STUDIES ON THE SUCROSE UTILIZATION SYSTEM OF *CLOSTRIDIUM BEIJERINCKII* NCIMB 8052

by

Neil Leat

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Faculty of Science, University of Cape Town, South Africa

July 1997

CERTIFICATION OF SUPERVISOR

In terms of paragraph **GP 8** of "General Rules for the degree of Doctor of Philosophy (Ph.D.)" I, as supervisor of the candidate Neil Leat, certify that I approve of the incorporation in this thesis of material that has already been published or submitted for publication.

Signed by candidate

Dr S.J. Reid
Department of Microbiology
University of Cape Town

ACKNOWLEDGMENTS

A simple and honest thanks to my supervisors Shez Reid and Dave Woods for their positive attitudes, ideas and constant encouragement. I would also like to thank Suhail Rafudeen, Helen Collett and Helen Stutz for creating an enjoyable working environment. Thanks to Ernest Clark for his tenacious ability to keep the lab stocked at all times. A special thanks goes to Di James for help with nucleotide sequencing and Greg Blatch and Val Abratt for proof-reading this thesis.

I am grateful to Wilf Mitchell and Kees Leenhouts for providing information on work prior to publication.

CONTENTS

ABSTRACT

Solventogenic clostridial strains were used extensively in the industrial acetone-butanol-ethanol (ABE) fermentation during the first half of this century. This fermentation was based predominantly on two substrates: corn mash (rich in starch) and molasses (rich in sucrose). Taxonomically distinct strains were used for ABE fermentations based on the two substrates. Although starch utilization by amylolytic solventogenic clostridial strains had been investigated, sucrose utilization by saccharolytic solventogenic clostridial strains had not been comprehensively studied. This provided the impetus for the present study. The primary aim of the project was to provide a molecular and physiological characterization of the sucrose utilization system of *Clostridium beijerinckii* NCIMB 8052.

The *C. beijerinckii* sucrose utilization operon, *scrARBK*, was cloned in three stages. In the first stage a *C. beijerinckii* genomic library was screened in *Escherichia coli* for clones able to utilize sucrose. Using this approach, a truncated *scr* operon was isolated. The remainder of the operon was isolated in the second and third stages by recovering a plasmid integrated into the *scrB* gene, and inverse-PCR respectively. The four genes of the *scrARBK* operon were proposed to encode an EnzymeIIBC^{sucrose} protein of the PTS (ScrA), a transcriptional repressor of the GalR-LacI family (ScrR), a sucrose-6-phosphate hydrolase of the glucosyl hydrolase family 32 (ScrB) and an ATP-dependent fructokinase of the ribokinase/*pfkB* sugar kinase family (ScrK).

Northern blot analysis confirmed that the individual genes of the *scr* operon were co-transcribed and that transcription was induced by sucrose. Disruption of the operon, by targeted plasmid integration into the *scrR* and *scrB* genes, resulted in *C. beijerinckii* strains unable to utilize sucrose. This suggested that the operon encoded the only sucrose inducible sucrose metabolic pathway in *C. beijerinckii*. Northern blot analysis also indicated that the disruption of *scrR* resulted in constitutive transcription of the 5'-undisrupted region of the operon. This suggested that *scrR* encoded a transcriptional repressor of the operon, and was therefore itself negatively autoregulated.

At the physiological level sucrose hydrolase and ATP-dependent fructokinase activities were found to be induced only in sucrose-grown *C. beijerinckii* cells. Induction of these activities by sucrose was consistent with transcriptional induction of the *scr* operon by sucrose.

Glucose-mediated repression of sucrose uptake, sucrose hydrolase and fructokinase activities was also demonstrated, however, the mechanism involved was not established.

From the results presented in this dissertation, and by analogy with previously studied sucrose utilization systems, it was proposed that sucrose utilization by *C. beijerinckii* involved PTS-dependent sucrose transport, yielding intracellular sucrose-6-phosphate. Sucrose-6-phosphate would in turn be hydrolyzed by a sucrose-6-phosphate hydrolase (sucrose hydrolase), releasing glucose-6-phosphate and fructose. Fructose would then be phosphorylated by an intracellular fructokinase. It was proposed that this pathway is encoded by the *scr* operon, and that transcription of the operon is regulated in response to sucrose by the *scrR* gene product.

ABBREVIATIONS

aa	amino acids
ABE	acetone-butanol-ethanol fermentation
Ap	ampicillin
ATCC	American Type Culture Collection
ATP	adenosine 5'-triphosphate
BSA	bovine serum albumin
bp	base pairs
C-	carboxy-(terminal)
CBM	Clostridial basal medium
CcpA	Catabolite control protein
CFE	cell free extract
CMM	Clostridial minimal medium
CoA	Co enzyme A
Da	Dalton
DDBJ	DNA database of Japan
DIG	digoxygenin
DNA	deoxyribonucleic acid
DNAse	deoxyribonuclease
DNS	dinitrosalicylic acid
DSM	Deutsche Sammlung von Mikroorganismen (German collection of Microorganisms)
EDTA	ethylenediaminetetra-acetic acid
EI	Enzyme I of the PTS
EIIA	Enzyme IIA domain of an EnzymeII^{sugar} protein of the PTS
EIIB	Enzyme IIB domain of an EnzymeII^{sugar} protein of the PTS
EIIC	Enzyme IIC domain of an EnzymeII^{sugar} protein of the PTS
Em	erythromycin
EMBL	European Molecular Biology Laboratory
EtBr	ethidium bromide
g	gram
g	standard gravitational acceleration
gb	GenBank (National Center for Biotechnology Information)
Glc	(superscript) glucose
h	hour(s)
HPr	histidine-containing or heat stable phosphocarrier protein of the PTS
kb	kilobase pair(s)
kDa	kilodaltons(s)
K_m	Michaelis constant
Km	kanamycin
LB	Luria-Bertani medium

MCS	multiple cloning site
min	minute(s)
mRNA	messenger RNA
MW	molecular weight in Dalton
N-	amino-(terminal)
NCBI	National Center for Biotechnology Information.
NCIMB	National Collection of Industrial and Marine Bacteria (Aberdeen, Scotland)
NCP	National Chemical Products Ltd. (Germiston, South Africa)
nm	nanometers
NRRL	National Center for Agricultural Utilization Research US Department of Agriculture.
OD_{600}	optical density measured at a wavelength of 600nm
ORF	open reading frame
p	plasmid
PCR	polymerase chain reaction
PEP	phosphoenolpyruvate
PIR	Protein Information Resource (National Biomedical Research Foundation Washington, USA)
P_{lac}	*lacZ* promoter on plasmid pSK Bluescript (Stratagene, San Diego)
P_R	phage lambda rightward promoter
PTS	phosphoenolpyruvate-dependent carbohydrate phosphotransferase system
R	(superscript) resistance
RNA	ribonucleic acid
RNAse	ribonuclease
SDS	sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
Scr	(superscript) sucrose
sp	SWISS-PROT
TAE	Tris-acetate-EDTA electrophoresis buffer
Tris	tris(hydroxymethyl)aminomethane
w/v	weight per volume (in grams per 100ml)
YT	yeast tryptone broth
α	alpha
β	beta
Δ	delta
λ	lambda
μ	micro
σ	sigma
::	novel joint fusion

CHAPTER 1

GENERAL INTRODUCTION

Contents

1.1. INTRODUCTION TO *C. BEIJERINCKII* AND THE SOLVENTOGENIC CLOSTRIDIA

Clostridium beijerinckii NCIMB 8052 is a mesophilic, obligately anaerobic, spore-forming bacterium, that produces acetone, butanol and ethanol when grown on sugary substrates. It is closely related to clostridial strains that were used in industrial acetone-butanol-ethanol (ABE) fermentations based on molasses. To place *C. beijerinckii* NCIMB 8052 in context, the history of the ABE fermentation and taxonomy of the solventogenic clostridia will be briefly considered. The general characteristics of *C. beijerinckii* will then be discussed, and this will be followed by a review of bacterial sucrose utilization. It should be noted that the designation *C. beijerinckii* NCIMB 8052 has only recently been assigned, and that this strain was previously classified as *Clostridium acetobutylicum* NCIMB 8052 (Keis *et al.*, 1995). Hereafter the designation *C. beijerinckii* NCIMB 8052 or simply *C. beijerinckii* will be used.

1.1.1. History of the ABE fermentation

The outbreak of the First World War resulted in a demand for acetone, which was used as a colloidal solvent for nitrocellulose in the production of munitions. This lead to the initiation of the first industrial ABE fermentation using a solventogenic clostridial strain, at the facilities of Strange and Graham Ltd. (United Kingdom), in 1916 (for reviews see Johnson and Chen (1995); Jones and Keis (1995); Jones and Woods (1986)). The strain used was initially designated BY and subsequently renamed *Clostridium acetobutylicum* (McCoy *et al.*, 1926). After the first world war, ABE fermentations were operated by the Commercial Solvents Corporation (USA) under a world-wide patent, valid until 1935. All of the early ABE fermentations were based on corn mash, a substrate rich in starch. In the 1930's, however, an abundant supply of low-cost molasses provided the impetus for the development of a molasses-based ABE fermentation. Since the original strains were unable to completely ferment the sucrose in molasses, new solventogenic clostridial strains were isolated. The lapse of the Commercial Solvents Corporation patent in 1935, further stimulated the search for new strains, as several companies implemented ABE fermentations. New strains were named primarily on the basis of patent considerations, and this resulted in a plethora of strain names emerging with no clear understanding of the relationships between the strains.

During the 1950's the ABE fermentation faced increasing competition for solvent production from the developing petrochemical industry. In addition, molasses prices continued to increase, eventually resulting in the demise of the fermentation. At present, it is operated only

in the People's Republic of China, although there has been a renewal of interest in the fermentation as a means of producing solvents from renewable resources (Woods, 1995). The development of a molasses-based fermentation using *Clostridium saccharoperbutylacetonicum* was considered for Brazil in the early 1990's, where low cost molasses and improved fermentation processes, were proposed to make the fermentation economically viable (Afschar *et al.*, 1990).

1.1.2. Taxonomy of the solventogenic clostridia

With the demise of the ABE fermentation many of the strains isolated in the 1930's and 1940's were lost and many of those remaining were classified as strains of *C. acetobutylicum*. Some of the more commonly studied strains include: *C. acetobutylicum* ATCC 824 (Serving as the ATCC *C. acetobutylicum* type strain); *C. acetobutylicum* DSM 792 (Serving as the DSM *C. acetobutylicum* type strain); *C. beijerinckii* NCIMB 8052 (which until recently, incorrectly served as the NCIMB *C. acetobutylicum* type strain); *Clostridium saccharoperbutylacetonicum* N1-4 (also referred to as *C. acetobutylicum* N1-4) and *C. acetobutylicum* NCP 262.

As a result of extensive research performed on the various solventogenic clostridial strains, many physiological and genetic differences were observed. These studies suggested that the various strains designated as *C. acetobutylicum*, may in fact represent several different species (Wilkinson *et al.*, 1995a). These differences have been noted by Johnson and Chen (1995) and include:

(i) Differences in the degree to which strains are able to sustain solventogenesis in continuous culture. *C. beijerinckii* NCIMB 8052 rapidly loses the ability to produce solvents and sporulate, in a process referred to as strain degeneration (Kashket and Cao, 1995; Stephens *et al.*, 1985). In contrast, *C. acetobutylicum* ATCC 824 and DSM 1731 are able to sustain solvent production in continuous culture (Woolley and Morris, 1990).

(ii) Differences in the pH at which solventogenesis is initiated and sustained. For *C. acetobutylicum* strains ATCC 824 and DSM 792, sustained solvent production occurs at or below pH 5.5. In contrast, *C. beijerinckii* NCIMB 8052 is capable of producing solvents in cultures maintained at pH 7 (Holt *et al.*, 1984)

(iii) A specific restriction endonuclease system is present in *C. acetobutylicum* ATCC 824 cleaving at 5'-GCNGC-3'(Mermelstein and Papoutsakis, 1993a). No evidence exists for a similar system in *C. beijerinckii* NCIMB 8052.

(iv) Significant differences in genome size and restriction endonuclease fragment profiles were revealed, using pulsed-field gel electrophoresis of genomic DNA extracted from *C. saccharoperbutylacetonicum* N1-4, *C. acetobutylicum* ATCC 824 and *C. beijerinckii* NCIMB 8052 (Wilkinson and Young, 1993).

Differences identified between the strains stimulated research into the taxonomic relationships between them. These studies have used biotyping, involving rifampin and bacteriocin resistance assays, as well as bacteriophage susceptibility typing (Keis *et al.*, 1995). In addition, reassociation of genomic DNA from the different strains was used as a measure of overall DNA sequence similarity (Johnson and Chen, 1995). Genomic DNA fingerprint analysis, using pulsed-field gel electrophoresis (Keis *et al.*, 1995; Wilkinson and Young, 1993) and 16S ribosomal RNA sequence analysis were also employed (Collins *et al.*, 1992; Keis *et al.*, 1995; Wilkinson *et al.*, 1995a). These studies allowed for the identification of four taxonomic groups, with the proposal that each group represents a distinct species (Keis *et al.*, 1995).

Relationships between representatives of the four taxonomic groups are illustrated in Figure 1.1. Taxonomic group 1 is represented by *C. acetobutylicum* ATCC 824 and consists of amylolytic strains used in corn mash fermentations. Taxonomic group 2 is represented by *C. acetobutylicum* NCP262, taxonomic group 3 is represented by *C. saccharoperbutylacetonicum* N1-4 and taxonomic group 4 is represented by *C. beijerinckii* NCIMB 8052. It is noteworthy that the species isolated reflect the changing requirements of the ABE fermentation during the century. Initially, amylolytic solventogenic strains from group 1 were used in corn mash fermentations. These strains were replaced by saccharolytic strains, from groups 2, 3 and 4, as molasses became the preferred fermentation substrate.

Research has been conducted into the utilization of starchy substrates by some of the members of taxonomic group 1, particularly *C. acetobutylicum* ATCC 824 (Verhasselt *et al.*, 1989; Verhasselt and Vanderleyden, 1993; Annous and Blaschek, 1990, 1991; Paquet *et al.*, 1991). In contrast, comprehensive studies on sucrose utilization by saccharolytic solventogenic

clostridial strains, have not been reported. This provided the impetus for the present study, which was aimed at characterizing the sucrose utilization system of *C. beijerinckii* NCIMB 8052. In the sections which follow, the general characteristics of *C. beijerinckii* NCIMB 8052 will be considered. This will be followed by a review of bacterial sucrose utilization.

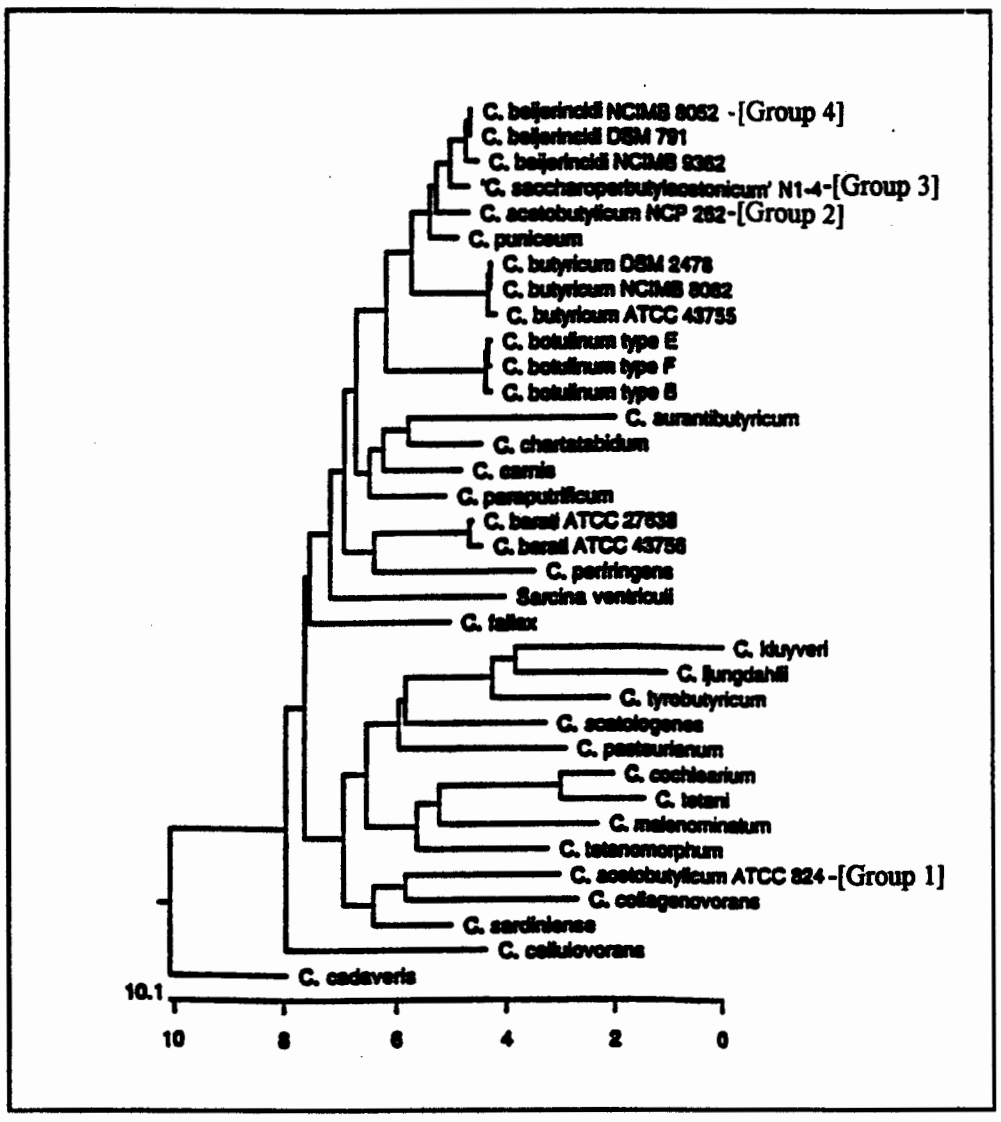

Fig. 1.1 Unrooted phylogenetic dendrogram showing the taxonomic positions of: group 1 representative *C. acetobutylicum* ATCC 824; group 2 representative *C. acetobutylicum* NCP 262; group 3 representative *C. saccharoperbutylacetonicum* N1-4 and group 4 representative *C. beijerinckii* NCIMB 8052. The dendrogram was based on 16S ribosomal RNA gene sequences (positions 100 to 1434; *E. coli* numbering). The scale bar indicates evolutionary distance (taken from Kies *et al.*, 1995).

1.1.3. *C. beijerinckii* growth, solventogenesis and differentiation.

The growth of solventogenic clostridial strains, including *C. beijerinckii,* is typically associated with three phases: an initial growth phase, involving highly motile rod-shaped cells and the production of acetic and butyric acids; a solventogenic phase, involving nongrowing, nonmotile, clostridial-shaped cells, which reassimilate acetic and butyric acids, and produce

acetone and butanol; and a third phase involving sporulation and the inhibition of solventogenesis (Jones and Woods, 1986; Woods, 1995).

Acidogenic and solventogenic pathways are relatively well understood and have been reviewed (Jones and Woods, 1986; Rogers and Gottschalk, 1993; Dürre *et al.*, 1995). These pathways will be considered in three components: firstly central reactions, involving intermediates common to acidogenesis and solventogenesis (c to f in Fig. 1.2); secondly, reactions involved in acidogenesis (g to j in Fig. 1.2) and thirdly, reactions involved in solventogenesis (k to p in Fig. 1.2). The central pathway intermediates acetyl-CoA, acetoacetyl-CoA and butyryl-CoA are the key metabolites from which both solvents and acids are derived. A thiolase (acetyl-CoA acetyltransferase) catalyses the formation of acetoacetyl-CoA from acetyl-CoA. Butyryl-CoA is derived from acetoacetyl-CoA as a result of the activities of three enzymes *viz:* 3-hydroxybutyryl-CoA dehydrogenase; crotonase and butyryl-CoA dehydrogenase. During acidogenic metabolism, acetate and butyrate are derived from acetyl-CoA and butyryl-CoA respectively in parallel reactions. A phosphotransacetylase and phosphotransbutyrylase generate acetyl phosphate and butyryl phosphate respectively. The relevant acids are derived from their phosphorylated intermediates by acetate and butyrate kinases, in reactions yielding ATP. In the case of *C. beijerinckii* two clustered genes designated *ptb* and *buk* have been cloned and sequenced and encode a phosphotransbutyrylase and butyrate kinase respectively (Oultram *et al.*, 1993).

During solventogenic metabolism acetate and butyrate are reassimilated, yielding central pathway intermediates from which acetone and butanol are derived. CoA-transferase (Acetoacetyl-CoA:acetate/butyrate:CoA transferase) plays a central role in catalyzing the reassimilation of acetate and butyrate, in reactions generating central pathway intermediates acetyl-CoA and butyryl-CoA. CoA-transferase also plays an important role in the formation of acetone, by catalyzing the formation of acetoacetate from the central pathway intermediate acetoacetyl-CoA. Acetoacetate is then decarboxylated by acetoacetate decarboxylase yielding acetone and CO_2. Acetone may be further metabolized to isopropanol by alcohol dehydrogenases. Butanol is generated by the reduction of butyryl-CoA involving butyraldehyde and butanol dehydrogenases. Ethanol is generated by the reduction of acetyl-CoA, involving acetyldehyde and ethanol dehydrogenases.

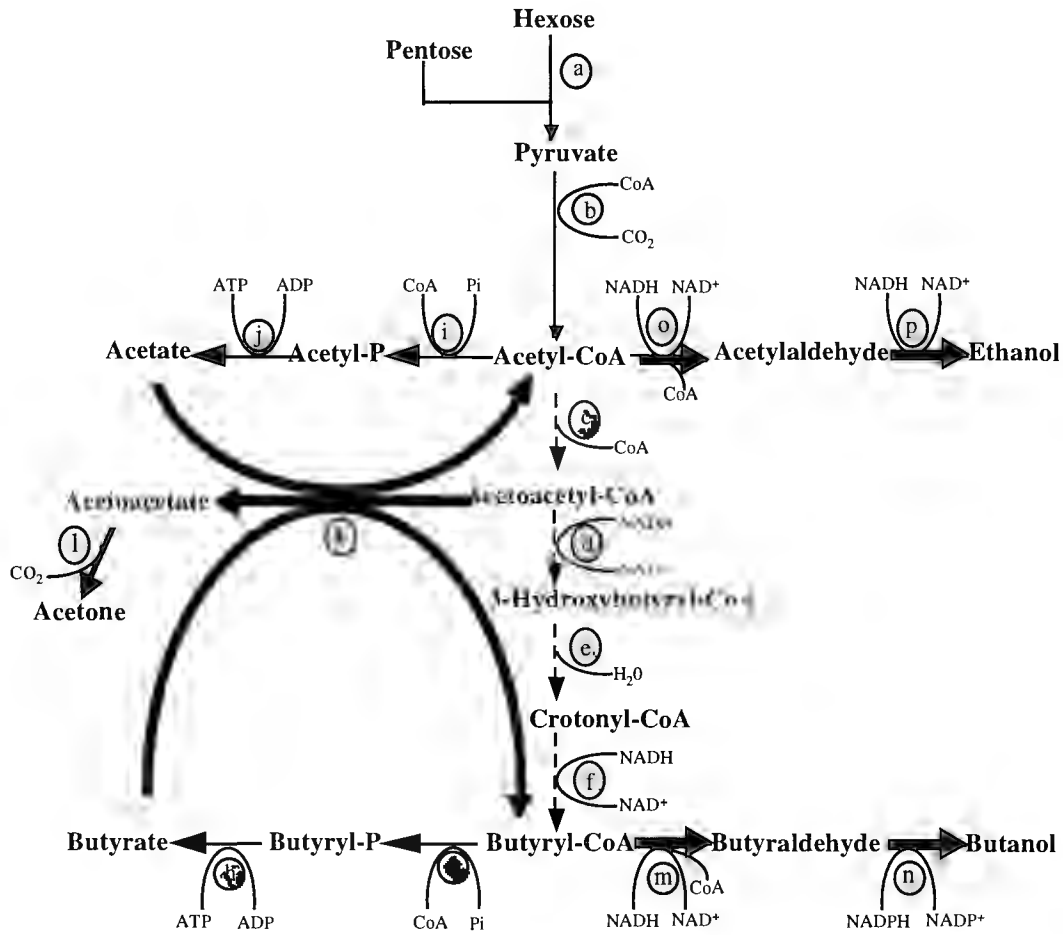

Fig. 1.2 Simplified representation of the branched acidogenic and solventogenic pathways in solventogenic clostridia.

[a] Initial metabolism of hexoses and pentoses occurs via the Embden-Meyerhof and pentose phosphate pathways respectively, ultimately yielding pyruvate. [b] Acetyl-CoA is generated by the activity of a Pyruvate-Ferredoxin Oxidoreductase

Central pathways are indicated by vertical dashed arrows and catalyzed by: [c] thiolase (acetyl-CoA acetyltransferase); [d] 3-hydroxybutyryl-CoA dehydrogenase; [e] crotonase; [f] butyryl-CoA dehydrogenase.

Acidogenic pathways are indicated by thin horizontal arrows and catalyzed by: [g] phosphotransbutyrylase; [h] butyrate kinase; [i] phosphotransacetylase; [j] acetate kinase.

Solventogenic pathways are indicated by thick arrows and catalyzed by: [k] acetoacetyl-CoA:acetate/butyrate:CoA transferase; [l] acetoacetate decarboxylase; [m] butyraldehyde dehydrogenase; [n] butanol dehydrogenase; [o] acetyldehyde dehydrogenase; [p] ethanol dehydrogenase. (After Woods, 1995; Collett, 1996; and Chen, 1993).

Factors involved in the regulation of solventogenic metabolism and sporulation are only beginning to be elucidated for *C. beijerinckii* (Wilkinson *et al.*, 1995b). *C. beijerinckii* often fails to initiate solventogenesis and sporulation in a process referred to as strain degeneration (Kashket and Cao, 1995). This process is of interest because it may reflect on regulatory mechanisms controlling the initiation of solventogenesis and sporulation. In addition, it poses a limitation on the potential use of *C. beijerinckii* in ABE fermentations.

Strain degeneration occurs in both batch and continuous culture. In batch culture, it is likely that the rate of acid production exceeds the rate at which solventogenesis is initiated, with the result that the culture pH drops to lethal levels. Evidence in support of this comes from the fact that if *C. beijerinckii* is cultured under conditions leading to a reduced growth rate, such as in minimal medium, strain degeneration is less likely to occur (Kashket and Cao, 1995). Presumably, a reduced growth rate allows for solventogenesis to be initiated, before lethal acidification of the growth medium can occur. In continuous culture, however, *C. beijerinckii* also ultimately gives rise to a population of degenerate mutants. These mutants vary with respect to their capacity to initiate solventogenesis, sporulation and the accumulation of granulose (a glucose polymer). In one particular class of mutants, all three parameters were affected (Woolley and Morris, 1990). These mutants completely reverted to the wild-type phenotype at a frequency of about $5x\ 10^{-6}$. This lead to the suggestion that a single global regulatory locus was involved (Woolley and Morris, 1990).

Attempts to characterize such a regulatory locus have employed transposon mutagenesis and an analysis of a *C. beijerinckii* homologue of the *B. subtilis spo0A* (Kashket and Cao, 1993; Brown *et al.*, 1994). Transposon mutagenesis was used to produce a degeneration resistant *C. beijerinckii* strain carrying a single Tn*1545* insertion (Kashket and Cao, 1993). It was proposed that a global regulatory locus had been affected. However, evidence has yet to be reported, clearly explaining the mechanism whereby the transposon insertion results in a resistance to strain degeneration. More promising is the work conducted on the *C. beijerinckii spo0A* gene (Brown *et al.*, 1994; Wilkinson *et al.*, 1995b). The study was prompted by the fact that the *B. subtilis spo0A* encodes a phosphorylation-activated transcription factor involved in the regulation of stationary phase metabolism (Brown *et al.*, 1994). A *spo0A* homologue was isolated from *C. beijerinckii* using PCR. Its deduced product was proposed to consist of an N-terminal phospho-acceptor domain and a C-terminal DNA-binding-effector domain (Brown *et al.*, 1994; Wilkinson *et al.*, 1995b). Disruption of *spo0A* in *C. beijerinckii* resulted in a strain unable to initiate granulose accumulation, solventogenesis and sporulation and this suggested that its gene product played a role in the regulation of these functions (Wilkinson and Young, 1994; Wilkinson *et al.*, 1995b). It was hypothesized that the *spo0A* gene product would regulate gene expression by binding to '0A box' sequences (5'-TGNCGAA) present in the 5'-regions of regulated genes. Putative '0A boxes' were identified for several genes involved in solventogenic metabolism (Wilkinson *et al.*, 1995b).While *spo0A* appears to encode a

regulator of solventogenesis and sporulation, the precise signals to which it responds, and the details of the regulatory mechanisms involved, remain to be established.

1.1.4. Genetic aspects of *C. beijerinckii*

A physical map of the *C. beijerinckii* genome has been generated using a "Top Down" approach (Wilkinson and Young, 1995). This initially involved the mapping of extremely rare restriction endonuclease sites, *Sfi*I and *Rsr*II, (3 and 6 sites respectively) to give a low resolution map. Subsequently, an attempt was made to map more common sites, *Apa*I and *Sma*I (more than 30 sites each). Twenty-seven genetic loci, corresponding to thirty-two protein-encoding genes, were located on the map. This work indicated that *C. beijerinckii* has a relatively large 6.7-Mbp circular genome with approximately fourteen ribosomal RNA operons (*rrn*) operons, nine of which were clustered around a region proposed to encode the origin of replication. High numbers of *rrn* operons are thought to facilitate a high growth rate, reflecting the situation in *C. beijerinckii*, which has a mean generation time of approximately 30 minutes under optimal conditions (Wilkinson and Young, 1995).

Although *C. beijerinckii* is devoid of endogenous plasmids, it supports replication of vectors containing replicons from pAMβ1 (*Enterococcus faecalis*), pCB101 (*Clostridium butyricum*) and pWV01 (*Streptococcus cremoris*) (Williams *et al.*, 1990b). It has been argued that, where plasmid replication takes place via a single-stranded intermediate, as in the case of pCB101 and pWV01, the plasmids are structurally less stable than when replication does not involve a single stranded intermediate (pAMβ1) (Minton *et al.*, 1993; Oultram *et al.*, 1988). The only reliable antibiotic for use with *C. beijerinckii* is erythromycin or erythromycin derivatives such as clarithromycin, which has the added benefit of being stable in low pH environments (Minton *et al.*, 1993; Mermelstein and Papoutsakis, 1993b).

Transformation of *C. beijerinckii* has been achieved using electroporation, where transformation efficiencies ranged between 8.0×10^1 and 2.9×10^3 transformants/μg plasmid DNA (Oultram *et al.*, 1988). Unfortunately these relatively low transformation efficiencies limited the usefulness of this technique. Subsequently, a more efficient conjugative plasmid transfer system, based on the broad-host-range conjugation system encoded by IncP plasmids, was developed (Williams *et al.*, 1990b). This technique relied on an IncP helper plasmid, R702 (Hedges and Jacob, 1974), providing conjugation functions *in trans* to a series of mobilizable *E. coli*/*C. beijerinckii* shuttle plasmids (Williams *et al.*, 1990b). The shuttle

plasmids carried: Gram-negative and Gram-positive replicons; Ap^R and Em^R markers; and *oriT* sites (the sites at which the DNA duplex is nicked, prior to transfer of a single strand from donor to recipient). This approach provided for the efficient mobilization of shuttle plasmids from *E. coli* to *C. beijerinckii* and facilitated the subsequent development of a protocol for the targeted integration of plasmids into the *C. beijerinckii* genome.

Integration of plasmids into the *C. beijerinckii* genome was achieved using a mobilizable shuttle plasmid (pMTL30) as described above, except for the omission of a Gram-positive replicon. Short (0.34-3.8kb) regions of *C. beijerinckii* 'target' DNA were cloned into pMTL30 and conjugated into *C. beijerinckii*. Plasmid maintenance was dependent on a Campbell-like recombination event between the 'target' DNA, carried by pMTL30, and its equivalent on the *C. beijerinckii* genome (Campbell, 1962; Wilkinson and Young, 1994). Using this approach, both single and amplified plasmid copies became established in the *C. beijerinckii* genome. The frequency of nonreplicative plasmid establishment ranged between 2.6×10^{-7} to 1.7×10^{-6} transconjugants per recipient, with no clear correlation between the size of homologous DNA provided and the frequency of plasmid establishment. Since Campbell-like integration events generate a duplication of target DNA sequences, a reversal of the integration event was possible. It was found that in the absence of erythromycin selection, erythromycin-sensitive recombinants segregated at a frequency of between 3.7×10^{-4} to 1.3×10^{-3} per generation.

The *C. beijerinckii gutD* and *spo0A* genes, encoding a sorbitol 6-phosphate dehydrogenase and a phosphorylation-activated transcription factor respectively, were disrupted by targeted plasmid integration. In both cases disruption occurred because the 'target' DNA corresponded to a region entirely within the coding sequence of the respective gene. The strain carrying a disrupted *gutD* was unable to grow on media where sorbitol was supplied as the sole carbon source. As indicated above, the strain carrying a disrupted *spo0A* was unable to initiate granulose accumulation, solventogenesis and sporulation. A similar approach has been used to disrupt the phosphotransacetylase, butyrate kinase and aldehyde/alcohol dehydrogenase encoding genes from *C. acetobutylicum* ATCC 824, where plasmid transfer was achieved by electroporation (Green and Bennett, 1996; Green *et al.*, 1996).

1.1.5. Carbohydrate utilization by *C. beijerinckii*.

Clostridial carbohydrate metabolism, and the cloning of carbohydrate catabolic genes from solventogenic clostridial strains, has been reviewed (Mitchell *et al.*, 1995; Verhasselt and

Vanderleyden, 1993). The objective here is focus on work conducted on *C. beijerinckii* NCIMB 8052. It should be noted that although the PTS of *C. beijerinckii* is considered below, a more detailed discussion of the PTS is presented, in the context of sucrose uptake, in section 1.2.1. This section should be consulted for a basic explanation of the relationships between PTS components.

A glucose PTS was indicated for *C. beijerinckii*, by the presence of PEP-dependent glucose phosphorylation activity in cell free extracts (Mitchell *et al.*, 1991). The K_m for glucose was found to be 34μM. Extract reconstitution assays were conducted to characterize the glucose PTS. These experiments involved the preparation of membrane and cytoplasmic extracts by ultracentrifugation. Cytoplasmic extracts were then fractionated with respect to size, by gel filtration chromatography. The various fractions were combined with extracts from *pts* mutants of *E. coli* and *B. subtilis* and assayed for sugar phosphorylation. The work demonstrated, firstly, that the *C. beijerinckii* glucose PTS consisted of four components (Enzyme I, HPr, $EIIA^{Glc}$ and a membrane associated EII^{Glc}), secondly, that the components of the *C. beijerinckii* PTS were able to interact with the PTS systems of *E. coli* and *B. subtilis*, and thirdly, that the cytoplasmic components of the *C. beijerinckii* glucose PTS were approximately the same size as their equivalents in other bacteria.

Studies on the metabolism of several other sugars by *C. beijerinckii* have been aimed at: (i) establishing whether a PTS-dependent or non-PTS system is involved (usually by assaying for PEP-dependent sugar phosphorylation in *C. beijerinckii* extracts); (ii) establishing the relationship between supplied carbohydrates and the induction of sugar uptake systems and (iii) examining the effect of glucose on the utilization of other sugars. The results of these experiments have been presented in three reports (Albasheri and Mitchell, 1995; Mitchell *et al.*, 1995 and Mitchell, 1996) and an attempt has been made to summarize the information in T ables 1.1. and 1.2. PTS-mediated sugar uptake appears to be the predominant carbohydrate transport mechanism in *C. beijerinckii*, and in addition to the glucose PTS, PTS-dependent transport mechanisms have been specifically proposed for fructose, sucrose, glucitol, mannitol and lactose. Non-PTS uptake systems, however, are thought to play a role in galactose and maltose uptake.

Mitchell (1996) examined the relationship between supplied carbohydrates and the induction of sugar uptake systems (Table 1.1). The data obtained indicated that in cases such as glucitol,

the corresponding uptake system was induced only by the substrate involved, while in contrast, uptake systems for glucose appeared to be constitutively expressed. This work provides the first step towards identifying which carbohydrate utilization systems are regulated in *C. beijerinckii*. Further investigation is required to identify the physiological and molecular genetic mechanisms resulting in the observed regulation.

Table 1.1. Uptake of carbohydrates by whole cells of *C. beijerinckii*[a]

Growth Substrate	Presence of uptake activity for indicated carbohydrate			
	Glucose	Fructose	Glucitol	Galactose
Glucose	+	+	-	-
Fructose	+	+	-	-
Glucitol	+	+	+	-
Galactose	+	-	-	+

[a] The information presented here has been compiled from Mitchell (1996)

The effect of glucose, on the utilization of other sugars, has also been investigated (Albasheri and Mitchell, 1995; Mitchell *et al.*, 1995; and Mitchell, 1996). This was achieved by growing 'induced' precultures in presence of a sugar of interest and 'uninduced' precultures in presence of glucose. Precultures were then inoculated into media containing both glucose and the sugar of interest. The disappearance of glucose and the sugar of interest was then monitored during growth. Three patterns were observed, and are summarized here as complete repression, delayed repression and no repression. Complete repression refers to a situation in which glucose completely repressed utilization of a second sugar. Delayed repression refers to a situation where glucose and the sugar of interest were initially utilized simultaneously. However, as growth proceeded, glucose utilization continued, while utilization of the second sugar was repressed (This typically occurred when an 'induced' preculture was used). In the third scenario, no repression was observed and both glucose and the carbohydrate of interest were utilized simultaneously.

The results for various sugars are summarized in Table 1.2., however, it must be stressed that original reports should be consulted for an appreciation of detail. It is evident that the utilization of various sugars is affected differently by the presence of glucose. Fructose utilization, for example, is not repressed by the presence of glucose, while glucitol utilization is completely repressed. This work provides the first step towards identifying which

carbohydrate utilization systems are subject to glucose-mediated catabolite repression. However, further analysis is required to establish the mechanisms involved.

Table 1.2. Repression of carbohydrate uptake by the presence of glucose[b].

	Uninduced preculture[a]			Induced preculture[a]		
Carbohydrate of interest mixed with glucose	Complete repression	Delayed repression	No effect	Complete repression	Delayed repression	No effect
Fructose			+			+
Sucrose	+				+	
Maltose	+				+	
Lactose	+			+		
Cellobiose	+			+		
Glucitol	+			+		

[a]'Uninduced' precultures were prepared by growth in the presence of glucose. 'Induced' precultures were prepared by growth in the presence of the carbohydrate of interest (Fructose, Sucrose, Maltose, Lactose, Cellobiose, Glucitol). Precultures were then inoculated into media containing glucose and the carbohydrate of interest. The disappearance of glucose and the carbohydrate of interest was then assayed over several hours.
[b] The information presented here has been compiled from three reports (Albasheri and Mitchell, 1995; Mitchell *et al.*, 1995; and Mitchell, 1996)

Catabolite repression mechanisms, in low-GC Gram-positive bacteria, are only beginning to be clearly elucidated (Saier *et al.*, 1995 and 1996). The PTS appears to play a central role, and since this may well be the case in *C. beijerinckii*, it will be considered briefly. HPr, from low-GC Gram-positive bacteria, can be phosphorylated at both a conserved histidine and a conserved serine residue (Ser-46 for the *B. subtilis* HPr). Histidyl phosphorylation is involved in the transfer of phosphoryl groups, ultimately to PTS sugars. Serine phosphorylation plays a role in catabolite repression and is catalyzed by an HPr(ser) kinase. HPr(ser) kinase activity is stimulated by the presence of metabolites such as fructose-1,6-bisphosphate and gluconate-6-phosphate (Saier *et al.*, 1996). The phosphorylated form of HPr, HPr(ser-P), has been implicated in: (i) direct repression of sugar permeases, possibly via an allosteric binding mechanism; (ii) activation of sugar phosphatases as a prelude to inducer expulsion and (iii) direct repression of catabolic gene transcription. The latter is thought to occur as HPr(ser-P) interacts with a catabolite control protein (CcpA) of the GalR-LacI family. CcpA in turn represses catabolic gene expression by binding to catabolite-responsive elements (CREs) present in the 5'-regions of regulated genes (Saier *et al.*, 1996). It is worth noting that genes encoding CcpA-like proteins have been cloned from *C. acetobutylicum* P262 and *C. beijerinckii* NCIMB 8052 (Davison *et al.*, 1995; Rafudeen, unpublished results, University of Cape Town, Microbiology Dept.). These genes and their gene products are currently under investigation.

1.2. GENERAL ASPECTS OF BACTERIAL SUCROSE UTILIZATION

It has been suggested that bacterial sucrose utilization systems evolved by modular evolution, where individual genes sharing a common origin became independently associated into regulons and operons (Titgemeyer *et al.*, 1996). Individual components common to bacterial sucrose utilization systems include: (i) sucrose transport systems, typically involving the PTS (ii) enzymes catalyzing sucrose hydrolysis, which may also be involved in the synthesis of glucans and fructans; (iii) ATP-dependent fructokinases and (iv) regulatory proteins, typically of the GalR-LacI family of transcriptional regulators. A schematic representation of how these components are involved in sucrose utilization is presented in Figure 1.3. The essential features of these components will be considered briefly. This will be followed by a more detailed analysis of well characterized bacterial sucrose utilization systems. Less common components of bacterial sucrose utilization systems, such as regulatory transcriptional antiterminator proteins, and non-PTS sucrose uptake systems, will be considered in the context of the systems in which they occur.

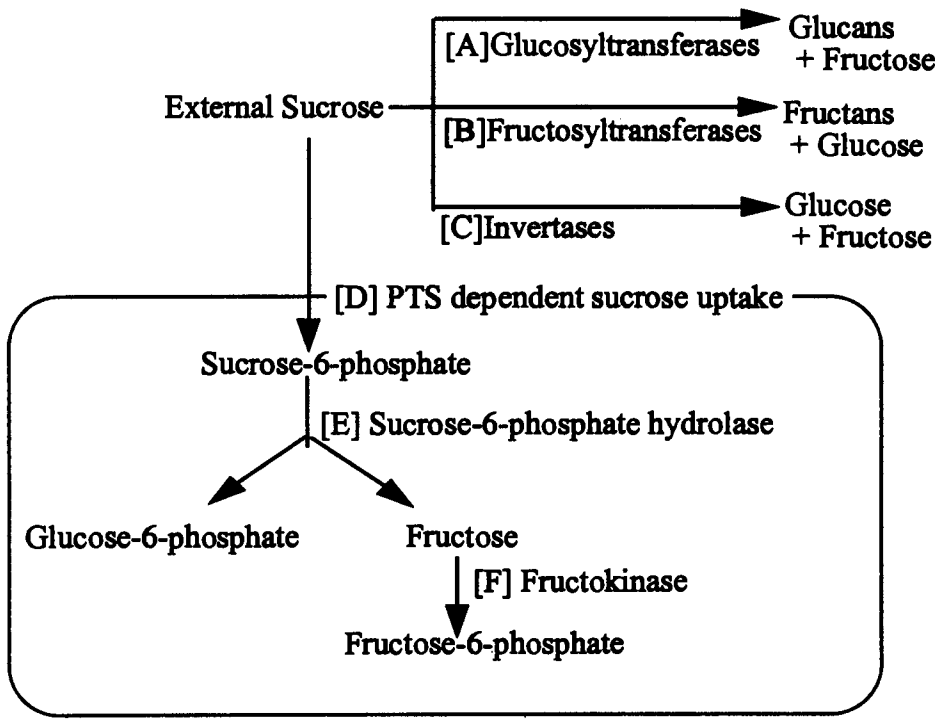

Fig 1.3 Schematic representation of typical bacterial sucrose metabolic pathways. [A] Extracellular sucrose may be hydrolyzed by glucosyltransferases with the subsequent production of glucans and free fructose. [B] Extracellular sucrose may be hydrolyzed by fructosyltransferases with the consequent formation of fructans and free glucose. [C] Extracellular sucrose may hydrolyzed by extracellular invertases yielding extracellular glucose and fructose. [D] Extracellular sucrose may be taken up by the PTS yielding intracellular sucrose-6-phosphate. (It should be noted that non-PTS sucrose uptake systems such as the *E. coli* EC3132 sucrose/H^+ symporter exist (Bockmann *et al.*, 1992)). [E] Sucrose-6-phosphate is hydrolyzed by a sucrose-6-phosphate hydrolase yielding glucose-6-phosphate and fructose. [F] Free intracellular fructose is then phosphorylated by a fructokinase. Many of the bacterial sucrose utilization systems studied to date are permutations of the generalized system outlined above.

1.2.1. The carbohydrate phosphotransferase system (PTS) and sucrose transport

The PTS appears to be the most common system facilitating sucrose transport. The PTS catalyses the concomitant uptake and phosphorylation (group translocation) of its sugar substrates (PTS sugars), and has been studied extensively (For reviews see Lengeler *et al.*, 1994; Saier and Reizer, 1992; Saier and Reizer, 1994; Postma *et al.*, 1993). The PTS typically consists of the central components, Enzyme I and HPr (histidine-containing phosphocarrier protein), shared by several sugar specific membrane associated proteins (Enzyme II proteins). Enzyme II proteins typically consist of three domains EIIA, EIIB and EIIC. Transport and phosphorylation of PTS sugars occurs as follows: PEP is the source of a phosphoryl group which is passed from PEP via Enzyme I, HPr, $EIIA^{sugar}$, $EIIB^{sugar}$ to the incoming sugar (Fig. 1.4). The passage of the respective sugar through the membrane is facilitated by the integral membrane domain ($EIIC^{sugar}$).

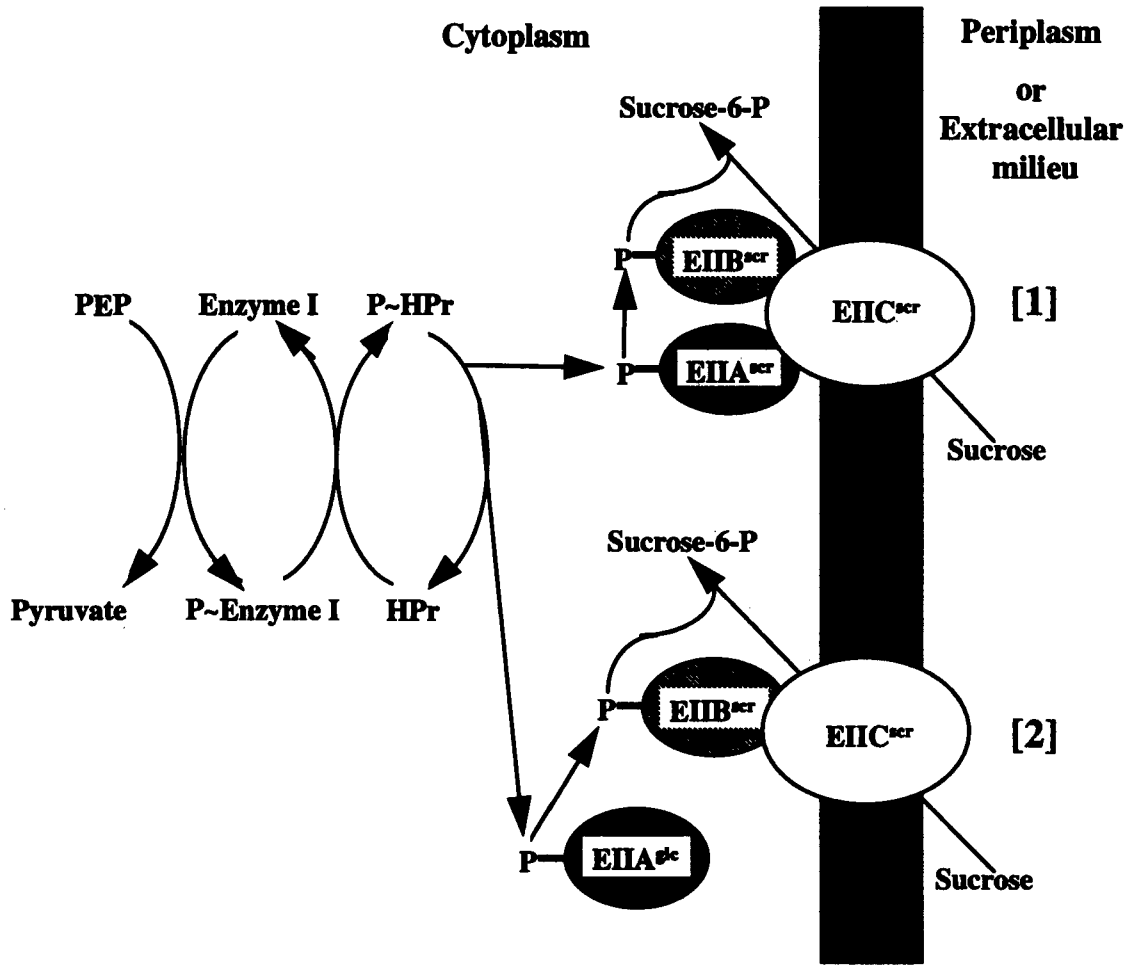

Fig. 1.4 Organization of sucrose specific EII^{Scr} proteins. The EII^{Scr} proteins correspond to [1] those found in *Streptococcus mutans, Pediococcus pentosaceus and Streptococcus sobrinus,* where the EIIA EIIB and EIIC domains are fused to form $EIIBCA^{Scr}$ proteins and [2] those found in *B. subtilis, Staphylococcus xylosus, Vibrio alginolyticus, Klebsiella pneumoniae,* where only the EIIB and EIIC domains are fused to form $EIIBC^{Scr}$ proteins. It should be noted that only sucrose-specific EII proteins are considered here. For EII proteins, specific for other sugars, several EIIA EIIB and EIIC domain combinations exist. (After Postma *et al.*, 1993)

Enzyme I proteins range between 65000-85000Da per monomer and autophosphorylate a conserved histidyl residue with PEP as the phosphoryl donor. HPr proteins range between 9000 and 10000Da and are phosphorylated, by the phosphorylated form of Enzyme I, at a conserved histidyl residue. In addition, HPr from low-GC Gram-positive bacteria can be phosphorylated at a conserved serine residue. The role of serine phosphorylation in the regulation of catabolite repression, was considered in section 1.1.5.

The organization of EnzymeII protein domains varies, and may consist of a single fused protein or fused and unfused domains. In the case of sucrose-specific Enzyme II proteins, two varieties exist (Fig. 1.4). Those of *S. mutans, P. pentosaceus and S. sobrinus*, where the $EIIA^{Scr}$ $EIIB^{Scr}$ and $EIIC^{Scr}$ domains are fused to form $EIIBCA^{Scr}$ proteins, and those of *B. subtilis, V. alginolyticus, S. typhimurium, K. pneumoniae* and *S. xylosus* where only the $EIIB^{Scr}$ and $EIIC^{Scr}$ domains are fused, forming $EIIBC^{Scr}$ proteins. In cases where $EIIBC^{Scr}$ proteins exist, evidence has been presented suggesting that the $EIIB^{Scr}$ subunit is phosphorylated by the $EIIA^{Glc}$ domain of the respective glucose PTS (Blatch, 1990; Lengeler *et al.*, 1982; Sprenger and Lengeler, 1988; Sutrina *et al.*, 1990).

A two dimensional model for the *K. pneumoniae* $EIIBC^{Scr}$ protein has been proposed, comprising of six transmembrane helices with three short periplasmic loops and three large intracellular domains (Fig. 1.5) (Titgemeyer *et al.*, 1996). The N-terminal loop 1 corresponds to the EIIB domain, containing the highly conserved sequence (HCATRLR). This conserved sequence includes the cysteine residue phosphorylated by a phosphorylated EIIA domain. Since this domain interacts with a soluble EIIA, loop 1 should be on the cytoplasmic side of the inner membrane and loop 2 on the periplasmic side. This is in agreement with the periplasmic location of a *phoA* (encoding alkaline phosphatase) fusion at residue 152 of the $EIIBC^{Scr}$ of *V. alginolyticus* (Blatch *et al.*, 1990). For unknown reasons, the residues of helix 1 are highly conserved while the residues of the remaining helices are not (Titgemeyer *et al.*, 1996). Intracellular loop 5 contains a highly conserved histidine residue as well as a conserved GITE motif. Mutations which uncouple transport and phosphorylation are located on this loop and it has been suggested that it is involved in substrate binding, phosphorylation and perhaps substrate translocation (Lengeler *et al.*, 1994).

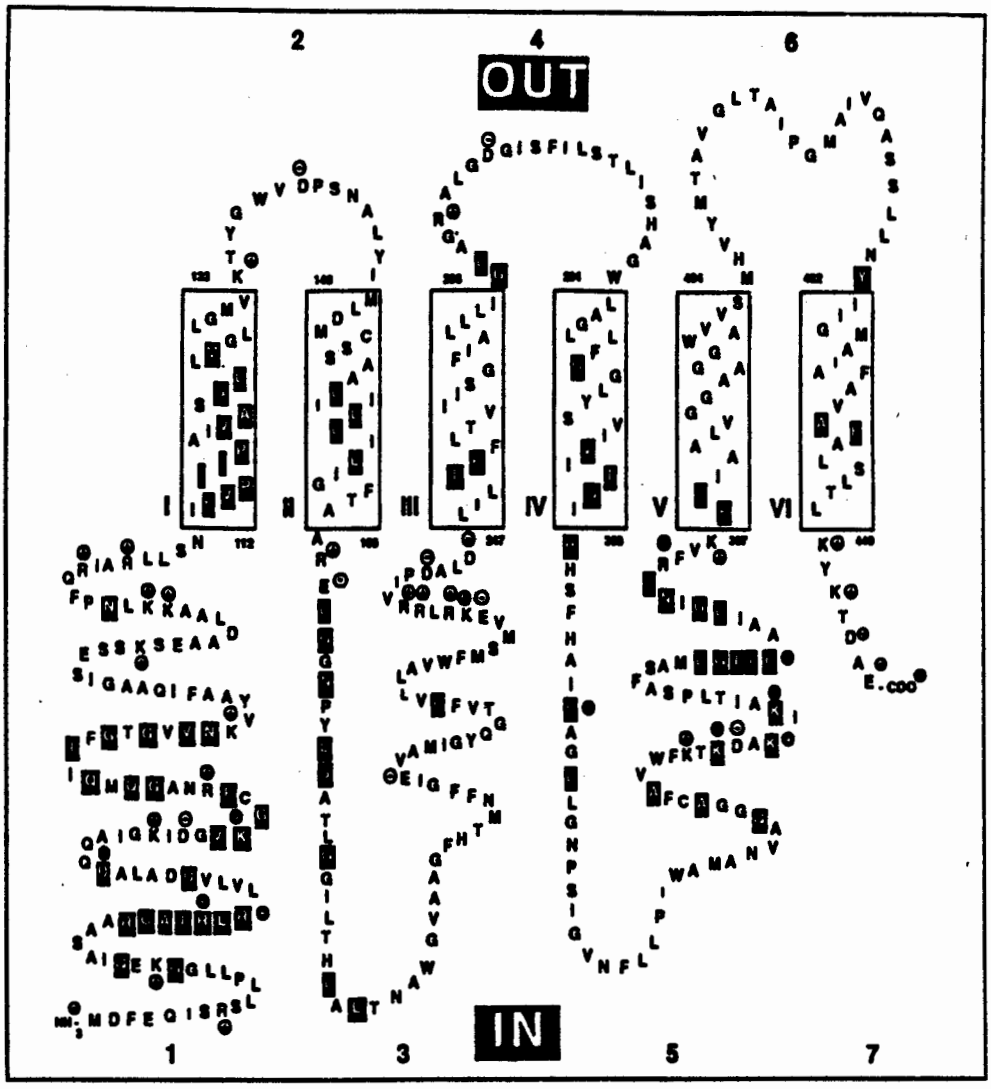

Fig. 1.5 Two-dimensional model of the EIIBC^{Scr} of *K. pneumoniae*. The model is based on the comparison of seven functional EIIBC^{Scr} proteins from *K. pneumoniae*, *S. typhimurium* (encoded by plasmid pUR400), *B. subtilis*, *V. alginolyticus*, *S. sobrinus*, *S. mutans* and *S. xylosus*. Residues conserved in all EIIBC^{Scr} proteins are highlighted in black boxes. Six putative transmembrane helices are boxed and numbered in Roman numerals (I-VI). Cytoplasmic (1,3,5,7) and periplasmic (2,4,6) loops are numbered in Arabic numerals. The positions of the first and last residues in the helices are indicated. The numbering is for the EIIBC^{Scr} sequence from the *K. pneumoniae* sucrose utilization system. (Taken from Titgemeyer *et al.*, 1996)

1.2.2. Bacterial sucrose hydrolysis

The β-fructofuranosidases of the glucosyl hydrolase family 32 constitute possibly the largest group of enzymes capable of sucrose hydrolysis. Examples are included from archaebacteria, eubacteria and eukaryotes (Henrissat, 1991; Henrissat and Bairoch, 1993; Henrissat and Bairoch, 1996; 'http: //expasy.hcuge.ch/cgi-bin/lists?glycosid.txt'). These enzymes catalyze the hydrolysis of terminal fructose moieties from fructose-containing disaccharides, trisaccharides and polymers, with varying specificity. Some examples such as levanases and inulinases, have increased specificity against high molecular weight fructose polymers (β(2-6)-linked levans and β(2-1)-linked inulins). Others specifically hydrolyze low molecular

weight fructose containing sugars such as sucrose and raffinose. Within this group are sucrose-6-phosphate hydrolases associated with PTS-dependent sucrose utilization systems. These enzymes typically hydrolyze both sucrose and sucrose-6-phosphate, but have a significantly lower K_m for sucrose-6-phosphate (Lunsford and Macrina, 1986; Thompson *et al.*, 1991a; Schmid *et al.*, 1982).

Several bacterial enzymes catalyze not only sucrose hydrolysis, but additional reactions as well. Glucosyltransferases catalyze the hydrolysis of sucrose, with the formation of glucose polymers (glucans) and release of free fructose. The nature of the glucan synthesized differs depending on the glucosyltransferase involved and both water-soluble (predominantly α(1-6) linked glucans) and water-insoluble (predominantly α(1-3) linked glucans) are produced. Insoluble glucans are resistant to microbial degradation and in the case of oral streptococci, are thought to play a role in the colonization of smooth dental surfaces. Soluble glucans are susceptible to microbial degradation and are thought to provide an extracellular energy reserve.

Fructosyltransferases catalyze the hydrolysis of sucrose with the subsequent production of high molecular weight fructose polymers. Depending on the enzyme involved, the individual residues of the polymer may be linked predominately by β(2-6) bonds (levans) or β(2-1) bonds (inulin). High molecular weight fructose polymers, like soluble glucans, serve as an extracellular energy reserve.

Sucrose phosphorylases catalyze the simultaneous hydrolysis of sucrose and phosphorylation, using inorganic phosphate, of the resulting glucose, to yield glucose-1-phosphate and fructose. Examples of sucrose phosphorylases have been identified in several bacteria, including *S. mutans* and *Leuconostoc mesenteroides* (Koga *et al.*, 1991; Russell *et al.*, 1988).

1.2.3. Fructokinases associated with sucrose metabolism

Kinases which catalyze the phosphorylation of sugars have been divided into three non-homologous families, referred to as the hexokinase, ribokinase and galactokinase families, based on differences in conserved sequence patterns (Bork *et al.*, 1993). The ribokinase family has also been referred to as the *pfkB* family in the Prosite collection of sequence patterns, and therefore, the designation ribokinase/*pfkB* family will be used. Members of the hexokinase family have been included in what Titgemeyer *et al.* (1994), referred to as the

Repressor, ORF, Kinase (ROK) family. Therefore the designation ROK/hexokinase will be used. Fructokinase activity has evolved independently, by convergent evolution, in both the ribokinase/*pfkB* and the ROK/hexokinase families (Bork *et al.*, 1993). Fructokinases from both families catalyze the same basic reaction (ATP + fructose → ADP + fructose-6-phosphate) and examples from both families have been shown to be require the presence of divalent cations, usually Mg^{2+}, for activity (Fennington and Hughes, 1996, Thompson *et al.*, 1992, Thompson *et al.*, 1991b).

Ribokinase/*pfkB* fructokinases, range in size between 307 and 331aa and include several examples encoded within sucrose utilization regulons or operons (*K. pneumoniae, S. typhimurium* and *V. alginolyticus*) as well as examples encoded independently (*Rhizobium leguminosarium* and *B. subtilis*) (Aulkemeyer *et al.*, 1991; Blatch *et al.*, 1990; Fennington and Hughes, 1996; Yoshida *et al.*, 1994). In addition, eukaryotic fructokinases from *Beta vulgaris* (GenBank accession U37838), *Lycopersicon esculentum* (GenBank accession U62329) and *Solanum tuberosum* (Smith *et al.*, 1993) are members of the ribokinase/*pfkB* family.

The ROK/hexokinase family includes fructokinases from: *Zymomonas mobilis*; *P. pentosaceus*; *S. mutans*; *Fusobacterium mortiferum* and *Lactococcus lactis*. In addition, a glucokinase from *Streptomyces coelicolor* and transcriptional repressor proteins from *B. subtilis* (XylR), *S. xylosus* (XylR) and *E. coli* (NagC) are included (Titgemeyer *et al.*, 1994). It should be noted that the inclusion of fructokinases from *F. mortiferum* and *L. lactis* is tentative. They have been included here because the short N-terminal sequences available for these proteins are homologous to the N-terminal sequence of a ROK/hexokinase family fructokinase from *S. mutans* (Sato *et al.*, 1993b).

ROK/hexokinase fructokinases range in size between 288 and 301aa and have been proposed to associate as dimers (Thompson *et al.*, 1991b; Thompson *et al.*, 1992). In the cases of *P. pentosaceus*, *S. mutans* and *L. lactis*, fructokinases are encoded by genes clustered in sucrose utilization regulons. In the case of *Z. mobilis*, the fructokinase encoding gene, *frk*, is not associated with a sucrose utilization regulon (Zembrzuski *et al.*, 1992). In the case of *F. mortiferum*, complete sequence information is not yet available (Thompson *et al.*, 1992).

Although the transcriptional repressors and sugar kinases of the ROK/hexokinase family share extensive homology, they differ in that the repressors all have an 80aa N-terminal helix-turn-

helix DNA binding domain. A parallel has been drawn between the evolution of ROK/hexokinase transcriptional repressors and the evolution of the GalR-LacI family of regulatory proteins. In the case of the GalR-LacI family, it has been proposed that the family arose as a result of the association of a periplasmic sugar binding protein with a helix-turn-helix DNA binding motif (Nguyen and Saier, 1996). In the case of the ROK/hexokinase transcriptional repressors, it was hypothesized that a hexokinase-like sugar kinase domain became associated with a helix-turn-helix DNA binding domain, yielding a family of sugar binding regulatory proteins (Titgemeyer *et al.*, 1994).

1.2.4. Features of the GalR-LacI family of transcriptional regulators

Transcriptional regulation by GalR-LacI family regulatory proteins constitutes the predominant mechanism regulating the expression of sucrose catabolic genes. The GalR-LacI family of transcriptional regulators was initially reviewed by Weickert and Adhya (1992), and has been more recently reviewed by Nguyen and Saier (1995). To date, more than 25 members have been identified and crystal structures have been determined for the LacI and purine biosynthesis (PurR) repressors of *E. coli* (Schumacher *et al.*, 1994; Friedman *et al.*, 1995). The members have a characteristic size of 320-341aa, except for the ScrR protein of *V. alginolyticus* which has been reported to be only 94aa in size (Blatch, 1990; Blatch and Woods 1991; Weickert and Adhya, 1992). All of the members have an N-terminal helix-turn-helix DNA binding motif and the remainder of the protein is involved in ligand binding, dimerization and, in the case of LacI, tetramerization.

Operator sites for GalR-LacI regulators are typically imperfectly palindromic nucleotide sequences, with the central bases ($A_4A_3N_2C_1/G_1'N_2'T_3'T_4'$) more highly conserved than outer bases (Weickert and Adhya, 1992). Regulatory dimers bind operator sites via their helix-turn-helix domains, with each subunit binding a half site. Regulation of gene expression is typically achieved by transcriptional repression where unliganded regulator dimers bind to operator sites downstream or within the promoters involved. Ligand binding ultimately results in a conformational change of the helix-turn-helix domain and dissociation of the regulator/DNA complex. It should be noted, however, that in some cases the liganded form of the regulator binds its operator site (e.g. *E. coli* PurR) (Schumacher *et al.*, 1994). In addition, examples exist which are able to effect both transcriptional repression and activation, such the Catabolite Repressor/Activator (Cra) protein of Enteric Bacteria (Saier and Ramseier, 1996).

In this case, the position of the operator site relative to the respective promoter appears to determine whether Cra acts as a repressor or activator of transcription.

As indicated above, the family is thought to have originated as a result of a fusion between a precursor of periplasmic sugar binding proteins and a helix-turn-helix DNA binding motif (Vartak *et al.*, 1991; Tam and Saier, 1993; Nguyen and Saier, 1995). Phylogenetic tree construction for 25 sequenced GalR-LacI family members, revealed twenty branches, all stemming from a point close to the center of the tree (Nguyen and Saier, 1995). This lead to the suggestion that shortly after the DNA binding domain became associated with the ligand binding domain, there was a marked pressure for gene duplication events, giving rise to the GalR-LacI family. Subsequently, however, the members of the family appear to have evolved independently of one another with relatively few gene duplication events. It was also proposed that the highly conserved helix-turn-helix motif has diverged less from the primordial DNA binding sequence than ligand binding domains, which have diversified to bind a range of ligands (Nguyen and Saier, 1995).

1.3. SUCROSE UTILIZATION SYSTEMS IN BACTERIA

1.3.1 *Bacillus subtilis*

Sucrose utilization by *B. subtilis* has been studied extensively (Kunst *et al.*, 1990; Steinmetz, 1993). Two pathways exist which are directly involved in sucrose metabolism. The first involves an extracellular fructosyltransferase (levansucrase/SacB). This enzyme catalyzes the hydrolysis of sucrose, and synthesis of levans (β(6-2)-linked fructose polymers with β(1-2) branches) (Chambert and Gonzy-Treboul., 1976; Steinmetz *et al.*, 1985). The second pathway involves an EIIBC^{Scr} protein (SacP), and a sucrose-6-phosphate hydrolase of the glucosyl hydrolase family 32. A third sucrose-hydrolyzing enzyme, a non-specific β-D-fructofuranosidase (levanase/SacC), is also expressed (Kunst *et al.*, 1990; Wanker *et al.*, 1995). Since the levanase is capable of hydrolyzing high molecular weight fructose polymers and is encoded as the distal gene in a fructose inducible fructose-PTS operon, it seems likely that its primary role would be in the metabolism of fructose polymers. For detailed studies on the fructose-PTS/levanase operon see Martin-Verstraete *et al.* (1990); Martin-Verstraete *et al.* (1994); Martin-Verstraete *et al.* (1995) and Stülke *et al.* (1995). The remainder of this section will deal with the molecular structure and regulation of the *B. subtilis sac* genes, directly involved in sucrose metabolism.

The molecular organization of *B. subtilis sac* genes is well understood (Fig. 1.6). The fructosyltransferase (levansucrase/SacB) is encoded by *sacB* while the EIIBC^{Scr} protein (SacP) and the sucrose-6-phosphate hydrolase (SacA) are encoded by the *sacPA* operon (Gay *et al.*, 1983; Steinmetz *et al.*, 1985; Fouet *et al.*, 1986; Fouet *et al.*, 1987). Expression of *sacB* is fully induced by sucrose concentrations higher than 30mM, while *sacPA* operon expression is induced by sucrose concentrations lower than 1mM (Steinmetz *et al.*, 1989). In addition, *sacPA* expression is subject to repression by glucose (Arnaud *et al.*, 1992). Regulation in response to sucrose occurs predominately by transcriptional antitermination at ribonucleotide antiterminator (RAT)/terminator sequences (Debarbouille *et al.*, 1990; Aymerich and Steinmetz., 1992; Rutberg, 1997). These sequences are located in the leader regions of the *sacB* and *sacPA* operons (Fig. 1.6 and Fig. 1.7). The SacT and SacY proteins are responsible for antitermination at the *sacPA* and *sacB* RAT/terminator sequences respectively. SacT and SacY are 48% identical to each other and 35% identical to the BglG antiterminator protein of *E. coli* (Debarbouille *et al.*, 1990). The *sacT* gene lies upstream of the *sacPA* operon (Fig. 1.6). The *sacY* gene is encoded as part of the *sacXY* operon where *sacX* specifies an EIIBC^{Scr}-like protein with 56% identity to SacP (Zukowski *et al.*, 1990).

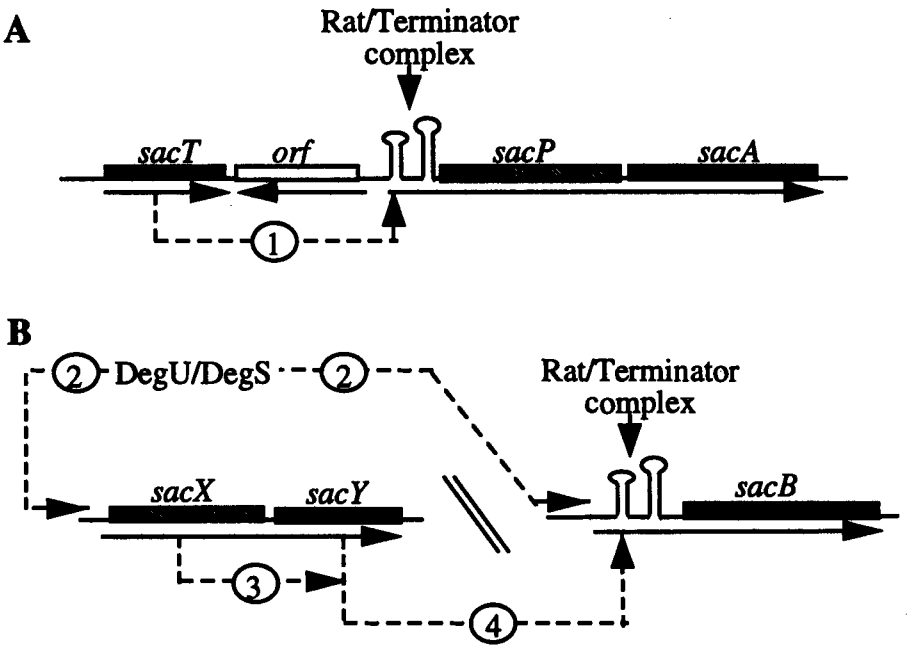

Fig. 1.6 Structure of [A] the *B. subtilis sacPA* operon and adjacent *sacT* gene and [B] the *sacB* gene and unlinked *sacXY* operon. Solid arrows indicate transcriptional polarity and, where a single arrow extends over two genes, co-transcription is thought to occur. Dashed lines and arrows indicate regulatory relationships and roles of: (1) SacT as an antiterminator active at the *sacPA* RAT/Terminator complex; (2) the DegU/DegS signaling system in increasing transcription off the *sacXY* and *sacB* promoters; (3) SacX as a negative regulator of the antitermination activity of SacY and (4) SacY as an antiterminator at the *sacB* RAT/terminator complex. It should be noted that SacT and SacY also regulate *sacXY* transcription in response to sucrose. (Not drawn to scale. Adapted from Steinmetz, 1993)

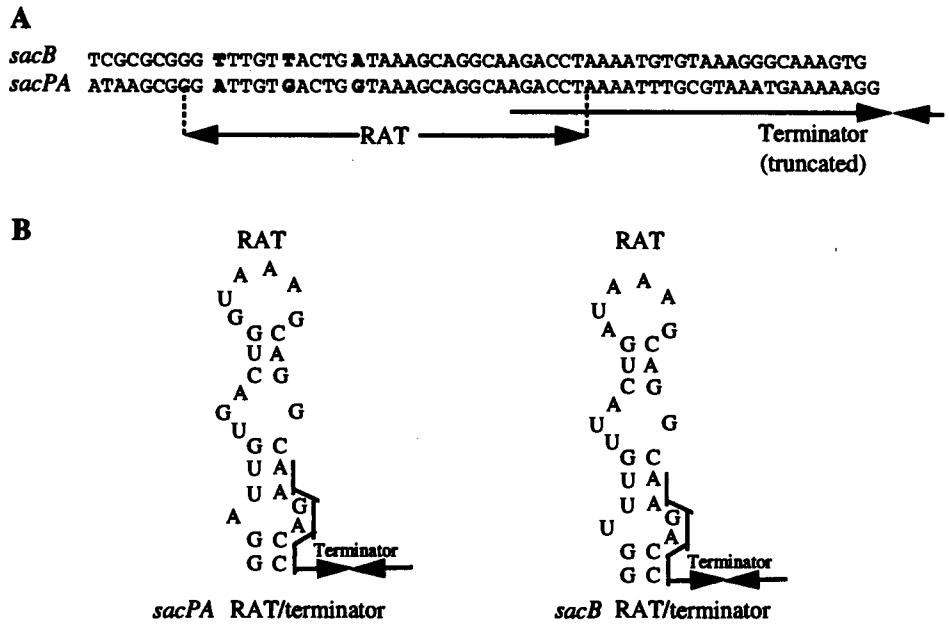

Fig 1.7 [A] Sequence of *sacB* and *sacPA* RAT/terminator elements. Note that the terminator component overlaps the RAT element and that the *sacB* and *sacPA* RAT sequences differ in only three positions, indicated by bold text [B] Predicted secondary structure of the *sacB* and *sacPA* RAT elements. In each case the adjacent terminator is indicated by opposed arrows. After Aymerich and Steinmetz (1992)

Mechanisms by which RAT/terminator sequences operate, have been proposed (Aymerich and Steinmetz., 1992). Both the RAT and adjacent terminator are predicted to form stem-loop structures (Fig. 1.7). The predicted RAT stem loop structures are usually substantially weaker (ΔG = -4 to -5 kcal/mol) than the adjacent terminator stem loop structures (ΔG = -18 to -30 kcal/mol) (Aymerich and Steinmetz, 1992). It has been suggested that an activated antiterminator protein (SacY or SacT) stabilizes the RAT stem-loop in the nascent mRNA, precluding the formation of the terminator structure. In this state, full length transcripts are produced. In the absence of an activated antiterminator protein, the terminator structure would form, resulting in early transcription termination.

As indicated above, the SacY and SacT proteins effect transcriptional termination at the *sacB* and *sacPA* RAT/terminator complexes respectively. However, factors influencing SacY and SacT activity have not been completely characterized. Analogy with the homologous BglG antitermintor of *E. coli* suggests that phosphorylation/dephosphorylation may play a role (Rutberg, 1997). In the case of SacY, a single site is thought to be phosphorylated by the PTS via the EIIBC^{Scr}-like protein, SacX (Crutz *et al.*, 1990; Rutberg, 1997). It seems likely that the availability of sucrose would determine the extent to which SacX phosphorylates SacY (Crutz *et al.*, 1990). In the presence of sucrose, SacX would not phosphorylate SacY. The dephosphorylated form of SacY would be active and facilitate antitermination at the *sacB*

RAT/terminator sequence. In the absence of sucrose, SacX would phosphorylate SacY. The phosphorylated form of SacY would be inactive and early termination would occur at the *sacB* RAT/terminator sequence (Crutz *et al.*, 1990).

Factors affecting the regulation of SacT are slightly more complicated. Unlike SacY, it was demonstrated that an intact PTS was essential for *in vivo* activity of SacT (Arnaud *et al.*, 1992). Models developed suggested the presence of two phosphorylation sites (Rutberg, 1997; Arnaud *et al.*, 1992). One of the sites would be analogous to that described for SacY above, modulating antitermination activity in response to sucrose concentration. The phosphoryl donor in this case has not been clearly identified. The other site would act as an 'activation' site and would be phosphorylated by the PTS proteins, Enzyme I and HPr. Recently, it was demonstrated that purified SacT could be phosphorylated, *in vitro*, by Enzyme I and HPr (Arnaud *et al.*, 1996). However, the phosphorylation state of SacT did not affect its activity. Therefore, the precise role of HPr/Enzyme I mediated phosphorylation at an 'activation' site remains unclear.

SacT and SacY are also involved in the induction of the *sacXY* operon, in response to sucrose (Crutz and Steinmetz, 1992). However, the mechanism involved is not entirely clear, since the RAT sequence identified ahead of the *sacXY* transcriptional start site, is some distance (100bp) from the nearest transcriptional terminator structure.

Mechanisms other than transcriptional antitermination are also involved in the regulation of sucrose metabolism in *B. subtilis*. Expression off both the *sacB* and *sacXY* promoters is induced by the DegU/DegS signaling system (Crutz and Steinmetz, 1992; Shimotsu and Henner, 1986). In addition, sucrose uptake appears to be regulated by a form of inducer exclusion. It has been demonstrated, *in vitro*, that the $EIIB^{Scr}$ domain of SacP, is phosphorylated by the $EIIA^{glucose}$ domain of the $EIICBA^{glucose}$ protein (Sutrina *et al.*, 1990). Furthermore, glucose was found to be preferentially phosphorylated over sucrose, in an *in vitro* milieu containing both carbohydrates. It would follow then that glucose may be preferentially transported over sucrose, as the EIIB domains of the corresponding EII proteins compete for phosphorylated $EIIA^{glucose}$ (Sutrina *et al.*, 1990).

1.3.2. *Staphylococcus xylosus*

A PTS-dependent sucrose metabolic pathway has been studied in *S. xylosus*, a Gram-positive nonpathogenic bacterium, isolated from human skin (Brückner *et al.*, 1993; Gering and Brückner, 1996; Wagner *et al.*, 1993). It was established that PTS-dependent sucrose uptake and intracellular sucrose hydrolase activities were induced by the presence of sucrose and that sucrose hydrolase activity was repressed by glucose. Genes involved in sucrose metabolism include: *scrA*, encoding a 480aa $EIIBC^{Scr}$ of the PTS; *scrR*, encoding a 320aa GalR-LacI family regulatory protein and *scrB*, encoding a 494aa sucrose-6-phosphate hydrolase of the glucosyl hydrolase family 32 (Fig. 1.8) (Gering and Brückner, 1996; Wagner *et al.*, 1993). The *scrA* gene has a transcriptional start site 46 to 49bp upstream of its initiation codon and is not associated with the *scrR and scrB* genes. The *scrR* and *scrB* genes are clustered, with the termination codon of the *scrR* ORF 84bp from the start codon of the *scrB* ORF. Despite the proximity of the *scrR* and *scrB* genes, they are essentially independently transcribed (Gering and Brückner, 1996). Transcriptional start sites have been identified, 33bp upstream of the *scrR* ORF and 59bp in upstream of the *scrB* ORF.

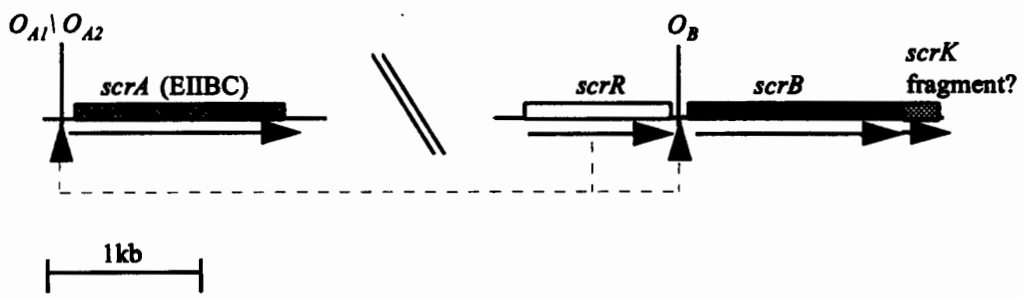

Fig. 1.8 Molecular organization of the *S. xylosus* sucrose utilization system. The genes involved are essentially independently transcribed, with transcriptional polarity indicated by solid arrows. Dashed lines and arrows indicate regulatory relationships. In the absence of sucrose, ScrR prevents transcription of *scrB* and *scrA* by binding to the O_B and the O_{A1}/O_{A2} operator sites. In the presence of sucrose ScrR is released from the operator sites and transcription proceeds. During the preparation of this dissertation the 140bp nucleotide sequence reported for the region downstream of *scrB* was analyzed. It was proposed that this region may encode a truncated ribokinase/*pfkB* family fructokinase, *scrK*.

During the preparation of this dissertation, the 140bp nucleotide sequence, reported for the region downstream of *scrB*, was examined. The region specified a truncated protein (46aa), with an initiation codon overlapping the *scrB* termination codon. The deduced sequence had distinct homology with the N-terminal regions of ribokinase/*pfkB* family fructokinases (29.7% identity with the *K. pneumoniae* ScrK to 39.5% identity with a fructokinase from *Solanum tuberosum*). Furthermore, two domains, conserved in the N-terminal regions of ribokinase/*pfkB* family fructokinases, (GEALID) and (GGAPCNVA), were identified (see Chapter 2, Figure 2.12 for a multiple sequence alignment of ribokinase/*pfkB* family fructokinases). It would follow that a gene encoding a fructokinase lies immediately

downstream of *scrB*. By convention this gene would be designated *scrK*. If this analysis is correct, it would follow that the *scr* gene cluster consists of, not two, but three genes, *scrR*, *scrB* and *scrK*.

The *scrA*, *scrR* and *scrB* genes have been individually disrupted in the genome of *S. xylosus*. The *scrA* gene was disrupted using transposon mutagenesis and the *scrB* and *scrR* genes were disrupted by insertion of the *ermB* gene into the respective ORFs. The *scrA* mutant was deficient in sucrose uptake while the *scrB* mutant was deficient in sucrose hydrolase activity. A residual sucrose hydrolase activity was, however, still present and ascribed to an α-glucosidase. (Brückner *et al.*, 1993; Wagner *et al.*, 1993). The *scrR* mutant expressed sucrose uptake and sucrose hydrolase activity constitutively, suggesting that its gene product was a negative regulator of *scrA* and *scrB* expression (Gering and Brückner, 1996).

An imperfect inverted repeat, designated O_B, was identified +6 to +21bp relative to the *scrB* transcription initiation site, and proposed to be the site at which ScrR might act. An involvement of O_B in *scrB* regulation was confirmed by a 4bp deletion derivative of this site, which resulted in constitutive expression of sucrose hydrolase activity (Gering and Brückner, 1996). Gel mobility shift assays demonstrated that purified ScrR specifically bound to DNA fragments containing the O_B site, thereby confirming an interaction between ScrR and O_B (Gering and Brückner, 1996). It would follow that *in vivo*, transcriptional repression would occur as ScrR binds O_B in the absence of sucrose.

The precise inducer molecule resulting in the release of ScrR from O_B was not identified. The authors did, however, indicate that native sucrose, as well as fructose and glucose and their phosphorylated derivatives failed to prevent *in vitro* binding of ScrR to O_B. Although not tested, sucrose-6-phosphate would seem to be the most likely inducer molecule, given that sucrose would be present in this form after PTS-dependent transport (Gering and Brückner, 1996).

Evidence has also been presented to suggest that the O_B operator site, plays a role in glucose-mediated repression of *scrB*. This was evident in the strain carrying the 4bp deletion derivative of O_B, where glucose-mediated repression of sucrose hydrolase activity was not detected. Gering and Brückner (1996) noted that O_B fits perfectly to the catabolite-responsive

element (CRE) consensus sequence reported by Hueck *et al.* (1994). This raised the possibility that a catabolite control protein (CcpA) may also exert an effect at O_B.

Transcription of the *scrA* gene is almost certainly also regulated by ScrR. Two O_B-like sites O_{A1} and O_{A2} were identified in the *scrA* promoter region. O_{A1} has three mismatches relative to O_B and covers four bases of the predicted *scrA* -10 promoter region. O_{A2} differs at two positions from O_B and is located between positions +11 and +24 downstream of the *scrA* major transcriptional start point. Gel mobility shift assays demonstrated that ScrR bound to the *scrA* promoter region, but it was not established whether O_{A1} or O_{A2} or both operator sites were involved.

1.3.3. *Streptococcus mutans*, *Streptococcus sobrinus*, *Lactococcus lactis* and *Pediococcus pentosaceus*.

A number of similarities exist between the sucrose PTS regulons of *S. mutans*, *S. sobrinus*, *L. lactis* and *P. pentosaceus*. Shared features include: a similarity in molecular organization; a similarity in *scrA* encoded EII^{scr} structure; and an association with ROK/hexokinase family fructokinases. The similarities raise the possibility of a common evolutionary origin for these regulons. Sucrose utilization by these organisms will be considered individually, after which the common elements between their sucrose PTS regulons will again be emphasized.

1.3.3.1. *Streptococcus mutans* and *Streptococcus sobrinus*

S. mutans possesses multiple extracellular and intracellular sucrose metabolic pathways, and evidence has accumulated to indicate that extracellular sucrose metabolism plays a critical role in the formation of dental caries. Aspects of extracellular sucrose metabolism will be discussed first, after which intracellular sucrose metabolic pathways will be considered. The sucrose PTS regulon of *S. sobrinus* will also be briefly considered.

Extracellular sucrose metabolism: Glucosyltransferases play a predominant role in extracellular sucrose metabolism. At least three glucosyltransferases, encoded by the *gtfB*, *gtfC* and *gtfD* genes are produced (Shiroza *et al.*, 1987; Ueda *et al.*, 1988; Honda *et al.*, 1990). Homology between the *gtf* genes suggests that they share a common origin (Honda *et al.*, 1990). The *gtfB* and *gtfC* genes are arranged in tandem and constitute an operon while *gtfD* is not genetically linked to the *gtfBC* locus. GtfB catalyzes the synthesis of water-insoluble glucans with glucose residues linked predominantly by α(1-3) bonds. GtfC catalyzes the

synthesis of glucans with varying solubility in water. GtfD catalyzes the synthesis of water-soluble glucans with glucose residues linked predominantly by α(1-6) bonds. Water-insoluble glucans are thought to contribute to the dental plaque matrix and facilitate the colonization of smooth dental surfaces. Non-diffusing water-soluble glucans are thought to provide an extracellular energy reserve, important in the oral environment where sugar concentrations are likely to fluctuate.

A fructosyltransferase, encoded by the *ftf* gene, is also involved in extracellular sucrose metabolism. The *ftf* gene specifies a protein with homology to the *B. subtilis* levansucrase (Shiroza and Kuramitsu, 1988), but differs from the *B. subtilis* levansucrase in that it catalyzes the synthesis of β(2-1) linked fructose polymers (inulins), rather than β(2-6) linked fructose polymers (levans). Fructose polymers, like water-soluble glucans, are thought to act as an extracellular energy reserve. Fructans are most likely hydrolyzed by an exo-β-D fructosidase, encoded by the *fruA* gene (Burne and Penders, 1992; 1994). FruA is a member of the glucosyl hydrolase family 32 (Henrissat and Bairoch, 1993) and is capable of hydrolyzing sucrose as well as β(2-1) and β(2-6) linked fructose polymers.

Studies on the regulation of glucosyltransferase and fructosyltransferase activities have been complicated by several factors. At least three glucosyltransferases are produced. Therefore, activity attributable to a specific glucosyltransferase cannot easily be distinguished from overall glucosyltransferase activity. Furthermore, endodextranases are partly able to degrade glucans, while fructanases degrade fructans, with the consequence that total extracellular glucan and fructan synthesis may be underestimated. To overcome these difficulties, Wexler *et al.* (1993) constructed fusions between the *cat* gene (encoding chloramphenicol acetyltransferase) and the *gtfBC* and *ftf* genes. Using this system, the authors were able to demonstrate significant increases in CAT activity, for both *gtfBC::cat* and *ftf::cat* fusions, when sucrose was added to steady state cultures. An independent study has also reported sucrose induction of *ftf::cat* fusions (Kiska and Macrina, 1994). Clearly, mechanisms exist regulating the expression of *gtfBC* and *ftf* in response to sucrose.

It has been proposed that regulation of *ftf* expression is mediated at two inverted repeat structures, within 135bp of the *ftf* initiation codon (Kiska and Macrina, 1994). Regulation at these sites is thought to be facilitated by the product of the *frp* gene, which lies immediately downstream and in the opposite orientation to *ftf* (Shiroza and Kuramitsu, 1988; Shibata and

Kuramitsu, 1996). Gel mobility shift assays have indicated that purified Frp was able to bind to the *ftf* promoter region, containing the two inverted repeat structures (Shibata and Kuramitsu, 1996). While these data suggest a role for Frp in the regulation of *ftf* expression, the factors modulating this regulation remain to be established.

Possibly the most important aspect of extracellular sucrose metabolism, is the role it plays in dental decay (Loesche, 1986). This has been investigated using a gnotobiotic rat model and *S. mutans* strains carrying mutations in the *gtf* and *ftf* genes (Munro *et al.*, 1995). *S. mutans* strains, unable to produce water-insoluble glucans, were depressed in the formation of smooth dental surface lesions (Munro *et al.*, 1995). This indicated that water-insoluble glucans contributed to the formation of dental caries, possibly by facilitating the adhesion of cells to dental surfaces. The presence of water-soluble glucans and fructans may also play a role in the formation of dental caries. This may occur as monomers from these polymers are fermented, during periods of nutrient deprivation. This, in turn, would result in sustained acid production on dental surfaces.

Intracellular sucrose metabolism: Only a small proportion of sucrose (<10%) is metabolized in the extracellular milieu (Loesche, 1986). The remainder is metabolized intracellularly. Intracellular sucrose metabolism is mediated by at least two systems: a high affinity PTS-dependent system (K_m 70μM for sucrose) and a low affinity multiple sugar metabolism (msm) system (St. Martin and Wittenberger, 1979b; Tao *et al.*, 1993). In addition, a trehalose-PTS system has been shown to transport sucrose (K_m 250μM for sucrose) (Poy and Jacobson, 1990). The discussion which follows will focus firstly on the msm system and then on PTS-dependent sucrose metabolism.

The msm system is involved in the uptake and initial metabolism of several sugars, including melibiose, raffinose, and isomaltosaccharides, and is induced by raffinose and melibiose. At the molecular level the msm system consists of eight genes: *msmR*, encoding a positive effector protein; *aga*, encoding an α-galactosidase; *msmE*, encoding a sugar-binding lipoprotein; *msmF* and *msmG*, encoding integral membrane proteins; *gtfA*, encoding a sucrose phosphorylase; *msmK*, encoding an ATP-binding protein and *dexB*, encoding a dextran glucosidase (Russell *et al.*, 1992). Except for *msmR*, the remaining genes are present in the same orientation and are co-transcribed (McLaughlin and Ferretti, 1996). The sugar-binding lipoprotein MsmE is thought to reside on the cell surface, where it binds target sugars. Integral

membrane components MsmF and MsmG are involved in permitting sugar entry into the cell, while blocking exit.

The msm was initially only thought to generate intracellular sucrose as a result of α-galactosidase activity on raffinose (galactose linked to sucrose) (Tao *et al.*, 1993). Intracellular sucrose would then be metabolized by GtfA (sucrose phosphorylase), yielding glucose-1-phosphate and fructose. Evidence for a direct role in sucrose uptake, was suggested by the fact that sucrose was able to block ^{3}H-melibiose transport by 70% (Tao *et al.*, 1993). This was further substantiated by the finding that ^{14}C-sucrose uptake was increased, on induction of the msm, in a sucrose PTS deficient strain. While these data suggested a role in sucrose uptake, it was proposed that the msm would have a low affinity for sucrose and would only play a part in the presence of high sucrose concentrations (Tao *et al.*, 1993).

A sucrose PTS system is likely to facilitate the utilization of sucrose present in low concentrations. Characterization of this system, at the physiological level, has revealed the presence of sucrose phosphotransferase, sucrose-6-phosphate hydrolase and fructokinase activities (St Martin and Wittenberger, 1979a; St Martin and Wittenberger, 1979b; Sato *et al.*, 1993b). Sucrose phosphotransferase activity was found to be induced by sucrose while sucrose-6-phosphate hydrolase activity was constitutive (St Martin and Wittenberger, 1979a). Both sucrose-6-phosphate hydrolase and sucrose phosphotransferase activities were repressed in fructose-grown cultures.

At the molecular level, the sucrose PTS regulon is encoded by the *scrK*, *scrA*, *scrB* and *scrR* genes (Fig. 1.9). The *scr* genes encode: a ROK/hexokinase family fructokinase (ScrK); an EIIBCA^{Scr} protein (ScrA); a sucrose-6-phosphate hydrolase of the glucosyl hydrolase family 32 (ScrB) and a GalR-LacI family regulatory protein (ScrR) (Sato and Kuramitsu, 1988; Sato *et al.*, 1989; Sato *et al.*, 1993b; Titgemeyer *et al.*, 1994; Henrissat, 1991; GenBank accession number: U46902). The *scrA* and *scrK* genes are transcribed divergently from the *scrB* and *scrR* genes. Sequence analysis revealed putative, *E. coli*-like promoters and ribosome-binding-sites for the divergent *scrA* and *scrB* genes (Sato *et al.*, 1989; Sato and Kuramitsu, 1988). The absence of a terminator-like structure after *scrB* suggested that *scrB* and *scrR* may be co-transcribed. The presence of a putative transcriptional terminator after *scrA*, as well as a putative promoter and ribosome-binding-site for *scrK*, suggested that these two genes may be independently transcribed (Sato *et al.*, 1989; Sato *et al.*, 1993b). A phosphomannose

isomerase encoding gene (*pmi*), was identified downstream of *scrK* (Sato *et al.*, 1993a). Disruption of *scrK* reduced *pmi* activity, suggesting that *scrK* and *pmi* may be co-transcribed. Further experimental analysis is required to confirm the transcriptional organization of the *S. mutans scr* system.

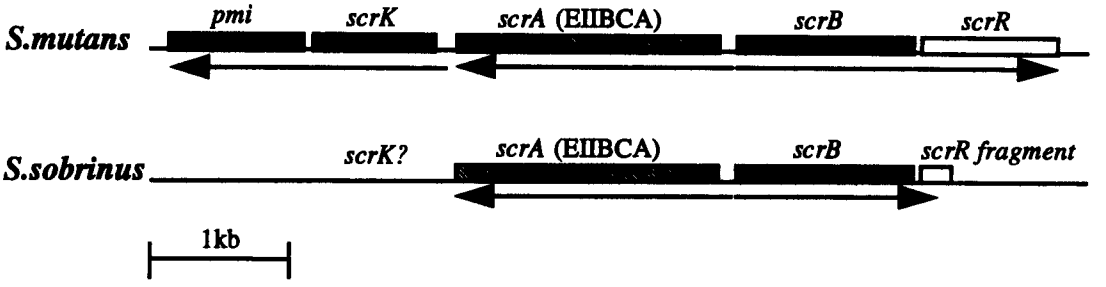

Fig. 1.9 Organization of the *S. mutans* and *S. sobrinus* sucrose utilization regulons. Arrows indicate transcriptional polarities. Where a single arrow extends over two genes, co-transcription has been inferred from nucleotide sequence analysis.

The expression of *scrA* has been studied using *scrA::lacZ* fusions (Sato *et al.*, 1991). These studies demonstrated that extracellular sucrose induced *scrA::lacZ* expression, although the precise sucrose metabolite involved in the induction was not determined. It was also noted that glucose, fructose and maltose inhibited the induction of *scrA::lacZ* by sucrose, while sorbitol and lactose did not. It is tempting to suggest that the *scrR* gene product is involved in the regulation of *scrA* expression, however, this remains to be experimentally established.

A truncated sucrose PTS regulon has also been cloned from *S. sobrinus* (Chen *et al.*, 1992). The *scrA* and *scrB* genes of this regulon encode an EIIBCA^{Scr} and sucrose-6-phosphate hydrolase respectively, and are arranged identically to their equivalents in *S. mutans* (Fig. 1.9) (Chen *et al.*, 1993). A truncated open reading frame lies immediately downstream of the *S. sobrinus scrB* gene and appears to encode a helix-turn-helix motif similar to that of the *S. mutans* ScrR (28aa identical amino acids over a 31 amino-terminal overlap). This suggests that, in *S. sobrinus* as in *S. mutans*, a gene encoding a regulatory protein lies immediately downstream of *scrB*.

1.3.3.2. *Lactococcus lactis*

Studies on sucrose utilization by *Lactococcus lactis* have revealed that the ability to utilize sucrose is associated with the production and resistance to the small protein antibiotic nisin (Gasson, 1984). It has been established that the traits for sucrose utilization and nisin production are encoded on transposons (Donkersloot and Thompson, 1990; Rauch and de Vos, 1992a; Rauch and de Vos, 1992b; Thompson *et al.*, 1991a, Thompson *et al.*, 1991b). The

Nisin-Sucrose transposons of *L. lactis* strains are approximately 70kb in size. They differ with regard to the nature of the nisin antibiotic produced and in their ability to conjugate to other *L. lactis* strains (Rauch and de Vos, 1992b; Rauch *et al.*, 1994). Due to the 30% G + C content of the transposons as opposed to the 38% G + C content of the *L. lactis* genome, it has been proposed that the transposons originated outside the genus *Lactococcus* (Rauch *et al.*, 1994).

Sucrose utilization by *L. lactis* KI occurs via a typical sucrose PTS pathway. Sucrose transport yields intracellular sucrose-6-phosphate and this is hydrolyzed by a sucrose-6-phosphate hydrolase releasing glucose-6-phosphate and fructose (Thompson and Chassy, 1981). Free intracellular fructose is phosphorylated by a fructokinase yielding fructose-6-phosphate. The sucrose-6-phosphate hydrolase of *L. lactis KI* has been purified and was reported to have a relative molecular mass of 52,000 after ultracentrifugal analysis (Thompson *et al.*, 1991a). Although it was capable of hydrolyzing both sucrose-6-phosphate and sucrose, its K_m for sucrose-6-phosphate (0.1mM) was significantly lower than its K_m for sucrose (100mM).

Two electrophoretically distinct fructokinases, fructokinase I and fructokinase II, have been identified for *L. lactis* KI (Thompson *et al.*, 1991b). Fructokinase I was found to be induced by growth on sucrose, while fructokinase II was detected during growth on ribose, galactose, maltose and lactulose. Fructokinase I was purified to electrophoretic homogeneity and found to be dimeric with a subunit size of 33.5kDa. The N-terminal sequence of purified fructokinase I was homologous to the *S. mutans* ROK/hexokinase family fructokinase (18 identical residues out of 25aa) (Sato *et al.*, 1993b), suggesting that it too is a member of the ROK/hexokinase family. Fructokinase I was able to utilize a range of nucleoside triphosphates as phosphoryl donors, however activity was highest in the presence of ATP. Fructokinase I activity also required the presence of a divalent cation, with activity being highest in the presence of $MgCl_2$.

Sucrose PTS regulons have been studied from *L. lactis* NIZO R5 and *L. lactis* K1, where they are encoded on Tn*5276* and Tn*5306* respectively. The studies involved partial sequencing of the *L. lactis* NIZO R5 system and a Southern blot analysis of the *L. lactis* K1 system (Rauch and de Vos, 1992a; Thompson *et al.*, 1991a; 1991b). The southern blot analysis used DNA oligonucleotide probes, derived from the N-terminal sequences of fructokinase I (*scrK* probe) and the purified sucrose-6-phosphate hydrolase (*scrB* probe). In addition, a *scrA* probe was

prepared from a conserved region of EII^{Scr} encoding genes. The analysis revealed that the *scr* genes were clustered in the order *scrK*, *scrA*, *scrB* (Thompson *et al.*, 1991b) (Fig. 1.10)

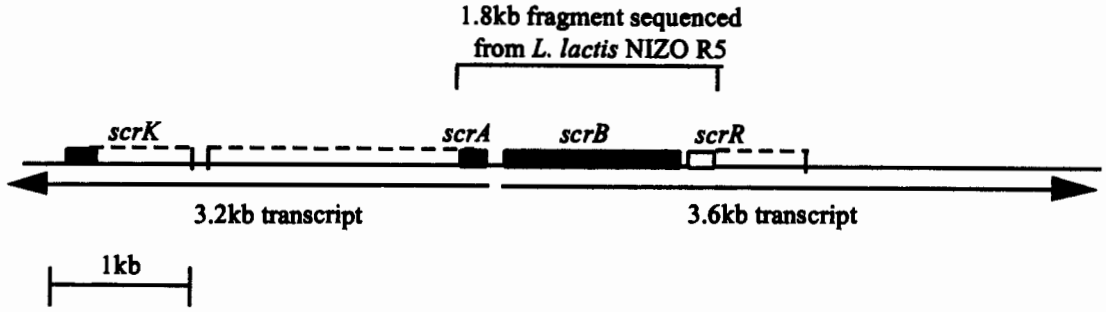

Fig. 1.10 Organization of sucrose utilization regulons from *L. lactis* transposons. The diagram was composed from the sequence of the Tn5276 *scrB* and adjacent DNA (Rauch and de Vos, 1992b), as well as a Southern blot analysis of the *scrK*, *scrA* and *scrB* genes of Tn5306 (Thompson *et al.*, 1991b). Arrows indicate transcriptional polarity. Where a single arrow extends over two genes, transcriptional coupling has been inferred from Northern blot analysis (Rauch and de Vos, 1992b). It should be noted that the exact position and orientation of *scrK* has not yet been established.

A 1.8kb fragment, from the *L. lactis* NIZO R5 system was sequenced. The fragment contained a *scrB* gene flanked by a truncated *scrA* and *scrR* (Fig. 1.10) (Rauch and de Vos, 1992a). The truncated *scrR* was present downstream and in the same orientation as *scrB*, while the truncated *scrA* was present upstream and in the opposite orientation. The *scrB* gene encoded a sucrose hydrolase, while the truncated *scrA* and *scrR* were proposed to encode an EII^{Scr} protein and a GalR-LacI-like regulatory protein respectively (Rauch and de Vos, 1992a).

Transcription of the *scrA* and *scrB* genes was investigated by northern blot and primer extension analysis. Primer extension analysis revealed that *scrA* and *scrB* were transcribed off sucrose inducible promoters, positioned back-to-back, in the 180bp intergenic region. Northern blot analysis indicated that *scrA* was expressed on a 3.2kb transcript, while *scrB* and *scrR* were co-transcribed on a 3.6kb transcript. The 3.2kb and 3.6kb transcripts were detected only in RNA extracted from sucrose-grown cells (Rauch and de Vos, 1992b). An imperfect palindrome ($T_8G_7T_6C_5A_4A_3G_2C_1/G'_1T'_2T'_3T'_4G'_5C'_6C'_7A'_8$) was identified within the intergenic region between *scrA* and *scrB*. It is tempting to suggest that the *scrR* gene product regulates transcription of the *scr* genes at this site. However, experimental evidence for this has yet to be presented.

1.3.3.3. *Pediococcus pentosaceus*

The adjacent sucrose and raffinose operons of *P. pentosaceus* have been cloned and sequenced. This information is available as a nucleotide sequence submission (GenBank

accession number L32093). The sucrose PTS regulon consists of what are thought to be two transcriptional units with opposite polarities (K. Leenhouts personal communication). The first involves the *scrB*, *scrR* and *agl* genes, encoding a sucrose hydrolase of the glucosyl hydrolase family 32, a GalR-LacI-like regulatory protein and an α-glucosidase respectively. The second transcriptional unit is thought to involve the *scrA*, *agaS* and *scrK* genes which most likely encode a PTS EIIBCA^{Scr} protein, a putative α-galactosidase and a ROK/hexokinase family fructokinase respectively (Fig. 1.11).

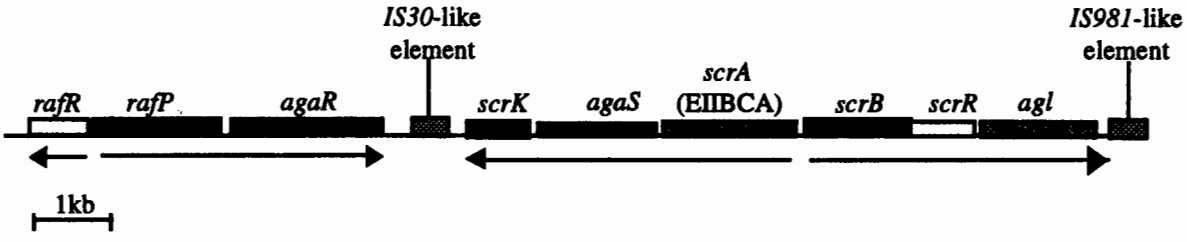

Fig. 1.11 Organization of the adjacent sucrose and raffinose operons from *P. pentosaceus*. Horizontal arrows indicate transcriptional polarity. Where a single arrow extends over two genes transcriptional coupling is thought to occur (K. Leenhouts personal communication).

Transcription of the *scrB-scrR-agl* and *scrA-agaS-scrK* transcriptional units was found to be repressed by fructose but not by glucose (K. Leenhouts personal communication). Putative transcriptional terminators flank the two transcriptional units, and these are in turn, flanked by IS elements similar to the *E. coli* IS30 and *L. Lactis* IS981. This has lead to speculation that the regulon is located on a transposon (K. Leenhouts personal communication). The adjacent raffinose regulon, like the sucrose regulon, appears to consist of two transcriptional units with opposite polarities. The first transcriptional unit involves *rafP* and *agaR* encoding a putative raffinose permease and α-galactosidase. The second transcriptional unit involves the *rafR* gene, which has homology to *msmR* activator of the *S. mutans* multiple sugar metabolism system. Further reported research is required to discuss this system in more detail.

1.3.3.4. Similarities between the sucrose PTS regulons of *S. mutans*, *S. sobrinus*, *P. pentosaceus*, and *L. lactis*.

As indicated at the beginning of section 1.3.3. the sucrose PTS regulons of *S. mutans*, *S. sobrinus*, *P. pentosaceus* and *L. lactis* share several features. Indeed, Rauch and de Vos (1992a) have already noted the similarities between the *S. mutans* and *L. lactis* systems and suggested a common origin for the two systems. The shared features are highlighted below.

Molecular organization: In each case, the *scrA* and *scrB* genes are divergently transcribed and separated by between 180 and 256bp. In all cases, evidence has been presented suggesting

that a *scrR* gene, encoding a regulatory protein, lies downstream and in the same orientation the respective *scrB*. Similarities also exist in the positions of *scrK* genes. In the case of *S. mutans*, the *scrK* gene lies immediately downstream and in the same orientation as *scrA*. In *P. pentosaceus*, the *scrK* gene lies downstream and in the same orientation as the *scrA* gene, but is separated from it by a 2259bp region proposed to encode an α-galactosidase. In *L. lactis*, it was established that the *scrK* gene is associated with the *scrA* and *scrB* genes and lies downstream of *scrA*, however, its exact orientation and position remains to be established. The association of a *scrK* with the *S. sobrinus* system has not been established.

Structure of EII^{Scr} proteins: In the cases of *S. mutans*, *S. sobrinus* and *P. pentosaceus*, the *scrA* genes encode EIIBCA^{Scr} proteins. In this respect the sucrose PTS systems of these three organisms are alike, and differ from the remaining systems, which are characterized by EIIBC^{Scr} proteins. Since the complete sequence of the EII^{Scr} protein of *L. lactis* has not been reported, it could not be established whether an EIIBC^{Scr} or EIIBCA^{Scr} protein was involved.

Association with ROK/hexokinase family fructokinases: In the cases of *S. mutans* and *P. pentosaceus*, where the entire *scrK* genes have been sequenced, it has been established that the encoded proteins belong to the ROK/hexokinase family. In *L. lactis*, the N-terminal sequence of fructokinase I is homologous to members of the ROK/hexokinase family. The association of these systems with ROK/hexokinase family fructokinases, distinguishes them from other bacterial sucrose utilization systems, which are typically associated with ribokinase/*pfkB* family fructokinases.

When considering a common origin for the regulons discussed here, it is important to note that the sucrose utilization systems of *L. lactis* strains are encoded on conjugative transposons and that these transposons have a GC content which differs from that of the *L. lactis* genome. This led to the proposal that the *L. lactis* sucrose/nisin transposons originated outside of the genus *Lactococcus* (Rauch and de Vos, 1992a). In the light of this, it must be noted that the sucrose utilization system of *P. pentosaceus* is flanked on one side by an *E. coli* IS30-like element and on the other by a *L. lactis* IS981-like insertion sequence. It has not yet been established whether the *S. mutans* and *S. sobrinus* sucrose utilization systems are encoded on transposons. Further investigation is required to clearly establish whether the sucrose PTS regulons of *S. mutans S. sobrinus*, *L. lactis* and *P. pentosaceus* do, in fact, share a common origin.

1.3.4. *Salmonella typhimurium* and *Klebsiella pneumoniae*

The conjugative plasmid pUR400, isolated from a clinical strain of *S. typhimurium*, encodes a sucrose PTS pathway (Titgemeyer *et al.*, 1996). The sucrose PTS regulon carried by this plasmid is almost identical to the chromosomally encoded system of *K. pneumoniae* and both systems have been studied extensively (Aulkemeyer *et al.*, 1991; Ebner, and Lengeler, 1988; Jahreis and Lengeler, 1993; Schmid *et al.*, 1982, 1988, 1991; Sprenger and Lengeler, 1988; Titgemeyer *et al.*, 1996).

The sucrose PTS regulons of *K. pneumoniae* and pUR400 are unique, in that both encode sucrose-specific outer membrane porins (Forst *et al.*, 1993; Hardesty *et al.*, 1991; Schmid *et al.*, 1991; Schülein *et al.*, 1991). In both organisms, sucrose in the periplasmic milieu is further metabolized by typical sucrose PTS pathways. (Lengeler *et al.*, 1982; Schmid *et al.*, 1982; Sprenger and Lengeler, 1988). At the molecular level the *K. pneumoniae* and pUR400 systems consist of five clustered genes (Fig. 1.12). Gene designations for the systems are identical, with *scrK*, *scrY*, *scrA*, *scrB* and *scrR* encoding ribokinase/*pfkB* family fructokinases, outer membrane sucrose porins, EIIBC^{Scr} proteins, sucrose-6-phosphate hydrolases and GalR-LacI-like regulatory proteins respectively. The similarity between the systems is indicated by the high percentage of identical amino acids in the corresponding proteins *viz*: ScrK (76.7%), ScrY (86%); ScrA (95%), ScrB (76%) and ScrR (91%) (Aulkemeyer *et al.*, 1991; Jahreis and Lengeler, 1993; Schmid *et al.*, 1991; Titgemeyer *et al.*, 1996).

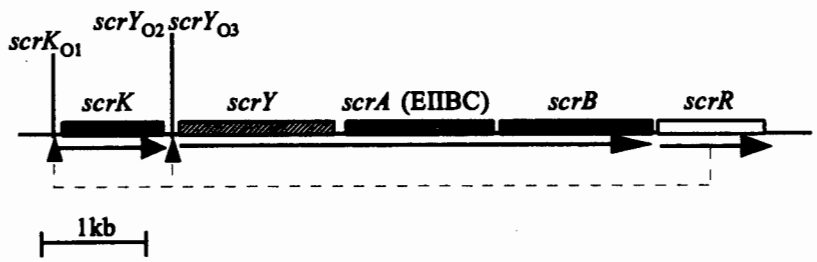

Fig 1.12 Molecular organization of the pUR400 and *K. pneumoniae* sucrose utilization systems. Solid arrows indicate transcriptional polarity. Where a single arrow extends over several genes, transcriptional coupling is thought to occur. Dashed lines/arrows indicate established regulatory relationships. ScrR binds to the operators ($scrK_{O1}$, $scrY_{O2}$ and $scrY_{O3}$) in the absence of fructose or fructose-1-phosphate, preventing transcription of the respective genes.

Three physiological promoters have been identified in the sucrose PTS regulons of *K. pneumoniae* and pUR400. These are those of *scrK*, *scrY* (facilitating expression of the *scrYAB* operon) and *scrR* (Cowan *et al.*, 1991; Hardesty *et al.*, 1991; Jahreis and Lengeler, 1993;

Schmid *et al.*, 1991). A putative cyclic AMP receptor protein binding site, was identified immediately upstream of *scrY* in both the pUR400 and *K. pneumoniae* systems, as well as in a highly similar plasmid encoded system from *Salmonella thompson* (Cowan *et al.*, 1991; Jahreis and Lengeler, 1993).

Regulation of the *K. pneumoniae* and pUR400 systems is mediated by the respective ScrR proteins. ScrR operator sites are identical for the two systems, with one operator site lying upstream of the respective *scrK* gene ($scrK_{01}$) and two operator sites lying upstream of the respective *scrY* gene ($scrY_{02}$ and $scrY_{03}$) (Fig. 1.12) (Jahreis and Lengeler, 1993). Gel mobility shift assays have confirmed that the ScrR proteins bind to regions of DNA containing the $scrK_{01}$ and $scrY_{02O3}$ operator sites (Jahreis and Lengeler, 1993). Furthermore, it has been demonstrated that the binding of ScrR proteins to the respective operators is reversed by the presence of fructose and fructose-1-phosphate, implicating these metabolites as inducers (Jahreis and Lengeler, 1993). In this regard it is intriguing that ScrR proteins respond to a sucrose metabolite present after sucrose hydrolysis, rather than to sucrose-6-phosphate itself. As a result, the systems are induced in environments containing fructose, even when sucrose is not present (Schmid *et al.*, 1982; Sprenger and Lengeler, 1988).

1.3.5. *Escherichia coli*

Laboratory strains of *E. coli* are typically unable to metabolize sucrose, however, several sucrose metabolizing *E. coli* strains have been isolated. In two isolates *E. coli* EC3132 and *E. coli* B-62, sucrose catabolic genes appear to be chromosomally encoded, while in a third clinical isolate, the traits for sucrose utilization appear to be encoded on a transposon designated *Tn2555* (Bockmann *et al.*, 1992; Doroshenko *et al.*, 1988; Tsunekawa *et al.*, 1992). The most well studied of these systems is the chromosomally encoded system of *E. coli* EC3132, which will be the focus in this review.

A physiological analysis of sucrose metabolism by *E. coli* EC3132 indicated that sucrose was taken up by a non-PTS mechanism, followed by hydrolysis by a sucrose hydrolase and phosphorylation of fructose by a fructokinase (Bockmann *et al.*, 1992). It was also demonstrated that sucrose uptake, sucrose hydrolase and fructokinase activities were induced by the presence of sucrose. At the molecular level, the system has been cloned and sequenced and consists of four clustered genes: *cscA*, encoding a 477aa sucrose hydrolase of the glucosyl hydrolase family 32; *cscR*, encoding a 331aa GalR-LacI-like regulatory protein; *cscK*,

encoding a 305aa ribokinase/*pfkB* family fructokinase and *cscB*, encoding a 415aa sucrose/H^+ symporter (Fig. 1.13) (*csc* is a mnemonic for chromosomally coded sucrose genes). A *cscB* promoter was not identified suggesting that it may be co-transcribed with the adjacent *cscK* (Bockmann *et al.*, 1992).

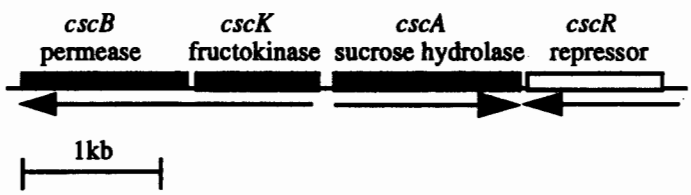

Fig 1.13 Organization of the chromosomally encoded sucrose utilization regulon of *E. coli* EC3132. Arrows indicate transcriptional polarities of the respective genes. Where a single arrow extends over two genes, transcriptional coupling has been inferred from nucleotide sequence analysis.

The most unique feature of the *E. coli* EC3132 system, is that it involves a sucrose/H^+ symporter. CscB has been classified as a member of the major facilitator superfamily, containing antiporters, uniporters and H^+ symporters (but not Na^+ symporters) (Marger and Saier, 1993). More specifically, CscB is a member of Cluster 5 of this family which contains oligosaccharide/H^+ symporters, including the lactose permeases of *E. coli, K. pneumoniae* and *Citrobacter freundii,* and the raffinose permease of *E. coli.* The members of the major facilitator superfamily are thought to have a characteristic structure involving 12 transmembrane α-helices connected by hydrophilic loops, a structure which has been specifically proposed for CscB (Bockmann *et al.*, 1992).

Regulation of the *csc* gene cluster is mediated by CscR. In this respect, CscR is most likely a negative regulator of *cscA* and *cscKB* transcription, since a mutation in *cscR* resulted in constitutive expression of sucrose hydrolase, sucrose uptake and fructokinase activities (Bockmann *et al.*, 1992). It has been reported that within the GalR-LacI family, CscR has evolved independently of other ScrR proteins (Nguyen and Saier, 1995). This may reflect a situation in which the different GalR-LacI regulatory proteins, involved in sucrose metabolism, respond to different sucrose metabolites. In this regard, it should be noted that in *E. coli* EC3132, sucrose enters the cell in its native form, while in organisms where sucrose is transported via the PTS, the product of sucrose transport is sucrose-6-phosphate.

1.3.6. *Vibrio alginolyticus*

A typical sucrose PTS system has been characterized for a halotolerant, collagenolytic *V. alginolyticus* strain (Blatch, 1990; Blatch *et al.*, 1990; Blatch and Woods, 1991; Scholle *et al.*,

1987; Scholle *et al.*, 1989). It has been established that sucrose hydrolase activity expressed by this organism is intracellular, and that both sucrose hydrolase and sucrose uptake activities are induced by sucrose and repressed by glucose (Scholle *et al.*, 1987). Cloning and sequencing of the sucrose utilization system from *V. alginolyticus*, revealed the presence of five clustered genes (Fig 1.14). These include: *scrR*, encoding a 94aa regulator of the GalR-LacI family; *scrA*, encoding a 479aa EIIBC^{Scr} protein; *scrK*, encoding a 307aa fructokinase of the ribokinase/*pfkB* family and *scrB*, encoding a 484aa sucrose hydrolase of the glucosyl hydrolase family 32. In studies where the system was expressed in *E. coli*, sucrose transport was reported to be dependent on the presence of functional *E. coli* EIIA^{Glc}. This suggested that an EIIA$^{Glc/Scr}$ equivalent, in *V. alginolyticus*, may act as a phosphoryl donor for EIIB^{Scr} (Blatch, 1990).

Fig 1.14 Organization of the *V. alginolyticus* sucrose utilization regulon. Arrows indicate transcriptional polarity. Where a single arrow extends over two genes, transcriptional coupling has been inferred from nucleotide sequence analysis. However, experimental evidence for this has not been reported

Nucleotide sequence analysis revealed putative *E. coli*-like promoters upstream of *scrR*, *scrA*, and *scrK*, with no promoter-like sequences identified between *scrK* and *scrB*. The fact that the putative *scrK* promoter was not well conserved, lead to the suggestion that the *scrA*, *scrK* and *scrB* genes may be co-transcribed (Blatch, 1990). Cyclic AMP receptor protein binding sites were not identified in the *V. alginolyticus* system, which is curious, given the fact that the traits for sucrose utilization were repressed by the presence of glucose. Catabolite repression may act at other levels, possibly involving inducer exclusion, as a result of competition for a shared EIIA$^{Scr/Glc}$.

The deduced ScrR protein is a particularly unusual member of the GalR-LacI family, in that it is approximately one third the size of the remaining members (Weickert and Adhya, 1992). The role of the *scrR* gene product, in the regulation of the *V. alginolyticus* sucrose utilization system, was investigated by deletion analysis of the system cloned into *E. coli* (Blatch and Woods, 1991). In plasmids where the system remained intact, regulation in response to sucrose was observed. However, in plasmids carrying a truncated *scrR*, expression of sucrose

hydrolase activity became constitutive, indicating that the ScrR protein is a negative regulator of at least the *scrB* gene. Two putative repressor-binding sites were identified in the intergenic region between the *scrA* and *scrR* genes, and these may be the sites at which ScrR acts (Blatch and Woods, 1991).

1.3.7. *Zymomonas mobilis*

Z. mobilis, a Gram-negative facultative anaerobe, has attracted interest due to its ability to produce ethanol at a high yield with glucose, fructose or sucrose as a carbon source (Ingram *et al.*, 1989). Sucrose utilization has been studied in several strains of *Z. mobilis viz*: *Z. mobilis* Z6C; *Z. mobilis* ZM1; *Z. mobilis* CP4. A sucrose PTS has not been identified for these strains and sucrose hydrolyzing activity, is localized primarily in the periplasm or extracellular milieu (Mortatte *et al.*, 1983; Preziosi *et al.*, 1990; O'mullan *et al.*, 1991). Extracellular sucrose is either metabolized to form high-molecular-weight fructose polymers (Park *et al.*, 1993) or is hydrolyzed, with glucose and fructose entering cells via a carrier-mediated diffusion system (Di Marco and Romano, 1985). Three genes involved in sucrose metabolism, and their corresponding products, have been characterized. The genes specify: an extracellular fructosyltransferase, catalyzing levan synthesis; an extracellular 'sucrose hydrolase', which appears to be a mutated fructosyltransferase; and an intracellular sucrose hydrolase. Each will be discussed below.

A fructosyltransferase encoding gene *levU* has been cloned from *Z. mobilis* ZM1 and a corresponding protein has been purified from *Z. mobilis* Z6C (Song *et al.*, 1993; Yanase *et al.*, 1992). The first 19 N-terminal amino acid residues of the purified fructosyltransferase were a perfect match with the deduced LevU. Five relatively well conserved domains were identified between the deduced LevU and fructosyltransferases from *B. subtilis* and *S. mutans*. In addition, the *levU* gene conferred a sucrose-positive phenotype and transfructosylation activity on *E. coli*.

Identical genes, thought to encode the extracellular 'sucrose hydrolase', have been cloned by two groups from *Z. mobilis* ZM1, and designated *sacC* and *invB* (Kannan *et al.*, 1995; Song *et al.*, 1994). The *sacC/invB* gene was found to lie immediately downstream of *levU* and its deduced product was found to be homologous to LevU. However, the *sacC/invB* gene product was unable to catalyze transfructosylation reactions. It appears, therefore, that the *sacC/invB* gene arose as a result of a duplication of *levU*, followed by a mutation preventing

transfructosylation but leaving sucrose hydrolase activity unaffected (Kannan *et al.*, 1995). It must be emphasized, however, that conditions conducive to transfructosylation by this enzyme may not yet have been identified.

Genes thought to encode intracellular sucrose hydrolases have been cloned by two groups. The gene isolated from *Z. mobilis* ZM1 was designated *sacA* while the gene isolated from *Z. mobilis* Z6C was designated *invA* (Gunasekaran *et al.*, 1990 and Yanase *et al.*, 1991). The genes isolated by the two groups are virtually identical and encode β-D-fructofuranosidases of the glucosyl hydrolase family 32. The role of the intracellular sucrose hydrolase is obscure, since most sucrose hydrolysis occurs extracellularly and sucrose uptake systems have yet to be identified.

1.4. AIMS OF THIS STUDY

The systems discussed above provide a background against which studies on sucrose utilization by *C. beijerinckii* NCIMB 8052 can be presented. *C. beijerinckii* is a member of a taxonomic group of solventogenic clostridial strains, specifically selected for their ability to ferment molasses, in ABE fermentations (section 1.1.1) (Keis *et al.*, 1995). This implies that efficient sucrose utilization systems exist. However, until recently, sucrose metabolism by these strains has remained essentially unstudied. This provided the impetus for the work reported in this dissertation. The objective of the study was to provide an understanding of sucrose utilization by *C. beijerinckii*, at both the molecular and physiological levels. At the molecular level, the initial aim was to clone and characterize genes involved in sucrose metabolism. The transcriptional regulation of these genes and their roles in sucrose metabolism could then be examined. At the physiological level, the aim was to examine how sucrose metabolism was regulated in response to supplied carbohydrates. The work represents the first comprehensive study of a sucrose utilization system, from a saccharolytic solventogenic clostridial strain.

CHAPTER 2

CLONING AND SEQUENCING OF THE *CLOSTRIDIUM BEIJERINCKII* SUCROSE UTILIZATION SYSTEM

Contents

2.1. SUMMARY

The sucrose utilization system of *C. beijerinckii* NCIMB 8052 was cloned in three stages. In the first stage a *C. beijerinckii* genomic library was screened for plasmids conferring a sucrose-positive phenotype on *E. coli*. Using this approach, a plasmid designated pCBS1 was isolated and found to carry a 6.8kb insert of *C. beijerinckii* chromosomal DNA. Preliminary sequence analysis indicated that the insert contained a truncated sucrose utilization gene cluster. The remainder of the sucrose utilization gene cluster was cloned in the second and third stages by plasmid rescue and inverse PCR respectively.

The *C. beijerinckii* sucrose utilization system was found to contain four closely linked genes, designated *scrA*, *scrR*, *scrB* and *scrK*. On the basis of homology studies, *scrA* was proposed to encode an EIIBC^{Scr} protein of the PTS, *scrR*, a regulatory protein of the GalR-LacI family of transcriptional regulators, *scrB*, a sucrose-6-phosphate hydrolase of the glucosyl hydrolase family 32 and *scrK*, a fructokinase of the ribokinase/*pfkB* sugar kinase family. The putative protein products of the *scrA*, *scrB* and *scrK* genes would specify a typical PTS-dependent sucrose metabolic pathway. The presence of the *scrR* gene suggested that sucrose metabolism would be regulated.

2.2. INTRODUCTION

Despite the fact that solventogenic clostridial strains were used extensively in molasses-based ABE fermentations, sucrose metabolism by these strains remained largely unstudied. This provided the impetus for the present study, which aimed to provide a molecular and physiological characterization of sucrose utilization by a solventogenic clostridial strain. *C. beijerinckii* NCIMB 8052 was selected since it is amenable to genetic manipulation and represents a group of strains widely used in molasses-based ABE fermentations (Kies *et al.*, 1995).

Well characterized bacterial sucrose utilization systems were reviewed in the general introduction (section 1.3). From the review, it is clear that PTS-dependent sucrose metabolic pathways are common. These pathways are typically encoded as regulons or operons, present on transposons or plasmids, or on the chromosome of the respective organism. Sucrose PTS regulons have been isolated from several bacterial species, by screening genomic libraries for clones conferring a sucrose-positive phenotype on *E. coli*. Common laboratory strains of *E. coli* have been used, where the development of a sucrose-positive phenotype required the simultaneous isolation of a sucrose transport system and a sucrose hydrolase (Scholle *et al.*, 1987; Li and Ferenci, 1996). The requirement for a sucrose transport system has also been avoided by screening for raffinose fermentation in *E. coli* strains unable to metabolize galactose (Brückner *et al.*, 1993; Lunsford and Macrina, 1986). In these cases, raffinose was internalized by the LacY permease and cleaved by an α-galactosidase, yielding intracellular galactose and sucrose. A positive fermentation phenotype was dependent on the presence of an enzyme catalyzing the hydrolysis of sucrose.

The tendency for sucrose catabolic genes to be clustered, raised the possibility that an entire *C. beijerinckii* sucrose utilization gene cluster could be cloned in one step. If this was not achieved, chromosome walking would most likely yield additional genes involved in sucrose metabolism. Sequence analysis would then provide an indication of the molecular organization of the components involved, while the identification of genes encoding putative regulatory proteins would allow for the development of hypotheses about regulation. Therefore, as a starting point for this project, it was decided to clone and sequence sucrose catabolic genes from *C. beijerinckii*. This chapter describes the steps taken to isolate the *C. beijerinckii* sucrose utilization (*scr*) gene cluster and presents its subsequent sequence analysis.

2.3. MATERIALS AND METHODS

2.3.1. Bacterial strains, plasmids and culture conditions

A laboratory spore stock of *C. beijerinckii* NCIMB 8052 was maintained in distilled water at 4°C. *C. beijerinckii* was routinely cultured using Clostridial Basal Medium (CBM) (O'Brien and Morris, 1971). CBM liquid medium was cooled to approximately 50°C after autoclaving and transferred to an anaerobic cabinet with a gas phase of 5% H_2 , 10% CO_2 and 85% N_2 (Forma Scientifica Inc., Marietta, Ohio, USA). CBM broths were typically inoculated with a 1/1000 volume of heat shocked spores (70°C for 3 minutes) and cultured at 37°C. CBM agar plates, solidified with agar (1.5%, w/v), were prepared under aerobic conditions and then preincubated under anaerobic conditions for at least 12 hours before use. Where appropriate, erythromycin was used at a final concentration of 10μg.ml^{-1}.

E. coli JM109 (Yanisch-Perron *et al.*, 1985) was used for all cloning purposes and was routinely cultured, at 37°C, on Yeast Tryptone (YT) or Luria-Bertani (LB) medium (Sambrook *et al.*, 1989). YT and LB media were solidified with agar (1.5%, w/v) as appropriate. MacConkey agar medium (Difco Laboratories, Detroit, USA), supplemented with either sucrose (1%, w/v) or lactose (1%, w/v), was used to assess carbohydrate fermentation phenotypes. Ampicillin (100μg.ml^{-1}) and kanamycin (100μg.ml^{-1}) were included in media where appropriate.

Plasmids used for subcloning and sequencing of the *C. beijerinckii* sucrose utilization system include the M13-derived Bluescript SK plasmid (pSK) (Stratagene, La Jolla, California, USA), pEcoR251 (Zabeau and Stanley, 1982), and pACYC177 (Chang and Cohen, 1978). In addition, chromosomally integrated derivatives of plasmid pMTL30 (Williams *et al.*, 1990b) were recovered from the genome of *C. beijerinckii* CBSCRB. The construction of *C. beijerinckii* CBSCRB is discussed in Chapter 4, sections 4.3.2 and 4.4.2. Further strain and plasmid details are supplied in Appendices A and B.

2.3.2. General DNA manipulation and Southern blot hybridization analysis

Competent *E. coli* cells were prepared by rubidium chloride treatment, as described by Armitage *et al.* (1988). Plasmid DNA was extracted and purified from *E. coli* using the Nucleobond® AX KIT (Macherey-Nagel, Germany). T4 DNA ligase and DNA restriction endonucleases were obtained from Boehringer Mannheim and were used according to the manufacturer's recommendations. Where DNA fragments with incompatible ends were to be ligated, protruding 5' and 3' ends were first converted to blunt ends, using Klenow fragment

and T4 DNA polymerase, as described by Sambrook *et al.* (1989). Both Klenow fragment and T4 DNA polymerase were obtained from Boehringer Mannheim.

Gel electrophoresis of DNA was conducted in agarose gels (0.8%, w/v), using a Tris-Acetate EDTA buffer as described by (Sambrook *et al.*, 1989). For Southern blot analysis, DNA fragments were separated by agarose gel electrophoresis and capillary blotted with 0.4M NaOH onto positively charged nylon membranes (Amersham). Non radioactive DIG-labelled probes were prepared using the Boehringer Mannheim DIG random primer labelling kit (catalog number: 1175033). Hybridization and detection of DIG-labelled probes was conducted according to standard protocols supplied by Boehringer Mannheim.

2.3.3. *C. beijerinckii* genomic DNA extraction

One litre *C. beijerinckii* CBM cultures were grown to an OD_{600} of 0.4 units and cells were harvested by centrifugation (7min at 15000*g*). Cells were resuspended in a 20ml volume of CBM containing: 10% (w/v) sucrose; 12.5mM $MgCl_2$; 12.5mM $CaCl_2$ and 5mg.ml^{-1} lysozyme (Boehringer Mannheim). After anaerobic incubation at 37°C for 1 hour, SDS and EDTA were added to a final concentration of 2%(w/v) and 100mM respectively, bringing the total volume of the mixture to 27ml. A 20ml volume of hot phenol (50°C), equilibrated with 0.1M Tris-HCl (pH 8), was added and mixed gently. The aqueous phase was recovered by centrifugation (10 minutes at 12000g) and extracted, firstly with one volume of chloroform:iso-amyl alcohol (24:1 v/v) and then with one volume of water-saturated ether. RNA was removed using ribonuclease A (Sigma Chemical Company, St. Louis, Missouri). Chromosomal DNA was precipitated using isopropanol (Sambrook *et al.*, 1989). The DNA pellet was resuspended in distilled water and stored at 4°C.

2.3.4. Construction and screening of a *C. beijerinckii* genomic library

C. beijerinckii chromosomal DNA was partially digested with *Sau*3AI and subjected to agarose gel electrophoresis. Portions of the gel containing DNA fragments between 5kb and 11kb were excised and the DNA fragments were recovered using the Geneclean® III kit (Bio 101 Inc.). The recovered DNA fragments were ligated into the *Bgl*II site of the positive selection vector pEcoR251 and transformed into *E. coli* JM109. Transformants were screened for sucrose-positive colonies on MacConkey agar plates containing 1% (w/v) sucrose and ampicillin (100μg.ml^{-1}).

2.3.5. Recovery of integrated plasmids from the *C. beijerinckii* genome

During the study, a *C. beijerinckii* strain was constructed carrying a pMTL30 derivative integrated into the *scrB* gene. Details of this are provided in Chapter 4, sections 4.3.2 and 4.4.2. Chromosomal DNA adjacent to the *scrB* gene, was isolated by recovery of the integrated plasmid. The specific details of the plasmid recovery strategy are presented in the results and discussion (Fig. 2.4). The technical aspects of the plasmid recovery are presented below.

Genomic DNA was isolated from the *C. beijerinckii* strain carrying chromosomally integrated copies of a pMTL30 derivative. The isolated DNA was digested with *Hin*dIII, and the resulting fragments were self-ligated (10ng DNA per 30ul ligation reaction). *E. coli* cells were transformed with the self-ligated fragments and transformants were screened for ampicillin resistance on YT agar medium containing ampicillin ($100\mu g.ml^{-1}$).

2.3.6. PCR amplification of DNA from the *C. beijerinckii* genome

An inverse PCR strategy, as described by Sambrook *et al.* (1989), was used to amplify the 5'-region of the *C. beijerinckii* sucrose utilization system. The specific details of the inverse PCR strategy used in this study are presented in the results and discussion (Fig. 2.5). The technical aspects are presented below.

The strategy involved the digestion of *C. beijerinckii* chromosomal DNA with *Xmn*I, followed by self-ligation of the digested fragments (200ng DNA per 50ul ligation reaction). Self-ligated chromosomal DNA formed the template on which PCR was carried out. Two primers, primer *a* (5'-TATTATAGGTACAAATACATCTCC) and primer *b* (5'-ATATGTGAGCGAAGAAAATGC) were used for inverse PCR. Their positions relative to the *C. beijerinckii* sucrose utilization system are indicated in Figures 2.5 and 2.6. The following optimized PCR regime was used for 30 cycles: Step 1, template denaturation at 96°C for 30 seconds; Step 2, primer annealing at 50°C for 30 seconds and Step 3, strand elongation at 70°C for 90 seconds. PCR reactions were conducted using a OmniGene thermocycler (Hybaid, Middlesex, UK) and *Taq* DNA polymerase (Advanced Biotechnologies Ltd., U.K.). The amplified fragment was cloned into the *Sma*I site of the low copy number vector pACYC177.

2.3.7. Exonuclease III digestion and nucleotide sequencing.

Overlapping nested deletions were generated in DNA to be sequenced, by Exonuclease III digestion (Henikoff, 1984). Double-stranded templates were sequenced by the dideoxy chain termination method of Sanger *et al.* (1977). Initially, sequencing was conduced using the Sequenase kit version (2.0) (U.S. Biochemical Corp., Cleveland, Ohio) and [^{35}S]-dATP. Subsequently, sequencing was conducted using the Sequitherm™ kit (Epicentre Technologies, Madison, USA) using CY-5™ labeled M13 primers and an OmniGene thermocycler (Hyband, Middlesex, UK).

The 5'-region of the *C. beijerinckii* sucrose utilization system was sequenced using primer *a* (described above) and primer *c* (5'-CATATGAATAGGGAGG). Their positions relative to the *C. beijerinckii* sucrose utilization system are indicated in figure 2.5. Where primers a and c were used, sequencing was conducted using the AutoRead ™ sequencing kit (Pharmacia Biotech), in conjunction with CY-5 ™ labelled dATP(Pharmacia Biotech). Where CY-5 ™ label was used, nucleotide sequence was resolved using an Alfexpress™ automated DNA sequencer (Pharmacia Biotech).

Amino acid identities and similarities between protein sequences were determined using the BESTFIT component of the Genetics Computer Group (GCG) sequence analysis package. The minimum free energy of RNA stem loop structures was determined using the FOLD component of GCG (Zuker and Stiegler, 1981). Multiple sequence alignments were conducted using the PILEUP component of GCG. Protein hydrophobicity analysis was conducted according to the algorithm of Kyte and Doolittle (1982), as implemented in the DNAMAN sequence analysis package (Lynnon Biosoft, Vaudreuil, Canada). Phylogenetic analysis was conducted using the ClustalW program of Thompson *et al.* (1994), and the results were rendered using the TreeView software package (R.D.M. Page, 1996, Institute of Biomedical and Life Sciences, University of Glasgow)

2.3.8. DNA-directed cell-free protein synthesis.

Plasmid encoded proteins were investigated using an *E. coli* DNA-directed cell-free transcription/translation system (Promega, Madison, USA). The manufacturer's protocol was followed throughout. The proteins were labeled with L-[^{35} S]methionine and resolved on 12% SDS-PAGE gels as described by Laemmli (1970). The *in vitro* translated proteins were visualized after SDS-PAGE electrophoresis by autoradiography. Molecular weight standards were obtained from Pharmacia Biotech

2.4. RESULTS AND DISCUSSION

2.4.1. Isolation and preliminary sequence analysis of the truncated *scr* gene cluster

The initial aim of the study was to clone sucrose catabolic genes from *C. beijerinckii*. This was initially achieved by screening a *C. beijerinckii* genomic library for clones conferring a sucrose-positive phenotype on *E. coli*. This lead to the isolation of a truncated sucrose utilization (*scr*) gene cluster, which was then subjected to a preliminary sequence analysis.

The *C. beijerinckii* genomic library constructed in pEcoR251, was transformed into *E. coli* JM109 and transformants were screened on sucrose MacConkey agar plates for sucrose fermenting clones. After a 36 hour incubation period, a single red colony was identified out of approximately 3000 colonies screened. The plasmid isolated from this colony was designated pCBS1 and was found to contain a 6.8kb insert (Fig. 2.1). In order to clone the sucrose utilization system on a smaller fragment more suitable for sequence analysis, the insert in pCBS1 was digested with several restriction endonucleases and subcloned into pSK. A 3kb *Sca*I subclone was found to confer a sucrose-positive phenotype on *E. coli* and the corresponding plasmid was designated pCBS2 (Fig. 2.1). Sucrose hydrolase assays confirmed that pCBS2 conferred sucrose hydrolase activity on *E. coli*. A study of the sucrose hydrolase activity is reported in section 3.4.1.

To characterize the gene encoding the sucrose hydrolase and identify other closely linked genes, the insert in pCBS2 was sequenced. This revealed the presence of a truncated open reading frame (ORF) followed by two complete ORFs (Fig. 2.1). The ORFs were present in the same orientation as the one another and the pSK *lac* promoter. The truncated ORF specified a protein with homology to members of the GalR-LacI family of transcriptional regulators (e.g. 31.8% amino acid identity with the *S. xylosus* ScrR; full details are supplied in Table 2.2) and a preliminary designation *scrR* was assigned. The ORF following *scrR* specified a protein with homology to sucrose hydrolases of the glucosyl hydrolase family 32 (e.g. 35.6% amino acid identity with the *B. subtilis* SacA. Full details are supplied in Table 2.3) and a preliminary designation *scrB* was assigned. The ORF following *scrB* specified a protein with homology to ribokinase/*pfkB* family fructokinases (e.g. 36.9% amino acid identity with a fructokinase from *Beta vulgaris*. Full details are supplied in Table 2.4) and a preliminary designation *scrK* was assigned. No ORFs were detected in the 300 bases of sequence downstream of the *scrK* gene. Therefore it appeared that a truncated *scr* gene clusterhad been cloned, consisting of three genes *viz*: a truncated *scrR* and complete *scrB* and *scrK* genes.

A gene specifying a protein involved in sucrose transport, was not present on the insert in pCBS2. This was surprising, since pCBS2 was able to confer a sucrose-positive phenotype on *E. coli*. It is possible that during the 36 hour incubation period required for the sucrose-positive phenotype to develop, some cell lysis occurred, releasing the cloned sucrose hydrolase. Extracellular sucrose hydrolysis would have released glucose and fructose and utilization of these sugars would have resulted in a positive fermentation phenotype.

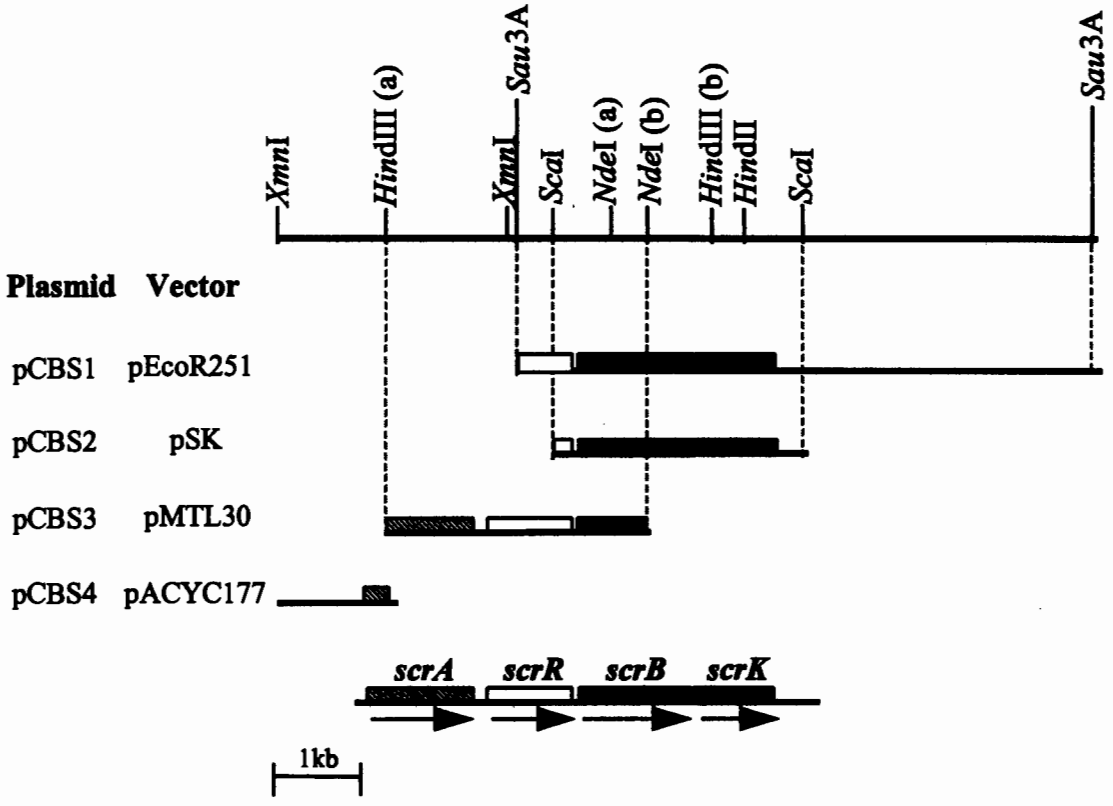

Fig. 2.1 Relationship between subclones of the *C. beijerinckii* sucrose utilization system. Transcriptional polarities are indicated by solid arrows. pCBS1 was isolated from a *C. beijerinckii* genomic library, by its ability to confer a sucrose-positive phenotype on *E. coli*. It contains a 6.8kb insert of *C. beijerinckii* chromosomal DNA cloned into the unique *Bgl*II site of p*Eco*R251. pCBS2 was constructed by cloning the 3kb *Sca*I fragment from the insert in pCBS1 into the unique *Eco*RV site of pSK. pCBS3 was recovered from the genome of a *C. beijerinckii* strain CBSCRB, carrying multiple copies of a pMTL30 derivative integrated into the *scrB* gene. It contains the 3.2kb region of *C. beijerinckii* DNA extending from *Hin*dIII (a) to *Nde*I (b) cloned into the *Hin*dIII and *Stu*I sites of the pMTL30 multiple cloning site (The strategy for the recovery of pCBS3 is illustrated in Figure 2.4). pCBS4 was constructed, first by amplifying a 1.2kb fragment containing the 5'-region of *scrA* by inverse PCR and then cloning this fragment into the unique *Sma*I site in pACYC177 (The inverse PCR strategy is illustrated in Figure 2.5).

2.4.2. Southern blot analysis of the *scrB* and *scrK* genes

Before continuing to isolate the remainder of the *C. beijerinckii* sucrose utilization system, a Southern blot analysis was conducted to confirm that *C. beijerinckii* DNA had indeed been cloned. This was achieved by hybridizing DIG-labeled DNA probes prepared from the cloned *scrB* and *scrK* genes against genomic DNA from *C. beijerinckii* (Fig. 2.2). In addition, genomic DNA from *C. acetobutylicum* NCP262 was included in the analysis, since this strain was used extensively in molasses-based ABE fermentations implemented by National Chemical Products (South Africa) and it was hypothesized that a similar sucrose utilization system might be present. The *scrB* and *scrK* probes hybridized to *C. beijerinckii* genomic fragments of 435 and 420bp respectively, indicating that the insert in pCBS2 originated from *C. beijerinckii*. No hybridization was detected against *C. acetobutylicum* NCP262 DNA. This result serves to emphasize the difference between these two species, which until recently were thought to be members of single species. However, it should be noted that the hybridizations were conducted under stringent conditions and it is entirely likely that under less stringent conditions *scrB* and *scrK* equivalents would have been identified for *C. acetobutylicum* NCP262.

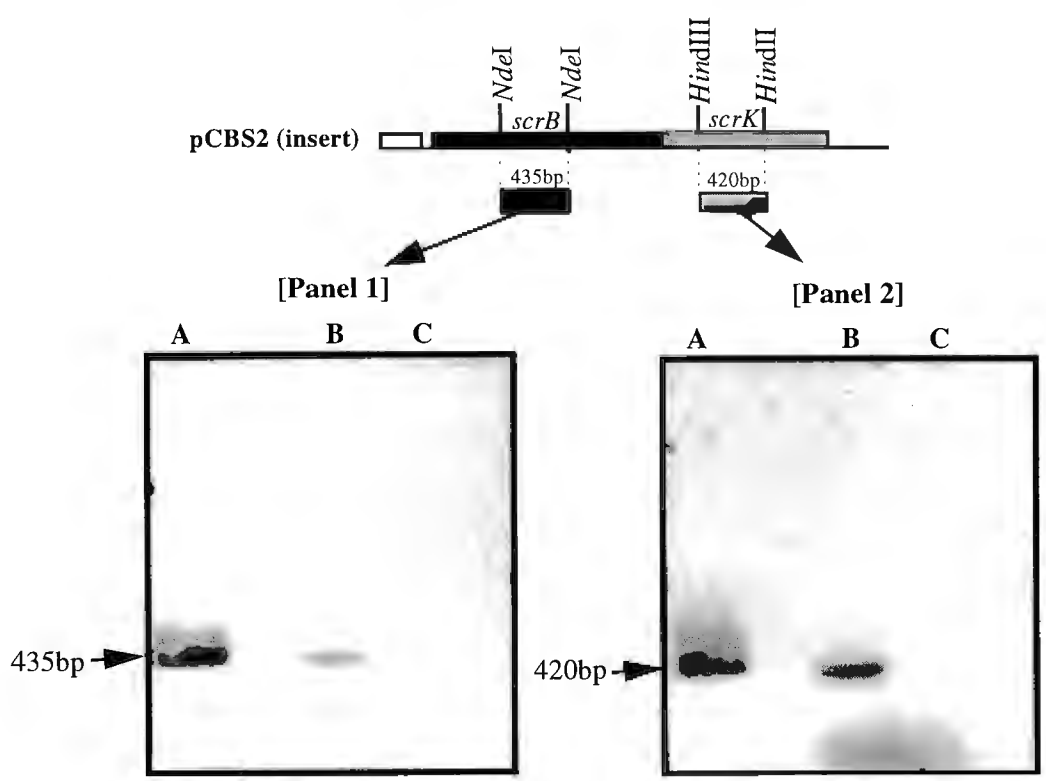

Fig. 2.2 Southern blot analysis of the *scrB* and *scrK* genes. Dig-labelled DNA probes were prepared from a 435bp *Nde*I fragment internal to *scrB* and a 420bp *Hin*dIII/*Hin*dII fragment internal to *scrK*.
Panel 1: The *scrB* probe was hybridized against DNA digested with *Nde*I. The three lanes represent pCBS2 DNA (lane A), *C. beijerinckii* chromosomal DNA (lane B) and *C. acetobutylicum* NCP262 chromosomal DNA (lane C).
Panel 2: The *scrK* probe was hybridized against DNA digested with *Hin*dIII and *Hin*dII. The three lanes represent pCBS2 DNA (lane A), *C. beijerinckii* chromosomal DNA (lane B) and *C. acetobutylicum* NCP262 chromosomal DNA (lane C).

2.4.3. Analysis of translation products produced by pCBS2.

The translation products of pCBS2 and its derivatives were investigated using an *E. coli* DNA-directed cell-free transcription/translation system. In the case of pCBS2, two proteins were produced, corresponding to proteins encoded by its 3kb insert (Fig. 2.3). The molecular weights of these two proteins, determined after PAGE analysis, were 53.2kDa and 32kDa respectively. As regions corresponding to the *scrK* ORF were deleted, so the 32kDa protein was replaced by consecutively smaller proteins (Fig. 2.3, lanes 2 to 4). As regions corresponding to the *scrB* ORF were deleted, so the 53.2kDa protein was replaced by consecutively smaller proteins (Fig. 2.3, lanes 4 and 6). This indicated that the *scrB* and *scrK* genes encoded proteins, with molecular weights approximating the values calculated for the deduced ScrB (57128Da) and ScrK (34297Da) proteins. It is intriguing that the 53.2kDa protein corresponding to the *scrB* gene product became visible only after *scrK* had been almost entirely deleted (compare lanes 1 and 4 in Fig. 2.3.). However, the significance of this result is questionable given the fact that it was observed using an *E. coli* derived *in vitro* transcription/translation system.

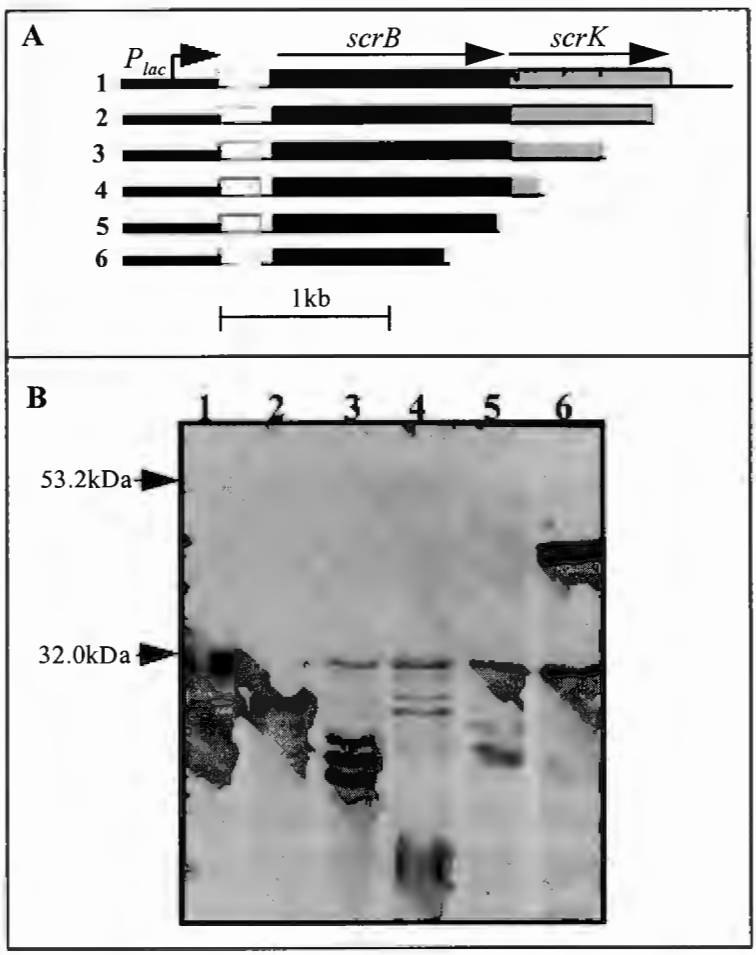

Fig. 2.3 [A] Illustration of inserts in pCBS2 (1) and derivatives generated by exonuclease III digestion (2-6). [B] Autoradiogram of polyacrylamide gel showing translation products of plasmids 1 to 6 (Plasmid and lane numbers correspond). Translation products were generated by *in vitro* coupled transcription-translation and separated by polyacrylamide gel electrophoresis, as indicated in methods.

2.4.4. Isolation of the complete *scrR* gene by plasmid rescue.

Screening of the *C. beijerinckii* plasmid library resulted in the isolation of a *scr* gene cluster containing a truncated *scrR* gene. A thorough analysis of the gene cluster required that the entire *scrR* gene be cloned and sequenced. In concurrent work, reported in Chapter 4, sections 4.3.2 and 4.4.2, a *C. beijerinckii* strain (CBSCRB) was constructed carrying multiple copies of a pMTL30 derivative integrated into the *scrB* gene. The existence of this strain provided for the rapid isolation of the complete *scrR* gene, by recovery of the integrated plasmid. The integrated plasmid was recovered by digesting genomic DNA from *C. beijerinckii* CBSCRB with *Hin*dIII, self-ligating the digested fragments and screening for self-ligated fragments conferring ampicillin resistance on *E. coli* (Fig. 2.4). Only cells transformed with self-ligated chromosomal fragments, containing previously integrated plasmids, would replicate and confer ampicillin resistance on *E. coli*. The recovered plasmid, containing the adjacent chromosomal DNA, was designated pCBS3 (Fig. 2.1 and 2.4).

Sequence analysis of the insert in pCBS3 indicted that the complete *scrR* gene had been isolated. In addition, however, a second truncated ORF was identified upstream and in the same orientation as *scrR*. This ORF specified a protein with homology to EIIBC^{Scr} PTS proteins (Table 2.1.) and a preliminary designation *scrA* was assigned.

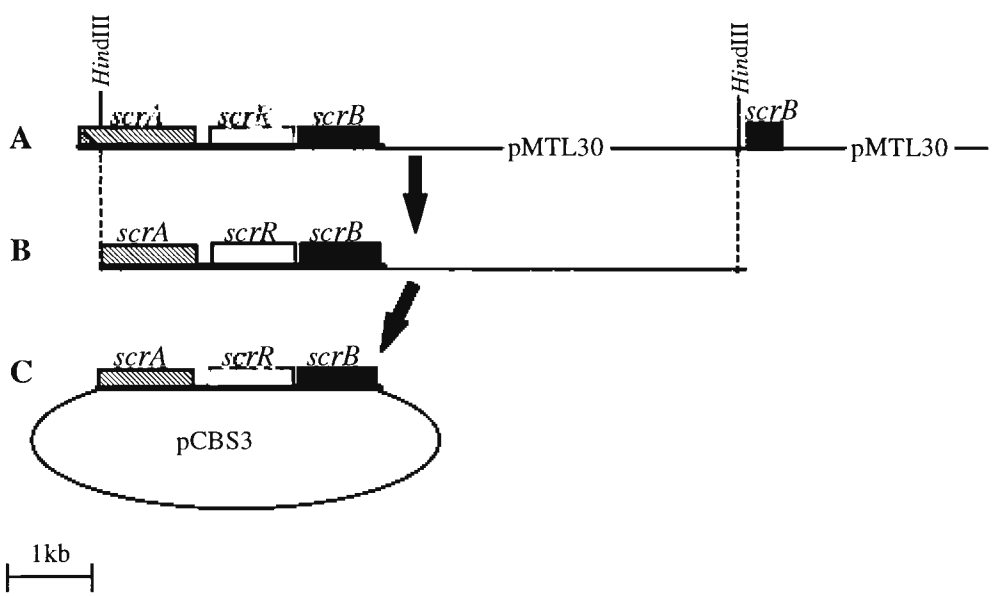

Fig. 2.4 Plasmid recovery strategy used to clone the complete *scrR* gene. (A) Chromosomal DNA was isolated from a *C. beijerinckii* mutant strain CBSCRB carrying multiple copies of a pMTL30 derivative integrated into the *scrB* gene. (B) The isolated DNA was subject to restriction endonuclease digestion with *Hin*dIII. (C) The digested DNA was self-ligated and transformed into *E. coli* JM109. Cells carrying self-ligated regions of DNA containing derivatives of pMTL30 were selected on YT agar plates containing ampicillin. The recovered plasmid, containing adjacent chromosomal DNA, was designated pCBS3.

2.4.5. Isolation of the 5'-region of the *scrA* gene by inverse PCR.

In order to conduct a comprehensive analysis of the *scr* gene cluster, it was necessary to clone and sequence the entire *scrA* gene. Attempts were made to clone *scrA* using partial *Hin*dIII digests of *C. beijerinckii* CBSCRB DNA, and the plasmid recovery strategy outlined above. While plasmids equivalent to pCBS3 were readily recovered, regions further upstream were never obtained. This suggested that the region upstream of the *scr* genes contained a DNA motif or encoded a protein that was lethal when transformed into *E. coli*. Such an observation has been made for the sucrose utilization system of *L. lactis*, where the promoter region of the *scrB* gene was found to be unstable when cloned on high copy number plasmids in *E. coli* (Rauch and de Vos, 1992b). Maintenance of the *L. lactis scrB* promoter could, however, be achieved on low copy number plasmids in *E. coli*. It has also been noted that overexpression of membrane-bound PTS proteins may be lethal to host cells (Lengeler *et al.*, 1994). This has been specifically observed for the *P. pentosaceus scrA* gene, which was unstable when cloned in *E. coli* (K. Leenhouts pers. comm.). In the light of this, an inverse PCR strategy was developed to amplify the 5'-region of the *scrA* gene and clone it on a low copy number plasmid (pACYC177). The details of the inverse PCR strategy are indicated in Figure 2.5. Sequencing of the fragment recovered by inverse PCR, using primers *a* and *c*, revealed that the 5'-region of the *scrA* ORF had been cloned.

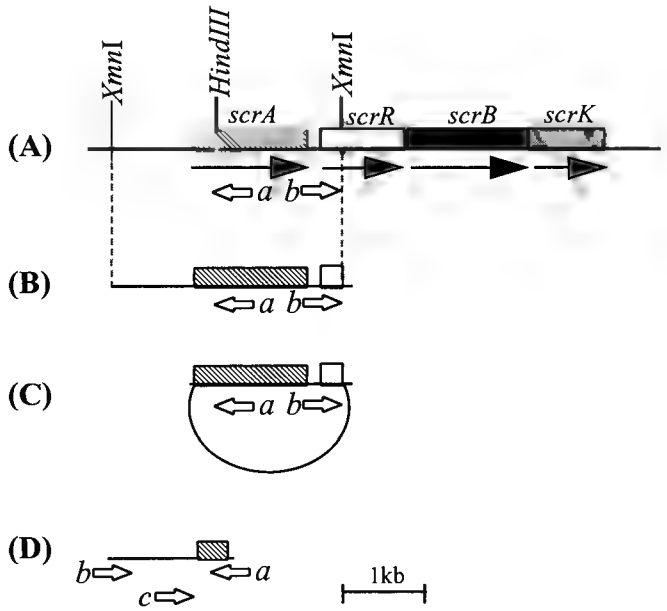

Fig. 2.5 Inverse PCR strategy used to clone the 5'-region of the *scrA* gene. (A) & (B) *C. beijerinckii* chromosomal DNA was isolated and digested with *Xmn*I. (C) & (D) The digested DNA was self-ligated and used as a template for PCR reactions. This resulted in the amplification of a 1.2kb fragment containing the 5'-region of *scrA* and adjacent upstream DNA. The amplified PCR product was cloned into the unique *Sma*I site of pACYC177, yielding plasmid pCBS4. Short open arrows correspond to the approximate positions of primers *a*, *b* and *c*. Primers *a* and *b* were used for inverse PCR. Primer *a* was used to sequence the 5'-region of the *scrA* gene, in one direction. This sequence was used to prepare primer *c*, which was used to sequence in the opposite direction.

2.4.6. Nucleotide sequence analysis of the complete *scrARBK* gene cluster.

Having cloned and sequenced all four *scr* genes, it was possible to conduct a comprehensive sequence analysis. The *scrA, scrR, scrB* and *scrK* genes were clustered and present in the same orientation on the *C. beijerinckii* genome. As indicated above, *scrA* was proposed to encode an EIIBC^{Scr} protein, *scrR*, a GalR-LacI family regulatory protein, *scrB*, a sucrose hydrolase and *scrK*, a fructokinase. The largest intergenic region was present between the *scrA* and *scrR* genes, which were separated by 190bp (Fig. 2.6). The *scrR* and *scrB* genes were separated by only 26bp, while the putative termination and initiation codons of the *scrB* and *scrK* genes overlapped. Attempts to identify promoters within the sequenced DNA, resulted in the identification of two putative promoters, both of which were similar to *E. coli* σ^{70} and clostridial consensus promoter sequences (Young *et al.*, 1989). The first putative promoter (TTGACA-N18-TATCAT) was present from -51 to -80bp upstream of the proposed *scrA* initiation codon. The second putative promoter (TTGACA-N17-TATATT) was identified from -129 to -155bp upstream of the proposed *scrK* initiation codon, within the *scrB* ORF. Promoters have previously been identified within the 3'-regions of *scrB* genes, as is the case with the *scrR* promoter of the pUR400 sucrose utilization system (Jahreis and Lengeler, 1993). Putative promoters were not identified upstream of the *scrR* and *scrB* genes.

Attempts to identify potential transcriptional terminators resulted in the identification of a stem loop structure 21bp to 59bp downstream of the proposed *scrK* termination codon (-15 kcal/mol) (Zuker and Stiegler, 1981) (Fig. 2.6). The presence of this structure immediately downstream of the last gene of the *scr* gene cluster is consistent with the proposal that it acts as a transcriptional terminator. Putative transcriptional terminators were not identified between the individual genes of the *scr* cluster, suggesting that they may be transcribed as an operon. This hypothesis was further investigated by Northern blot analysis (section 4.4.1.).

Putative ribosome-binding-sites were identified for each of the *scr* genes. The sites are: (5'-AGGAGG) 8bp upstream of the putative *scrA* initiation codon; (5'-TGGAGG) 5bp upstream of the putative *scrR* initiation codon; (5'-AGGAGA) 9bp upstream of the putative *scrB* initiation codon and (5'-AGGAGT) 7bp upstream of the putative *scrK* initiation codon (Fig. 2.6). Clearly, for the *scrA, scrB* and *scrK* genes, putative ribosome-binding-sites are separated from their initiation codons by 7 to 9bp. In the case of the *scrR* gene, however, the equivalent

distance is only 5bp. Vellanoweth and Rabinowitz (1992) demonstrated that for *B. subtilis*, translational yield can be substantially reduced if the spacing between a ribosome-binding-site and initiation codon is less than 6bp. Therefore it is possible that the sub-optimal spacing between the putative *scrR* ribosome-binding-site and initiation codon, may play a role in limiting the amount of ScrR translated.

It was proposed that *scrR* specifies a transcriptional regulator of the GalR-LacI family. It was therefore of interest to identify potential sites at which the *scrR* gene product might act. GalR-LacI family regulators typically act at imperfectly palindromic operator sequences centered around a ($A_4A_3N_2C_1/G'_1N'_2T'_3T'_4$) consensus (Weickert and Adhya, 1992). An imperfect palindrome ($G_6A_5A_4A_3A_2C_1/G'_1G'_2T'_3T'_4T'_5C'_6$) was identified between the putative *scrA* promoter and initiation codon, as might be expected for an operator site (Fig. 2.6). However, further experimental analysis is required to establish whether this site is involved in the regulation of *scr* gene transcription. In this regard, it is also important to note that GalR-LacI family regulators other than the *scrR* gene product may regulate *scr* gene transcription. More specifically, catabolite control proteins (CcpA), which are members of the GalR-LacI family, have been characterized for several low-GC Gram-positive bacteria (Saier *et al.*, 1996). It is entirely possible that a CcpA equivalent exists in *C. beijerinckii*, which could act at palindromic operator sites within the *scr* gene cluster.

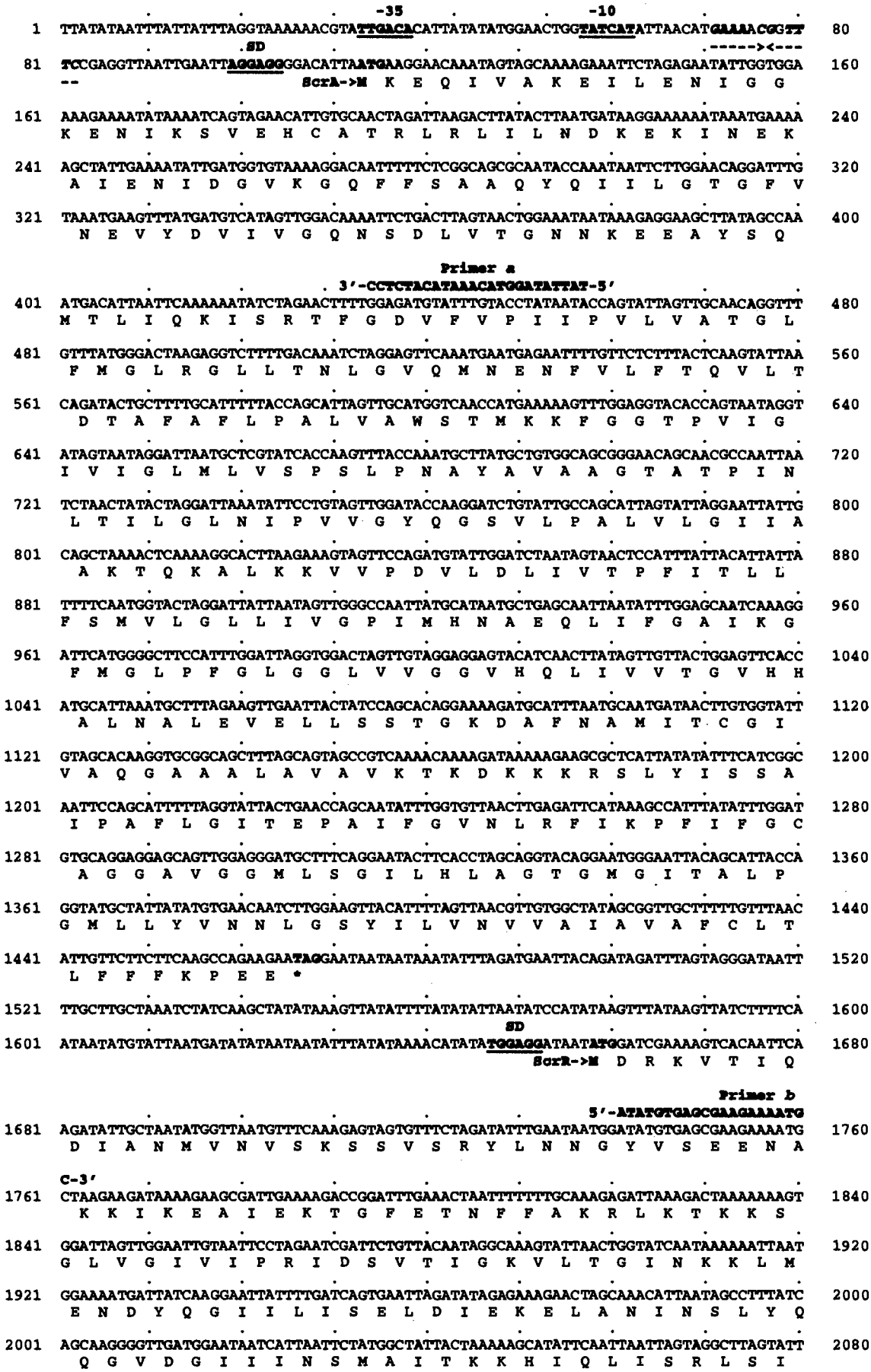

Fig. 2.6 Nucleotide sequence of the *scr* gene cluster. Full legend appears on page 64

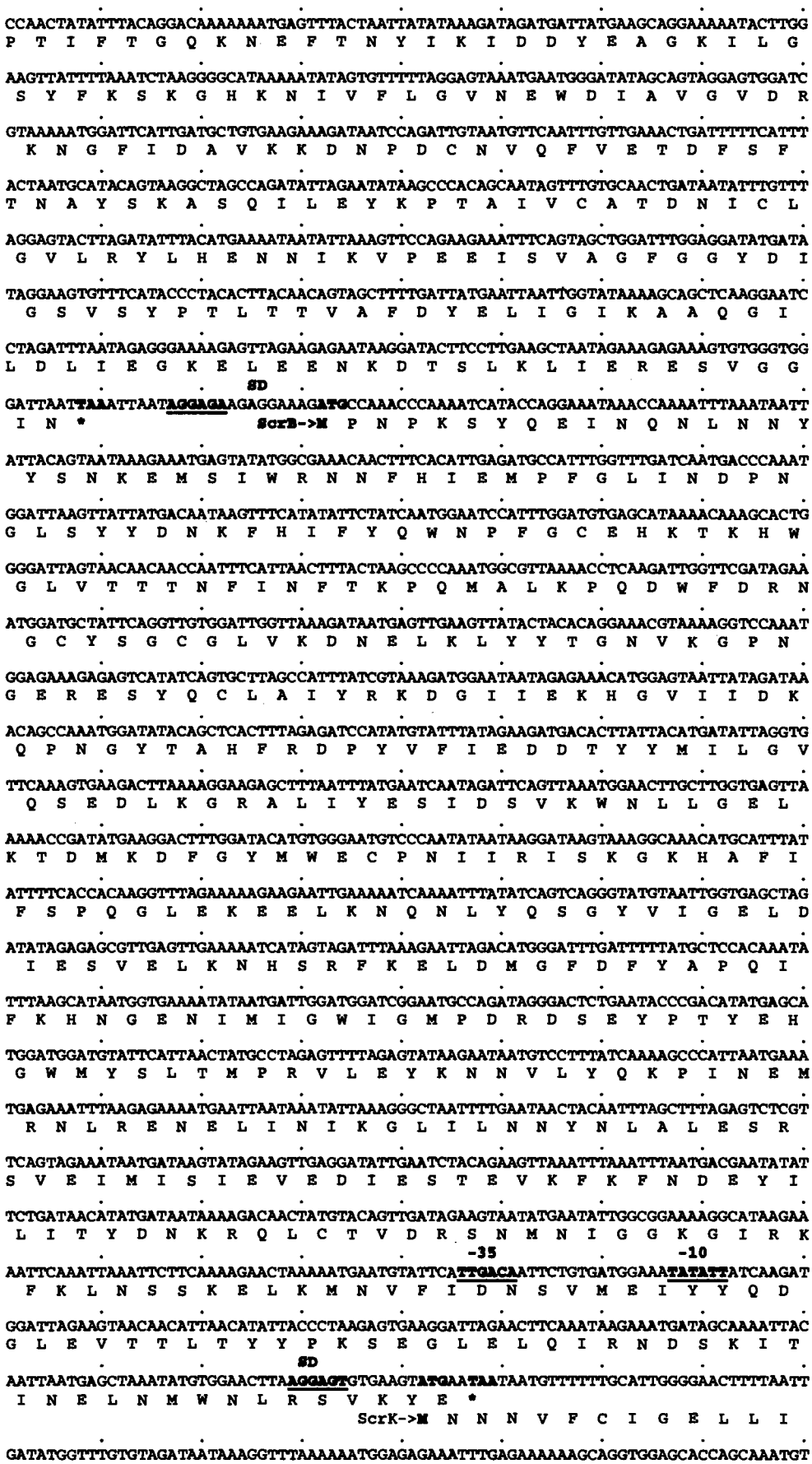

Fig. 2.6 Nucleotide sequence of the *scr* gene cluster. Full legend appears on page 64

```
4241  AGCGGCGAGTATTTGTAAATTAGAAGGTAATGCTTATTTTTTAGGACAAGTAGGTAATGATTTTTTTGGTAAGCATTTAG  4320
       A  A  S  I  C  K  L  E  G  N  A  Y  F  L  G  Q  V  G  N  D  F  F  G  K  H  L  V

4321  TTGAGCTGCTAAAAAGCTTGAATATAAATACTGATATGACTGTAGAAAGAGGAAGTACAACAATTGCATTAGTTGGAATT  4400
       E  L  L  K  S  L  N  I  N  T  D  M  T  V  E  R  G  S  T  T  I  A  L  V  G  I

4401  GATGCAAATGGTGAACGTAACTTTGACTTCCTAAGAGGTAGTGATGGGGAATATTCGTTTGATAATATAGATTTATCAAA  4480
       D  A  N  G  E  R  N  F  D  F  L  R  G  S  D  G  E  Y  S  F  D  N  I  D  L  S  K

4481  GATAAGTAATTCAGATATAATTCATTTTGGGTCTGCAACGGGATTTTTAGATGGAGAGTTAAAAAAAGACATATTTTAAAC  4560
       I  S  N  S  D  I  I  H  F  G  S  A  T  G  F  L  D  G  E  L  K  K  T  Y  F  K  L

4561  TATTAGAATATGCGAAATCTAATAACATATATGTATCATTTGATCCTAATTACCGCGATGCACTAATCACTCAAGACAAG  4640
       L  E  Y  A  K  S  N  N  I  Y  V  S  F  D  P  N  Y  R  D  A  L  I  T  Q  D  K

4641  TTAGAGTTGTTTGTTGAGGATTGTGTTAAATTCTTGAGATATAGTAACTTTACGAAACTAAGTGATGAAGAATTATATTT  4720
       L  E  L  F  V  E  D  C  V  K  F  L  R  Y  S  N  F  T  K  L  S  D  E  E  L  Y  L

4721  AATAACAAAAGAAAAAGATATAGAAAGCGCTGTTAACAAGTTACATGAAATTGGAGTTAAGGTTGTAACAGTTACTTTAG  4800
       I  T  K  E  K  D  I  E  S  A  V  N  K  L  H  E  I  G  V  K  V  V  T  V  T  L  G

4801  GGTCTAAGGGCACATACTTAAGTGTTAATGGAAAAAATGAAATAATACCATCTATAGAAATTAAGCAAGTAGATTCCACT  4880
       S  K  G  T  Y  L  S  V  N  G  K  N  E  I  I  P  S  I  E  I  K  Q  V  D  S  T

4881  GGTGCAGGAGATGCCTTTGTTGGAGCAGTATTAAGACAGGTATCTGATATTGATGACAAAGAGAGTATAAGTTTTGATAG  4960
       G  A  G  D  A  F  V  G  A  V  L  R  Q  V  S  D  I  D  D  K  E  S  I  S  F  D  R

4961  ATGGAAAGAGATAGTTTCATTTGCAAATAAGGTAGGTGCAATAACATGCACCAACTACGGGGCTATAGCCTCAATGCCGA  5040
       W  K  E  I  V  S  F  A  N  K  V  G  A  I  T  C  T  N  Y  G  A  I  A  S  M  P  T

5041  CACTAAACGAATTAAATTAAGCCGAGTTAGATGAATTCATATATAGGAGGGAGATATTTTAAATATATCCCTTCTATATT  5120
       L  N  E  L  N  *                        ------ ----- -----> <----- ----- ------

5121  TTTGGATTTTTATTTGATTATTGAATATAAGTTAATAATAATGAGTATGAAACTTATGATTTTATATTTTTATTATATGG  5200

5201  TAGAATATTGAGAAAGGAACATATATAATATTAAAAATATTTAGTCAAGAAAATTCATGATGGGAGCATTACAAATGGTT  5280

5281  GATTTTCTAAAAATTGCAAAAGAAGAAGCTAAGATGGCAATGAAAAAAGGGGAGATTCCAGTAGGAGCAGTAATTGTATT  5360

5361  AGATGATAAAATAATAG  5377
```

Fig. 2.6 Nucleotide and deduced amino acid sequence of the *C. beijerinckii scrARBK* gene cluster. Deduced aa sequences are shown in single-letter code below the coding sequences. Putative initiation and termination codons for the *scrA*, *scrR*, *scrB* and *scrK* genes are indicated in bold type. Putative promoters (-35 and -10 regions) and ribosome-binding sites (SD) are in bold type and underlined. The facing arrows, -31 to -42bp upstream of the proposed *scrA* initiation codon, indicate an imperfect palindrome. This site may act as an operator site for a GalR-LacI family regulatory protein. The facing arrows 21 to 59bp downstream of *scrK* indicate an inverted repeat sequence, representing a potential transcriptional terminator, with a predicted minimum free energy of (-15 kcal/mol) (Zuker and Stiegler, 1981). The positions and sequences of Primers *a* and *b* used in inverse PCR are indicated in bold type above the corresponding nucleotide sequence. The position of primer *c* has not been indicated, since it lies in a region sequenced only in one direction.

2.4.7. Analysis of the deduced *C. beijerinckii* ScrA and homologous proteins.

Analysis of several sucrose utilization systems has indicated that in almost all cases, sucrose transport occurs via the PTS. In the cases of *S. mutans*, *S. sobrinus* and *P. pentosaceus*, EII^{Scr} proteins consist of EIIB, EIIC and EIIA domains, fused by linker regions ($EIIBCA^{Scr}$ proteins) (Sato *et al.*, 1989; Chen *et al.*, 1993; GenBank accession L32093). In the remaining examples the EII^{Scr} proteins simply consist of fused $EIIBC^{Scr}$ domains. In all cases, the $EIIBC^{Scr}$ domains are highly homologous to one another (Fig. 2.7) (Titgemeyer *et al.*, 1996).

The *C. beijerinckii scrA* gene specifies a deduced protein of 451aa, with a calculated molecular weight of 47641Da and homology to previously studied EIIBC^{Scr} proteins (Table 2.1). Several particularly well conserved regions have been identified for EIIBC^{Scr} proteins and were discussed in section 1.2.1. It is evident from a multiple sequence alignment of EIIBC^{Scr} proteins that these regions have been conserved in the deduced *C. beijerinckii* ScrA (Fig. 2.7). The most highly conserved regions include: the (HCATRLRL) motif, containing the proposed EIIB cysteine phosphorylation site; the region corresponding to the first putative transmembrane α-helix; and the GITE motif, present on what is proposed to be an intracellular loop towards the C-terminal end of the EIIC domain (Fig. 1.5 and 2.7). The GITE motif, or similar motifs (GIHE, FASE, QINE, KLTE, GVNE), have been found in all EIIC PTS domains and are characterized by the presence of the "E" Glu amino acid residue (Titgemeyer *et al.*, 1996). The GITE motif has been proposed to be involved in phosphate and/or substrate binding (Lengeler *et al.*, 1994).

TABLE 2.1. Percent identity and similarity of the of the amino acid sequence of the deduced *C. beijerinckii* ScrA to known protein sequences.

Putative *C. beijerinckii* protein	Homologous proteins	Accession number[a]	%identity	%Similarity
ScrA	*Streptococcus sobrinus* ScrA	(PIR: S68599)	55.1	77.4
	Klebsiella pneumoniae ScrA	(sp: P27219)	40.6	67.6
	Salmonella typhimurium ScrA	(sp: P08470)	40.4	67.1
	Vibrio alginolyticus ScrA	(sp: 22825)	38.8	67.1
	Staphylococcus xylosus ScrA	(sp: P51184)	36.2	66.6
	Pediococcus pentosaceus ScrA	(sp: P43470)	34.5	61.6
	Streptococcus mutans ScrA	(sp: P12655)	34.8	63.1
	Bacillus subtilis SacP	(sp: P05306)	35.7	61.0

[a] sp: Swiss Prot; PIR: Protein Information Resource

An analysis of overall amino acid sequence identity indicates that the deduced *C. beijerinckii* ScrA is particularly homologous to the *S. sobrinus* ScrA (55.1% identity over the corresponding EIIBC domains) (Table 2.1). However, the deduced *C. beijerinckii* ScrA represents an EIIBC^{Scr} protein, while the *S. sobrinus* ScrA represents an example of a fused EIIBCA^{Scr} protein. Furthermore, the sucrose utilization system of *S. sobrinus* is structurally similar to the *S. mutans* and *P. pentosaceus* systems, which also encode fused EIIBCA^{Scr} proteins (Section 1.3.3.4). In these cases, however, the *C. beijerinckii* EIIBC^{Scr} shares only 34.5% to 34.8% identity with equivalent EIIBC^{Scr} domains. This raises questions about the relationship between the ScrA proteins of *C. beijerinckii* and *S. sobrinus*. They no doubt share

a common origin with the remaining ScrA proteins, however, the high degree of homology between these two proteins, raises the possibility of an even closer evolutionary relationship.

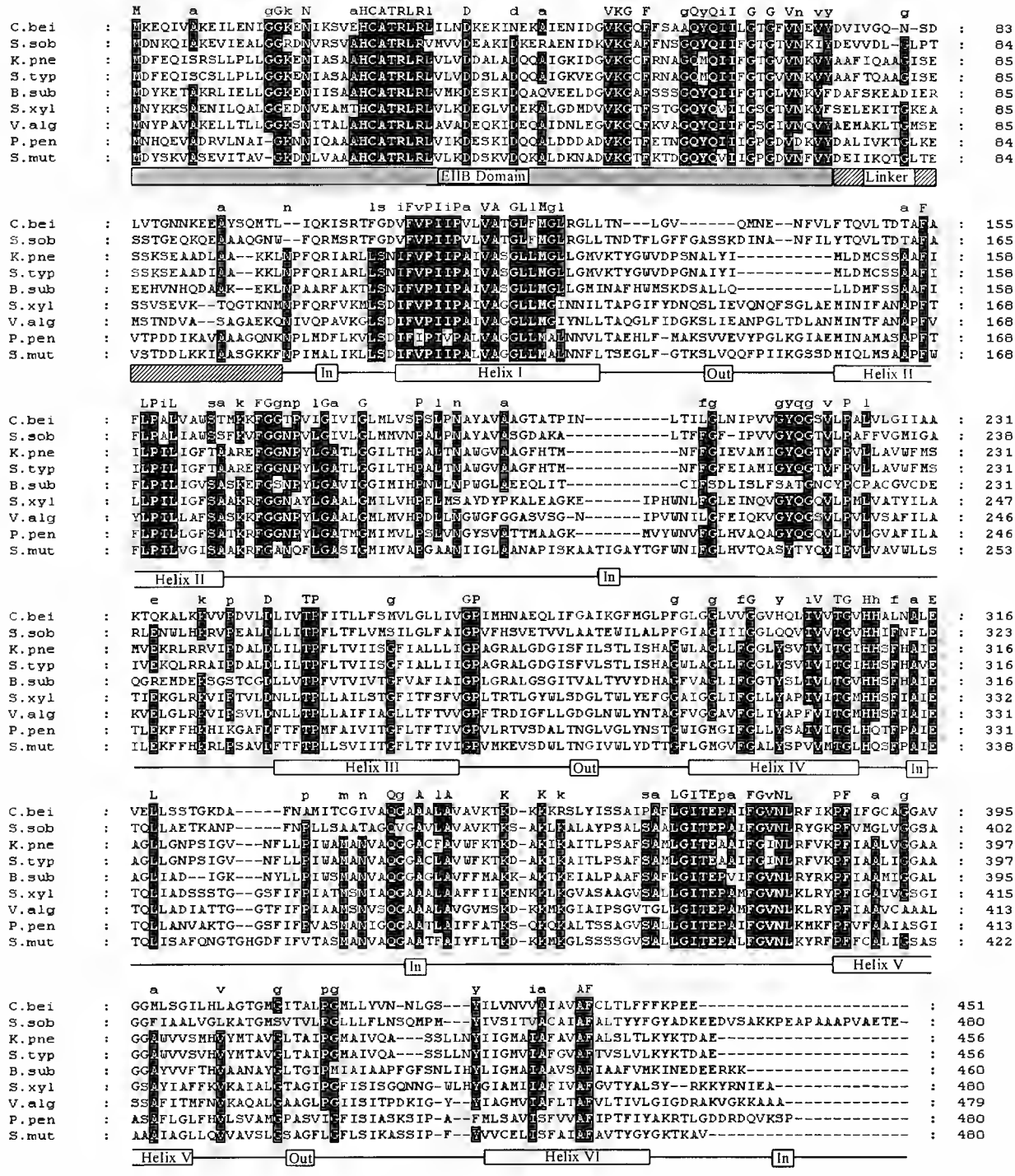

Fig. 2.7 Multiple sequence alignment of bacterial EIIBC^{Scr} domains. Sequences included are: *C. beijerinckii* ScrA (C.bei), *S. sobrinus* ScrA (S.sob), *K. pneumoniae* ScrA (K.pne), *S. typhimurium* ScrA (S.typ), *B. subtilis* SacP (B.sub), *S. xylosus* ScrA (S.xyl), *V. alginolyticus* ScrA (V.alg), *P. pentosaceus* ScrA (P.pen), and *S. mutans* ScrA (S.mut). Sequence accession numbers are supplied in Table 2.1. Shaded areas represent identical amino acids. The two dimensional structure, proposed by Titgemeyer *et al.* (1996) for the *K. pneumoniae* ScrA, is represented below the aligned sequences. Intracellular loops [IN] extracellular loops [OUT] and six putative transmembrane helices [Helix I to VI].

Hydrophobicity analysis revealed that the deduced *C. beijerinckii* ScrA is characterized by an N-terminal hydrophilic domain of approximately 112aa, followed by an essentially hydrophobic domain (Fig. 2.8). This hydrophobicity profile reflects the fact that the N-terminal region corresponds to the EIIB^{Scr} domain and linker region, and is likely to have an intracellular location, while the remainder of the protein corresponds to the EIIC^{Scr} domain, and is likely to be an integral membrane component.

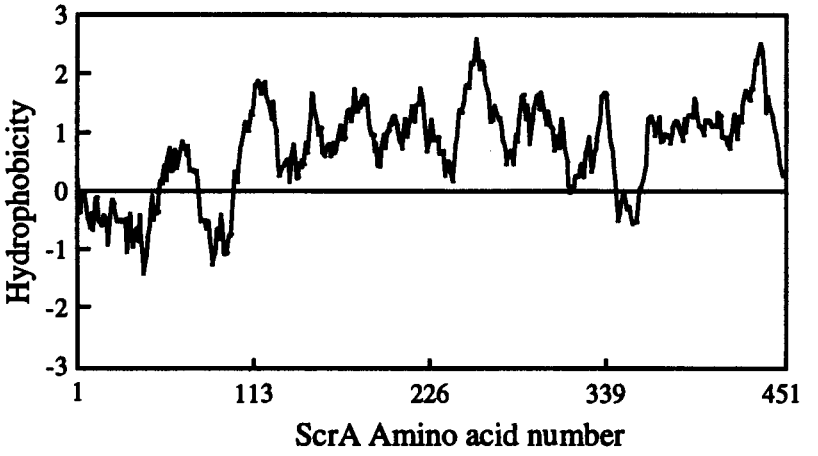

Fig. 2.8 Hydropathy plot of the deduced *C. beijerinckii* ScrA (EIIBC^{Scr}) protein (Kyte and Doolittle, 1982). The window size chosen was 20aa.

Since the *scrA* of *C. beijerinckii* appears to encode an EIIBC^{Scr} protein, it must therefore interact with an EIIA domain encoded elsewhere on the chromosome. The possibility exists that a separate EIIA^{Scr} domain is encoded elsewhere on the chromosome. However, in *B. subtilis* the EIIA^{Glc} domain of the EIICBA^{Glc} protein has been shown to phosphorylate EIIB^{Scr} (Sutrina *et al.*, 1990). In studies with the *K. pneumoniae*, *S. typhimurium* and *V. alginolyticus* sucrose utilization systems, cloned into *E. coli*, it was demonstrated that EIIBC^{Scr} dependent sucrose uptake required the presence of a functional *E. coli* EIIA^{Glc} (Blatch *et al.*,1990; Lengeler *et al.*, 1982; Sprenger and Lengeler, 1988). It is therefore possible that the *C. beijerinckii* EIIB^{Scr} domain interacts with a shared EIIA domain, possibly that of the glucose PTS.

2.4.8. Analysis of the deduced *C. beijerinckii* ScrR and homologous proteins.

The *C. beijerinckii scrR* gene specifies a deduced protein of 330aa, with a calculated molecular weight of 36602Da and homology to members of the GalR-LacI family of transcriptional regulators (Table 2.2). In order to establish the relationship between the deduced *C. beijerinckii* ScrR and other GalR-LacI family members, a phylogenetic analysis was conducted using the ClustalW algorithm of Thompson *et al.* (1994). To place the present

study in context, the findings of a similar study, conducted by Nguyen and Saier (1995), will be considered.

TABLE 2.2. Percent identity and similarity of the of the amino acid sequence of the deduced *C. beijerinckii* ScrR to known protein sequences.

Putative *C. beijerinckii* protein	Homologous proteins	Accession number[a]	%Identity	%Similarity
ScrR	*Staphylococcus xylosus* ScrR	(EMBL: X67744)	31.8	57.5
	C. acetobutylicum NCP 262 RegA	(gb: L14685)	30.8	52.4
	Bacillus subtilis CcpA	(sp: P25144)	29.4	51.5
	Pediococcus pentosaceus ScrR	(sp: P43472)	26.0	46.6
	Escherichia coli PurR	(sp: P15039)	25.6	51.7
	Streptococcus mutans ScrR	(gb: U46902)	24.6	50.0
	Salmonella typhimurium ScrR	(sp: P37077)	22.1	45.0
	Klebsiella pneumoniae ScrR	(sp: P37076)	21.7	45.0
	Escherichia coli CscR	(sp: P40715)	22.5	46.3

[a] sp: Swiss Prot; gb: GenBank; EMBL: European Molecular Biology Laboratory

Nguyen and Saier (1995) conducted a phylogenetic analysis involving twenty-five GalR-LacI members. This resulted in a phylogenetic tree, with approximately equal branch lengths, stemming from a point close to the center of the tree. Consequently, it was proposed that most GalR-LacI family members evolved by gene duplication events, occurring at a specific time in evolutionary history, and that further duplication events were rare. The authors included two GalR-LacI family members, involved in the regulation of sucrose catabolic genes, the *K. pneumoniae* ScrR and the *E. coli* CscR. Despite both being involved in sucrose metabolism, it appeared that these two proteins had evolved separately within the family. This prompted the phylogenetic analysis conducted in the present study, with the aim of identifying whether subsequently characterized ScrR proteins would also represent independent lines of evolution within the GalR-LacI family.

The same GalR-LacI regulatory proteins selected by Nguyen and Saier (1995), were subjected to a phylogenetic analysis in conjunction with recently isolated ScrR proteins. The resulting phylogenetic tree was similar to that presented by Nguyen and Saier (1995), with almost all branches stemming from a point close to the center of the tree (Fig. 2.9). The phylogenetic tree suggests that within the GalR-LacI family, proteins involved in the regulation of sucrose catabolic genes have evolved at least four times (A to D in Figure 2.9).

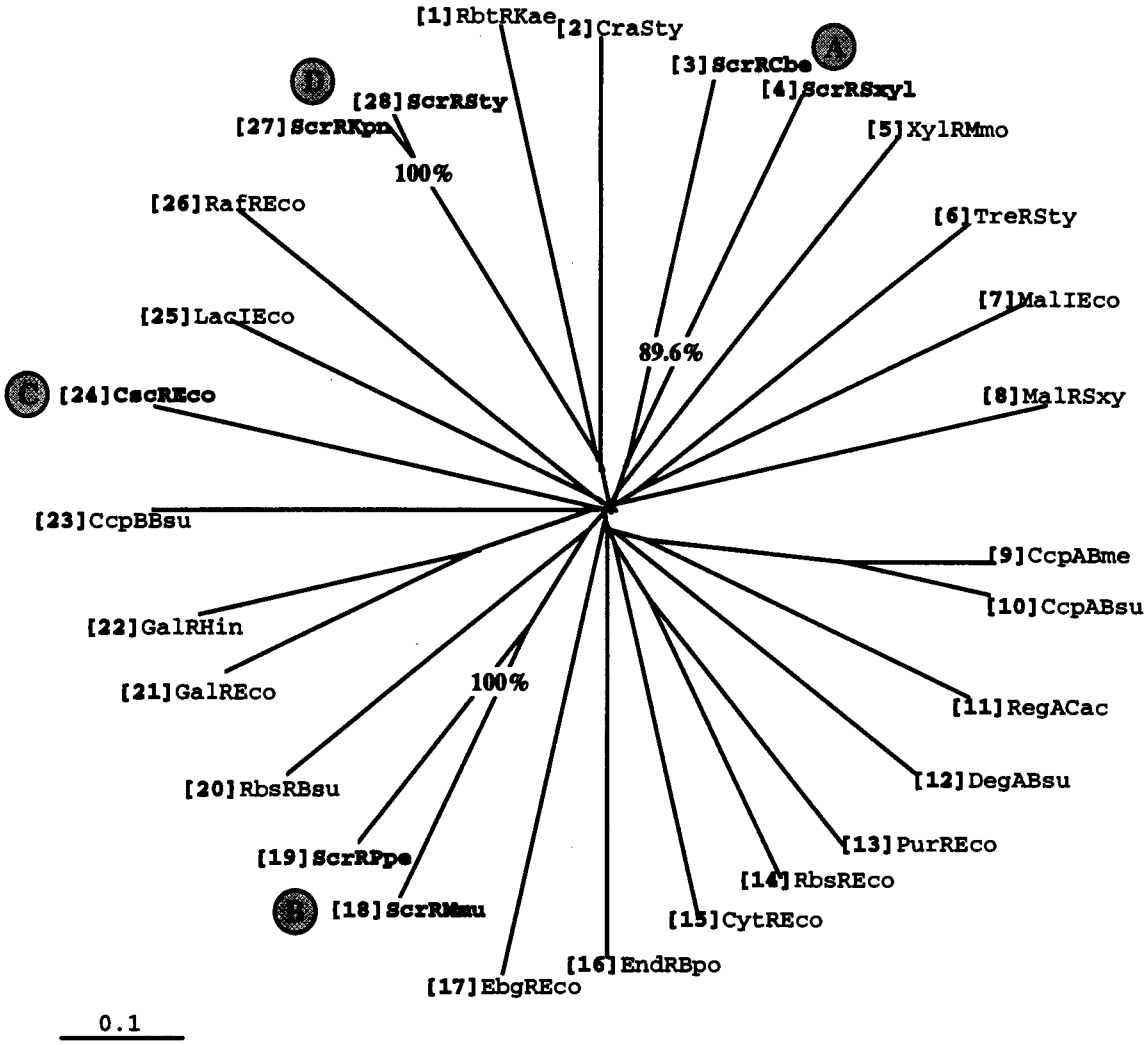

Fig. 2.9 Phylogenetic tree for the proteins of the GalR-LacI family. The tree was constructed using the ClustalW program of Thompson *et al.* (1994). Branch lengths are approximately proportional to the amount of sequence difference. The scale of 0.1 indicates 10% amino acid sequence divergence. In the interests of clarity, bootstrap values, as a percentage of 1000 trials, are presented only at branch-points involving ScrR proteins. The four evolutionary lines involving regulators of sucrose catabolic genes are indicated (A to D). Sequence abbreviations and accession numbers (gb: GenBank and sp: SWISS-PROT) are **[1]** RbtRKae: *Klebsiella aerogenes* Ribitol repressor; sp: P07760 **[2]** CraSty: *Salmonella typhimurium* Catabolite Repressor/Activator protein, sp: P21930 **[3]** ScrRCbe: *Clostridium beijerinckii* ScrR **[4]** ScrRSxyl: *Staphylococcus xylosus* ScrR, EMBL: X67744 **[5]** XylRMmo: *Morganella morganii* Xylitol repressor, gb: L34345 **[6]** TreRSty: *S. typhimurium* Trehalose repressor, sp: P36674 **[7]** MalIEco: *E. coli* maltose repressor, sp: P18811 **[8]** MalRSxy: *S. xylosus* maltose repressor, PIR: S44187 **[9]** CcpABme: *Bacillus megaterium* Catabolite control protein A (CcpA), gb: L26052 **[10]** CcpABsu: *Bacillus subtilis* CcpA sp: P25144 **[11]** RegACac *Clostridium acetobutylicum* RegA, gb: L14685 **[12]** DegABsu: *B. subtilis* Degradation enzyme activator, sp: P37947 **[13]** PurREco: *E. coli* purine repressor, sp: P15039 **[14]** RbsREco: *E. coli* ribose repressor, gb: L10328 **[15]** CytREco: *E. coli* cytidine repressor, sp: P06964 **[16]** EndRBpo: *Bacillus polymyxa* endogluconase repressor, sp: P27871 **[17]** EbgREco: *E. coli* evolved β-galactosidase repressor, sp: P06846 **[18]** ScrRMmu: *Streptococcus mutans* ScrR, gb: U46902 **[19]** ScrRPpe: *Pediococcus pentosaceus* ScrR, sp: P43472 **[20]** RbsRBsu: *B. subtilis* Ribose repressor, sp: P36944 **[21]** GalREco: *E. coli* GalR, sp: P03024 **[23]** GalRHin: *Haemophilus influenzae*, sp: P31766 **[23]** CcpBBsu: *B. subtilis* Catabolite control protein B (CcpB), sp: P37517 **[24]** CscREco: *E. coli* CscR, sp: P40715 **[25]** LacIEco: *E. coli* lactose repressor, sp: P03023 **[26]** RafREco: *E. coli* raffinose repressor, sp: P21867 **[27]** ScrRKpn: *K. pneumoniae* ScrR, sp: P37076 **[28]** ScrRSty: *S. typhimurium* ScrR, sp: P37077.

The first evolutionary line involves the *S. xylosus* and *C. beijerinckii* ScrR proteins (Fig. 2.9 (A)). The bootstrap value at the *S. xylosus* and *C. beijerinckii* ScrR branch, is 89.6%, providing confidence that they are more closely related to one another, than to other GalR-

LacI members. The second line involves the *P. pentosaceus* and *S. mutans* ScrR proteins (Fig. 2.9 (B)). The bootstrap value at the *P. pentosaceus* and *S. mutans* ScrR branch, is 100% and these two proteins appear to have diverged from one another more recently than the ScrR proteins of *S. xylosus* and *C. beijerinckii*. The third and fourth lines lead to the CscR of *E. coli* (Fig. 2.9 (C)). and the closely related ScrR proteins of *K. pneumoniae* and *S. typhimurium* respectively (Fig. 2.9 (D)). Despite almost certainly being involved in sucrose metabolism, the proteins of the four evolutionary lines are no more related to each other than to other GalR-LacI family members, responding to a diverse range of ligands. In addition, it should also be noted that although the ScrR protein of *V. alginolyticus* was not included in the analysis, since it is less than one third the size of other GalR-LacI proteins, it does have homology to members of this family (Weickert and Adhya, 1992), and may represent an additional evolutionary line leading to a regulatory protein involved in sucrose metabolism.

A possible explanation for the lack of homology between the various GalR-LacI regulators involved in sucrose metabolism, may be that they respond to different metabolites signaling the presence of sucrose. In the case of the *K. pneumoniae* and *S. typhimurium* ScrR proteins, the inducer molecule has been demonstrated to be fructose or fructose-1-phosphate (Jahreis and Lengeler, 1993). In the remaining cases, inducer molecules may be sucrose or sucrose-6-phosphate, or derivatives of these sugars.

A multiple sequence alignment for ScrR proteins, emphasizes the lack of conserved domains along the length of these proteins (Fig. 2.10). The only region of particularly high homology, conserved throughout the ScrR proteins, corresponds to the N-terminal helix-turn-helix DNA binding motifs (Fig. 2.10). The protein most homologous to the *C. beijerinckii* deduced ScrR is the ScrR of *S. xylosus* (31.8% identity, Table 2.2). This is particularly evident in the helix-turn-helix domains of these two proteins. In addition, the *C. beijerinckii* and *S. xylosus* ScrR proteins share several other conserved domains. These domains correspond to the regions of the LacI protein susceptible to inducer sensitive mutations, and may be involved in inducer binding (Fig. 2.10) (Kleina and Miller, 1990; Weickert and Adhya, 1992). This suggests that the *C. beijerinckii* and *S. xylosus* ScrR proteins, respond to the same ligand, which has been proposed to be sucrose-6-phosphate for the *S. xylosus* ScrR (Gering and Brückner, 1996).

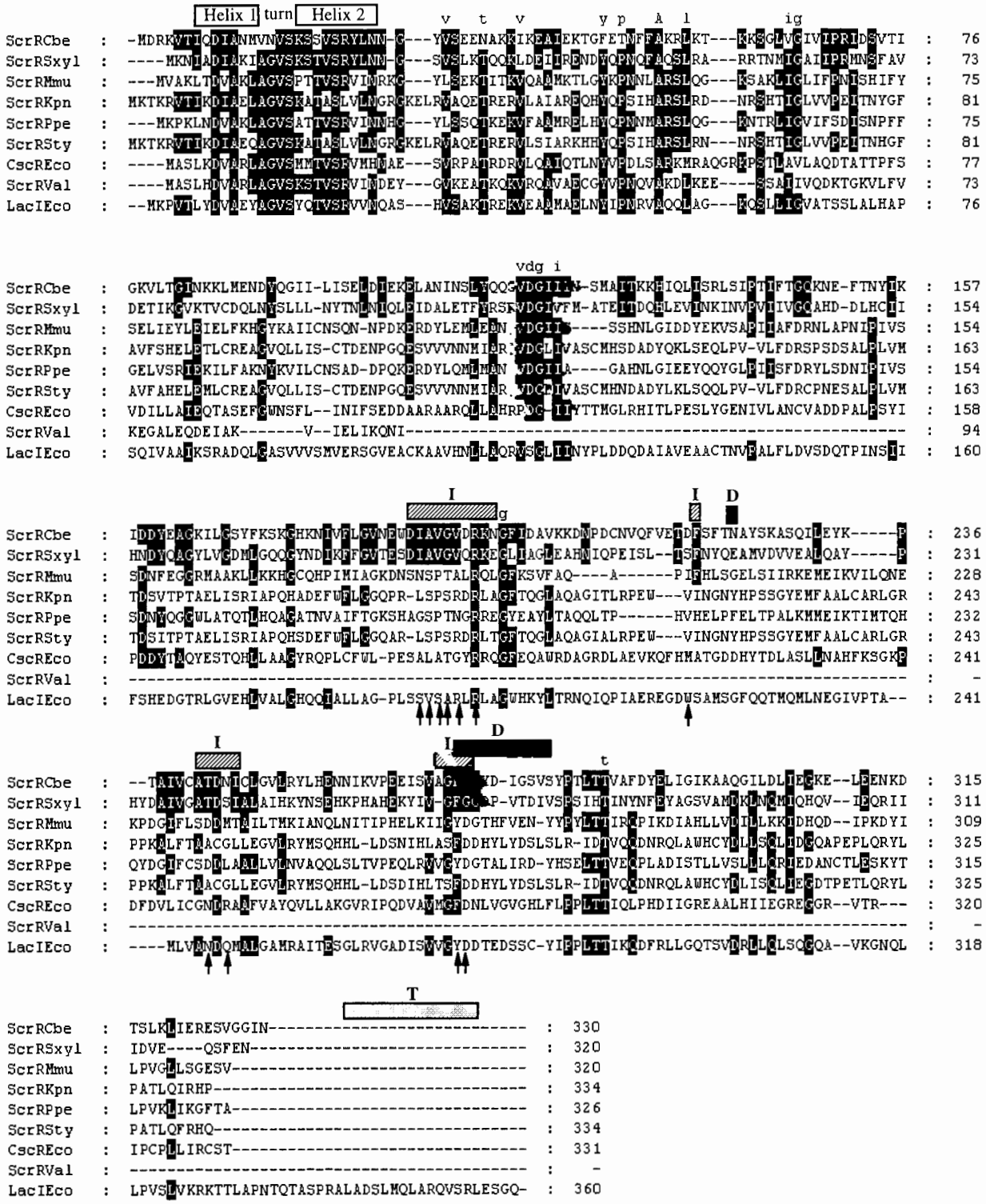

Fig. 2.10 Multiple sequence alignment of bacterial ScrR proteins. *E. coli* LacI was included to orientate the inducer binding (I; cross-hatched), dimerization (D; Dark filled box) and tetramerization (T; light shaded box) domains of LacI (Weickert and Adhya, 1992) to the ScrR proteins. Arrowheads under the LacI sequence indicate specific amino acids, at which alterations result in inducer sensitive mutations (Kleina and Miller, 1990). Shaded areas represent identical amino acids and have been manipulated to emphasize similarities between the ScrR proteins of *C. beijerinckii* and *S. xylosus*. Protein sequences included are *C. beijerinckii* ScrR (ScrRCbe), *S. xylosus* ScrR (ScrRSxyl), *S. mutans* ScrR (ScrRMmu), *K. pneumoniae* ScrR (ScrRKpn), *P. pentosaceus* ScrR (ScrRPpe), *S. typhimurium* (ScrRSty), *E. coli* CscR (CscREco), *V. alginolyticus* ScrR (ScrRVal) and the *E. coli* LacI (LacIEco). Sequence accession numbers are supplied in Table 2.2.

2.4.9. Analysis of the deduced *C. beijerinckii* ScrB and homologous proteins.

The *scrB* gene specifies a deduced protein of 485aa, with a molecular weight of 57128Da and homology to β-fructofuranosidases of the glucosyl hydrolase family 32 (Table 2.3.). Seven regions of distinct homology have been reported for sucrose hydrolases from various organisms (Gunasekaran *et al.*, 1990; Brückner *et al.*, 1993). For convenience, a multiple sequence alignment illustrating these regions and the extent to which they are conserved in the *C. beijerinckii* ScrB, is presented in Figure 2.11. Functions have been proposed for domains A and E. Specifically the conserved Asp (D) residue, in homology domain A, has been proposed to act as a catalytic nucleophile in the hydrolysis of sucrose, while the conserved Glu (E) residue, of homology domain E, has been proposed to act as an acid-base catalyst in sucrose hydrolysis (Reddy and Maley, 1996). Replacement of the Glu residue in homology region E results in a 3000 fold decrease in the activity of the yeast invertase (Reddy and Maley, 1996).

TABLE 2.3. Percent identity and similarity of the of the amino acid sequence of the deduced *C. beijerinckii* ScrB to known protein sequences.

Putative *C. beijerinckii* protein	Homologous proteins	Accession number[a]	%identity	%Similarity
ScrB	*Bacillus stearothermophilus* SurA	(gb: U34872)	36.5	62.4
	Bacillus subtilis SacA	(sp: P07819)	35.6	59.6
	Staphylococcus xylosus ScrB	(sp: Q05936)	35.2	57.3
	Lactococcus lactis ScrB	(sp: Q04937)	35.6	57.0
	Vibrio alginolyticus ScrB	(sp: P13394)	35.5	59.8
	Streptococcus mutans ScrB	(sp: P13522)	35.1	57.8
	Salmonella typhimurium ScrB	(sp: P37075)	34.2	60.0
	Klebsiella pneumoniae ScrB	(sp: P27217)	34.0	59.1
	Escherichia coli CscA	(sp: P40714)	31.3	53.0
	Streptococcus sobrinus ScrB	(PIR: S68598)	33.8	56.9
	Pediococcus pentosaceus ScrB	(sp: P43471)	30.1	50.6
	Zymomonas mobilis SacA	(sp: P22632)	29.3	53.1

[a] sp: Swiss Prot; gb: GenBank; PIR: Protein Information Resource

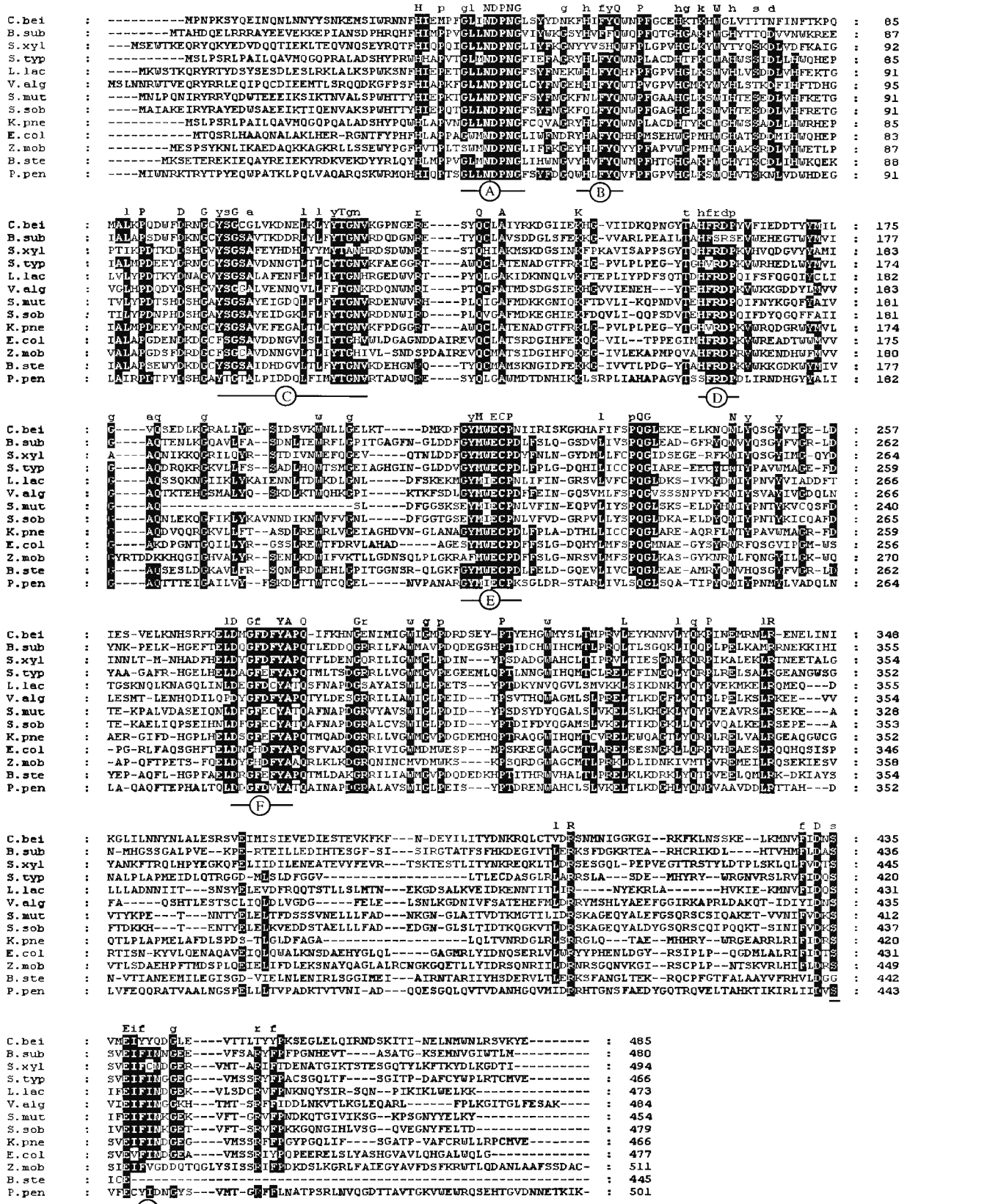

Fig. 2.11 Multiple sequence alignment of bacterial sucrose and sucrose-6-phosphate hydrolases of the glucosyl hydrolase family 32. Sequences involved are *C. beijerinckii* ScrB (C.bei), *B. subtilis* SacA (B.sub), *S. xylosus* ScrB (S.xyl), *S. typhimurium* ScrB (S.typ), *L. lactis* ScrB (L.lac), *V. alginolyticus* ScrB (V.alg), *S. mutans* ScrB (S.mut), *S. sobrinus* ScrB (S.sob), *K. pneumoniae* ScrB (K.pne), *E. coli* CscA (E.col), *Z. mobilis* SacA (Z.mob), *B. stearothermophilus* SurA (B.ste) and the *P. pentosaceus* ScrB (P.pen). Sequence accession numbers are supplied in Table 2.3. Shaded areas represent identical amino acids. Conserved regions A through G have been identified previously (Gunasekaran *et al.*, 1990; Brückner *et al.*, 1993).

2.4.10. Analysis of the deduced *C. beijerinckii* ScrK and homologous proteins.

The *scrK* gene specifies a deduced protein of 312aa, with a calculated molecular weight of 34297Da and homology to several prokaryotic as well as three eukaryotic ribokinase/*pfkB* family fructokinases (Table 2.4). A multiple sequence alignment of ribokinase/*pfkB* family fructokinases, revealed several relatively well conserved regions, some of which have been previously identified (Wu *et al.*, 1991; Fennington and Hughes, 1996; Bork *et al.*, 1993; Spychala *et al.*, 1996) (Fig. 2.12). Regions 2, 6 and 7 were identified by Bork *et al.* (1993) as being common to all members of the ribokinase/*pfkB* sugar kinase family, and are likely to facilitate functions common to its members, such as ATP binding. Regions 1, 3, 4 and 5 appear to be relatively well conserved among the fructokinases and may be responsible for activities specific to fructokinases. Indeed, Fennington and Hughes, (1996) suggested that since regions 4 and 5 were conserved primarily in the fructokinases, they may play a role in fructose binding. However, given the lack of structural information on members of the ribokinase/*pfkB* family, caution should be exercised before ascribing specific functions to the conserved domains (Spychala *et al.*, 1996).

TABLE 2.4. Percent identity and similarity of the deduced *C. beijerinckii* ScrK to known protein sequences.

Putative protein	Homologous protein	Accession number[a]	%identity	%Similarity
ScrK	*Staphylococcus aureus* Hypothetical protein	(PIR: S58482)	36.1	57.9
	Beta vulgaris Fructokinase	(gb: U37838)	36.9	60.3
	Solanum tuberosum Fructokinase	(sp: P37829)	36.0	59.1
	Lycopersicon esculentum Fructokinase	(gb: U62329)	35.0	58.6
	Vibrio alginolyticus ScrK	(sp: P22824)	29.9	53.5
	Klebsiella pneumoniae ScrK	(sp: P26420)	30.3	54.6
	Salmonella typhimurium ScrK	(sp: P26984)	29.2	51.4
	Synechocystis sp. Fructokinase	(DDBJ: D90907)	28.5	49.6
	Escherichia coli CscK	(sp :P40713)	29.2	55.1
	Rhizobium leguminosarum FrK	(sp: P42720)	26.0	51.8
	Bacillus subtilis Putative fructokinase	(sp: P42414)	22.5	49.1

[a] sp: Swiss Prot; gb: GenBank; PIR: Protein Information Resource; DDBJ: DNA Database of Japan

It was indicted in section 1.2.3. that ATP-dependent fructokinases occur in two nonhomologous families: the ribokinase/*pfkB* family (Bork *et al.*, 1993) and the ROK/hexokinase family (Bork *et al.*, 1993, Titgemeyer *et al.*, 1994). Genes encoding examples from both families have been found to be associated with sucrose utilization gene clusters. However, it is possibly worth noting that the *C. beijerinckii* ScrK represents the first complete sequence of a ribokinase/*pfkB* family fructokinase, associated with a sucrose utilization gene cluster, in a Gram-positive organism. In other Gram-positive systems, fructokinase encoding genes are either not clustered with sucrose

catabolic genes or encode ROK/hexokinase family fructokinases (Thompson *et al.*, 1991b; Sato *et al.*, 1993b; GenBank accession L32093).

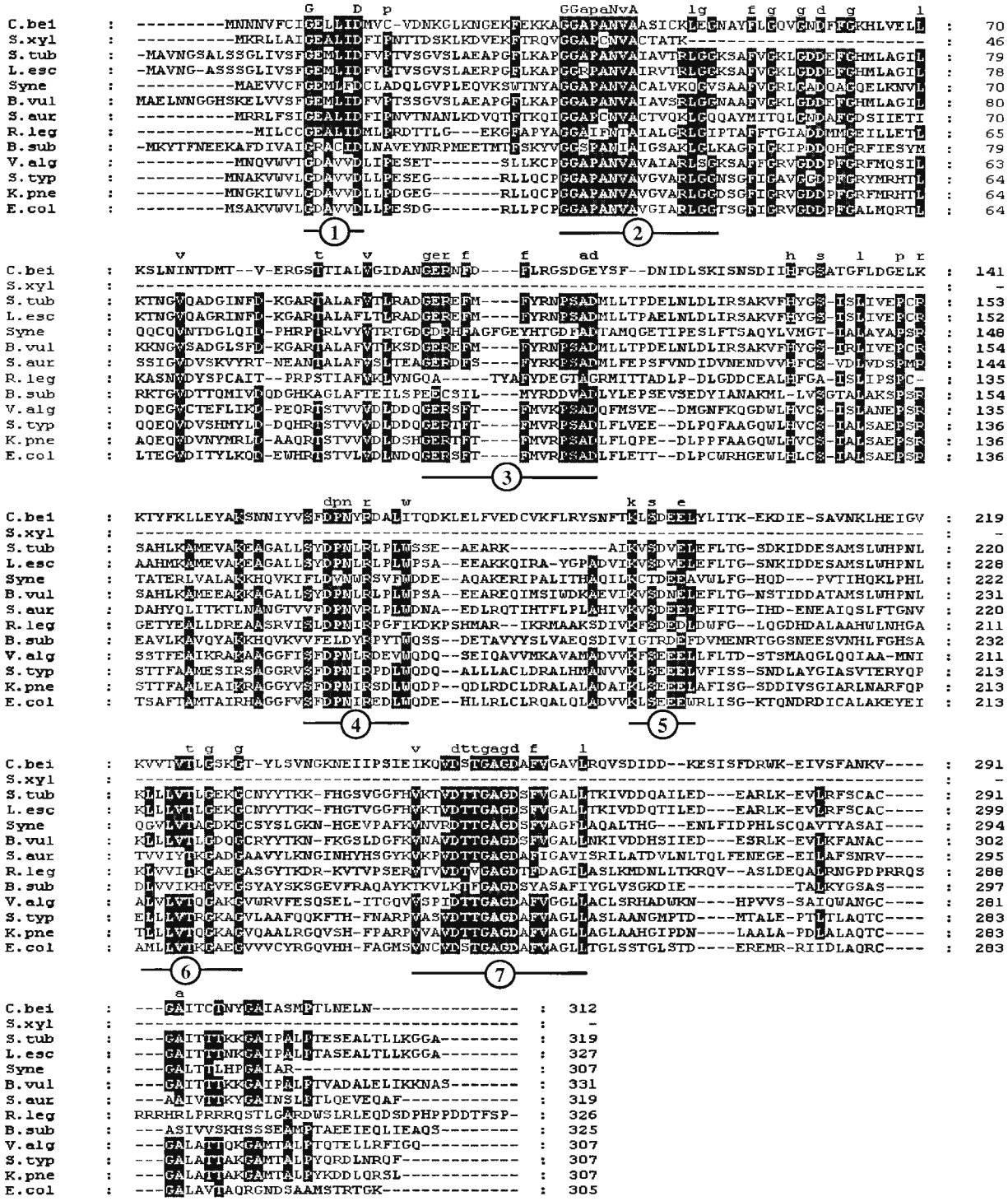

Fig. 2.12 Multiple sequence alignment of proteins homologous to the deduced *C. beijerinckii* ScrK. Sequences involved are: *C. beijerinckii* ScrK (C.bei); *S. xylosus* putative ScrK fragment (S.xyl) (the amino acid sequence was translated during this study from the reported nucleotide sequence downstream of the *scrB* gene (Brückner *et al.*, 1993)); *Solanum tuberosum* fructokinase (S.tub), *Lycopersicon esculentum* fructokinase (L.esc), *Synechocystis* sp. fructokinase (Syne), *Beta vulgaris* fructokinase (B.vul), *Staphylococcus aureus* hypothetical protein (S.aur), *Rhizobium leguminosarum* Frk (R.leg), *B. subtilis* putative fructokinase (B.sub), *V. alginolyticus* ScrK (V.alg), *S. typhimurium* ScrK (S.typ), *K. pneumoniae* ScrK (K.pne), *E. coli* CscK (E.col). Sequence accession numbers are supplied in Table 2.4. Shaded areas represent identical amino acids. Conserved regions 1 to 7 are discussed in the text.

2.5. CONCLUSIONS

Nucleotide sequence analysis indicated that the *C. beijerinckii* sucrose utilization system consisted of four genes: *scrA*, *scrR*, *scrB* and *scrK*. The genes were closely linked and present in the same orientation. Putative transcriptional terminator structures were not identified between the *scr* genes, which raised the possibility that they may be transcribed as an operon. Homology analysis suggested that the *scrA* gene encoded an EIIBC^{Scr} of the PTS, *scrR*, a regulatory protein of the GalR-LacI family, *scrB*, a sucrose hydrolase of the glucosyl hydrolase family 32 and *scrK*, a fructokinase of the ribokinase/*pfkB* sugar kinase family. The presence of these genes suggested that *C. beijerinckii* would transport sucrose via the PTS, and that this would generate intracellular sucrose-6-phosphate. Sucrose-6-phosphate would then be hydrolyzed by a sucrose-6-phosphate hydrolase, yielding glucose-6-phosphate and fructose. Fructose would then be phosphorylated by a fructokinase. The presence of *scrR* encoding a GalR-LacI-like regulatory protein, suggested that the transcription of the *C. beijerinckii scr* system would be subject to regulation by the *scrR* gene product.

Having the complete sequence of the *C. beijerinckii scr* gene cluster, allowed for a comparison with previously studied systems. The *C. beijerinckii scr* gene cluster has a molecular organization similar to the sucrose PTS regulon of *S. xylosus*. The *S. xylosus* system has been reported to consist of an independent *scrA* and a *scrRB* gene cluster (Brückner *et al.*, 1993; Gering and Brückner, 1996; Wagner *et al.*, 1993). However, it was noted during the preparation of this dissertation, that a gene encoding a ribokinase/*pfkB* family fructokinase appears to lie immediately downstream of the *S. xylosus scrB* (Chapter 1, section 1.3.2). It would follow therefore, that the *S. xylosus scr* gene cluster consists of, not two, but three genes *viz*: *scrRBK* (Fig. 1.8). This is precisely the same order as the equivalent genes of the *C. beijerinckii* system. While the *scr* gene clusters of *C. beijerinckii* and *S. xylosus* differ in that the *scrA* of *S. xylosus* is not associated with the remaining *scr* genes, the similarity in the molecular organization of the remaining genes raises the possibility of a common origin for the two gene clusters. A comparison of the *C. beijerinckii scr* gene cluster with other sucrose PTS regulons failed to identify clear similarities in molecular organization.

The basic molecular analysis of the *C. beijerinckii scr* gene cluster, allowed for speculation about how sucrose is metabolized and provided clues to the regulation of this metabolism. This provided the background against which a physiological analysis of sucrose metabolism could be conducted, and the molecular characterization extended.

CHAPTER 3

PHYSIOLOGICAL ANALYSIS OF SUCROSE UTILIZATION BY *CLOSTRIDIUM BEIJERINCKII*.

Contents

3.1. SUMMARY

A preliminary characterization was carried out on sucrose hydrolase and fructokinase activities, conferred on *E. coli* by the cloned *scrB* and *scrK* genes. The sucrose hydrolase had a pH optimum slightly below pH 6, and a temperature optimum of between 35°C and 40°C, and was unstable at temperatures higher than 30°C. Substrate specificity assays confirmed that the sucrose hydrolase was a β-fructofuranosidase able to hydrolyze sucrose and raffinose, but not the β(2-1) bonds of inulin, or the β (2-6) bonds of levans. The *scrK* gene conferred ATP/Mg^{2+} dependent fructokinase activity on *E. coli* and this activity was unstable at temperatures above 20°C.

Sucrose hydrolase and fructokinase activities expressed by *C. beijerinckii,* were also investigated. This work demonstrated that sucrose hydrolase and fructokinase activities were induced in cultures grown on sucrose, but not in cultures grown on glucose, fructose, maltose and xylose. Sucrose hydrolase and fructokinase activities were partially repressed in a mixed sucrose/glucose batch culture. Assays of carbohydrate depletion in the same culture indicated that, although both glucose and sucrose were co-metabolized, glucose appeared to be the preferred carbohydrate.

3.2. INTRODUCTION

Carbohydrate utilization systems are often regulated to exploit a changing range of available carbohydrates. This is likely to involve induction of carbohydrate utilization systems by their substrates and repression when more readily metabolized substrates are available. In this regard, bacterial sucrose utilization systems are typically induced by the presence of sucrose. This has been specifically reported for the sucrose PTS systems of *B. subtilis*, *S. xylosus*, *L. lactis* and *V. alginolyticus* and the non-PTS system of *E. coli* EC3132 (Arnaud *et al.*, 1992; Bockmann *et al.*, 1992; Gering and Brückner, 1996; Scholle *et al.*, 1987; Rauch and de Vos., 1992b). In other cases, such as the sucrose PTS systems of *K. pneumoniae* and *S. typhimurium* (plasmid pUR400), a sucrose metabolite (fructose) acts as an inducer (Jahreis and Lengeler, 1993). Transcriptional regulation in response to the presence of sucrose, is typically mediated by GalR-LacI family regulatory proteins, although exceptions do exist, such as the *B. subtilis* sucrose PTS (*sacPA* operon), which is regulated by transcriptional antitermination (Arnaud *et al.*, 1996).

In the case of *C. beijerinckii*, the *scrA*, *scrB* and *scrK* genes would specify a typical sucrose PTS pathway and the presence of *scrR* suggested that the system would be regulated in response to sucrose. The primary objective of the work presented in this chapter was to examine the regulation of the *C. beijerinckii* sucrose utilization system in response to supplied carbohydrates. In order to do this, it was first necessary to establish appropriate conditions under which sucrose hydrolase and fructokinase activities could be assayed. To this end, a preliminary characterization of these activities was carried out.

3.3. MATERIALS AND METHODS

3.3.1. Bacterial strains and media

C. beijerinckii was cultured as described in section 2.3.1., except that a Clostridial Minimal Medium (CMM), adapted from *Clostridium acetobutylicum* Minimal Medium (Long, 1984) was used (see appendix C for full details on CMM preparation). Unless otherwise stated, carbohydrates (sucrose, glucose, fructose, maltose, xylose) were supplied in CMM medium at a concentration of 0.25%(w/v). *E. coli* JM109 was cultured as described in section 2.3.1.

3.3.2. Processing of bacterial cultures for sucrose hydrolase and fructokinase assays

Cell free extracts (CFEs) were prepared in one of three buffer types. To establish the relationship between pH and sucrose hydrolase activity, CFEs were prepared in Tris-Acetate buffers with pH values ranging between pH5 and pH8.5 (100mM Tris, 100mM Acetate, adjusted to the appropriate pH with 1M HCl or 1M NaOH). For subsequent sucrose hydrolase and substrate specificity assays, CFEs were prepared in a Citrate-Phosphate (CP) buffer (pH 6.0) (17.9mM citric acid, 64.2mM Na_2HPO_4). For fructokinase assays, CFEs were prepared in a 50mM Tris-HCl buffer (pH 7.2).

The preparation of CFEs involved harvesting *E. coli* or *C. beijerinckii* cells by centrifugation and resuspending the cells in the relevant buffer (10ml buffer/gram wet weight of cells). Cell disruption was achieved by sonication of 2ml volumes for three thirty second bursts, interspersed by thirty second intervals (Soniprep 150 sonicator, MSE Scientific Instruments, Sussex, U.K.). Cell debris was removed by centrifugation in an eppendorf® 5415C microfuge (14000rpm for 10 minutes). The supernatant was then used as a CFE. Further fractionation of CFEs into soluble and membrane components, was achieved by ultracentrifugation at 175 000 *g* for 90 minutes (Beckman, TL-100), as described by Mitchell and Booth (1984). All procedures were conducted on ice or at 4°C.

In order to investigate the possibility of extracellular sucrose hydrolase activity, 50ml culture volumes were harvested, and the cells were removed by centrifugation. The supernatant was concentrated 10-fold using an Amicon 8050 Ultrafiltration cell and a PM 10 membrane. The concentrated culture medium was dialyzed overnight, at 4°C, against a CP buffer (pH 6.0) and tested for sucrose hydrolase activity.

3.3.3. Enzyme assays

Sucrose hydrolase activity was assayed using an adaptation of the method described by Blatch and Woods (1993). This involved incubating a 15µl aliquot of 0.88M sucrose with 35µl of an appropriately diluted CFE, for 15 minutes. The enzyme reaction was terminated by the addition of 150µl dinitrosalicylic acid reagent (DNS) (Miller, 1959), and boiling for five minutes. The amount of reaction product was determined spectrophotometrically at 510nm, using glucose as a standard. Sucrose hydrolase activities were expressed as micromoles of reducing sugar produced per minute per milligram of protein. The substrate specificity assays were identical to the standard saccharolytic assays, except that all substrates (sucrose, raffinose, melezitose, levan and inulin) were supplied at a concentration of 1%(w/v). Where reaction product was not detected after 15 minutes, the assays were repeated with a 60 minute incubation period. Unless otherwise specified, assays were conducted at 30°C.

Fructokinase activity was assayed using an adaptation of the method described by Sprenger and Lengeler (1988). The assay is a coupled assay, in which the production of fructose-6-phosphate is linked to the production of NADH by the presence of two auxiliary enzymes: phosphoglucose-isomerase and glucose-6-phosphate dehydrogenase. The assay was conducted in two stages. Stage 1 allowed for the production of fructose-6-phosphate, after which the reaction was stopped. Stage 2 involved the addition of the auxiliary enzymes and production of NADH. Fructose-6-phosphate (Sigma Chemical Company, St. Louis, Missouri) was used as a standard for these assays.

Stage 1: An assay mixture was prepared in a 50mM Tris-HCl buffer (pH 7.2), with a total volume of 200µl. The 200µl reaction mixture contained a 40µl aliquot of an appropriately diluted CFE and final concentrations of 10mM-$MgCl_2$, 5mM-ATP and 5mM-fructose. The preparation of this mixture involved preincubating all of the components, except fructose, for 10 minutes at 20°C. Addition of fructose initiated the reaction, which was allowed to proceed for up to 10 minutes at 20°C. The reaction was stopped by boiling for 5 minutes.

Stage 2: The stopped reaction was cooled on ice and diluted to 1ml with the addition of NAD (5mM final concentration) (Sigma Chemical Company), phosphoglucose-isomerase (final concentration 3U.ml^{-1}) (Boehringer Mannheim, GmbH) and glucose-6-phosphate dehydrogenase (final concentration 3U.ml^{-1}). (Type XXIII from *Leuconostoc mesenteriodes*,

Sigma Chemical Company). This was maintained at 37°C for 30 minutes, before the amount of NADH produced was determined spectrophotometrically at 340nm. Fructokinase activities were expressed as nanomoles or micromoles of fructose-6-phosphate produced per milligram of protein per minute.

3.3.4. Protein assays

Protein concentrations were determined using the method of Bradford (1976), with bovine serum albumin as the standard.

3.3.5. Carbohydrate utilization assays

For the determination of sucrose and glucose concentrations in the culture medium, 1ml volumes were harvested and the cells removed by centrifugation (5 minutes at 7000rpm in an eppendorf® 5415C microfuge). Glucose was assayed using the Sigma glucose assay kit (Catalog no. 510-A). Sucrose was assayed, by hydrolysis to glucose and fructose using Grade V yeast invertase (Sigma Chemical Company). Glucose was then assayed using the Sigma glucose assay kit.

3.3.6. Repetition of Assays.

Results are presented as the average and standard deviation of triplicate or duplicate assays, conducted on each sample.

3.4. RESULTS AND DISCUSSION

3.4.1. Characterization of sucrose hydrolase and fructokinase activities conferred on *E. coli* by pCBS2.

The initial aim was to establish the physical parameters under which the cloned sucrose hydrolase and fructokinase activities could be assayed. This in turn, would lead to the selection of appropriate conditions for subsequent assays. Plasmid pCBS2 conferred a sucrose-positive phenotype on *E. coli*. In addition, the *scrB* and *scrK* genes carried by pCBS2 were present in the same orientation as the pSK *lac*Z promoter, suggesting that they would be expressed (Fig. 2.1). CFEs were prepared from *E. coli* cells carrying pCBS2 and were used for a preliminary analysis of sucrose hydrolase and fructokinase activities. CFEs from *E. coli* cells carrying pSK were used as a negative control.

Before considering the work presented here, it is important to note that *scrA* encodes a putative EIIBC^{Scr} PTS protein. This suggests that intracellular sucrose would be accumulated as sucrose-6-phosphate. It would follow that sucrose-6-phosphate, and not sucrose, would be the *in vivo* substrate for the *C. beijerinckii* sucrose hydrolase. However, due to the cost and limited availability of sucrose-6-phosphate, the experiments which follow were conducted using sucrose. In cases where sucrose-6-phosphate hydrolases have been characterized, both sucrose-6-phosphate and sucrose were hydrolyzed, with the difference that K_m values for sucrose-6-phosphate were lower than those for sucrose (Lunsford and Macrina, 1986; Thompson *et al.*, 1991a; Schmid *et al.*, 1982).

Preliminary studies confirmed that pCBS2 conferred sucrose hydrolase activity on *E. coli*. In an initial study, its substrate specificity was investigated. This indicated that it catalyzed the hydrolysis of sucrose and raffinose, but not melezitose, inulin and levan (Table 3.1). These data are consistent with the presence of a β-fructofuranosidase, able to hydrolyze the fructose moieties off the low molecular weight sugars (sucrose and raffinose), but not off the β(2-1) linked fructose polymer, inulin, or the β(2-6) linked fructose polymer, levan. Sucrose hydrolase activity conferred on *E. coli* by pCBS2 was also characterized with respect to pH and temperature. The pH optimum was slightly below pH6.0 (Fig. 3.1), and this corresponds almost exactly with that reported for the intracellular invertase activity of *C. acetobutylicum* IFP 912 (Looten *et al*,. 1987). As regards thermal stability, the sucrose hydrolase was found to

be relatively stable at temperatures up to 30°C for two hours. However, exposure to temperatures higher than 30°C resulted in a substantial loss of activity (Fig. 3.1). Maximum sucrose hydrolase activity was recorded between 35 and 40°C.

Table 3.1. Substrate specificity of sucrose hydrolase activity conferred on *E. coli* JM109 by pCBS2. CFEs were prepared from stationary phase YT broths of *E. coli* JM109 carrying either pCBS2 or pSK. CFEs were exposed to the indicated substrate for 15 minutes and the release of reducing sugars was assayed. Where no activity could be detected, additional assays were conducted with an incubation period of 60 minutes.

Strain	Specific Enzyme activity[a]				
	Sucrose	Raffinose	Melezitose	Levan	Inulin
E. coli JM109 (pCBS2)	15.60 (sd: 0.31)	8.84 (sd 0.51)	ND	ND	0.24 (sd 0.08)
E. coli JM109 (pSK)	ND	0.22 (sd 0.04)	ND	ND	0.18 (sd 0.07)

[a]Specific activity expressed as µmol reducing sugar.mg protein^{-1}.min^{-1}
ND- indicates that no reducing sugars were detected after exposure to the indicated substrate for 60 minutes
sd- Standard Deviation.

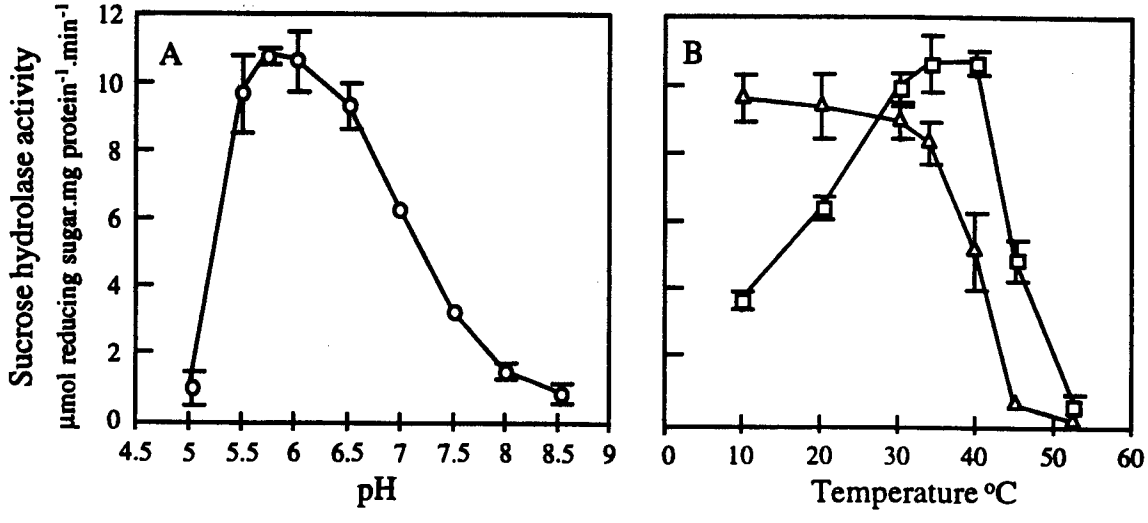

Fig. 3.1 [A] pH optimum of sucrose hydrolase activity (○) conferred on *E. coli* by pCBS2. CFEs were prepared in Tris-acetate buffers at the indicated pH values, from stationary phase YT cultures of *E. coli* JM109, carrying pCBS2. Sucrose hydrolase assays were conducted at 30°C. [B]. Temperature optimum (□) and stability (△) of sucrose hydrolase activity conferred on *E. coli* by pCBS2. CFEs were prepared in a CP buffer (pH 6) from stationary phase YT cultures of *E. coli* JM109 carrying pCBS2. The temperature optimum was determined by assaying for sucrose hydrolase activity at the indicated temperatures. Temperature stability was determined by incubating CFEs at the indicated temperatures, for two hours, followed by sucrose hydrolase assays conducted at 30°C.

The preliminary characterization of sucrose hydrolase activity was used to select conditions for subsequent assays. For these assays, CFEs were prepared in a CP buffer at pH6, and assays

were conducted at 30°C. This facilitated high levels of activity, without the risk of denaturation.

The *scrK* gene product was proposed, on the basis of homology, to be a fructokinase of the ribokinase/*pfkB* sugar kinase family. Fructokinases of this family have been demonstrated to use ATP as a phosphoryl donor, and activity has been found to be dependent on the presence of divalent cations, with Mg^{2+} being commonly supplied in assays (Fennington and Hughes, 1996; Sprenger and Lengeler, 1988). The aim of the work presented here, was to confirm that pCBS2 conferred ATP/Mg^{2+} dependent fructokinase activity on *E. coli* and to characterize this activity with respect to temperature.

As expected pCBS2 conferred fructokinase activity on *E. coli* (Fig. 3.2). No fructokinase activity could be detected when ATP or $MgCl_2$ were omitted from the assay mixture, or for CFEs prepared from *E. coli* cells carrying pSK. In similar experiments, a low endogenous fructokinase activity has been reported for *E. coli*. Where detected, this endogenous activity did not interfere with the measurement of fructokinase activities conferred by cloned genes (Fennington and Hughes, 1996; Sato *et al.*, 1993b). In the present study, no such endogenous fructokinase activity was observed for *E. coli*. It is possible that the dilution factor required to bring the *scrK* encoded fructokinase activities into a measurable range, was so great, that when equivalent dilutions were made with CFEs from *E. coli* cultures carrying pSK, the low endogenous fructokinase activity was diluted below the level of detection.

Fructokinase activity was characterized with respect to temperature but not pH, because changes in pH would have affected the activity of the auxiliary enzymes. Maximum fructokinase activities were recorded between 20 and 35°C, and did not vary greatly within that range (Fig. 3.3). In terms of temperature stability, fructokinase activity was found to be unstable, particularly when exposed to temperatures higher than 20°C for two hours. Subsequent fructokinase assays were conducted at 20°C ,which facilitated relatively high levels of activity, while minimizing enzyme denaturation.

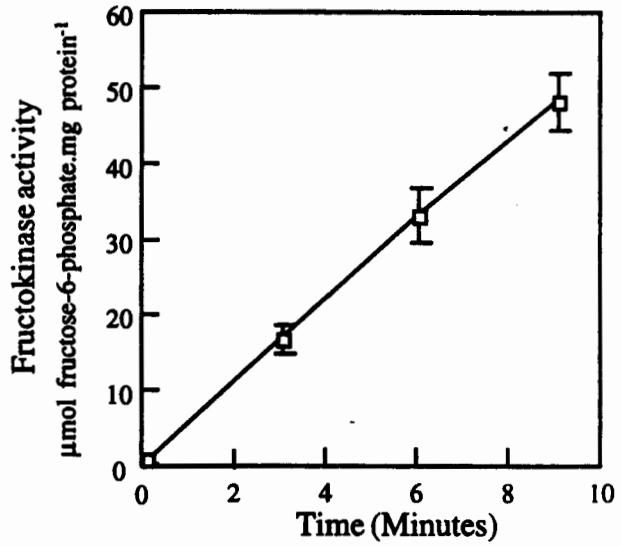

Fig. 3.2 Accumulation of fructose-6-phosphate, as a result of fructokinase activity conferred on *E. coli* by pCBS2. CFEs were prepared from stationary phase YT cultures of *E. coli* JM109 carrying pCBS2. Fructokinase assays were also conducted in the absence of either $MgCl_2$ or ATP and in both cases, no detectable activity was observed. Fructokinase activity was also not observed for CFEs from stationary phase YT cultures of *E. coli* JM109 carrying pSK.

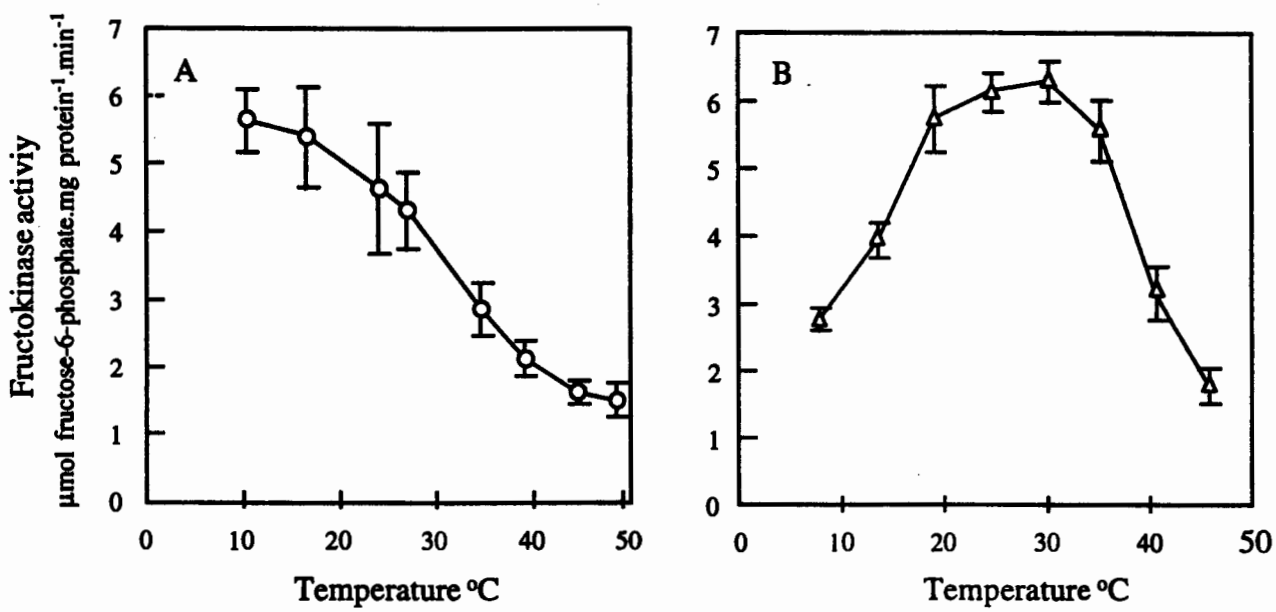

Fig. 3.3 [A] Temperature stability (O) and [B] optimum (Δ) of fructokinase activity conferred on *E. coli* by pCBS2. CFEs were prepared from stationary phase YT cultures of *E. coli* JM109 carrying pCBS2. Temperature stability was determined by incubating CFEs at the indicated temperatures, for two hours. The extracts were then assayed for fructokinase activity at 20°C. The temperature optimum was determined by assaying for fructokinase activity at the indicated temperatures.

3.4.2. Expression of sucrose hydrolase and fructokinase activities in *C. beijerinckii*.

Having established appropriate conditions under which the *scrB* and *scrK* encoded sucrose hydrolase and fructokinase activities could be assayed, the equivalent activities were investigated in *C. beijerinckii*. The primary objective was to examine the regulation of these activities in response to supplied carbohydrates. In a preliminary experiment, however, the localization of sucrose hydrolase activity expressed by *C. beijerinckii* was investigated.

C. beijerinckii cultures were grown in the presence of 0.25% (w/v) sucrose to OD_{600} = 0.4. Both cells and the extracellular medium were processed for sucrose hydrolase assays. No activity could be detected in the extracellular medium. The CFE was separated into membrane and cytosol fractions by ultracentrifugation. Each component was assayed for sucrose hydrolase activity. The activities were expressed as a percentage of that present in the CFE, prior to ultracentrifugation. Only a small proportion (<5%) of the original sucrose hydrolase activity was associated with the membrane fraction, while 90% was identified in the cytosolic component. These data are consistent with the presence of an intracellular sucrose hydrolase in *C. beijerinckii*.

In order to establish whether sucrose hydrolase and fructokinase activities were regulated, *C. beijerinckii* cultures were grown to OD_{600} = 0.4, in the presence of sucrose, glucose, fructose, maltose and xylose. CFEs were then prepared and assayed for sucrose hydrolase and fructokinase activities. Both activities were induced in the presence of sucrose, but were low or absent where cells were grown on the remaining carbohydrates (Fig. 3.4). PEP-dependent sucrose phosphorylation activity has also been found to be induced in sucrose-grown *C. beijerinckii* cells, but not in xylose- or fructose-grown cells (personal communication from Tangney and Mitchell, Department of Biological Sciences, Heriot-Watt University, Edinburgh). Clearly, enzyme activities associated with sucrose metabolism are regulated and it is likely that sucrose or a sucrose metabolite acts as an inducer of these activities. These results do not, however, indicate the extent to which other carbohydrates may be involved in the repression of sucrose hydrolase and fructokinase activities. This prompted an investigation of glucose-mediated repression of sucrose metabolism.

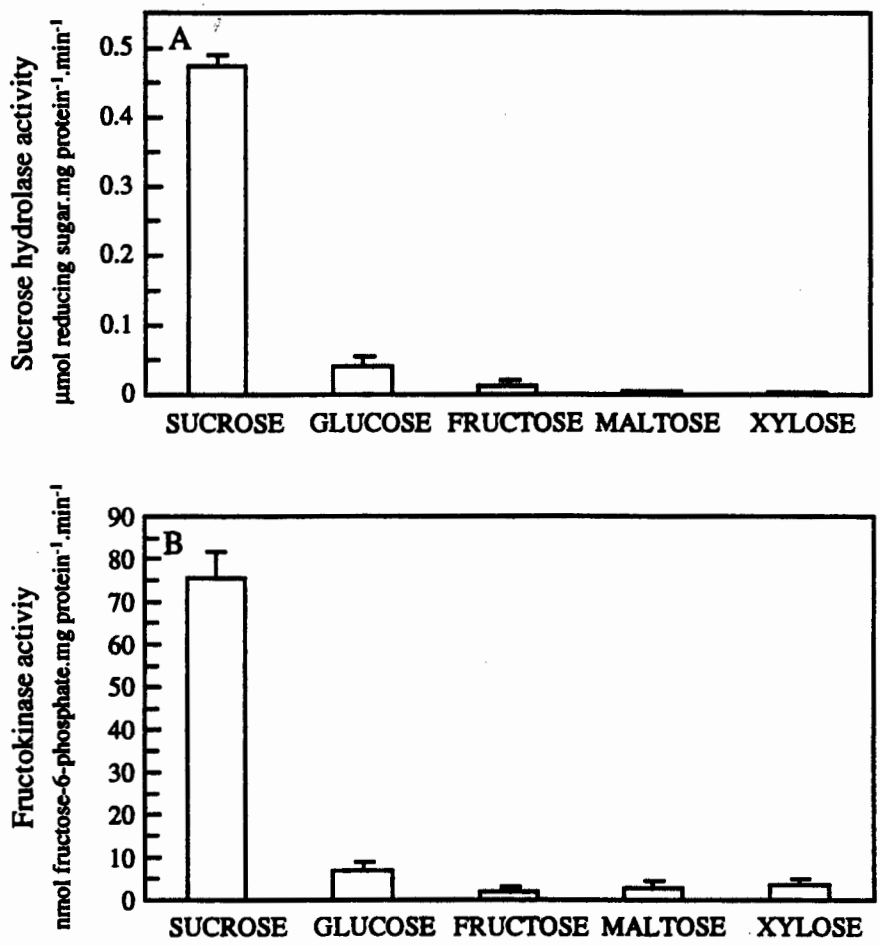

Fig. 3.4 Relationship between carbohydrate carbon source and [A] sucrose hydrolase activity [B] fructokinase activity. *C. beijerinckii* cultures were grown in CMM broth containing 0.25% of the indicated carbohydrate. CFEs were prepared from cultures during logarithmic growth (OD_{600} = 0.4).

3.4.3. Glucose-mediated repression of sucrose utilization in *C. beijerinckii*

The possibility that glucose exerted a negative effect on the *C. beijerinckii* sucrose utilization system, was investigated. The experimental design will be considered briefly. Precultures were grown to an OD_{600} of 0.15 in CMM broth, containing 6.5mM sucrose. Precultures were then subcultured into CMM broth, containing either sucrose or glucose or both glucose and sucrose. Glucose was supplied at a concentration of 13mM and sucrose at 6.5mM. This ensured that the total amount of carbohydrate supplied as the monosaccharide glucose, approximated the total amount supplied as the disaccharide sucrose. Carbohydrate concentrations were referred to as a percentage of that originally supplied, where 100% was equivalent to 13mM glucose or 6.5mM sucrose. Since conditions in batch culture change during growth, the parameters being monitored *viz*: culture medium carbohydrate concentration, sucrose hydrolase and fructokinase activities, were assayed from early to late growth stages (Fig. 3.5).

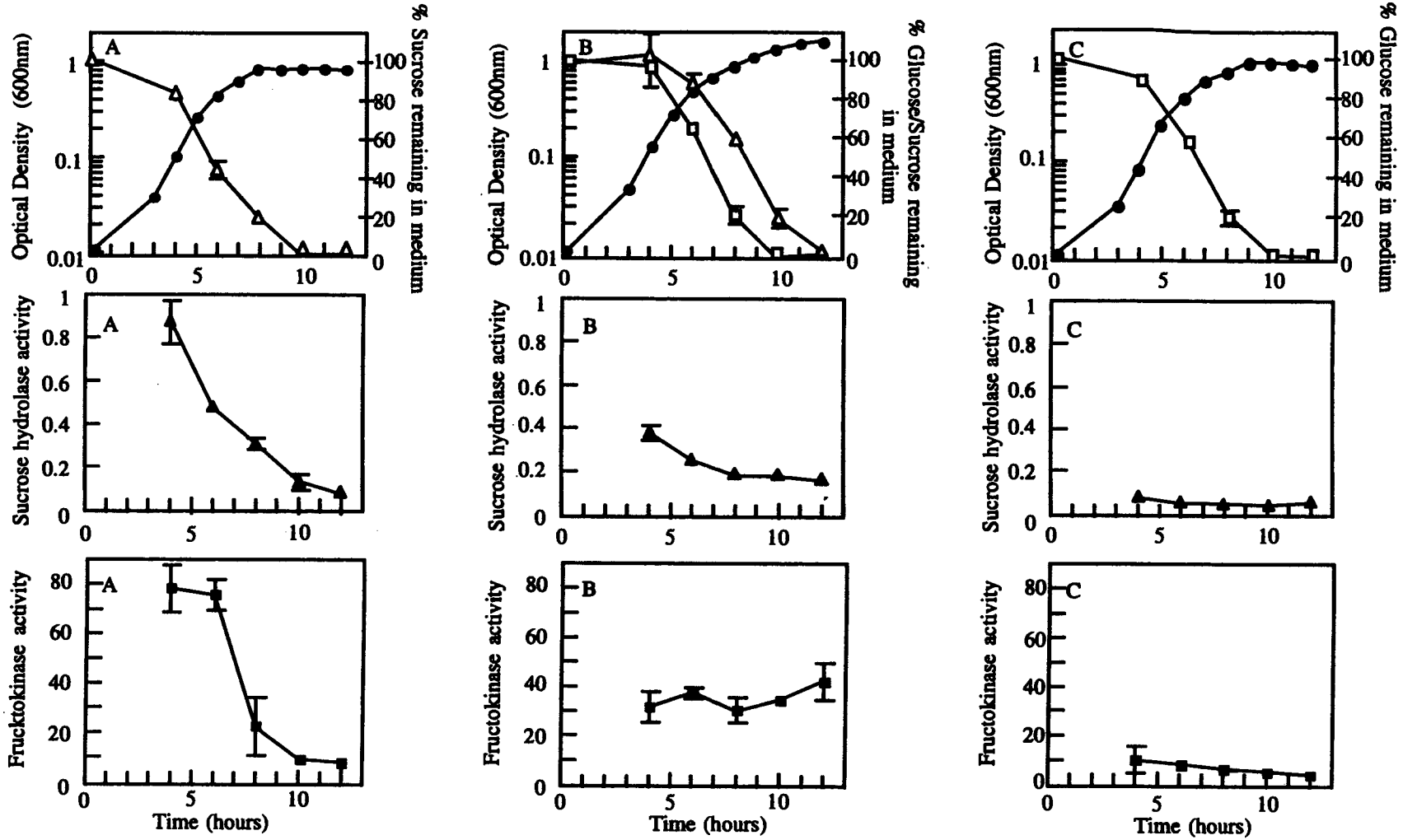

Fig. 3.5 Relationship between growth, carbohydrate carbon source, sucrose hydrolase and fructokinase activities expressed by *C. beijerinckii*. Precultures were grown to an OD_{600nm} = 0.15 in CMM medium containing 6.5mM sucrose, before being diluted 1/200 into CMM broth containing [A] 6.5mM sucrose [B] 13mM glucose + 6.5 mM sucrose and [C] 13mM glucose. (●) Optical density 600nm, (▲) Sucrose concentration as a percentage of the initial concentration (100% = 6.5mM sucrose) (□) Glucose concentration as a percentage of the initial concentration (100% = 13mM glucose) (▲) sucrose hydrolase activity µmol reducing sugar.mg protein^{-1}.min^{-1}, ( ) Fructokinase activity nmol fructose-6-phosphate.mg protein^{-1}.min^{-1}. Error bars have been manually removed where they were obscured by the respective data point.

Growth in the mixed glucose/sucrose culture increased to a maximum OD_{600} of 1.4 units, while growth on either carbohydrate alone reached a maximum OD_{600} of 0.9 units (Fig. 3.5). Assays of carbohydrate depletion for the mixed glucose/sucrose culture indicated that although glucose and sucrose were utilized simultaneously, glucose was the preferred carbohydrate. This is particularly clear if the levels of carbohydrate remaining in the medium are compared after eight hours of growth (Fig. 3.5). Where glucose and sucrose were present alone, only 20% of each carbohydrate remained after eight hours. In the mixed glucose/sucrose culture, however, approximately 20% of the original glucose remained, as opposed to 60% of the original sucrose.

Similar experiments have been conducted previously (Mitchell *et al.*, 1995). The study indicated that preculture conditions influenced glucose-mediated repression of sucrose utilization. If precultures were prepared with glucose as the sole carbohydrate, subsequent repression of sucrose utilization in a mixed sucrose/glucose medium was almost complete. If precultures were prepared with sucrose as the sole carbohydrate, then sucrose and glucose were initially metabolized at the same rate, however, as growth proceeded, sucrose utilization was retarded and resumed only after glucose depletion. Together, the results reported by Mitchell *et al.* (1995) and those presented in the present study, demonstrate that glucose exerts a negative effect on sucrose utilization, however, the extent to which this occurs varies under different conditions. Where the sucrose utilization system is initially uninduced, glucose-mediated repression appears to be absolute (Mitchell *et al.*, 1995). Where the sucrose utilization system is induced, glucose-mediated repression is manifest either as a delayed repression of sucrose utilization, as reported by Mitchell *et al.* (1995), or as a decrease in the rate of sucrose utilization, as reported here. Reasons for the subtle differences between the results observed by Mitchell *et al.* (1995) and those observed here, are not immediately obvious. It is possible that the differences are the result of a complex set of catabolite repression mechanisms, which may be sensitive to small changes in the growth environment.

An important part of assessing the effect of glucose-mediated repression on sucrose utilization, was to determine the effect of glucose on sucrose hydrolase and fructokinase activities. In cultures grown on sucrose alone, sucrose hydrolase and fructokinase activities were initially relatively high but dropped rapidly during growth (Fig. 3.5). This may reflect the fact that the concentration of extracellular inducer (sucrose) dropped rapidly during growth. Where sucrose-grown precultures were inoculated into media containing glucose as the sole

carbohydrate, only low levels of sucrose hydrolase and fructokinase activities were observed. Where sucrose-grown precultures were inoculated into media containing both glucose and sucrose, initial sucrose hydrolase and fructokinase activities were less than half the values observed for cultures grown on sucrose alone (Fig. 3.5). This correlated with the fact that sucrose was depleted more slowly in the mixed glucose/sucrose culture and indicates that a mechanism facilitating glucose-mediated repression of these activities exists.

3.5. CONCLUSIONS

A preliminary characterization of *scrB* and *scrK* encoded sucrose hydrolase and fructokinase activities, was carried out. This was conducted on extracts from *E. coli* cells carrying pCBS2, since in *C. beijerinckii*, several enzymes may have contributed to the same activity. The findings of this work essentially bear out what would be expected from *scrB* and *scrK* sequence analysis. The *scrB* gene conferred sucrose hydrolase, but not inulinase or levanase activity on *E. coli*. Sucrose hydrolase activity had a pH optimum slightly below pH 6 and a temperature optimum between 35°C and 40°C. The *scrK* gene conferred Mg^{2+}/ATP dependent fructokinase activity on *E. coli*. This preliminary characterization was used to select conditions for subsequent assays and provides a starting point for further characterization of these enzymes.

The possibility that *C. beijerinckii* expresses extracellular sucrose hydrolases was investigated. In cultures grown on 0.25% (w/v) sucrose, almost all sucrose hydrolase activity was intracellular, and no activity could be detected in the extracellular milieu. While this suggests that only intracellular sucrose hydrolases exist, it is possible that extracellular enzymes capable of sucrose hydrolysis are induced by substrates other than sucrose. Indeed, extracellular β-fructofuranosidases (inulinases) have been characterized for *C. acetobutylicum* strains IFP9 and ABKn8 (Efstathiou *et al.*, 1986; Looten *et al.*, 1987). In both cases the inulinases were induced by the presence of inulin and not sucrose, but were active against both substrates.

Of the sucrose utilization systems discussed in the general introduction (section 1.3), almost all are regulated to some degree. Typically the systems are induced by sucrose, however, sucrose derivatives such as fructose may also act as inducers (Schmid *et al.*, 1982; Sprenger and Lengeler, 1988). In the case of *C. beijerinckii*, sucrose hydrolase and fructokinase

activities were induced in sucrose-grown cells and were low or absent in cells grown on glucose, fructose, maltose and xylose.

The degree to which glucose mediates repression of sucrose utilization, was investigated in mixed sucrose/glucose batch cultures. Although glucose and sucrose were co-metabolized, sucrose was used more slowly than when present as the sole carbon source, and both sucrose hydrolase and fructokinase activities were partially repressed. The mechanisms facilitating catabolite repression in *C. beijerinckii*, have not been elucidated, and it is entirely possible that the observed repression was exercised at several levels. One possible level could be inducer exclusion as a result of competition for common PTS intermediates. The *scrA* gene was proposed to encode an $EIIBC^{Scr}$ protein. It is possible the $EIIB^{Scr}$ domain is phosphorylated by a shared EIIA domain. If the shared EIIA is an $EIIA^{Glc}$, as has been proposed for other bacterial $EIIBC^{Scr}$ proteins (Blatch, 1990; Lengeler *et al.*, 1982, Sprenger and Lengeler, 1988; Sutrina *et al.*, 1990), competition for phosphorylated $EIIA^{Glc}$ may result in preferential glucose uptake at the expense of sucrose uptake. Precisely such a mechanism has been proposed for the sucrose utilization system of *B. subtilis* (Sutrina *et al.*, 1990). Glucose-mediated repression may act at other levels. An HPr(ser)kinase/HPr/CcpA mediated catabolite repression pathway has been characterized for several low-GC Gram-positive bacteria (Saier *et al.*, 1996). While genes encoding CcpA-like proteins have been cloned from *C. acetobutylicum* NCP 262 (Davison *et al.*, 1995) and more recently from *C. beijerinckii* (Rafudeen, unpublished results, University of Cape Town, Microbiology Dept.) the existence of a CcpA-mediated catabolite repression pathway needs to be established before its role in sucrose metabolism can be considered.

CHAPTER 4

STUDIES ON TRANSCRIPTION OF THE *scrARBK* OPERON AND TARGETED DISRUPTION OF *scrR* AND *scrB* IN *CLOSTRIDIUM BEIJERINCKII*

Contents

4.1. SUMMARY

Northern blot analysis indicated that the *scrARBK* genes were transcribed as an operon and that transcription was induced by the presence of sucrose. The operon was disrupted in *C. beijerinckii,* by targeted integration of plasmids into the *scrR* and *scrB* genes. Mutant strains carrying disrupted *scrR* and *scrB* genes were unable to grow where sucrose was supplied as the sole carbohydrate. This suggested that the *scrARBK* operon encoded the only sucrose utilization system in *C. beijerinckii.* In the strain carrying a disrupted *scrR,* expression off the *scrA* promoter occurred under non-inducing conditions. This indicated that the *scrR* gene product is a negative regulator of *scrARBK* transcription, and is therefore itself negatively autoregulated.

4.2. INTRODUCTION

In almost all well characterized bacterial sucrose utilization systems, GalR-LacI-like regulatory proteins play a role in transcriptional regulation. This has been clearly demonstrated in the cases of *S. typhimurium*, *K. pneumoniae*, *V. alginolyticus* and *S. xylosus* (Blatch and Woods, 1991; Gering and Brückner, 1996; Jahreis and Lengeler, 1993). In the cases of *L. lactis*, *S. mutans* and *P. pentosaceus*, *scrR* genes are clustered with sucrose catabolic genes, suggesting an involvement in their regulation (Rauch and de Vos. 1992b; Sato *et al.*, 1989) (GenBank accession numbers U46902 and L32093).

Where regulation by ScrR proteins has been studied, negative regulatory mechanisms have invariably been proposed. These typically involve repression of transcription, in the absence of sucrose, as the respective ScrR proteins bind palindromic operator sites within or immediately downstream of the regulated promoters. The presence of sucrose, or sucrose derivatives, prevents repressor operator interaction and transcription proceeds.

Evidence has already been presented indicating induction of sucrose hydrolase and fructokinase activities in *C. beijerinckii,* in response to sucrose. The overall objective of the work presented here, was to examine how this regulation was facilitated. The *C. beijerinckii scrR* specified a protein with distinct homology to GalR-LacI-like regulatory proteins. In order to gain insight into the role that the *scrR* gene product might play, it was necessary firstly, to identify whether transcriptional regulation occurred, secondly, to identify transcriptional units within the *scr* gene cluster and thirdly, to establish whether the *scrR* gene product was responsible for the observed regulation. These three aspects were investigated by Northern blot analysis, using wild type *C. beijerinckii* strains, as well as strains in which the *scrR* and *scrB* genes had been disrupted. *C. beijerinckii* strains carrying disrupted *scrR* and *scrB* genes, were also used to establish the importance of the *scr* genes to sucrose metabolism in *C. beijerinckii.*

4.3. MATERIALS AND METHODS

4.3.1. Bacterial strains media and plasmids.

C. beijerinckii was cultured essentially as described in sections 2.3.1 and 3.3.1. except that Reinforced Clostridial Medium (RCM) (Difco Laboratories, Detroit, USA) was also used. Where appropriate erythromycin was added at a final concentration of 10 μg/ml.

E. coli CA474, a *λcI857* lysogen of *E. coli* HB101 (Wilkinson and Young, 1994) was used as a donor in transconjugation experiments and was cultured on Brain Heart Infusion (BHI) medium (Difco Laboratories, Detroit, USA). *E. coli* JM109 was used for all cloning purposes and was cultured as described in section 2.3.1. Where appropriate, ampicillin and kanamycin were used at a final concentration of 100μg/ml. Additional strain details are provided in Appendix A.

Plasmids R702, pCTC1 and pMTL30 were used for targeted integration experiments and have been described previously (Hedges and Jacob, 1974; Minton *et al.*, 1990; Minton *et al.*, 1993; Wilkinson and Young, 1994). Briefly R702 is an autonomous IncP plasmid able to facilitate plasmid mobilization *in trans*, and was used as a helper plasmid in transconjugation experiments. Plasmid R702 confers kanamycin resistance on *E. coli*. Plasmids pCTC1 and pMTL30 carry *oriT* regions from the IncP plasmid RK2, Em^R and Ap^R markers, and ColE1 replicons, facilitating maintenance in Gram-negative strains. These two plasmids differ in that pCTC1 also carries a pAMβ1 replicon, facilitating maintenance in Gram-positive strains. Additional plasmid details are provided in Appendix B.

Plasmids pCBMU1 and pCBMU3 are pMTL30 derivatives constructed for the disruption of the *scrB* and *scrR* genes respectively. Plasmid pCBMU1 carries a 435bp fragment internal to *scrB*, while pCBMU3 carries a 456bp fragment internal to *scrR*. Details of plasmid construction are presented in Figure 4.1.

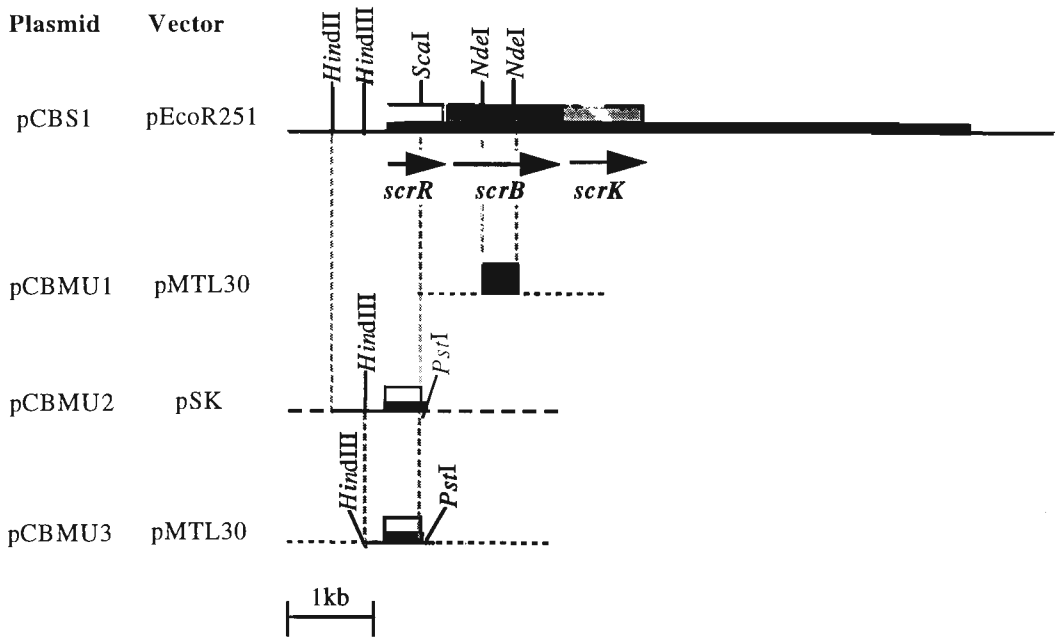

Fig. 4.1 Construction of plasmids for targeted disruption of the *scrR* and *scrB* genes. Thick horizontal lines correspond to *C. beijerinckii* DNA. Thin horizontal lines represent p*Eco*R251. Dotted horizontal lines represent pMTL30. Dashed horizontal lines represent pSK. pCBMU1 was constructed for the disruption of *scrB*. A 435bp *Nde*I fragment internal to *scrB* was cloned from pCBS1 into the *Stu*I site of the pMTL30 MCS. The fragment encodes amino acids 162 to 330 of the 486aa deduced ScrB. pCBMU2 was initially constructed for the purposes of sequencing, by cloning the *Hin*dII-*Sca*I fragment from pCBS1 into the *Eco*RV site of the pSK MCS. pCBMU3 was then constructed by cloning the *Hin*dIII-*Pst*I fragment from pCBMU2 into the equivalent sites of the pMTL30 MCS. pCBMU3 carries a 456bp fragment of *C. beijerinckii* DNA encoding amino acids 98 to 249 of the 331aa deduced ScrR. Only those restriction endonuclease sites used in the construction of pCBMU1, pCBMU2 and pCBMU3 are shown.

4.3.2. Targeted integration of plasmids into the *C. beijerinckii* genome

The targeted integration system developed by Wilkinson and Young (1994), was reviewed in section 1.1.4. The technique used pMTL30 derivatives, carrying short regions of *C. beijerinckii* 'target' DNA, which were mobilized from *E. coli* CA474 to *C. beijerinckii*. Since pMTL30 lacked a Gram-positive replicon, maintenance in *C. beijerinckii* was dependent on a homologous recombination event between the 'target' DNA and its equivalent on the *C. beijerinckii* genome. The Em^R marker of pMTL30 facilitated selection of *C. beijerinckii* transconjugants. Incubation at 42°C counter-selected the donor strain by induction of a λ prophage.

E. coli donor strains, containing R702 and the respective pMTL30 derivative, were grown to stationary phase in BHI broth, supplemented with ampicillin (100μg/ml) and kanamycin (100μg/ml). Recipient *C. beijerinckii* cultures were grown in CBM broth, to an OD_{600} of between 0.6 and 1.2. Donor and recipient strains were mixed by the addition of 0.2ml donor

culture to 2ml recipient culture. The cells were recovered by centrifugation and plated onto RCM plates. The plates were incubated anaerobically, overnight, at 30°C. Accumulated bacteria were then resuspended in 700µl volumes of CBM and plated onto CBM agar plates, containing 10µg/ml erythromycin. Overnight anaerobic incubation at 42°C facilitated counter-selection of the donor strain, while the presence of erythromycin facilitated the selection of *C. beijerinckii* transconjugants. Viable counts of recipient bacteria were obtained on CBM plates, in the absence of selective antibiotics. Frequencies of plasmid establishment were recorded as the number of transconjugants per recipient.

4.3.3. Preparation of total RNA from *C. beijerinckii*

C. beijerinckii cells were harvested from 30ml culture volumes by centrifugation. Cell pellets were then resuspended in 500µl aliquots of a 0.3M sucrose/0.01M sodium acetate solution (pH 4.5). In each case, a 500µl aliquot of a 2% SDS/0.01M sodium acetate solution (pH4.5) was added and the mixture was maintained at 65°C for 1.5 minutes. Water-saturated phenol (1ml per sample) was then added and the samples were mixed and maintained at 65°C for a further 3 minutes. The samples were then cooled at -78°C for 15 seconds. Aqueous phase recovery was achieved by centrifugation (10000 rpm for 5min in an eppendorf® 5415C microfuge). Two further hot phenol extractions were conducted. This was followed by ethanol precipitation of DNA and RNA and resuspension in a 200µl volume of a 20mM sodium acetate/10mM $MgCl_2$/10mM NaCl solution (pH4.5). DNA was removed using DNAse I (30U/per sample) (Boehringer Mannheim). The remaining RNA was ethanol precipitated and resuspended in distilled water. RNA samples were stored at -70°C.

4.3.4. Northern blot analysis

Denaturing agarose gel electrophoresis of RNA was conducted as described by Sambrook *et al.*, (1989). RNA was then capillary transferred for nine hours onto positively charged nylon membranes (Boehringer Mannheim), using a 0.05M NaOH solution as a transfer medium. DIG-labelled RNA probes, specific for the transcripts of the *scrA*, *scrR*, *scrB* and *scrK* genes, were prepared using the Boehringer Mannheim RNA labeling kit (Catalogue no: 1175 025). Hybridization and Detection of DIG-labeled RNA probes was conducted according to standard techniques specified by Boehringer Mannheim.

4.4. RESULTS AND DISCUSSION

4.4.1. Northern blot analysis of *scrA*, *scrR*, *scrB* and *scrK* expression.

The fact that the *scr* genes were clustered and present in the same orientation on the *C. beijerinckii* genome, raised the possibility that they may be transcribed as an operon. Furthermore the presence of *scrR*, specifying a GalR-LacI-like regulatory protein, suggested that *scr* gene expression would be transcriptionally regulated. Experimental evidence was required in order to assess these proposals. Northern blot analysis was selected, since it would unequivocally identify transcriptional coupling, and would also provide evidence for transcriptional regulation. Although primer extension analysis was considered, it was reserved for future work, since failure to identify a transcription initiation site would not unequivocally indicate the absence of such a site *in vivo*. In addition, continued difficulties were being experienced in the cloning of the 5'-region of the *scr* gene cluster (section 2.4.5). The cloning of this region was required before a complete primer extension analysis could be conducted.

Probes specific for the *scrA*, *scrR*, *scrB*, and *scrK* transcripts were hybridized against RNA extracted from *C. beijerinckii* cells grown on sucrose, glucose, fructose, xylose and maltose. A predominant 5.0kb transcript was identified by each of the *scr* probes in RNA isolated from sucrose-grown cells (Fig. 4.2). An equivalent transcript was not identified in RNA extracted from cells grown on glucose, fructose, xylose or maltose. It is important to note that not only do all *scr* probes identify a transcript of the same size, but the size of the transcript (5.0kb) correlates with the size of the four genes involved (4946bp separate the putative initiation and termination codons of the *scrA* and *scrK* genes respectively). These data confirm that the *scr* genes are expressed as an operon and that transcription is induced by sucrose. The findings also correspond with observations made at the physiological level, where sucrose hydrolase and fructokinase activities were induced by sucrose. PEP-dependent sucrose phosphorylation has also been found to be induced by sucrose (personal communication from Tangney and Mitchell, Department of Biological Sciences, Heriot-Watt University, Edinburgh). The fact that the *scr* genes are transcribed as an operon is important in the context of regulation. It indicates that the entire operon is transcribed off the *scrA* promoter(s), which must therefore be the site at which transcriptional regulation occurs.

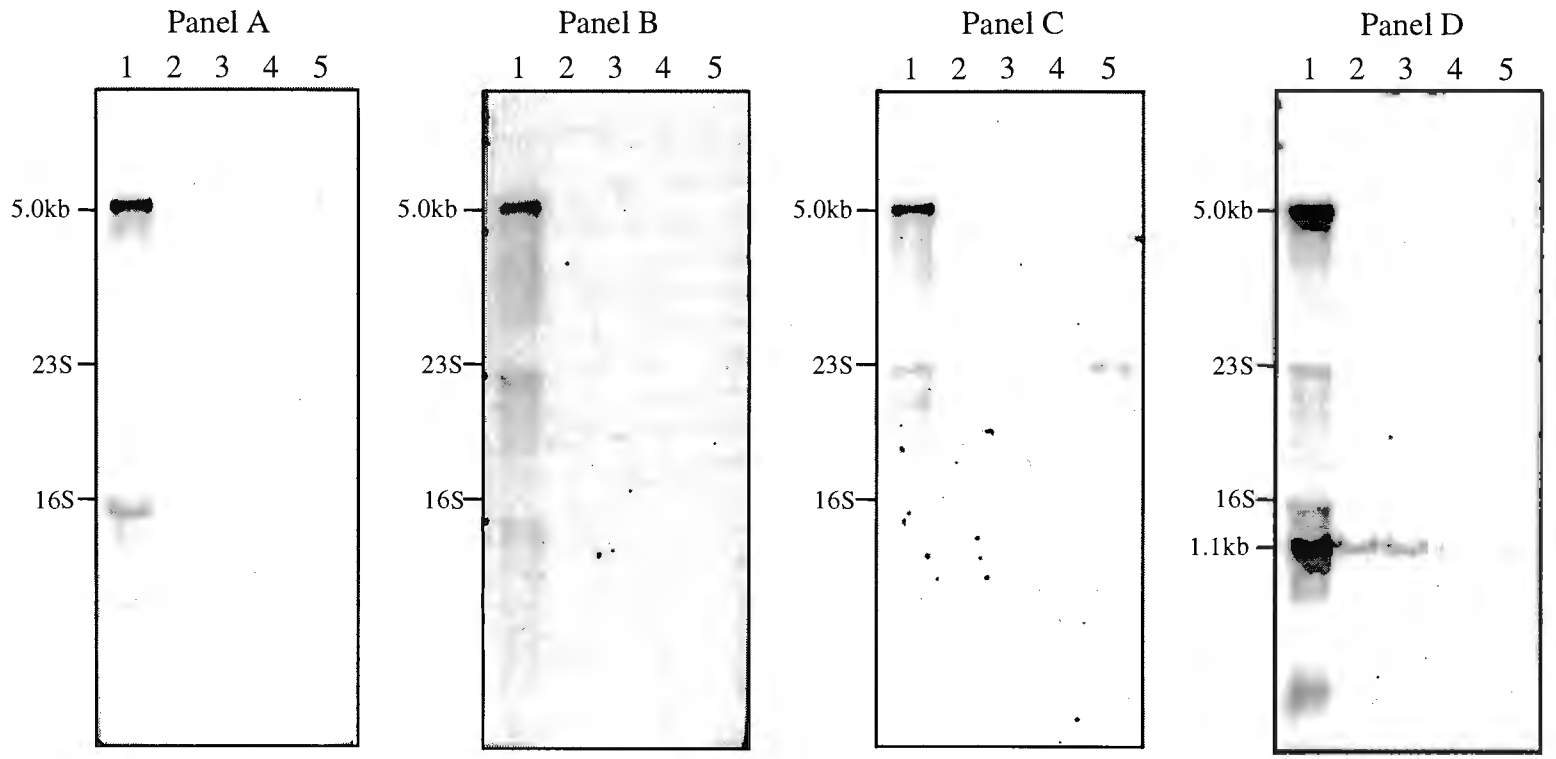

Fig. 4.2 Northern blot analysis of the *C. beijerinckii scrARBK* operon. Panels A, B, C, and D represent Northern blots conducted with probes specific for the *scrA*, *scrR*, *scrB* and *scrK* transcripts respectively. The five lanes in each panel represent RNA extracted from wild type *C. beijerinckii* cells, grown to logarithmic phase in CMM medium, containing 0.25% w/v of the respective carbohydrate: (1) Sucrose (2) Glucose (3) Fructose (4) Xylose (5) Maltose.

In addition to the predominant 5.0kb transcript, other distinct signals are also present. In lanes representing RNA isolated from sucrose-grown cells, signals corresponding in size to the 16S and 23S ribosomal RNA species are present. It is likely that these signals represent *scr* operon mRNA co-migrating with high concentrations of ribosomal RNA. This explanation does not, however, account for a 1.1 kb transcript identified with the *scrK* probe (Fig. 4.2, Panel D). This transcript is smaller than the 16S RNA species and although it is constitutively expressed, its intensity is significantly greater in RNA from sucrose-grown cells. The transcript could be accounted for in several ways. It could represent a second distinct gene with homology to *scrK*, or a processed derivative of the 5.0kb *scr* mRNA transcript. The fact that the 1.1kb transcript is expressed in the absence of the 5.0kb transcript argues against the latter. Possibly the most likely explanation is that it represents an additional *scrK* transcript, transcribed off a promoter within the *scr* operon. This is tentatively supported by the fact that a putative promoter sequence (TTGACA-N17-TATATT) was identified -130 to -156bp upstream of the proposed *scrK* initiation codon (Chapter 2, Fig. 2.6).

4.4.2. Targeted disruption of *scrR* and *scrB* in *C. beijerinckii*.

In order to investigate the role of the *scrR* gene product in transcriptional regulation, and establish the importance of the *scr* operon to sucrose metabolism, the *scrR* and *scrB* genes were disrupted in *C. beijerinckii*. This was achieved using the targeted integration system of Wilkinson and Young (1994), as described in Materials and Methods. *C. beijerinckii* strains carrying disrupted *scrR* and *scrB* genes were designated CBSCRR and CBSCRB respectively. Frequencies of non-replicative plasmid establishment are presented below (Table 4.1)

Table 4.1. Frequencies of plasmid establishment in *C. beijerinckii* after conjugative mobilization from *E. coli*

Plasmid	Size of *C. beijerinckii* fragment carried (bp)	Replication	Frequency per recipient
pCTC1		+	2.6×10^{-7}
pCBMU1	435	-	1.2×10^{-10}
pCBMU3	456	-	3×10^{-9}

Wilkinson and Young (1994) reported the establishment of non-replicative plasmids at frequencies of between 2.6 x 10^{-7} to 1.7 x 10^{-6} per recipient In this study, significantly lower frequencies were recorded (Table 4.1). This may reflect the fact that conjugation was conducted directly on RCM agar plates, as opposed to being conducted on cellulose filters placed on RCM plates. Whilst the approach used in this study was found to be more convenient, it may have lead to an overall decrease in conjugation efficiency. It should be noted that the values reported in Table 4.1 were obtained from a single experiment. Having obtained the desired mutants, an investigation of plasmid establishment in itself was deemed unnecessary.

Southern blot analysis was used to verify that the *scrR* and *scrB* genes had been disrupted in *C. beijerinckii* CBSCRR and CBSCRB respectively. Before considering the results of the Southern blot analysis, the various possibilities need to be considered. Wilkinson and Young (1994) reported that targeted integration most likely occurred via a Campbell-like recombination event between the target DNA, carried by the non-replicative plasmid, and its equivalent on the *C. beijerinckii* genome. In addition, Wilkinson and Young (1994) reported the establishment of single integrated plasmids, as well as the amplification of integrated plasmids (Fig. 4.3). The two alternatives would yield different results on Southern blot analysis.

A hypothetical Southern blot analysis will be considered where the 'target' DNA is prepared as a probe and hybridized against *Hin*dIII digested chromosomal DNA (A single *Hin*dIII site exists in pMTL30). In the case of a single integrated plasmid, two fragments would be identified (Fig. 4.3). These would be the left junction fragment (containing plasmid DNA, target DNA and adjacent chromosomal DNA) and the right junction fragment (containing only target DNA and adjacent chromosomal DNA) (Fig. 4.3). If multiple tandem repeats of the plasmid were present, three fragments should be identified. These would be the left junction fragment (containing plasmid DNA, target DNA and adjacent chromosomal DNA), multiple copies of the integrated plasmid (containing only plasmid and target DNA) and the right junction fragment (containing only target DNA and the adjacent chromosomal DNA)

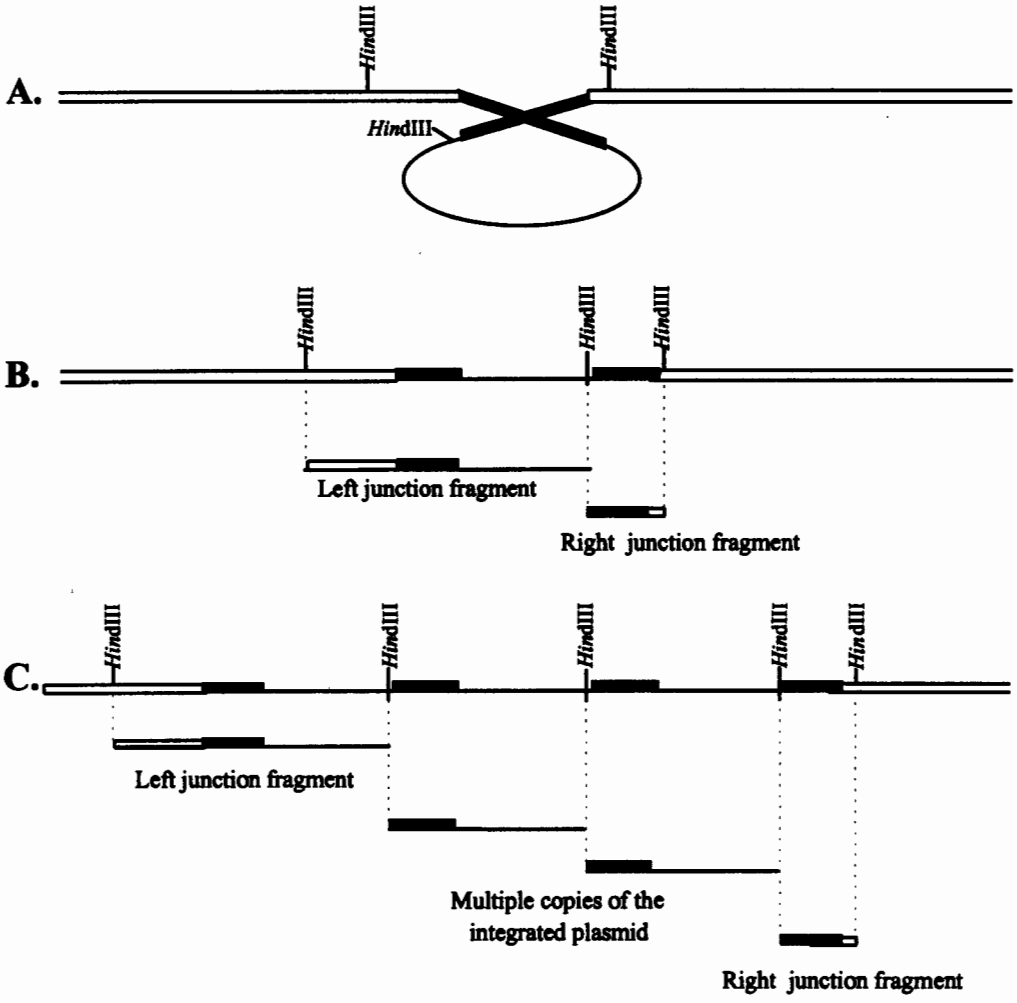

Fig. 4.3 Schematic representation of pMTL30 derivatives integrated into the *C. beijerinckii* genome. Thin lines represent pMTL30 DNA. The 'target' DNA carried by pMTL30 and its equivalent on the *C. beijerinckii* genome are represented by filled black lines. Open horizontal lines represent *C. beijerinckii* genomic DNA. A single *Hin*dIII site is present in the pMTL30 MCS. [A.] Campbell-like recombination event between the target DNA and its equivalent on the *C. beijerinckii* genome. [B.] The product of a single integration event, with a duplicated target area and hypothetical fragments generated after digestion with *Hin*dIII. [C.] The result of multiple insertions into the genome of *C. beijerinckii* and hypothetical fragments generated after digestion with *Hin*dIII.

As indicated above, Southern blot analysis was used to verify *scrB* and *scrR* disruption in *C. beijerinckii* CBSCRB and CBSCRR respectively. For the analysis, probes were prepared from the *scrB* and *scrR* 'target' regions (Fig. 4.1). These probes were hybridized against *Hin*dIII digested chromosomal DNA from wild type and mutant *C. beijerinckii* strains. Both probes hybridized to an equivalent 3.9kb fragment in *C. beijerinckii* wild type DNA (Fig. 4.4). In the case of the CBSCRB mutant strain, the *scrB* 'target' probe hybridized to three fragments. These correspond to a left junction fragment (7.6kb), multiple copies of pCBMU1 (4.8kb) and a right junction fragment (1.2kb) (Fig. 4.4). It is possibly worth noting that the 7.6kb fragment corresponds to

plasmid pCBS3, rescued from the genome of CBSCRB (Section 2.4.4). In the case of the CBSCRR mutant strain, the *scrR* 'target' probe hybridized to three fragments. These correspond to a left junction fragment (6.4kb), multiple copies of pCBMU3 (5.1kb) and a right junction fragment (2.4kb). These results are consistent with the presence of multiple tandem copies of non-replicative plasmids integrated into the *scrR* and *scrB* genes respectively. It should also be noted that since the *scrARBK* genes are transcribed as an operon, disruption of either *scrR* or *scrB* genes would disrupt the transcription of downstream genes.

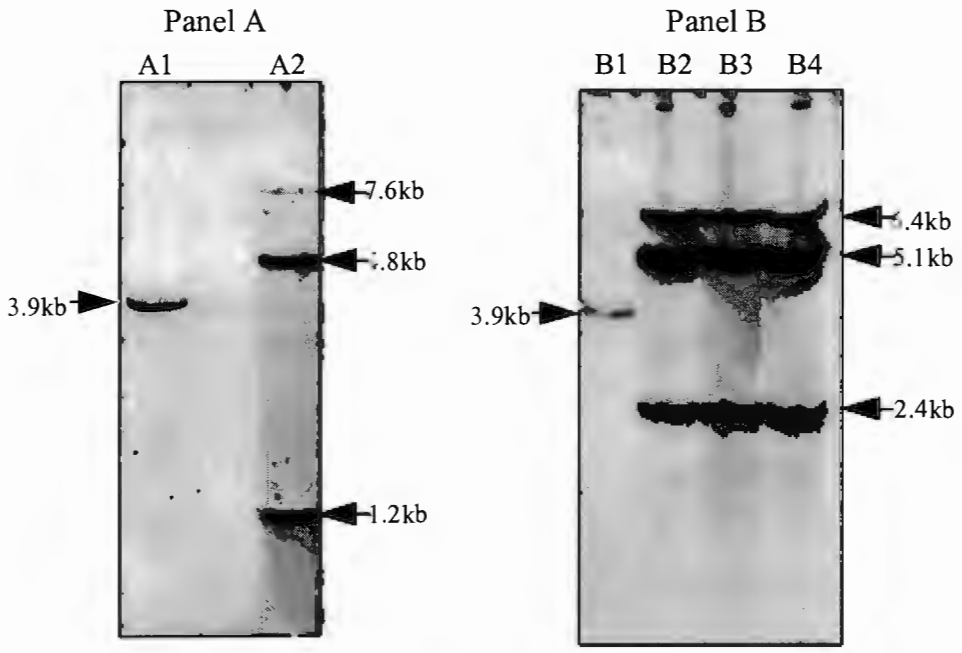

Fig. 4.4 Southern blot analysis of *C. beijerinckii* strains carrying disrupted *scrB* and *scrR* genes. Differences in signal intensity are most likely due to the presence of multiple copies of integrated plasmids, and differences in the extent to which large and small fragments are transferred by capillary blotting.
Panel A represents a Southern blot using the *scrB* 'target' DNA, carried by pCBMU1, as a probe. Lanes A1 and A2 represent *Hin*dIII digested chromosomal DNA from *C. beijerinckii* wild type and CBSCRB strains respectively.
Panel B represents a Southern blot using the *scrR* 'target' DNA, carried by pCBMU3, as a probe. Lane B1 represents *Hin*dIII digested chromosomal DNA from wild type *C. beijerinckii*. Lanes B2 to B4 represent *Hin*dIII digested chromosomal DNA extracted from equivalent strains of *C. beijerinckii* CBSCRR.

4.4.3. Effect of *scrB* and *scrR* disruptions on sucrose utilization by *C. beijerinckii*.

The construction of *C. beijerinckii* mutant strains made it possible to assess whether the *scrARBK* operon encoded the only sucrose metabolic pathway in *C. beijerinckii*. In other bacterial species such as *B. subtilis*, *S. mutans* and *Z. mobilis*, multiple sucrose metabolic pathways exist (Steinmetz, 1993; Tao *et al.*, 1993; St. Martin and Wittenberger, 1979b; Poy and Jacobson, 1990; Wexler *et al.*, 1993; Kannan *et al.*, 1995). Indeed, individual pathways may be induced by different sucrose concentrations, for example the *B. subtilis* PTS-dependent sucrose metabolic

pathway is induced at sucrose concentrations of less than 1mM while the fructosyltransferase (levansucrase) pathway is induced at concentrations over 30mM (Steinmetz *et al.*, 1989). In order to determine whether *C. beijerinckii* possessed multiple sucrose metabolic pathways, the CBSCRB and CBSCRR mutant strains were inoculated onto glucose and sucrose CMM agar plates, supplemented with erythromycin (10μg/ml). Glucose was supplied at a concentration of 10mM, while sucrose concentrations ranged between 10mM to 100mM. The plates were incubated anaerobically at 37°C overnight. Both mutant strains grew on media containing glucose as the sole carbon source, however, neither could grow where sucrose was supplied as the sole carbon source. The result clearly indicated that an intact *scr* operon was essential for growth on sucrose.

4.4.4. Northern blot analysis of *scr* operon expression in the *scrB* and *scrR* mutant strains.

One of the primary objectives of this work was to assess the role of the *scrR* gene product in transcriptional regulation of the *scr* operon. The *scrR* gene specifies a protein with homology to GalR-LacI-like regulatory proteins which typically act as transcriptional repressors. Therefore it would follow that disruption of *scrR* would result in constitutive expression of the *scr* operon. Northern blot analysis provided a convenient way of testing this.

The analysis was conducted with a probe specific for the *scrA* transcript. This was done because *scrA* lies upstream of the *scrR* and *scrB* genes and would therefore not be subject to the polar effects of disruptions in either of these genes. Total RNA was extracted from wild type *C. beijerinckii* cells grown on 0.25% (w/v) sucrose, glucose and xylose. Since the *C. beijerinckii* CBSCRR and CBSCRB strains were unable to grow on sucrose, RNA could only be extracted from cells grown on 0.25% (w/v) glucose and xylose.

As expected, in the case of the wild type *C. beijerinckii* strain, the *scrA* transcript was detected only in RNA from sucrose-grown cells (Fig. 4.5, Panel A). While the 5.0kb *scr* operon transcript was clearly visible, the signal corresponding in size to 16S RNA was more pronounced than previously observed (Fig. 4.5, Panel A). This may be accounted for by the fact that a different RNA preparation was used, which may have contained a higher proportion of degraded *scr* operon transcripts. In the case of the CBSCRR mutant, *scrA* transcripts were detected in RNA extracted from both glucose- and xylose-grown cells. Clearly, a disruption of *scrR* had resulted in

expression of the *scr* operon under non-inducing conditions. This result was consistent with the argument that *scrR* encoded a negative regulator of *scr* operon transcription. An equivalent of the 5.0kb *scr* operon transcript was not present and transcripts with a wide range in size were detected (Fig. 4.5, Panel A). This would was not unexpected since the *scr* operon had been disrupted and transcription initiated at the *scrA* promoter would have extended into multiple tandem repeats of plasmids integrated into the *scrR* gene. In the case of the CBSCRB mutant, carrying a disrupted *scrB*, a *scrA* transcript was not detected. In this strain the *scrR* gene would still have been intact and its gene product would most likely still have repressed *scr* operon transcription.

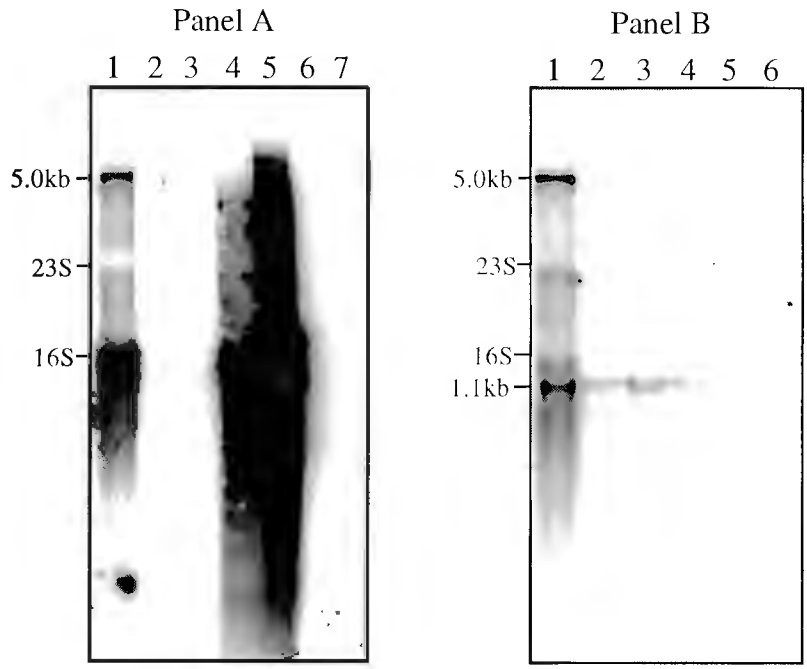

Fig. 4.5 Northern blot analysis of *C. beijerinckii* wild type, CBSCRR and CBSCRB strains. RNA was extracted from cells grown in CMM with carbohydrates supplied at 0.25% (w/v).
Panel A Northern blot conducted with a *scrA* specific probe. Lanes 1 to 3 represent RNA extracted from wild type *C. beijerinckii* cells grown on (1) sucrose (2) glucose and (3) xylose. Lanes 4 and 5 represent RNA extracted from *C. beijerinckii* CBSCRR cells grown on (4) glucose and (5) xylose. Lanes 6 and 7 represent RNA extracted from *C. beijerinckii* CBSCRB cells grown on (6) glucose and (7) xylose.
Panel B Northern blot conducted with a *scrK* specific probe. Lanes 1 and 2 represent RNA extracted from wild type *C. beijerinckii* cells grown on (1) sucrose and (2) glucose. Lanes 3 and 4 represent RNA extracted from *C. beijerinckii* CBSCRR cells grown on (3) glucose and (4) xylose. Lanes 5 and 6 represent RNA extracted from *C. beijerinckii* CBSCRB cells grown on (5) glucose and (6) xylose.

While the initial northern blot analysis had indicated that the individual *scr* genes were transcribed as an 5.0kb operon, an additional 1.1kb transcript was detected with a *scrK* specific probe (Fig. 4.2, Panel D). In the case of the CBSCRR and CBSCRB strains, northern blot

analysis, indicated that the 1.1kb transcript was expressed, even when the *scr* operon had been disrupted (Fig. 4.5, Panel B). This demonstrated that expression of the 1.1kb transcript was not dependent on the presence of an intact *scr* operon.

4.5. CONCLUSIONS

Northern blot analysis demonstrated that the *scrARBK* genes were co-transcribed as a sucrose inducible operon. It follows therefore that transcriptional regulation must be exercised at the *scrA* promoter(s). The proposed *scrR* gene product (ScrR) had homology to members of the GalR-LacI family of transcriptional regulators and was proposed to be involved in *scr* operon regulation. The role of the *scrR* gene product was investigated by construction of a *C. beijerinckii* strain carrying a disrupted *scrR*. In this strain, the *scr* operon was transcribed under non-inducing conditions. This confirmed that the *scrR* gene product played a role in regulation and suggested a negative regulatory mechanism. While the precise details of this mechanism remain unclear, analogy with other GalR-LacI-like regulatory proteins, particularly the *S. xylosus* ScrR, suggests that in the absence of sucrose the *C. beijerinckii* ScrR may repress transcription by binding to imperfectly palindromic operator sites in the vicinity of the *scr* operon promoter. Such a site was identified ($G_6A_5A_4A_3A_2C_1/G'_1G'_2T'_3T'_4T'_5C'_6$) and lies between the putative *scrA* promoter and initiation codon (Chapter 2, Fig. 2.6) However, a detailed analysis of the *scrA* promoter region, followed by gel mobility shift assays, using purified ScrR, would be required to confirm the mechanism by which ScrR acts.

In most other sucrose-utilization regulons, with the exception of *L. lactis* (Rauch and de Vos, 1992b), the corresponding *scrR* genes are independently transcribed. In contrast, the *C. beijerinckii scrR* is transcribed as part of the *scr* operon, and it follows that it is negatively autoregulated. The *scrR* gene product (ScrR) is likely to act at few, possibly only one, site in the genome. Therefore it would seem unnecessary for large amounts of ScrR to be synthesized as *scr* operon transcripts accumulate under inducing conditions. Indeed, if large amounts of ScrR were produced, it could conceivably lead to the repression of *scr* operon transcription, even in the presence of inducer. Therefore, it seems likely that other mechanisms may exist to limit the translation of *scrR*. Gram-positive bacteria are stringent in their requirements for high-level initiation of protein synthesis. In *B. subtilis* translational yield can be substantially reduced if ribosome-binding-sites and initiation codons are separated by less than 6bp (Vellanoweth and

Rabinowitz, 1992). In the case of *scrR*, only 5bp separate the proposed ribosome-binding-site and initiation codon. It is possible that this sub-optimal spacing may limit translation of *scrR*. This in turn may allow for optimal expression of the *scr* operon, and the build-up of gene products other than ScrR, while the inducer is present.

While Northern blot analysis indicated that *scrK* was part of the *scr* operon, an additional 1.1kb transcript was identified with the *scrK* probe. Although this transcript appeared to be induced by sucrose, it was constitutively expressed at a low level. The transcript was also expressed in *C. beijerinckii* CBSCRR and CBSCRB strains, indicating that an intact *scr* operon was not necessary for its expression. The 1.1kb transcript may represent the mRNA of an independent gene with homology to *scrK*. However, a likely explanation is that it represents an additional *scrK* transcript expressed off a promoter within the *scr* operon. Indeed, a putative promoter was identified upstream of *scrK*. It is not inconceivable that such a promoter could be a vestigial promoter, which facilitated expression of *scrK* prior to its association with the *scr* operon.

The importance of the *scr* operon, to sucrose utilization by *C. beijerinckii*, was also established. This was achieved using *C. beijerinckii* strains CBSCRR and CBSCRB carrying disrupted *scrR* and *scrB* genes respectively. Both strains were unable to grow when sucrose was supplied as the sole carbohydrate. Since both *scrR* and *scrB* disruptions would have polar effects, caution should be exercised before attributing this result directly to the disrupted gene. What is clear, however, is that an intact *scr* operon is essential to sucrose metabolism. If other sucrose utilization systems exist, they are either not induced by sucrose, or are unable to act independently of the *scr* operon.

CHAPTER 5

GENERAL CONCLUSION

During the study a sucrose utilization operon, *scrARBK*, was cloned and sequenced from *C. beijerinckii* NCIMB 8052. Sequence analysis indicated that the operon consisted of four genes proposed to encode an EIIBC^{Scr} PTS protein (ScrA), a GalR-LacI-like transcriptional regulatory protein (ScrR), a sucrose-6-phosphate hydrolase of the glucosyl hydrolase family 32 (ScrB) and a fructokinase of the ribokinase/*pfkB* family (ScrK). In the cases of the *scrB*, *scrK* and *scrR* genes, experimental evidence was presented to substantiate proposals made on the basis of sequence analysis, *viz*: *scrB* and *scrK* conferred sucrose hydrolase and fructokinase activity on *E. coli*, while disruption of *scrR* in *C. beijerinckii* resulted in constitutive expression of the *scr* operon. In the case of *scrA*, its deduced gene product had distinct homology with EIIBC^{Scr} proteins and its association with the *scr* operon strongly suggested a role in PTS-dependent sucrose uptake. The importance of the *scr* operon to sucrose metabolism in *C. beijerinckii* was demonstrated by the fact that disruption of the operon resulted in an inability to utilize sucrose.

Regarding the question of induction of the sucrose utilization system by sucrose, equivalent effects were observed at the molecular and physiological levels. At the molecular level northern blot analysis indicated that the *scr* operon was transcriptionally induced in sucrose-grown cells but not in cells grown on other carbohydrates. At the physiological level it was demonstrated that sucrose hydrolase and fructokinase activities were induced in sucrose-grown cells but not in cells grown on other carbohydrates. This has also been found to be the case for PEP-dependent sucrose phosphorylation activity (personal communication from Tangney and Mitchell, Department of Biological Sciences, Heriot-Watt University, Edinburgh). The observed regulation is almost certainly mediated by the *scrR* gene product, since disruption of *scrR* resulted in expression of the operon under non-inducing conditions.

The work presented in this thesis provides a basic molecular and physiological analysis of the *C. beijerinckii* sucrose utilization system. It seems appropriate that further work should focus on establishing the mechanisms by which sucrose metabolism is regulated. It was proposed that ScrR negatively regulates *scr* operon transcription, by binding to its promoter region in

the absence of sucrose. However, details of this mechanism need to be experimentally established. Mapping of the transcription initiation site for the *scr* operon, by primer extension analysis, would allow for a precise identification of the promoter region. Gel mobility shift assays using purified ScrR and the *scr* operon promoter region could then verify ScrR interaction with the *scr* operon promoter region. Gel mobility shift assays would also allow for the identification of ScrR operator sites and confirm that ScrR binds its operator sites in the unliganded state. Testing sucrose, glucose and fructose and their phosphorylated derivatives for their ability to prevent interaction of ScrR and its operator sites, would provide a means of identifying the *in vivo* inducer molecule. The issue of glucose mediated repression of the sucrose utilization system also requires further analysis. However, it is possible that glucose-mediated repression acts at several levels. The mechanisms facilitating catabolite repression in *C. beijerinckii* require further investigation before their effects on the sucrose utilization system can be established.

APPENDIX A

Bacterial strains used in this study

Strain	Genetic characteristic(s)	Reference
Eschericia coli		
JM109	*recA1 endA1 gyrA96 thi hsdR17 supE44 relA1* λ^- *Δ(lac-proAB)* (F', *traD36 proAB lacI*^{q}*ZΔM15*)	Yanisch-Perron *et al* (1985)
CA474	F$^-$ *λcI857 hsdS20*($r_B^- m_B^-$) *supE44 aro-14 galK2 lacY1 proA2 rpsL20*(Sm^R) *xyl-5 mtl-1 recA13*	*λcI857* lysogen of HB101 containing R702 Wilkinson and Young (1994)
Clostridium beijerinckii		
NCIMB 8052	Prototrophic	Laboratory strain
CBSCRB	*scrB::erm*	Insertion of pCBMU1 in NCIMB 8052
CBSCRR	*scrR::erm*	Insertion of pCBMU3 in NCIMB 8052

APPENDIX B

CLONING VECTORS

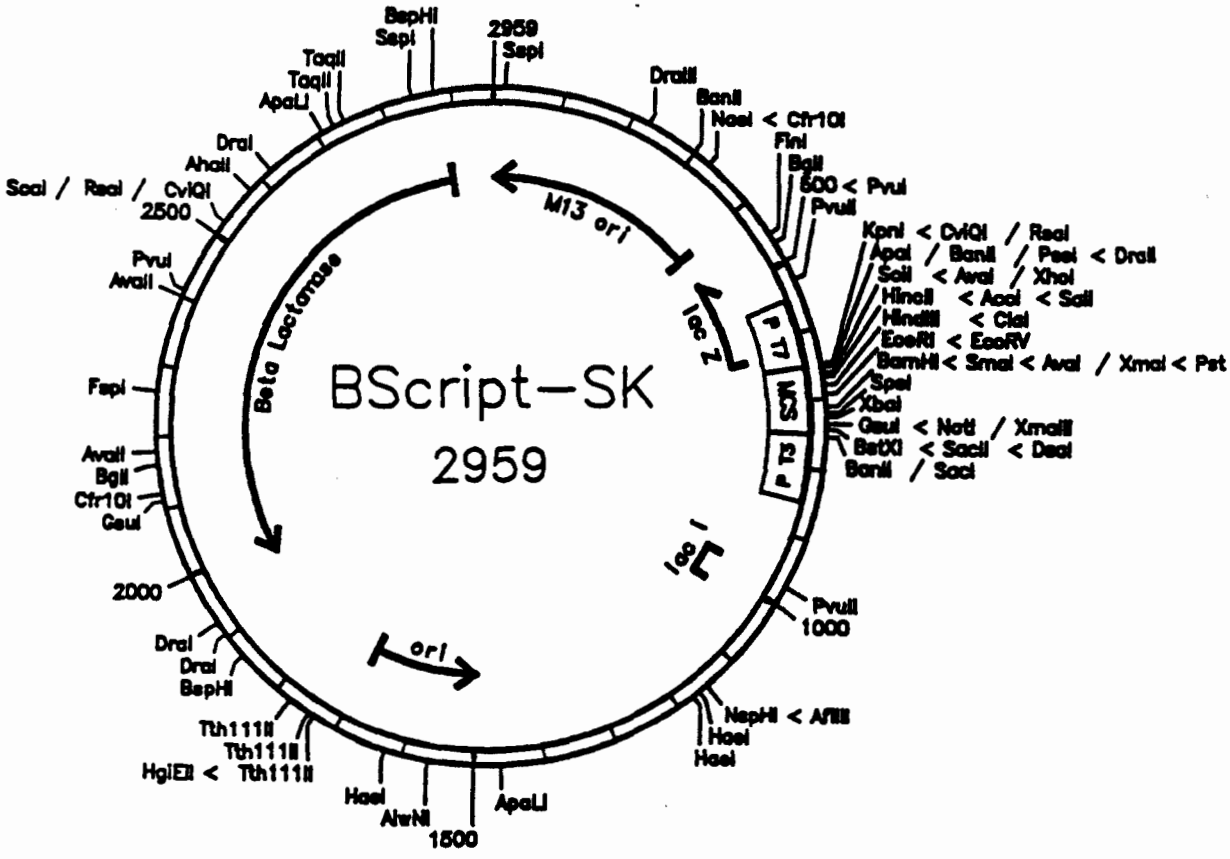

Fig. B1. Restriction map of Bluescript SK (Stratagene, San Diego)

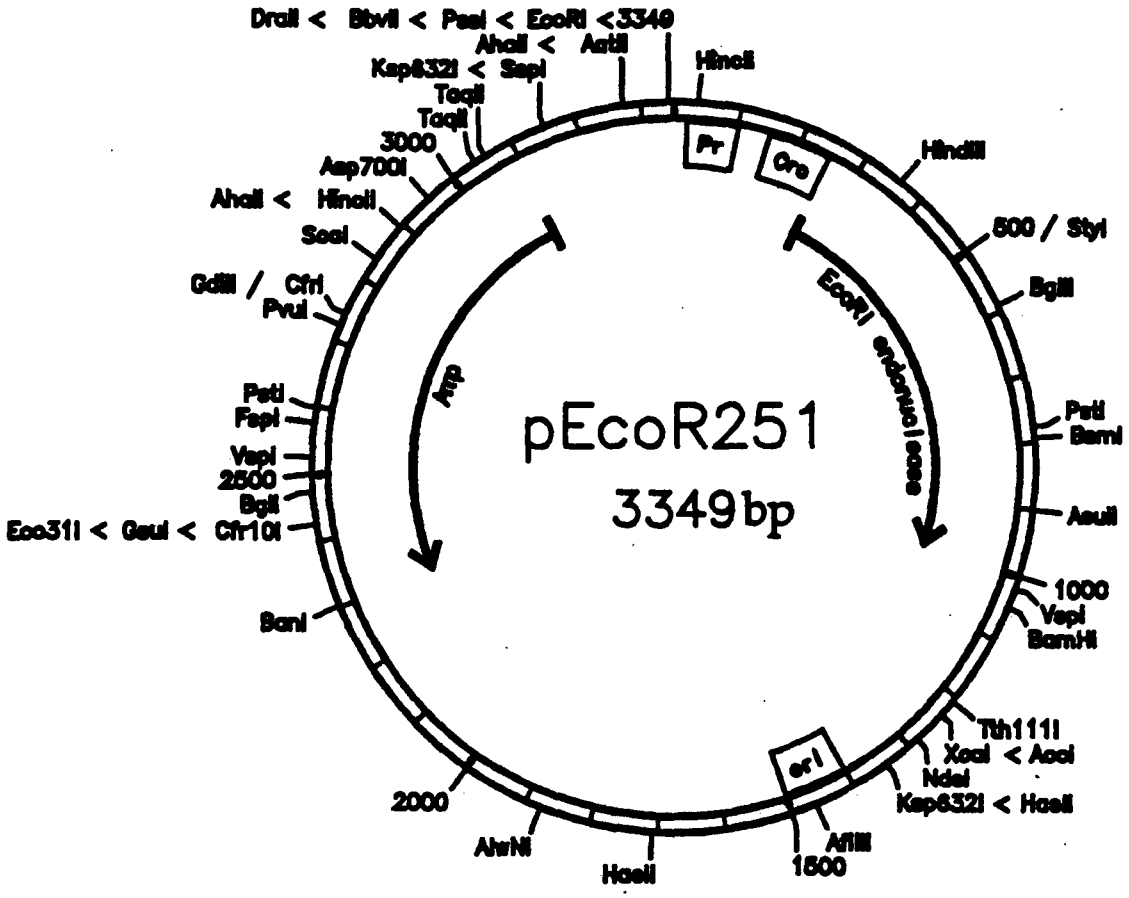

Fig. B.2 Restriction map of pEcoR251 (Zabeau and Stanley, 1982).

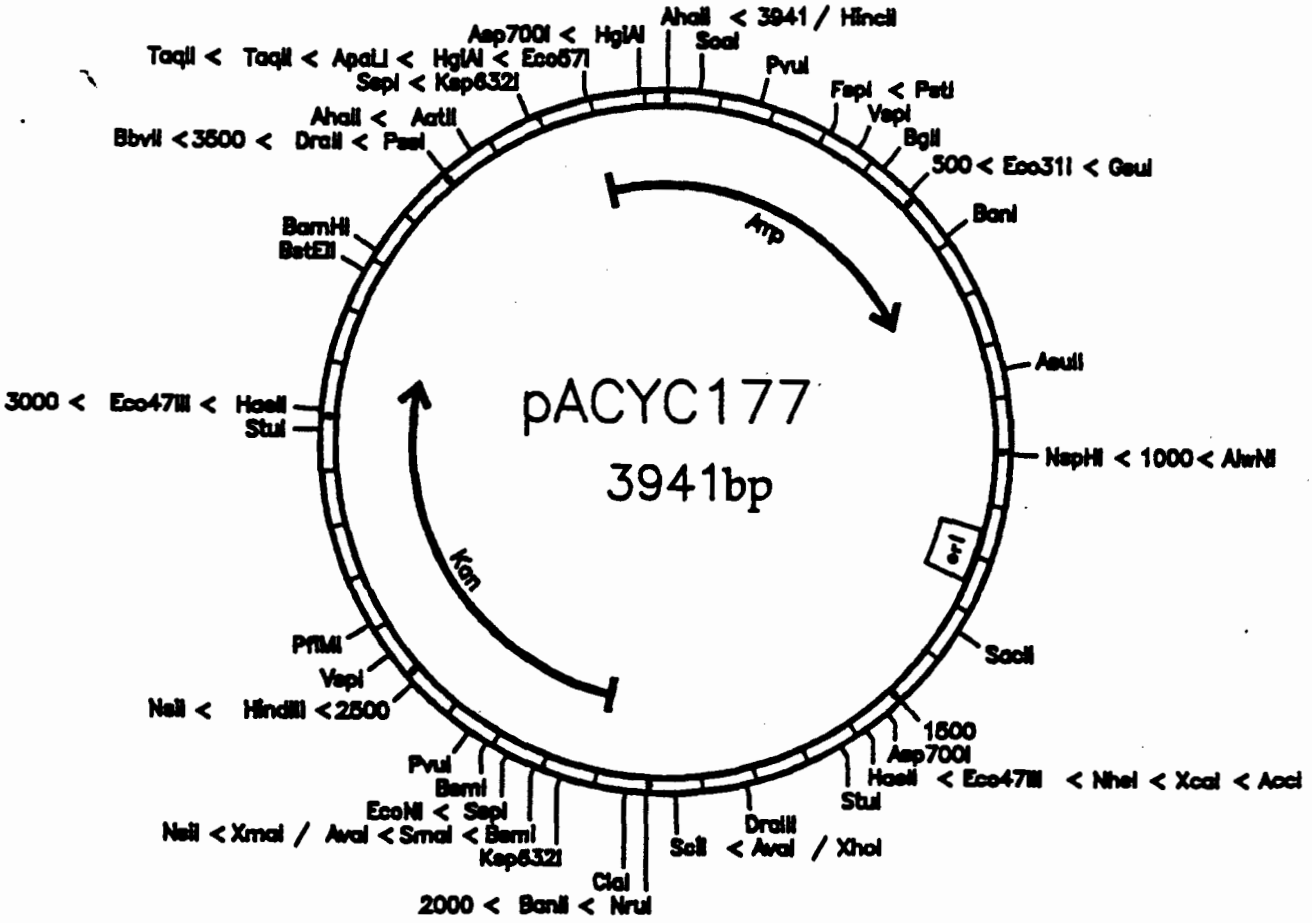

Fig. B.3 Restriction map of pACYC177 (Chang and Cohen, 1978).

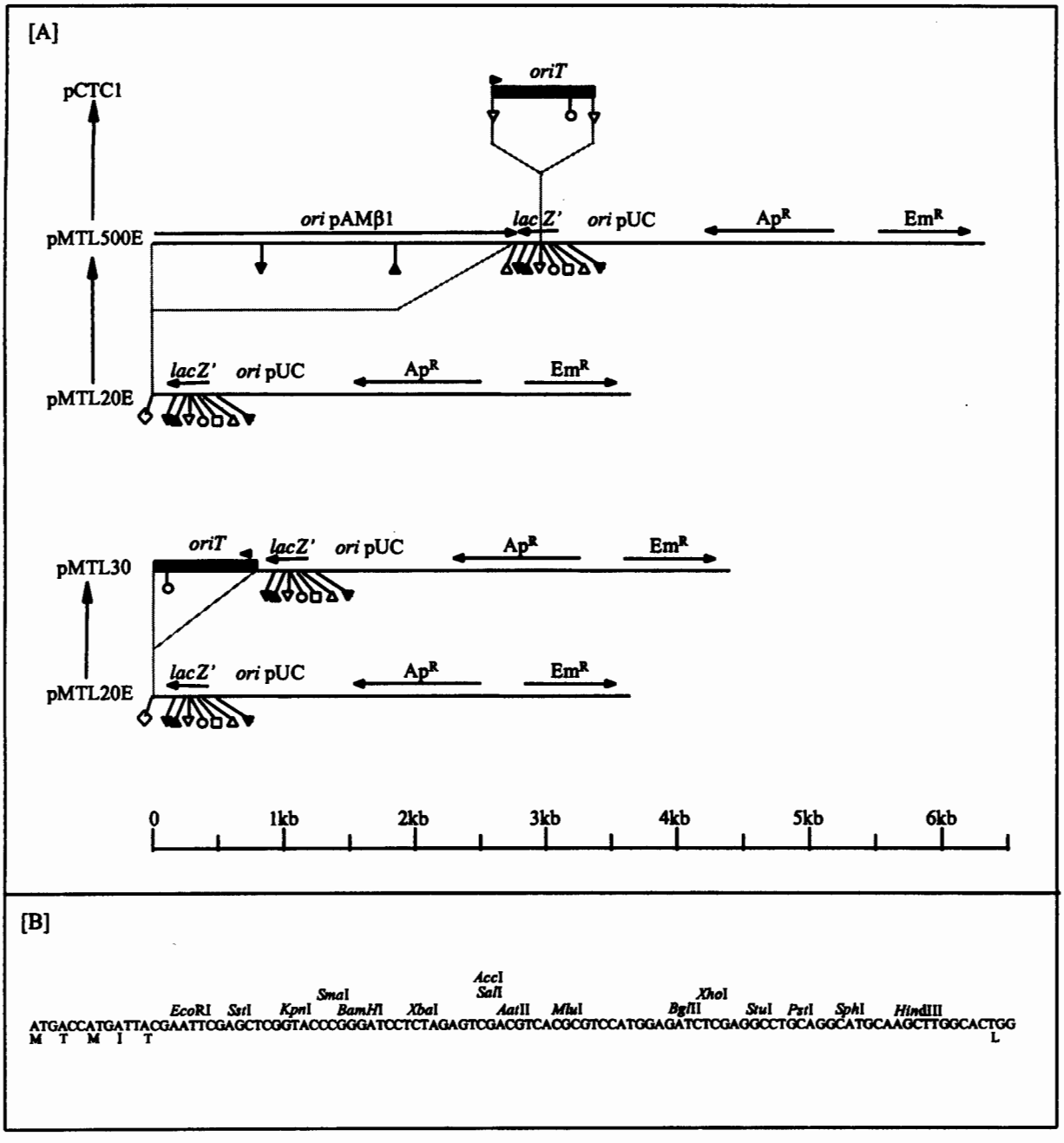

Fig. B.4 (A) Relationships between pMTL20E and its derivatives: pMTL30, pMTL500E and pCTC1 (For complete details the following references should be consulted: Oultram *et al.* (1988); Williams *et al.* (1990a and b) and Minton *et al.* (1990). The construction of pMTL20E has been described previously (Oultram *et al.*, 1988) and involved cloning the erythromycin resistance gene of the streptococcal plasmid pAMβ1 into the unique *Eco*RV site of pMTL20 (Chambers *et al.*, 1988). Williams *et al.* (1990a) constructed pMTL30 by cloning the *oriT* region of RK2 into the unique *Nhe*I(◇) site of pMTL20E. Oultram *et al.* (1988) constructed pMTL500E by cloning the replication origin of the streptococcal pAMβ1 into the unique *Nhe*I site of pMTL20E. Williams *et al.* (1990b) then constructed pCTC1 by cloning the *oriT* site into the *Pst*I site of the MCS of pMTL500E. Restriction sites are denoted as follows: ○, *Ava*I; □,*Bam*HI, △,*Eco*RI, ▲,*Hin*dIII; ▽,*Pst*I, ▼,*Pvu*II, ◇, *Nhe*I (adapted from Williams *et al.*, 1990b) (B) Detailed representation of the multiple cloning site (MCS) of pMTL20 (Chambers *et al.*, 1988). The encoded amino acid sequence of *lacZ'* is indicated below the illustrated nucleotide sequence.

Appendix C

Preparation of Clostridial Minimal Medium (CMM)

(Modified from *Clostridium acetobutylicum* minimal medium (Long, 1984))

The following components were diluted in distilled water to a volume of 1 litre.

Carbohydrate source (sucrose, glucose, fructose, maltose or xylose) ..... Typically 2.5 g
NH_4 acetate ..... 2 g
Casein hydrolysate ..... 2 g
Agar (for solid media only) ..... 15 g
$MgSO_4.7H_2O$ stock solution (20%, w/v) ..... 1 ml
$MnSO_4.4H_2O$ stock solution (1%, w/v) ..... 1 ml
$FeSO_4.7H_2O$ stock solution (1%, w/v) ..... 1 ml
Para-aminobenzoic acid stock solution (0.1%, w/v) ..... 1 ml
Thiamine.HCl stock solution (0.1%, w/v) ..... 1 ml
Biotin stock solution (0.02%, w/v) ..... .0.25 ml
Salt stock solution (See below) ..... 40 ml
Resazurin stock solution (0.025%, w/v) ..... 2 ml

After autoclaving, while still relatively hot, the following sterile solutions were added.

$NaHCO_3$ stock solution (10%, w/v) 10ml
Cysteine stock solution (5%, w/v) 10ml

Liquid CMM was transferred directly to an anaerobic cabinet while solid CMM was allowed to solidify under aerobic conditions before being transferred to an anaerobic cabinet. In both cases media were allowed to equilibrate under anaerobic conditions for at least 12 hours before use.

Salt stock solution:

$CaCl_2$ (anhydrous) ..... 0.2 g
$MgSO_4.7H_2O$ ..... 0.2 g
Dissolved in 300ml

KH_2PO_4 ..... 1 g
K_2HPO_4 ..... 1 g
$NaHCO_3$ ..... 10 g
NaCl ..... 2 g
Dissolved in 700ml

The two solutions were added slowly together and neutralized to prevent formation of a precipitate

Literature Cited

Afschar, A.S., C.E. Vaz Rossell, and K. Schaller. 1990. Bacterial conversion of molasses to acetone and butanol. Appl. Microbiol. Biotechnol. **34:** 168-171.

Albasheri, K.A., and W.J. Mitchell. 1995. Identification of two α-glucosidase activities in *Clostridium acetobutylicum* NCIB 8052. Journal of Applied Bacteriology. **78:** 149-156.

Annous, B.A., and H.P. Blaschek. 1990. Regulation and localization of amylolytic enzymes in *Clostridium acetobutylicum* ATCC 824. Appl. Environ. Microbiol. **56:** 2559-2561.

Annous, B.A., and H.P. Blaschek. 1991. Isolation and characterization of *Clostridium acetobutylicum* mutants with enhanced amylolytic activity. Appl. Environ. Microbiol. **57:** 2544-2548.

Armitage, P, R. Walden and J. Draper. 1988. Vectors for the transformation of plant cells using agrobacterium. *In* Plant Genetic Transformation and gene expression, A laboratory manual. Draper, J., R. Scott, P. Armitage, and W. Walden. Blackwell Scientific Publications Ltd. Oxford.

Arnaud, M., M. Debarbouille, G. Rapoport, M.H. Saier, Jr., and J. Reizer. 1996. In vitro reconstitution of transcriptional antitermination by the SacT and SacY proteins of *Bacillus subtilis*. J. Biol. Chem. **271:** 18966-18972.

Arnaud, M., P. Vary, M. Zagorec, A. Klier, M. Debarbouille, P. Postma, and G. Rapoport. 1992. Regulation of the *sacPA* operon of *Bacillus subtilis*: identification of phosphotransferase system components involved in SacT activity. J. Bacteriol. **174:** 3161-3170.

Aulkemeyer, P., Ebner, R., Heilenmann, G., Jahreis, K., Schmid, K., Wrieden, S., and Lengeler, J.W. 1991. Molecular analysis of two fructokinases involved in sucrose metabolism of enteric bacteria. Mol. Microbiol. **5:** 2913-2922.

Aymerich, S., and M. Steinmetz. 1992. Specificity determinants and structural features in the RNA target of the bacterial antiterminator proteins of the BglG/SacY family. Proc. Natl. Acad. Sci. U S A. **89:** 10410-10414.

Blatch, G.l. 1990. Molecular analysis of the *Vibrio alginolyticus* sucrose utilization system cloned into *Escherichia coli*. PhD thesis, Department of Microbiology, University of Cape Town.

Blatch, G.L, and D.R. Woods. 1993. Molecular characterization of a fructanase produced by *Bacteroides fragilis* BF-1. J. Bacteriol. **175:** 3058-3066.

Blatch, G.L., R.R. Scholle, and D.R. Woods. 1990. Nucleotide sequence and analysis of the *Vibrio alginolyticus* sucrose uptake-encoding region. Gene. **95:** 17-23.

Blatch, G.L., and D.R. Woods. 1991. Nucleotide sequence and analysis of the *Vibrio alginolyticus scr* repressor-encoding gene (*scrR*). Gene. **101:** 45-50.

Bockmann, J., H. Heuel, and J.W. Lengeler. 1992. Characterization of a chromosomally encoded, non-PTS metabolic pathway for sucrose utilization in *Escherichia coli* EC3132. Mol. Gen. Genet. **235:** 22-32.

Bork, P., C. Sander, and A. Valencia. 1993. Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases. Protein Sci. **2:** 31-40.

Bradford, M. M. 1976. A rapid and sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72:**248-254..

Brown, D.P., L. Ganova-Raeva, B.D. Green, S.R. Wilkinson, M. Young, and P. Youngman. 1994. Characterization of spo0A homologues in diverse Bacillus and Clostridium species identifies a probable DNA-binding domain. Mol. Microbiol. **14:** 411-426.

Brückner, R., E. Wagner, and F. Gotz. 1993. Characterization of a sucrase gene from *Staphylococcus xylosus*. J. Bacteriol. **175:** 851-857.

Burne, R.A., and J.E. Penders. 1992. Characterization of the *Streptococcus mutans* GS-5 *fruA* gene encoding exo-beta-D-fructosidase. Infect. Immun. **60:** 4621-4632.

Burne, R.A., and J.E. Penders. 1994. Differential localization of the *Streptococcus mutans* GS-5 fructan hydrolase enzyme, FruA. FEMS Microbiol. Lett. **121:** 243-249.

Campbell, A.M. 1962. Episomes. Adv. Genet. **11:** 101-146.

Chambers, S.P., S.E. Prior, D.A. Barstow, and N.P. Minton. 1988. The pMTL*nic*⁻ cloning vectors. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. Gene. **68:** 139-149.

Chambert, R., and G. Gonzy-Treboul. 1976. Levansucrase of *B.subtilis*: kinetic and thermodynamic aspects of the transfructosylation process. Eur. J. Biochem. **62:** 55-64.

Chen, J-S. 1993. Properties of acid- and solvent-forming enzymes of clostridia. p 51-76. *In* Woods D.R. (ed.) The clostridia and biotechnology, Butterworth-Heinemann, Stoneham, MA, USA.

Chen, Y.Y., L.N. Lee, and D.J. LeBlanc. 1992. Genetic analysis of *scrA* and *scrB* from *Streptococcus sobrinus* 6715. Infect. Immun. **60:** 3739-3746.

Chen, Y.Y., L.N. Lee, and D.J. LeBlanc. 1993. Sequence analysis of *scrA* and *scrB* from *Streptococcus sobrinus* 6715. Infect. Immun. **61:** 2602-2610.

Collett, H. 1996. Characterization of a transposon-induced pleiotropic metronidazol resistant mutant of *Clostridium acetobutylicum* P262. PhD Thesis, University of Cape Town, South Africa.

Collins, M.D., U.M. Rodrigues, R.H. Dainty, R.A. Edwards, and T.A. Roberts. 1992. Taxonomic studies on a psychrophilic Clostridium from vacuum-packed beef: description of *Clostridium estertheticum* sp.nov. FEMS Microbiol. Lett. **96:** 235-240.

Cowan, P.J., Nagesha, H., Leonard, L., Howard, J.L., and Pittard, A.J. 1991. Characterization of the major promoter for the plasmid-encoded sucrose genes *scrY*, *scrA*, and *scrB*. J. Bacteriol. **173:** 7464-7470.

Crutz, A.M., and M. Steinmetz. 1992. Transcription of the *Bacillus subtilis sacX* and *sacY* genes, encoding regulators of sucrose metabolism, is both inducible by sucrose and controlled by the DegS-DegU signalling system. J. Bacteriol. **174:** 6087-6095.

Crutz, A.M., M. Steinmetz, S. Aymerich, R. Richter, and D. Le Coq. 1990. Induction of levansucrase in *Bacillus subtilis*: an antitermination mechanism negatively controlled by the phosphotransferase system. J Bacteriol. **172:** 1043-1050.

Davison, S.P., J.D. Santangelo, S.J. Reid, and D.R. Woods. 1995. A *Clostridium acetobutylicum* regulator gene (*regA*) affecting amylase production in *Bacillus subtilis*. Microbiology. **141:** 989-996.

Debarbouille, M., M. Arnaud, A. Fouet, A. Klier, and G. Rapoport. 1990. The *sacT* gene regulating the *sacPA* operon in *Bacillus subtilis* shares strong homology with transcriptional antiterminators. J. Bacteriol. **172:** 3966-3973.

Di Marco, A.A., and A.H. Romano. 1985. D-Glucose transport system of *Zymomonas mobilis*. Appl. Environ. Microbiol. **49:** 151-157.

Donkersloot, J. A., and J. Thompson. 1990. Simultaneous loss of N5-(carboxyethyl)ornithine synthase, nisin production, and sucrose-fermenting ability by *Lactococcus lactis* K1. J. Bacteriol. **172:** 4122-4126.

Doroshenko, V.G., V.N. Danilevich, G.I. Karataev, and V.A. Livshits. 1988. Structural and functional organization of the transposon Tn2555, carrying genes of sucrose utilization. Mol. Biol. (Mosk) **22:** 645-658.

Dürre, P., R.J. Fischer, A. Kuhn, K. Lorenz, W. Schreiber, B. Sturzenhofecker, S. Ullmann, K. Winzer, and U. Sauer. 1995. Solventogenic enzymes of *Clostridium acetobutylicum*: catalytic properties, genetic organization, and transcriptional regulation. FEMS Microbiol. Rev. **17:** 251-262.

Ebner, R., and J. W. Lengeler. 1988. DNA sequence of the gene *scrA* encoding the sucrose transport protein EnzymeII^{Scr} of the phosphotransferase system from enteric bacteria: homology of the EnzymeII^{Scr} and EnzymeII^{Bgl} proteins. Mol. Microbiol. **2:** 9-17.

Efstathiou, I., G. Reyesset, and N. Truffet. 1986. A study of inulinase activity in the *Clostridium acetobutylicum* strain ABKn8. Appl. Microbiol. Biotechnol. **25:** 143- 149.

Fennington, G.J., and T.A.Hughes. 1996. The fructokinase from *Rhizobium leguminosarum* biovar *trifolii* belongs to group I frutokinase enzymes and is encoded separately from other carbohydrate metabolism enzymes. Microbiology **142:** 321-330.

Forst, D., K. Schulein, T. Wacker, K. Diederichs, W. Kreutz, R. Benz, and W. Welte. 1993. Crystallization and preliminary X-ray diffraction analysis of ScrY, a specific bacterial outer membrane porin. J. Mol. Biol. **229:** 258-262.

Fouet, A., M. Arnaud., A. Klier, and G. Rapoport. 1987. *Bacillus subtilis* sucrose-specific enzymeII of the phosphotransferase system: Expression in *Escherichia coli* and homology to enzymes II from enteric bacteria. Proc. Natl. Acad. Sci. **84:** 8773-8777.

Fouet, A., A. Klier, and G. Rapoport. 1986. Nucleotide sequence of the sucrase gene of *Bacillus subtilis*. Gene. **45:** 221-225.

Friedman, A.M., T.O. Fischmann, and T.A. Steitz. 1995. Crystal structure of *lac* repressor core tetramer and its implications for DNA looping. Science. **268:** 1721-1727.

Gasson, M. J. 1984. Transfer of sucrose fermenting ability, nisin resistance and nisin production in *Streptococcus lactis* 712. FEMS Microbiol. Lett. **21:** 7-10.

Gay, P., D. Le Coq, M. Steinmetz, E. Ferrari, and J.A. Hoch. 1983. Cloning structural gene *sacB*, which codes for exoenzyme levansucrase of *Bacillus subtilis*: expression of the gene in *Escherichia coli*. J. Bacteriol. **153:** 1424-1431.

Gering, M., and R. Brückner. 1996. Transcriptional regulation of the sucrase gene of *Staphylococcus xylosus* by the repressor ScrR. J. Bacteriol. **178:** 462-469.

Green, M.G., Z.l. Boynton, L.M Harris, F.B. Rudolph, E.T. Papoutsakis, and G.N. Bennett. 1996. Genetic manipulation of acid formation pathways by gene inactivation in *Clostridium acetobutylicum* ATCC 824. Microbiology **142:** 2079-2086.

Green, M.G., and G.N. Bennett. 1996. Inactivation of an aldehyde/alcohol dehydrogenase gene from *Clostridium acetobutylicum* ATCC 824. Appl. Biochem. Biotechnol. **57:** 213-221.

Gunasekaran, P., T. Karunakaran, B. Cami, A.G. Mukundan, L. Preziosi, and J. Baratti. 1990. Cloning and sequencing of the *sacA* gene: characterization of a sucrase from *Zymomonas mobilis*. J. Bacteriol. **172:** 6727-6735.

Hardesty, C., C. Ferran, and J.M. DiRienzo. 1991. Plasmid-mediated sucrose metabolism in *Escherichia coli:* characterization of *scrY*, the structural gene for a phosphoenolpyruvate-dependent sucrose phosphotransferase system outer membrane porin. J. Bacteriol. **173:** 449-456.

Hedges, R.W., and A.E. Jacob. 1974. Transposition of ampicillin resistance from RP4 to other replicons. Mol. Gen. Genet. **132:** 31-34.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28:** 351-359

Henrissat, B. 1991. A classification of glucosyl hydrolases based on amino acid sequence similarities. Biochem. J. **280:** 309-316.

Henrissat, B., and A. Bairoch. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J. **293:** 781-788.

Henrissat, B., and A. Bairoch. 1996. Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. **316:** 695-696.

Holt, R.A., Stephens, G.M., and J.G. Morris. 1984. Production of solvents by *Clostridium acetobutylicum* cultures maintained at neutral pH. Appl. Environ. Microbiol. **48:** 1166-1170.

Honda, O., C. Kato, and H.K. Kuramitsu. 1990. Nucleotide sequence of the *Streptococcus mutans gtfD* gene encoding the glucosyltransferase-S enzyme. J. Gen. Microbiol. **136:** 2099-2105.

Hueck, C. J., W. Hillen, and M. H. Saier, Jr. 1994. Analysis of a cis-active sequence mediating catabolite repression in Gram-positive bacteria. Res. Microbiol. **145:** 503-518.

Ingram, L.O., C.K. Eddy, K.F. Mackenzie, T. Conway, and F. Alterthum. 1989. Genetics of *Zymomonas mobilis* and ethanol production. Dev. Ind. Microbiol. **30:** 53-69

Jahreis, K., and J.W. Lengeler. 1993. Molecular analysis of two ScrR repressors and of a ScrR-FruR hybrid repressor for sucrose and D-fructose specific regulons from enteric bacteria. Mol. Microbiol. **9:** 195-209.

Johnson, J.L., and J-S. Chen. 1995. Taxonomic relationships among strains of *Clostridium acetobutylicum* and other phenotypically similar organisms. FEMS Microbiol. Rev. **17:**233-240.

Jones, D.T., and S. Keis. 1995. Origins and relationships of industrial solvent-producing clostridial strains. FEMS Microbiol. Rev. **17:**223-232.

Jones, D.T., and D.R. Woods. 1986. Acetone-butanol fermentation revisited. Microbiol. Rev. **50:** 484-524.

Kannan, R., G. Mukundan, N. Ait-Abdelkader, V. Augier-Magro, J. Baratti, and P. Gunasekaran. 1995. Molecular cloning and characterization of the extracellular sucrase gene (*sacC*) of *Zymomonas mobilis*. Arch. Microbiol. **163:** 195-204.

Kashket, E.R., and Z. Cao. 1993. Isolation of a degeneration-resistant mutant of *Clostridium acetobutylicum* NCIMB 8052. Appl. Environ. Microbiol. **59:** 4198-4202.

Kashket, E.R., and Z. Cao. 1995. Clostridial strain degeneration. FEMS Microbiol. Rev. **17:** 307-315.

Keis, S., C.F. Bennett, V.K. Ward, and D.T. Jones. 1995. Taxonomy and phylogeny of industrial solvent-producing clostridia. Int. J. Syst. Bacteriol. **45:** 693-705.

Kiska, D.L., and F.L. Macrina. 1994. Genetic regulation of fructosyltransferase in *Streptococcus mutans*. Infect. Immun. **62:** 1241-1251.

Kleina, L.G., and J.H. Miller. 1990. Genetic studies of the *lac* repressor. Extensive amino acid replacements generated by the use of natural and synthetic nonsense suppressors. J. Mol. Biol. **212:** 295-318.

Koga, T., K. Nakamura, Y. Shirokane, K. Mizusawa, S. Kitao, and M. Kikuchi. 1991. Purification and some properties of sucrose phosphorylase from *Leuconostoc mesenteroides*. Agric. Biol. Chem. **55:** 1805-1810.

Kunst, F., T. Msadek, and I. Martin-Verstraete. 1990. The sucrose system of *Bacillus subtilis*. Proceedings: 6th International Conference on Genetics of Industrial Microorganisms. **vol 2:** 702-713.

Kyte, J., and R.F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:** 105-132.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:** 680-685

Lengeler, J.W., K. Jahreis and U.F. Wehmeier. 1994. Enzymes II of the phosphoenolpyruvate-dependent phosphotransferase systems: their structure and function in carbohydrate transport. Biochim. Biophys. Acta. **1188:** 1-28.

Lengeler, J.W., R.J. Mayer, and K. Schmid. 1982. Phosphoenolpyruvate-dependent phosphotransferase system enzyme III and plasmid-encoded sucrose transport in *Escherichia coli* K-12. J. Bacteriol. **151:** 468-471.

Li, Y., and T. Ferenci. 1996. The *Bacillus stearothermophilus* NUB36 *surA* gene encodes a thermophilic sucrase related to *Bacillus subtilis* SacA. Microbiology. **142:** 1651-1657.

Loesche, W.J. 1986. Role of *Streptococcus mutans* in human dental decay. Microbiol. Rev. **50:** 353-380.

Long, S. 1984. Studies on the regulation of solvent production and endospore formation in *Clostridium acetobutylicum* P262. PhD Thesis, University of Cape Town, South Africa.

Looten, P., D. Blanchet, and J.P. Vandecasteele. 1987. The β-fructofuranosidase activities of a strain of *Clostridium acetobutylicum* grown on inulin. Appl.. Microbiol. Biotechnol. **25:** 419-425.

Lunsford, R.D., and F.L. Macrina. 1986. Molecular cloning and characterization of *scrB*: the structural gene for the *Streptococcus mutans* PTS sucrose-6-phosphate hydrolase. J. Bacteriol. **166:** 426-434.

Marger, M.D., and M.H. Saier. 1993. A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. Trends Biochem. Sci. **18:** 13-20.

Martin-Verstraete, I., M. Debarbouille, A. Klier, and G. Rapoport. 1990. Levanase operon of *Bacillus subtilis* includes a fructose-specific phosphotransferase system regulating the expression of the operon. J. Mol. Biol. **214:** 657-671.

Martin-Verstraete, I., M. Debarbouille, A. Klier, and G. Rapoport. 1994. Interactions of wild-type and truncated LevR of *Bacillus subtilis* with the upstream activating sequence of the levanase operon. J. Mol. Biol. **241:** 178-192.

Martin-Verstraete, I., J. Stülke, A. Klier, and G. Rapoport. 1995. Two different mechanisms mediate catabolite repression of the *Bacillus subtilis* levanase operon. J. Bacteriol. **177:** 6919-6927.

McLaughlin, R.E., and J.J. Ferretti. 1996. The multiple-sugar metabolism (msm) gene cluster of *Streptococcus mutans* is transcribed as a single operon. FEMS Microbiol. Lett. **140:** 261-264.

McCoy, E., E.B.Fred, W.H. Peterson, and E.G. Hastings. 1926. A cultural study of the acetone butyl alcohol organism. J. Infect. Dis. **39:** 437-484.

Mermelstein, L.D., and E.T. Papoutsakis. 1993a. In vivo methylation in *Escherichia coli* by the *Bacillus subtilis* phage 3T I methyltransferase to protect plasmids from restriction upon transformation of *Clostridium acetobytylicum* ATCC 824. Appl. Environ. Microbiol. **59:** 1077-1081.

Mermelstein, L.D., and E.T. Papoutsakis. 1993b. Evaluation of macrolide and lincosamide antibiotics for plasmid maintenance in low pH *Clostridium acetobutylicum* ATCC824 fermentations. FEMS Microbiol. Lett. **113:** 71-76.

Miller, G. L. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal. Chem. **31:**426-428.

Minton, N.P., J.K.Brehm, J.D.Oultram, D.E. Thompson, T-J Swinfield, A. Pennock, S. Schimming, S.M.Whelan, U. Vetter, M. Young, and W.L. Staudenbauer. 1990. Vector systems for the genetic analysis of *Clostridium acetobutylicum*. p 187-201. *In* Borriello S.P. (ed.) Clinical and molecular aspects of anaerobes, Wrightson Biomedical Publishing Ltd.

Minton, N.P., J.K.Brehm, T-J Swinfield, S.M.Whelan, M.L.Mauchline, N. Bodsworth, and J.D.Oultram. 1993. Clostridial cloning vectors. p 119-150. *In* Woods D.R. (ed.) The clostridia and biotechnology, Butterworth-Heinemann, Stoneham, MA, USA.

Mitchell, W.J. 1996. Carbohydrate uptake and utilization by *Clostridium beijerinckii* NCIMB 8052. Anaerobe **2:** 379-384.

Mitchell, W.J., and I.R. Booth. 1984. Characterization of the *Clostridium pasteurianum* phosphotransferase system. J. Gen. Micro. **130:** 2193-2200.

Mitchell, W.J., K.A. Albasheri, and M. Yazadanian. 1995. Factors affecting utilization of carbohydrates by clostridia. FEMS Microbiol. Rev. **17:** 317-329.

Mitchell, W.J., J.E. Shaw, and L. Andrews. 1991. Properties of the glucose phosphotransferase system of *Clostridium acetobutylicum* NCIB 8052. Appl. Environ. Microbiol. **57:** 2534-2539.

Mortatte, M.P.L., H.H. Sato, and Y.K. Park. 1983. Induction and some characteristics of sucrose hydrolyzing enzymes from *Zymomonas mobilis*.. Biotechnol. Lett. **5:** 229-232.

Munro, C.L., S.M. Michalek, and F.L. Macrina. 1995. Sucrose-derived exopolymers have site-dependent roles in *Streptococcus mutans*-promoted dental decay. FEMS Microbiol. Lett. **128:** 327-332

Nguyen, C.C., and M.H. Saier, Jr. 1995. Phylogenetic, structural and functional analyses of the LacI-GalR family of bacterial transcription factors. FEBS Lett. **377:** 98-102.

O'Brien, R.W., and J.G. Morris. 1971. Oxygen and the growth and metabolism of *Clostridium acetobutylicum*. J. Gen. Micro. **68:** 307-318.

O'mullan, P., M. Szakacs-Dobozi, and D.E. Eveleigh. 1991. Identification of saccharolytic enzymes of *Zymomonas mobilis* CP4. Biotechnol. Lett. **13:** 137-142.

Oultram, J.D., I.D. Burr, M.J. Elmore, and N.P.Minton. 1993. Cloning and sequence analysis of the genes encoding phosphotransbutyrylase and butyrate kinase from *Clostridium acetobutylicum* NCIMB 8052. Gene **131:** 107-112.

Oultram, J.D., M. Loughlin, T-J. Swinfield, J.K. Brehm, D.E.Thompson, and N.P. Minton. 1988. Introduction of plasmids into whole cells of *Clostridium acetobutylicum* by electroporation. FEMS Microbiol. Lett. **56:** 83-88.

Park, Y.K., M.P.L. Mortatte, and H.H. Sato. 1983. Study of levan formation during fermentation of *Zymomonas mobilis* on sucrose. Biotechnol. Lett. **5:** 515-518.

Paquet, V., C. Croux, G. Goma and P. Soucaille. 1991. Purification and characterization of the extracellular α-amylase from *Clostridium acetobutylicum* ATCC 824. Appl. Environ. Microbiol. **57:** 212-218.

Postma, P.W., J.W. Lengeler, and G.R. Jacobson. 1993. Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. Microbiol. Rev. **57:** 543-594

Poy, F., and G.R. Jacobson. 1990. Evidence that a low-affinity sucrose phosphotransferase activity in *Streptococcus mutans* GS-5 is a high-affinity trehalose uptake system. Infect. Immun. **58:** 1479-1480.

Preziosi, L., G.P.F. Michel, and J. Baratti. 1990. Sucrose metabolism in *Zymomonas mobilis*: production and localization of sucrase and levansucrase activities. Can. J. Microbiol. **36:** 159-163.

Rauch, P. J., M. M. Beerthuyzen, and W.M. de Vos. 1994. Distribution and evolution of nisin-sucrose elements in *Lactococcus lactis*. Appl. Environ. Microbiol. **60:** 1798-1804.

Rauch, P.J., and W. M. de Vos. 1992a. Characterization of the novel nisin-sucrose conjugative transposon Tn5276 and its insertion in *Lactococcus lactis*. J. Bacteriol. **174:** 1280-1287.

Rauch, P. J., and W. M. de Vos. 1992b. Transcriptional regulation of the Tn5276-located *Lactococcus lactis* sucrose operon and characterization of the *sacA* gene encoding sucrose-6-phosphate hydrolase. Gene **121:** 55-61.

Reddy, A., and F. Maley. 1996. Studies on identifying the catalytic role of Glu-204 in the active site of yeast invertase. J. Biol. Chem. **271:** 13953-13957.

Rogers, P., and G. Gottschalk. 1993. Biochemistry and regulation of acid and solvent production in clostridia. p 25-50. *In* Woods D.R. (ed.) The clostridia and biotechnology, Butterworth-Heinemann, Stoneham, MA, USA.

Russell, R.R., H. Mukasa, A. Shimamura, and J.J. Ferretti. 1988. *Streptococcus mutans gtfA* gene specifies sucrose phosphorylase. Infect. Immun. **56:** 2763-2765.

Russell, R.R., J. Aduse-Opoku, I.C. Sutcliffe, L. Tao, and J.J. Ferretti. 1992. A binding protein-dependent transport system in *Streptococcus mutans* responsible for multiple sugar metabolism. J. Biol. Chem. **267:** 4631-4637.

Rutberg, B. 1997 Antitermination of transcription of catabolic operons. Mol. Microbiol. **23:** 413-421.

Sanger, F., S. Nicklen, and A.R. Coulson.. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. **74:** 5463-5467.

Saier, M.H., S. Chauvax, G.M. Cook, J. Deutscher, I. T. Paulsen, J. Reizer, and J-J. Ye. 1996. Catabolite repression and inducer control in gram-positive bacteria. Microbiology **142:** 217-230.

Saier, M.H., S. Chauvax, J. Deutscher, J. Reizer, and J-J. Ye. 1995. Protein phosphorylation and regulation of carbon metabolism in gram-negative versus gram positive bacteria. TIBS. **20:** 267-271.

Saier, M.H., and T.M. Ramseier. 1996. The Catabolite Repressor/Activator (Cra) Protein of enteric bacteria. J. Bacteriol. **178:** 3411-3417.

Saier, M.H., and J. Reizer. 1992. Proposed uniform nomenclature for the proteins and protein domains of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. J. Bacteriol. **174:** 1433-1438

Saier, M.H., and J. Reizer. 1994. The bacterial phosphotransferase system: new frontiers 30 years later. Mol. Microbiol. **13:** 755-764.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: A laboratory Manual 2 ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Sato, Y., and H.K. Kuramitsu. 1988. Sequence analysis of the *Streptococcus mutans scrB* gene. Infect. Immun. **56:** 1956-1960.

Sato, Y., F. Poy, G.R. Jacobson, and H.K. Kuramitsu. 1989. Characterization and sequence analysis of the *scrA* gene encoding enzyme II^{Scr} of the *Streptococcus mutans*

phosphoenolpyruvate-dependent sucrose phosphotransferase system. J. Bacteriol. **171:** 263-271.

Sato, Y., Y. Yamamoto, H. Kizaki, and H.K. Kuramitsu. 1993a. Isolation and sequence analysis of the *pmi* gene encoding phosphomannose isomerase of *Streptococcus mutans*. FEMS Microbiol. Lett. **114:** 61-66.

Sato, Y., Y. Yamamoto, H. Kizaki, and H.K. Kuramitsu. 1993b. Isolation, characterization and sequence analysis of the *scrK* gene encoding fructokinase of *Streptococcus mutans*. J. Gen. Microbiol. **139:** 921-927.

Sato, Y., Y. Yamamoto, R. Suzuki, H. Kizaki, and H.K. Kuramitsu. 1991. Construction of *scrA::lacZ* gene fusions to investigate regulation of the sucrose PTS of *Streptococcus mutans*. FEMS Microbiol. Lett. **63:** 339-345.

Schmid, K., R. Ebner, J. Altenbuchner, R. Schmitt, and J.W. Lengeler. 1988. Plasmid-mediated sucrose metabolism in *Escherichia coli* K12: mapping of the scr genes of pUR400. Mol. Microbiol. **2:** 1-8.

Schmid, K., R. Ebner, K. Jahreis, J.W. Lengeler, and F. Titgemeyer. 1991. A sugar-specific porin, ScrY, is involved in sucrose uptake in enteric bacteria. Mol. Microbiol. **5:** 941-950.

Schmid, K., M. Schupfner, and R. Schmitt. 1982. Plasmid-mediated uptake and metabolism of sucrose by *Escherichia coli* K- 12. J. Bacteriol. **151:** 68-76.

Scholle, R.R., V.E. Coyne, R. Maharaj, F.T. Robb, and D.R. Woods. 1987. Expression and regulation of a *Vibrio alginolyticus* sucrose utilization system cloned in *Escherichia coli*. J. Bacteriol. **169:** 2685-2690.

Scholle, R.R., S.M. Robb, F.T. Robb, and D.R. Woods. 1989. Nucleotide sequence and analysis of the *Vibrio alginolyticus* sucrase gene (*scrB*). Gene **80:** 49-56.

Schulein, K., K. Schmid, and R. Benzl. 1991. The sugar-specific outer membrane channel ScrY contains functional characteristics of general diffusion pores and substrate-specific porins. Mol. Microbiol. **5:** 2233-2241.

Schumacher, M.A., K.Y. Choi, H. Zalkin, and R.G. Brennan. 1994. Crystal structure of LacI member PurR, bound to DNA: minor groove binding by α helices. Science **266:** 763-769.

Shibata, Y., and H.K. Kuramitsu. 1996. Identification of the *Streptococcus mutans frp* gene as a potential regulator of fructosyltransferase expression. FEMS Microbiol. Lett. **140:** 49-54.

Shimotsu, H., and D.J. Henner. 1986. Modulation of *Bacillus subtilis* levansucrase gene expression by sucrose and regulation of the steady-state mRNA level by *sacU* and *sacQ* genes. J. Bacteriol. **168:** 380-388.

Shiroza, T., and H.K. Kuramitsu. 1988. Sequence analysis of the *Streptococcus mutans* fructosyltransferase gene and flanking regions. J. Bacteriol. **170:** 810-816.

Shiroza, T., S. Ueda, and H.K. Kuramitsu. 1987. Sequence analysis of the *gtfB* gene from *Streptococcus mutans.* J. Bacteriol. **169:** 4263-4270.

Smith, S.B., M.A. Taylor, L.R. Burch, and H.V. Davies. 1993. Primary structure and characterization of a cDNA clone of fructokinase from potato (*Solanum tuberosum* L. cv record). Plant Physiol. **102:** 1043.

Song, K.B., H.K. Joo, and S.K. Rhee. 1993. Nucleotide sequence of levansucrase gene (*levU*) of *Zymomonas mobilis* ZM1 (ATCC10988). Biochim. Biophys. Acta. **1173:** 320-324.

Song, K.B., S.K. Lee, H.K. Joo, and S.K. Rhee. 1994. Nucleotide and derived amino acid sequences of an extracellular sucrase gene (*invB*) of *Zymomonas mobilis* ZM1 (ATCC10988). Biochim. Biophys. Acta. **1219:** 163-166.

Sprenger, G.A., and J.W. Lengeler. 1988. Analysis of sucrose catabolism in *Klebsiella pneumoniae* and in Scr+ derivatives of *Escherichia coli* K12. J. Gen. Microbiol. **134:** 1635-1644.

Spychala, J., N.S. Datta, K. Takabayashi, M. Datta, I.H. Fox, T. Gribbin, and B.S. Mitchell. 1996. Cloning of human adenosine kinase cDNA: Sequence similarity to microbial ribokinases and fructokinases. Proc. Natl. Acad. Sci. **93:** 1232-1237.

Steinmetz, M. 1993. Carbohydrate catabolism: pathways, enzymes, genetic regulation, and evolution. p 157-170. *In* Sonenshein A.L. (ed.) *Bacillus subtilis* and other Gram-positive bacteria, American Society for Microbiology, Washington, D.C., USA.

Steinmetz, M., D. Le Coq, and S. Aymerich. 1989. Induction of saccharolytic enzymes by sucrose in *Bacillus subtilis*: evidence for two partially interchangable pathways. J. Bacteriol. **171:** 1519-1523.

Steinmetz, M., D. Le Coq, S. Aymerich, G. Gonzy-Treboul, and P. Gay. 1985. The DNA sequence of the gene for the secreted *Bacillus subtilis* enzyme levansucrase and its genetic control sites. Mol. Gen. Genet. **200:** 220-228.

Stephens, G.M., R.A. Holt, J.C. Gottschal, and J.G. Morris. 1985. Studies on the stability of solvent production by *Clostridium acetobutylicum* in continuous cultures. J. Appl. Bacteriol. **59:** 597-605.

St. Martin, E.J., and C.L. Wittenberger. 1979a. Regulation and function of sucrose 6-phosphate hydrolase in *Streptococcus mutans.* Infect. Immun. **26:** 487-491.

St. Martin, E.J., and C.L. Wittenberger. 1979b. Characterization of a phosphoenolpyruvate-dependent sucrose phosphotransferase system in *Streptococcus mutans.* Infect. Immun. **24:** 865-868.

Stülke, J., I. Martin-Verstraete, V. Charrier, A. Klier, J. Deutscher, and G. Rapoport. 1995. The HPr protein of the phosphotransferase system links induction and catabolite repression of the *Bacillus subtilis* levanase operon. J. Bacteriol. **177:** 6928-6936.

Sutrina, S.I., P. Reddy, M.H. Saier, and J. Reizer. 1990. The glucose permease of *Bacillus subtilis* is a single polypeptide chain that functions to energize the sucrose permease. J. Biol. Chem. **265:** 18581-18589.

Tam, R., and M.H. Saier. 1993. Structural, functional and evolutionary relationships among extracellular solute-binding receptors of bacteria. Microbiol. Rev. **57:** 320-346.

Tao, L., I.C. Sutcliffe, R.R. Russell, and J.J. Ferretti. 1993. Transport of sugars, including sucrose, by the *msm* transport system of *Streptococcus mutans*. J. Dent. Res. **72:** 1386-1390.

Thompson, J., and B. Chassy. 1981. Uptake and metabolism of sucrose by *Streptococcus lactis*. J. Bacteriol. **147:** 543-551.

Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Research. **22:** 4673-4680.

Thompson, J., N. Y. Nguyen, D. L. Sackett, and J. A. Donkersloot. 1991a. Transposon-encoded sucrose metabolism in *Lactococcus lactis*. Purification of sucrose-6-phosphate hydrolase and genetic linkage to N^5- (L-1-carboxyethyl)-L-ornithine synthase in strain K1. J. Biol. Chem. **266:** 14573-14579.

Thompson, J., D. L. Sackett, and J. A. Donkersloot. 1991b. Purification and properties of fructokinase I from *Lactococcus lactis*. Localization of *scrK* on the sucrose-nisin transposon Tn5306. J. Biol. Chem. **266:** 22626-22633.

Thompson, J., N.Y. Nguyen, and S.A. Robrish. 1992. Sucrose fermentation by *Fusobacterium mortiferum* ATCC25557: transport, catabolism and products. J. Bacteriol. **174:** 3227-3235

Titgemeyer, F., K. Jahreis, R. Ebner, and J.W. Lengeler. 1996. Molecular analysis of the *scrA* and *scrB* genes from *Klebsiella pneumoniae* and plasmid pUR400, which encode the sucrose transport protein Enzyme II^{Scr} of the phosphotransferase system and a sucrose-6-phosphate invertase. Mol. Gen. Genet. **250:** 197-206.

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

Tsunekawa, H., S. Azuma, M. Okabe, R. Okamoto, and S. Aiba. 1992. Acquisition of a sucrose utilization system in *Escherichia coli* K-12 derivatives and its application to industry. Appl. Environ. Microbiol. **58:** 2081-2088.

Ueda, S., T. Shiroza, and H.K. Kuramitsu. 1988. Sequence analysis of the *gtfC* gene from *Streptococcus mutans* GS-5. Gene **69:** 101-109.

Vartak, N.B., J. Reizer, A. Reizer, J.T. Gripp, E.A.Groisman, L.-F. Wu, J.M. Tomich, and M.H.Saier. 1991. Sequence and evolution of the FruR protein of *Salmonella typhimurium*: a pleiotropic transcriptional regulatory protein possessing both activator and

repressor functions which is homologous to the periplasmic ribose-binding protein. Res. Microbiol. **142:** 951-963.

Vellanoweth, R.L., and J.C. Rabinowitz. 1992. The influence of ribosome-binding-site elements on translation efficiency in *Bacillus subtilis* and *Escherichia coli in vivo*. Mol. Microbiol. **6:** 1105-1114.

Verhasselt, P., and J. Vanderleyden. 1993. Cloning and expression of *Clostridium acetobutylicum* genes involved in carbohydrate utilization. p.301-316. *In* Sebald, M. (ed.) Genetics and molecular biology of anaerobic bacteria. Springer-Verlag. N.Y.

Verhasselt, P., F. Poncelet, K. Vits, A. Van Gool, and J. Vanderleyden. 1989. Cloning and expression of a *Clostridium acetobutylicum* α-amylase gene in *Escherichia coli*. FEMS Microbiol. Lett. **50:** 135-140.

Wagner, E., F. Gotz and R. Brückner. 1993. Cloning and characterization of the *scrA* gene encoding the sucrose- specific Enzyme II of the phosphotransferase system from *Staphylococcus xylosus*. Mol. Gen. Genet. **241:** 33-41.

Wanker, E., A. Huber, and H. Schwab. 1995. Purification and characterization of the *Bacillus subtilis* levanase produced in *Escherichia coli*. Appl. Environ. Microbiol. **61:** 1953-1958.

Weickert, M.J., and S. Adhya. 1992. A family of bacterial regulators homologous to Gal and Lac repressors. J. Biol. Chem. **267:** 15869-15874.

Wexler, D., M.C. Hudson, and R.A. Burne. 1993. *Streptococcus mutans* fructosyltransferase (*ftf*) and glucosyltransferase (*gtfBC*) operon fusion strains in continuous culture. Infect. Immun. **61:** 1259-1267.

Wilkinson, S.R., and M. Young. 1993. Wide diversity of genome size among different strains of *Clostridium acetobutylicum*. J. Gen. Microbiol. **139:** 1069-1076.

Wilkinson, S.R., and M. Young. 1994. Targeted integration of genes into the *Clostridium acetobutylicum* chromosome. Microbiology. **140:** 89-95.

Wilkinson, S.R., and M. Young. 1995. Physical map of the *Clostridium beijerinckii* (formerly *Clostridium acetobutylicum*) NCIMB8052 chromosome. J. Bacteriol. **177:** 439-448.

Wilkinson, S.R., M. Young, R. Goodacre, J.G. Morris, J.A.E. Farrow, and M.D. Collins. 1995a. Phenotypic and genotypic differences between certain strains of *Clostridium acetobutylicum*. FEMS Microbiol. Lett. **125:**199-204.

Wilkinson, S.R., D. I. Young, J.G. Morris, and M. Young. 1995b. Molecular genetics and the initiation of solventogenesis in *Clostridium beijerinckii* (formerly *Clostridium acetobutylicum*) NCIMB 8052. FEMS Microbiol. Rev. **17:** 275-285

Williams, D.R., D.I. Young, J.D. Oultram, N.P. Minton, and M. Young. 1990a. Development and optimization of conjugative plasmid transfer from *Escherichia coli* to

Clostridium acetobutylicum NCIB 8052. p239-246. *In* Clinical and molecular aspects of anaerobes. Edited by S.P. Borriello. Wrightson Biomedical Publishing Ltd.

Williams, D.R. D.I. Young, and M. Young. 1990b. Conjugative plasmid transfer from *Escherichia coli* to *Clostridium acetobutylicum*, J. Gen. Micro. **136:** 819-826.

Woods, D.R. 1995. The genetic engineering of microbial solvent production. Trends Biotechnol. **13:** 259-264.

Woolley, R.C., and J.G. Morris. 1990. Stability of solvent production *by Clostridium acetobutylicum* in continuous culture: strain differences. J.Appl. Bacteriol. **69:** 718-728.

Wu, L-F, A. Reizer, J. Reizer, B. Cai, J.M. Tomich, and M.H. Saier. 1991. Nucleotide sequence of the *Rhodobacter capsulatus fruK* gene, which encodes fructose-1-phosphate kinase: Evidence for a kinase superfamily including both phosphofructokinases of *Escherichia coli*. J. Bacteriol. **173:** 3117-3127.

Yanase, H., H. Fukushi, N. Ueda, Y. Maeda, A. Toyoda, and K. Tonomura. 1991. Cloning, sequencing, and characterization of the intracellular invertase gene from *Zymomonas mobilis*. Agric. Biol. Chem. **55:** 1383-1390.

Yanase, H., M. Iwata, R. Nakahigashi, K. Kita, N. Kato, and K. Tonomura. 1992. Purification, crystalization and properties of the extracellular levansucrase from *Zymomonas mobilis*. Biosci. Biotechnol. Biochem. **56:** 1335-1337.

Yanisch-Perron, D., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:** 103-119.

Young, M., N.P. Minton, and W.L. Staudenbauer. 1989. Recent advances in the genetics of the clostridia. FEMS Microbiol Rev. **5:** 301-325.

Yoshida, K., H. Sano, Y. Miwa, N. Ogasawara, and Y. Fujita. 1994. Cloning and nucleotide sequencing of a 15 kb region of the *Bacillus subtilis* genome containing the iol operon. Microbiology. **140:** 2289-2298

Zabeau, M., and K.K. Stanely. 1982. Enhanced expression of Cro-β-galactosidase fusion proteins under the control of the P_R promoter of the bacteriophage lambda. EMBO J. **1:** 1217-1224.

Zembrzuski, B., P. Chilco, X.L. Liu, J. Liu, T. Conway, and R. Scopes. 1992. Cloning, sequencing, and expression of the *Zymomonas mobilis* fructokinase gene and structural comparison of the enzyme with other hexose kinases. J. Bacteriol. **174:** 3455-3460.

Zuker, M., and P. Stiegler. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxilary information. Nucleic Acids Res. **9:** 133-148.

Zukowski, M.M., L. Miller, P. Cosgwell, K. Chen, S. Aymerich, and M. Steinmetz. 1990. Nucleotide sequence of the *sacS* locus of *Bacillus subtilis* reveals the presence of two regulatory genes. Gene. **90:** 153-155.