful onto minimal plates supplemented with 40 μg/ ml X-gal and 200 μg/ml IPTG. This resulted in a number of blue colonies on the plates, of which 20 were picked and restreaked onto the same selective medium to generate single colonies. Testing revealed that IPTG was not required for expression of the ß-galactosidase. Approximately 20 of these blue colonies were isolated, the DNA was extracted, and digested using restriction enzyme EcoRI as described in section 2.9. This showed that they contained three common digest patterns. Three of these colonies were chosen and designated DRI4997-DRI4999 (pKH100-pKH102). Each of these three cosmids shared three common insert bands of 5.2, 7.2, and 28 kb (Fig. 5.3, lanes 3-5). The colonies were tested for kanamycin resistance to ensure the absence of pRK2013. None could grow in the presence of kanamycin. These three recombinant clones were transferred back into HB101 via tri-parental crossing (section 2.22) and plated onto supplemented minimal medium containing lactose. However, no growth was detected. This led to the conclusion that the equivalent lactose permease gene was not present upon the cosmid, or was unable to complement the lacY mutation in HB101. Testing these three recombinant clones for phospho-ß-galactosidase activity was also

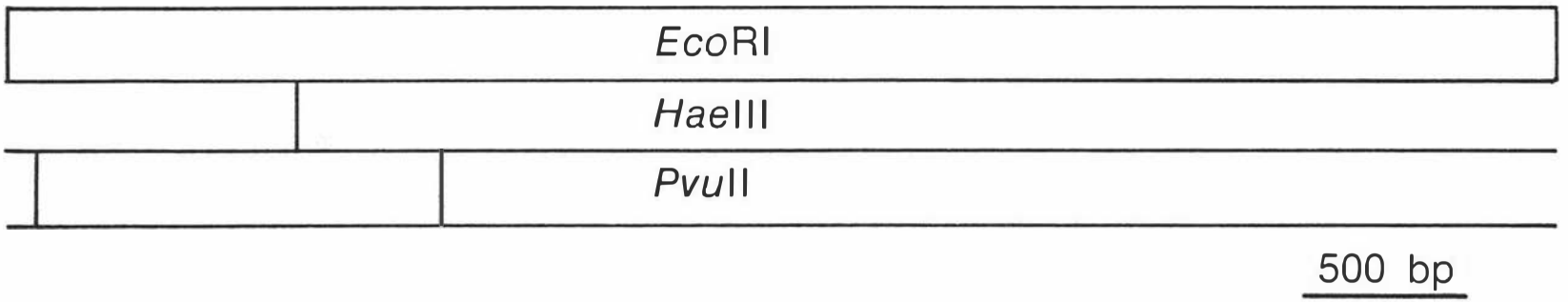
negative, showing that the pbg gene was not situated close to the cbg gene.

The relative position of the cbg gene within the clones pKH100-102 was determined by a combination of subcloning. To clone the fragment that contained the C. acetobutylicum cbg gene, plasmid DNA (20 μg) from pKH102 was digested with EcoRI restriction enzyme (section 2.9) and run on a gel as described in section 2.12. The common fragments were cut out from the seaplaque agarose and isolated via the freeze/thaw technique as described in section 2.16. This DNA was ligated (section 2.20) into EcoRI-digested pSUP202 vector DNA (containing the mob site of pRK2), transformed into E. coli HB101 (section 2.21) and plated onto LB agar plates containing tetracycline. Twelve Tc^R colonies were picked from each plate, grown up overnight in selective LB medium and the plasmid DNA was isolated (section 2.7.3), and digested with EcoRI as described in section 2.9. This analysis enabled selection of several clones which were shown to contain an 8.0 kb EcoRI fragment corresponding to pSUP202 DNA, and either a 5.2 kb or 7.2 kb EcoRI fragment of insert DNA. Others were shown to contain only an 8.0kb EcoRI band corresponding to pSUP202.

These two smaller, common EcoRI restriction fragments (as present in the recombinant plasmids pKH106-pKH107) were crossed into E. coli strain PB2959 (DRI5013, DRI5014) and screened on the selective medium. The results showed that only when the 5.2 kb EcoRI fragment was present (cloned in pSUP202) could ß-galactosidase activity be detected in the isogenic host strain PB2959. This signified that the cbg gene was present upon the 5.2 kb EcoRI fragment. Since PB2959, with or without pSUP202, forms white colonies on the selective medium containing X-gal, this result indicated that the cbg gene had been cloned into pSUP202 on the 5.2 kb EcoRI fragment (Fig. 5.3, lane 1). One such transformant DRI5014, was kept and the recombinant plasmid in HB101 was designated pKH107. To test whether the expression of the cloned cbg gene was

Fig.5.4: Restriction enzyme map of the 5.2 kb EcoRI fragment containing the cbg gene region of C. acetobutylicum NCIB2951.

EcoRI (E)
HaeIII (H)
PvuII (P)


*Eco*RI
*Hae*III
*Pvu*II
500 bp

orientation specific, i.e. under the control of a vector promoter, this fragment was then recloned in the opposite orientation in pSUP202 (pKH179), and identified to be in the opposite orientation by PvuII restriction enzyme digestion. When crossed into PB2959(section 2.22) (DRI5015) and re-tested upon the selective medium containing X-gal, the fragment still possessed an active ß-galactosidase, thus confirming that the 5.2 kb EcoRI fragment not only contained the structural gene for ß-galactosidase, but also the required clostridial promoter.

5.2.5 Restriction enzyme mapping of pKH107

Once pKH107 had been isolated and characterized with respect to the size of insert and the enzyme for which it coded, a physical map of the pKH107 insert was developed by restriction enzyme analysis as described in section 2.9 in an attempt to localize the position of the cbg gene. The map was constructed from cleavage of both the plasmid pKH107 and the purified insert (Fig 5.4). No enzyme sites were detected for BamHI, BclI, BglII, ClaI, HindIII, HpaI, KpnI, MboI, PstI, PvuI, XbaI, or XhoI. The enzymes, HaeIII and PvuII had one and two cleavage sites in the insert, respectively, and were the only useful 4 and 6bp restriction enzymes found. The insert contained multiple cleavage sites for AluI, RsaI and Sau3A, but the positions of these cleavage sites are not indicated on the map as they were far too numerous, the fragments ranging from 1000 to less than 200bp long. Sub-cloning of the HaeIII/EcoRI fragments into pUC18 totally interrupted the ß-galactosidase activity and therefore did not help to further localize the boundaries of the gene.

5.2.6 Characterization of the coded ß-galactosidase gene on pKH107 and pKH179

Experiments were carried out to investigate the level of activity of the C. acetobutylicum ß-galactosidase gene (cbg) in E. coli strains PB2959 and JM101. Permeabilized cells (glucose-grown) were tested using ONPG as substrate as

described in section 2.24. ONPG is hydrolysed by ß-galactosidase to galactose and o-nitro phenol (yellow).

DRI5015 and DRI8520, containing the recombinant plasmids pKH107 and pKH179 respectively in PB2959, expressed ß-galactosidase activity irrespective of the orientation of the 5.2 kb EcoRI fragment (390U and 344U). Similar results were observed in JM101 (Table 5.1) where 416U and 400U of activity were recorded for the two orientations. PB2959, however, with or without pSUP202, yielded very low units of activity, as did JM101. These results showed that the activities were well above the level of background activities of either strain as well as that of HB101 (when not induced with IPTG).

Table 5.1: Expression of C. acetobutylicum ß-galactosidase in E. coli

Bacteria	Units of ß-galactosidase[a]			
	Glucose	Glucose/ IPTG	Succinate	Succinate/ IPTG
HB101	31.7	1106	33.3	877
JM101	3.9	9.0	3.3	6.2
JM101/pSUP202	5.0	5.2	4.5	4.1
JM101/pKH107	416	377	418	451
JM101/pKH179	400	424	485	480
PB2959	8.9	2.5	3.8	4.5
PB2959/pSUP202	7.5	4.7	5.5	5.3
PB2959/pKH107	390	365	390	321
PB2959/pKH179	344	381	376	333

[a] Units of ß-galactosidase = $\frac{1000 \times OD_{420}}{\text{time} \times \text{vol} \times OD_{600}}$

When grown in the presence of the inducer IPTG, the activities were 365U and 381U for DRI5015 and 8520, respectively, indicating that there was little significant increase in activities. These results show that the cbg gene is not induced by IPTG, the activities being approximately equal in the presence or absence of the inducer. It was also noticed that the clostridial ß-galactosidase gene was expressed constitutively in E. coli strain HB101 when grown in the presence or absence of glucose as blue colonies were observed from the metabolism of X-gal.

Strain JM101 overproduces the E. coli lacI repressor protein. However, there was no detectable change in the levels of the ß-galactosidase activity in strains harbouring pKH107 and pKH179. This demonstrated that the lacI repressor of JM101 had no effect on the ß-galactosidase promoter. When the cells were grown on succinate, the activities of the clostridial cbg gene in either orientation were at equivalent levels to those seen when grown on glucose. This showed that activity is neither repressed nor induced when cloned in E. coli.

5.2.7 Detection of ß-galactosidase genes in other C. acetobutylicum strains using Southern blot hybridization

Yu et al. (1987) detected the presence of ß-galactosidase activity in C. acetobutylicum strains NRRL 594, NRRL598, NRRL2490, NCIB2951 and P262, but no activity was detected in strain ATCC824. Hence, an attempt was made to identify the presence of a cbg gene in this strain via Southern blot hybridization. A range of other strains were included for comparison. Total chromosomal DNA was isolated from each strain and digested with restriction enzyme EcoRI as described in section 2.9. The resultant gels were Southern blotted (section 2.19) and probed with the [^{32}P]-labelled 5.2 kb EcoRI fragment containing the cbg gene from pKH107 (Fig. 5.5). The results showed that all the strains tested contained a single fragment which showed homology to the NCIB2951 cbg gene probe. Also, the hybridization patterns showed that the size of the comparable fragments differed from the fragment size used as a probe

Fig.5.5: Hybridization of the C. acetobutylicum NCIB2951 cbg gene region from pKH107 to EcoRI genomic digests of other C. acetobutylicum strains. Size markers are in kilobases.

(a) Agarose gel electrophoresis of EcoRI genomic digests of C. acetobutylicum DNA.

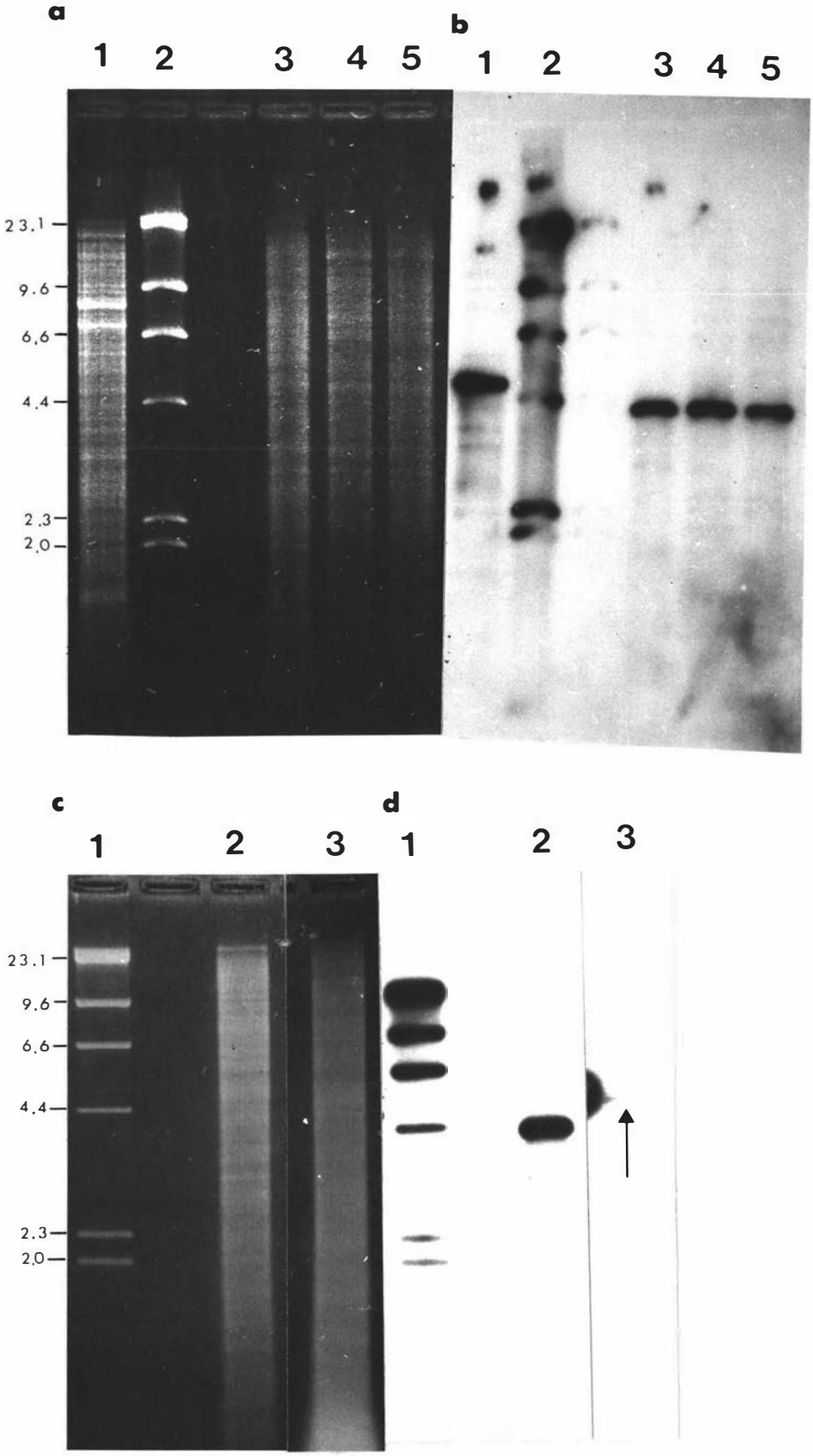
Lane 1 C. acetobutylicum NCIB2951
Lane 2 Lambda HindIII size standards
Lane 3 C. acetobutylicum NRRL594
Lane 4 C. acetobutylicum NRRL598
Lane 5 C. acetobutylicum NRRL2490

(b) Autoradiograph of the same gel hybridized with [^{32}P]-labelled 5.2 kb EcoRI fragment of pKH107 carrying the C. acetobutylicum NCIB2951 cbg gene region. [^{32}P]-labelled lambda DNA was included to identify the position of the lambda standards.

(c) Agarose gel electrophoresis of EcoRI genomic digests of C. acetobutylicum DNA.

Lane 1 Lambda HindIII size standards
Lane 2 C. acetobutylicum P262
Lane 3 C. acetobutylicum ATTC824

(d) Autoradiograph of the same gel hybridized with [^{32}P]-labelled 5.2 kb EcoRI fragment of pKH107 carrying the C. acetobutylicum NCIB2951 cbg gene region. [^{32}P]-labelled lambda DNA was included to identify the position of the lambda standards.


a
1
2
3
4
5
23.1
9.6
6.6
4.4
2.3
2.0
b
1
2
3
4
5
c
1
2
3
23.1
9.6
6.6
4.4
2.3
2.0
d
1
2
3

(NRRL594, 4.6 kb; NRRL598, 4.6 kb; ATTC824, 5.6 kb; NRRL2490, 4.6 kb; P262, 4.6 kb). Since the amount of DNA (2 μg) loaded was equal for each strain, the results also demonstrated that the degree of homology was very strong with all other strains, except for ATCC 824, which showed a lesser degree of homology.

The result above showing that ATCC824 possessed DNA homology to the cbg gene probe was unexpected since Yu et al. (1987) had reported that no detectable ß-galactosidase activity existed in this strain. Hence, this strain, and the others, were plated onto CGM medium plates, containing lactose as the sole carbon source, plus X-gal, and tested using ONPG. All strains were shown to express the gene, except for ATCC824, which failed repeatedly to give blue colonies on the CGM medium plates or to hydrolyse ONPG. This confirmed the previous work (Yu et al., 1987).

5.2.8 Degree of homology of the C. acetobutylicum NCIB2951 cbg gene region with DNA from other bacteria

Homology studies were performed using the [^{32}P]-labelled 5.2 kb EcoRI fragment containing the cbg gene to S. lactis ATCC7962, S. thermophilus DRI1242, L. bulgaricus DRI20056, L. helviticus DRI20064, E. coli DC272 and R. loti PN2213.

Table 5.2: Detection of ß-galactosidase activity in various strains of bacteria using selective medium supplemented with X-gal

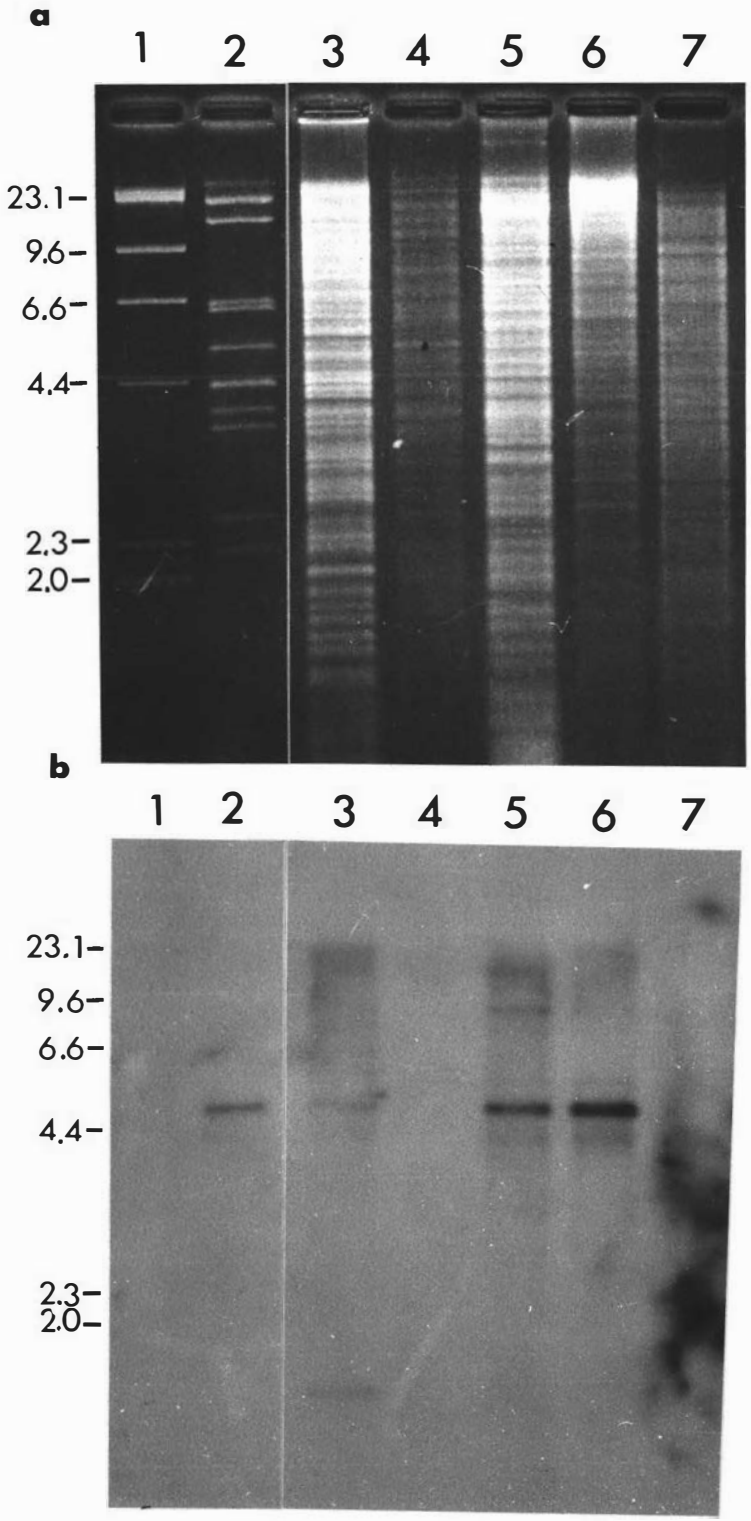
Medium	Bacteria	Temperature	Result
M17 lactose	S. lactis ATCC7962	30°C	+
M17 lactose	S. thermophilus DRI1424	37°C	+
MRS lactose	L. bulgaricus DRI20056	30°C	+
MRS lactose	L. helviticus DRI20064	30°C	+
MM56 lactose	E. coli DC272	37°C	+
MM56 lactose	E. coli PB2959	37°C	-

Fig.5.6: Hybridization of the 5.2 kb EcoRI fragment containing the cbg gene region of C. acetobutylicum NCIB2951 to genomic and plasmid DNA of other Gram-positive and Gram-negative bacteria.
Size markers are in kilobases.

(a) Agarose gel of EcoRI restriction enzyme digests of DNA from various strains.

Lane 1	Lambda HindIII size standards
Lane 2	p7962 from S. lactis ATCC7962
Lane 3	S. thermophilus DRI1242
Lane 4	R. loti PN2213
Lane 5	L. bulgaricus DRI20056
Lane 6	L. helveticus DRI20064
Lane 7	E. coli DC272

(b) Autoradiograph of the same gel hybridized with [^{32}P]-labelled 5.2 kb EcoRI fragment of pKH107 carrying the C. acetobutylicum NCIB2951 cbg gene region.


a
1 2 3 4 5 6 7
23.1–
9.6–
6.6–
4.4–
2.3–
2.0–
b
1 2 3 4 5 6 7
23.1–
9.6–
6.6–
4.4–
2.3–
2.0–

All strains (except R. loti) were first tested on a suitable selective lactose-containing medium supplemented with X-gal to confirm the presence of an active ß-galactosidase. All strains produced blue colonies.

Total chromosomal DNA was isolated (section 2.8.1), digested with the restriction enzyme EcoRI (section 2.9), and the resultant gel was Southern blotted and hybridized (section 2.19) at low (60°C, 6xSSC) and high (65°C, 6xSSC) stringency using [^{32}P]-labelled 5.2 kb EcoRI fragment from pKH107 (Fig. 5.6). At a high stringency there was no detectable homology between the clostridial cbg gene and the equivalent gene in all other strains (results not shown). However, at the lower stringency there was some detectable homology between the Gram-positive bacterial DNA digests (S. lactis, S. thermophilus, L. bulgaricus, and L. helviticus), but none between the probe and the total DNA digests of E. coli or R. loti (which was used as a negative control). S. lactis p7962, known to carry the gene for ß-galactosidase activity, was also prepared as described above. This enabled identification of a single EcoRI fragment of 5.6 kb which demonstrated some detectable homology.

5.3 DISCUSSION

5.3.1 Complementation of E. coli auxotrophs with the C. acetobutylicum gene library

The results obtained confirm those reported by Efstathiou and Truffaut (1986) and Usdin et al. (1986) that various auxotrophic lesions in E. coli can be successfully complemented. In E. coli the proline biosynthesis pathway proceeds by three enzymatic steps, one of which is spontaneous. The genes are located as an operon proceeding from proB to proA (Deutch et al., 1984). The proA mutation in HB101 prevents production of the glutamate semi-aldehyde dehydrogenase. The recombinant clones which were isolated produced this enzyme, enabling the biosynthesis of proline to proceed. The same

complementation has been reported by Efstathiou and Truffaut (1986) with C. acetobutylicum ABKn8.

In E. coli, the leucine genes are clustered as an operon (Somers et al., 1973). In HB101 the lesion in the leucine biosynthesis pathway is in the leuB6 gene which encodes for α-isopropyl malate hydrogenase. The same gene was cloned from C. butyricum (Ishii et al., 1983), and was able to complement this mutation in E. coli C600. The biosynthesis of thiamine in E. coli involves the condensation of independently synthesized pyrimidine and thiazole moieties (Estramareix et al., 1977). Complementation of the thi mutation present in HB101 by the recombinant clones allowed biosynthesis of thiamine. The exact mutation causing thiamine auxotrophy in HB101 is unknown.

5.3.2 Attempts to isolate the C. acetobutylicum pbg gene

The lack of detectable hybridization between the pbg gene probe of L. casei and the DNA of C. acetobutylicum shows that these genes are dissimilar in the two bacteria. Although L. casei has a reported G+C content of 45-47% (Kandler and Weiss, 1986), compared to the 28% G+C reported for C. acetobutylicum, the two bacteria are, taxonomically, closely related. Thus, this result was rather surprising.

ONPG-6-P is hydrolysed by phospho-ß-galactosidase to produce the yellow o-nitro phenol. Use of this substrate has enabled the isolation of pbg genes from L. casei (Lee et al., 1982), S. lactis (Maeda and Gasson, 1986; Boizet et al., 1988), S. cremoris (Inamine et al., 1986), and S. aureus (Breidt and Stewart, 1986). Hence, the failure to detect the gene from C. acetobutylicum was, again, rather surprising. If the pbg gene is present in the genomic library the inability to detect it could be due to low expression levels and/or lack of induction. First, the problem may be due to the level of phospho-ß-galactosidase. All other groups using this method of detection have utilized a high copy number vector

(such as pBR322, pUC18) for the cloning of insert DNA (Lee et al., 1982; Maeda and Gasson, 1986; Breidt and Stewart, 1986). The levels of the cloned pbg genes were then equal to the uninduced levels found in the native bacteria (Breidt and Stewart, 1986; Maeda and Gasson, 1986; Lee et al., 1982). In the present work, the vector used, pLAFR1, has only a relatively low copy number.

Secondly, it is not known if lactose or galactose themselves are the inducers or a product from their catabolism. Since HB101 has a fully operational galactose utilization pathway, if the latter is true the product should have been present. Conversely, the addition of galactose to the LB medium should have caused induction if the regulation that exists in C. acetobutylicum functions in E. coli. However, it is possible that the regulation observed in the host strain disappears upon cloning of the gene into E. coli as reported for many Gram-positive genes cloned into E. coli (Limsowtin et al., 1986; Gilbert and Hall, 1987).

Expression and regulation of pbg genes appears to be complex, differing greatly from one particular organism to another. Expression and induction of the pbg in C. acetobutylicum occurs in the presence of lactose or galactose. However, the actual inducer is unknown.

5.3.3 Cloning of the cbg gene from C. acetobutylicum

Cloning of the cbg gene met with greater success than for the pbg. Detection of ß-galactosidase activity in strains with the pLAFR1 cosmids took four days on minimal medium, but only two days for strains containing pSUP202 recombinant plasmids. However, when the relatively high copy number of pSUP202 is taken into account, the level of ß-galactosidase/plasmid is low. It could be hypothesized that the clostridial promoter is not utilized efficiently in E. coli, as this has been reported for the promoter of the S. bovis of ß-galactosidase gene cloned in E. coli (Gilbert and Hall, 1987).

In E. coli, cbg was expressed constitutively, and no repression by glucose or succinate was observed. This constitutive expression is similar to that reported for S. bovis (Gilbert and Hall, 1987). The relatively low level of expression was presumably due to the fact that the cbg gene was being expressed from its own promoter as indicated by the orientation independent manner in which it was expressed. The lack of regulation (repression and induction) may be due to the regulation genes being situated elsewhere on the Clostridium genome, as reported for the K. lactis ß-galactosidase gene where the regulation gene is unlinked to the ß-galactosidase structural gene (Dickson and Markin, 1978).

In C. acetobutylicum glucose does not appear to repress either the pbg or the cbg genes. Cell extracts from strains P262 and ATCC824 grown on glucose showed no detectable levels of ß-galactosidase or phospho-ß-galactosidase activity (Yu et al., 1987). However, glucose seemed to have no effect on the expression of the cbg gene in E. coli, neither repressing nor inducing. Gilbert and Hall (1987) have reported that although the E. coli lacZ gene is repressed by glucose, this metabolite has no effect on the corresponding S. bovis gene when present in E. coli.

JM101 possesses a lacIq mutation and therefore overproduces the repressor protein. However, even with these elevated levels of the repressor protein, there was no detectable decrease in the ß-galactosidase activity in the presence or absence of IPTG. This shows that there is no intraspecific function between the E. coli repressor and the equivalent region of the C. acetobutylicum gene. The ß-galactosidase gene of K lactis cloned in E. coli also showed no regulation by the E. coli lacI repressor and was not affected by the presence of IPTG. In contrast, intraspecific regulation between E. coli and S. bovis lac regions was detected (Gilbert and Hall, 1987).

All strains of C. acetobutylicum tested displayed homology to the NCIB2951 cbg gene region, although ATCC824 seemed to show the weakest homology. Yu et

al. (1987) have reported that ATCC824 has no detectable level of ß-galactosidase during any stage of the growth period, and a similar absence of activity has been demonstrated in the present study. This is in contrast to the other strains tested, which all possessed some degree of ß-galactosidase activity. The reason for the lack of activity in ATCC824 may be due to a mutation in the cbg structural gene, or to a mutation in the regulation region involved in control of its synthesis. The regulation system controlling this gene in C. acetobutylicum is totally unknown, but since there is differential induction between the phospho-ß-galactosidase and the ß-galactosidase enzymes during the growth cycle, it is most likely that the corresponding genes are tightly regulated.

Similar results have been reported for Aeromonas caviae, where strains which were cryptic for lactose utilization were shown by hybridization to possess a ß-galactosidase gene region homologous to A. caviae ß-galactosidase gene. Insertion of a cloned lactose permease into these strains still did not allow growth on lactose, nor expression of the A. caviae ß-galactosidase gene (Rodgers et al., 1987).

Only a low degree of homology exists between the cbg gene region of C. acetobutylicum and DNA from other Gram-positive bacteria tested. No detectable homology was observed between E. coli and the C. acetobutylicum cbg gene region. These results support earlier results of differences between the E. coli and C. acetobutylicum ß-galactosidase genes. The K. lactis ß-galactosidase gene has also been shown via immunoprecipitation and hybridization to be dissimilar to E. coli ß-galactosidase (lacZ) gene (Dickson and Markin, 1978). Gilbert and Hall (1987), however, found that there was strong homology between the operator region and the 5' end of the S. bovis lacZ gene and the equivalent E. coli gene. The other ß-galactosidase genes cloned from various organisms have not been compared to the E. coli ß-galactosidase gene.

The G+C content of the Gram-positive bacteria (S. thermophilus, 37-40%; S.

lactis ,35.5%; L. helviticus, 37-40%; L. bulgaricus, 49-51%) and E. coli (48-52%) are more similar to themselves than to C. acetobutylicum which is reported to have a G+C content of only 26-28% (Cato et al., 1986; Hardie, 1986; Kandler and Weiss, 1986; Orshov, 1986). However, from a taxonomic viewpoint, C. acetobutylicum is much closer related to the Gram-positive bacteria than to the Gram-negative E. coli.

5.4 SUMMARY

Complementation of the E. coli auxotrophic mutations (proA2, leuB6 and thi) was achieved, showing that C. acetobutylicum genes are expressed in E. coli.

The C. acetobutylicum pbg gene was not isolated, although detection and hybridization using ONPG-6-P were attempted. The lacY mutation of HB101 could not be complemented by the genomic library. Recombinant clones coding for the ß-galactosidase of NCIB2951 were isolated from the genomic library using the chromogenic substance X-gal. The cbg gene was shown to be located on a 5.2 kb EcoRI fragment. This fragment possessed few restriction sites. Expression of the gene occurred in E. coli when the 5.2 kb fragment was cloned in either orientation, showing that the cbg gene was expressed from its own promoter. The cbg gene was not induced by IPTG, in contrast to the E. coli gene. The cbg gene was expressed in E. coli in the presence of glucose, in contrast to the gene in C. acetobutylicum where no ß-galactosidase activity was measured in cells grown in the presence of glucose. Six strains of C. acetobutylicum were shown to possess a sequence sharing high homology to the cbg containing fragment, including ATCC824 which has no detectable levels of ß-galactosidase activity. The cbg gene region of NCIB2951 shared only low homology to the Gram-positive bacteria (S. lactis ATCC7962, S. thermophilus DRI1424, L. bulgaricus DRI20056, and L. helviticus DRO20064) but no detectable homology to E. coli DC272 or R. loti PN2231.

CHAPTER 6: TRANSPOSON TN5 SITE-DIRECTED MUTAGENESIS OF C. ACETOBUTYLICUM NCIB2951 ß-GALACTOSIDASE GENE REGION

6.1 INTRODUCTION

Chapter 5 describes how the cbg gene from the C. acetobutylicum NCIB2951 genomic library was isolated and subcloned into pSUP202 (pKH107) on a 5.2 kb EcoRI fragment. Restriction analysis of this fragment revealed too few sites to enable identification of a smaller region carrying the cbg gene and therefore site-directed mutagenesis was chosen as the method to physically map the exact location of the gene.

6.1.1 Transposons

Transposons have been used to introduce mutations into bacterial genomes and for site-specific mutagenesis of cloned DNA fragments (Kleckner et al., 1977; de Bruijn and Lupski, 1984). A commonly used transposon is the well characterized Tn5 (Kleckner et al., 1977; Jorgenson et al., 1979). Tn5 (5.7 kb) is a composite transposable element, composed of inverted repeats of the 1533 bp transposable sequence IS50 (Auserswald et al., 1981; Berg et al., 1982) surrounding a central 2650 bp unique segment. Tn5 encodes resistance to a number of aminoglycoside antibiotics, such as kanamycin, upon its host and thus is suitable for direct selection in the isolation of mutants (Kleckner et al., 1977).

Transposition of Tn5 appears to be a singular (and relatively non-specific) event, hence multiple mutations are avoided. If Tn5 insertion occurs within an operon it has a polar effect on genes downstream of the insertion (Kleckner et al., 1977). The absence of certain restriction sites, such as EcoRI, in Tn5 facilitates identification and molecular characterization of the Tn5-mutagenised genes (Jorgensen et al., 1979).

Site-directed mutagenesis, first described by Berg et al. (1975) is a method for introducing Tn5 into plasmids containing cloned DNA fragments, by using the suicide bacteriophage vector lambda 467
(b221 rex::Tn5 cI857, Oam29, Pam80). This system of site-specific mutagenesis has been used successfully in the characterization of Tn5-cloned DNA in the Klebsiella pneumoniae glnALG (glnAR) operon (de Bruijn and Ausubel, 1981); K. pneumoniae hisDGO region (de Bruijn and Ausubel., 1983); K. pneumoniae glnF (ntrA) gene (de Bruijn et al., 1983); E. coli dnaG region (Lupski et al., 1982); Salmonella typhimurium genomic DNA (Palva et al., 1981) and Streptococcus conjugative transposon Tn916 (Yamamoto et al., 1987).

This system involves infection of a suitable strain of bacterium (E. coli HB101) carrying a recombinant plasmid with bacteriophage lambda 467 which, due to the b221 deletion and amber mutations in the O and P genes, is incapable of integration or replication at 30°C. This causes Tn5 to transpose into the host cell genomic or plasmid DNA. Isolation of plasmid DNA and transformation allows selection of plasmids containing Tn5 insertions (de Bruijn and Lupski, 1984).

6.1.2 Aim

The aim of this phase of the project is to use the lambda 467 system to isolate site-specific mutants of plasmid pKH107 (containing the cbg gene of C. acetobutylicum NCIB2951) and to subsequently identify the position of the cbg gene. Secondly, the aim is to identify the protein products coded for by the 5.2 kb fragment.

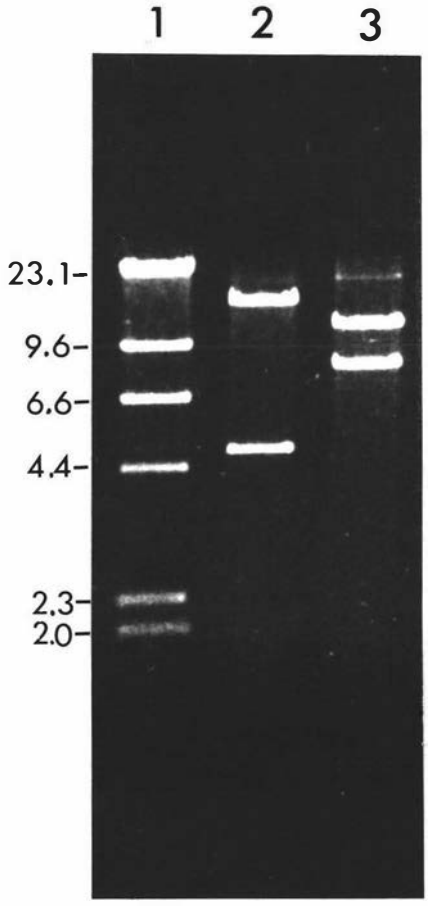
6.2 RESULTS

6.2.1 Site-specific mutagenesis of plasmid pKH107

To enable location of the C. acetobutylicum NCIB2951 cbg gene on the 5.2 kb EcoRI fragment, Tn5 mutagenesis was carried out using the bacteriophage lambda 467 (Berg et al., 1975). Since the b221 deletion in lambda 467 prevents lysogenation by this bacteriophage, Km^R colonies represent Tn5 transposition into either chromosomal or plasmid DNA. E. coli HB101 was transformed with pKH107 (section 2.21), mutagenised (section 2.26), and selection was carried out on LB plates containing tetracycline and kanamycin. The total number of cells was 1.05 $x10^9$/ml. The frequency of transfer of Km^R was 6.3 $x10^{-5}$ since the number of Km^R cells was 1.0 $x10^4$ per plate or 6.6 $x10^4$/ml of cells. The plasmid DNA isolated from the resultant colonies was used to transform E. coli HB101 and selection was made on LB plates containing tetracycline and kanamycin. From these colonies, plasmid DNA was prepared (section 2.7.3) and the isolated DNA was digested with restriction enzyme EcoRI (section 2.9) to determine whether the Tn5 had inserted in the 5.2 kb EcoRI fragment or within the vector pSUP202 itself (Fig. 6.1). Essentially all derivatives examined had Tn5 inserted into the plasmid. Since EcoRI does not cleave Tn5, the presence of Tn5 in any of the two EcoRI bands of pKH107 leads to the disappearance of one of the two expected fragments on the gel and the appearance of a new higher molecular weight fragment corresponding to a DNA size of the missing fragment plus 5.7 kb (the size of Tn5). Of the 118 recombinant plasmids tested 65 (56%) were shown to have Tn5 inserts in the 5.2 kb EcoRI fragment (demonstrated by the presence of a 10.9 kb band), plus an 8 kb band corresponding to the vector pSUP202 alone (Fig. 6.1, lane 3). These 65 pKH107::Tn5 derivatives were then plated onto LB plates containing tetracycline and kanamycin and designated a number (Table 6.1). The remainder were shown to contain the transposon in the vector, and possessed the 5.2 kb EcoRI insert band and a 13.7 kb band corresponding to pSUP202 containing Tn5 (Fig. 6.1, lane 2).

Fig.6.1: Agarose gel electrophoresis of EcoRI digests of Tn5-mutagenized pKH107 derivatives.

Lane 1 Lambda HindIII size standards
Lane 2 Plasmid with Tn5 in vector (pSUP202)
Lane 3 Plasmid with Tn5 in 5.2 kb insert


1
2
3
23.1-
9.6-
6.6-
4.4-
2.3-
2.0-

6.2.2 Physical analysis of plasmid derivatives

For precise mapping, the restriction enzymes that cleave within the inverted repeats of Tn5 are most useful, as they allow orientation-independent localisation of Tn5. A restriction enzyme from this group which cleaves the plasmid pKH107 once or twice provides a reference point relative to the Tn5 insertions, and thus allows exact positioning of the inserted Tn5. The location of Tn5 in the plasmid was therefore accurately mapped using the restriction enzyme HindIII. Tn5 has two HindIII sites, equidistant from the ends of the inverted repeats, giving three fragments; the internal 3.4 kb plus two outside 1.2 kb fragments. HindIII cuts the vector pSUP202 once, 3.2 kb from the EcoRI site of pSUP202. There are no HindIII sites in the insert.

Thus, HindIII digestion of a Tn5-mutagenised plasmid would free the internal Tn5 fragment and result in attachment of each outer Tn5 HindIII fragment to the remaining two insert and vector fragments (Fig. 6.2). To determine the position of the Tn5 in each of the recombinant plasmids (pKH108-pKH172), HindIII digests of the plasmid DNA (section 2.9) from each derivative were Southern blotted (section 2.19) and hybridised with the [^{32}P]-labelled 3.2 kb HindIII/EcoRI fragment of pSUP202 (see Figs. 6.3 - 6.7). The lower band corresponds to the internal HindIII fragment of Tn5 (3.4 kb) while each of the two remaining fragments contains one end of the inverted repeat (1.2 kb) plus a portion of the 5.2 kb EcoRI insert. Therefore, the value of 1.2 kb has to be subtracted from the measured size of the two larger fragments. This allowed identification of the right hand arm of the 5.2 kb fragment joined to the 3.2 kb HindIII fragment of pSUP202. The subsequent location of each Tn5 insert along the 5.2 kb EcoRI fragment was determined using the HindIII fragment sizes for each clone coupled with the hybridisation results. By measuring from the HindIII site of the vector, the exact position of each Tn5 was determined. These results are summarized in Fig. 6.7.

Fig.6.2: Diagram of a generalized pKH107 derivative containing a Tn5 insertion within the 5.2 kb EcoRI fragment, showing the area used as probe (see arrow).

EcoRI (E)
HindIII (H)

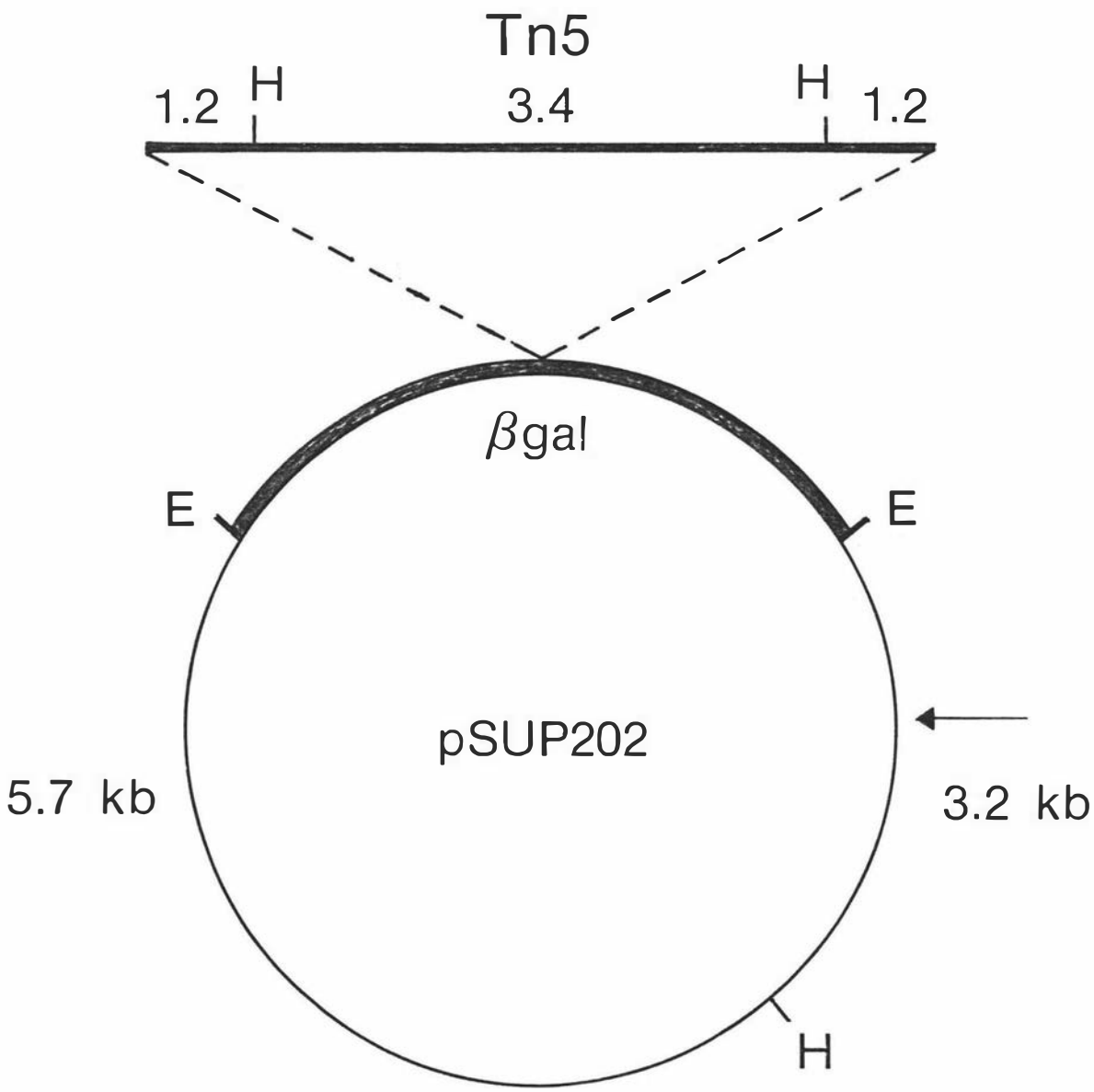

Tn5
1.2
H
3.4
H
1.2
βgal
E
E
pSUP202
5.7 kb
3.2 kb
H

Fig.6.3: Hybridization of the 3.2 kb EcoRI/HindIII fragment from pSUP202 to plasmids carrying the mutagenized 5.2 kb EcoRI ß-galactosidase fragment of NCIB2951.

(a) Agarose gel of HindIII digests of pKH107::Tn5 plasmid derivatives.
Size markers are in kilobases.

Lane 1	Lambda HindIII size standards
Lane 2	pSUP202 EcoRI/HindIII
Lane 3	pKH107 EcoRI
Lane 4	3/3 (pKH133)
Lane 5	A1 (pKH108)
Lane 6	A2 (pKH109)
Lane 7	A3 (pKH110)
Lane 8	A4 (pKH111)
Lane 9	A6 (pKH112)
Lane 10	B1 (pKH113)
Lane 11	B4 (pKH114)
Lane 12	C3 (pKH115)
Lane 13	C4 (pKH116)
Lane 14	D1 (pKH117)
Lane 15	D2 (pKH118)

(b) Autoradiograph of the same gel hybridized with the [^{32}P]-labelled 3.2 kb EcoRI/HindIII fragment from pSUP202. Size markers are in kilobases. [^{32}P]-labelled lambda DNA was included to identify the positions of the lambda standards.

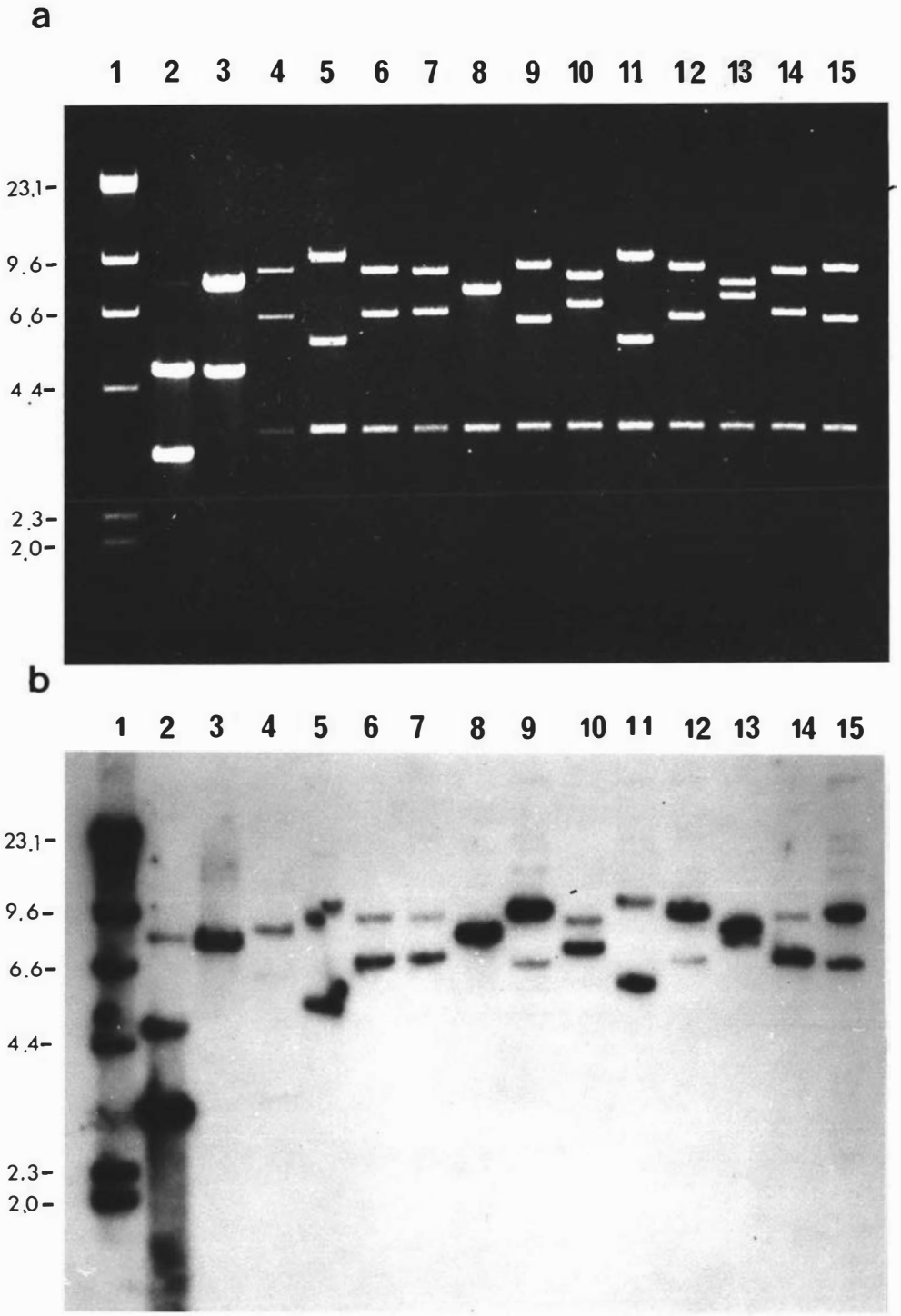

a
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
23.1-
9.6-
6.6-
4.4-
2.3-
2.0-
b
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
23.1-
9.6-
6.6-
4.4-
2.3-
2.0-

Fig.6.4: Hybridization of the 3.2 kb EcoRI/HindIII fragment from pSUP202 to plasmids carrying the mutagenized 5.2 kb EcoRI ß-galactosidase fragment of NCIB2951.

(a) Agarose gel of HindIII digests of pKH107::Tn5 plasmid derivatives.
Size markers are in kilobases.

Lane 1	D2 (pKH118)
Lane 2	Lambda HindIII size standards
Lane 3	D3 (pKH118)
Lane 4	D5 (pKH120)
Lane 5	E1 (pKH121)
Lane 6	E4 (pKH122)
Lane 7	F1 (pKH123)
Lane 8	F3 (pKH124)
Lane 9	F5 (pKH125)
Lane 10	G1 (pKH126)
Lane 11	G2 (pKH127)
Lane 12	G3 (pKH128)
Lane 13	G4 (pKH129)
Lane 14	H1 (pKH130)
Lane 15	H2 (pKH131)
Lane 16	H5 (pKH132)

(b) Autoradiograph of the same gel hybridized with the [^{32}P]-labelled 3.2 kb EcoRI/HindIII fragment from pSUP202.
Size markers are in kilobases. [^{32}P]-labelled lambda DNA was included to identify the position of the lambda standards.

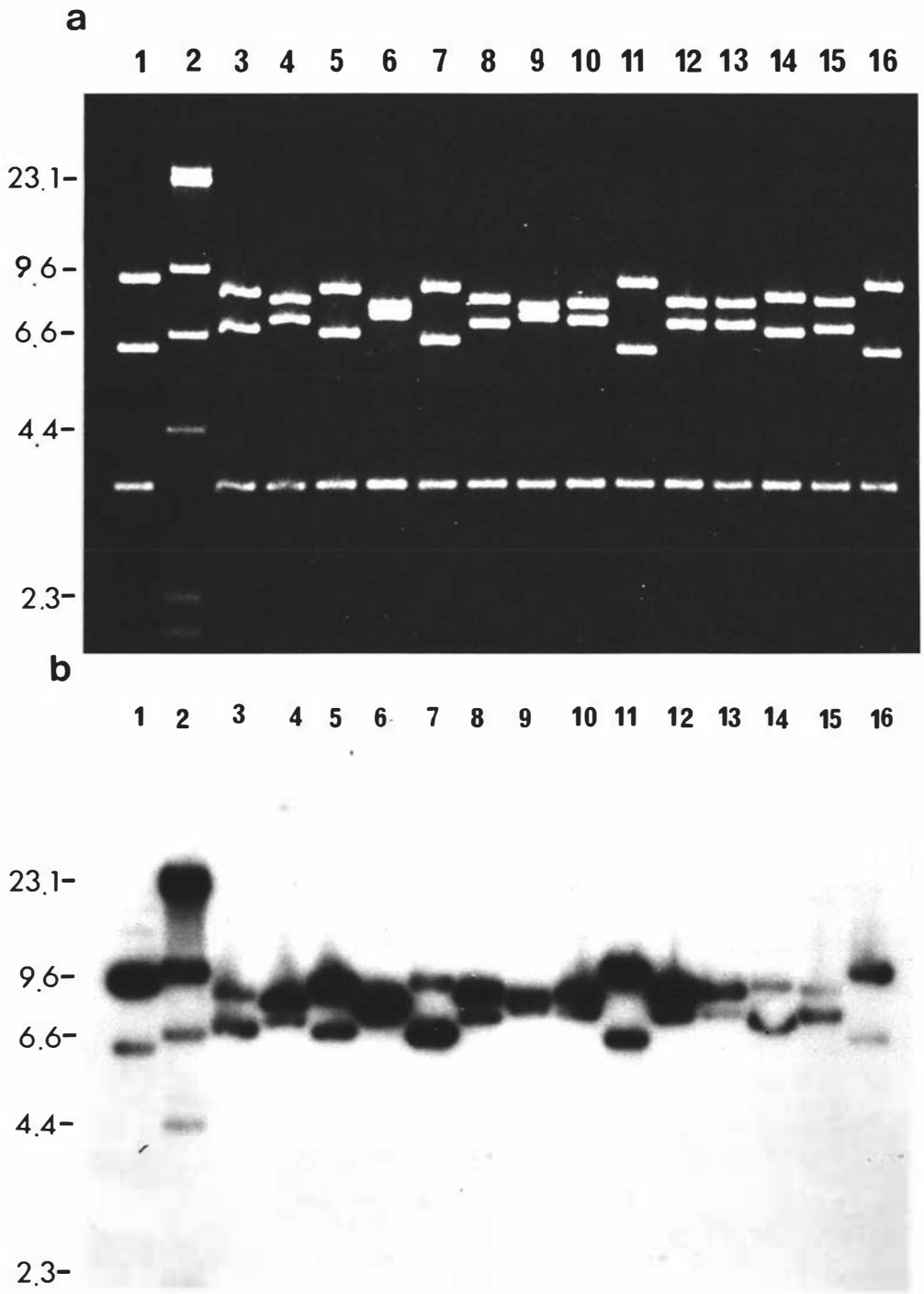

a
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
23.1-
9.6-
6.6-
4.4-
2.3-
b
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
23.1-
9.6-
6.6-
4.4-
2.3-

Fig.6.5: Hybridization of the 3.2 kb EcoRI/HindIII fragment from pSUP202 to plasmids carrying the mutagenized 5.2 kb EcoRI ß-galactosidase fragment of NCIB2951.

(a) Agarose gel of HindIII digests of pKH107::Tn5 plasmid derivatives.
Size markers are in kilobases.

Lane 1	I2 (pKH134)
Lane 2	I3 (pKH135)
Lane 3	Lambda HindIII size standards
Lane 4	I6 (pKH136)
Lane 5	J4 (pKH137)
Lane 6	J5 (pKH138)
Lane 7	K4 (pKH139)
Lane 8	K5 (pKH140)
Lane 9	L1 (pKH141)
Lane 10	L2 (pKH142)
Lane 11	L3 (pKH143)
Lane 12	L5 (pKH144)
Lane 13	M1 (pKH145)
Lane 14	N1 (pKH146)
Lane 15	N2 (pKH147)
Lane 16	N5 (pKH148)
Lane 17	O1 (pKH149)
Lane 18	O2 (pKH150)
Lane 19	O3 (pKH151)
Lane 20	O4 (pKH152)

(b) Autoradiograph of the same gel hybridized with the [^{32}P]-labelled 3.2 kb EcoRI/HindIII fragment from pSUP202.
Size markers are in kilobases. [^{32}P]-labelled lambda DNA was included to identify the position of the lambda standards.

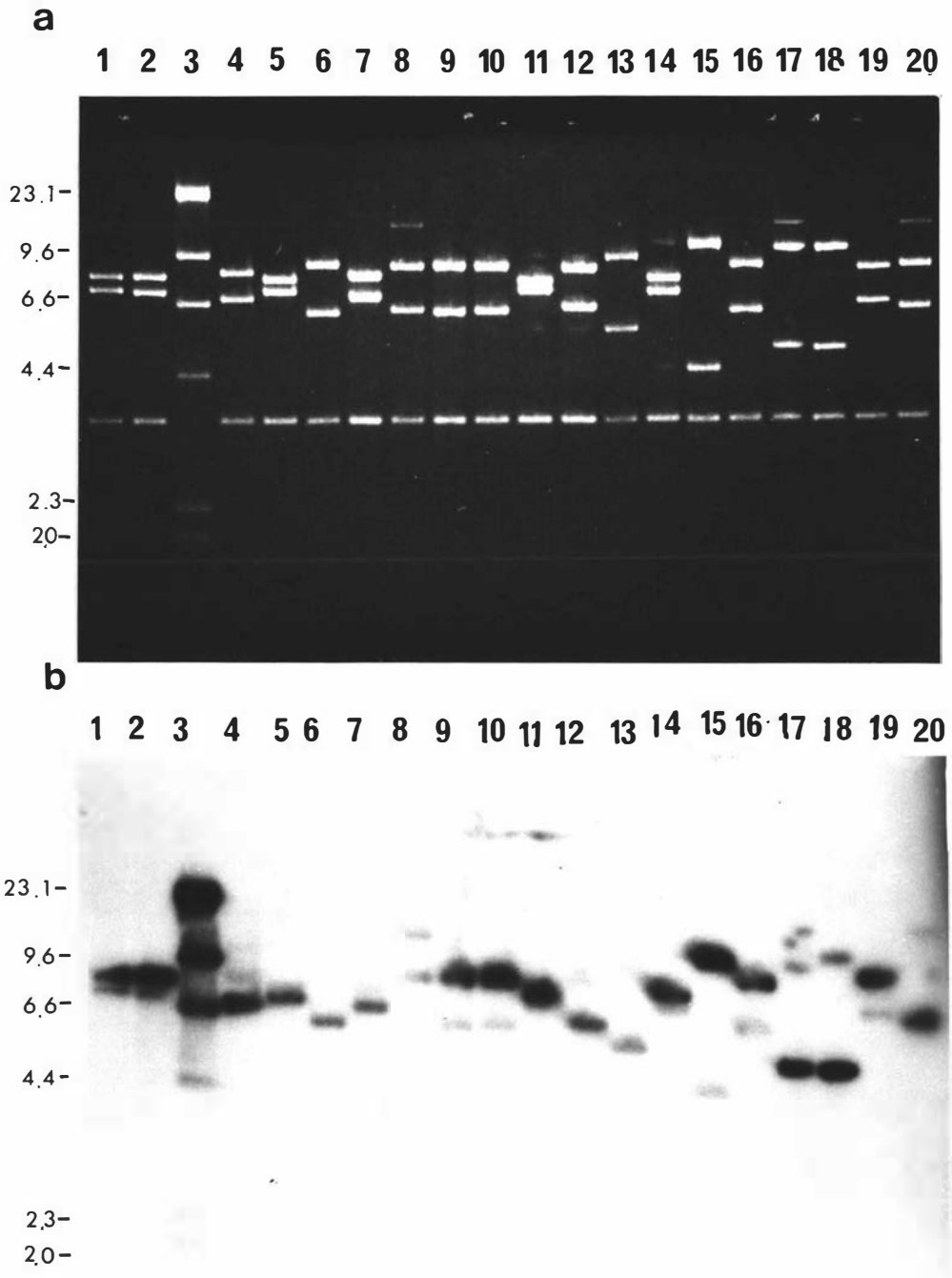

a
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
23.1-
9.6-
6.6-
4.4-
2.3-
2.0-
b
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
23.1-
9.6-
6.6-
4.4-
2.3-
2.0-

Fig.6.6: Hybridization of the 3.2 kb EcoRI/HindIII fragment from pSUP202 to plasmids carrying the mutagenized 5.2 kb EcoRI ß-galactosidase fragment of NCIB2951.

(a) Agarose gel of HindIII digests of pKH107::Tn5 plasmid derivatives.
Size markers are in kilobases.

Lane 1	O5 (pKH153)
Lane 2	O6 (pKH154)
Lane 3	P6 (pKH155)
Lane 4	T6 (pKH172) (Lambda HindIII size standards)
Lane 5	Q1 (pKH156)
Lane 6	Q2 nil
Lane 7	Q3 (pKH158)
Lane 8	Q3 (pKH159)
Lane 9	R2 (pKH160)
Lane 10	R3 (pKH161)
Lane 11	R4 (pKH162)
Lane 12	R6 (pKH163)
Lane 13	S1 (pKH164)
Lane 14	S3 (pKH165)
Lane 15	S4 (pKH166)
Lane 16	S6 (pKH167)
Lane 17	T1 (pKH168)
Lane 18	T2 (pKH169)
Lane 19	T3 (pKH170)
Lane 20	T4 (pKH171)

(b) Autoradiograph of the same gel hybridized with the [^{32}P]-labelled 3.2 kb EcoRI/HindIII fragment from pSUP202.
Size markers are in kilobases. [^{32}P]-labelled lambda DNA was included to identify the position of the lambda standards.

Fig.6.7: Map of pKH107 containing the 5.2 kb EcoRI ß-galactosidase fragment of NCIB2951 with the positions of the Tn5 inserts shown.
Those Tn5 inserts which did not inactivate the function of the cbg gene are shown above the line (+). Those Tn5 insertions which did inactivate the function of the cbg gene are shown below the line (-).

EcoRI (E)
HindIII (H)

The two genetic loci identified are shown as Locus I and II.

H
5.7 kb
N2
E
5.2 kb
E
3.2 kb
H
E
N5
T6
G2
A6
H5
D2
S6
R6
S4
S3
C3
K5
L1
L2
P6
T3
O3
E1
Q1
I3
I2
3-3
D5
F3
G3
G4
G1
C4
Q6
Q3
L3*
F5
E4
R2
L3*
J4
A4
T1*
T1*
N1*
I6
K4
O5
H2
T4
R3
B1
L5
O6
H1
A3
O4
A2
D1
F1
J5
M1
B4
S1
A1
R4
O1
T2
O2
E
H

Table 6.1: ß-galactosidase activity in Tn5 mutagenized derivatives of pKH107

Activity of ß-galactosidase was detected on plates using X-gal. Blue colonies, resulting from the hydrolysis of X-gal were scored as positive (+). White colonies, with no evident hydrolysis of X-gal, were scored as negative (-).

No.	Strain No.	Plasmid No.	Activity
(a)	Controls		
	PB2959	-	-
	DC272	-	+
	DRI5014	pKH107	+
(b)	Tn5 derivatives		
	DRI		
A1	(5020)	pKH108	+
A2	(5021)	pKH109	-
A3	(5022)	pKH110	-
A4	(5023)	pKH111	-
A6	(5024)	pKH112	-
B1	(5025)	pKH113	-
B4	(5026)	pKH114	+
C3	(5027)	pKH115	+
C4	(5028)	pKH116	-
D1	(5029)	pKH117	-
D2	(5030)	pKH118	-
D3	(5031)	pKH119	+
D5	(5032)	pKH120	-
E1	(5033)	pKH121	-
E4	(5034)	pKH122	-
F1	(5035)	pKH123	-
F3	(5036)	pKH124	-
F5	(5037)	pKH125	-
G1	(5038)	pKH126	-
G2	(5039)	pKH127	-
G3	(5040)	pKH128	-
G4	(5041)	pKH129	-
H1	(5042)	pKH130	-
H2	(5043)	pKH131	-
H5	(5044)	pKH132	-

Table 6.1 cont'd

No.	Strain No.	Plasmid No.	Activity
I2	(5060)	pKH134	-
I3	(5061)	pKH135	-
I6	(5062)	pKH136	-
J4	(5063)	pKH137	-
J5	(5064)	pKH138	-
K4	(5065)	pKH139	-
K5	(5066)	pKH140	+
3/3	(5046)	pKH133	-
L1	(5067)	pKH141	+
L2	(5068)	pKH142	+
L3	(5069)	pKH143	-
L5	(5070)	pKH144	-
M1	(5049)	pKH145	-
N1	(5050)	pKH146	-
N2	(5051)	pKH147	+
N5	(5052)	pKH148	+
O1	(5053)	pKH149	+
O2	(5054)	pKH150	+
O3	(5055)	pKH151	-
O4	(5056)	pKH152	-
O5	(5057)	pKH153	-
O6	(5058)	pKH154	-
P6	(5059)	pKH155	+
Q1	(8503)	pKH156	-
Q2	(8504)	pKH157	-
Q3	(8505)	pKH158	-
Q6	(8506)	pKH159	-
R2	(8507)	pKH159	-
R3	(8508)	pKH160	-
R4	(8509)	pKH162	+
R6	(8510)	pKH163	+
S1	(8511)	pKH164	+
S3	(8512)	pKH165	+
S4	(8513)	pKH166	+
S6	(8514)	pKH167	+
T1	(8515)	pKH168	-
T2	(8516)	pKH169	+
T3	(8517)	pKH170	+
T4	(8518)	pKH171	-
T6	(8519)	pKH172	+

6.2.3 ß-galactosidase activity in tn5 mutagenized derivatives of pKH107

In order to characterise whether the Tn5 insertions within the 5.2 kb EcoRI fragment had interrupted the ability of the cbg gene to function, each of the pKH107::Tn5 derivative plasmids were directly screened for their ability to metabolise X-gal. The derivative plasmids in E. coli strain HB101 were therefore crossed into E. coli PB2959 using the helper plasmid pRK2013 (section 2.22), and, after overnight incubation at 37°C, a loopful was plated for single colonies on supplemented minimal medium agar plates containing kanamycin, tetracycline and X-gal, and incubated overnight at 37°C. The various recombinant clones were then scored for activity upon the presence or absence of the blue indigo dye on single colonies. These results demonstrated that 20 (31%) still possessed normal gene function while 45 (69%) had no apparent ß-galactosidase activity and were thus scored negative (Table 6.1). Summary of these results and map positions are shown in Fig. 6.7. Two genetic loci (I and II) were identified as necessary for expression of ß-galactosidase in E. coli. Locus I is approximately 400 bp long and lies within the small EcoRI/HaeIII fragment. Locus II is approximately 3.2 kb long and lies within the large HaeIII/EcoRI fragment.

6.2.4 Maxi-cell analysis

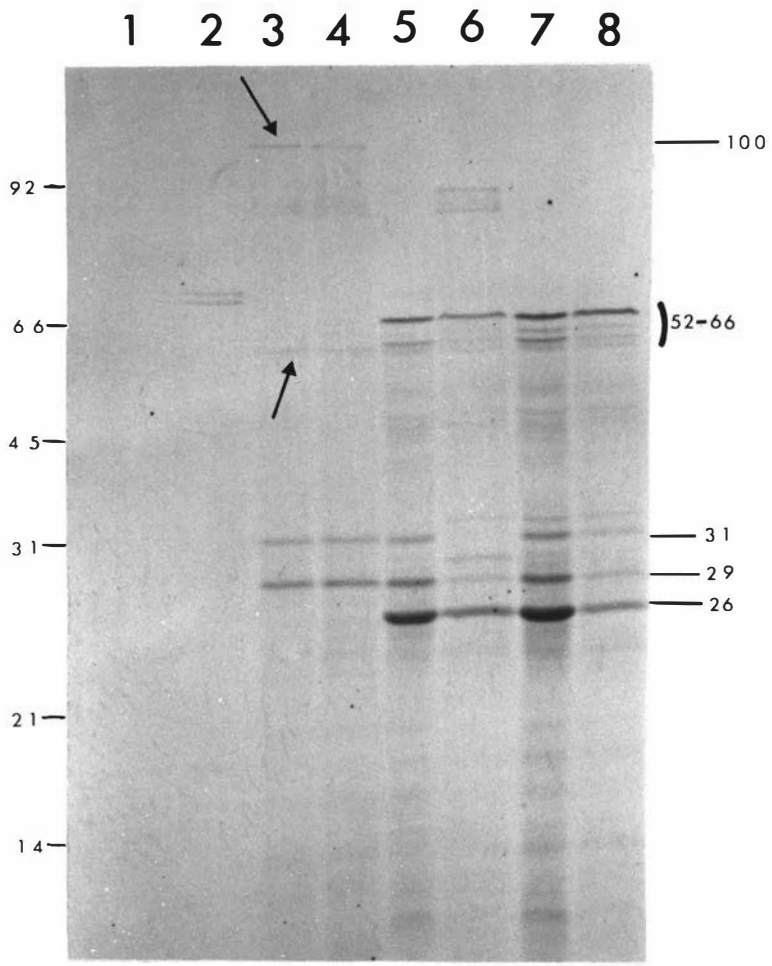
In order to investigate the protein products encoded by genes on the 5.2 kb fragment of NCIB2951, various inserts containing the NCIB2951 5.2 kb EcoRI region were cloned into the vector pAD10 and analyzed in E. coli maxi-cells as described in Section 2.27. The plasmid pAD10 is an expression vector derivative of pAD7 containing the S. typhimurium trp promoter and the E. coli rpoC terminator in pUC8 (Egelhoff and Long, 1985). The plasmid pKH107 was digested with EcoRI (section 2.9) and ligated into EcoRI digested pAD10 as described in section 2.27. The ligated DNA was transformed into E. coli HB101 (section 2.21), plated onto LB plates containing Ap and X-gal, and incubated overnight at 37°C. To enable isolation of those transformants which carried the

5.2 kb insert in the vector pAD10, 50 of the blue colonies were then replica-plated onto LB plates (containing Ap and Tc respectively), and incubated overnight at 37°C. Colonies containing plasmids resistant to Ap but sensitive to Tc should contain inserts in pAD10. Therefore, six of these colonies were selected, plasmid DNA was isolated from the colonies (section 2.7.3), digested with EcoRI (section 2.9), and analyzed by gel electrophoresis as described in section 2.12. All of the six clones contained the 5.2 kb EcoRI insert carrying the ß-galactosidase gene region plus the 3.2 kb expression vector pAD10. To determine the orientation of the insert these clones were digested with PvuII and examined by electrophoresis. The results established that three of the six transformants had the insert in one orientation, and three in the opposite orientation. Two transformants, (DRI8533 and DRI8534), one representing each orientation, were chosen for maxi-cell analysis, and the plasmids were designated pKH184 and pKH185.

Four pKH107::Tn5 plasmid derivatives (pKH112, pKH127, pKH124 and pKH153) (from the ß-galactosidase locus I and locus II respectively; Fig. 6.7) were also digested with EcoRI (section 2.9) and the resulting fragments were ligated into EcoRI digested pAD10 as described in section 2.20. The resulting ligated DNA was transformed into E. coli HB101 (section 2.21), plated onto LB plates containing Ap and Km, and incubated overnight at 37°C. Only those transformants carrying a vector containing the insert with the Tn5 would be able to grow in the presence of Km. Fifty of the resulting colonies from each ligation were replica-plated onto LB plates (containing Ap, Km and Ap, Km and Tc respectively) and incubated overnight at 37°C. Six Tc sensitive colonies were chosen from each plate, and plasmid DNA was isolated (section 2.7.3), digested with EcoRI (section 2.9) and examined by gel electrophoresis as described in section 2.12. The digests showed that all four groups of six plasmid profiles consisted of a 3.2 kb vector fragment (pAD10) plus a large 10.9 kb fragment corresponding to the 5.2kb::Tn5 insert. The transformants were then digested

Fig.6.8: Autoradiograph of a 6% SDS-PAGE gel showing the plasmid-encoded proteins produced by in vivo [^{35}S]-methionine labelling of maxi-cells.
Migration of protein standards (left) and positions of plasmid encoded proteins (right) are indicated in kilodaltons (kDa).

Lane 1: E. coli HB101
Lane 2: E. coli HB101 plus pAD10
Lane 3: E. coli HB101 containing pKH184
Lane 4: E. coli HB101 containing pKH185
Lane 5: E. coli HB101 containing pKH189
Lane 6: E. coli HB101 containing pKH188
Lane 7: E. coli HB101 containing pKH187
Lane 8: E. coli HB101 containing pKH186


1
2
3
4
5
6
7
8
100
92
66
52-66
45
31
31
29
26
21
14

with PvuII to determine the orientation of the insert within the vector and to confirm the identity of the insert. Four transformants (DRI8535 to DRI8538), one representing each Tn5:insert derivative, were chosen for maxi-cell analysis. The plasmids within these strain were designated (pKH186-pKH189).

Maxi-cells were then prepared from these E. coli strains as described in section 2.27, and analyzed as described in section 2.28. The results obtained are shown in Figure 6.8. The major non-plasmid encoded protein is the 100 kDa protein, presumably corresponding to the ß-galactosidase (lanes 3 and 4). This protein is absent in the Tn5 inserts in both Locus I and II (lanes 5 to 8). With the exception of a protein band at 54 kDa (lanes 3 and 4) all other major protein bands observed correspond to known plasmid (pAD10) and transposon (Tn5) gene products. These include the 29 kDa processed ß-lactamase and 31 kDa unprocessed ß-lactamase gene products from pAD10 (lanes 3 to 8; Egelhoff and Long, 1985). The 26 kDa neomycin phospho-transferase (NPT II) gene product plus a group of proteins (in the range of 52 to 66 kDa) corresponding to the transposase gene product from Tn5 are also evident (lanes 5 to 8: Fuhrmann and Hennecke, 1982). The identity of the 54 kDa protein visible in lanes 3 and 4 is not known. In the experimental results shown in Figure 6.8, no ß-lactamase proteins were observed in the plasmid control (lane 2). However, in other experiments these were shown to be synthesised. The reason for the absence of these proteins in lane 2 in this experiment could be due to the poor growth observed with the cells used in this sample.

6.3 DISCUSSION

6.3.1 Tn5 insertion specificity

By using the Tn5 carrying vector lambda 467 (De Bruijn and Lupski, 1984) the transposon Tn5 was successfully introduced into the bacterial host DRI5014 containing pKH107 at a high frequency. Upon isolation and transformation of

plasmid DNA, 106 Tn5-mutagenised derivatives of pKH107 were isolated. Analysis of the plasmid profiles via EcoRI digests of plasmid DNA confirmed that all these mutants had Tn5 inserted into either the insert or the vector of pKH107. Of the 106 plasmid derivatives studied (65) 56% had transposed into the 5.2 kb EcoRI fragment known to carry the C. acetobutylicum NCIB2951 ß-galactosidase gene. The pKH107::Tn5 derivatives known to have the transposon positioned into the 5.2 kb EcoRI fragment were characterized further via HindIII restriction enzyme digestion, followed by Southern blotting and hybridization using [^{32}P]-labelled 3.2 kb HindIII/EcoRI fragment of the vector pSUP202 as probe.

Tn5 has been ranked among the transposon elements with relatively low insertional specificity, (de Bruijn et al., 1983), and is classified as essentially random in transposition. Tn5 has been shown to be capable of inserting into numerous alternative sites in target DNA, but does show some degree of preferential target regulation selection (Shaw and Berg, 1979; Miller et al., 1980; Berg et al, 1980; Bossi and Ciampi, 1981). Theoretically, if transposition of Tn5 insertions within pKH107 were totally random, one would expect percentages of 39% and 61% into the 5.2 kb EcoRI insert and the 8 kb pSUP202 vector respectively. However, the results obtained were not as expected, showing a far greater number of Tn5 transposons (56%) inserted into the smaller fragment (5.2kb EcoRI insert). Of these 65, the map positions in Fig. 6.7 show that the distribution of targets for Tn5 insertions within the insert of pKH107 was also non-random. There was an over-representation of insertions into the left of the fragment and very few insertions, in the right hand side of the fragment. Also, in some areas, "hot spots" occurred with a high number of insertions, whereas in other places no insertions occurred at all (Fig. 6.7). A possible explanation is the A+T richness of C. acetobutylicum DNA, which has a G+C content of only 28% (Cummins and Johnson, 1971).

Miller et al. (1980), using Tn5 mutagenesis of the E. coli lac operon, noted that

Tn5 insertions followed a similar pattern to that observed with Tn9 (though with less preference), and identified moderate correlation between A+T richness and regions of preferential integration, especially the region extending from the end of the lacZ through the entire lacY gene. These results suggest that, under the conditions used here, there are preferred areas of transposition. This has also been reported by McKlinnon et al. (1985), who found non-random distribution of insertion sites within the 5.9 kb fragment of human adenovirus DNA.

Several types of evidence have raised the possibility of a role for transcription in the regulation of Tn5 transposition. Berg et al. (1983) showed that there is existence of "hot spots" immediately downstream from the promoter of the tetracycline gene on pBR322. McKlinnon et al. (1985) also reported that transcriptional activity may have influenced Tn5 transposition, as the majority of insertions mapped downstream of a fortuitous promoter sequence. Removal of this promoter sequence enabled Tn5 transposition into previously unused upstream target sequences. However, identification of any promoter sequences and their positions on the insert has not as yet been investigated. Similar results have been reported by Miller et al. (1980) in the E. coli lacZYA genes. Distribution of Tn5 insertions showed that Tn5 inserted at a higher frequency than the other transposons that they used. Insertions were more uniformly distributed in the lac operon, with integration being preferred in the entire lacY gene, and at the very end of the lacZ gene (Berg et al., 1980).

6.3.2 Identification of two gene loci for ß-galactosidase expression

Herman and McKay (1986), upon cloning and characterizing the ß-galactosidase of S. thermophilus ATCC19258, reported that the ß-galactosidase gene was located on a portion of the 3.4 kb HindIII fragment and that the 0.45 kb PstI/HindIII fragment which adjoins it on the left was required for stable expression of the ß-galactosidase gene. This would suggest two loci responsible for the ß-galactosidase function in S. thermophilus. This is similar to the findings

in the present study, which also show evidence for two gene loci, separated by approximately 500 bp. Since region II is totally on the large HaeIII/EcoRI gene region, yet is unable to function in E. coli, there must be some regulatory or structural gene product encoded by smaller fragment which is required for expression of the ß-galactosidase.

Tn1000 mutagenesis was used by Gilbert and Hall (1987) to analyze the lac genes of S. bovis. The lac genes in this organism are organized as a single operon as found in E. coli. They reported that Tn1000 insertions which had inserted in the region where the repressor resided resulted in constitutive synthesis of all the lac gene products (lacZYA). Tn 1000 insertions were also shown to have a polar effect on the other lac genes downstream in the operon. However, no Tn5 insertions in the cbg of NCIB2951 caused an increase in ß-galactosidase activity, nor gave a "leaky" mutation or constitutive synthesis. Since the expression of the C. acetobutylicum cbg gene is not repressed by glucose in E. coli, it is difficult to identify the isolation of any constitutive mutations.

The identification by Tn5 transposon mutagenesis of two cbg loci on the NCIB2951 5.2 kb EcoRI fragment raises the possibility that locus I, corresponding to a DNA region of less than 400 bp is a positive regulatory element. There is no report to date for a positive regulatory element associated with bacterial lac genes. However, the yeast Kluyveromyces lactis has been shown to be positively regulated (Ruzzi et al., 1987). Another possibility is that locus I is a structural sub-unit of the C. acetobutylicum ß-galactosidase enzyme. While E. coli (Campbell et al., 1973), Klebsiella (Hall and Reeve, 1977) and B. stearothermophilus (Hirata et al., 1984) may contain two different ß-galactosidase genes, there is no evidence to date for a bacterial ß-galactosidase being composed of two non-identical sub-units. The E. coli α-complementation selection system demonstrates that the E. coli ß-galactosidase can be spatially separated into two genetic components (Ullman and Perri, 1970). However, neither the 1.2 kb nor the 4.2 kb HaeIII/EcoRI fragments, when cloned into E.

coli JM101, were able to complement the defected E. coli ß-galactosidase and produce a hybrid, functional ß-galactosidase. The cbg locus II region is of a size, 3.2 kb, that would suggest that it encodes for the ß-galactosidase structural gene.

6.3.3 Gene products of the ß-galactosidase gene region

Cloning of the insert in either orientation in pAD10 did not affect the expression of the ß-galactosidase, providing further evidence that a native clostridial promoter is responsible for initiating transcription of the ß-galactosidase on the 5.2 kb fragment. The observed Mr (100 kDa) for the presumed ß-galactosidase structural gene is similar to those reported by other groups (105 kDa from S. thermophilus (Somkuti and Steinberg, 1979; Herman and McKay, 1987); 135 kDa from K. lactis (Dickson and Markin, 1978); 135 kDa from E. coli (Craven et al., 1965); (ß-gal I) 135 kDa, (ß-gal II) 120 kDa and (ß-gal III) 68 kDa from Klebsiella strain RE1544 (Hall and Reeve, 1977); (bgaA) 120 kDa, (bgaB) 70 kDa from B. stearothermophilus (Hirata et al., 1984).

6.4 SUMMARY

Using Tn5 mutagenesis of the 5.2kb ß-galactosidase gene region, two loci were identified that are necessary for ß-galactosidase expression in E. coli. The first locus (I) was 400 bp long and situated approximately 500 bp from the second locus (II), which was 3.2 kb long. There was preferential insertion of Tn5 towards the left of the cloned 5.2 kb insert. Maxi-cell analysis of the gene products encoded by the 5.2 kb EcoRI fragment identified a 100 kDa protein being coded for by the cbg gene. Tn5 insertion mutants of ß-galactosidase failed to produce this protein.

CHAPTER 7: NUCLEOTIDE SEQUENCE AND GENETIC ORGANIZATION OF THE 5.2 KB ECORI FRAGMENT

7.1 INTRODUCTION

7.1.1 M13 biology

M13 is a male specific filamentous bacteriophage which can adsorb to strains of E. coli containing the F sex plasmid (Fig. 7.1). The bacteriophage particles contain one circular single-stranded (SS) DNA molecule, called the (+) strand. Following infection via the F pilus, synthesis of a complementary (-) strand occurs, converting the SS DNA into the double-stranded (DS) circular replicative form (RF). The RF molecules are replicated throughout the life cycle to reach approximately 200 M13 per cell. Replication of the RF form then ceases, and progeny (+) strands are produced via the 'rolling ball' method, packaged in the bacteriophage coat, and extruded from the cell by a non-lytic mechanism. Hence, M13 allows simple and efficient generation of SS DNA templates (Yanisch-Perron et al., 1985). Even though cell lysis does not occur, the host cell growth rate is reduced by the presence of M13. This results in the production of detectable turbid plaques.

7.1.2 Modification of M13 for production of templates

This M13 system has been modified to offer several advantageous features for the generation of DNA templates for sequencing (Messing et al., 1977; Schreier and Cortese, 1979; Sanger et al., 1980). These are the insertion of the α-peptide of the E. coli lacZ gene to enable easy detection of insertions within M13, and the addition of multiple cloning plus universal primer sites (Fig. 7.2).

Fig.7.1: Diagram of the M13 life cycle.

(1) The bacteriophage enters the bacterial host cell via the F pilus.
(2) Single-stranded (SS) DNA is converted to double-stranded replicative form (RF).
(3) Replication of the RF DNA occurs to give progeny RF (200 per cell).
(4) Progeny (+) strand synthesis occurs via the 'rolling ball' method.
(5) The (+) SS DNA is packaged into the bacteriophage coat, and extruded from the cell via a non-lytic mechanism.

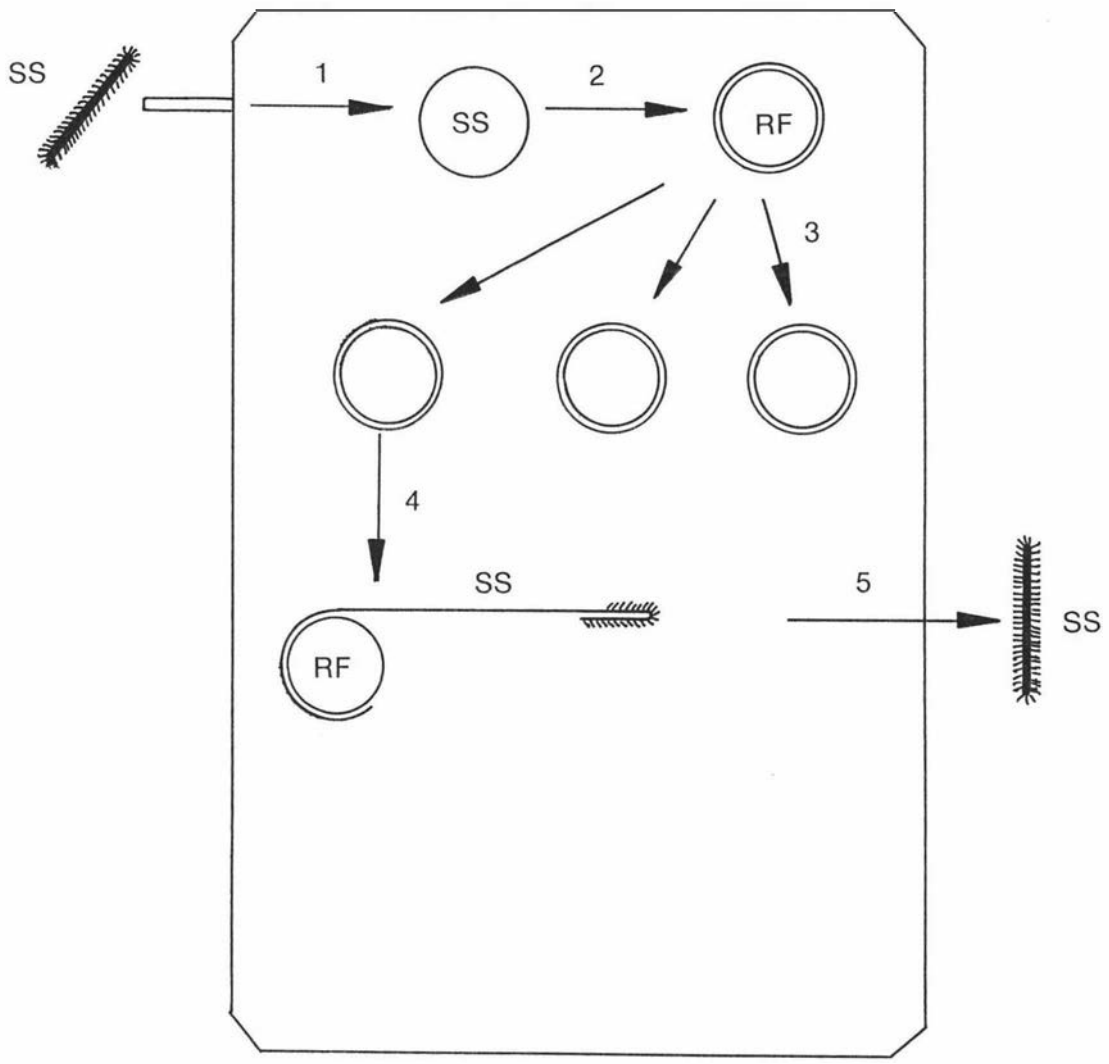
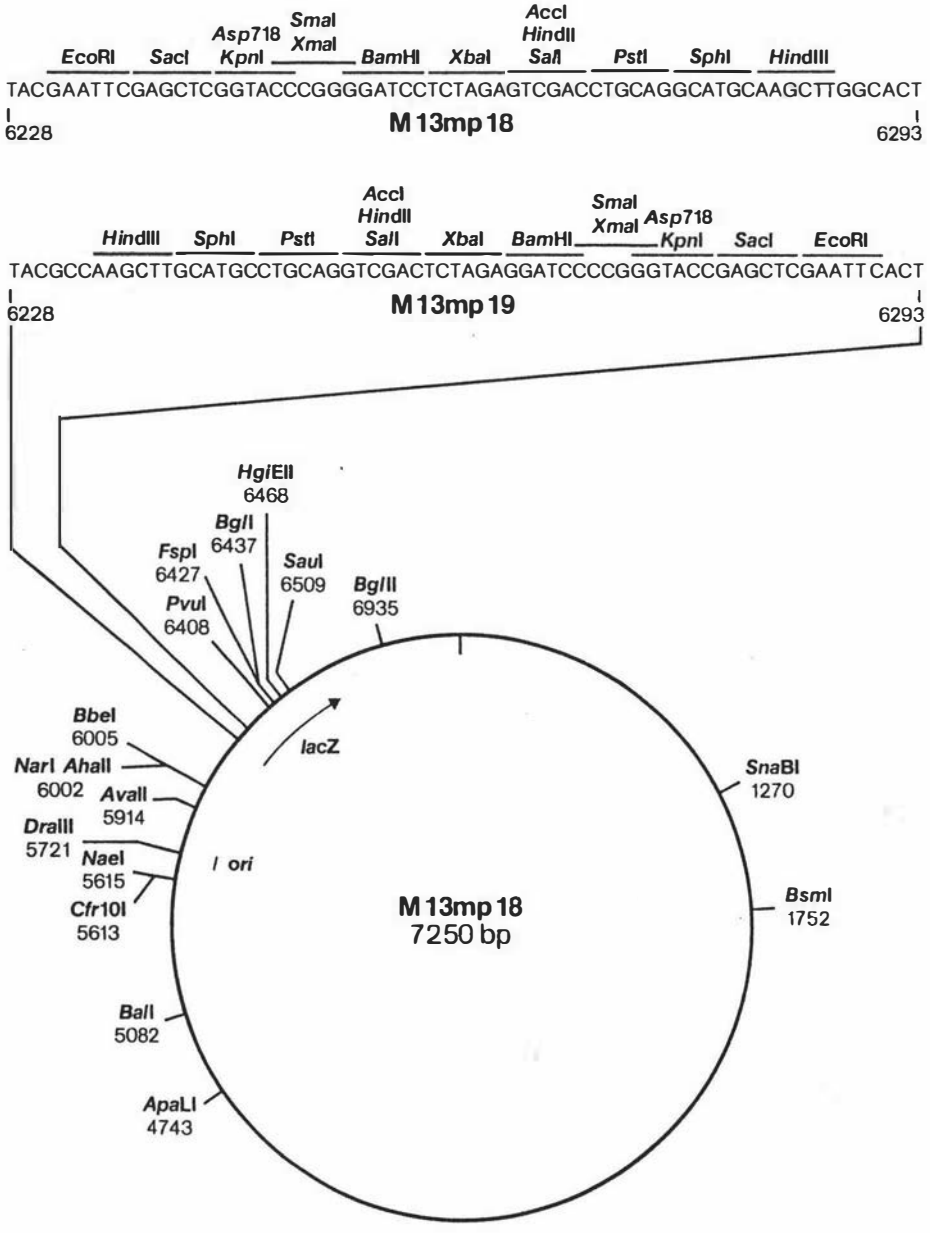

SS
1
SS
2
RF
3
4
SS
5
SS
RF

Fig.7.2: Diagram of modified M13 for use in the production of SS templates for sequencing.

The lacZ α-peptide is shown, with the pair of polylinkers used in the project (mp18 and mp19). The primer sites are found to the left and the right of the polylinker position (Boehringer Mannheim, 1987).


EcoRI SacI Asp718 KpnI SmaI XmaI BamHI XbaI AccI HindII SalI PstI SphI HindIII
TACGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCACT
6228
M13mp18
6293
HindIII SphI PstI AccI HindII SalI XbaI BamHI SmaI XmaI Asp718 KpnI SacI EcoRI
TACGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCACT
6228
M13mp19
6293
HgiEII 6468
BglI 6437
FspI 6427
PvuI 6408
SauI 6509
BglII 6935
BbeI 6005
NarI AhaII 6002
AvaII 5914
DraIII 5721
NaeI 5615
Cfr10I 5613
f ori
lacZ
SnaBI 1270
BsmI 1752
BalI 5082
ApaLI 4743
M13mp18
7250 bp

The α-peptide of the E. coli lacZ gene was inserted into an intergenic region of M13. A mutated form of the lacZ gene lacking the α-peptide was introduced on a F plasmid into a lac deletion strain, JM101 (Messing et al., 1977). Parental M13 plaques on JM101 therefore allow production of an active ß-galactosidase via α-complementation, which is detectable by the cleavage of X-gal. Hence, parental M13 plaques turn blue.

A number of different polylinker sites have been engineered into the lacZ α-peptide with the M13 genome in such a way as not to interfere with α-complementation. These polylinkers, engineered in pairs with respect to the primer site, permit easy cloning of insert DNA within M13 (Messing and Vieira, 1982; Vieira and Messing , 1982; Messing, 1983; Norrander et al., 1983). Recombinant clones containing insert DNA in the polylinker site interfere with the open reading frame (ORF) of the α-peptide. This results in no α-complementation, nor production of an active ß-galactosidase. Hence, recombinant M13 plaques remain white. Universal primers were established upstream of the polylinker site, the primers being independent of the insert DNA. This allows easy 5' to 3' extension into insert DNA (Heidecker et al., 1980; Anderson et al., 1980; Norrander et al., 1983).

Due to the filamentous nature of M13 packaging, there are no constraints on the size of DNA that can be packaged. Although very large inserts can be cloned, their presence can cause deletions. This disadvantage has been overcome by the construction of new E.coli strains (JM105, JM109) which give improved stability to large DNA inserts or to those containing repetitive sequences (Yanisch-Perron et al., 1985).

7.1.3. Generation of insert templates

Several strategies have been developed for the generation of templates which allow sequencing of DNA. These include random (shotgun cloning, DNase

treatment, and sonication), and non-random (Bal31 and exonuclease III digestion) methods (Smith, 1979; Poncz et al., 1982; Deininger, 1983). The random methods provide an array of fragments representing the entire DNA fragment to be cloned. This random generation of clones suffers from a number of problems. There may be regions which are totally devoid of sites, and small fragments clone much more readily into M13 and therefore get sequenced more often.

The generation of clones via a non-random method overcomes these disadvantages. Nuclease Bal31 possesses a double-stranded exonuclease activity which removes base pairs from both ends of linear DNA (Poncz et al., 1982). Successive deletions of a linear fragment of DNA can therefore be achieved.

7.1.4 Aim

The aim of this phase of the project was to sequence the 5.2 kb EcoRI fragment in order to obtain more insight into the genetic organisation of the cloned ß-galactosidase region.

7.2 RESULTS

7.2.1 Preparation of plasmid DNA

In order to isolate successive deletions of the 5.2 kb EcoRI fragment, a series of Bal-31 digests was performed. Four separate plasmids were used which provided entire coverage of the 5.2 kb fragment (Fig. 7.3). The construction of pKH173 (containing the large HaeIII/EcoRI fragment in pUC18), and pKH174 (containing the small EcoRI/HaeIII fragment in pUC18) have been described in section 5.2.7. The 1.8 kb PvuII internal fragment, and the 3.2 kb PvuII/EcoRI end fragment were also cloned (Table 7.1). Plasmid pKH107 was digested with PvuII and PvuII/EcoRI, separated by gel electrophoresis (section 2.12), and the

Fig.7.3: Sequencing strategy for the 5.2 kb EcoRI fragment containing the cbg gene of C. acetobutylicum NCIB2951. A restriction map is also shown.

(E) EcoRI
(H) HindIII
(P) PvuII

The arrows indicate origin, direction and extent of the individual sequencing reactions.

E
P
H
P
E
E
E
H
H
H
E
E
P
P
P
P
E
E

Table 7.1

Table describing the plasmids used for the preparation of Bal31 deletion derivatives of the 5.2 kb EcoRI fragment.

Plasmid	Insert	Length (kb)	DNA lot	Enzyme I	Enzyme II	M13	Cloning site
pKH173	HaeIII/EcoRI	4.2	A	EcoRI	HindIII	mp19	SmaI/HindIII
			B	HindIII	EcoRI	mp18	SmaI/EcoRI
pKH174	EcoRI/HaeIII	1.0	C	EcoRI	HindIII	mp19	SmaI/HindIII
			D	HindIII	EcoRI	mp18	SmaI/EcoRI
pKH183	PvuII	1.8	E	EcoRI	HindIII	mp19	SmaI/HindIII
			F	HindIII	EcoRI	mp18	SmaI/EcoRI
pKH182	PvuII/EcoRI	3.2	G	EcoRI	HindIII	mp19	SmaI/HindIII
			H	HindIII	EcoRI	mp18	SmaI/EcoRI

required fragments were eluted by the phenol-freeze method (section 2.16), and ligated into SmaI digested pUC18 as described in section 2.20. The ligation mixes were transformed into HB101 (section 2.21), plated on LB plates supplemented with Ap, X-gal and IPTG, and incubated overnight at 37°C. Six white colonies from each plate were chosen, and plasmid DNA was isolated using the rapid boiling method (section 2.7.3). The DNA isolated from the PvuII ligation clones were digested with EcoRI and EcoRI/HindIII (section 2.9) to confirm the presence of the 1.8 kb PvuII band. All EcoRI/HindIII digests gave two bands, a 3.2 kb band corresponding to pUC18, plus a 1.8 kb band corresponding to the internal PvuII fragment. All EcoRI digests gave a single band corresponding to the 1.8 kb plus 3.2 kb bands together. One such plasmid was chosen and designated pKH183.

Since the large PvuII/EcoRI insert is the same size as the pUC18 vector (3.2 kb), only EcoRI digests of the EcoRI/PvuII ligation clones were carried out. All digests resulted in the production of a single 6.4 kb band corresponding to pUC18 (3.2 kb) plus the 3.2 kb EcoRI/PvuII fragment. One such plasmid was chosen and designated pKH182.

7.2.2 Nuclease Bal-31 digestion

Since enzyme proportionality exists over a narrow DNA concentration range, activity was carefully tested with each DNA preparation using suitable DNA markers (Table 7.1).

For each of the four plasmids used, 40 μg of DNA was digested with EcoRI or HindIII separately to open the plasmids on the right or the left of the insert. The enzyme activity necessary for each experiment was determined by varying the enzyme quantity as described in section 2.29. A series of digests were set up for each of the 8 lots of DNA (A to H: Table 7.1) using different dilutions of Bal31 (1/2 to 1/32). These were incubated at 37°C for 30 min, stop dye added (section 2.9), and the DNA was examined by gel electrophoresis as described in section 2.12. From these results, the amount of enzyme required to give complete digestion of each linear plasmid was determined. The time steps required to give 200 to 400 bp deletion steps were then calculated.

Each of the DNA lots was then subjected to a large (30μg) Bal31 digest (section 2.29) under the same conditions, and sampled at the determined time intervals as described above. Samples from each of these time intervals was digested with a second enzyme to release the insert DNA (HindIII or EcoRI). These fragments were ligated into M13 prepared vector (section 2.21), transformed into JM101 or JM109 (section 2.31), plated out, and incubated overnight at 37°C. The plates containing the recombinant plaques were replica-plated onto nitrocellulose filters, dried under vacuum for 2h at 80°C, and hybridised overnight to the [^{32}P]-labelled 5.2 kb EcoRI fragment as described in section 2.32. Positive plaques were identified, and SS template DNA prepared as described in section 2.33. These templates were then sequenced using the chain termination method described in sections 2.34 to 2.38.

Several Tn5:pKH107 derivatives from both ß-galactosidase loci, as well as some from the borders of each locus, were also cloned into M13 for sequencing. These derivatives were digested with EcoRI/HindIII, cloned into M13mp18, and treated as described above. When sequenced, the Tn5 primer was used which extends from 15 to 31 bases from the ends of Tn5 (see section 2.37). This enabled sequencing from the Tn5-DNA junction into the insert DNA. The Tn5 inserts were sequenced in only one direction.

7.2.3 Sequence results

M13 SS templates containing DNA homologous to the 5.2 kb EcoRI fragment were sequenced by the dideoxy sequencing approach as described in section 2.37. The restriction map and sequencing strategy for the 5.2 kb EcoRI fragment are shown in Fig. 7.3. Both strands were totally sequenced in both directions from the overlapping clones. The cloned Tn5:pKH107 derivatives were also sequenced. Preliminary sequence data obtained for the 5.2 kb EcoRI fragment was assembled using the computer program GEL ASSEMBLE (UWGCG; Staden, 1980). The resultant complete DNA sequence showing the main restriction enzyme sites is presented in Fig. 7.4.

In order to help identify potential coding regions, the sequence data was analyzed using FRAMES (Devereux et al., 1984) and TESTCODE (Fickett, 1982). Because of the high dA+dT content (and hence, low dC+dG bias) otherwise useful programmes available on UWGCG which utilized the high dC+dG bias (such as CODON PREFERENCE and THIRD POSITION GC BIAS) were uninformative. From these results and those presented below (i.e. presence of Shine-Dalgarno sequences and positions of Tn5 insertions) three separate ORF's (ORF1, ORF2, ORF3) are predicted to be coding (Fig. 7.5).

Fig. 7.4: The complete DNA sequence of the 5.2 kb EcoRI fragment from C. acetobutylicum. The various restriction enzymes used for subcloning are as indicated.

Note: This is in reverse orientation from that used in all other chapters.

EcoRI

```
    1  GAATTCCTTTTCATATATATCTTTAATATTTCTACTGGAATAGAAGAGGTTGCTCAATAC   60
       CTTAAGGAAAAGTATATATAGAAATTATAAAGATGACCTTATCTTCTCCAACGAGTTATG

   61  AAAAAATGCTTCTTTAAAACTATTTGAAACTACTTCTGAAATATTTTCTAGCTTACTAAA  120
       TTTTTTACGAAGAAATTTTGATAAACTTTGATGAAGACTTTATAAAAGATCGAATGATTT

  121  TAGAGAATTATAATTTTTATCATCAAAATTTAGAATTACAACTATGATTTCGTTTTCAAT  180
       ATCTCTTAATATTAAAAATAGTAGTTTTAAATCTTAATGTTGATACTAAAGCAAAAGTTA

  181  ATTAGCAATTTGTATATTATAATTGCTATTTAATCCGTCTAAAGAAAATTCTTTGCCGAT  240
       TAATCGTTAAACATATAATATTAACGATAAATTAGGCAGATTTCTTTTAAGAAACGGCTA

  241  TTCTGAAATTGTAAAATCAATAATTTCATGGCGTTTGCTATAATTATCATATATTTCTTT  300
       AAGACTTTAACATTTTAGTTATTAAAGTACCGCAAACGATATTAATAGTATATAAAGAAA

  301  GCGTTTAAACCAAATAAGCAAAATGATTGAAAAGTAAATATGTATCAAAGTAGTTAAAGT  360
       CGCAAATTTGGTTTATTCGTTTTACTAACTTTTCATTTATACATAGTTTCATCAATTTCA

  361  CAGGATCATGTCAAAACCTGATATAAGGCGATTTAAGGCGCTATTAGTGAGACTTAAAGA  420
       GTCCTAGTACAGTTTTGGACTATATTCCGCTAAATTCCGCGATAATCACTCTGAATTTCT

  421  GTTTCCTTCTAAAGTATTTCTTTTCATTTTTATTGAAATCTTTTTTAGAGTACTTAATAA  480
       CAAAGGAAGATTTCATAAAGAAAAGTAAAAATAACTTTAGAAAAAATCTCATGAATTATT

  481  CTCAGAAGGATTTAGAGAAGGTTTTAAAATATAATCAACAGCACCATTTTGAAAAGATGA  540
       GAGTCTTCCTAAATCTCTTCCAAAATTTTATATTAGTTGTCGTGGTAAAACTTTTCTACT

  541  TTTAACATATTCAAAATCGCTATAACTACTTAAGATGATAATTCTTATCTTAGGATATTT  600
       AAATTGTATAAGTTTTAGCGATATTGATGAATTCTACTATTAAGAATAGAATCCTATAAA

  601  GTCCTGCACAAATTTAGCTAATTCAACCCCATTTATTTGGGGCATTACAACATCAGAAAT  660
       CAGGACGTGTTTAAATCGATTAAGTTGGGGTAAATAAACCCCGTAATGTTGTAGTCTTTA

  661  TATAATGTCAGGAATATCCTTTTTTATCATTTCCAGAGCTTCTTGACCATTAGAAGCCTG  720
       ATATTACAGTCCTTATAGGAAAAAATAGTAAAGGTCTCGAAGAACTGGTAATCTTCGGAC

  721  TCCTATAATTTGAAAGCCTTCTTTTTCCCAATCAATCATATGAGTTATGCCTTGCCGCAT  780
       AGGATATTAAACTTTCGGAAGAAAAAGGGTTAGTTAGTATACTCAATACGGAACGGCGTA
```

```
781   AATAAATTCATCATCAACAACTAAATTTTTACTATATTCGTTCAATAGTATAGCACCCCT   840
      TTATTTAAGTAGTAGTTGTTGATTTAAAAATGATATAAGCAAGTTATCATATCGTGGGGA

841   TATTCTAAAATTACCACAACATAGATAAATATTGCTTAATACTATTATACCTTATAGATT   900
      ATAAGATTTTAATGGTGTTGTATCTATTTATAACGAATTATGATAATATGGAATATCTAA

901   TATTGTATGTATCTGTATACGTTACGTTAATTCATCTACAAATTTATATGAGTTTTGGTT   960
      ATAACATACATAGACATATGCAATGCAATTAAGTAGATGTTTAAATATACTCAAAACCAA

961   GCACTTTTAGAGAAAATCTTTTTGTCTATGGTCTTATTGTCCTATAATGGTCAAATCATC   1020
      CGTGAAAATCTCTTTTAGAAAAACAGATACCAGAATAACAGGATATTACCAGTTTAGTAG

1021  TTTACCAAAGTCTCTTGATTTAAAGAGATAAAAACACCACTGATCCATTATTCCTCATTT   1080
      AAATGGTTTCAGAGAACTAAATTTCTCTATTTTTGTGGTGACTAGGTAATAAGGAGTAAA

1081  TGGTAATGAACCTATGCGGTTGAAGATATTAATCAGATGTCTAAATACTTTAGAAAAAAA   1140
      ACCATTACTTGGATACGCCAACTTCTATAATTAGTCTACAGATTTATGAAATCTTTTTTT

1141  GACCTTTACTAATATCTTCAATATTTACACCCCTATTCTAAAATTACCACAAGATAGATA   1200
      CTGGAAATGATTATAGAAGTTATAAATGTGGGGATAAGATTTTAATGGTGTTCTATCTAT

1201  AATATTGCTTAATACTGATTATACCTTATAGATTAAAGGTTTTCAATTAAACAATAAATT   1260
      TTATAACGAATTATGACTAATATGGAATATCTAATTTCCAAAAGTTAATTTGTTATTTAA

1261  ACTTTAGTAAAGTTTAGTAAAATATAATTGATTTTTTACTAAAAAAGATAATAAAATGAAA   1320
      TGAAATCATTTCAAATCATTTTATATTAACTAAAAAAATGATTTTTCTATTATTTTACTTT

1321  CTATAAATTTAGTTAATAGCATAAATCTAACATCAGAAGATAGGATAAATTAAAGAAGTA   1380
      GATATTTAAATCAATTATCGTATTTAGATTGTAGTCTTCTATCCTATTTAATTTCTTCAT

1381  ATGTAATTGATTACGAAACAAAATCTCATATTAATATTAGCCCATAATTTTTTTATTCTC   1440
      TACATTAACTAATGCTTTGTTTTAGAGTATAATTATAATCGGGTATTAAAAAAATAAGAG

1441  ATATATGTTTAAGTATTAATTAAATGTGACTTTATAAAAAGGTTGCATTTAGTTAATACG   1500
      TATATACAAATTCATAATTAATTTACACTGAAATATTTTTCCAACGTAAATCAATTATGC

1501  ATTAACAACTTTAATTTAAAAAAGCAATAACTCTACAAAGTGAAAGTGAGGGGGTAAGTA   1560
      TAATTGTTGAAATTAAATTTTTTCGTTATTGAGATGTTTCACTTTCACTCCCCCATTCAT
```

```
1561  TGATTAATAATAAACCGTCATTAGATTGGCTAGAAAATCCGGAAATATTTAGAGTTAATA  1620
      ACTAATTATTATTTGGCAGTAATCTAACCGATCTTTTAGGCCTTTATAAATCTCAATTAT

1621  GAATAGATGCTCATTCTGATACTTGGTTTTATGAAAAATTTGAGGATGTTAAATTAGAAG  1680
      CTTATCTACGAGTAAGACTATGAACCAAAATACTTTTTAAACTCCTACAATTTAATCTTC

1681  ACACCATGCCTCTTAAGCAAAATTTAAATGGAAAATGGAGATTTTCATATAGTGAAAATT  1740
      TGTGGTACGGAGAATTCGTTTTAAATTTACCTTTTACCTCTAAAAGTATATCACTTTTAA

1741  CATCATTAAGAATTAAAGAGTTTTATAAGGATGAGTTTGACGTAAGTTGGATTGATTATA  1800
      GTAGTAATTCTTAATTTCTCAAAATATTCCTACTCAAACTGCATTCAACCTAACTAATAT

1801  TTGAAGTTCCAGGTCATATTCAGCTTCAAGGATATGATAAATGTCAATATATTAATACTA  1860
      AACTTCAAGGTCCAGTATAAGTCGAAGTTCCTATACTATTTACAGTTATATAATTATGAT

1861  TGTATCCTTGGGAAGGTCACGATGAATTAAGACCACCTCATATTTCAAAAACATATAATC  1920
      ACATAGGAACCCTTCCAGTGCTACTTAATTCTGGTGGAGTATAAAGTTTTTGTATATTAG

1921  CGGTGGGAAGCTATGTAACATTTTTTGAAGTTAAAGATGAACTCAAAAATAAGCAGACTT  1980
      GCCACCCTTCGATACATTGTAAAAAACTTCAATTTCTACTTGAGTTTTTATTCGTCTGAA

1981  TTATTTCTTTTCAAGGTGTTGAAACAGCATTTTACGTATGGGTAAATGGAGAATTTGTAG  2040
      AATAAAGAAAAGTTCCACAACTTTGTCGTAAAATGCATACCCATTTACCTCTTAAACATC

2041  GATATAGCGAAGATACATTTACACCATCAGAATTTGATATTACTGATTATTTAAGAGAGG  2100
      CTATATCGCTTCTATGTAAATGTGGTAGTCTTAAACTATAATGACTAATAAATTCTCTCC

2101  GAGAAAATAAACTTGCAGTTGAGGTTTATAAAAGGAGTAGCGCAAGTTGGATAGAAGATC  2160
      CTCTTTTATTTGAACGTCAACTCCAAATATTTTCCTCATCGCGTTCAACCTATCTTCTAG

2161  AAGATTTCTGGAGATTTTCAGGCATCTTTAGAGATGTATATTTATATGCAGTTCCAGAAA  2220
      TTCTAAAGACCTCTAAAAGTCCGTAGAAATCTCTACATATAAATATACGTCAAGGTCTTT

2221  CTCATGTAAATGATATATTTATAAAAACAGATTTATATGACGATTTCAAAAACGCAAAGT  2280
      GAGTACATTTACTATATAAATATTTTTGTCTAAATATACTGCTAAAGTTTTTGCGTTTCA

2281  TAAATGCTGAACTTAAAATGATTGGAAATTCAGAAACAACAGTTGAAACATATTTAGAAG  2340
      ATTTACGACTTGAATTTTACTAACCTTTAAGTCTTTGTTGTCAACTTTGTATAAATCTTC
```

```
2341  ATAAAGAAGGAAATAAAATAGCTATATCTGAAAAGATTCCGTTCTCTGATGAGTTGACTT  2400
      TATTTCTTCCTTTATTTTATCGATATAGACTTTTCTAAGGCAAGAGACTACTCAACTGAA

2401  TATATTTAGATGCGCAAAATATAAACCTATGGAGTGCAGAAGAGCCTAACTTATATACAC  2460
      ATATAAATCTACGCGTTTTATATTTGGATACCTCACGTCTTCTCGGATTGAATATATGTG

2461  TTTATATTTTAGTGAATAAAAAAGATGGTAATTTAATTGAGGTTGTAACTCAAAAGATAG  2520
      AAATATAAAATCACTTATTTTTTCTACCATTAAATTAACTCCAACATTGAGTTTTCTATC

2521  GGTTTAGGCACTTTGAAATGAAGGATAAAATTATGTGTCTAAAATGGAAACGTATTATCT  2580
      CCAAATCCGTGAAACTTTACTTCCTATTTTAATACACAGATTTTACCTTTGCATAATAGA

2581  TTAAAGGCGTAAACCGTCACGAATTTAGCGCAAGACGTGGACGCTCAATTACGAAAGAGG  2640
      AATTTCCGCATTTGGCAGTGCTTAAATCGCGTTCTGCACCTGCGAGTTAATGCTTTCTCC

2641  ACATGTTGTGGGATATTAAGTTCTTGAAACAACACAATATTAATGCTGTTAGAACATCAC  2700
      TGTACAACACCCTATAATTCAAGAACTTTGTTGTGTTATAATTACGACAATCTTGTAGTG

2701  ATTATCCAAATCAAAGTTTATGGTACAGACTTTGCGATGAATACGGGATTTATTTAATAG  2760
      TAATAGGTTTAGTTTCAAATACCATGTCTGAAACGCTACTTATGCCCTAAATAAATTATC

2761  ATGAAACAAATTTAGAAAGCCATGGTTCATGGCAAAAGATGGGGCAGATTGAACCATCAT  2820
      TACTTTGTTTAAATCTTTCGGTACCAAGTACCGTTTTCTACCCCGTCTAACTTGGTAGTA

2821  GGAATGTGCCAGGAAGTCTTCCACAGTGGCAGGCAGCAGTTTTAGATCGAGCATCATCAA  2880
      CCTTACACGGTCCTTCAGAAGGTGTCACCGTCCGTCGTCAAAATCTAGCTCGTAGTAGTT

2881  TGGTTGAAAGAGATAAAAATCATCCATCTGTACTTATTTGGTCATGTGGTAATGAATCCT  2940
      ACCAACTTTCTCTATTTTTAGTAGGTAGACATGAATAAACCAGTACACCATTACTTAGGA

2941  ATGCGGGTGAAGATATTTATCAGATGTCTAAATACTTTAGAAAAAAAGATCCTTCACGTT  3000
      TACGCCCACTTCTATAAATAGTCTACAGATTTATGAAATCTTTTTTTCTAGGAAGTGCAA

3001  TAGTGCACTATGAAGGGGTAACTAGATGCAGAGAATTTATGACACGACGACATGAAAGTA  3060
      ATCACGTGATACTTCCCCATTGATCTACGTCTCTTAAATACTGTGCTGCTGTACTTTCAT

3061  GAATGTATGCAAAGGCAGCAGAAATAGAAGAATATCTTAATGATAATCCGAAGAAACCTT  3120
      CTTACATACGTTTCCGTCGTCTTTATCTTCTTATAGAATTACTATTAGGCTTCTTTGGAA
```

```
                .         .         .         .         .         .
3121  ATATACAGCTGCGATACATGCACTCAATGGGTAACTCAACTGGTGGAATGATGAAATACA  3180
      TATATGTCGACGCTATGTACGTGAGTTACCCATTGAGTTGACCACCTTACTACTTTATGT
           PvuII
                .         .         .         .         .         .
3181  CAGAACTTGAAGATAAATATTTGATGTATCAAGGTGGATTCATTTGGGATTACGGCGATC  3240
      GTCTTGAACTTCTATTTATAAACTACATAGTTCCACCTAAGTAAACCCTAATGCCGCTAG

                .         .         .         .         .         .
3241  AGGCGTTGTATAGAAAACTTCCAGATGGAAAAGAAGTTCTAGCTTATGGAGGAGACTTTA  3300
      TCCGCAACATATCTTTTGAAGGTCTACCTTTTCTTCAAGATCGAATACCTCCTCTGAAAT

                .         .         .         .         .         .
3301  CAGATCGTCCAACAGACTATAATTTCTCTGGAAATGGTTTGATTTATGCAGATAGAACTA  3360
      GTCTAGCAGGTTGTCTGATATTAAAGAGACCTTTACCAAACTAAATACGTCTATCTTGAT

                .         .         .         .         .         .
3361  TATCACCTAAAGCACAGGAAGTTAAGTATCTATATCAAAACGTAAAATTAGAACCAGATG  3420
      ATAGTGGATTTCGTGTCCTTCAATTCATAGATATAGTTTTGCATTTTAATCTTGGTCTAC

                .         .         .         .         .         .
3421  AAAAAGGGGTGACTATTAAGAATCAAAATCTTTTTGTTAATACTGATAAATATGATTTAT  3480
      TTTTTCCCCACTGATAATTCTTAGTTTTAGAAAAACAATTATGACTATTTATACTAAATA

                .         .         .         .         .         .
3481  ACTATATCGTTGAAAGAGATGGAAAACTAATAAAAGATGGTTATCTAAATGTATCTGTAG  3540
      TGATATAGCAACTTTCTCTACCTTTTGATTATTTTCTACCAATAGATTTACATAGACATC

                .         .         .         .         .         .
3541  CTCCAGATGAAGAAAAATATATAGAACTTCCAATAGGAAATTACAATTTTCCTGAAGAAA  3600
      GAGGTCTACTTCTTTTTATATATCTTGAAGGTTATCCTTTAATGTTAAAAGGACTTCTTT

                .         .         .         .         .         .
3601  TTGTACTTACAACCTCATTAAGATTAGCACAAGCTACACTTTGGGCAGAAAAAGGATATG  3660
      AACATGAATGTTGGAGTAATTCTAATCGTGTTCGATGTGAAACCCGTCTTTTTCCTATAC

                .         .         .         .         .         .
3661  AAATAGCATTTGGACAAAAGGTTATTAAAGAAAAATCAGATATGAATAATCATAATTCAG  3720
      TTTATCGTAAACCTGTTTTCCAATAATTTCTTTTTAGTCTATACTTATTAGTATTAAGTC

                .         .         .         .         .         .
3721  AGTCTAAAATGAAGATCATTCATGGAGATGTAAACATAGGGGTTCACGGAAAAGATTTCA  3780
      TCAGATTTTACTTCTAGTAAGTACCTCTACATTTGTATCCCCAAGTGCCTTTTCTAAAGT

                .         .         .         .         .         .
3781  AGGCTATATTCTCTAAACAAGAGGGAGGAATCGTATCCTTGAGATATAATAATAAGGAGT  3840
      TCCGATATAAGAGATTTGTTCTCCCTCCTTAGCATAGGAACTCTATATTATTATTCCTCA

                .         .         .         .         .         .
3841  TTATAACGAGAACGCCAAAAACTTTCTATTGGAGAGCAACAACAGATAATGATAGAGGAA  3900
      AATATTGCTCTTGCGGTTTTTGAAAGATAACCTCTCGTTGTTGTCTATTACTATCTCCTT
```

```
                .         .         .         .         .         .
3901  ATAGACATGAATTTAGATGCAGTCAATGGCTGGCTGCTACTATGGGGCAGAAGTATGTGG  3960
      TATCTGTACTTAAATCTACGTCAGTTACCGACCGACGATGATACCCCGTCTTCATACACC

                .         .         .         .         .         .
3961  ATTTTTCAGTTGAGGAATTTGATGAGAAGATTACATTATATTATACTTATCAATTGCCAA  4020
      TAAAAAGTCAACTCCTTAAACTACTCTTCTAATGTAATATAATATGAATAGTTAACGGTT

                .         .         .         .         .         .
4021  CAGTGCCATCTACTAATGTTAAGATAACTTATGAAGTATCTGGAGAAGGAATAATTAAAG  4080
      GTCACGGTAGATGATTACAATTCTATTGAATACTTCATAGACCTCTTCCTTATTAATTTC

                .         .         .         .         .         .
4081  TAAATGTTAAGTATAAAGGAGTTAGCGGATTACCTGAATTGCCTGTACTAGGAATGGATT  4140
      ATTTACAATTCATATTTCCTCAATCGCCTAATGGACTTAACGGACATGATCCTTACCTAA

                .         .         .         .         .         .
4141  TTAAATTATTAGCCGAATTTAATTCATTTAGCTGGTATGGAATGGGGCCAGAAGAAAACT  4200
      AATTTAATAATCGGCTTAAATTAAGTAAATCGACCATACCTTACCCCGGTCTTCTTTTGA
                                                  HaeIII

                .         .         .         .         .         .
4201  ATATAGACAGATGTGAAGGTGCAAAACTTGGAATATATGAGAGTACACAATAGAAAATCT  4260
      TATATCTGTCTACACTTCCACGTTTTGAACCTTATATACTCTCATGTGTTATCTTTTAGA

                .         .         .         .         .         .
4261  ATCAAGGTATTTAGTACCACAAGAATGTGGTAACAGGATAGGAACTAGATGGGTAGTAGT  4320
      TAGTTCCATAAATCATGGTGTTCTTACACCATTGTCCTATCCTTGATCTACCCATCATCA

                .         .         .         .         .         .
4321  TAAAAATCATAAGAATGAAGGTCTTAAATTTACTTATGTTAAAGTTCCATTTGAATTTAG  4380
      ATTTTTAGTATTCTTACTTCCAGAATTTAAATGAATACAATTTCAAGGTAAACTTAAATC

                .         .         .         .         .         .
4381  TGTTTTACCATACAGCAGCATGGAATTAGAAAATTCACTTCATATAGAAGAATTACCATC  4440
      ACAAAATGGTATGTCGTCGTACCTTAATCTTTTAAGTGAAGTATATCTTCTTAATGGTAG

                .         .         .         .         .         .
4441  TGTTAATTTTACACATTGTGAATATAATAGGTAAACAAATGGGTGTTGGCGGAGATGCAA  4500
      ACAATTAAAATGTGTAACACTTATATTATCCATTTGTTTACCCACAACCGCCTCTACGTT

                .         .         .         .         .         .
4501  TGCTGGGGAGCACCATGATACCTAAATTCTGTATAGATTCAAGTAAGGATTTAGAATATA  4560
      ACGACCCCTCGTGGTACTATGGATTTAAGACATATCTAAGTTCATTCCTAAATCTTATAT

                .         .         .         .         .         .
4561  GTTTTATAATTTCTAAAATTATACTACGCACATATGGGAACTATAGATATCCAAAACAAA  4620
      CAAAATATTAAAGATTTTAATATGATGCGTGTATACCCTTGATATCTATAGGTTTTGTTT

                .         .         .         .         .         .
4621  ACTTAGACTTATGCAATAATTTACGAAAGGACAGGTACTCTGTTGTTTCGGTTACTAAGA  4680
      TGAATCTGAATACGTTATTAAATGCTTTCCTGTCCATGAGACAACAAAGCCAATGATTCT
```

```
                .         .         .         .         .         .
4681  ATAAGTTGAGGCTTTCTAACATCATAAGTTGCACCATTTCAGCATGCTCCCGAGACAAGC   4740
      TATTCAACTCCGAAAGATTGTAGTATTCAACGTGGTAAAGTCGTACGAGGGCTCTGTTCG

                .         .         .         .         .         .
4741  TCGTGACAAGCAAAAATGGAACAACTTATGATGAAGAAATGCCTGCAACATATTCTTTAA   4800
      AGCACTGTTCGTTTTTACCTTGTTGAATACTACTTCTTTACGGACGTTGTATAAGAAATT

                .         .         .         .         .         .
4801  TGTAACACTGCACAAAAGAGTACCTGTCCTTTCTGATATAGCAGATTTTTCAAGCTATAA   4860
      ACATTGTGACGTGTTTTCTCATGGACAGGAAAGACTATATCGTCTAAAAAGTTCGATATT

                .         .         .         .         .         .
4861  GTATATCTCACGAAATCATAAATATTTTGATTCCGAAAAGCTATGAAAATATCGCTGAAG   4920
      CATATAGAGTGCTTTAGTATTTATAAAACTAAGGCTTTTCGATACTTTTATAGCGACTTC

                .         .         .         .         .         .
4921  GTTCTAAGCAGCTGGTTGTGTGCACCTTAGCATGCTCCAACTTTCAGTTTGACAAGCTAA   4980
      CAAGATTCGTCGACCAACACACGTGGAATCGTACGAGGTTGAAAGTCAAACTGTTCGATT
              PvuII
                .         .         .         .         .         .
4981  AATGGAACAATCTACAGCTCAAGAAACTTTAACAGCTCATTTTCAAATGTTTTCTACACA   5040
      TTACCTTGTTAGATGTCGAGTTCTTTGAAATTGTCGAGTAAAAGTTTACAAAAGATGTGT

                .         .         .         .         .         .
5041  AATATATTTATATTTCTAGTGAAGATATGAAATTAAATTTTTAGCGACTTTGTAAATATG   5100
      TTATATAAATATAAAGATCACTTCTATACTTTAATTTAAAAATCGCTGAAACATTTATAC

                .         .
5101  TTAATCTAATATACGAATTC  5120
      AATTAGATTATATGCTTAAG
                    EcoRI
```

Fig. 7.5: Analysis of the sequence presented in Fig. 7.4, showing the inferred ORF's for the six possible frames from the sequence of the 5.2 kb EcoRI fragment (FRAMES: Devereux et al., 1984).

The bars represent the possible ORF's. The three ORF's that potentially code (labelled ORF1, ORF2, ORF3) were chosen of the basis of the analysis presented here as well as by an analysis using TESTCODE (Fickett, 1982) together with other information in this chapter.

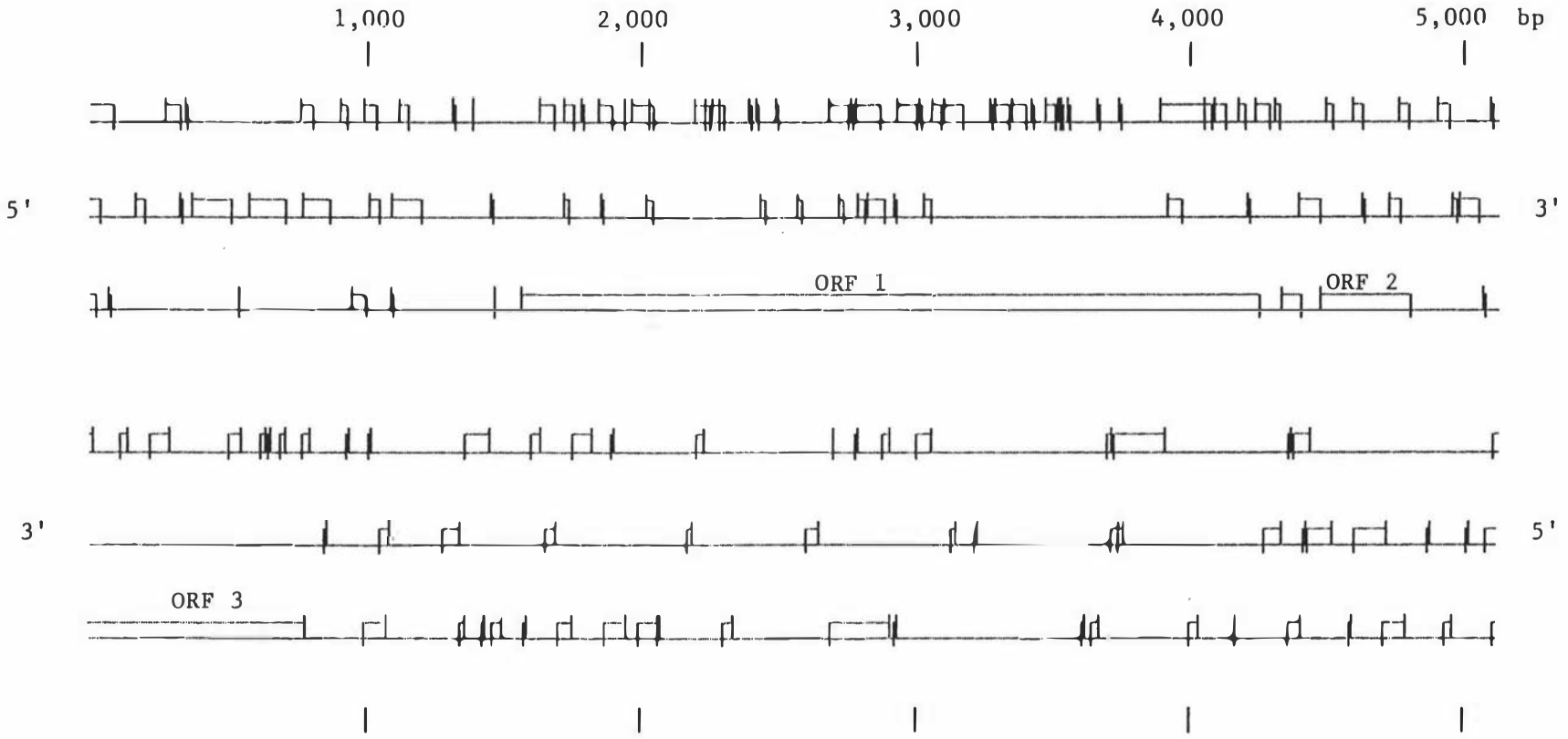

1,000
2,000
3,000
4,000
5,000
bp
5'
3'
ORF 1
ORF 2
3'
5'
ORF 3

Table 7.2

Location of the ORF's on the 5.2 kb EcoRI fragment, putative promoter sequences plus potential translation initiation signals.

ORF	Gene	Location	-35	-10
I	**Promoter sequences**			
C. pasteurianum consensus			TTGACA--16/19 bp--TATAAT[a]	
E. coli consensus			TTGACA--17 bp	--TATATT[b]
1	cbgA	1559-4253	(-173)TTGATT-----19-----TAATAT(-143)	
			(-58)TTAACA-----19-----TAACTCT(-26)	
2	cbgR	4500-4805	(-81)TTCATA-----18-----TTAATT(-52)	
			(-59)TGTTAA-----10-----TATAAT(-32)	
3	ORF3	796-1	(-54)TAAGCA-----19-----TAATTT(-23)	
			(-87)GATACA-----18-----TATAAT(-59)	
II	**Translation initiation signals**			
3' end of B. subtilis 16S rRNA			3'-OH.UCUUUCCUCCACUAG-5'[c]	
3' end of E. coli 16S rRNA			3'-OH.AUUCCUCC-5'[d]	
1	cbgA ORF1		GAAAGT**GAGGGG**GTAAGTATG	
2	cbgR ORF2		TT**GGCGGA**GATGCAATG	
			ATGCT**GGGGAG**CACCATG	
3	ORF3		AGAATA**GGGGTG**CTAACTATTG	

[a] Graves and Rabinowitz (1986)

[b] Hawley and McClure (1983)

[c] Moran et al. (1982)

[d] Shine and Dalgarno (1974)

The main ORF identified, ORF1, is assumed to be the structural gene for the Clostridium ß-galactosidase (designated cbgA: clostridial ß-galactosidase) gene. The sequence of this ORF together with the 5' upstream sequences and the deduced amino acid sequence is shown in Fig. 7.6. Sequencing of the Tn5 derivatives in Locus II which inactivated the synthesis of ß-galactosidase mapped into this large ORF1 with no other ORF of significant length present in this region. The position of ORF1 is also consistent with the Tn5 insertional inactivation data presented in Chapter 6. This ORF consisted of a coding sequence of 2691 bp, starting with an ATG codon at position 1560 and ending with a TAG (stop) codon at position 4253 (Fig. 7.5).

The nucleotide sequences upstream of the 5' end of ORF1 were screened for sequences complementary to the 3' end of the 16S rRNA of B. subtilis and E. coli (Shine and Dalgarno, 1974; Moran et al., 1982; Stormo et al., 1982). A strong Shine-Dalgarno (SD) sequence (GAGGGG) containing 5 C/G's was located at position -12 to -6, 5 bp upstream of the ATG initiation codon (Table 7.2). The region upstream of this presumptive translational start site was analyzed for potential promoter sequences homologous to E. coli and C. pasteurianum consensus sequences (Hawley and McClure, 1983; Graves and Rabinowitz, 1986) using BESTFIT (Smith and Waterman, 1981) and GAP (Needleman and Wunsch, 1970). The two potential Clostridium promoter sequences identified are shown in Table 7.2. Several inverted repeat sequences that have the potential to form stem-loop structures were identified downstream of the stop codon. The nucleotide sequence of the ORF contained 67% dA+dT. The resulting DNA sequence of the ORF was analyzed using CODONFREQUENCY (Devereux et al., 1984) revealing a strong bias towards codons containing dA+dT (Discussion; Table 7.4, 7.5).

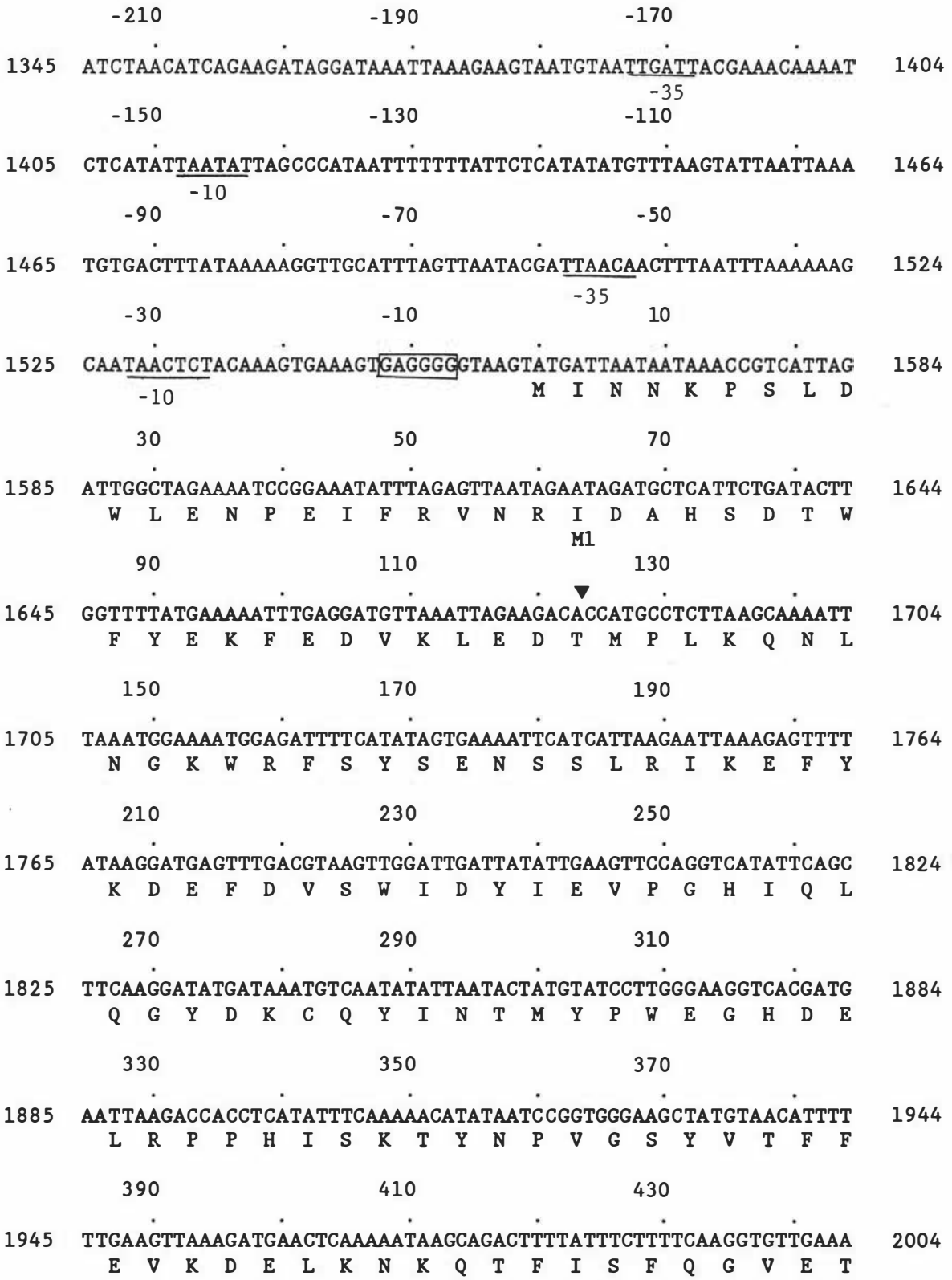
The ORF was sufficient in size to code for an unmodified protein of 897 amino acid residues with a calculated Mr of 105 kDa (Fig. 7.6). This predicted molecular weight of the translated protein is in agreement with the size (100

Fig. 7.6: The complete DNA sequence and deduced amino acid sequence of the region of the 5.2 kb EcoRI fragment containing the ß-galactosidase structural gene (cbgA). The -10 and -35 promoter regions are underlined. The Shine-Dalgarno (SD) sequence is boxed.

The positions of a number of Tn5 inserts shown by ▼ (ß-galactosidase+) and ▽(ß-galactosidase-) are as indicated. The number given in brackets above the symbol corresponds to the number assigned to the Tn5 insertion in Chapter 6.

Possible stem-loop structures in the terminal region are indicated by arrows. The amino acids are identified by the following single letter code.

A Ala
C Cys
D Asp
E Glu
F Phe
G Gly
H His
I Ile
K Lys
L Leu
M Met
N Asn
P Pro
Q Gln
R Arg
S Ser
T Thr
V Val
W Trp
Y Tyr


-210
-190
-170
1345 ATCTAACATCAGAAGATAGGATAAATTAAAGAAGTAATGTAATTGATTACGAAACAAAAT 1404
-35
-150
-130
-110
1405 CTCATATTAATATTAGCCCATAATTTTTTTATTCTCATATATGTTTAAGTATTAATTAAA 1464
-10
-90
-70
-50
1465 TGTGACTTTATAAAAAGGTTGCATTTAGTTAATACGATTAACAACTTTAATTTAAAAAAG 1524
-35
-30
-10
10
1525 CAATAACTCTACAAAGTGAAAGTGAGGGGGTAAGTATGATTAATAATAAACCGTCATTAG 1584
-10
M I N N K P S L D
30
50
70
1585 ATTGGCTAGAAAATCCGGAAATATTTAGAGTTAATAGAATAGATGCTCATTCTGATACTT 1644
W L E N P E I F R V N R I D A H S D T W
M1
90
110
130
1645 GGTTTTATGAAAAATTTGAGGATGTTAAATTAGAAGACACCATGCCTCTTAAGCAAAATT 1704
F Y E K F E D V K L E D T M P L K Q N L
150
170
190
1705 TAAATGGAAAATGGAGATTTTCATATAGTGAAAATTCATCATTAAGAATTAAAGAGTTTT 1764
N G K W R F S Y S E N S S L R I K E F Y
210
230
250
1765 ATAAGGATGAGTTTGACGTAAGTTGGATTGATTATATTGAAGTTCCAGGTCATATTCAGC 1824
K D E F D V S W I D Y I E V P G H I Q L
270
290
310
1825 TTCAAGGATATGATAAATGTCAATATATTAATACTATGTATCCTTGGGAAGGTCACGATG 1884
Q G Y D K C Q Y I N T M Y P W E G H D E
330
350
370
1885 AATTAAGACCACCTCATATTTCAAAAACATATAATCCGGTGGGAAGCTATGTAACATTTT 1944
L R P P H I S K T Y N P V G S Y V T F F
390
410
430
1945 TTGAAGTTAAAGATGAACTCAAAAATAAGCAGACTTTTATTTCTTTTCAAGGTGTTGAAA 2004
E V K D E L K N K Q T F I S F Q G V E T

```
          450                 470                 490
           .         .         .         .         .         .
2005  CAGCATTTTACGTATGGGTAAATGGAGAATTTGTAGGATATAGCGAAGATACATTTACAC  2064
        A  F  Y  V  W  V  N  G  E  F  V  G  Y  S  E  D  T  F  T  P

          510                 530                 550
           .         .         .         .         .         .
2065  CATCAGAATTTGATATTACTGATTATTTAAGAGAGGGAGAAAATAAACTTGCAGTTGAGG  2124
        S  E  F  D  I  T  D  Y  L  R  E  G  E  N  K  L  A  V  E  V

          570                 590                 610
           .         .         .         .         .         .
2125  TTTATAAAAGGAGTAGCGCAAGTTGGATAGAAGATCAAGATTTCTGGAGATTTTCAGGCA  2184
        Y  K  R  S  S  A  S  W  I  E  D  Q  D  F  W  R  F  S  G  I

          630                 650                 670
           .         .         .         .         .         .
2185  TCTTTAGAGATGTATATTTATATGCAGTTCCAGAAACTCATGTAAATGATATATTTATAA  2244
        F  R  D  V  Y  L  Y  A  V  P  E  T  H  V  N  D  I  F  I  K

          690                 710                 730
           .         .         .         .         .         .
2245  AAACAGATTTATATGACGATTTCAAAAACGCAAAGTTAAATGCTGAACTTAAAATGATTG  2304
        T  D  L  Y  D  D  F  K  N  A  K  L  N  A  E  L  K  M  I  G

          750                 770                 790
           .         .         .         .         .         .
2305  GAAATTCAGAAACAACAGTTGAAACATATTTAGAAGATAAAGAAGGAAATAAAATAGCTA  2364
        N  S  E  T  T  V  E  T  Y  L  E  D  K  E  G  N  K  I  A  I

          810                 830                 850
           .         .         .         .         .         .
2365  TATCTGAAAAGATTCCGTTCTCTGATGAGTTGACTTTATATTTAGATGCGCAAAATATAA  2424
        S  E  K  I  P  F  S  D  E  L  T  L  Y  L  D  A  Q  N  I  N

          870                 890                 910
           .         .         .         .         .         .
2425  ACCTATGGAGTGCAGAAGAGCCTAACTTATATACACTTTATATTTTAGTGAATAAAAAAG  2484
        L  W  S  A  E  E  P  N  L  Y  T  L  Y  I  L  V  N  K  K  D
                                                               J5
          930                 950                 970
           .         .         .         .         .         .  ▼
2485  ATGGTAATTTAATTGAGGTTGTAACTCAAAAGATAGGGTTTAGGCACTTTGAAATGAAGG  2544
        G  N  L  I  E  V  V  T  Q  K  I  G  F  R  H  F  E  M  K  D

          990                1010                1030
           .         .         .         .         .         .
2545  ATAAAATTATGTGTCTAAAATGGAAACGTATTATCTTTAAAGGCGTAAACCGTCACGAAT  2604
        K  I  M  C  L  K  W  K  R  I  I  F  K  G  V  N  R  H  E  F
```

```
        1050                1070                1090
         .         .         .         .         .         .
2605  TTAGCGCAAGACGTGGACGCTCAATTACGAAAGAGGACATGTTGTGGGATATTAAGTTCT  2664
        S  A  R  R  G  R  S  I  T  K  E  D  M  L  W  D  I  K  F  L

        1110                1130                1150
         .         .         .         .         .         .
2665  TGAAACAACACAATATTAATGCTGTTAGAACATCACATTATCCAAATCAAAGTTTATGGT  2724
        K  Q  H  N  I  N  A  V  R  T  S  H  Y  P  N  Q  S  L  W  Y

        1170                1190                1210
         .         .         .         .         .         .
2725  ACAGACTTTGCGATGAATACGGGATTTATTTAATAGATGAAACAAATTTAGAAAGCCATG  2784
        R  L  C  D  E  Y  G  I  Y  L  I  D  E  T  N  L  E  S  H  G

        1230                1250                1270
         .         .         .         .         .         .
2785  GTTCATGGCAAAAGATGGGGCAGATTGAACCATCATGGAATGTGCCAGGAAGTCTTCCAC  2844
        S  W  Q  K  M  G  Q  I  E  P  S  W  N  V  P  G  S  L  P  Q

        1290                1310                1330
         .         .         .         .         .         .
2845  AGTGGCAGGCAGCAGTTTTAGATCGAGCATCATCAATGGTTGAAAGAGATAAAAATCATC  2904
        W  Q  A  A  V  L  D  R  A  S  S  M  V  E  R  D  K  N  H  P

        1350                1370                1390
         .         .         .         .         .         .
2905  CATCTGTACTTATTTGGTCATGTGGTAATGAATCCTATGCGGGTGAAGATATTTATCAGA  2964
        S  V  L  I  W  S  C  G  N  E  S  Y  A  G  E  D  I  Y  Q  M

        1410                1430                1450
         .         .         .         .         .         .
2965  TGTCTAAATACTTTAGAAAAAAAGATCCTTCACGTTTAGTGCACTATGAAGGGGTAACTA  3024
        S  K  Y  F  R  K  K  D  P  S  R  L  V  H  Y  E  G  V  T  R

        1470                1490                1510
         .         .         .         .         .         .
3025  GATGCAGAGAATTTATGACACGACGACATGAAAGTAGAATGTATGCAAAGGCAGCAGAAA  3084
        C  R  E  F  M  T  R  R  H  E  S  R  M  Y  A  K  A  A  E  I

        1530                1550                1570
         .         .         .         .         .         .
3085  TAGAAGAATATCTTAATGATAATCCGAAGAAACCTTATATACAGCTGCGATACATGCACT  3144
        E  E  Y  L  N  D  N  P  K  K  P  Y  I  Q  L  R  Y  M  H  S

        1590                1610                1630
         .         .         .         .         .         .
3145  CAATGGGTAACTCAACTGGTGGAATGATGAAATACACAGAACTTGAAGATAAATATTTGA  3204
        M  G  N  S  T  G  G  M  M  K  Y  T  E  L  E  D  K  Y  L  M
```

```
        1650                1670                1690
          .         .         .         .         .         .
3205  TGTATCAAGGTGGATTCATTTGGGATTACGGCGATCAGGCGTTGTATAGAAAACTTCCAG  3264
        Y  Q  G  G  F  I  W  D  Y  G  D  Q  A  L  Y  R  K  L  P  D

        1710                1730                1750
          .         .         .         .         .         .
3265  ATGGAAAAGAAGTTCTAGCTTATGGAGGAGACTTTACAGATCGTCCAACAGACTATAATT  3324
        G  K  E  V  L  A  Y  G  G  D  F  T  D  R  P  T  D  Y  N  F
                    N1
        1770                1790                1810
          .         .▼        .         .         .         .
3325  TCTCTGGAAATGGTTTGATTTATGCAGATAGAACTATATCACCTAAAGCACAGGAAGTTA  3384
        S  G  N  G  L  I  Y  A  D  R  T  I  S  P  K  A  Q  E  V  K

        1830                1850                1870
          .         .         .         .         .         .
3385  AGTATCTATATCAAAACGTAAAATTAGAACCAGATGAAAAAGGGGTGACTATTAAGAATC  3444
        Y  L  Y  Q  N  V  K  L  E  P  D  E  K  G  V  T  I  K  N  Q

        1890                1910                1930
          .         .         .         .         .         .
3445  AAAATCTTTTTGTTAATACTGATAAATATGATTTATACTATATCGTTGAAAGAGATGGAA  3504
        N  L  F  V  N  T  D  K  Y  D  L  Y  Y  I  V  E  R  D  G  K

        1950                1970                1990
          .         .         .         .         .         .
3505  AACTAATAAAAGATGGTTATCTAAATGTATCTGTAGCTCCAGATGAAGAAAAATATATAG  3564
        L  I  K  D  G  Y  L  N  V  S  V  A  P  D  E  E  K  Y  I  E

        2010                2030                2050
          .         .         .         .         .         .
3565  AACTTCCAATAGGAAATTACAATTTTCCTGAAGAAATTGTACTTACAACCTCATTAAGAT  3624
        L  P  I  G  N  Y  N  F  P  E  E  I  V  L  T  T  S  L  R  L

        2070                2090                2110
          .         .         .         .         .         .
3625  TAGCACAAGCTACACTTTGGGCAGAAAAAGGATATGAAATAGCATTTGGACAAAAGGTTA  3684
        A  Q  A  T  L  W  A  E  K  G  Y  E  I  A  F  G  Q  K  V  I

        2130                2150                2170
          .         .         .         .         .         .
3685  TTAAAGAAAAATCAGATATGAATAATCATAATTCAGAGTCTAAAATGAAGATCATTCATG  3744
        K  E  K  S  D  M  N  N  H  N  S  E  S  K  M  K  I  I  H  G

        2190                2210                2230
          .         .         .         .         .         .
3745  GAGATGTAAACATAGGGGTTCACGGAAAAGATTTCAAGGCTATATTCTCTAAACAAGAGG  3804
        D  V  N  I  G  V  H  G  K  D  F  K  A  I  F  S  K  Q  E  G
```

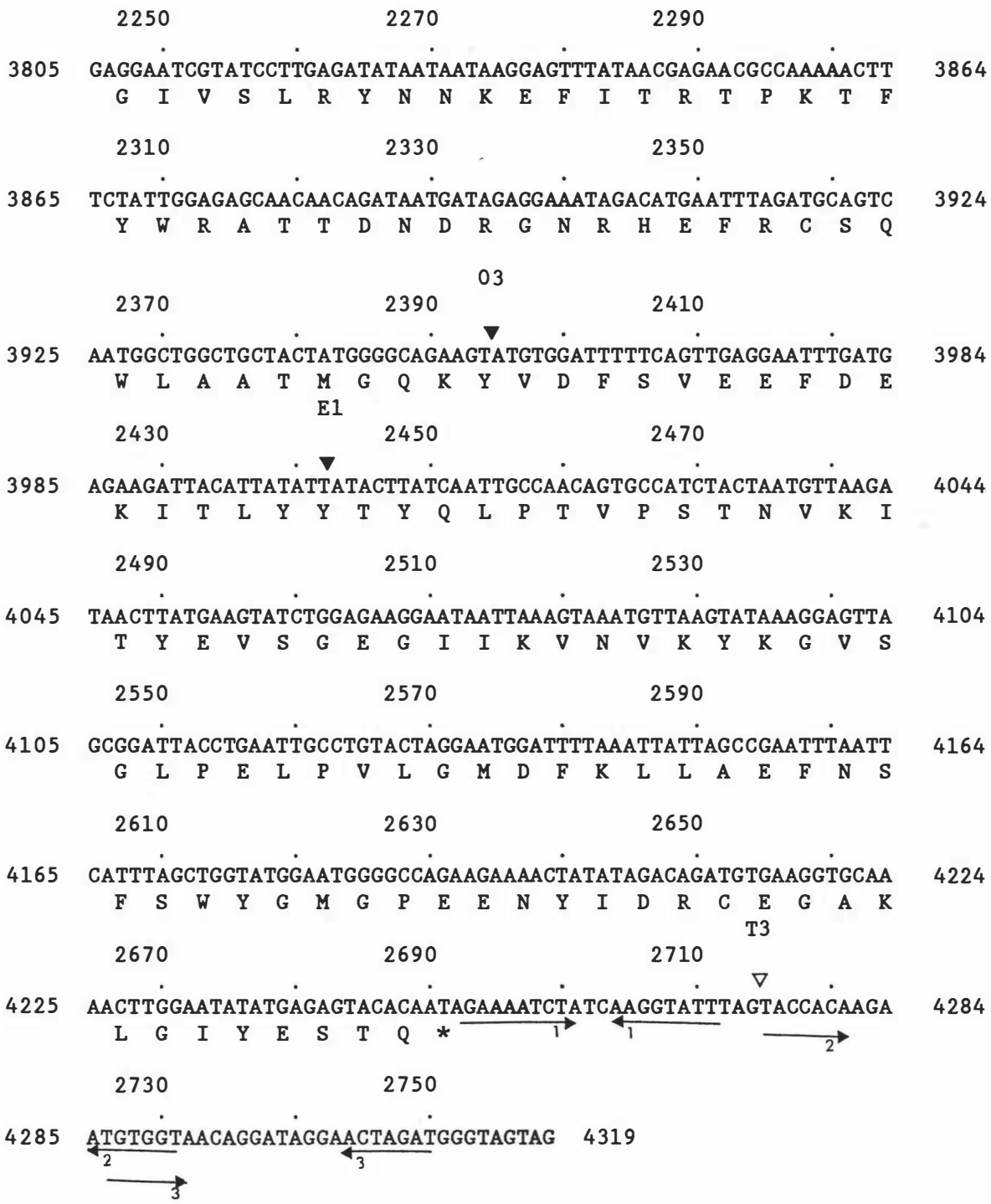

2250
2270
2290
3805 GAGGAATCGTATCCTTGAGATATAATAATAAGGAGTTTATAACGAGAACGCCAAAAACTT 3864
G I V S L R Y N N K E F I T R T P K T F
2310
2330
2350
3865 TCTATTGGAGAGCAACAACAGATAATGATAGAGGAAATAGACATGAATTTAGATGCAGTC 3924
Y W R A T T D N D R G N R H E F R C S Q
03
2370
2390
2410
3925 AATGGCTGGCTGCTACTATGGGGCAGAAGTATGTGGATTTTTCAGTTGAGGAATTTGATG 3984
W L A A T M G Q K Y V D F S V E E F D E
E1
2430
2450
2470
3985 AGAAGATTACATTATATTATACTTATCAATTGCCAACAGTGCCATCTACTAATGTTAAGA 4044
K I T L Y Y T Y Q L P T V P S T N V K I
2490
2510
2530
4045 TAACTTATGAAGTATCTGGAGAAGGAATAATTAAAGTAAATGTTAAGTATAAAGGAGTTA 4104
T Y E V S G E G I I K V N V K Y K G V S
2550
2570
2590
4105 GCGGATTACCTGAATTGCCTGTACTAGGAATGGATTTTAAATTATTAGCCGAATTTAATT 4164
G L P E L P V L G M D F K L L A E F N S
2610
2630
2650
4165 CATTTAGCTGGTATGGAATGGGGCCAGAAGAAAACTATATAGACAGATGTGAAGGTGCAA 4224
F S W Y G M G P E E N Y I D R C E G A K
T3
2670
2690
2710
4225 AACTTGGAATATATGAGAGTACACAATAGAAAATCTATCAAGGTATTTAGTACCACAAGA 4284
L G I Y E S T Q *
1
1
2
2730
2750
4285 ATGTGGTAACAGGATAGGAACTAGATGGGTAGTAG 4319
2
3
3

Fig. 7.7: **A** Alignment of the amino acid sequence of the ß-galactosidase gene of C. acetobutylicum cbgA (CacbgA) with the amino acid sequences of E. coli ebgA (EcebgA: Stokes et al., 1985); E. coli lacZ (EclacZ: Kalnins et al., 1983); K. pneumoniae lacZ (KplacZ: Buvinger and Riley, 1985); and B. stearothermophilus bgaB (BsbgaB: Hirata et al., 1986). Sequences have been aligned using the programmes WORDSEARCH and SEGMENTS (Devereux et al., 1984).

The amino acids are identified by the single letter code (see Fig. 7.6) and the regions of identical homology to one or more of the sequences are underlined.

B Alignment of the amino acid sequence of the region surrounding the E. coli LacZ active site as identified by Lerrchen and Legler (1983) with the regions of highest homology of the above sequences plus E. coli ß-glucuronidase (EcuidA: Jefferson et al., 1986). The active site is indicated by an arrow.

A

```
KplacZ   MQISDTGRSH TPDFHAVLAR EDWHNQTITH LNRLPAHPVF ASWRDELAAR 50
EclacZ   ........MI TDSLAVVLQR RDWENPGVTQ LNRLAAHPPF ASWRNSEEAR 42
EcebgA   .......... .......... .......... .......... ..........
CacbgA   ........MI NNKPSLDW.. ..LENPEIFR VNRIDAHSDT WFYEKFEDVK 38

KplacZ   DNLPSSRRRQ LDGS...GSS LTPAARLPSM RVVTQDLPDC RGTPVPSNWQ 97
EclacZ   TDRPSQQLRS LNGEWRFAWF PAPEAVPESW L..ECDLPEA DTVVVPSNWQ 90
EcebgA   .......... .......... .......... ......MADW GHITVPAMWQ 14
CacbgA   LEDTMPLKQN LNGKWRFSYS ENSSLRIKEF YKDEFDVSWI DYIEVPGHIQ 88

KplacZ   MEGYDAPIYT NVRYPIDT.. ..TPPRVPE. DNPTGCYSLH FTVEDTWREN 142
EclacZ   MHGYDAPIYT NVTYPITVN. ...PPFVP.T ENPTGCYSLT FNVDESWLQE 135
EcebgA   MEGHGKLQYT DEGFPFPI.D VPFVPSD... .NPTGAYQRI FTLSDG.WQG 58
CacbgA   LQGYDKCQYI NTMYPWEGHD ELRPPHISKT YNPVGSYVTF FEVKDE.LKN 137

BsbgaB                                              MNV LSSICYGGDY 13
KplacZ   GQTQIIFDGV NSAFHLWCNG VWVGYSQDSR LPAAFDLSPF LRPGDNRLCV 192
EclacZ   GQTRIIFDGV NSAFHLWCNG RWVGYGQDSR LPSEFDLSAF LRAGENRLAV 185
EcebgA   KQTLIKFDGV ETYFEVYVNG QYVGFSKGSR LTAEFDISAM VKTGDNLLCV 108
CacbgA   KQTFISFQGV ETAFYVWVNG EFVGYSEDTF TPSEFDITDY LREGENKLAV 187

BsbgaB   NPEQWPEEIW YEDAKLMQKA GVNLVSLGIF SWSKIEPSDG VFDFEWLDKV 63
KplacZ   MVMRWSAGSW LEDQDMWRMS GIFRSVWLL. NKPQQRLCDV QLTPALDALY 241
EclacZ   MVLRWSDGSY LEDQDMWRMS GIFRDVSLL. HKPTTQISDF HVATRFNDDF 234
EcebgA   RVMQWADSTY VEDQDMWWSA GIFRDVYLV. GKHLTHINDF TVRTDFDEAY 157
CacbgA   EVYKRSSASW IEDQDFWRFS GIFRDVYLY. AVPETHVNDI FIKTDLYDDF 236

BsbgaB   IDILYDHGVY INLGTATATT PAWFVKKYPD SLPI...... .DESGVILSF 106
KplacZ   RDGTLQVQAT IEATEAALAG LSVGVSLWRG EEQFAAGRQP LGTPTVDERG 291
EclacZ   SRAVLEAEVQ MCGELRDYLR VTVSLW..QG ETQVASGTAP FGGEIIDERG 282
EcebgA   CDATLSCEVV LENLAASPVV TTLEYT..LF DGERVVHSSA IDHLA....I 201
CacbgA   KNAKLNAELK MIGNSETTVE TYLEDK.... EGNKIAISEK IPFSD..... 277

BsbgaB   GSRQHYCPNH PQLIT...HI KRLVRAIAER YKNHPALKMW HVNNEYACHV 153
KplacZ   HYAERVDFSL AVATPAHWSA ETPNCYRAVV TLWRGDE.LL EAEAWDIGFR 340
EclacZ   GYADRVTLRL NVENPKLWSA EIPNLYRAVV ELHTADGTLI EAEACDVGFR 332
EcebgA   EKLTSATFAF TVEQPQQWSA ESPYLYHLVM TLKDANGNVL EVVPQRVGFR 251
CacbgA   ....ELTLYL DAQNINLWSA EEPNLYTLYI LVNKKDGNLI EVVTQKIGFR 323

BsbgaB   SKCFCENCAV AFRKWLKERY KTIDELNERW GTNFWGQRYN HWDEINPPRK 203
KplacZ   RIEIADGLLR LNGKPLLIRG VNRHEHHHLR GQVVTEADMV QDILLMKQNN 390
EclacZ   EVRIENGLLL LNGKPLLIRG VNRHEHHPLH GQVMDEQTMV QDILLMKQNN 382
EcebgA   DIKVRDGLFW INNRYVMLHG VNRHDNDHRK GRAVGMDRVE KDLQLMKQHN 301
CacbgA   HFEMKDKIMC LKWKRIIFKG VNRHEFSARR GRSITKEDML WDIKFLKQHN 373
```

```
BsbgaB  APTFINPSQE LDYYRFMNDS ILKLFLTEKE ILREVTPDIP EVSTNFMGSF 253
KplacZ  FNAVRCSHYP NAPRWYELCN RYGLYVVDEA NIETHG.... ....MVPMNR 432
EclacZ  FNAVRCSHYP NHPLWYTLCD RYGLYVVDEA NIETHG.... ....MVPMNR 424
EcebgA  INSVRTAHYP NDPRFYELCD IYGLFVMAET DVESHG.FAN VGDISRI... 347
CacbgA  INAVRTSHYP NQSLWYRLCD EYGIYLIDET NLESHGSWQK MGQIEPSWNV 423

BsbgaB  KPLNYFQWAQ HVDIVTWDSY PDPREGLPIQ HAMMNDLMRS LRKGQPFILM 303
KplacZ  LSDDPAWLPA FSARVTRMVQ SNRNHPCIII WSLGNESGGG GNHEALYHWL 482
EclacZ  LTDDPRWLPA MSERVTRMVQ RDRNHPSVII WSLGNESGHG ANHDALYRWI 474
EcebgA  .TDDPQWEKV YVERIVRHIH AQKNHPSIII WSLGNESGYG CNIRAMYHAA 396
CacbgA  PGSLPQWQAA VLDRASSMVE RDKNHPSVLI WSCGNESYAG EDIYQMSKYF 473

BsbgaB  EQVTSHVNWR DINVPKPPGV MRLWSYATIA RGADGIMFFQ WRQSRAGAEK 353
KplacZ  KRNDPSRPVQ YEGGG..ADT TATDIICPMY ARVERDQPIP AVPKWGIKKW 530
EclacZ  KSVDPSRPVQ YEGGG..ADT TATDIICPMY ARVDEDQPFP AVPKWSIKKW 522
EcebgA  KRLDDTRLVH YEED...RDA EVVDIISTMY TRVPLMNEFG E......... 434
CacbgA  RKKDPSRLVH YEGVTRCREF MTRRHESRMY AKAAEIEEYL N......... 514

BsbgaB  FHGAMVPHFL NENNRIYREV TQLGQELKKL DCLVGSRIKA EVAIIFDWEN 403
KplacZ  ISLPGEQRPL ILCEYAHAMG NSLGNFADYW QAFREYPRLQ GGFIWDWADQ 580
EclacZ  LSLPGETRPL ILCEYAHAMG NSLGGFAKYW QAFRQYPRLQ GGFVWDWVDQ 472
EcebgA  ...YPHPKPR IICEYAHAMG NGPGGLTEYQ NVFYKHDCIQ GHYVWEWCDH 481
CacbgA  ...DNPKKPY IQLRYMHSMG NSTGGMMKYT ELEDKYLMYQ GGFIWDYGDQ 561

BsbgaB  WWAVELSSKP HNKLRYIPIV EAYYRELYKR NIAVDFVRPS DDLTKYKVVI 453
KplacZ  AIRKTFADGS VGWAYGGDFG DKPNDRQFCM NGLVFPDRTP HPSLVEAKHA 630
EclacZ  SLIKYDENGN PWSAYGGDFG DTPNDRQFCM NGLVFADRTP HPALTEAKHQ 622
EcebgA  GIQAQDDHGN VWYKFGGDYG DYPNNYNFCL DGLIYSDQTP GPGLKEYKQV 531
CacbgA  ALYRKLPDGK EVLAYGGDFT DRPTDYNFSG NGLIYADRTI SPKAQEVKYL 611

BsbgaB  APMLYMVKEG EDENLRQFVA NGGTLIVSFF SGIVDENDRV HLGGYPGPLR 503
KplacZ  QQYFQFTLLS TSPLRVRIIS EYLFRPTDNE VVRWQVQAAG EPLYHGDLTL 680
EclacZ  QQFFQFRLSG ...QTIEVTS EYLFRHSDNE LLHWMVALDG KPLASGEVPL 669
EcebgA  IAPVKIHARD LTRGELKVEN KLWFTTLDDY TLHAEVRAEG ESLATQQIKL 581
CacbgA  YQNVKLEPDE ...KGVTIKN QNLFVNTDKY DLYYIVERDG KLIKDGYLNV 658

BsbgaB  DILGIFVEEF VPYPETKVNK IYSNDGEYDC TTWADIIRLE GAEPLATFKG 553
KplacZ  ALPPEGSDEI TLLDSLILPE GARAVWLTLE VTQPQATAWS EAEHRVAWQQ 730
EclacZ  DVAPQGKQLI ELPELPQ.PE SAGQLWLTVR VVQPNATAWS EAGHISAWQQ 718
EcebgA  PDVAPNSEAP LQITCRS.WT PAKRSLNITV TKDSRTRYSE AGHPIATYQF 630
CacbgA  SVAPDEEKYI ELPIGNY.NF PEEIVLTTSL RLAQATLWAE KGYEIAFGQK 707
```

```
BsbgaB  DWYAGLPAVT RNCYGKGEGI YVGTYPDSNY LGRLLEQVFA KHHINPILEV 603
KplacZ  FPLPAPLGCR RPPCLPALPD LIVSDEVWQI RAGSQCWTID RRTGLLSRWS 780
EclacZ  WRLAENLSVT LPAASHAIPH LTTSEMDFCI ELGNKRWQFN RQSGFLSQMW 768
EcebgA  PLKEN.TAQP VPFAPNKCAS VTLEDDRLSC TVRGYNFAIT FSKMSGKPTS 679
CacbgA  VIKEK.SDMN ...NHNSESK MKIIHGDVNI GVHGKDFKAI FSKQEGGIVS 753

BsbgaB  AENVEVQQRE TDEWKYLIII NHNDYEVTLS LPEDKIYQNM IDGKCFRGGE 653
KplacZ  VGGQEQ.LLT PLRDQ.FIR. APLDNDIGVS EVERIDPNAW VERWRSAGLY 827
EclacZ  IGDKKQ.LLT PLRDQ.FTR. APLDNDIGVS EATRIDPNAW VERWKAAGHY 815
EcebgA  WQVNGESLLT REPKINFFKP MMIDNHKQEY E..... ....GLWQPN 716
CacbgA  LRYNNKEFIT RTPKT.FYWR ATTDNDRGNR HE..... ....FRCSQW 790

BsbgaB  LRIQGVDVAV LREHDEAGKV                               673
KplacZ  DLEAHCVQCD AQRLANETLV DCRWHYLRGE EVVIVSHWRM HFTADGTLRL 877
EclacZ  QAEAALLQCT ADTLADAVLI TTAHAWQHQG KTLFISRKTY RIDGSGQMAI 865
EcebgA  HLQIMQEHLR DFAVEQSDGE VLIISRTVIA PPVFDFGMRC TYIWRIAADG 766
CacbgA  LAATMGQKYV DFSVEEFDEK I.TLYYTYQL PTVP..STNV KITYEVSGEG 837

KplacZ  AVDGERAETL PPLPRVGLHF QVADQQAPVS WLGLGPHENY PDRRSSACFA 927
EclacZ  TVDVEVASDT PHPARIGLNC QLAQVAERVN WLGLGPQENY PDRLTAACFD 915
EcebgA  QVNVALSGER YGDYPHIIPC IGFTMGINGE YDQVAYYGRG PGENYADSQQ 816
CacbgA  IIKVNVKYKG VSGLPE.LPV LGMDFKLLAE FNSFSWYGMG PEENYIDRCE 886

KplacZ  RWEQPLAAMT TPYIFPTENG LRCDTQALDW GRWHISGHFH FSVQPWSTRQ 977
EclacZ  RWDLPLSDMY TPYVFPSENG LRCGTRELNY GPHQWRGDFQ FNISRYSQQQ 965
EcebgA  ANIIDIWRQA VDAMFENYPF PQNNGNRQHV RWTALTNRHG NGLLVVPQRP 866
CacbgA  GAKLGIYEST Q                                       897

KplacZ  LMETDHWHKM QAEDGVWITL DGLHMGVGGD DSWTPSVLPQ WLLSQTRWQY 1027
EclacZ  LMETSHRHLL HAEEGTWLNI DGFHMGIGGD DSWSPSVSAE FQLSAGRYHY 1015
EcebgA  INFSAWHYTQ ENIHAAQHCN ELQRSDDITL GTSITSCLAS APTPGAARCW 916

KplacZ  EVSLRSL                                             1034
EclacZ  QLVWCQK                                             1022
EcebgA  TPGASGSVTS ATALRCCRFL AEKLPRKAWR RMSSAQGSFP RICTRRIS   964
```

```
B   EcuidA  395 LIARDKNHPSVVMWSIANEPDT  416
    BsbgaB   81 ATTPAWFVKKYPDSLPIDESGV  102
    EcebgA  364 HIHAQKNHPSIIIWSLGNESGY  385
    KplacZ  450 MVQSNRNHPCIIIWSLGNESGG  471
    CacbgA  441 MVERDKNHPSVLIWSCGNESYA  462
    EclacZ  444 MVQRDRNHPSVIIWSLGNESGH  465
```

Fig. 7.8: Alignment of the amino acid sequence of the ß-galactosidase gene of C. acetobutylicum cbgA (CacbgA)with the amino acid sequences of E. coli ebgA (EcebgA: Stokes et al., 1985); E. coli lacZ (EclacZ: Kalnins et al., 1983); K. pneumoniae lacZ (KplacZ: Buvinger and Riley, 1985); and B. stearothermophilus bgaB (BsbgaB: Hirata et al., 1986).

Amino acids have been placed into groups of similar function using the programme SIMPLIFY (Devereux et al., 1984) and aligned using WORDSEARCH and SEGMENTS (Devereux et al., 1984). Groups are identified by the single letter codes below. Regions of identical homology to one or more of the sequences are underlined.

A P A G S T - neutral, weakly hydrophobic
D Q N E D - hydrophobic, acid amine
H H K R - hydrophobic, basic
I L I V M - hydrophobic
F F Y W - hydrophobic, aromatic
C C - cross-link forming

```
KplacZ  IDIADAAHAH AADFHAIIAH DDFHDD.AIA HIDHIAAHAI FAAFHDDIAA 49
EclacZ  ........AI IADAIAIIID HHDFDDAAIA DIDHIAAHAA FAAFHDADDA 42
EcebgA  .......... .......... .......... .......... ..........
CacbgZ  ........II DDHAAIDFID DADIFHIDHI DAHADAFFFD HFDDIHIDDA 42

KplacZ  HDDIAAAHHH DIDA..AAAA IAAAAHIAAI H.IIADDIAD CHAAAIAADF 96
EclacZ  HADHAADDIH AIDADFHFAF FAAADAIADA ..FIDCDIAD ADAIIIAADF 90
EcebgA  .......... .......... .......... .......IAD FAHIAIAAIF 13
CacbgA  IAIHDDIDAH FHFAFADDAA IHIHDFFHDD FDIAFIDFID IAAHIDIDAF 92

KplacZ  DIDAFDAAIF ADIHFAIDAA AAHIADDDAA ACFAIHFAID DAFHDDADAD 146
EclacZ  DIHAFDAAIF ADIAFAIAID AAFIAADDAA ACFAIAFDID DAFIDDADAH 140
EcebgA  DIDAHAHIDF ADDAFAFAID IAFIAADDAA AAFDHIFAIA DAF.DAHDAI 62
CacbgA  DHCDFIDAIF AFDAHDDIHA AHIAHAFDAI AAFIAFFDIH DDI.HDHDAF 141

KplacZ  IIFDAIDAAF HIFCDAIFIA FADDAHIAAA FDIAAFIHAA DDHICIIIIH 196
EclacZ  IIFDAIDAAF HIFCDAHFIA FADDAHIAAD FDIAAFIHAA DDHIAIIIIH 190
EcebgA  IHFDAIDAFF DIFIDADFIA FAHAAHIAAD FDIAAIIHAA DDIICIHIID 112
CacbgA  IAFDAIDAAF FIFIDADFIA FADDAFAAAD FDIADFIHDA DDHIAIDIFH 191

KplacZ  FAAAAFIDDD DIFHIAAIFH AIFIIDHADD HICDIDIAAA .IDAIFHDAA 248
EclacZ  FADAAFIDDD DIFHIAAIFH DIAIIHHAAA DIADFHIAAH .FDDDFAHAI 239
EcebgA  FADAAFIDDD DIFFAAAIFH DIFIIAHHIA HIDDFAIHAD .FDDAFCDAA 161
CacbgA  HAAAAFIDDD DFFHFAAIFH DIFIFAIADA HIDDIFIHAD IFDD.FHDAH 240

KplacZ  IDIDAAIDAA DAAIAAIAIA IAIFHADDDF AAAHDAIAAA AIDDHAHFAD 295
EclacZ  IDADIDICAD IHDFIHIAIA I..FDADADI AAAAAAFAAD IIDDHAAFAD 287
EcebgA  IACDIIIDDI AAAAIIAAID FAIFDADHII HA.....AAI DHIAIDHIAA 206
CacbgA  IDADIHIIAD ADAAIDAFID DHDADHIAIA DH..... .....IAFAD 277

KplacZ  HIDFAIAIAA AAHFAADAAD CFHAII.AIF HADDIIDADA FDIAFHHIDI 344
EclacZ  HIAIHIDIDD AHIFAADIAD IFHAIIDIHA ADAAIIDADA CDIAFHDIHI 337
EcebgA  AAFAFAI.DD ADDFAADAAF IFHIIIAIHD ADADIIDIIA DHIAFHDIHI 255
CacbgA  DIAIFIDADD IDIFAADDAD IFAIFIIIDH HDADIIDIIA DHIAFHHFDI 327

KplacZ  ADAIIHIDAH AIIIHAIDHH DHHHIHADII ADADIIDDII IIHDDDFDAI 394
EclacZ  DDAIIIIDAH AIIIHAIDHH DHHAIHADII DDDAIIDDII IIHDDDFDAI 387
EcebgA  HDAIFFIDDH FIIIHAIDHH DDDH.HHAHA IAIDHIDHDI DIIH....DH 300
CacbgA  HDHIICIHFH HIIFHAIDHH DFAAHHAHAI AHDDIIFDIH FIHDHDIDAI 377

KplacZ  HCAHFADAAH FFDICDHFAI FIIDDADIDA HAI..... IAI.DHIADD 437
EclacZ  HCAHFADHAI FFAICDHFAI FIIDDADIDA HAI..... IAI.DHIADD 429
EcebgA  DIDAIHAAHF ADDAHFFDIC DIFAIFIIAD ADIDAHAFAD IADIAHIADD 350
CacbgA  HAAHFADDAI FFHICDDFAI FIIDDADIDA HAAFDHIADI DAAFDIAAAI 427
```

```
KplacZ  AAFIAAFAAH IAHIIDADHD HACIIIFAIA DDAAAAADHD AIFHFIHHDD 486
EclacZ  AHFIAAIADH IAHIIDHDHD HAAIIIFAIA DDAAHAADHD AIFHFIHAID 479
EcebgA  ADFDHIFIDH IIHHIHADHD HAAIIIFAIA DDAAFACDIH AIFHAAHHID 400
CacbgA  ADFDAAIIDH AAAIIDHDHD HAAIIIFACA DDAFAADDIF DIAHFFHHHD 477

KplacZ  AAHAIDFDAA AADAAAADII CAIFAHIDHD DAIAAIAHFA IHHFIAIAAD 536
EclacZ  AAHAIDFDAA AADAAAADII CAIFAHIDDD DAFAAIAHFA IHHFIAIAAD 529
EcebgA  DAHIIHFDDD .......... ...HDADIID IIAAIFAHIA IIDDFADFAH 437
CacbgA  AAHIIHFDA. .........I AHCHDFIAHH HDAHIFAHAA DIDDFIDDDA 517

KplacZ  DHAIIICDFA HAIADAIADF ADFFDAFHDF AHIDAAFIFD FADDAIHHAF 586
EclacZ  AHAIIICDFA HAIADAIAAF AHFFDAFHDF AHIDAAFIFD FIDDAIIHFD 579
EcebgA  AHAHIICDFA HAIADAAAAI ADFDDIFFHH DCIDAHFIFD FCDHAIDADD 487
CacbgA  HHAFIDIHFI HAIADAAAAI IHFADIDDHF IIFDAAFIFD FADDAIFHHI 567

KplacZ  ADAAIAFAFA ADFADHADDH DFCIDAIIFA DHAAHAAIID AHHADDFFDF 636
EclacZ  DDADAFAAFA ADFADAADDH DFCIDAIIFA DHAAHAAIAD AHHDDDFFDF 629
EcebgA  DHADIFFHFA ADFADFADDF DFCIDAIIFA DDAAAAAIHD FHD...IIAA 534
CacbgA  ADAHDIIAFA ADFADHAADF DFAADAIIFA DHAIAAHADD IHF...IFDD 614

KplacZ  AIIAAAAIHI H..IIADFIF HAADDDIIHF DIDAAADAIF HADIAIAIAA 684
EclacZ  HIAADAI... ..DIAADFIF HHADDDIIHF IIAIDAHAIA AADIAIDIAA 674
EcebgA  IHIHAHDIAH ADIHIDDHIF FAAIDDFAIH ADIHADADAI AADDIHIADI 584
CacbgA  IHIDADD..H AIAIHDDDIF IDADHFDIFF IIDHDAHIIH DAFIDIA..I 660

KplacZ  DAADDIAIID AIIIADAAHA IFIAIDIADA DAAAFADADH HIAFDDFAIA 734
EclacZ  DAHDII.DIA DIADADAAAD IFIAIHIIDA DAAAFADAAH IAAFDDFHIA 723
EcebgA  AADADAAIDI ACHAFAAAHH AIDIAIAHDA HAHFADAAHA IAAFDFAIHD 634
CacbgA  AADDDHFIDI AIADFDFADD IIIAAAIHIA DAAIFADHAF DIAFADHIIH 710

KplacZ  AAIACHHAAC IAAIADIIIA DDIFDIHAAA DCFAIDHHAA IIAHFAIAAD 784
EclacZ  DDIAIAIAAA AHAIAHIAAA DIDFCIDIAD HHFDFDHDAA FIADIFIADH 773
EcebgA  DAADAIAFAA DHCAAIAIDD DHIACAIHAF DFAIAFAHIA AHAAAFDIDA 684
CacbgA  DHADID..DH DADAHIHIIH ADIDIAIHAH DFHAIFAHDD AAIIAIHFDD 758

KplacZ  DDIIAAI.HD DFIHAAIDDD IAIA...DID HIDADAFIDH FHAAAIFDID 830
EclacZ  HDIIAA.IHD DFAHAAIDDD IAIA...DAA HIDADAFIDH FHAAAHFDAD 819
EcebgA  DAIIAHDAHI DFFHAIIIDD HHDDFDAIFD ADHIDIIDDH IHDFAIDDAD 734
CacbgA  HDFIAHAAHA FFFHAAADDD HADHHDFHCA DFIAAAIADH FIDFAIDDFD 808

KplacZ  AHCIDCDADH IADDAIIDCH FHFIHADDII IIAHFHIHFA ADAAIHIAID 880
EclacZ  AAIIDCAADA IADAIIIAAA HAFDHDAHAI FIAHHAFHID AAADIAIAID 869
EcebgA  ADIIIIAHAI IAAAIFDFAI HCAFIF.... .......HIA ADADIDIAIA 773
CacbgA  DHIAIFFAFD IAAIA..AAD IHIAF..... .......DIA ADAIIHIDIH 844
```

```
KplacZ   ADHADAIAAI AHIAIHFDIA DDDAAIAFIA IAAHDDFADH HAAACFAHFD 930
EclacZ   IDIAADAAHA AHIAIDCDIA DIADHIDFIA IAADDDFADH IAAACFDHFD 919
EcebgA   ADHFADFAHI IACIAFAIAI DADFDDIAFF AHAAADDFAD ADDADIIDIF 823
CacbgA   FHAIAAIAD. .IAIIAIDFH IIADFDAFAF FAIAADDDFI DHCDAAHIAI 892

KplacZ   DAIAAIAAAF IFAADDAIHC DADAIDFAHF HIAAHFHFAI DAFAAHDIID 980
EclacZ   IAIADIFAAF IFAADDAIHC AAHDIDFAAH DFHADFDFDI AHFADDDIID 969
EcebgA   HDAIDAIFDD FAFADDDADH DHIHFAAIAD HHADAIIIIA DHAIDFAAFH 873
CacbgA   FDAAD*                                                  897

KplacZ   ADHFHHIDAD DAIFIAIDAI HIAIAADDAF AAAIIADFII ADAHFDFDIA 1030
EclacZ   AAHHHIIHAD DAAFIDIDAF HIAIAADDAF AAAIAADFDI AAAHFHFDII 1019
EcebgA   FADDDIHAAD HCDDIDHADD IAIAAAIAAC IAAAAAAAAA HCFAAAAAAA 923

KplacZ   IHAI                                                    1034
EclacZ   FCDH                                                    1022
EcebgA   IAAAAAIHCC HFIADHIAHH AFHHIAAADA AFAHICAHHI A           964
```

kDa) estimated on the basis of SDS-PAGE in Chapter 6. An examination of the charged residues would suggest that the protein has a net negative charge of 25 (Kyte and Doolittle, 1982).

The available nucleotide sequences of other ß-galactosidase genes were compared to the ß-galactosidase nucleotide sequence from C. acetobutylicum using the programmes WORDSEARCH and SEGMENTS (Devereux et al., 1984). Results of this comparison showed that there was no obvious homology to the three Gram-negative ß-galactosidase sequences of E. coli (lacZ, ebgA) and K. pneumoniae (lacZ). Since the B. stearothermophilus bgaB nucleotide and amino acid sequences were not in the data bases, these were added separately. No obvious homology was detected to the bgaB nucleotide sequence from the Gram-positive B. stearothermophilus. No obvious homology was detected to any DNA sequence in the bacterial component of Genbank (Version 56: Bilofsky and Burks, 1988), or European Molecular Biology Laboratory (EMBL: version 15: Cameron, 1988) DNA sequence data or to the National Biochemical Research Foundation (NBRF) Nucleotide Sequence Data Library (Version 33: Sidman et al., 1988).

The S. bovis lacOP nucleotide sequence (Gilbert and Hall, 1987) and the E. coli lacOP and CAP nucleotide sequences (Manly and Matthews, 1984) were compared to the upstream region of the cbgA gene using BESTFIT (Smith and Waterman, 1981) and GAP (Needleman and Wunsch, 1970). No significant homology was detected between these two sequences and the upstream region, nor was any significant homology detected throughout the 5.2 kb EcoRI fragment.

The deduced amino acid sequence of the C. acetobutylicum ß-galactosidase protein was also compared to the amino acid sequences of the other ß-galactosidase proteins and those within the NBRF resource data base (Version 16: Sidman et al., 1988) using WORDSEARCH and SEGMENTS (Devereux et

al., 1984). Extensive homology was detected at the amino acid level (Fig. 7.7). Homology was observed throughout the amino acid sequences of the C. acetobutylicum CbgA and the ß-galactosidase proteins for E. coli (LacZ, EbgA), as well as the K. pneumoniae LacZ (Fig 7.7a).

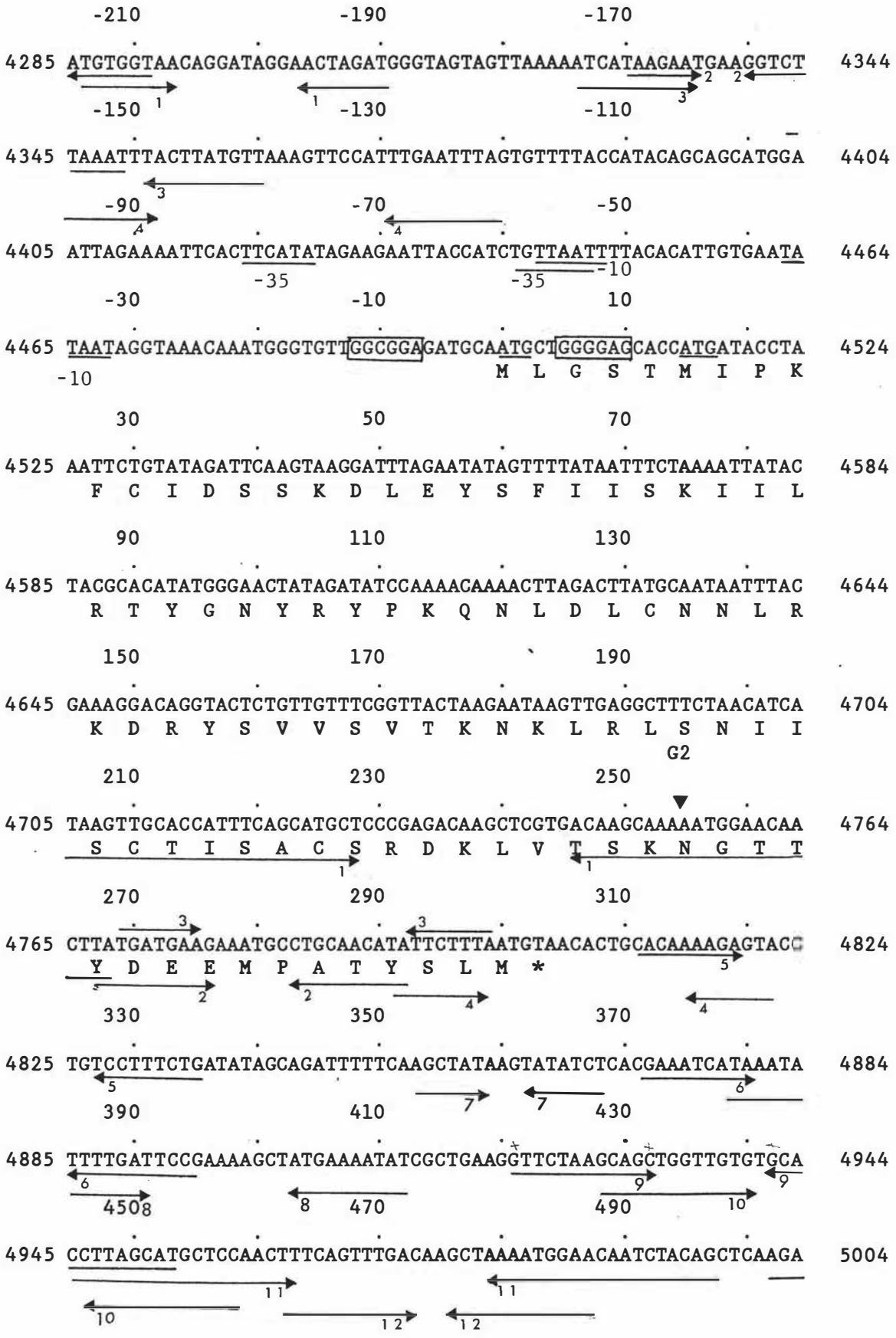
High homology was noted between the region containing the active site of E. coli LacZ (Glu-461: Lerrchen and Legler, 1983) and all other ß-galactosidase amino acid sequences (Fig. 7.7b). The region of homology between the CbgA and the LacZ was between residues 441 through 461 of CbgA and residues 444 through 465 of LacZ. In addition to the ß-galactosidases, homology about the active site was also detected to a site within the E. coli glucuronidase enzyme (Fig. 7.7b; LIARDKNHPSVVMWSIANEPDT). Comparison of the amino acid sequence of the C. acetobutylicum CbgA to that of B. stearothermophilus BgaB showed less homology than of the gene products of other ß-galactosidase genes. The amino acid sequences of these ß-galactosidase enzymes were further compared to the CbgA (C. acetobutylicum) using SIMPLIFY and WORDSEARCH (Devereux et al., 1984). This increased the degree of homology (Fig. 7.8), revealing that where the actual amino acids differed, the substitution was often by another amino acid from the same group. To determine if the amino acid sequence deduced from the C. acetobutylicum cbgA gene shared homology to other amino acid sequences, a search of the NBRF protein resource data base (Version 16: Sidman et al., 1988) was carried out. This yielded no new amino acid sequences which shared reasonable homology to the CbgA protein.

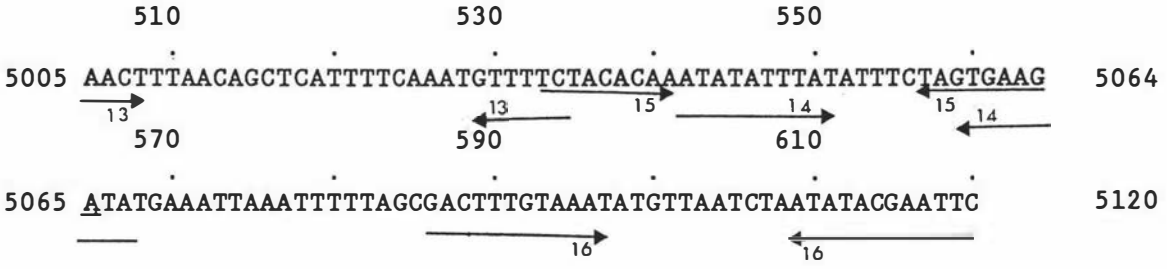
The second ORF that was potentially coding corresponded to the locus I region identified by Tn5 mutagenesis. Evidence that this ORF is coding is supported by TESTCODE (Fickett, 1982), by the detection of SD sequences and promoter sites in the appropriate position with respect to the translational start site, and by the presence of the Tn5 insertion G2 within the ORF. The other presumptive ORF's identified in this region (Fig. 7.5) did not possess SD sequences, potential promoter sites, and did not correspond to the region where the Tn5 mapped.

Fig. 7.9: The complete DNA sequence and deduced amino acid sequence of the region of the 5.2 kb EcoRI fragment containing the regulatory ß-galactosidase ORF (cbgR). The -10 and -35 promoter regions are underlined. The Shine-Dalgarno (SD) sequences are boxed.

The positions of a number of Tn5 inserts shown by ▼(ß-galactosidase +) and ▽(ß-galactosidase-) are as indicated. The number given in brackets above the symbol corresponds to the number assigned to the Tn5 insertion in Chapter 6.

Possible stem-loop structures in the terminal region are indicated by arrows. The amino acids are identified by the single letter code (see Fig. 7.6).


-210
-190
-170
4285 ATGTGGTAACAGGATAGGAACTAGATGGGTAGTAGTTAAAAATCATAAGAATGAAGGTCT 4344
-150
-130
-110
4345 TAAATTTACTTATGTTAAAGTTCCATTTGAATTTAGTGTTTTACCATACAGCAGCATGGA 4404
-90
-70
-50
4405 ATTAGAAAATTCACTTCATATAGAAGAATTACCATCTGTTAATTTTACACATTGTGAATA 4464
-35
-35
-10
-30
-10
10
4465 TAATAGGTAAACAAATGGGTGTTGGCGGAGATGCAATGCTGGGGAGCACCATGATACCTA 4524
-10
M L G S T M I P K
30
50
70
4525 AATTCTGTATAGATTCAAGTAAGGATTTAGAATATAGTTTTATAATTTCTAAAATTATAC 4584
F C I D S S K D L E Y S F I I S K I I L
90
110
130
4585 TACGCACATATGGGAACTATAGATATCCAAAACAAAACTTAGACTTATGCAATAATTTAC 4644
R T Y G N Y R Y P K Q N L D L C N N L R
150
170
190
4645 GAAAGGACAGGTACTCTGTTGTTTCGGTTACTAAGAATAAGTTGAGGCTTTCTAACATCA 4704
K D R Y S V V S V T K N K L R L S N I I
G2
210
230
250
4705 TAAGTTGCACCATTTCAGCATGCTCCCGAGACAAGCTCGTGACAAGCAAAAATGGAACAA 4764
S C T I S A C S R D K L V T S K N G T T
270
290
310
4765 CTTATGATGAAGAAATGCCTGCAACATATTCTTTAATGTAACACTGCACAAAAGAGTACC 4824
Y D E E M P A T Y S L M *
330
350
370
4825 TGTCCTTTCTGATATAGCAGATTTTTCAAGCTATAAGTATATCTCACGAAATCATAAATA 4884
390
410
430
4885 TTTTGATTCCGAAAAGCTATGAAAATATCGCTGAAGGTTCTAAGCAGCTGGTTGTGTGCA 4944
4508
470
490
4945 CCTTAGCATGCTCCAACTTTCAGTTTGACAAGCTAAAATGGAACAATCTACAGCTCAAGA 5004


510
530
550
5005 AACTTTAACAGCTCATTTTCAAATGTTTTCTACACAAATATATTTATATTTCTAGTGAAG 5064
13
13
15
14
15
14
570
590
610
5065 ATATGAAATTAAATTTTTAGCGACTTTGTAAATATGTTAATCTAATATACGAATTC 5120
16
16

Facing page 207

Fig. 7.10: Alignment of the amino acid sequence of the ß-galactosidase regulator (cbgR) of C. acetobutylicum cbgR (CacbgR) with the amino acid sequence of the "0.3 kb gene" from Bacillus subtilis (Bs.3kb: Stephens et al., 1984). Sequences have been aligned using the programmes WORDSEARCH and SEGMENTS (Devereux et al., 1984).

The amino acids are identified by the single letter code (see Fig. 7.6) and the regions of identical homology are underlined.

Facing page 207

Fig. 7.11: Alignment of the amino acid sequence of the ß-galactosidase regulator (cbgR) of C. acetobutylicum cbgR (CacbgR) to the amino acid sequence of the "0.3 kb gene" from Bacillus subtilis (Stephens et al., 1984).

Amino acids have been placed into groups of similar function using the programme SIMPLIFY (Devereux et al., 1984) and aligned using WORDSEARCH and SEGMENTS (Devereux et al., 1984).

The amino acids are identified by the single letter code (see Fig. 7.8) and the regions of identical homology are underlined.

A

```
Bs.3kb  MLGRRRGVMS DEFKYELAKD LGF...YDTV KNGGWGEIRA R....DAGNM 43
CaCbgR  MLG...STMI PKFCIDSSKD LEYSFIISKI ILRTYGNYRY PKQNLDLCNN 47

Bs.3kb  VKRA....IE IAEQQMAQNQ NNR                                62
CaCbgR  LRKDRYSVVS VTKNKLRLSN IISCTISACS RDKLVTSKNG TTYDEEMPAT YSLM 102
```

B

```
Bs.3kb  IIAHHHAIIA DDFHFDIAHD IAF...FDAI HDAAFADIH. AHDAADI... 43
CaCbgR  IIA...AAII AHFCIDAAHD IDFAFIIAHI IIHAFADFHF AHDDIDICDD 47

Bs.3kb IHHA....ID IADDDI..AD DDDDH                              62
CaCbgR IHHDHFAIIA IAHDHIHIAD IIACAIAACA HDHIIAAHDA AAFDDDIAAA FAII 102
```

The nucleotide sequences adjacent to the 5' end of the ORF were analyzed for sequences complementary to the 3' end of the 16S rRNA of B. subtilis and E. coli. Two potential start codons were identified, both possessing good SD sequences (GGGGAG; GCGGAG: Table 7.2). It is impossible to distinguish which is used for translation of the cbgR gene. Taking the first ATG (second ATG) as the start site, an ORF (designated cbgR: clostridial ß-galactosidase regulator) consisting of a coding sequence of 303 bp (288 bp) was identified, starting with an ATG codon at position 4500 (4515) and ending with a TAA (stop) codon at position 4805 (Fig. 7.9). The region upstream of the presumptive translational start sites was also analyzed for potential promoter sequences using BESTFIT (Smith and Waterman, 1981) and GAP (Needleman and Wunsch, 1970). Two regions upstream of the presumptive start sites were identified as potential Clostridium promoter sequences (Table 7.2).

A number of inverted repeats that have the potential to form stem-loop structures were located at the end of the cbgR gene and downstream of the stop codon. The nucleotide sequence of the ORF contained 73%dA+dT. The resulting DNA sequence was analyzed using CODONFREQUENCY (Devereux et al., 1984), revealing that this ORF possessed a codon bias towards dA+dT. The ORF was sufficient in size to code for an unmodified protein of 101 (96) amino acids with a calculated Mr of 11 (10.5) kDa (Fig. 7.10). An examination of the charged residues would suggest that this protein has a net positive charge of 6 (Kyte and Doolittle, 1982).

No significant homology was detected to any DNA sequence in the Genbank (Version 56: Bilofsky and Burks, 1988) or EMBL (Version 15: Cameron, 1988) DNA sequence data bases or to the NBRF Nucleotide Sequence Data Library (Version33: Sidman et al., 1988). The deduced amino acid sequence of the cbgR present at Locus I was also compared to other known protein sequences in the NBRF protein resource data base (Version 16: Sidman et al., 1988) using WORDSEARCH and SEGMENTS (Devereux et al., 1984). Moderate homology

was detected to the amino acid sequence of the '0.3 kb gene' from B. subtilis (Fig. 7.10). Analysis of these sequences using SIMPLIFY, WORDSEARCH, and SEGMENTS Devereux et al., 1984) increased the extent of the homology (Fig 7.11).

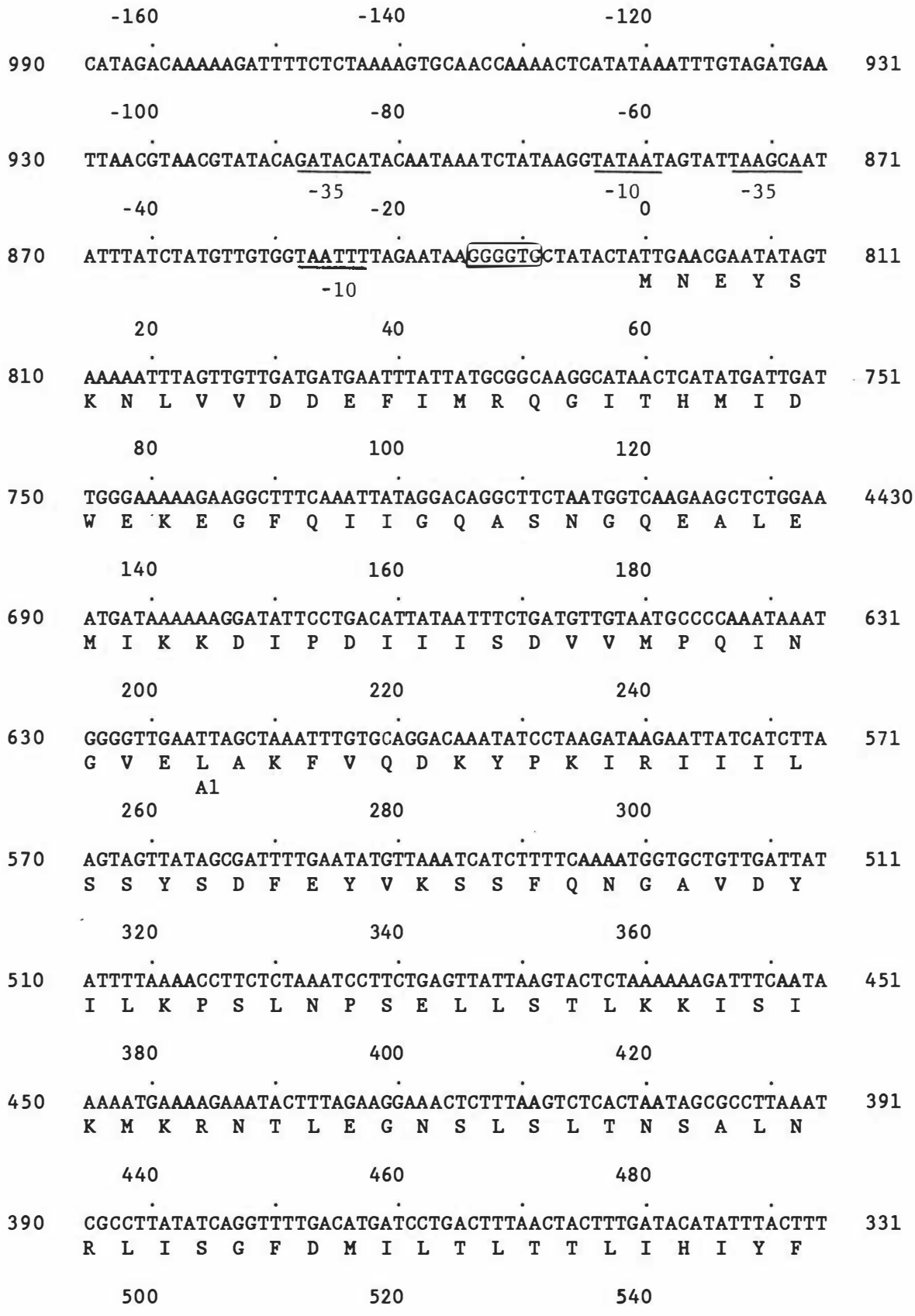
The third potential coding sequence identified (ORF3) consisted of a coding sequence of 796 bp, located adjacent to cbgA but transcribed divergently from it (Fig. 7.12). Evidence that this ORF is coding is supported by the results from TESTCODE (Fickett, 1982), and by the detection of a SD sequence and promoter sites in the appropriate position with respect to the translational start site. The coding sequence did not contain an in-frame ATG start codon, but rather possessed an in-frame TTG (met) start codon. No stop codon was present on this ORF and so presumably it terminates outside the region sequenced.

The nucleotide sequences upstream of the 5' end of the ORF were screened for sequences complementary to the 3' end of the 16S rRNA of B. subtilis and E. coli (Table 7.2). A potential SD sequence similar to that found for the other two ORF's (GGGGTG) was located at position -3 to -8 bp upstream of the TTG start codon (Table 7.2). The region upstream of this presumptive translational start site was analyzed for potential promoter sequences using BESTFIT (Smith and Waterman, (1981) and GAP (Needleman and Wunsch, 1970). The two potential Clostridium promoter sequences identified are shown in Table 7.2. The nucleotide sequence of the ORF contained 73% dA+dT. The resulting DNA sequence was analyzed using CODONFREQUENCY (Devereux et al., 1984). This ORF also exhibited a similar bias towards codons containing dA+dT as observed for the other two ORF's (Table 7.4). This ORF codes for an unmodified truncated protein (275 amino acids long) with a calculated Mr of 32 kDa.

Fig. 7.12: The complete DNA sequence and deduced amino acid sequence of the region of the 5.2 kb EcoRI fragment containing ORF3. The -10 and -35 promoter regions are underlined. The Shine-Dalgarno (SD) sequence is boxed.

The positions of a number of Tn5 inserts are as indicated. The number given in brackets above the symbol corresponds to the number assigned to the Tn5 insertions in Chapter 6.

The amino acids are identified by the single letter code (see Fig. 7.6).


-160 -140 -120
990 CATAGACAAAAAGATTTTCTCTAAAAGTGCAACCAAAACTCATATAAATTTGTAGATGAA 931
-100 -80 -60
930 TTAACGTAACGTATACAGATACATACAATAAATCTATAAGGTATAATAGTATTAAGCAAT 871
-35 -10 -35
-40 -20 0
870 ATTTATCTATGTTGTGGTAATTTTAGAATAAGGGGTGCTATACTATTGAACGAATATAGT 811
-10
M N E Y S
20 40 60
810 AAAAATTTAGTTGTTGATGATGAATTTATTATGCGGCAAGGCATAACTCATATGATTGAT 751
K N L V V D D E F I M R Q G I T H M I D
80 100 120
750 TGGGAAAAAAGAAGGCTTTCAAATTATAGGACAGGCTTCTAATGGTCAAGAAGCTCTGGAA 4430
W E K E G F Q I I G Q A S N G Q E A L E
140 160 180
690 ATGATAAAAAAGGATATTCCTGACATTATAATTTCTGATGTTGTAATGCCCCAAATAAAT 631
M I K K D I P D I I I S D V V M P Q I N
200 220 240
630 GGGGTTGAATTAGCTAAATTTGTGCAGGACAAATATCCTAAGATAAGAATTATCATCTTA 571
G V E L A K F V Q D K Y P K I R I I I L
A1
260 280 300
570 AGTAGTTATAGCGATTTTGAATATGTTAAATCATCTTTTCAAAATGGTGCTGTTGATTAT 511
S S Y S D F E Y V K S S F Q N G A V D Y
320 340 360
510 ATTTTAAAACCTTCTCTAAATCCTTCTGAGTTATTAAGTACTCTAAAAAAGATTTCAATA 451
I L K P S L N P S E L L S T L K K I S I
380 400 420
450 AAAATGAAAAGAAATACTTTAGAAGGAAACTCTTTAAGTCTCACTAATAGCGCCTTAAAT 391
K M K R N T L E G N S L S L T N S A L N
440 460 480
390 CGCCTTATATCAGGTTTTGACATGATCCTGACTTTAACTACTTTGATACATATTTACTTT 331
R L I S G F D M I L T L T T L I H I Y F
500 520 540

```
330  TCAATCATTTTGCTTATTTGGTTTAAACGCAAAGAAATATATGATAATTATAGCAAACGC  271
     S  I  I  L  L  I  W  F  K  R  K  E  I  Y  D  N  Y  S  K  R
                                                   02
        560                 580                 600
                                                  ▽
270  CATGAAATTATTGATTTTACAATTTCAGAAATCGGCAAAGAATTTTCTTTAGACGGATTA  211
     H  E  I  I  D  F  T  I  S  E  I  G  K  E  F  S  L  D  G  L

        620                 640                 660

210  AATAGCAATTATAATATACAAATTGCTAATATTGAAAACGAAATCATAGTTGTAATTCTA  151
     N  S  N  Y  N  I  Q  I  A  N  I  E  N  E  I  I  V  V  I  L

        680                 700                 720

150  AATTTTGATGATAAAAATTATAATTCTCTATTTAGTAAGCTAGAAAATATTTCAGAAGTA   91
     N  F  D  D  K  N  Y  N  S  L  F  S  K  L  E  N  I  S  E  V

        740                 760                 780

90   GTTTCAAATAGTTTTAAAGAAGCATTTTTTGTATTGAGCAACCTCTTCTATTCCAGTAGA   31
     V  S  N  S  F  K  E  A  F  F  V  L  S  N  L  F  Y  S  S  R

        800                 820

30   AATATTAAAGATATATATGAAAAGGAATTC  1
     N  I  K  D  I  Y  E  K  E  F
```

Fig. 7.13: Alignment of the amino acid sequence of ORF3 of C. acetobutylicum (CaORF3) with the amino acid sequences of spo0A (Bsspo0A: Kudoh et al., 1985) and spo0F (Bsspo0F: Trach et al., 1985) from Bacillus subtilis. Sequences have been aligned using the programmes WORDSEARCH and SEGMENTS (Devereux et al., 1984).

The amino acids are identified by the single letter code (see Fig. 7.6) and the regions of identical homology are underlined.

```
BsspoOA  .M.EKIKVCV ADDNRELVS. LLSEYIEGQE DMEVIGVAYN GQECLSLFKE 47
BsspoOF  MMNEKI..LI VDDQYGIRI. LLNEV.FNKE GYQTF.QAAN GLQALDIVTK 45
 CaORF3  .MNEYSKNLV VDDEFIMRQG ITHMIDWEKE GFQIIGQASN GQEALEMIK. 48

BsspoOA  KD.PDVLVL. DIIMPHLDGL AVLERLRESD LKKQPNVIML TAFGQEDVTK 95
BsspoOF  ER.PD.LVLL DMKIPGMDGI EILKRMKVID E..NIRVIIM TAYGELDMIQ 91
 CaORF3  KDIPDIIIS. DVVMPQINGV ELAKF..VQD KYPKIRIIIL SSYSDFEYVK 95

BsspoOA  KAVDLGASYF ILKPFDMENL VGHIRQVSGN ASSVTHRAPS SQSSIIRSSQ 145
BsspoOF  ESKELGALTH FAKPFDI.DE IRDAVKKY.L PLKSN                 124
 CaORF3  SSFQNGAVDY ILKPSLNPSE LLSTLKKISI KMKRNTLEGN SLSLTNSALN 145

BsspoOA  PEPKKKNLDA SITSIIHEIG VPAHIKGYLY LREAISMVYN DIELLGSITK 195
 CaORF3  RLISGFDMIL TLTTLIHIYF SIILLIWFKR KEIYDNYSKR HEIIDFTISE 195

BsspoOA  VLYPDIAKKF NTTASRVERA IRHAIEVAWS RGNIDS.ISS LFGYTVSMTK 244
 CaORF3  IGKEFSLDGL NSNYNIQIAN IENEIIVVIL NFD.DKNYNS LFSKLENISE 244

BsspoOA  AKPTNSEFIA .MVADKLRLE HKAS                           267
 CaORF3  .VVSNSFKEA FFVLSNLFYS SRNIKDIYEK EF                   275
```

Fig. 7.14: The aligned simplified amino acid sequence of ORF3 of C. acetobutylicum (CaORF3) to the amino acid sequences of spo0A (Bsspo0A: Kudoh et al., 1985) and spo0F (Bsspo0F: Trach et al., 1985) from Bacillus subtilis.

Amino acids have been placed into groups of similar function using the programme SIMPLIFY (Devereux et al., 1984) and aligned using WORDSEARCH and SEGMENTS (Devereux et al., 1984).

The amino acids are identified by the single letter code (see Fig. 7.8), and regions of identical homology are underlined.

```
BsspoOA  .I.DHIHICI ADDDHDIIAI .IADFIDADD DIDIIAIAFD ADDCIAIFHD 47
BsspoOF  IIDDHI..II IDDDFAIHII .IDDI.FDHD AFDAF.DAAD AIDAIDIIAH 45
 CaORF3  .IDDFAHDII IDDDFIIHDA IAHIIDFDHD AFDIIADAAD ADDAIDII.H 48

BsspoOA  HD.ADIIIID IIIAHIDAIA IIDHIHDADI HHDADIIIIA AFADDDIAHH 96
BsspoOF  DH.ADIIIID IHIAAIDAID IIHHIHIIDD D..IHIIIIA AFADIDIIDD 92
 CaORF3  HDIADIIIAD IIIADIDAID IAHFI..DDH FAHIHIIIIA AFADFDFIHA 96

BsspoOA  AIDIAAAFFI IHAF.DIDDI IAHIHDIAAD AAAIAHHAAA ADAAIIHAAD 145
BsspoOF  AHDIAAIAHF AHAF.DIDDI HDAIHHFIAI HAD                   124
 CaORF3  AFDDAAIDFI IHAAIDAADI IAAIHH.IAI HIHHDAIDAD AIAIADAAID 145

BsspoOA  ADAHHHDIDA AIAAIIHDIA IAAHIHAFIF IHDAIAIIFD DIDIIAAIAH 195
 CaORF3  HIIAAFDIII AIAAIIHIFF AIIIIIFFHH HDIFDDFAHH H.DIIDFAIA 194

BsspoOA  IIFADIAHHF DAAAAHIDHA IHHAIDIAFA HADIDAIAAI FAFAIAIAHA 245
 CaORF3  DIAHDFAIDA IDADFDIDIA DIDDDIIIII IDFDDHDFDA IFAHIDDIAD 244

BsspoOA  HAADADF..I .AIIADHIHI DHHAA                            267
 CaORF3  IIADAFHDEA FFIIADIFFA AHDIHDIFDH DF                    275
```

No significant homology was detected to any DNA sequence in the Genbank (version 56: Bilofsky and Burks, 1988), or EMBL (version 15: Cameron, 1988) DNA sequence data bases or to the NBRF Nucleotide Sequence Data Library (version 33: Sidman et al., 1988). The deduced amino acid sequence of ORF3 shared homology to a group of proteins belonging to the known family of regulatory genes, including ntrB (K. pneumoniae), cheY (E. coli), and stage 0 sporulation protein A(spo0A) plus stage 0 sporulation protein F(spo0F) from B. subtilis (Nixon et al., 1986; Ronson et al., 1987). The greatest homology was detectable to spo0A and spo0F. Further analysis revealed that this homology was strongest between the N-terminal regions (1-150 amino acids) of these proteins (Fig. 7.13). When these two spo0 proteins were compared to the amino acid sequence of ORF3 using SIMPLIFY, WORDSEARCH, and SEGMENTS (Devereux et al., 1984) extensive homology was detected, (Fig. 7.14).

7.3 DISCUSSION

The results of this phase of the project have provided the complete sequence of a 5.2 kb EcoRI region from C. acetobutylicum known to contain a ß-galactosidase gene. Analysis of this sequence revealed three potential coding sequences; a single large ORF for the ß-galactosidase gene (cbgA), a small ORF for a regulatory gene (cbgR), and a third ORF coding for a protein that is homologous to spo0A and spo0F.

Several other prokaryotic ß-galactosidase genes have been sequenced. These include the E. coli lacZ gene (Kalnins et al., 1983), the E. coli ebgA gene (Stokes et al., 1985), the K. pneumoniae lacZ gene (Buvinger and Riley, 1985), and the B. stearothermophilus bgaB gene (Hirata et al., 1986). The lacOP region from S. bovis (Gilbert and Hall, 1987) has also been sequenced, and the partial sequence of a lacZ gene from L. bulgaricus and L. casei has been reported (De Vos and Simons, 1988).

Table 7.3

Percentage of the dC+dG content of genomic DNA of specific bacteria plus that obtained within the ß-galactosidase genes.

Species	Nucleotide content			
	Genomic DNA %dC+dG (%dA+dT)		ß-galactosidase sequence %dC+dG (%dA+dT)	
B. stearothermophilus	43-62	(57-38)[a]	bgaB 40	(60)[d]
C. acetobutylicum	28	(72)[b]	cbgA 30	(70)
E. coli	48-52	(52-48)[c]	lacZ 56	(44)[e]
			ebgA 55	(45)[f]
K. pneumoniae	56	(44)[c]	lacZ 64	(36)[g]

[a] Claus and Berkeley (1986)
[b] Cato et al . (1986)
[c] Orshov (1986)
[d] Hirata et al. (1986)
[e] Kalnins et al. (1983)
[f] Stokes et al. (1985)
[g] Buvinger and Riley (1985)

7.3.1. Analysis of the sequence of the 5.2 kb fragment

The nucleotide sequence of the ß-galactosidase gene from C. acetobutylicum shares little homology to the other sequenced ß-galactosidase genes. This is in agreement with the hybridization results reported in Chapter 5, where no homology was detected between C. acetobutylicum and E. coli. Since the K. pneumoniae lacZ gene shares strong homology to the equivalent E. coli lacZ gene, and both bacteria possess a relatively high dC+dG% content, little homology would be expected between the K. pneumoniae and C. acetobutylicum genes (Buvinger and Riley, 1985). B. stearothermophilus possesses a similar, if not higher dC+dG content to the other Gram-negative bacteria (43.5-62.2%; Table 7.3). This is much higher than that observed for C. acetobutylicum (28%dC+dG). Hence, the lack of significant nucleotide homology to this organism is again not surprising. In contrast, Gilbert and Hall (1987) reported

strong homology between the S. bovis and E. coli 5' ends of the lacZ structural genes. Since S. bovis possesses 36%dC+dG (Hardie, 1986), this detectable homology is unexpected. Gilbert and Hall (1987) suggested that the lacZ gene may have been acquired relatively recently from an enteric organism.

Although no significant homology was observed between the various nucleotide sequences, considerable amino acid homology was detected. Considerable amino acid homology has been reported for the ß-galactosidase proteins from E. coli (LacZ and EbgA), and K. pneumoniae (LacZ) (Buvinger and Riley, 1985; Stokes et al., 1985). Hence, the C. acetobutylicum CbgA amino acid sequence shares the highest homology to the amino acid sequences from the Gram-negative bacteria. In contrast, Hirata et al. (1986) reported no overall homology between the ß-galactosidase sequences from the Gram-negative bacteria and the B. stearothermophilus BgaB sequence.

The highest homology reported between the B. stearothermophilus BgaB and the E. coli LacZ gene was in the area of the active site (Hirata et al., 1986). This region was also strongly conserved between the E. coli LacZ and the EbgA (Stokes et al ., 1985). When this region was compared to the CbgA amino acid sequence of C. acetobutylicum, very high homology was also detected. The detection of a region within the ß-glucuronidase which shares high homology to the active site of the E. coli ß-galactosidase protein was not previously noted (Jefferson et al., 1986). Hence, there is considerable conservation surrounding the probable active-site of this enzyme. The cbgA gene product (897 amino acids; Mr of 105 kDa) is similar in size to the E. coli lacZ gene product (1023 amino acids; Mr of 116 kDa), ebgA gene product (964 amino acids; 101 Mr of kDa), and K. pneumoniae lacZ (1034 amino acids; Mr of 119 kDa). The B. stearothermophilus gene product is significantly smaller (672 amino acids; Mr of 78 kDa).

Unlike the enteric bacteria that possess the ß-galactosidase as part of a

polycistronic operon that is both positively (CAP and cAMP) and negatively (lacI) regulated, the results here for C. acetobutylicum would suggest that the gene is monocistronic, and controlled by a small gene product, CbgR, which shares homology to a B. subtilis sporulation regulatory protein. There is no evidence for homology in the 5' region of cbgA to lac control regions, i.e. lacOP and the CAP binding sequence. The lack of the CAP binding sequence upstream of the C. acetobutylicum cbgA gene is similar to that reported for S. bovis by Gilbert and Hall (1987). A complete absence of the E. coli CAP binding sequence in the lacOP promoter region of the lac operon in S. bovis was reported. This absence of a CAP binding sequence also associated with the C. acetobutylicum cbgA gene could explain the apparent lack of glucose repression observed in E. coli.

However, sequences homologous to known E. coli and C. pasteurianum ribosome binding sites and promoters were observed in the 5' region of these C. acetobutylicum genes sequenced here. Further work would be required to demonstrate that these sites are involved in transcriptional and translational initiation. The SD sequences detected for the three ORF's resemble those reported for S. aureus (McLaughlin et al.,1981); C. perfringens (Garnier and Cole, 1988); and C. acetobutylicum (Janssen et al., 1988; Zappe et al., 1988). The position of the SD sequences are also in agreement with the previously described positions, which have been found to vary from 4 to 14 nucleotides in distance from the start codon in Gram-positive bacteria (McLaughlin et al., 1981; Garnier and Cole, 1988).

The potential promoter sequences identified upstream of the three ORF's share strong homology to the promoter sequences identified in both Gram-positive (Graves and Rabinowitz, 1986), and Gram-negative (Hawley and McClure, 1983) bacteria. All three ORF's possess extended dA+dT-rich regions upstream of the promoter sites. Fujita et al. (1986) have suggested that such dA+dT-rich regions upstream from Gram-positive and Gram-negative promoters are somehow

involved in the promoter function.

Tn5 mutagenesis had previously revealed a second region (Locus I) within the 5.2 kb EcoRI fragment that affected ß-galactosidase expression in E. coli. Sequence analysis of this region revealed an ORF downstream of cbgA. This ORF of 303 nucleotides would appear to be transcribed in the same direction as cbgA but is separated from the stop codon of cbgA by a 300 bp region that has the potential to form considerable secondary structure. The presence of these repeats together with the fact that the intergenic Tn5 inserts had no effect on the expression on cbgR would suggest that cbgR is transcribed independently of cbgA. The amino acid sequence of this ORF possessed moderate homology to the amino acid sequence of the "0.3 kb gene" from B. subtilis (Stephens et al., 1984).

In B. subtilis, the 0.3 kb gene is involved in gene activation during sporulation. Stephens et al. (1984) investigated the regulation of this gene by lacZ fusion analysis. They showed that the gene had an active promoter and that this promoter was independently regulated, and under developmental control, the activity being detectable during the early phases of sporulation. Therefore, the activity of the ß-galactosidase gene of C. acetobutylicum may be under the developmental regulation of a gene similar to one which is known to regulate the beginning of sporulation in B. subtilis. Previous work (Yu et al., 1987) showed that the expression of the ß-galactosidase was maximum during the switch from acid to solvent production suggesting that it might be developmentally regulated.

The third potential coding region identified in this work (ORF3) was shown to possess homology to spo0A and spo0F from B. subtilis (Trach et al., 1985; Kudoh et al., 1985). In B. subtilis, sporulation is divided into 6 to 7 stages (0 to VII); (Fitz-James and Young,1969; Losick and Youngman, 1984). Growing and vegetative cells are said to be in stage 0. At least 10 stage 0 sporulation (spo0) genes exist, the onset of sporulation being dependent upon the products of the

spo0 loci. Mutations in spo0 genes arrest the onset of sporulation via the interference of a biochemical or morphological event. They exhibit high pleiotrophy, the degree depending on the particular spo0 locus, causing a group of effects (loss of antibiotic production and genetic competence, conference of sensitivity to antibiotics and bacteriophage, and loss of sporulation; Brehm et al., 1975; Linn and Losick, 1976). They have a distinct involvement at the level of gene expression, influencing the expression of a variety of promoters whose products are directly involved in the initiation of the sporulation process (Losick, 1981). Hence, ORF3 may be coding for a clostridial spo0 gene.

However, the homology is highest at the N-terminal region of these two genes, and recently it has been shown that these two genes belong to a family of proteins which comprise a two-component system that senses specific changes in the environment and transduces that information to the transcriptional apparatus (Nixon et al., 1986; Ronson et al., 1987). In this system, one component acts as an environmental sensor (sensor component) that transmits a signal to the second component (regulatory component) which then results in activation of a gene regulon that effects a response. Spo0A and spo0F are transcriptional regulatory components, and, like all regulators, share in common approximately 150 amino acids at the N-terminal end of the protein. The fact that this N-terminal domain is shared by so many other proteins makes it difficult to assign a function to ORF3. However, its location adjacent to the ß-galactosidase gene (cbgA) that is probably regulated by cbgR, a gene homologous to a sporulation regulatory gene from B. subtilis, would suggest that ORF3 may code for a spo0 gene. Sporulation genes in B. subtilis are grouped together in several groups throughout the chromosome (Piggot and Hoch, 1985). Hence, ORF3 probably maintains a regulatory function during the differentiation of C. acetobutylicum through its two phases of growth.

7.3.2 Nucleotide content and codon preference analysis

One of the striking features of the 5.2 kb EcoRI fragment, as is observed in Clostridium in general, is the high dA+dT content (70%). This content is not unexpected, corresponding to the general dA+dT level seen in the genomic DNA of members of the Genus Clostridium (Cato et al., 1986). As seen in Table 7.3, the dA+dT% of the various ß-galactosidase genes resembled closely the overall dA+dT% of the genomic DNA of each specific species.

The differences between the dA+dT content of the coding and non-coding region of the 5.2 kb EcoRI fragment has been reported by other groups. Chen et al. (1986) reported high levels of dA+dT (83%) in the non-coding regions, compared to the 64% found in the coding region of the nifH genes in C. pasteurianum. Garnier and Cole (1988) also noted that there was a higher dA+dT content (82.1%) in the non-coding regions of the 10 kb C. perfringens plasmid pIP404 when compared to that observed in the 10 ORF's identified (72.9%). The same trend has also been reported for another gene sequenced from C. acetobutylicum. Zappe et al. (1988) reported that the ORF of the endo-glucanase gene contained 67.2% dA+dT while the upstream region possessed a dA+dT content of 80.1%.

Although the genetic code provides an array of codons for each amino acid (degenerate), all the codons of a particular amino acid are not used with equal frequency. Rather there is a non-random usage. The codon usage observed is characteristic of the overall base composition for any one species. Therefore, the genes from one species often share similarities in codon usage. However, there can be considerable differences amongst genes ranging from highly biased, to unbiased codon usage (Bibb et al ., 1984; Sharp et al., 1988). This effect was investigated using the different C. acetobutylicum genes available.

There is definite bias in codon usage towards those codons which contain dA+dT

within the three ORF's from C. acetobutylicum (Table 7.4). The codon bias is similar to that reported by Zappe et al. (1988) within the ORF of the endoglucanase gene of C. acetobutylicum, and by Janssen et al. (1988) within the glnA gene from C. acetobutylicum.

Similar bias in codon usage has been reported in other Clostridium species (Table 7.4). Chen et al. (1986) reported a very biased codon usage (towards dA+dT) within the ORF's coding for the nifH1 and nifH2 of C. pasteurianum, with only 38 out of the 61 possible codons being utilized. Garnier and Cole (1988) also noted that, within the 10 ORF's on the C. perfringens plasmid pIP404, there was a pronounced preference for codons containing dA+dT.

The results observed here for the C. acetobutylicum genes sequenced are in agreement with those reported for genes from other Clostridium species. This is not surprising since Clostridium species have a very low dC+dG content (Cato et al., 1986), which significantly influences the codons specified. The bias is, however, totally different to that observed in E. coli which possesses a higher dC+dG% content. There is also less bias observed in the Gram-positive B. subtilis which has a dC+dG content of 41-48% (Table 7.4; Ogasawara, 1985).

The codon usage observed between the ß-galactosidase genes from the different species was also very different, each showing a trend towards the dC+dG content of the species genomic DNA (Table 7.5). In E. coli, the lacZ gene possess no major bias (Kalnins et al., 1983). The bgaB ß-galactosidase gene from B. stearothermophilus has a biased codon usage, biased towards codons containing dA+dT in the third position. The biased codon usage observed with the ß-galactosidase gene from C. acetobutylicum shares more similarity to that observed in B. stearothermophilus, than that observed in E. coli or K. pneumoniae ß-galactosidase genes.

Table 7.4

Codon usage (%) in different ORF's of genes from various bacteria.

Amino acid		C. acetobutylicum A cbgA	B GluA[a]	C glnA[b]	D ORF3	C. per.[c] E	C. pas.[d] F	B. sub.[c] G	E. coli[c] H
Gly	GGG	14	0	3	10	7	1	14	7
	GGA	57	52	72	30	58	58	32	40
	GGT	23	28	16	30	30	39	25	47
	GGG	5	21	9	5	2	30	41	30
Glu	GAG	19	15	9	5	23	6	21	27
	GAA	81	85	91	95	77	94	69	73
Asp	GAT	88	80	87	75	90	72	64	51
	GAC	12	20	13	25	10	28	36	49
Val	GTG	14	0	4	8	3	0	19	27
	GTA	40	42	48	31	50	47	25	23
	GTT	46	50	44	62	45	50	31	38
	GTC	0	8	4	0	2	3	25	13
Ala	GCG	7	0	2	0	3	0	26	31
	GCA	59	21	60	14	58	46	23	27
	GCT	29	72	38	71	33	54	28	28
	GCC	3	7	0	14	7	0	20	19
Lys	AAG	29	25	22	23	19	34	25	23
	AAA	71	75	78	77	80	66	75	77
Asn	AAT	85	87	83	83	87	61	53	24
	AAC	15	13	17	17	13	39	47	76
Met	ATG	100	100	100	100	100	100	100	100
Ile	ATA	41	66	14	33	61	59	11	1
	ATT	51	29	71	51	36	12	50	37
	ATC	8	6	14	15	3	29	39	62
Thr	ACG	7	0	0	0	5	0	28	20
	ACA	51	43	60	13	49	55	43	6
	ACT	38	53	40	88	41	41	15	23
	ACC	4	4	0	0	5	4	14	51

Table 7.4 (cont'd):

Amino acid		C. acetobutylicum A cbgA	B GluA[a]	C glnA[b]	D ORF3	C. per.[c] E	C. pas.[d] F	B. sub.[e] G	E. coli[e] H
Trp	TGG	100	100	100	100	100	100	100	100
Cys	TGT	57	50	83	-	88	92	46	42
	TGC	43	50	17	-	12	8	54	58
Tyr	TAT	85	73	76	92	94	77	62	41
	TAC	15	27	24	8	6	23	36	59
Phe	TTT	79	94	76	88	83	27	64	44
	TTC	21	6	24	12	17	73	36	57
Ser	TCG	0	2	7	0	1	4	10	11
	TCA	44	37	29	24	30	44	19	8
	TCT	22	27	43	28	18	0	25	27
	TCC	4	2	0	3	3	11	12	26
	AGT	18	24	21	28	42	28	11	7
	AGC	13	8	0	17	7	12	24	22
Arg	AGG	5	0	0	0	14	0	9	1
	AGA	69	60	100	43	80	100	28	1
	CGG	0	0	0	14	1	0	11	3
	CGA	10	0	0	0	2	0	9	2
	CGT	13	40	0	0	3	0	25	58
	CGC	3	0	0	43	0	0	18	35
Gln	CAG	34	35	0	25	10	29	49	73
	GAA	66	65	100	75	90	71	54	27
His	CAT	61	83	100	100	75	75	69	39
	CAC	39	17	0	0	25	25	31	61
Leu	TTG	14	14	3	15	6	1	14	8
	TTA	44	55	82	44	71	47	22	6
	CTG	3	5	3	7	1	0	22	69
	CTA	13	0	10	7	9	10	22	69
	CTT	25	18	3	7	12	35	26	9
	CTC	2	9	0	7	0	0	10	7

Table 7.4 (cont'd):

Amino acid	C. acetobutylicum A cbgA	B GluA[a]	C glnA[b]	D ORF3	C. per.[c] E	C. pas.[d] F	B. sub.[e] G	E. coli[e] H
Pro CCG	15	0	0	0	5	0	38	65
CCA	55	40	85	0	56	78	19	20
CCT	30	60	15	80	40	22	34	9
CCG	0	0	0	20	0	0	10	6

[a] Zappe et al. (1988) [b] Janssen et al. (1988)
[c] Garnier and Cole (1988) [d] Chen et al. (1986)
[e] Ogasawara (1985)

Table 7.5

Codon usage (%) in different ORF's of ß-galactosidase genes from various bacteria.

Amino acid	A cbgA	B bgaB[a]	C lacZ[b]	D lacZ[c]	E ebgA[d]
Gly GGG	14	28	13	19	12
GGA	57	28	6	3	12
GGT	23	30	34	18	28
GGG	5	14	48	60	48
Glu GAG	19	31	24	58	28
GAA	81	69	76	42	72
Asp GAT	88	80	66	46	52
GAC	12	20	34	54	48
Val GTG	14	18	36	47	45
GTA	40	33	16	8	3
GTT	46	33	19	9	19
GTC	0	16	30	36	33

Table 7.5 (cont'd):

Amino acid		A cbgA	B bgaB[a]	C lacZ[b]	D lacZ[c]	E ebgA[d]
Ala	GCG	0	19	36	48	34
	GCA	59	36	16	4	12
	GCT	29	36	12	1	18
	GCC	3	8	36	48	36
Lys	AAG	29	47	25	45	29
	AAA	71	53	75	55	71
Asn	AAT	85	68	36	14	23
	AAC	15	32	64	86	77
Met	ATG	100	100	100	100	100
Ile	ATA	41	17	0	3	2
	ATT	51	67	56	31	34
	ATC	8	17	44	67	64
Thr	ACG	7	20	36	27	31
	ACA	51	40	15	2	5
	ACT	38	36	7	9	15
	ACC	4	4	42	62	49
Trp	TGG	100	100	100	100	100
Cys	TGT	57	90	31	25	30
	TGC	43	10	69	75	70
Tyr	TAT	85	71	42	41	51
	TAC	15	29	58	59	49
Phe	TTT	79	69	50	42	46
	TTC	21	31	50	58	54
Ser	TCG	0	3	15	19	14
	TCA	44	24	13	5	8
	TCT	22	17	5	2	6
	TCC	4	14	10	14	26
	AGT	18	34	15	5	8
	AGC	13	7	42	55	38

Table 7.5 (cont'd):

Amino acid		A cbgA	B bgaB[a]	C lacZ[b]	D lacZ[c]	E ebgA[d]
Arg	AGG	5	16	0	1	2
	AGA	69	38	0	1	2
	CGG	0	5	11	19	7
	CGA	10	11	5	1	4
	CGT	13	27	30	13	29
	CGC	3	3	55	64	57
Gln	CAG	34	38	74	85	84
	GAA	66	62	26	15	16
His	CAT	61	68	56	31	42
	CAC	39	32	44	69	58
Leu	TTG	14	23	11	7	15
	TTA	44	35	7	1	2
	CTG	3	8	56	65	62
	CTA	13	17	6	4	5
	CTT	25	12	9	9	8
	CTC	2	6	9	14	8
Pro	CCG	15	19	58	56	58
	CCA	55	49	10	10	16
	CCT	30	22	18	10	12
	CCG	0	11	15	24	14

[a] Hirata et al. (1986)

[b] Kalnins et al. (1983)

[c] Buvinger and Riley (1985)

[d] Stokes et al. (1985)

This biased codon usage has important implications for gene expression in a heterologous host. Garnier and Cole (1988) noted that only one of the ten ORF's on pIP404 from C. perfringens was expressed in E. coli at any significant level. Since heterologous bacterial transcription signals are usually recognized in E. coli, they therefore should not cause a barrier to gene expression. Garnier and

Cole (1988) suggested that base composition may be influencing the expression due to codon usage, causing a major obstacle to translation. As observed, the codons preferably used by Clostridium species are rarely, if ever, used in E. coli (Ikemura, 1981; Grosjean and Fiers, 1982). This provides a possible explanation for the low ß-galactosidase activity observed in E. coli (Chapter 5). Similar low levels are evident for various other Gram-positive genes when plasmid copy number are considered (Lee et al., 1982; Efstathiou and Truffaut, 1986).

7.3.3 Amino acid content

A relationship between DNA base and protein composition has been proposed by Elton (1973). The theoretical limits for a DNA sequence coding for a protein were set at 27 to 71% dC+dG. At the extremes of these theoretical limits, codons and hence amino acids often show extreme biases. In organisms rich in dA+dT (such as Clostridium) high levels of the amino acids phe, ile, tyr, asn, and lys are expected to be found (Table 7.6). This theory, when tested against the amino acid content of the ß-galactosidase gene of C. acetobutylicum supports the hypothesis of Elton (1973), the above amino acids being more abundant in this gene than that observed in the other ß-galactosidase genes from the Gram-negative bacteria (Table 7.6). This is also observed in ORF3. There is also corresponding scarcity of those amino acids which would be expected to be abundant in an dC+dG-rich organism (pro, trp, gly, arg, and ala; Table 7.6). Incorporation of these amino acids may be rate-limiting during translation. The endo-glucanase and glnA genes of C. acetobutylicum (Zappe et al., 1988; Janssen et al., 1988) also share higher levels of the expected amino acids. This correlates well with the higher levels of dA+dT in the codons representing these amino acids. Similar results have been reported by Garnier and Cole (1988) using the complete sequences of the ten ORF's on pIP404 from C. perfringens.

Table 7.6. Analysis of the amino acid content of various genes.

Group	acid	pIP404[a]	glnA[b]	gluA[c]	cbgA[d]	ORF3[d]	bgaB[e]	lacZ[f]	lacZ[g]
1	Met	1.8	2.9	2.5	2.4	2.1	2.2	1.9	2.2
	Leu	9.4	8.8	4.9	7.1	9.8	7.7	9.7	9.4
	*Ile	9.7	4.7	7.8	6.5	14.2	7.2	3.8	3.8
	Val	5.7	6.1	5.3	5.6	4.7	7.3	6.4	6.3
	Total	26.6	29.7	20.5	21.7	30.8	24.4	21.8	21.7
2	=Gly	6.3	7.2	6.5	6.2	3.6	6.4	7.1	6.9
	=Ala	4.7	9.0	6.5	3.7	2.6	5.3	7.9	7.5
	=Pro	2.4	4.5	3.3	3.7	1.8	5.5	7.0	6.0
	Ser	6.9	3.4	11.4	6.1	10.5	4.4	5.6	5.9
	Thr	4.3	4.5	10.9	5.0	2.9	3.7	5.3	5.4
	Total	24.6	28.5	38.6	24.7	21.4	25.3	32.9	31.7
3	His	0.9	1.8	1.3	2.0	1.1	2.8	3.5	3.3
	*Lys	10.1	7.2	7.2	8.0	8.0	5.6	1.1	1.9
	=Arg	3.7	4.3	1.1	4.3	2.5	5.5	8.0	6.5
	Total	14.7	13.3	9.6	14.3	11.6	13.9	12.6	11.7
4	Asp	5.3	6.8	5.6	6.5	5.8	5.9	6.6	6.3
	Glu	7.1	7.7	2.9	8.7	7.6	7.5	5.5	6.1
	*Asn	7.5	7.9	6.9	6.0	8.3	5.6	3.6	5.0
	Gln	2.6	2.5	3.8	3.3	2.9	3.1	5.2	5.6
	Total	22.5	24.9	19.2	24.4	24.6	22.1	20.9	22.6
5	*Phe	4.3	4.7	4.0	4.8	6.1	4.4	3.0	3.6
	*Tyr	5.3	3.8	4.9	6.7	4.4	5.2	2.6	3.0
	=Trp	0.9	0.9	2.3	2.3	0.7	3.1	4.3	3.8
	Total	10.5	9.4	11.2	13.8	11.2	12.7	9.9	10.4
6	Cys	1.0	1.4	0.9	0.9	0.0	1.6	1.9	1.6

C. perfringens[a]: Garnier and Cole (1988)

C. acetobutylicum[b]: Janssen et al. (1988)

C. acetobutylicum[c]:Zappe et al. (1988)

C. acetobutylicum[d]: this study

B. stearothermophilus[e]: Hirata et al. (1986)

K. pneumoniae[f]: (Stokes et al. (1985)

E. coli[g]: (Kalnins et al. (1983)

(*) Amino acids coded for by A+T-rich codons

(=) Amino acids coded for by G+C-rich codons

7.4 SUMMARY

The 5.2 kb EcoRI fragment known to contain a ß-galactosidase gene from C. acetobutylicum NCIB2951 was sequenced. Analysis of the sequence enabled identification of three ORF's. A long ORF (cbgA; structural ß-galactosidase) was 2691 nucleotides in length (897 amino acids long) and coded for an unmodified protein of 105 kDa. While no significant homology was detected at the DNA level, extensive homology was detected at the amino acid level to the sequenced ß-galactosidase genes from E. coli and K. pneumoniae. The second region (Locus I) identified by Tn5 mutagenesis to be involved in the expression of the ß-galactosidase gene was found to contain a small ORF (303 bp long) coding for a protein 101 amino acids long with a predicted Mr of 11 kDa. Moderate amino acid homology was detected to the "0.3 kb gene" from B. subtilis. The third potential coding region identified extends beyond the region sequenced. The ORF was 825 nucleotides in length, and coded for a truncated protein of 275 amino acids long with a calculated Mr of 32 kDa. The amino acid sequence of this ORF possessed significant homology to the stage 0 spo0A and spo0F sporulation proteins from B. subtilis. The classical lac operon system does not exist in C. acetobutylicum. Rather, a single monocistronic ß-galactosidase (cbgA) gene exists under the control of cbgR.

CHAPTER 8: FINAL DISCUSSION AND CONCLUSIONS

The aim of this project was to isolate and characterise genes involved in the early steps of lactose utilization in C. acetobutylicum. Discussion and summary sections have been reported for each chapter, so the information contained in those chapters will not be repeated here. Instead, this discussion will examine possible models for the induction and regulation of the two lactose utilizing systems that operate in C. acetobutylicum.

As described in Chapter 1, there are two major lactose utilization systems known in bacteria. In Gram-negative bacteria, such as E. coli the common system is the lac operon which involves transport of the lactose into the cell by a lactose permease and utilization by a ß-galactosidase (Dills et al., 1980). In Gram-positive bacteria, the major lactose utilization system identified involves the PEP:PTS system for lactose transport into the cell, and phospho-ß-galactosidase for hydrolysis of the phosphorylated disaccharide (Postma and Lengeler, 1985). However, some Gram-positive bacteria utilize lactose by way of the lactose permease and ß-galactosidase pathway rather than the PEP:PTS system. The presence of both phospho-ß-galactosidase and ß-galactosidase in C. acetobutylicum would suggest that both systems operate in this organism. Yu et al. (1987) have shown differential induction of ß-galactosidase and phospho-ß-galactosidase in C. acetobutylicum, with high levels of phospho-ß-galactosidase during the acidogenic phase followed by high levels of ß-galactosidase during the solventogenic phase. Conditions which bring about the switch from phospho-ß-galactosidase activity to ß-galactosidase activity are not known, and the demonstration that these two genes are induced at the transcriptional level still needs to be confirmed.

While I failed to isolate the phospho-ß-galactosidase gene I did succeed in isolating the ß-galactosidase gene. I have shown that in E. coli ß-galactosidase (Locus II: cbgA) requires Locus I (cbgR) for expression. DNA sequencing identified a small ORF (cbgR) which shared homology to the 0.3 kb spo regulatory gene (Stephens

et al., 1984). It would appear that this gene serves as a positive activator for the developmental expression of the cbgA. This is in contrast to the known system of regulation of the lacZ gene within the lac operon system of E. coli, which is both positively and negatively regulated by lacI, CAP and cAMP (Loomis and Magasanik, 1964; Perlman and Pastan, 1968; Goeddel et al., 1978).

During the acidogenic phase only phospho-ß-galactosidase can be detected suggesting that at this stage the other pathway is shut down. As growth proceeds, phospho-ß-galactosidase is induced, rises to a plateau, and then declines. Then as the cell switches from acidogenic to solventogenic phase, ß-galactosidase activity rises separately indicating that both genes are very tightly controlled during the growth cycle of C. acetobutylicum (Yu et al., 1987).

Utilization of lactose by phospho-ß-galactosidase during the acidogenic phase presumably requires transport by way of a PEP:PTS system. In species of Enterobacteriaceae, two methods are known whereby enzymes from several PEP:PTS systems play a central role in the regulation of non-PTS transport systems (e.g. lactose permease) and catabolic pathways (Postma and Lengeler, 1987). EnzymeIII^{glucose} (EnzIII^{glu}) is the best known enzyme, inhibiting a number of non-PTS uptake systems, including the lactose operon system. The level of EnzIII^{glu} can also be controlled indirectly by other PEP:PTS systems. The dephosphorylation of EnzIII^{glu} occurs due to the utilization of P-HPr by an EnzIII^{substrate} (such as EnzIII^{lactose}. This reverses the equilibrium between HPr and EnzIII^{glu}, resulting in high levels of non-phosphorylated EnzIII^{glu} which inhibit other non-PTS system such as lactose (Nelson et al., 1983; Postma and Lengeler, 1985).

The first method involves the P-EnzIII^{glu} as a regulator of adenyl cyclase expression. In the phosphorylated form, EnzIII^{glu} activates adenylate cyclase, which in turn controls the level of transcription of non-PTS systems by modulating the level of intracellular cAMP and CAP. The presence of non-phosphorylated EnzIII^{glu} inhibits the non-PTS systems by the low level of cAMP and CAP. Direct coupling between

CAP and the PTS system can also occur, modulating transcription without the involvement of cAMP (Saier and Feucht, 1975; Postma and Lengeler, 1985). In C. acetobutylicum, no E. coli-like CAP consensus sequence is present upstream of the cbgA gene, and there is no evidence to date demonstrating a role for cAMP in catabolite repression in Gram-positive bacteria. The existence of cAMP in Gram-positive bacteria has also been questioned (Botsford, 1981; Lee et al., 1982; Gilbert and Hall, 1987). Hence, this system probably does not operate in C. acetobutylicum.

The second method, known as inducer exclusion, is mediated through non-PTS uptake systems (including lactose permease) by the direct binding of $EnzIII^{glu}$ to the permease in the cell membrane, thus preventing the transportation of lactose into the cell and in turn preventing induction of the ß-galactosidase (Nelson et al., 1983; Postma and Lengeler, 1985). When the levels of $EnzIII^{glu}$ drop, lactose can then be transported by lactose permease and would then allow induction of ß-galactosidase (Saier and Roseman, 1976; Mitchell et al., 1982). The existence of such a regulation system has been observed for the non-PTS glycerol uptake system in B. subtilis (Reizer et al., 1984). The possibility exists therefore, that when C. acetobutylicum is utilizing lactose as the sole carbon source, the $EnzIII^{lactose}$ of the PEP:PTS lactose pathway exerts the same inhibitory effect on the non-PTS lactose pathway (lactose permease) during the acidogenic phase, this inhibitory effect declining as the cell proceeds into the solventogenic phase. However, the presence of a lactose permease in C. acetobutylicum still remains to be demonstrated.

The demonstration that ß-galactosidase expression requires expression of a second gene (cbgR) suggests that induction of ß-galactosidase is linked to some global signalling as the cell switches from acid to solvent production. The identification here of an ORF (ORF3) homologous to the regulatory component of a two-component regulation system adjacent to cbgA is therefore interesting. However, it still remains to be demonstrated whether there is any control exerted by this gene on the expression of ß-galactosidase in C. acetobutylicum. In E. coli it appears to

be unnecessary. The observed pattern of protein sequence homology observed between the amino and carboxyl ends of proteins involved in this two-component regulation system suggest that during prokaryotic evolution, a two-component system has continually been adapted for sensing specific changes in the environment, and transducing that information to transcriptional apparatus (Ronson et al., 1987). This system is responsible for the regulation of nitrogen assimilation in K. pneumoniae, and initiation of sporulation in B. subtilis (Nixon et al., 1986). This protein, ORF3, could serve as an environmental sensor for initiating transcription of sporulation and solvent genes.

The switch from acidogenic to the solventogenic phase is known to be accompanied by a corresponding shift in the levels of activities of a large number of other enzymes involved in these pathways. Many acidogenic phase enzymes show a coordinated decrease, while enzymes involved in solvent production show a coordinate activation or increase in activity (Andersch et al., 1983; Hartmanis et al., 1984; Hartmanis and Gatenbeck, 1984). These results suggest that a common regulatory signal such as observed in the two-component systems described to date, may be involved in the induction of groups of enzymes functioning during the solventogenic phase (Jones and Woods, 1986). In C. acetobutylicum the two lactose utilization systems may hence be regulated by the two-component system via growth parameters, causing activation of the ß-galactosidase system during the solventogenic phase, with a coincident decrease in the PEP:PTS system. The nature of this signal needs to be determined.

However, the protein from ORF3 is not regulating ß-galactosidase expression in E. coli as Tn5 inserts in this region had no effect of ß-galactosidase expression. In any case, the incomplete ORF would produce a truncated protein which presumably would be inactive. In contrast, regulation of ß-galactosidase activity in C. acetobutylicum is more complex, i.e. although ß-galactosidase activity is detectable in the presence of glucose in E. coli, there is no ß-galactosidase activity detectable in C. acetobutylicum (Yu et al., 1987). The complete protein from ORF3 may

therefore play an important role during lactose utilization in C. acetobutylicum where all the components of the regulatory control are present.

Further work needs to be carried out, such as the completion of the gene sequence of ORF3, identification of the function of the protein from ORF3, and isolation of the environmental sensor which activates the regulator (ORF3). Analysis of mutations in these genes would also be useful, allowing identification of their effects on the normal growth cycle of C. acetobutylicum. While gene transfer in C. acetobutylicum is possible, there is no report of the reintroduction of a native known gene back into C. acetobutylicum. Also, no technique has been developed to carry out gene marker exchange in this species, which would greatly aid the analysis of the role of genes in vivo.

In conclusion, this project has provided a detailed molecular and genetical analysis of a small section of the lactose utilization pathways present in C. acetobutylicum; namely the analysis of the ß-galactosidase gene. This system is very different from the known lac operon system previously studied in other bacteria. It has also enabled the identification of a potential two-component regulation system in C. acetobutylicum, similar to that observed in other Gram-positive and Gram-negative bacteria. The existence of this system may be very important in the regulation of the two-stage growth cycle. Further studies need to be undertaken to resolve the molecular mechanisms which control and regulate lactose uptake, utilization and solvent production in C. acetobutylicum. Such knowledge could allow the genetic manipulation of these pathways enabling an increase in the efficiency of solvent production when C. acetobutylicum is grown in the presence of lactose (whey permeate).

BIBLIOGRAPHY

Abou-Zeid, A.A., M. Fouad and M. Yassein. (1978). Microbiological production of acetone-butanol by Clostridium acetobutylicum. Zentalbl. Bakteriol. Parasitenkd. Infektionskr. Ayg. Abt. 133: 125-134.

Akrigg, A. (1978). Purification and properties of a manganese stimulated deoxyribonuclease produced during sporulation of Bacillus subtilis. Biochem. J. 172: 69-76.

Akrigg, A., J. Mandelstam. (1977). Repair of hydrogen peroxide-induced single strand breaks in Escherichia coli deoxyribonucleic acid. J. Bacteriol. 130: 187-191.

Allcock, E.R., S.J. Reid, D.T. Jones, D.R. Woods. (1981). Autolytic activity and an autolytic-deficient mutant of Clostridium acetobutylicum. Appl. Environ. Microbiol. 42: 929-935.

Allcock, E.R., S.J. Reid, D.T. Jones, D.R. Woods. (1982). Clostridium acetobutylicum protoplast formation and regeneration. Appl. Environ. Microbiol. 43: 719-721.

Ananthaswamy, H.N., A. Eisenstark. (1977). Repair of hydrogen peroxide-induced single-stranded breaks in Escherichia coli deoxyribonucleic acid. J. Bacteriol. 130: 187-191.

Andersch, W., H. Bahl, G. Gottschalk. (1982). Acetone-butanol production by Clostridium acetobutylicum in an ammonium-linked chemostat at low pH values. Biotechnol. Lett. 4: 29-32.

Andersch, W., H. Bahl, G. Gottschalk. (1983). Level of enzymes involved in acetate, butyrate, acetone and butanol formation by Clostridium acetobutylicum. Eur. J. Appl. Microbiol. Biotechnol. 18: 327-332.

Anderson, D.G., L.L. McKay. (1983). Simple and rapid method for isolating large plasmid DNA from lactic streptococci. Appl. Environ. Microbiol. 46: 549-552.

Anderson, S., M.J. Gait, L. Mayol, I.G. Young. (1980). A short primer for sequencing DNA cloned in the single-stranded phage vector M13mp2. Nucl. Acids. Res. 8: 1731.

Andrews, K.J., E.C.C. Lin. (1976). Thiogalactoside transacetylase of the lactose operon as an enzyme for detoxification. J. Bacteriol. 128: 510-513.

Aronson, A., P. Fitz-James. (1976). Structure and morphogenesis of the bacterial spore coat. Bacterial Rev. 40: 360-402.

Auerswald, E-A., G. Ludwig, H. Schaller. (1981). Structural analysis of Tn5. Cold Spring Harbour Symp. Quant. Biol. 45: 107-113.

Baghdaf, A.O., N.A. Abou-Zeid, M. Yassein. (1980). Influence of nitrogen and phosphorous on the fermentative production of acetone and butanol by Clostridium acetobutylicum. Zentralblatt fur Bacteriologie Parasitenkunde Infektios md Hygeine Abteileing. 135: 515-522.

Bahl, H., W. Andersch, K. Braun, and G. Gottschalk. (1982a). Effect of pH and butyrate concentration on the production of acetone and butanol by Clostridium acetobutylicum grown in continuous culture. Eur. J. Appl. Microbiol. Biotechnol. 14: 17-20.

Bahl, H., W. Andersch, G. Gottschalk. (1982b). Continuous production of acetone and butanol by Clostridium acetobutylicum in a two-stage phosphate limited chemostat. Eur. J. Appl. Microbiol. Biotech. 15: 201-205.

Bahl, H., and G. Gottschalk. (1985). Parameters affecting solvent production by Clostridium acetobutylicum in continuous culture. Biotechnol. Bioeng. 514: 217-223.

Barber, J.M., F.T. Robb, J.R. Webster, D.R. Woods. (1979). Bacteriocin production by Clostridium acetobutylicum in an industrial fermentation process. Appl. Environ. Microbiol. 37: 433-437.

Bates, P.F., R.A. Swift. (1982). Double cos vectors: simplified cosmid cloning. Gene 26: 137-146.

Beckwith, J., T. Grodzicker, R. Arditti. (1972). Evidence for two sites in the lac promoter region. J. Mol. Biol. 69: 155-160.

Beesch, S.C. (1952). Acetone-butanol fermentation of sugars. Eng. Proc. Dev. 44: 1677-1682.

Beesch, S.C. (1953). A microbiological process report: acetone:butanol fermentation of starches. Appl. Microbiol. 1: 85-96.

Benton, W.D., R.W. Davis. (1977). Screening recombinant clones by hybridization to single plaques. Science 196: 180-182.

Berg, D.E., C.M. Berg. (1983). The prokaryotic transposable element Tn5. Biotechnol. 1: 417-435.

Berg, D.E., C. Egner, J.B. Lowe. (1983). Mechanism of F-factor enhanced excision of transposon Tn5. Gene 22: 1-7.

Berg, D.E., J. Davies, B. Allet, J.D. Rochaix. (1975). Transposition of R factor genes to bacteriophage lambda. Proc. Natl. Acad. Sci. USA 72: 3628-3632.

Berg, D.E., A. Weiss, L. Crossland. (1980). Polarity of Tn5 insertion in Escherichia coli. J. Bacteriol. 142: 439-446.

Berg, D.E., L. Johnsrud, L. McDivitt, R. Ramabhadran, B.J. Hirschel. (1982). Inverted repeats of Tn5 are transposable elements. Proc. Natl. Acad. Sci. USA 79: 2632-2635.

Bergquist, P.L., A.E. Adelberg. (1972). Abnormal excision and transfer of chromosomal segments by a strain of Escherichia coli K12. J. Bacteriol. 111: 119-128.

Bernier Jr. R., H. Driguez, M. Desrochers. (1983). Molecular cloning of a Bacillus subtilis xylanase gene in Escherichia coli. Gene 26: 59-65.

Bibb, M.J., P.R. Findlay, M.W. Johnson. (1984). The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein coding sequences. Gene 30: 157-166.

Biggin, M.O., J.J. Gibson, G.I. Hong. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci., USA 80: 3963-3965.

Bilofsky, H.S., C. Burks. (1988). The Genbank^{R} genetic sequence data bank. Nucl. Acids Res. 16: 1861-1863.

Bissett, D.L., R.L. Anderson. (1973). Lactose and D-galactose metabolism in Staphylococcus aureus: pathway of D-galactose 6-phosphate degradation. Biochem. Biophys. Res. Commun. 52: 641-647.

Blaschek, H.P. (1986). Genetic manipulation of Clostridium acetobutylicum for production of butanol. Food Technol. Oct. pp.84-87.

Blaschek, H.P., M.A. Klacik. (1984). Role of DNase in recovery of plasmid DNA from Clostridium perfringens. Appl. Environ. Microbiol. 48: 178-181.

Blaschek, H.P., M. Solberg. (1981). Isolation of a plasmid responsible for caseinase activity in Clostridium perfringens ATCC 3626B. J. Bacteriol. 147: 262-266.

Blumenthal, H.J. (1972). Glucose catabolism in Staphylococci, p.111-135. In J.O. Cohen (ed.). The Staphylococci. John Wiley and Sons. Inc., New York.

Boehringer Mannheim. (1987). Biochemicals for Molecular Biology. p.118, Boehringer Mannheim GmbH, Biochemica. West Germany.

Boizet, B., D. Villeval, P. Slos, M. Novel, G.N. Mercenier. (1988). Isolation and structural analysis of the phospho-ß-galactosidase gene from Streptococcus lactis. Gene 62: 249-261.

Bolivar, F., R.L. Rodriguez, P.J. Greene, M.C. Betlach, H.L. Heyneker, H.W. Boyer, J. Crosa, S. Falkow. (1977). Construction and characterisation of new cloning vehicles. II. A multi-purpose cloning system. Gene 2: 95-113.

Bossi, L., M.S. Ciampi. (1981). DNA sequence of the sites of three insertions of the transposable element Tn5 into the histidine operon of Salmonella. Mol. Gen. Genet. 183: 406-408.

Bowring, S.N., J.G. Morris. (1985). Mutagenesis of Clostridium acetobutylicum. J. Appl. Bacteriol. 58: 577-584.

Boyer, H.W., D. Roulland-Dussoix. (1969). A complementation analysis of the restriction and modification of DNA in Escherichia coli. J. Mol. Biol. 41: 459-472.

Botsford, J. (1981). Cyclic nucleotides in prokaryotes. Microbiol. Rev. 45: 620-642.

Brawn, K., I. Fridovich. (1981). DNA strand scission by enzymically generated oxygen radicals. Arch. Biochem. Biophys. 206: 414-419.

Brefort, G., M. Magot, H. Ionesco, M. Sebald. (1977). Characterization and transferability of Clostridium perfringens plasmids. Plasmid 1: 52-66.

Brehm, S.P., F. Le Hegarat, J.A. Hoch. (1975). Deoxyribonucleic acid-binding proteins in vegetative Bacillus subtilis: alterations caused by stage 0 sporulation mutations. J. Bacteriol. 124: 977-984.

Breidt, F., G.C. Stewart. (1986). Cloning and expression of the phospho-ß-galactosidase gene of Staphylococcus aureus in Escherichia coli. J. Bacteriol. 166: 1061-1066.

Brenner, D.J., S. Falkow. (1971). Molecular relationships among members of the Enterobacteriaceae. Ad. Genet. 16 81-118.

Bu'Lock, J.D., A.J. Bu'Lock. (1983). The AB fermentation and related topics 1980-1983. Sci. Technol. Lett. Kew, England.

Buvinger, W.E., M. Riley. (1985). Nucleotide sequence of Klebsiella pneumoniae lac genes. J. Bacteriol. 163: 850-857.

Cameron, G.N. (1988). The EMBL data library. Nucl. Acids Res. 16: 1865-1867.

Campbell, J.H., J.A. Lengyel, J. Langridge. (1973). Evolution of a second gene for ß-galactosidase in Escherichia coli. Proc. Natl. Acad. Sci. USA 70: 1841-1845.

Casida, L.E. (1968).Industrial Microbiology. Pub. John Wiley and Sons, Inc. New York, London and Sydney.

Cato, E.P., W.L. George, S.M.Finegold. (1986). Genus Clostridium. In P.H.A. Sneath, N.S. Mair, M.C. Sharpe, J.G. Holt (eds.). Bergeys manual of systemic bacteriology, pp.1141-1200. 9th ed. Vol.2. Williams and Wilkins Co., USA.

Chassy, B.M., E.V. Gibson, A. Gluffrida. (1978). Evidence for plasmid associated lactose metabolism in Lactobacillus casei subsp. casei. Curr. Microbiol. 1: 141-144.

Chassy, B.M., J. Thompson. (1983). Regulation of lactose-phosphoenol pyruvate-dependent phosphotransferase system and ß-D-phosphogalactoside galactohydrolase activities in Lactobacillus casei. J. Bacteriol. 154: 1195-1203.

Chen, K.C-K., J-S. Chen, J.L. Johnson. (1986). Structural features of multiple nifH-like sequences and very biased codon usage in nitrogenase genes of Clostridium pasteurianum. J. Bacteriol. 166: 162-172.

Cho, K.Y., C.H. Doy. (1973). Ultrastructure of the obligately anaerobic bacteria Clostridium kluyveri and Clostridium acetobutylicum. Aust. J.Biol. Sci. 26: 547-558.

Citti, J.E., W.E. Sandine, P.R. Elliker. (1965). ß-galactosidase of Streptococcus lactis . J. Bacteriol. 89: 937-942.

Clark, L., J. Carbon. (1979). Selection of specific clones from colony banks by suppression or complementation tests. Methods in Enzymology 68: 396-408. R. Wu (ed.), Academic Press, New York.

Clark, D.P., J.E. Cronan. (1980).Acetaldehyde coenzyme A dehydrogenase of Escherichia coli. J. Bacteriol. 144: 179-184.

Claus, D., R.C.W. Berkeley. (1986). Genus Bacillus. In P.H.A. Sneath, N.S. Mair, M.C. Sharpe, J.G. Holt (eds.). Bergeys manual of systemic bacteriology, pp.1105-1139. 9th ed. Vol.2. Williams and Wilkins Co., USA.

Clewell, D.B., D.R. Helinski. (1969). Supercoiled DNA-protein complex in Escherichia coli: purification and induced conversion to an open circular form. Proc. Natl. Acad. Sci. USA 62: 1159-1166.

Clewell, D.B., Y. Yagi, G.M. Dunny, S.K. Schultz. (1974). Characterization of three plasmid deoxyribonucleic acid molecules in a strain of Streptococcus faecalis: identification of a plasmid determining Em^R. J. Bacteriol. 117: 283-289

Cohen, S.N., A.C.Y. Chang, L. Hsu. (1972). Non-chromosomal antibiotic resistance in bacteria: genetic transformation of Escherichia coli by R-factor DNA. Proc. Natl. Acad. Sci. USA 69: 2110-2114.

Collins, M.E., B. Hohn. (1978). Cosmids: a type of plasmid gene-cloning vector that is packagable in vitro in bacteriophage heads. Proc. Natl. Acad. Sci. USA 75: 4242-4246.

Collins, M.E., J.D. Oultram, M. Young. (1985). Identification of restriction fragments from two cryptic Clostridium butyricum plasmids that promote the establishment of a replication-defective plasmid in Bacillus subtilis. J. Gen. Microbiol. 131: 2097-2105.

Compere, A.L., W.L. Griffith. (1979). Evaluation of substrates for butanol production. Dev. Ind. Microbiol. 20: 509-517.

Conway, T., C.W. Sewell, Y.A. Osman, L.O. Ingram. (1987). Cloning and sequencing of the alcohol dehydrogenase II gene from Zygomonas mobilis. J. Bacteriol. 769: 2591-2597.

Cords, B.R., L.L. McKay, P. Guerry. (1974). Extrachromosomal elements in Group N streptococci. J. Bacteriol. 117: 1149-1152.

Cornelis, G., D. Ghosal, H. Saedler. (1978). Tn951: a new transposon carrying a lactose operon. Mol. Gen. Genet. 160: 215-224.

Cornet, P., J. Millet, P. Beguin, J.P. Aubert. (1983a). Characterisation of two cel (cellulose degradation) genes of Clostridium thermocellum coding for endoglucanases. Biotechnol. 1: 589-594.

Cornet, P., D. Tronik, J. Millet, J.P. Aubert. (1983b). Cloning and expression in E. coli of Clostridium thermocellum genes coding for amino acid synthesis and cellulose hydrolysis. FEMS Microbiol. Lett. 16: 137-141.

Cota-Robles, E.H., S.M. Stein. (1973). Bacterial cell breakage and lysis. p.833-843. In Microbial composition A.I. Laskin, H.A. Lechevalie (ed.), CRC Handbook of Microbiology. CRC Press.

Craven, G.R., E. Steers Jr., C.B. Anfinsen. (1965). Purification, composition, and molecular weight of the ß-galactosidase of Escherichia coli K-12. J. Bio. Chem. 240: 2468-2477.

Crow, V.L., G.P. Davey, L.E. Pearce, T.D. Thomas. (1983). Plasmid linkage of the D-tagatose-6-phosphate pathway in Streptococcus lactis. Effect on lactose and galactose metabolism. J. Bacteriol. 153: 76-83.

Crow, V.L., T.D. Thomas. (1984). Properties of a Streptococcus lactis strain that ferments lactose slowly. J. Bacteriol. 157: 28-34.

Cummins, C.S., J.L. Johnson. (1971). Taxonomy of the Clostridia: wall composition, DNA homologies in Clostridium butyricum and other butyric acid-producing Clostridia. J. Gen. Microbiol. 67: 33-46.

Cummins, C.S. (1973). Bacterial cell wall structure. In CRC Handbook of Microbiology Vol. II: Microbial Composition. A.L. Laskin, H.A. Lechevalier (ed.).

Cynkin, M.A., E.A. Delwiche. (1958). Metabolism of pentoses by clostridia. I. Enzymes of ribose dissimilation in extracts of Clostridium perfringens. J. Bacteriol. 75: 331-334.

Cynkin, M.A., M. Gibbs. (1958). Metabolism of pentoses by clostridia. II. The fermentation of C^{14}-labelled pentoses by Clostridium perfringens, Clostridium beijerincki, and Clostridium butylicum. J. Bacteriol. 75: 335-338.

Dagert, M., S.D. Ehrlick. (1979). Prolonged incubation in calcium chloride improves the competence of Escherichia coli. Gene 6: 23-28.

Daldal, F., and J. Applebaum. (1985). Cloning and expression of Clostridium pasteurianum galactokinase gene in Escherichia coli K-12 and nucleotide sequence analysis of a region affecting the amount of enzyme. J. Mol. Biol. 186: 533-545.

Davies, I.W., D. Kay, J. Mandelstam. (1969).Sporulation in Bacillus subtilis. Establishment of a time scale for the morphological events. J. Gen. Micro. 56: 171-179.

Davies, R. (1943). Studies on the acetone-butanol fermentation: Acetoacetate decarboxylase of Clostridium acetobutylicum (BY). Biochem. J. 37: 230-238.

Dawes, R. and M. Stephenson. (1941). Studies on the acetone-butyl alcohol fermentation. I. Nutritional and other factors involved in the preparation of active suspensions of Clostridium acetobutylicum (Weizmann). Biochem. J. 35: 1320-1331.

De Bruijn, F.J., F.M. Ausubel. (1981). The cloning and transposon Tn5 mutagenesis of the glnA region of Klebsiella pneumoniae: Identification of glnR, a gene involved in the regulation of the nif and hut operons. Mol. Gen. Genet. 183: 289-297.

De Bruijn, F.J., F.M. Ausubel. (1983). The cloning and characterisation of the glnF (ntrA) gene of Klebsiella pneumoniae. Role of glnF (ntrA) in the regulation of nitrogen fixation (nif) and other nitrogen assimilation genes. Mol. Gen. Genet. 192: 342-353.

De Bruijn, F.J., J.R. Lupski. (1984). The use of transposon Tn5 mutagenesis in the rapid generation of correlated physical and genetic maps of DNA segments cloned into multicopy plasmids - a review. Gene 27: 131-149.

De Bruijn, F.J., I.L. Stroke, D.J. Marvel, F.M. Ausubel. (1983). Construction of a correlated physical and genetic map of the Klebsiella pneumoniae hisDGO region using transposon Tn5 mutagenesis. EMBO J. 2: 1831-1838.

Deininger, P.L. (1983). Random subcloning of sonicated DNA: applications to shotgun DNA sequence analysis. Anal. Biochem. 129: 216-219.

De Man, J.C., M. Rogosa, M.E. Sharpe. (1960). A medium for the cultivation of lactobacilli. J. Appl. Bacteriol. 23: 130-135.

Denhardt, D.A. (1966). A membrane-filter technique for the detection of complementary DNA. Biochem. Biophys. Res. Commun. 23 641-646.

De Vos, W.M., G. Simons. (1988). Molecular cloning of lactose genes in dairy lactic streptococci: the phospho-ß-galactosidase and ß-galactosidase genes and their expression products. Biochimie 70: 461-473.

Deutch, A.H., K.E. Rushlow, C.J. Smith. (1984). Analysis of the Escherichia coli proBA locus by DNA and protein sequencing. Nucl. Acids Res. 12: 6337-6355.

Deutscher, J.K., H.M. Beyreuther, H.M. Sobek, K. Stuber, W. Hengstenberg.(1982). Phosphoenolpyruvate-dependant phosphotransferase system in Staphylococcus aureus: factor III_{lac}, a trimeric phospho-carrier protein that also acts as a phase transfer catalyst. Biochem. 21: 4867-4873.

Devereux, J., P. Haeberli, O. Smithies. (1984). A comprehensive set of sequence analysis programs for the Vax. Nucl. Acids Res. 12: 387-395.

Dickson, R.C., J. Abelson, W.M. Barnes, W.S. Reznikoff. (1975). Genetic regulation: the lac control region. Science 187: 27-35.

Dickson, R.C., J.S. Markin. (1978). Molecular cloning and expression in Escherichia coli of a yeast gene coding for ß-galactosidase. Cell 15: 123-130.

Dills, S.S., A. Apperson, M.R. Schmidt, M.H. Saier. (1980). Carbohydrate metabolism in bacteria. Microbiol. Rev. 44: 385-418.

Ditta, G., S. Stanfield, D. Corbin, D.R. Helinski. (1980). Broad host range DNA cloning system for Gram-negative bacteria: constitution of a gene bank of Rhizobium meliloti. Proc. Natl. Acad. Sci. 77: 7347-7351.

Doelle, H.W. (1975). Bacterial metabolism. 2nd ed. Academic Press, Inc., New York.

Doelle, H.W. (1981). Basic metabolic processes. Biotechnol. 1: 175-210.

Duncan, C.L., E.A. Rokos, C.M. Christenson, J.I. Rood. (1978). Multiple plasmids in different toxigenic types of Clostridium perfringens: Possible control of ß-Toxin production. In: Schlessinger D(ea). Microbiology-1978. American Society for Microbiology D.C. pp.246-248.

Dyer, D.W., J.J. Iandolo. (1983). Rapid isolation of DNA from Staphylococcus aureus. Appl. Environ. Microbiol. 46: 283-285.

Efstathiou I., N. Truffaut. (1986). Cloning of Clostridium acetobutylicum genes and the expression in Escherichia coli and Bacillus subtilis. Mol. Gen. Genet. 204: 317-321.

Egelhoff, T.T., S.R. Long. (1985). Rhizobium meliloti nodulation genes: Identification of nodABC gene products, purification of nodA protein, and expression of nodA in Rhizobium meliloti. J. Bacteriol. 164: 591-599.

Egelhoff, T.T., R.F. Fisher, T.W. Jacobs, J.T. Mulligan, S.R. Long. (1985). Nucleotide sequence of Rhizobium meliloti 1021 nodulation genes: nodD is read divergently from nodABC. DNA 4: 241-248.

Ehrenberg, I., I. Fedorcsak, F. Solymosy. (1976). Diethyl pyrocarbonate in nucleic acid research. Prog. Nucl. Acid Res. Mol. Biol. 16: 189-262.

Eklund, M.W., F.T. Poysky, S.M. Reed, C.A. Smith. (1971). Bacteriophage and the toxigenecity of Clostridium botulinum type C. Science 172: 480-482.

Elton, R.A. (1973). The relationship of DNA base composition and individual protein composition in micro-organisms. J. Molec. Evol. 2: 263-276.

Ennis, B.M., I.S. Maddox. (1985). Use of Clostridium acetobutylicum P262 for production of solvents from whey permeate. Biotechnol. Lett. 7: 601-606.

Ennis, B.M., I.S. Maddox, G.H. Schoutens. (1986a). Immobilised Clostridium acetobutylicum for continuous butanol production from whey permeate. N.Z. J. Dairy Sci. Technol. 21: 99-109.

Ennis, B.M., N.A. Gutierrez, I.S. Maddox. (1986b). The acetone- butanol- ethanol fermentation, a current assessment. Pros. Biochem. 21:131-147.

Ennis, B.M., I.S. Maddox. (1987a). Use of Clostridium acetobutylicum P262 for production of solvents from whey permeate. Biotechnol. Lett. 7: 601-606.

Ennis, B.M., I.S. Maddox. (1987b). The effect of pH and lactose concentration on solvent production from whey permeate using Clostridium acetobutylicum. Biotechnol. Bioeng. 29: 329-334.

Errington, J. (1988). Regulation of sporulation. Nature 333: 399-400.

Estramareix, B., D. Gaudruy, M. Therisod. (1977). Biosynthesis of the thiamine thiazole in Escherichia coli. Biochemie (Paris) 59: 857-859.

Evans, R.P., F.L. Macrina. (1983). Streptococcal resistance plasmid pIP501. Endonuclease restriction map, resistance determinant location, and construction of novel derivatives. J. Bacteriol. 154: 1347-1355.

Farrow, J.A.E. (1980). Lactose hydrolysing enzymes in Streptococcus lactis and Streptococcus cremoris and also in some other species of streptococci. J. Appl. Bacteriol. 49: 493-503.

Fickett, J.W. (1982). Recognition of protein coding regions in DNA sequences. Nucl. Acids Res. 10: 5303-5318.

Figurski, D.H., D.R. Helinski. (1979). Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. Proc. Natl. Acad. Sci. USA. 76: 1648-1652.

Fisher, K., M.C. Johnson, R. Bibek. (1985). Lactose hydrolysing enzymes in Lactobacillus acidophilus strains. Food Microbiol. 2: 23-29.

Fisher, S.G., L.S. Lerman. (1979). Length-independent separation of DNA restriction fragments in two dimension gel electrophoresis. Cell 16: 191-200.

Fitz-James, P., E. Young. (1969). Morphology of sporulation. In The bacterial spores. G.W. Gould & A. Hurst (eds.). p.39-72. Academic Press; New York.

Fond, O., G. Matta-Ammouri, H. Petitdemange, J.M. Engasser. (1985). The role of acids on the production of acetone and butanol by Clostridium acetobutylicum. Appl. Microbiol. Biotechnol. 22: 195-200.

Fox, C.R., J.R. Beckwith, W. Epstein, E.R. Signer. (1966). Transposition of the lac region of Escherichia coli II. On the role of thiogalactoside transacetylase in lactose metabolism. J. Mol. Biol. 19: 576-579.

Fox, G.E., E. Stackebrandt, R.B. Hespell, J. Gibson, J. Maniloff, T.A. Dyer, R.S. Wolfe, W.E. Balch, S. Tanner, L.J. Magrum, L.B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B.J. Lewis, D.A. Stahl, K.R. Luehrsen, K.N. Chen, C.R. Woese. (1980). The Phylogeny of Prokaryotes. Science 209: 457-463.

Fridovich, I. (1978). The biology of oxygen radicals. Science (Washington D.C.) 201: 857-880.

Friedman, A.M., S.R. Long, S.E. Brown, W.J. Buikema, F.M. Ausubel. (1982). Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of Rhizobium mutants. Gene 18: 289-296.

Fuhrmann, M., H. Hennecke. (1982). Coding properties of cloned nitrogenase structural genes from Rhizobium japonicum. Mol. Gen. Genet. 187: 419-425.

Fujita, Y., T. Fujita, Y. Miwa, J. Nihashi, Y. Aratani. (1986). Organization and transcription of the gluconate operon gnt of Bacillus subtilis. J. Biol. Chem. 261: 13744-13753.

Gabriel, C.L. (1928). Butanol fermentation process. Ind. Eng. Chem. 20: 1063-1067.

Gabriel, C.L., F.M. Crawford. (1930). Development of the butyl-acetonic fermentation industry. Ind. Eng. Chem. 22: 1163-1165.

Gapes, J.R., V.F. Larsen, I.S. Maddox. (1982). Microbial production on n-butanol: some problems associated with the fermentation. p.90-101. In Energy from biomass. Proceedings of the 14th Biotechnology Conference, Palmerston North, New Zealand.

Gapes, J.R., V.F. Larsen, I.S. Maddox. (1983). A note on procedures for inoculum development for the production of solvents by a strain of Clostridium butylicum. J. Appl. Bacteriol. 55: 363-365.

Gardner, A.L., A.I. Aronson. (1984). Expression of the Bacillus subtilis glutamine synthetase gene in Escherichia coli. J. Bacteriol. 158: 967-971.

Garnier, T., S.T. Cole. (1988). Complete nucleotide sequence and genetic organisation of the bacteriogenic plasmid, pIP404, from Clostridium perfringens. Plasmid 19: 134-150.

Gasson, M.J., F.L. Davies. (1980). High-frequency conjugation associated with Streptococcus lactis donor cell aggregation. J. Bacteriol. 143: 1260-1264.

George, H.A., J.S. Chen. (1983). Acidic conditions are not obligatory for onset of butanol formation by Clostridium beijerinckii (synonym C. butylicum). Appl. Environ. Microbiol. 46: 321-327.

George, H.A., J.L. Johnson, W.E.C. Moore, L.V. Holdeman, J.S. Cher. (1983). Acetone, isopropanol and butanol production by Clostridium beijerinckii (syn. Clostridium butylicum) and Clostridium aurantibutyricum. Appl. Environ. Microbiol. 45: 1160-1163.

Ghuysen, J.M. (1974). Substrate requirements of glycosidases for lytic activity on bacterial walls. pp.185-194 In: Lysozyme E.F. Osserman, R.E. Canfield, S. Beychok (ed.), Academic Press, New York.

Ghuysen, J.M., J.L. Strominger, D.J. Tipper. (1968). Bacterial Cell walls. In: Comprehensive biochemistry. M. Florkin, E.H. Stotz (ed.). Vol. 26, Extracellular and supporting structures. pp.53-105. Elsevier Publishing Co., New York.

Ghuysen, J.M., and G.D. Shockman. (1973). Biosynthesis of peptidoglycan. In Bacterial membranes and cell walls Vol. 1: 37-130. L. Leive (ed.). Marcel Dekker, Inc., N.Y.

Gibbs, D.F. (1983). The rise and fall (.....and rise???) of acetone/butanol fermentations. Trends Biotechnol. 1: 12-15.

Gilbert, H.J., J. Hall. (1987). Molecular cloning of Streptococcus bovis lactose catabolite genes. J. Gen. Microbiol. 133: 2285-2293.

Glauert, A.M., M.J. Thornley, K.J.I. Thorne, U.B. Sleytr. (1976). The surface structure of bacteria. pp.31-48. In: Microbial ultrastructure. R. Fuller, D. Lovelock (ed.). Academic Press New York.

Goeddel,D.V., D.G. Yansura, M.H. Carruthers. (1978). How lac repressor recognises the lac operator. Proc. Natl. Acad. Sci. USA 75: 3578-3582.

Gottschal, J.C., J.G. Morris. (1981). Non-production of acetone and butanol by Clostridium acetobutylicum during glucose and ammonium limitation in continuous culture. Biotechnol. Lett. 3: 525-530.

Gottschal, J.C., J.G. Morris. (1981). The induction of acetone and butanol production in cultures of Clostridium acetobutylicum by elevated concentrations of acetate and butyrate. FEMS Microbiol. Lett. 12: 385-389.

Gottschalk, G. (1986). Bacterial metabolism. 2nd edition. Springer-Verlag, New York.

Gottschalk G., H. Bahl. (1981). Feasible improvements of the butanol production by Clostridium acetobutylicum, p.463-470. In Hollaender, R. Rabson, P. Rogers, A. San Pietro, R. Valentine, and R. Wolfe (ed.). Trends in the biology of fermentations for fuels and chemicals. Plenum Publishing Corp., New York.

Graves, M.C., G.T. Mullenbach, J.C. Rabinowitz. (1985). Cloning and nucleotide sequence determination of the Clostridium pasteurianum ferredoxin gene. Proc. Natl. Acad. Sci. USA 82: 1653-1657.

Graves. M.C., J.C. Rabinowitz. (1986). In vivo and in vitro transcription of the Clostridium pasteurianum ferredoxin gene. J. Biol. Chem. 261: 11409-11415.

Greenberg, N.A., Mahoney, R.R. (1982). Production and characterisation of ß-galactosidase from Streptococcus thermophilus. J. of Food Science. 47: 1824-1835.

Grosjean, H., W. Fiers. (1982). Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene 18: 199-209.

Guiso, N., A. Ullmann. (1976). Expression and regulation of lactose genes carried by plasmids. J. Bacteriol. 127: 691-697.

Gutierrez, N.A., I.S. Maddox. (1987). Role of chemotaxis in solvent production by Clostridium acetobutylicum. Appl. Environ. Microbiol. 53: 1924-1927.

Haggstrom, L. (1985). Acetone-butanol fermentation and its variants. Biotechnol. Adv. 3: 13-28.

Hall, B.G., E.C.R. Reeve. (1977). A third ß-galactosidase in a strain of Klebsiella that possesses two lac genes. J. Bacteriol 132: 219-223.

Hamilton, I.R., G.C.Y. Lo. (1978). Co-induction of ß-galactosidase and the lactose ß-phosphoenol pyruvate phosphotransferase system in Streptococcus salivarius and Streptococcus mutans. J. Bacteriol. 136: 900-908.

Hammes, W., K.H. Schleifer, O. Kandler. (1973). Mode of action of glycine on the biosynthesis of peptidoglycan. J. Bacteriol. 116: 1029-1053.

Hardie, J.M. (1986). Genus Streptococcus. In: Bergeys Manual of Systemic bacteriology. Vol. 2., 9th edition. P.H.A. Sneath, N.S. Mair, M.E. Sharpe, J.G.Hold (eds.). Williams and Wilkins Pub., USA.

Hartmanis, M.G.N., S. Gatenbeck. (1984). Intermediary metabolism in Clostridium acetobutylicum: levels of enzymes involved in the formation of acetate and butyrate. Appl. Environ. Microbiol. 47: 1277-1283.

Hartmanis, M.G.N., T.S. Klason, S. Gatenbeck. (1984). Uptake and activation of acetate and butyrate in Clostridium acetobutylicum. Appl. Microbiol. Biotechnol. 20: 66-71.

Hastings, J.H.J. (1978). Acetone-butyl alcohol fermentation. p.31-45 In A.H. Rose (ed.), Economic microbiology, vol. 2. Primary products of metabolism. Academic Press, Inc., New York.

Hawley, D.K., W.R. McClure. (1983). Compilation and analysis of Escherichia coli promoter DNA sequences. Nucl. Acids Res. 11: 2237-2255.

Hays, J.B., R.D. Simoni, S. Roseman. (1973). Sugar transport. V. A trimeric lactose-specific phosphocarrier protein of the Staphylococcus aureus phosphotransferase system. J. Biol. Chem. 248: 941-956.

Hediger, M.A., D.F. Johnson, D.P. Nierlich, I. Zabin. (1985). DNA sequence of the lactose operon: the lacA gene and the transcriptional termination region. Proc. Natl. Acad. Sci. USA 82: 6414-6418.

Heidecker, G., J. Messing, B. Gronenborn. (1980). A versatile primer for DNA sequencing in the M13mp2 cloning system. Gene 10: 68-73.

Hemme, D.H., V. Schmal, J.E. Auclair. (1981). Effect of the addition of extracts of thermophilic lactobacilli on acid production by Streptococcus thermophilus in milk. J. Dairy Res. 48: 139-148.

Hengstenberg, W., W.K. Penberthy, K.L. Hill, M.L. Morse. (1969). Phosphotransferase system of Staphylococcus aureus: its requirements for the accumulation and metabolism of galactosides. J. Bacteriol. 99: 383-388.

Hengstenberg, W., W.K. Penberthy, K.L. Hill, M.L. Morse. (1970). Metabolism of lactose by Staphylococcus aureus. J. Bacteriol. 96: 2187-2188.

Hengstenberg, W., W.K. Penberthy, M.L. Morse. (1970). Purification of the staphylococcal 6-phospho-ß-D-galactosidase. Eur. J. Biochem. 14: 27-32.

Hermann, M., F. Fayolle, R. Marchal, L. Podin, M, Sebald, J.P. Vandecasteele. (1985). Isolation and characterisation of butanol resistant mutants of Clostridium acetobutylicum. Appl. Environ. Microbiol. 50: 1238-1243.

Herman, R.E., L.L. McKay. (1986). Cloning and expression of the ß-D-galactosidase from Streptococcus thermophilus in Escherichia coli. Appl. Environ. Microbiol. 52: 45-50.

Herman, R.E., L.L. McKay. (1987). Characterisation of plasmids and cloning of the ß-galactosidase gene from Streptococcus thermophilus. In Streptococcal genetics. J.J Ferretti, R. Curtiss III (ed.). ASM, Washington, D.C.

Herrchen, M., G. Legler. (1983). Identification of an essential carboxylate group at the active site of lacZ ß-galactosidase from Escherichia coli. Eur. J. Biochem. 138: 527-531.

Hickey, M.W., A.J. Hillier, R. Jago. (1986). Transport and metabolism of lactose, glucose and galactose in homo-fermentative lactobacilli. Appl. Environ. Microbiol. 51: 825-831.

Hirata, H., S. Negoro, H. Okada. (1984). Molecular basis of isozyme formation of ß-galactosidases in Bacillus stearothermophilus. Isolation of two ß-galactosidase genes, bgaA and bgaB. J. Bacteriol. 160: 9-14.

Hirata, H., T. Fukazawa, S. Negoro, H. Okado. (1986). Structure of a ß-galactosidase gene of Bacillus stearothermophilus. J. Bacteriol. 166: 722-727.

Hobman, P.G. (1984). Review of processes and products for utilisation of lactose in deproteinated milk serum. J. Dairy Sci. 67: 2630-2653.

Hohn, B. (1979). In vitro packaging of lambda and cosmid DNA. In: Methods in Enzymology 68: 299-309. R. Wu (ed.). Academic Press, New York.

Hohn, B., J. Collins. (1980). A small cosmid for efficient cloning of large DNA fragments. Gene 11: 291-298.

Holmes, D.S., M. Quigley. (1981). A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. 114: 193-197.

Holt, R.A., G.M. Stephens, J.G. Morris. (1984). Production of solvents by Clostridium acetobutylicum cultures maintained at neutral pH. Appl. Environ. Microbiol. 48: 1166-1170.

Honda, H., T. Kudo, K. Horikoshi. (1985). Molecular cloning and expression of the xylanase gene of alkalophilic Bacillus spp. strain C-125 in Escherichia coli. J.Bacteriol. 161: 784-785.

Hopwood, D.A. (1981). Genetic studies with bacterial protoplasts. Annu. Rev. Microbiol. 35: 237-272.

Horinouchi, S., T. Uozumi, T. Beppu, K. Arima. (1977). A new isolation method of plasmid deoxyribonucleic acid from Staphylococcus aureus using a lytic enzyme of Achromobacter. Agric. Biol. Chem. 41: 2487-2489.

Howard, G.T., B.A. White. (1988). Molecular cloning and expression of cellulase genes from Ruminococcus albus 8 in Escherichia coli bacteriophage lambda. Appl. Environ. Microbiol. 54: 1752-1755.

Hsu, E.J., and Z.J. Ordal. (1970). Comparative metabolism of vegetative and sporulating cultures of Clostridium thermosaccharolyticum. J. Bacteriol. 102: 369-376.

Ikemura, T. (1981). Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes. J. Mol. Biol. 146: 1-21.

Inamine, J.M., L.N. Lee, D.J. LeBlanc. (1986). Molecular and genetic characterization of lactose metabolic genes of Streptococcus cremoris. J. Bacteriol. 167: 855-862.

Ionesco, H., G. Bieth, C. Dauget, D. Bouanchaud. (1975). Isolment et identification de deux plasmides d'une soache bacteriocinogene de Clostridium perfringens. Ann. Microbiol. (Paris) 127B: 283-294.

Ish-Horowicz, D., J.F. Burke. (1981). Rapid and efficient cosmid cloning. Nucl. Acid. Res. 9: 2989-2998.

Ishii, K., T. Kudo, H. Honda, K. Horikoshi. (1983). Molecular cloning of α-isopropylmalate dehydrogenase gene for Clostridium butyricum M588. Agric. Biol. Chem. 47: 2313-2317.

Jacob, F., D. Perrin, C. Sanchez, J. Monod. (1960). L'operon: groupe de genes à expression coordonée par un operateur. C.R. Seances Acad. Sci. 250: 1727-1729.

Jacob, F., J. Monod. (1961). On the regulation of gene activity. Cold Spring Harbour Symp. Quant. Biol. 26: 193-209.

Janssen, P.J., W.A. Jones, D.T. Jones, D.R. Woods. (1988). Molecular analysis and regulation of the glnA gene of the Gram positive anaerobe Clostridium acetobutylicum. J. Bacteriol. 170: 400-408.

Jefferson, R.A., S.M. Burgess, D. Hirsh. (1986). ß-glucuronidase from Escherichia coli as a gene-fusion marker. Proc. Natl. Acad. Sci. USA 83: 8447-8451.

Johnson, M.J. (1947). Industrial fermentations. Annu. Rev. Microbiol. 1: 159-172.

Johnson, J.L., R.S. Anderson, E.J. Ordal. (1970). Nucleic acid homologies among oxidase negative Moraxella species. J. Bacteriol. 101: 568-573.

Johnson, J.L., B.S. Francis. (1975). Taxonomy of the Clostridia: Ribosomal ribonucleic acid homologies among the species. J. Gen. Microbiol. 88: 229-244.

Johnstone, K., K.T. Holland. (1977). Ultrastructural changes during sporulation of Clostridium bifermentans. J. Gen. Microbiol. 100: 217-220.

Jones, D.T., A. van der Westhuizen, S. Long, E.R. Allcock, S.J. Reid, D.R. Woods. (1982). Solvent production and morphological changes in Clostridium acetobutylicum. Appl. Environ. Microbiol. 43: 1434-1439.

Jones, D.T., W.A. Jones, D.R. Woods. (1985). Production of recombinants after protoplast fusion in Clostridium acetobutylicum P262. J. Gen. Microbiol. 131: 1213-1216.

Jones, D.T. D.R. Woods. (1986). Gene transfer, recombination and gene cloning in Clostridium acetobutylicum. Microbiol Sci. 3:19-22.

Jones, D.T. and D.R. Woods. (1986). Acetone-butanol fermentation revisited. Microbiol. Rev. 50: 484-524.

Jorgensen, R.A., S.J. Rothstein, W.S. Reznikoff. (1979). A restriction enzyme cleavage map of Tn5 and location of a region encoding neomycin resistance. Mol. Gen. Genet. 177: 65-72.

Kado, C.I., S.T. Liu. (1981). Rapid procedure for detection and isolation of large and small plasmids. J. Bacteriol. 145: 1365-1373.

Kalnins, A., K. Otto, U. Ruther, B. Muller-Hill. (1983). Sequence of the lacZ of Escherichia coli. EMBO J. 2: 593-597.

Kandler, O., N. Weiss. (1986). Genus Lactobacillus. In: Bergeys Manual of Systemic bacteriology. Vol. 2., 9th edition. P.H.A. Sneath, N.S. Mair, M.E. Sharpe, J.G.Hold (eds.). Williams and Wilkins Pub., USA.

Karube, I., N. Urano, T. Yamada, H. Hirochika, K. Sakaguchi. (1983). Cloning and expression of the hydrogenase gene from Clostridium butyricum in Escherichia coli. FEBS Lett. 158: 119-122.

Kay, D., and S.C. Warren. (1968). Sporulation in Bacillus subtilis: morphological changes. Biochem. J. 109: 819-824.

Kepes, A. (1969). Transcription and translation in the lactose operon of Escherichia coli studied by in vivo kinetics. Prog. Biophys. Mol. Biol. 19: 201-236.

Killeffer, D.H. (1927). Butanol and acetone from corn. A description of the fermentation process. Ind. Eng. Chem. 19: 46-50.

Kim, B.H., P. Bellows, R. Datta, and J.G. Zeikus. (1984). Control of carbon and electron flow in Clostridium acetobutylicum fermentations: utilisation of carbon monoxide to inhibit hydrogen production and to enhance butanol yields. Appl. Environ. Microbiol. 48: 764-770.

Klacik, M.A., H.P. Blaschek. (1983). Cellular localization and inhibition of DNAse in plasmid bearing strains of Clostridium perfringens. Annu. Meet. Am. Soc. Microbiol. 83: H93.

Kleckner, N., J. Roth, D. Bostein. (1977). Genetic engineering in vivo using translocatible drug-resistance elements: new methods in bacterial genetics. J. Mol. Biol. 116: 125-159.

Kreig, N.R., P.S. Hoffmen. (1986). Microaerophily and oxygen toxicity. Ann. Rev. Microbiol. 40: 107-130.

Kudoh. J., T. Ikeuchi, K. Kurahashi. (1985). Nucleotide sequence of the sporulation gene spo0A and its mutant genes of Bacillus subtilis. Proc. Natl. Acad. Sci. USA 82: 2665-2668.

Kutzenok, A., M Aschner. (1952). Degenerative processes in a strain of Clostridium butylicum. J. Bacteriol. 64: 829-836.

Kyte, J., R.F. Doolittle. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157: 105-132.

Laird, W.J., W. Aaronson, R.P. Silver, W.H. Habig, M.C. Hardegree. (1980). Plasmid associated toxigenicity in Clostridium tetani. J. Infect. Disease 142: 623.

Lampen, J.O., W.H. Peterson. (1943). Growth factor requirements of clostridia. Arch. Biochem. 2: 443-449.

Langlykke, A.F., J.M. Van Lanen, D.R. Fraser. (1948). Butyl alcohol from xylose saccharification liquors from corncobs. Ind. Eng. Chem. 40: 1716-1719.

LeBlanc, D.J., L.N. Lee. (1984). Physical and genetic analyses of streptococcal plasmid pAMß1 and cloning of its replication region. J. Bacteriol. 157: 445-453.

Lederberg, J. (1950). The ß-galactosidase of Escherichia coli, strain K-12. J. Bacteriol. 60: 381-392.

Lee, C-K., P. Durre, H. Hippe, G. Gottschalk. (1987). Screening for plasmids in the genus Clostridium. Arch. Microbiol. 148: 107-114.

Lee, L.J., J.B. Hansen, E. Jagusztyn-Krynicka, B.M. Chassy. (1982). Cloning and expression of the ß-D-galactoside galactohydrolase gene of Lactobacillus casei in Escherichia coli K-12. J. Bacteriol. 152: 1138-1146.

Lee, S.F., C.W. Forsberg, L.N. Gibbons. (1985a). Cellulolytic activity of Clostridium acetobutylicum. Appl. Environ. Microbiol. 50: 220-228.

Lee, S.F., C.W. Forsberg, L.N. Gibbons. (1985b). Xylanolytic activity of Clostridium acetobutylicum. Appl. Environ. Microbiol. 50: 1068-1076.

Lemmel, S.A. (1985). Mutagenesis in Clostridium acetobutylicum. Biotechnol. Lett. 7: 711-716.

Lenz, T.G., A.R. Moreira. (1980). Economic evaluation of the acetone-butanol fermentation. Ind. Eng. Chem. Prod. Res. Dev. 19: 478-483.

Li, A.W., P.J. Krell, D.E. Mahony. (1980). Plasmid detection in a bacteriocinogenic strain of Clostridium perfringens. Can. J. Microbiol. 26: 1018-1022.

Limsowtin, G.K.Y., V.L. Crow, L.E. Pearce. (1986). Molecular cloning and expression of the Streptococcus lactis tagatose-1,6,- bisphosphate aldolase gene in Escherichia coli. FEMS Microbiol. Lett. 33: 79-83.

Lin, Y.L., H.P. Blaschek. (1984). Transformation of heat treated Clostridium acetobutylicum protoplasts with pUB110 plasmid DNA. Appl. Environ. Microbiol. 48: 737-742.

Linden, J.C., A.R. Moreira. (1982).In "Anaerobic production of chemicals". p.377-404 In A. Hollaender, A.I. Larkin, A.P. Rogers (ed.). Biological basis for new developments in biotechnology. Plenum Publishing Corp., New York.

Linn, T., R. Losick. (1976). The program of protein synthesis during sporulation in Bacillus subtilis. Cell 8: 103-114.

Ljungquist, S. (1977). A new endonuclease from Escherichia coli acting at apurinic sites in DNA. J. Biol. Chem. 252: 2808-2814.

Loomis, W.F. Jr, B. Magasanik. (1964). The relation of catabolite repression to the induction system for ß-galactosidase in Escherichia coli. J. Mol. Biol. 8: 417-426.

Long, S., D.T. Jones, D.R Woods. (1983). Sporulation of Clostridium acetobutylicum P262 in a defined medium. Appl. Environ. Microbiol. 45: 1389-1393.

Long, S., D.T. Jones, D.R Woods. (1984a). The relationship between sporulation and solvent production in Clostridium acetobutylicum p262. Biotechnol. Lett. 6: 529-534.

Long, S., D.T. Jones, D.R. Woods. (1984b). Initiation of solvent production, clostridial stage and endospore formation in Clostridium acetobutylicum P262. Appl. Microbiol. Biotechnol. 20: 256-261.

Losick, R., P. Youngman. (1984). Endospore formation in Bacillus. In Microbial development: Cold Spring Harbour Monograph Series: Monograph 16: 63-89. Ed., R. Losick, L. Shapiro.

Losick, R. (1982). Nucleotide sequences that signal the initiation of transcription and translation in Bacillus subtilis. Mol. Gen. Genet. 186: 339-346.

Luczak, H., H. Schwarzmoser, W.L. Staudenbauer. (1985). Construction of Clostridium butyricum hybrid plasmids and transfer into Bacillus subtilis. Appl. Microbiol. Biotechnol. 23: 114-122.

Ludwig, W., E. Seewaldt, R. Kilpper-Balz, K.H. Schleifer, L. Magrum, C.R. Woese, G.E. Fox, E. Stackebrandt. (1985). The phylogenetic position of Streptococcus and Enterococcus. J. Gen. Microbiol. 131: 543-551.

Lupski, J.R., B.L. Smiley, F.R. Blattner, G.N. Godson. (1982). Cloning and characterisation of the Escherichia coli chromosomal region surrounding the dnaG gene, with a correlated physical and genetic map of dnaG generated via transposon Tn5 mutagenesis. Mol. Gen. Genet. 185: 120-128.

MacDonald, C., M. Riley. (1983). Cloning of chromosomal lac genes of Klebsiella pneumoniae. Gene 24: 341-345.

Mackey, B.M., J.G. Morris. (1971). Ultrastructural changes during sporulation of Clostridium pasteurianum. J. Gen. Microbiol. 66: 1-13.

Macrina, F.L., J.A. Tobian, K.R. Jones, R.P. Evans, D.B. Clewell. (1982). A cloning vector able to replicate in Escherichia coli and Streptococcus sanguis. Gene 19: 345-353.

Maddox, I.S. (1980). Production of n-butanol form whey filtrate using Clostridium acetobutylicum NCIB 2951. Biotechnol. Lett. 2: 493-498.

Maddox, I.S., J.R. Gapes, V.F. Larsen. (1981). Production of n-butanol from whey ultrafiltrate. p.535-542. Chemeca 81 - 9th Australasian conference on chemical engineering, Christchurch, N.Z.

Maddox, I.S. (1982). Production of ethanol and n-butanol from hexose/pentose mixtures using consecutive fermentations with Saccharomyces cerevisiae and Clostridium acetobutylicum. Biotechnol. Lett. 4: 23-28.

Maddox, I.S., A.E. Murray. (1983). Production of n-butanol by fermentation of wood hydrolysate. Biotechnol. Lett. 5: 175-178.

Maeda, S., M.J. Gasson. (1986). Cloning, expression and location of the Streptococcus lactis gene for phospho-ß-galactosidase. J. Gen. Microbiol. 132: 331-340.

Magasanik, B. (1982). Genetic control of nitrogen assimilation in bacteria. Annu. Rev. Genet. 16: 135-168.

Mahony, D.E., G.A. Clark, M.F. Stringer, M.C. MacDonald, D.R. Duchesne, J.A. Mader. (1986). Rapid extraction of plasmids from Clostridium perfringens. Appl. Environ. Microbiol. 51: 521-523.

Majors, J. (1975). Specific binding of CAP factor to lac promoter DNA. Nature (London) 256: 672-674.

Maniatis, T., E.F. Fritsch, J. Sambrook. (1982). Molecular cloning - a laboratory manual. Cold Spring Harbour Laboratory.

Manly, S.P., K.S. Matthews. (1984). Lac operator DNA modification in the presence of proteolytic fragments of the repressor protein. J. Mol. Biol. 179: 315-333.

Marchal, R., M. Rebeller, J.P. Vandecasteele. (1984). Direct bioconversion of alkali-pretreated straw using simultaneous enzymatic hydrolysis and acetone-butanol fermentation. Biotechnol. Lett. 6: 523-528.

Marmur, J. (1961). A procedure for the isolation of DNA from microorganisms. J. Mol. Biol. 3: 208-218.

Martin, J.H., D.C. Savage. (1988). Degradation of DNA in cells and extracts of the obligately anaerobic bacterium Roseburia cecicola upon exposure to air. Appl. Environ. Microbiol. 54: 1619-1621.

Martin, J.R., H. Petitdemange, J. Ballongue, Gay, R. (1983). Effects of acetic and butyric acids on solvent production by Clostridium acetobutylicum. Biotech. Letters 5: 89-94.

Matta-El-Amouri, G., R. Janati-Idrissi, O. Assobhei, H. Petitdemange, R. Gay. (1985). Mechanism of the acetone formation by Clostridium acetobutylicum. FEMS Microbiol. Lett. 30: 11-16.

McClatchy, J.K., E.D. Rosenblum. (1963). Induction of lactose utilisation is Staphylococcus aureus. J. Bacteriol. 86: 1211-1215.

McCoy, E., E.B. Fred, W.H. Peterson, E.G. Hastings. (1926). A cultural study of certain anaerobic butyric acid forming bacteria. J. Infect. Dis. 39: 457-483.

McCoy, E., E.B. Fred, W.H. Peterson, E.G. Hastings. (1928). A cultural study of certain anaerobic butyric acid-forming bacteria. J. Infect. Dis. 46: 118-137.

McKay, L., A. Miller, W.E. Sandine, P.R. Elliker. (1970). Mechanisms of lactose utilisation by lactic acid streptococci: Enzymatic and genetic analysis. J. Bacteriol. 102: 804-809.

McKay, L.L., L.A. Walter, W.E. Sandine, P.R. Elliker. (1969). Involvement of phosphoenolpyruvate in lactose utilization by group N streptococci. J. Bacteriol. 99: 603-610.

McKinnon, R.D., J.S. Waye, D.S. Bautista, F.L. Graham. (1985). Non-random insertion of Tn5 into cloned human adenovirus DNA. Gene 40: 31-38.

McLaughlin, J.R., C.L. Murray, J.C. Rabinowitz. (1981). Unique features in the ribosome binding site sequence of the Gram-positive Staphylococcus aureus ß-lactamase gene. J. Biol. Chem. 256: 11283-11291.

Mead, G.C. (1971). The amino-acid fermenting clostridia. J. Gen. Microbiol. 67: 47-56.

Meinecke, B., H. Bahl, G. Gottschalk. (1984). Selection of an asporogenous strain of Clostridium acetobutylicum in continuous culture under phosphate limitation. Appl. Environ. Microbiol. 48: 1064-1065.

Mendez B.S., R.F. Gomez. (1982). Isolation of Clostridium thermocellum auxotrophs. Appl. Environ. Microbiol. 43: 495-496.

Mes-Hartree, M., J.N. Saddler. (1982). Butanol production of Clostridium acetobutylicum grown on sugars found in hemicellulose hydrolysates. Biotechnol. Lett. 4: 247-252.

Messing, J., B. Gronenborn, B. Muller-Hill, P. Hofschneider. (1977).Filamentous coliphage M13 as a cloning vehicle: Insertion of a HindII fragment of the lac regulatory region in M13 replicative form in vitro. Proc. Natl. Acad. Sci. 74: 3642-3646.

Messing, J., J. Vieira. (1982). A new pair of M13 vectors for selecting either strand of a double-digest restriction fragment. Gene 19: 269-276.

Messing, J. (1983). New M13 vectors for cloning. Met. Enzymol. 101: 20-79.

Meyer, C.L., J.K. McLaughlin, E.T. Papoutsakis. (1985). The effect of CO on growth and product formation in batch fermentations of Clostridium acetobutylicum. Biotechnol. Lett. 7: 37-42.

Mihelc, V.A., C.L. Duncan, G.H. Chambliss. (1978). Characterisation of a bacteriocinogenic plasmid in Clostridium perfringens CW55. Antimicrob. Agents. Chemother. 14: 771-779.

Miller, J.H. (1972). Experiments in molecular genetics. Cold Spring Harbour Laboratory, Cold Spring Harbour, New York.

Miller, J.H. (1978). The lacI gene: its role in lac operon control and its use as a genetic system, p.31-88. In J.H. Miller and W.S. Reznikoff (ed.), The Operon. 2nd edition. Cold Spring Harbour Laboratory, New York.

Miller, J.H., W.S. Reznikoff. (1980). The Operon, 2nd edition. Cold Spring Harbour Laboratory, Cold Spring Harbour, New York.

Miller, J.H. (1983). Mutational specificity in bacteria. Annu. Rev. Genet. 31: 732-736.

Miller, R.V., A.J. Clark. (1976). Purification and properties of two deoxyribonucleases of Pseudomonas aeruginosa. J. Bacteriol. 127: 794-802.

Miller, J.H., M.P. Calos, D. Galas, M.Hofer, D.E. Buchel, B. Muller-Hill. (1980). Genetic analysis of transposition in the lac region of Escherichia coli. J. Mol. Biol. 144: 1-18.

Millet, J., D. Petre, P. Beguin, O. Raynaud, J. Aubert. (1985). Cloning of ten distinct DNA fragments of Clostridium thermocellum coding for cellulases. FEMS Microbiol. Lett. 29: 145-149.

Minton, N., J.G. Morris. (1983). Regeneration of protoplasts of Clostridium pasteurianum ATCC6013. J. Bacteriol. 155: 432-434.

Mitchell, W.J., P. Misko, S. Roseman. (1982). Sugar transport by the bacterial phospho-transferase system. Regulation of other transport systems (lactose and melibiose). J. Biol. Chem. 257: 14553-14564.

Monot, F., J-R. Martin, H. Petitdemange, R. Gay. (1982). Acetone and butanol production by Clostridium acetobutylicum in a synthetic medium. Appl. Environ. Microbiol. 44: 1318-1324.

Monot, F., J.M. Engasser. (1983). Production of acetone and butanol by batch and continuous culture of Clostridium acetobutylicum under nitrogen limitation. Biotechnol. Lett. 5: 213-218.

Monot, F., J.M. Engasser, H. Petitdemange. (1983). Regulation of acetone butanol production in batch and continuous cultures of Clostridium acetobutylicum. Biotechnol. Bioeng. Symp. 13: 207-216.

Monot, F., J.M. Engasser, H. Petitdemange. (1984). Influence of pH and undissociated butyric acid on the production of acetone and butanol in batch cultures of Clostridium acetobutylicum. Appl. Microbiol. Biotechnol. 19: 422-426.

Moore, R.L., B.J. McCarthy. (1967). Comparative study of ribosomal ribonucleic acid cistrons in enterobacteria and myxobacteria. J. Bacteriol. 94: 1066-1074.

Moran, C.P., Jr., N. Lang, S.F.J. LeGrice, G. Lee, M. Stephens, A.L. Sorenshein, J. Pero, R. Losick. (1982). Nucleotide sequences that signal the initiation of transcription and translation in Bacillus subtilis. Mol .Gen .Genet. 186: 339-346.

Moreira, A.R. (1983). Acetone-butanol fermentation, p. 385-406 In D.L. Wise (ed.), Organic chemicals form biomass. Benjamin/ Cummings Publishing Co. Inc., California.

Morimyo, M. (1982). Anaerobic incubation enhances the colony formation of a polA recB strain of Escherichia coli. J. Bacteriol. 152: 209-214.

Morse, M.L., K.L. Hill, J.B. Egan, W. Hengstenberg. (1968). Metabolism of lactose by Staphylococcus aureus and its genetic basis. J. Bacteriol. 95: 2270-2274.

Muller-Hill, B., H.V. Rickenberg, K. Wallenfels. (1964). Specificity of the induction of the enzymes of the lac operon in Escherichia coli. J. Mol. Biol. 10: 303-318.

Murray, W.D., K.B. Wemyss, A.W. Khan. (1983). Increased ethanol production and tolerance by a pyruvate negative- mutant of Clostridium saccharolyticum. Eur. J. Appl. Microbiol. Biotechnol. 18: 71-74.

Nakamura, K., Y. Okazawa, M. Soejima, T. Masaki. (1973). Lysis of Micrococcus radiodurans. Agric. Biol. Chem. 37: 2667.

Nakazawa, T., R.D. Simoni, J.B. Hays, S. Roseman. (1971). Phosphorylation of a sugar-specific protein component of the lactose transport system in Staphylococcus aureus. Biochem. Biophys. Res. Commun. 42: 836-843.

Nakhmanovich, B.M., N.A. Shcheblykina. (1959). Fermentation of pentoses of corn cob hydrolysates by Clostridium acetobutylicum. Mikrobiologiya 28: 99-104.

Needleman, S.B., C.D. Wunsch. (1970). A general method applicable to the search for similarities in the amino acid sequences of two proteins. J. Mol. Biol. 48: 443-453.

Nelson, S.O., J.K. Wright, P.W. Postma. (1983). The mechanism of inducer exclusion. Direct interaction between purified III^{Glc} of the phosphoenol pyruvate: sugar phosphotransferase system and the lactose carrier of Escherichia coli. EMBO J. 2: 715-720.

Nestle, M., W.K. Roberts. (1969). An extracellular nuclease from Serratia marcescens I: Purification and some properties of the enzyme. J. Biol. Chem. 244: 5213-5218.

Nikaido, H. (1973). Biosynthesis and assembly of lipopolysaccharide and the outer membrane layer of Gram-negative cell walls. In Bacterial membranes and walls, p.131-208. Marcel Dekker Inc., New York.

Nishio, N., H. Biebl, M. Meiners. (1983). Effect of pH on the production of acetone and butanol by Clostridium acetobutylicum in a minimal medium. J. Fermen. Technol. 61: 101-104.

Nixon, B.T., C.W. Ronson, F.M. Ausubel. (1986). Two component regulatory systems responsive to environmental stimuli share strongly conserved domains with the nitrogen assimilation regulatory genes ntrB and ntrC. Proc. Natl. Acad. Sci. USA 83: 7850-7854.

Norrander, J., T. Kempe, J. Messing. (1983). Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26: 101-106.

N.Z.E.R.D.C. (1982). Production and utilisation of butanol-acetone as liquid fuel. Report by Badger Engineers N.Z. Ltd, No.3225 to the New Zealand Energy Research Development Council.

Ogasawara, N. (1985). Markedly unbiased codon usage in Bacillus subtilis. Gene 40: 145-150.

Ogata, S., M. Hongo. (1979). Bacteriophages of the genus Clostridium. Adv. Appl. Microbiol. 25: 241-273.

Okamoto, T., T. Morichi. (1979). Distribution of ß-galactosidase and ß-phosphogalactosidase activity among lactic streptococci. Agric. Biol. Chem. 43: 2389-2390.

Orskov, F. (1986). Escherichia coli. In Bergeys Manual of Systemic bacteriology. Vol. 1. P.H.A. Sneath, N.S. Mair, M.E. Sharpe, J.G.Hold (eds.). Williams and Wilkins Pub., USA.

Ounine, K., H. Petitdemange, G. Raval, R. Gay. (1983). Acetone-butanol production from pentoses by C. acetobutylicum. Biotechnol. Lett. 5: 605-610.

Oultram, J.D., M. Young. (1985). Conjugal transfer of plasmid pAMß1 from Streptococcus lactis and Bacillus subtilis to Clostridium acetobutylicum. FEMS Microbiol. Lett. 27: 129-134.

Oxford, A.E., J.O. Lampen, W.H. Peterson. (1940). Growth factor and other nutritional requirements of the acetone-butanol organism, Clostridium acetobutylicum. Biochem. J. 34: 1588-1597.

Pace, B., L.L. Campbell. (1971). Homology of ribosomal ribonucleic acid of diverse bacterial species with Escherichia coli and Bacillus stearothermophilus. J. Bacteriol. 107: 543-547.

Palva, C.T., P. Kiljestrom, S. Harayama. (1981). Cosmid cloning and transposon mutagenesis in Salmonella typhimurium using bacteriophage lambda vehicles. Mol. Gen. Genet. 181: 153-157.

Panbangred, W., A. Shinmyo, S. Kinoshita, H. Okada. (1983). Purification and properties of endo-xylanase produced by Bacillus pumilis. Agric. Biol. Chem. 47: 957-963.

Pan-Hou, H.S.K., M. Hosno, N. Imura. (1980). Plasmid controlled mercury biotransformation by Clostridium cochearium T-2. Appl. Environ. Microbiol. 40: 1007-1011.

Papoutsakis, E.T. (1984). Equations and calculations for fermentations of butyric acid bacteria. Biotechnol. Bioeng. 26: 174-187.

Perlman, R., I. Pastan. (1968). Cyclic 3'5'-AMP stimulation of ß-galactosidase and tryptophanase induction in Escherichia coli. Biochem. Biophys. Res. Commun. 30: 656-664.

Petitdemange, H., R. Gay. (1986). Biochemical mechanism of the acetone-butanol fermentation and future prospects. In Proceed. 3rd International Conference Biotechnology in the pulp and paper industry. pp.110-113. Stockholm.

Piggot, P.J., J.A. Hoch. (1985). Revised genetic linkage map of Bacillus subtilis. Microbiol. Rev. 49: 158-179.

Poncz, M., D. Solowiejczyk, M. Ballantine, E. Schwartz, S. Surrey. (1982). Nonrandom DNA sequence analysis in bacteriophage M13 by the dideoxy chain-termination method. Proc. Natl. Acad. Sci. USA 79: 4298-4301.

Postma, P.W., J.W. Lengeler. (1985). Phosphoenolpyruvate carbohydrate phosphotransferase systems in bacteria. Microbiol. Rev. 49: 232-269.

Premi, L., W.E. Sandine, P.R. Elliker. (1972). Lactose-hydrolysing enzymes of Lactobacillus species. Appl. Microbiol. 24: 51-57.

Prescott, S.G., G.C. Dunn. (1959). Industrial microbiology. 3rd ed. McGraw-Hill Book Co., New York.

Qadeer, M.A., F.M. Choudhry, S. Ahmad, S. Rashid, M.A. Akhtar. (1980). Acetone-butanol fermentation of cane molasses by Clostridium acetobutylicum. Pakistan J. Scient. Res. 32: 157-163.

Qureshi, N., I.S. Maddox. (1987). Continuous solvent production from whey permeate using cells of Clostridium acetobutylicum immobilised by absorption onto bonechar. Enzyme. Microb. Technol. 9: 668-671.

Reid, S.J., E.R. Allcock, D.T. Jones, D.R. Woods. (1983). Transformation of Clostridium acetobutylicum protoplasts with bacteriophage DNA. Appl. Environ. Microbiol. 45: 305-307.

Reizer, J., M.J. Novotny, I. Stuiver, M.H. Saier. (1984). Regulation of glycerol uptake by the phosphoenol pyruvate-sugar phosphotransferase system in Bacillus subtilis. J. Bacteriol. 159: 243-250.

Reshetnik, O.A., F.G. Kupriyanova, V.G. Vinter. (1978). Changes in the activity of intracellular deoxyribonucleases in the course of growth and division of cells of Bacillus subtilis. Mikrobiologiya 47: 37-40.

Reyes, T.R., M.N. Mickelson. (1945). Growth factor requirements of saccharolytic butyl alcohol acetone bacteria. Arch. Biochem. 6: 471-477.

Reysset, G., M. Sebald. (1985). Conjugal transfer of plasmid-mediated antibiotic resistance from streptococci to Clostridium acetobutylicum. Ann. Inst. Microbiol. 136B: 275-282.

Reysset, G., J. Hubert, L. Podvin, M. Sebald. (1987). Protoplast formation and regeneration of Clostridium acetobutylicum strain NI-4080. J. Gen. Microbiol. 133: 2595-2600.

Rickenberg, H.V., G.N. Cohen, G. Buttin, J. Monod. (1956). La galactoside permease d'Escherichia coli. Ann. Inst. Pasteur 91: 829-857.

Robson, P.M., D.T. Jones. (1982). Production of acetone-butanol by industrial fermentation. p.169-214 In O. Chaude & G. Durand (ed.), Industrielle et Biotechnologie. Production d'intermediates industriel par culture anaerobie. Societe Francaise de Microbiologie, Paris.

Robson, R.L., R.M. Robson, J.G. Morris. (1974). The biosynthesis of granulose by Clostridium pasteurianum. Biochem. J. 144: 503-511.

Rodgers, M.R., J.C. Cabelli, D.R. Nelson. (1987). Cloning and analysis of a lactose operon from Aeromonas caviae N-94 260. Abstracts of the Annual Meeting of American Society for Microbiology 1987. 1-6 March 1987.

Rogers, P. (1986). Genetics and biochemistry of Clostridium relevant to development of fermentation processes. Adv. Appl. Microbiol. 31: 1-60.

Rogers, P., N. Palosaari. (1987). Clostridium acetobutylicum mutants that produce butyraldehyde and altered quantities of solvents. Appl. Environ. Microbiol. 53: 2761-2766.

Ronson, C.W., B.T. Nixon, F.M. Ausubel. (1987). Conserved domains in bacterial regulatory proteins that respond to environmental stimuli. Cell 49: 579-581.

Rood, J.I., V.N. Scott, C.L. Duncan. (1978). Identification of a transferable tetracycline resistant plasmid (pCW3) from Clostridium perfringens. Plasmid 1: 563-570.

Rose, A.H. (1961). Industrial microbiology, p.160-166. Butterworths, London.

Rosenblum, E.D., P.W. Wilson. (1949). Fixation of isotopic nitrogen by Clostridium. J. Bacteriol. 57: 413-417.

Ross, D. (1961). The acetone-butanol fermentation. Prog. Ind. Microbiol. 3: 73-90.

Rubbo, S.D., M. Maxwell, R.A. Fairbridge, J.M. Gillespie. (1941). The biology, growth factor requirements and fermentation reactions of Clostridium acetobutylicum (Weizmann). Aust. J. Expt. Biol. Med. Sci. 19: 184-198.

Saddler, J.N., E.K.C. Yu, M. Mes-Hartree, N. Levitin, H.H. Brownell. (1983). Utilization of enzymatically hydrolysed wood hemicelluloses by microorganisms for production of liquid fuels. Appl. Environ. Microbiol. 45: 153-160.

Saier, Jr. M.H., B.U. Feucht. (1975). Co-ordinate regulation of adenylate cyclase and carbohydrate permeases by the phosphoenol pyruvate: sugar phosphotransferase system in Salmonella typhimurium. J. Biol. Chem. 250: 7078-7080.

Saier, Jr. M.H., S. Roseman. (1976). Sugar transport. Inducer exclusion and regulation of the melibiose, maltose, glycerol, and lactose transport systems by the phosphoenol-pyruvate: sugar phosphotransferase system. J. Biol. Chem. 251: 6606-6615.

Saier Jr. M.H. (1985). Mechanisms and regulation of carbohydrate transport in bacteria. Academic Press, Inc., New York.

Sanger, F., S. Nickleson, A.R. Coulson. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463-5467.

Sanger, F., A.R. Carlson, B.G. Barret, A.J.H. Smith, B. Roe. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. 143: 232-245.

Sanger, F., A.R. Coulson, G.F. Hong, D.F. Hill, G.B. Peterson. (1982). Nucleotide sequencing of bacteriophage lambda DNA. J. Mol. Biol. 162: 729-773.

Sankar, A., A.M. Hack, W.D. Rupp. (1979). Simple method for identification of plasmid-coded proteins. J. Bacteriol. 137: 692-693.

Schafer, A., O. Schrecker, W. Hengstenberg. (1981). The staphylococcal phospho pyruvate dependant phosphotransferase system. Purification and characterisation of the galactoside-specific membrane component Enzyme II. Eur. J. Biochem. 113: 289-294.

Schleifer, K.H., E. Stackebrandt. (1983). Molecular systematics of prokaryotes. Ann. Rev. Microbiol. 37: 143-187.

Schreier, H.J., S.H. Fisher, A.L. Sonenstein. (1985). Regulation of expression from the glnA promoter of Bacillus subtilis requires the glnA gene product. Proc. Natl. Acad. Sci. USA 82: 3375-3379.

Schreier, P.H., R. Cortese. (1979). A fast and simple method for sequencing DNA cloned in the single-stranded bacteriophage M13. J. Mol. Biol. 129: 169-172.

Schwartz, R.D., F.A. Keller. (1982). Isolation of a strain of Clostridium thermoaceticum capable of growth and acetic acid production at pH 4.5. Appl. Environ. Microbiol. 43: 117-123.

Scott, V.N., C.L. Duncan. (1978). Cryptic plasmids in Clostridium botulinum and Clostridium botulinum-like organisms. FEMS Microbiol. Lett. 4: 55-58.

Scurlock, T.R., R.V. Miller. (1979). PaeExoIX; a unique deoxyribonuclease from Pseudomonas aeruginosa active in the presence of EDTA. Nucl. Acid Res. 7: 167-177.

Sebald, M. (1982). In Woods D.R. (1982). Molecular genetic studies on obligate anaerobic bacteria. S. African J. Sci. 78: 448-450.

Sharp, P.M., E. Cowe, D.G. Higgins, D.C. Shields, K.H. Wolfe, F. Wright. (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. Nucl. Acids Res. 16: 8207-8217.

Shaw, K.J., C.M. Berg. (1979). Escherichia coli K-12 auxotrophs induced by insertion of the transposable element Tn5. Genetics 92: 741-744.

Shimotsu, H., F. Kawamura, Y. Kobayashi, H. Saito. (1983). Early sporulation gene spo0F: nucleotide sequence and analysis of the gene product. Proc. Natl. Acad. Sci. USA 80: 658-662.

Shine, J., L. Dalgarno. (1974). The 3' terminal sequence of Escherichia coli 16S ribosomal RNA: complementary to nonsense triplets and ribosomal binding sites. Proc. Natl. Acad. Sci. USA 71: 1342-1346.

Shockman, G.D., J.F. Barret. (1983). Structure, function and assembly of cell walls of Gram-positive bacteria. Ann. Rev. Microbiol. 37: 501-527.

Short, J.L. (1978). Prospects for the utilisation of deproteinised whey in New Zealand - a review. N.Z. J. Dairy Sci. Technol. 13: 181-194.

Sidman, K.E., D.G. George, W.C. Barker, L.T. Hunt. (1988). The protein identification resource (PIR). Nucl. Acids Res. 16: 1869-1871.

Siegel, J.L., S.F. Hurst, E.S. Liberman, S.E. Coleman, A.S. Bleiweis. (1981). Mutanolysin-induced spheroplasts of Streptococcus mutans are true protoplasts. Infect. Immun. 31: 808-815.

Simon, R., U. Priefer, A. Puhler. (1984). A broad host range mobilisation system for in vivo genetic engineering: random and site-specific transposon mutagenesis in Gram-negative bacteria. Biotechnol. 1: 784-791.

Simoni, R.D., T. Nakazawa, J.B. Hays, S. Roseman. (1973). Sugar transport. IV. Isolation and characterisation of the lactose phosphotransferase system in Staphylococcus aureus. J. Biol. Chem. 248: 932-940.

Simoni, R.D., M.F. Smith, S. Roseman. (1968). Resolution of a staphylococcal phosphotransferase system into four protein components and its relation to sugar transport. Biochem. Biophys. Res. Commun. 31: 804-811.

Smart, J.B., V.L. Crow, T.D. Thomas. (1985). Lactose hydrolysis in milk and whey using ß-galactosidase from Streptococcus thermophilus. N.Z. J. Dairy Sci. Technol. 20: 43-56.

Smith, A.J.H. (1979). The use of exonuclease III for preparing single-stranded DNA for use as a template in the chain terminator sequencing method. Nucl. Acids Res. 6: 831-848.

Smith, E. (1957). Cytological observations on the sporulation process of Clostridium perfringens. J. Bacteriol. 73: 1-7.

Smith, T.F. M.S. Waterman. (1981). Comparison of biosequences. Adv. Appl. Math. 2: 482-489.

Solberg, M., H.P. Blaschek, P. Kahn. (1981). Plasmids in Clostridium perfringens. J. of Food Safety 3: 267-289.

Somk uti, G.A., D.H. Steinberg. (1979). ß-D-galactoside galactohydrolase of Streptococcus thermophilus: induction, purification and properties. J. Appl. Biochem. 1: 357-368.

Somers, J.M., A. Amzallag, R.B. Middleton. (1973). Genetic fine structure of the leucine operon of Escherichia coli K12. J. Bacteriol. 113: 1268-1272.

Soni, B.K., K. Das, T.K. Ghose. (1982). Bioconversion of agro-wastes into acetone-butanol. Biotechnol. Lett. 4: 19-22.

Soucaille, P., G. Joliff, A. Izard, G. Goma. (1987). Butanol tolerance and autobacteriocin production by Clostridium acetobutylicum. Curr. Microbiol. 14: 295-299.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98: 503-517.

Spivey, M.J. (1978). The acetone/butanol/ethanol fermentation. Process Biochem. 13: 2-5.

Squires, C.H., D.L. Heefner, R.J. Evans, B.J. Kopp, M.J. Yarus. (1984). Shuttle plasmids for Escherichia coli and Clostridium perfringens. J. Bacteriol. 159: 465-471.

Stal, M.H., H.P. Blaschek. (1985). Protoplast formation and cell wall generation in Clostridium perfringens. Appl. Environ. Microbiol. 50: 1097-1099.

Staden, R. (1980). A new computer method for the storage and manipulation of DNA gel reading data. Nucl. Acids Res. 8: 3673-3694.

Steel, R. (1958). In Biochemical engineering unit process in fermentation. pp.125-148. Macmillan Press, New York.

Stephens, M.A., N. Lang, K. Sandman, R. Losick. (1984). A promoter whose utilization is temporally regulated during sporulation in Bacillus subtilis. J. Mol. Biol. 176: 333-348.

Stokes, H.W., P.W. Betts, B.G. Hall. (1985). Sequence of the ebgA gene of Escherichia coli: comparison with the lacZ gene. Mol. Biol. Evol. 2: 469-477.

Stormo, G.D., T.D. Schneider, L.M. Gold. (1982). Characterization of translational initiation sites in Escherichia coli. Nucl. Acids Res. 10: 2971-2996.

Streitfeld, M.M., E.M. Hoffmann, H.M. Jonklow. (1962). Evaluation of extracellular deoxyribonuclease activity in Pseudomonas. J. Bacteriol. 84: 77-80.

Strom, M.S., M.W. Eklund, F.T. Poysky. (1984). Plasmids in Clostridium botulinum and related Clostridium species. Appl. Environ. Microbiol. 48: 956-963.

Strominger, J.L., D.J. Tipper. (1974). Structure of bacterial cell walls: the lysozyme substrate In: Lysozyme. pp.169-184. E.E. Osserman, R.E. Canfield, S. Beychok (ed.). Academic Press, New York.

Strominger, J.L. (1975). Actions of penicillin and other antibiotics on bacterial cell wall synthesis. In Biochemistry of cell walls and membranes. pp.207-227, C.F. Fox (ed.). Butterworths University Park Press.

Tabor, S., C.C. Richardson. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA 84: 4767-4771.

Tally, F.P., D.R. Snydman, S.L. Gorbach, M.H. Malamy. (1979). Plasmid mediated transferable resistance to clindamycin and erythromycin in Bacteroides fragilis. J. Infect.Diseases 139: 83-88.

Tanner, R.S., E. Stackebrandt, G.E. Fox, C.R. Woese. (1981). A phylogenetic analysis of Acetobacterium woodii, Clostridium bakei, Clostridium butyricum, Clostridium lituseburense, Eubacterium limosum, and Eubacterium tenue. Curr. Microbiol. 8: 35-38.

Taylor, J.M., R. Illmensee, J.S. Summers. (1976). Efficient translation of RNA into DNA by avian sarcoma virus polymerase. Biochem. Biophys. Acta. 442: 324-330.

Terzarghi, B.K., W.E. Sandine. (1975). Improved medium for lactic streptococci and their bacteriophages. Appl. Environ. Microbiol. 29: 807-813.

Thauer, R.K., K. Jungermann, K. Dekker. (1977). Energy conservation in chemotropic anaerobic bacteria. Bacteriol. Rev. 41: 100-180.

Thuring, R.W.J., P.M. Sanders, P. Borst. (1975). A freeze-squeeze method for recovering long DNA from agarose gel. Analyt. Biochem. 66: 213-220.

Tinson, W., A.J. Hillier, G.R. Jago. (1985). Metabolism of Streptococcus thermophilus. Aust.J.Dairy Technol. March 8-13, pp.8-13.

Tipper, D.J., J.J. Gauthier. (1972). Structure of the bacterial endospore, p.3-12. In Spores V. Ed. H.O. Halverson, R. Hansen, L.L. Campbell. Pub. Am. Society for Microbiology.

Trach, K.A., J.W. Chapman, P.J. Piggot, J.A. Hoch. (1985). Deduced product of the stage 0 sporulation gene spo0F shares homology with the spo0A, ompR, and sfrA proteins. Proc. Natl. Acad. Sci. USA 82: 7260-7264.

Truffaut, N., M. Sebald. (1983). Plasmid detection and isolation in strains of Clostridium acetobutylicum and related species. Mol. Gen. Genet. 189: 178-180.

Truffaut, N., M. Sebald. (1988). Study of plasmids isolated from saccharolytic clostridia: Construction of hybrid plasmids and transfer to Escherichia coli and Bacillus subtilis. Biotechnol. Lett. 10: 1-6.

Uchida, J., H. Shigematu, K. Yamafuji. (1965). The mode of action of hydrogen peroxide on deoxyribonucleic acid. Enzymologia 29: 369-376.

Urano, N., I. Karube, S. Suzuki, T. Yamada, H. Hirochika, K. Sakaguchi. (1983). Isolation and characterisation of large plasmids in hydrogen-evolving bacterium Clostridium butyricum. Eur. J. Appl. Microbiol. 17: 349-354.

Usdin, K.P., H. Zappe, D.T. Jones, D.R. Woods. (1986). Cloning, expression and purification of glutamine synthetase from Clostridium acetobutylicum. Appl. Environ. Microbiol. 52: 413-419.

Van der Westhuizen, A., D.T. Jones, D.R. Woods. (1982). Autolytic activity and butanol tolerance of Clostridium acetobutylicum. Appl. Environ. Microbiol. 44: 1277-1281.

Vieira, J., J. Messing. (1982). The pUC plasmids, an m13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19: 259-268.

Volesky, B., A. Mulchandani, J. Williams. (1981). Biochemical production of industrial solvents (acetone-butanol-ethanol) from renewable resources. Ann. N.Y. Acad. Sci. 369: 205-219.

Volesky, B., T. Szczesny. (1983). Bacterial conversion of pentose sugars to acetone and butanol. Adv. Biochem. Eng. Biotechnol. 27: 101-117.

Wahl, G.M., S.M. Stern, G.R. Stark. (1979). Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl paper and rapid hybridization by using dextransulfate. Proc. Natl. Acad. Sci. USA 76: 3683-3687.

Wako Pure Chemicals, Ltd. (1986). Bacteriolytic enzyme for biochemical use: Achromopeptidase (TBL-1). Code 014-09661.

Walker, P.D. (1970). Cytology of spore formation and germination. J. Appl. Bacteriol. 33: 1-12.

Walker, G.C. (1983). Genetic strategies in strain design for fermentations. p.349-376 In A. Hollaender, A.I Laskin, P. Rogers (ed.). Basic biology of new developments in biotechnology. Plenum Publishing Corp., New York.

Walker, G.C. (1984). Mutagenesis and inducible responses to deoxyribonucleic acid damage in Escherichia coli. Microbiol. Rev. 48: 60-93.

Walton, M.T., J.L. Martin. (1979). Production of butanol-acetone by fermentation. p.187-209 In H.J. Peppler, D. Perlman (ed.). Microbial technology, 2nd ed., vol. 1, Academic Press, Inc., New York.

Webster, J.R., S.J. Reid, D.T. Jones, D.R. Woods. (1981). Purification and characterization of an autolysin from Clostridium acetobutylicum. Appl. Environ. Microbiol. 41: 371-374.

Weickert, M.J., G.H. Chambliss, H. Sugiyama. (1986). Production of toxins by Clostridium botulinum type A strains cured of plasmids. Appl. Environ. Microbiol. 51: 52-56.

Welsh, F.W., I.A. Veliky. (1984). Production of acetone-butanol from acid whey. Biotechnol. Lett. 6: 61-64.

Welsh, F.W., I.A. Veliky. (1986). The metabolism of lactose by Clostridium acetobutylicum. Biotechnol. Lett. 8: 43-46.

Welsh, F.W., R.E. Williams, I.A. Veliky. (1986). A note on the effect of nitrogen source on growth of and solvent production by Clostridium acetobutylicum. J. Appl. Bacteriol. 61: 413-419.

Whitehead, T.R., J.C. Rabinowitz. (1986). Cloning and expression in Escherichia coli of the gene for 10-formyl tetrahydrofolate synthetase from Clostridium acidiurcci. J. Bacteriol. 167: 205-209.

Whitfeld, P.L., P.H. Seeburg, J. Shine. (1982). The human pro-opiomalanocortin gene: organization sequence, and interspersion with repetitive DNA. DNA 1: 113-143.

Williamson, V.M., J. Bennetzen, E.T. Young. (1980). Isolation of the structural gene for alcohol dehydrogenase by genetic complementation in yeast. Nature 283: 214-216.

Wood, C.F. (1966). Host specificity of DNA produced by Escherichia coli bacterial mutations affected the restriction and modification of DNA. J. Mol. Biol. 16: 118-123.

Woods, D.R. (1982). Molecular genetic studies on obligate anaerobic bacteria. S. African J.Sci. 78: 448-450.

Woods, D.R., D.T. Jones. (1984). Genetics of anaerobic bacteria. In: Herbivore nutrition in the subtropics and tropics. pp.297-312. F.M.C. Gilchrist, R.I. Mackie (ed.). The Science Press Ltd, South Africa.

Woods, D.R., D.T. Jones. (1986). Physiological responses of Bacteroides and Clostridium strains to environmental stress factors. Adv. Microb. Physiol. 27: 1-64.

Woods, H.G., R.W. Brown, C.H. Werkman. (1945). Mechanism of the butyl alcohol fermentation with heavy carbon acetic and butyric acids and acetone. Arch. Biochem. 6: 243-260.

Yamamoto, M., J.M. Jones, E. Senghas, C. Gawron-Burke, D.B. Clewell. (1987). Generation of Tn5 insertions in streptococcal conjugative transposon Tn916. Appl. Environ. Microbiol. 53: 1069-1072.

Yanisch-Perron, C., J. Vieira, J. Messing. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103-119.

Yokogawa, K., S. Kawata, T. Takemura, Y. Yoshimura. (1975). Purification and properties of lytic enzymes from Streptomyces globisporus 1829. Agr. Biol. Chem. 39: 1533-1543.

Yoshikawa, H., J. Kazami, S. Yamashita, T. Chibazakura, H. Sone, F. Kawamura, M. Oda, M. Isaka, Y. Kobayashi, H. Saito. (1986). Revised assignment for the Bacillus subtilis spo0F gene and its homology with spo0A and with two Escherichia coli genes. Nucl. Acids Res. 14: 1063-1072.

Yoshino S., S. Ogata, S. Hayashida. (1984). Regeneration of protoplasts of Clostridium saccharoperbutylacetonicum. Agric. Biol. Chem. 48: 249-250.

Young, M., J. Mandelstam. (1979). Early events during bacterial endospore formation. Adv. Microb. Physiol. 20: 102-162.

Youngleson, J.S., J.D. Santangelo, D.T. Jones, D.R. Woods. (1988). Cloning and expression of a Clostridium acetobutylicum alcohol dehydrogenase gene in Escherichia coli. Appl. Environ. Microbiol. 54: 676-682.

Youngman, P., P. Zuber, J.B. Perkins, K. Sandman, M. Igo, R. Losick. (1985). New ways to study developmental genes in spore-forming bacteria. Science 228: 285-290.

Yu, E.K.C., L. Deschatelets, J.N. Saddler. (1984a). The bioconversion of wood hydrolysates to butanol and butanediol. Biotechnol. Lett. 6: 327-332.

Yu, E.K.C., N.Levitin, J.N. Saddler. (1984b). Utilisation of wood hemicellulose hydrolysates by microorganisms for the production of liquid fuels and chemicals. Dev. Ind. Microbiol. 25: 613-620.

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

Yu, P.L., L.E. Pearce. (1986). Conjugal transfer of streptococcal antibiotic resistance plasmids into Clostridium acetobutylicum. Biotechnol. Lett. 8: 469-474.

Yu, P.L., J.B. Smart, B.M. Ennis. (1987). Differential induction of ß-galactosidase and phospho-ß-galactosidase activities in the fermentation of whey permeate by Clostridium acetobutylicum. Appl. Microbiol. Biotechnol. 26: 254-257.

Yu, R.T.S., W.S.A. Kyle, A.A. Azad, T.V. Hung. (1984c). Conditions for the production and regeneration of protoplasts from lactic streptococci. Milchwissenschaft 39: 136-139.

Zappe, H., D.T. Jones, D.R. Woods. (1986). Cloning and expression of Clostridium acetobutylicum endoglucanase, cellobiase and amino acid biosynthesis genes in Escherichia coli. J. Gen. Microbiol. 132: 1367-1372.

Zappe, H., D.T. Jones, D.R. Woods. (1987). Cloning and expression of a xylanase gene from Clostridium acetobutylicum P262 in Escherichia coli. Appl. Microbiol. Biotechnol. 27: 57-63.

Zappe, H., W.A. Jones, D.T. Jones, D.R. Woods. (1988). Structure of an endo-ß-1,4-glucanase gene from Clostridium acetobutylicum P262 showing homology with endoglucanase genes from Bacillus spp. Appl. Environ. Microbiol. 54: 1289-1292.

Zeikus, J.G. (1980). Chemical and fuel production by anaerobic bacteria. Ann. Rev. Microbiol. 34: 423-464.