Studies on the regulation of nitrogen metabolism in *Clostridium acetobutylicum* NCP 262 and *Clostridium beijerinckii* NCIMB 8052

Keith Windsor McRostie Quixley

University of Cape Town

A dissertation submitted in fulfilment of the requirements for the degree of Master of Science in the Department of Microbiology, Faculty of Science, University of Cape Town, South Africa

September, 1999

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

CONTENTS

Acknowledgements........ ii

Abbreviations........ iii

Abstract........ iv

CHAPTER 1 **1**

GENERAL INTRODUCTION AND LITERATURE REVIEW

CHAPTER 2 **38**

STUDIES ON PHYSIOLOGICAL AND GENETIC ASPECTS OF NITROGEN METABOLISM REGULATION IN *Clostridium acetobutylicum* NCP 262 AND *Clostridium beijerinckii* NCIMB 8052

CHAPTER 3 **77**

DEVELOPMENT OF A REPORTER SYSTEM FOR *Clostridium beijerinckii* NCIMB 8052

CHAPTER 4 **107**

GENERAL CONCLUSIONS

Appendices........ 110

References........ 144

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere thanks to my supervisor, Dr Shez Reid, for her advice, guidance and encouragement during the course of this work. I would also like to thank the members of the anaerobe unit for their great company and their willingness to provide assistance whenever it was needed: Dr. Val Abratt, Suhail, Nina, Joy, Anton, Andisa, Wata, Careena, Ana, Garth, Marla and Kevin. Thank-you also to ex-anaerobes whose help has been invaluable: Neil and Helen. Thanks to Dr. Sagadevan Mundree for his help and advice, especially with the RNA work. To Di James, a big thank-you for help with DNA sequencing, PCR and primer extensions. Thank-you to Dr Paul Meyers for proofreading this dissertation.

Thank-you to Mrs Audrey Quixley and Mr Alan Quixley, and to Derek and Celia Walford for their moral support and encouragement, across the kilometres. Finally, to Sally-Ann Walford, the light of my life, a huge thank-you for all the support, encouragement, and cups of tea.

I acknowledge the financial support provided by the FRD and the Microbiology Department.

ABBREVIATIONS

A	Adenine	mg	milligram(s)
ABE	acetone-butanol-ethanol	ml	millilitre(s)
ADP	adenosine diphosphate	mRNA	messenger RNA
Ap	ampicillin	N	molal concentration
Ap^R	ampicillin resistance	NAD	nicotinamide adenine
ATP	adenosine triphosphate	NADP	dinucleotide
bp	base pairs		nicotinamide adenine
C	cytosine	nm	dinucleotide phosphate
CBM	clostridial basal medium	nt	nannometre(s)
cDNA	copy DNA	ORF	nucleotide(s)
CMC	carboxymethyl cellulose	OD	open reading frame
CsCl	caesium chloride	*p*	optical density
DMSO	dimethyl sulphoxide	PC	promoter
DNA	deoxribonucleic acid	PCR	phosphate-citrate(buffer)
DNAse	deoxyribonuclease	PEG	polymerase chain reaction
EDTA	ethylenediaminetetra-acetic acid	RBS	polyethylene glycol
Em	erythromycin	RNA	ribosome binding site
Em^R	erythromycin resistance	RNase	ribonucleic acid
EtBr	ethidium bromide	rpm	ribonucelase
EtOH	ethanol	σ	revolutions per minute
g	gram(s)	SD	sigma
G	guanine	SDS	Shine-Dalgarno sequence
GDH	glutamate dehydrogenase	T	sodium dodecyl sulphate
gln	glutamine	TA	thymine
glu	glutamate	TAE	tris-acetate buffer
GOGAT	glutamine oxoglutarate amido	Tris	tris-acetate EDTA buffer
	transferase		tris(hydroxymethyl)
GS	glutamine synthetase	tRNA	aminomethane
GSMM	glucose-salts minimal medium	Trp	transfer RNA
H_2	molecular hydrogen	Tyr	tryptophan
H^+	ionic hydrogen	Tween	tyrosine polyoxyethylene
IPTG	isopropyl-β-D-thiogalactopyranoside	U	sorbitan monolaurate
kb	kilobase(s)	U	uracil
kDa	kilodalton(s)	UTP	unit(s) of enzyme activity
λ	lambda	UV	uracil triphosphate
l	litre(s)	w/v	ultraviolet
LA	Luria-Bertani agar	X-gal	weight per volume
LB	Luria-Bertani broth		5-bromo-4-chloro-3-indolyl-β-
M	molar concentration	YT	D-galactopyranoside
min	minute(s)		yeast-tryptone broth

ABSTRACT

In order to produce *Clostridium acetobutylicum* strains with improved solventogenic capabilities through genetic manipulation, a thorough understanding of the mechanisms governing solvent production is required. Nitrogen metabolism has been shown to be important in the regulation of sporulation and solventogenesis, and the enzyme glutamine synthetase, encoded by the gene *glnA*, is central to this aspect of metabolism. This work represents a continuation of studies on the glutamine synthetase enzyme of *C. acetobutylicum* NCP 262. Previous work identified a gene, *glnR*, which is thought to encode an antiterminator protein involved in the regulation of *glnA* transcription. An attempt was made to demonstrate a regulatory role for *glnR* in a system using the cloned *C. acetobutylicum* NCP 262 *glnA* with and without *glnR* in the *E. coli glnAntrBC* mutant YMC11. Although the enzyme was efficiently expressed and fully functional, no evidence could be found of any influence of *glnR* on GS activity in the heterologous host. In an attempt to locate other genes involved in *glnA* regulation the region upstream of *glnA* in *C. acetobutylicum* NCP 262 was sequenced. Although no aceesory regulatory genes were found, an incomplete open reading frame encoding a putative aspartokinase or bifunctional aspartokinase/homoserine dehydrogenase was identified. As attempts to reconstitute the *C. acetobutylicum* NCP 262 *glnA* regulatory system in *E. coli* were unsuccessful, it was necessary to find an alternative to heterologous host. *C. beijerinckii* NCIMB 8052, which is closely related to *C. acetobutylicum* NCP 262, was evaluated as a model organism for the study of nitrogen metabolism in the latter. It was found that organic nitrogen in the form of casamino acids was preferred over ammonium, and that GS activity was induced under conditions of nitrogen limitation, and repressed in cultures grown in the presence of high concentration of casamino acids. Southern hybridisation experiments identified homologues of *glnA*, *glnR* and *gltA* in *C. beijerinckii* NCIMB 8052, and a clone was isolated from a partial gene library of this organism which complemented glutamine auxotrophy in *E. coli* YMC11. The clone was sequenced and found to carry two complete and one incomplete ORF's which shared a high degree of nucleotide sequence similarity with *glnA*, *glnR* and *gltA* genes of *C. acetobutylicum* NCP 262 (87.8%, 86.5.% and 86.8% respectively). In addition, the relative arrangement of the genes was similar. Primer extension experiments identified four transcriptional start sites, two of which corresponded approximately to those previously identified in *C. acetobutylicum* NCP 262.

Transcription from all four was found to increase under conditions of nitrogen limitation. Leaders of three of the transcripts were found to be capable of forming identical terminator structures upstream of the start codon of *glnA*, however potential antiterminator structures differed for each transcript. A reporter system was developed for use in *C. beijerinckii* NCIMB 8052, based on the *eglA* gene cloned from *C. acetobutylicum* NCP 262, cloned into the *Bacillus/Clostridium* shuttle vector pFNK1. When *eglA* was under the control of its own promoter, the endoglucanase activity was expressed and efficiently secreted into the culture medium by *C. beijerinckii* NCIMB 8052. The pH and temperature optima of the enzyme activity were found to be between 5.0 and 6.0 and 45^{0}C respectively. These parameters differed from those previously reported for an endoglucanase activity in *C. acetobutylicum* NCP 262. Expression of *eglA* appeared to be repressed by cellobiose and fructose, but not by glucose, sucrose or galactose. A promoterless *eglA* was placed under the control of *glnA* and *scr* promoters. Very little *eglA* expression was observed for the *scr* promoter-*eglA* in any of the above-mentioned carbon sources, indicating either a disruption of normal regulation in the cell, or the absence of essential regions on the cloned promoter fragment. The *glnA* promoter-*eglA* fusion produced large amounts of endoglucanase activity under conditions of nitrogen excess and limitation, indicating the absence of a negative regulatory factor. GS activity in the same cultures was regulated normally, indicating that the loss of regulation was most likely to be due to the absence of an essential factor on the fusion construct. It was shown, nevertheless, that efficient expression of the reporter could be obtained under the control of a heterologous promoter. Potential problems with the reporter system were discussed and suggestions were made for improvement of the system.

CHAPTER 1

GENERAL INTRODUCTION AND LITERATURE REVIEW

CONTENTS

1.1 INTRODUCTION **1**
1.1.1 The physiology of ABE fermentation 2
1.1.2 The role if Nitrogen in solventogenesis and sporulation 3
1.2 NITROGEN ASSIMILATION **4**
1.2.1 The enzymes of nitrogen assimilation 4
1.2.2 Glutamine synthetase 5
1.2.2.1 Classification 5
1.2.2.2 GSI structure 5
1.2.2.3 Regulation of GSI activity 6
1.2.3 Regulation of nitrogen metabolism in different bacteria 7
1.2.3.1 Regulation in the Enterobacteriaceae: The *ntr* system 7
1.2.3.2 Regulation in *B. subtilis* 10
1.2.3.3 Regulation in *C. acetobutylicum* NCP 262 13
1.3 TRANSCRIPTIONAL ANTITERMINATION **16**
1.3.1 Processive antitermination: The λ paradigm 16
1.3.2 Nonprocessive antitermination 18
1.3.2.1 Translation of a regulatory peptide 18
1.3.2.2 Control of termination by tRNA 20
1.3.2.3 Control of termination by RNA-binding proteins 23
1.4 AIMS OF THIS DISSERTATION **36**

1.1 INTRODUCTION

The genus *Clostridium* comprises Gram-positive, obligately anaerobic, rod-shaped, spore-forming bacteria. This designation does not reflect a valid taxonomic grouping and many studies have been conducted to refine our understanding of the phylogenetic relationships within this very diverse group of organisms. This situation is the same within the species *C. acetobutylicum*, and reclassification of the many strains bearing this name has resulted in their being divided into four distinct taxonomic groups (Keis et al., 1995). Consequently, *C. acetobutylicum* NCIMB 8052 being renamed *C. beijerinckii* NCIMB 8052 (Keis et al., 1995), while *C. acetobutylicum* NCP 262 is currently undergoing reclassification as *C.*

saccharobutylicum NCP 262. For many years after their discovery at the beginning of this century, solventogenic clostridia were used extensively in the production of the industrially important chemicals butanol and acetone. The advent of petrochemical synthesis of these substances, as well as the increasing cost of substrate for their production, has led to the decline and eventual demise of this process in the Western world. The recent increase in concern for the environment and for the depletion of fossil fuel resources has led to a resurgence in interest in the biological production of acetone and butanol. With molecular biological tools for genetically manipulating *C. acetobutylicum* becoming available, much work has been carried out with the aim of producing strains with enhanced solventogenic capabilities. Such strains could conceivably produce acetone and butanol in quantities which would make the process economically feasible once again (Woods, 1995).

1.1.1 The physiology of ABE fermentation

The fermentation of acetone, butanol and ethanol has been extensively characterised and reviewed (Jones & Woods, 1986; Jones & Woods, 1989; Bennett & Petersen, 1993). Glycolysis results in the production of pyruvate, which is catabolised to acetyl-CoA and either carbon dioxide or formate, depending on the organism. Acetyl-CoA is an essential intermediate and is the starting point of the pathways in acidogenic and solventogenic phases of metabolism which eventually lead to solvent production. During the acidogenic phase, acetyl-CoA is converted to acetate via the intermediate acetylphosphate, using the enzymes phosphotransacetylase and acetate kinase, respectively, and to butyrate via acetoacetyl-CoA. The conversion of acetyl-CoA to acetoacetyl-CoA is catalysed by acetyl-CoA acetyltransferase, and this intermediate is central to the ultimate production of acetone and butanol. During solventogenesis, ethanol is also produced from acetyl-CoA via acetaldehyde by the enzymes acetaldehyde and ethanol dehydrogenase, respectively.

Acetoacetyl-CoA is reduced to 3-hydroxybutyryl-CoA by 3-hydroxybutyryl-CoA dehydrogenase, then to crotonoyl-CoA through the action of crotonase, and then to butyryl-CoA by butyryl-CoA dehydrogenase. During acidogenesis, butyryl-CoA is phosphorylated by phosphotransbutyrylase and then hydrolysed to butyrate by butyrate kinase. Butyrate production proceeds simultaneously with acetate production.

The transition from acidogenic to solventogenic phase involves a switch in the carbon and electron flow from the acid-producing pathways to those responsible for producing solvents. The nature of the signal which triggers this shift appears to be related to both the pH of the culture and the concentration of acidic fermentation products. The transition also appears to be correlated with the onset of sporulation.

During solventogenesis, much of the acetate and butyrate produced during acidogenesis is retrieved from the medium and further metabolised by the cell to acetone and butanol, respectively. Acetone production requires the generation of acetoacetate which results from the uptake of acetate and butyrate from the medium, a process catalysed by acetyl-CoA:acetate/butyrate:CoA (CoAT) transferase. Acetoacetate is then converted to acetone by acetoacetate decarboxylase. Butyryl-CoA produced by the CoAT-dependent uptake reaction is converted to butanol, via butyraldehyde, by the enzymes butyraldehyde and butanol dehydrogenase.

1.1.2 The role of Nitrogen in solventogenesis and sporulation

In contrast to the well-characterised *Bacillus subtilis*, *C. acetobutylicum* will not sporulate in response to limitation of carbon or nitrogen sources. Instead, the cues for sporulation include other growth-limiting factors such as low phosphate levels, exposure to oxygen, excessive acid fermentation products and temperature changes. Sporulation will only occur if one or other of these stresses are applied in the presence of nitrogen and carbon excess (Long et al., 1984).

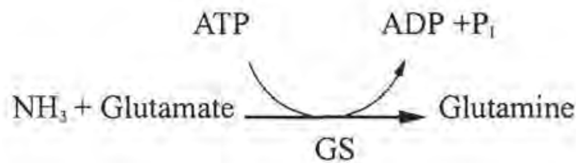
In *B. subtilis*, the protein TnrA, which is regulated by nitrogen availability, provides a link between nitrogen availability and sporulation, as it regulates an inhibitor of a protein kinase involved in sporulation induction (see section 1.2.3.2). In *C. acetobutylicum*, it has been shown that glutamine synthetase activity and the amount of nitrogen available to the cell plays a role in the induction of both sporulation and solventogenesis (Long et al., 1984). It is for this reason that the regulation of nitrogen metabolism in *C. acetobutylicum* has become the subject of extensive study, in addition to the work which has been aimed at the characterisation and understanding of the solvent fermentation pathways.

1.2 NITROGEN ASSIMILATION

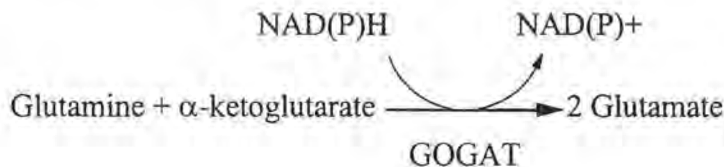
1.2.1 The enzymes of nitrogen assimilation

Although a wide variety of nitrogen-containing compounds are metabolised by different bacterial species, and the preferred nitrogen sources differ between species, the preferred source of nitrogen in the majority of bacteria studied so far is ammonia. Under conditions in which nitrogen is limiting, ammonia is metabolised to glutamine and glutamate. In most species these two compounds constitute the most important route for nitrogen assimilation, and in some it is the only route (Merrick & Edwards, 1995).

The enzyme which catalyses the formation of glutamine from ammonium and glutamate is glutamine synthetase (GS). In all species studied so far, it appears as if this enzyme is the only one capable of synthesising glutamine, and is thus one of the most important enzymes in nitrogen metabolism. The reaction is ATP-dependent and proceeds as follows:

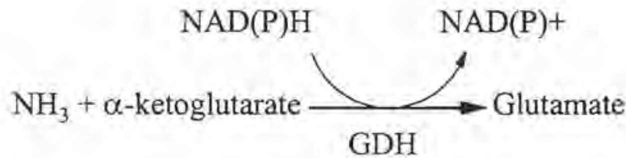


Glutamate synthase, previously known as **G**lutamine **O**xo**G**lutarate **A**mide **T**ransferase (GOGAT) produces 2 molecules of glutamate from a molecule of glutamine and a molecule of α-ketoglutarate in the following reaction:



In addition to the reactions described above, another common pathway exists for the synthesis of glutamate. It is catalysed by the enzyme glutamate dehydrogenase (GDH). This reaction allows ammonium to be incorporated directly into glutamate without the need for glutamine as an intermediate, and is more energy-efficient as it does not require ATP. It therefore has advantages over the previous pathways, however the K_m of glutamate dehydrogenase for glutamate is high, and the enzyme is therefore only effective at ammonium concentrations

higher than 1mM. Consequently, bacteria possessing this enzyme tend only to employ it when nitrogen is plentiful (Merrick & Edwards, 1995). The reaction it catalyses proceeds as follows:



1.2.2 Glutamine synthetase

1.2.2.1 Classification

Glutamine synthetase enzymes can be classified on the basis of their structure. It was long thought that there were two types of GS. GSI was thought to be the prokaryotic form, while GSII was believed to be its eukaryotic counterpart . It was subsequently established that GSI and GSII coexist in several bacterial species, and an additional form, GSIII, which is structurally distinct from both GSI and II has been found in *Bacteroides fragilis* (Hill et al., 1989) and *Butyrivibrio fibrisolvens* (Goodman et al., 1993). A fourth form has been isolated from *Rhizobium leguminosarum* (Chiurazzi et al., 1992) and *R. meliloti* (Shatters et al., 1993). Further subdivision of GSI-type enzymes can also be made on the basis of posttranslational modification, but this will be dealt with later in this chapter.

1.2.2.2 GSI structure

GSI enzymes have been isolated from several bacterial species, including the Enterobacteriaceae, *Vibrio alginolyticus*, *Thiobacillus ferrooxidans*, *Streptomyces cattleya*, *S. coelicolor*, and *B. subtilis* (Woods & Reid, 1993). GSI enzymes consist of 12 identical subunits arranged in two rings one on top of the other in a barrel-like arrangement. The entire structure is stabilized by both hydrophobic interactions and hydrogen bonding between the monomers. The subunits are encoded by the *glnA* gene, are between 444 and 447 amino acids in length, and have an average molecular weight of 55 kDa (Woods & Reid, 1993; Merrick & Edwards, 1995).

1.2.2.3 Regulation of GSI activity

As was stated earlier, a distinction can be made between different types of GSI enzymes on the basis of posttranslational modification. GSI enzymes of the Enterobacteriaceae, among others, are subject to modification by adenylylation of the tyrosine residue at position 401 of each subunit. Adenylylation of each subunit is non-cooperative and results in the inactivation of that subunit. The dodecameric enzyme is therefore able to exist in a wide range of intermediate activation states, depending on the amount of available nitrogen (Woods & Reid, 1993; Merrick & Edwards, 1995). The control of GSI adenylylation will be discussed in detail in section 1.2.3.1. In *C. acetobutylicum,* as well as in *B. subtilis*, no evidence exists of any posttranslational modification of GS under any nitrogen conditions tested so far (Janssen et al., 1988; Woods & Reid, 1993).

Interaction with an inactivating protein: a novel mechanism of GSI regulation

In cultures of *Synechocystis* PCC 6803 grown under conditions of nitrogen limitation, it has been observed that GS activity declines sharply after the addition of ammonium (Merida et al., 1991). It was subsequently demonstrated that in non-denaturing gel electrophoresis, the mobility of GS was retarded in ammonium-treated crude extracts of cells grown under ammonium-limiting conditions. The degree of retardation was indicative of a protein rather than a metabolite associating with the enzyme. Furthermore, only non-retarded bands displayed GS activity, indicating that the binding factor was also responsible for inactivation (Reyes & Florencio, 1995). Cross-linking studies identified two different-sized factors bound to individual GS. GS reactivation and simultaneous dissocation of the inactivating factor was observed in extracts whose pH was increased. Treatment with alkaline phosphatase also resulted in reactivation, suggesting that dephosphorylation of the binding factor was able to modulate its activity (Reyes & Florencio, 1995).

It was found that two polypeptides, IF7 (7kDa) and IF17 (17kDa), co-purified with the inactivated enzyme. N-terminal sequencing and analysis of the genome of the organism identified two ORF's, which were designated *gifA* (IF7) and *gifB* (IF17) (Garcia-Dominguez et al., 1999). Mutation of each of these genes resulted in impaired GS inactivation, and in a double mutant, inactivation was abolished (Reyes et al., 1999). *In vitro* studies showed that IF17 has a higher affinity for GS than IF7, and that the former binds co-operatively and is

able to displace the latter. Different degrees of retardation of inactivated GS suggested that intermediate states of GS activity could be achieved by different numbers of IF17 binding to the enzyme.

Northern blot analysis revealed that expression if *gifA* and *gifB* increases transiently after ammonium treatment (Reyes et al. 1999). These observations provided evidence for a model whereby the amount of IF7 and IF17, as well as their activity, are able to modulate the activity of GS rapidly in response to changes in ammonium availability. It also appears that this mechanism is fairly widely distributed amongst the cyanobacteria, with the relatively distantly related *Anabaena* sp. PCC 7120 possessing a protein homologous to IF7 and IF17 (Reyes et al., 1999).

1.2.3 Regulation of nitrogen metabolism in different bacteria

1.2.3.1 Regulation in the Enterobacteriaceae: The *ntr* system

The control of nitrogen metabolism has been most investigated and best characterised in the Enterobacteriaceae, and has been the subject of numerous reviews (Merrick & Edwards, 1995; Magasanik, 1996; Reitzer, 1996). Control in response to the nitrogen status of the cell is effected at both transcriptional and post-translational levels. This regulatory system is summarized in Figure 1.1.

Control of GS activity

As was stated earlier, the activity of GS in enteric bacteria is inversely proportional to the number of adenylylated subunits. Adenylylation and deadenylylation are catalysed by a bifunctional adenylyltransferase/deadenylylating enzyme. The decision between adenylyltransferase or deadenylylation activity is determined by the uridylylation state of the product of *glnB*, P_{II}. The uridylylation state of P_{II} is in turn controlled by the *glnD* product, which encodes an enzyme which has uridylyltransferase (UT) and deuridylylation (UR) activities depending on the nitrogen status of the cell. Nitrogen status is indicated by the ratio of glutamine to α-ketoglutarate, with nitrogen sufficiency indicated by a high glutamine :α-ketoglutarate ratio, and deficiency reflected in a low ratio. Under conditions of nitrogen excess, glutamine binds to the UR/UTase and stimulates the UR activity. P_{II} is therefore

deuridylylated and in this form stimulates the adenylylation and therefore the inactivation of GS. Under conditions of nitrogen limitation, α-ketoglutarate binds to P_{II} inducing a conformational change which results in an optimal configuration for its uridylylation. The UTase activity of the *glnD* product is thus stimulated, P_{II} is uridylylated and in turn stimulates the deadenylylation and hence the activation of GS.

Control of GS expression

The *ntrBC* system is responsible for the control of GS transcription. The *glnAntrBC* (also known as the *glnALG*) operon consists of the *glnA* gene, which encodes GS, and the downstream genes *ntrB* and *C* (or *glnL* and *G*), which encode the two components of the two-component *ntr* regulatory pathway. *ntrB* encodes NtrB (or NR_{II}), a histidine kinase sensor protein which phosphorylates or dephosphorylates the *ntrC* product, NtrC (or NR_{I}) and hence determines its activity.

Transcription of the *glnA* gene is initiated at one of two upstream promoters. The first of these, *glnAp1* is a σ^{70}-dependent promoter and directs low-level constitutive expression. Downstream of this sequence is *glnAp2* which directs σ^{54}-dependent transcription. The *ntrB* and *C* genes are located downstream of *glnA*, and are cotranscribed from the σ^{54}-dependent promoter *ntrBp1*. Between *glnA* and *ntrB* is a factor-independent transcription terminator which terminates approximately 75% of all transcripts initiated at the *glnA* promoters.

The signal to which NtrB responds is the uridylyation state of P_{II}. Uridylyated P_{II}, indicative of nitrogen limitation, binds to NtrB, causing it first to autophosphorylate a histidine residue at position 139 in an ATP-dependent reaction, and then to activate NtrC by phosphorylating it at an aspartate residue. Activated NtrC stimulates transcription from *glnAp2* and inhibits transcription from *glnAp1*. *glnAp2*-dependent transcription is at a considerably higher level than that from *glnAp1*, and the level of *glnA* expression is thus increased. Transcription from *ntrBp1* is also stimulated. Under conditions of nitrogen excess, deuridylylated P_{II} dissociates from NtrB, bringing about the dephosphorylation and hence the inactivation of NtrC, which in turn causes the repression of *glnA* and *ntrB* and *C* transcription.

Regulation by *nac*

In the enteric bacteria, several other genes involved in nitrogen metabolism are subject to control by the *ntr* system, for example the proline oxidase (*put*), urease (*ure*), GDH and GOGAT genes (Merrick & Edwards, 1995). In *Klebsiella aerogenes*, these genes are regulated by the protein Nac, a transcriptional regulator of the LysR family. Transcription of *nac* is subject to control by NtrC, and is therefore stimulated under conditions of nitrogen limitation. Nac acts to stimulate transcription of *hut*, *put* and *ure*, while repressing the transcription of *gdh* and the genes encoding the subunits of GOGAT. In *E.coli*, a *nac* homologue is thought to be involved in the control of asparagine synthesis (*asnAC*) and cytosine utilization (*codBA*) (Merrick & Edwards, 1995).

1.2.3.2 Regulation in *B. subtilis*

Among the Gram positive bacteria, the best understood in terms of nitrogen metabolism is *B. subtilis*. Although not as well characterised as the systems governing nitrogen metabolism in the enteric bacteria, the control systems operating in *B. subtilis* have nevertheless been the subject of considerable work, and have been reviewed by Schreier (1992), among others. Regulation on nitrogen metabolism in *B. subtilis* differs from that of the enteric bacteria in several ways. In *B. subtilis*, no assimilatory glutamate dehydrogenase activity has been identified to date. It appears, therefore, that the GS-GOGAT pathway is the only route by which ammonium is assimilated by the cell. The activity of the GS enzyme does not appear to be modulated by any post-translational mechanisms, although glutamine repressed its activity through a mechanism of feedback inhibition (Duel & Prusiner, 1974). There is no evidence of any system analogous to the enteric *ntr* system. The system is summarised in Figure 1.2.

Structure and Control of the *glnRA* Operon

The *glnRA* operon consists of two genes which are cotranscribed from a vegetative σ^A-dependent promoter (Schreier et al., 1989): The *glnA* gene encodes the GS enzyme, the GlnR protein is encoded by *glnR*, and is a member of the merR family of DNA-binding proteins. GlnR is responsible for negatively regulating transcription from the *glnRA* promoter under conditions of nitrogen excess. GlnR binds co-operatively to two operator sequences:

The first, *glnRA* O_1, is located upstream of the -35 region of the promoter, while *glnRA* O_2 overlaps the -35 region (Gutowski & Schreier, 1992).

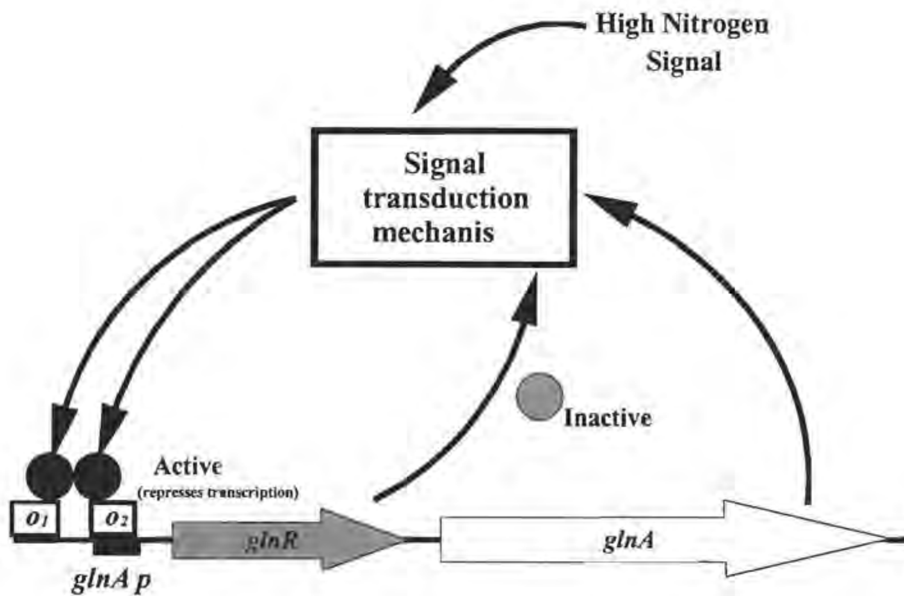


Figure 1.2: Regulation of *glnA* expression in *B. subtilis*. GlnR, activated by an as yet unknown mechanism, represses transcription from the *glnA* promoter (*glnAp*) by binding to operator sited *o1* and *o2* (Adapted from Fisher, 1999)

The cue to which GlnR responds remains unidentified (Fisher, 1999). It has been shown that in mutants of *B. subtilis* in which *glnA*, but not *glnR*, has been disrupted, GlnR-mediated repression of transcription under conditions of nitrogen excess is affected (Dean et al., 1997, Schreier et al., 1985). This indicates that GS is in someway required for regulation of its own synthesis. It is possible that GS modulates GlnR activity synthesising a regulatory ligand. This is unlikely to be glutamine alone, however, as it was observed that in a *glnA* mutant in which the intracellular glutamine concentration was six fold higher than that of the wild type, GlnR-mediated repression of transcription was not observed under conditions of nitrogen excess (Schreier & Sonenshein, 1986). Additionally, in vitro studies showed that DNA binding of GlnR was not affected by the presence of glutamine (Nakano & Kimura, 1991; Brown & Sonenshein, 1996). There are several lines of evidence that GS mediates GlnR activity through a direct interaction between the proteins: In experiments conducted using the cloned *glnR* and *glnA* in *E. coli*, it was found that transcription from the (Nakano & Kimura, 1991; Brown & Sonenshein, 1996). These observations, together with the feedback

inhibition of GS by glutamine raises an interesting possibility: It is conceivable that elevated concentrations of glutamine result in some change in the GS protein which allows it to relay a signal to GlnR to activate it.

Global nitrogen regulation in *B. subtilis*

It was long thought that GlnR was a pathway-specific regulator of the *glnRA* operon, and that there was no global nitrogen-regulatory system present in *B. subtilis* (Fisher, 1999). It has been discovered, however, that in addition to regulating transcription of the *glnRA* operon, GlnR is responsible for the repression of transcription of the *tnrA* and *ureABC* P3 promoters. The gene *tnrA* encodes another global nitrogen regulatory protein, TnrA (Wray et al., 1996), while the *ureABC* operon encodes a urease catabolic operon.

TnrA, is a protein which shares a high degree of amino acid sequence similarity with GlnR, particularly in its proposed N-terminal DNA-binding domain. It is responsible for positively regulating transcription of *gabP* (γ-aminobutyric acid permease), *ureABC* (urease), *nrgAB* (putative ammonium permease) and *nasABCDEF* (assimilatory nitrate and nitrite uptake system) under conditions of nitrogen limitation (Atkinson & Fisher, 1991; Wray & Fisher, 1994; Nakano et al., 1995, Nakano et al., 1998; Ferson et al., 1996). In addition to these nitrogen-related genes, TnrA has also been observed to increase expression of the sporulation kinase inhibitor KipI (Wang et al., 1997), as well as stimulate its own expression (Fisher, 1999). TnrA binds to an operator sequence almost identical to that of GlnR, and it has been proposed to exert its effect by stimulating transcription in a transcriptional enhancer-like manner (Wray et al., 1996; Wray et al., 1998).

It has been observed that TnrA negatively regulates *glnRA* transcription by binding to the operator which overlaps the -35 region of the promoter (Fisher, 1999). Although transcription of *gltAB*, which encodes the large and small subunits of glutamate synthase (GOGAT), respectively, is positively regulated by the GltC protein (Bohannon & Sonenshein, 1989), there appears to be an additional, independent control pathway facilitated by TnrA. However, no TnrA binding sites have been identified, so it appears that TnrA exerts its effect indirectly (Fisher, 1999). TnrA itself is repressed under conditions under which GlnR is active, namely nitrogen excess. The sensitivity of TnrA-mediated repression

of *glnRA* transcription to nitrogen availability is not as high as that of GlnR-mediated repression. TnrA may therefore provide a measure of fine-tuning of expression under conditions where GlnR is inactive. It is uncertain, however, whether the nature of the signal to which these proteins respond is the same (Fisher, 1999).

Although GlnR and TnrA regulate the expression of several genes involved in amino acid catabolism, utilisation of amino acids does not seem to be under the control of any single system analogous to the enteric *ntr/nac* system. The regulatory protein CodY responds to an as yet unidentified signal to repress several genes involved in amino acid utilisation (Slack et al., 1995; Ferson et al., 1996; Fisher et al., 1996; Wray et al., 1997), as well as some genes with functions unrelated to nitrogen metabolism (Serror & Sonenshein, 1996) under conditions of growth in carbon- and nitrogen-rich medium (Fisher et al., 1996). Other amino acid utilisation genes are subject to regulation dependent on modulation of transcription from σ^{54}-dependent promoters by transcriptional activators such as RocR, which regulates arginine and ornithine utilization (Garden et al., 1997), BkdR, which is involved in regulating isoleucine and valine utilisation (Fisher, 1999) and AhrC, which also regulates arginine and ornithine utilisation (Klingelt et al., 1995; Garda et al., 1997).

1.2.3.3 Regulation in *C. acetobutylicum* NCP 262

Studies on the regulation of nitrogen metabolism in *C. acetobutylicum* NCP 262 suggest that the processes in operation are markedly different to those of the well-characterized organisms discussed above. The *C. acetobutylicum glnA* has been cloned and expressed in *E. coli* (Usdin et al., 1986). Although initial experiments indicated that in the heterologous host, the enzyme was subject to similar regulation in response to nitrogen availability as in *C. acetobutylicum* (Usdin et al., 1986; Janssen et al., 1988), later work failed to confirm these findings (Fierro-Monti, Ph.D thesis)

Antisense RNA

Detailed study of the cloned DNA resulted in the identification of two promoters upstream of *glnA* (*p1* and *p2*), as well as a third promoter, *p3*, located downstream of, and oriented in the opposite direction to *glnA*. Transcription from *p3* results in the production of an oligoribonucleotide of 43 nucleotides, which exhibits sequence complementarity to the leader

region of the *glnA* mRNA transcript, including the ribosome binding site (Janssen et al., 1988). It has been postulated that this antisense RNA negatively regulates the expression of *glnA* at the level of translation by binding to the leader region of the nascent transcript and thereby preventing ribosomal binding (See Fig. 1.3). This model has been supported by evidence that in *E. coli*, expression of the cloned *C. acetobutylicum* NCP 262 *glnA* is suppressed in the presence of the antisense transcript (Janssen et al., 1990). In *C. acetobutylicum*, but not in *E. coli*, this antisense RNA has been found to be regulated in response to nitrogen (Fierro-Monti et al., 1992).

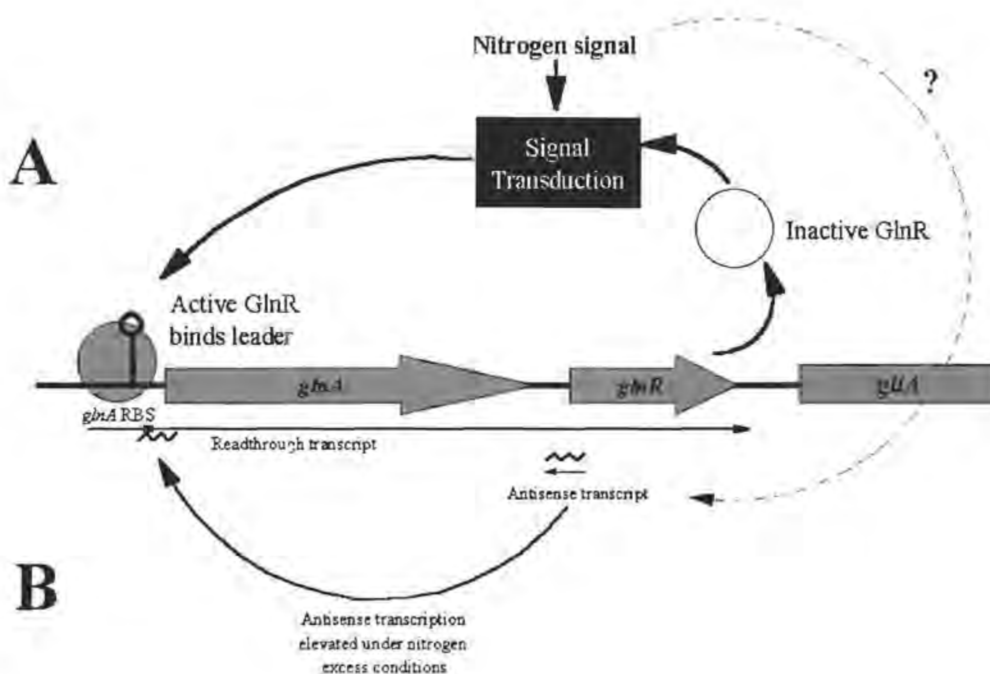


Figure 1.3: The proposed model of regulation of *glnA* expression in *C. acetobutylicum* NCP 262 (from Woods & Reid, 1995) (A) GlnR, a putative antiterminator protein, regulates transcription of *glnA* and *glnR* by binding to the leader sequence of the newly-transcribed mRNA and prevents termination at a factor-independent terminator. (B) An antisense transcript transcribed from a promoter on the antisense strand of *glnR* regulates translation of *glnA* (and probably also *glnR*) by binding to and sequestering the RBS.

Transcriptional antitermination

In addition to the proposed antisense control, an additional novel aspect of this system has been identified: Sequencing of the region downstream of *glnA* has revealed the presence of an ORF encoding a 189 amino acid polypeptide with a proposed molecular weight of 21,400

kDa. The ORF (referred to as *glnR*) has no promoter of its own, and is cotranscribed with *glnA* (Fierro-Monti, PhD thesis; Woods & Reid, 1995). The polypeptide has been found to share homology in its carboxy terminal region with the response regulators of two-component signalling systems (see below) and in the amino terminal region with proteins which regulate gene expression by facilitating antitermination (see below). The role of an antitermination function in the regulation of *glnA* is supported by the presence in the leader region of the *glnA* of an inverted repeat which appears to be capable of forming a Rho-independent terminator structure which would terminate transcription from *p1* (Fig. 1.3; Woods & Reid, 1995). It has also been found that transcription from *p1* and *p2* (and *p3*) are regulated differentially with respect to each other in response to varying nitrogen conditions, which would indicate the presence of additional regulatory elements (Fierro-Monti et al., 1992). Primer extension experiments have shown that in *C. acetobutylicum* there are two transcriptional start sites, corresponding to *p1* and *p2* (Janssen et al., 1990). In *E. coli* YMC11 (*glnALG*) mutant, when a construct was present which carried the cloned *glnA* as well as approximately 4.1 kb of downstream sequence, repression of transcription from both promoters was observed in the presence of excess nitrogen, as indicated by primer extension product, but this was not observed when only 600 bp of downstream sequence was present (Janssen et al., 1990). This is not entirely consistent with the proposed role of *glnR* as a positive regulator, however it is not known how transcription from the two promoters is regulated in *C. acetobutylicum* (Janssen et al., 1990). It is also possible that *glnR* is a negative regulator.

The ORF may represent an additional mechanism of control of *glnA* expression, acting at the level of transcription to up-regulate it in response to an as yet unknown signal indicating nitrogen availability. As the ORF is cotranscribed with *glnA*, its action would also result in increasing the rate of its own synthesis. It is also conceivable that the increased rate of *glnA* and thus ORF transcription could inhibit the synthesis of the antisense transcript. As *p3* is located on the strand complementary to the ORF coding region, the increased frequency with which RNA polymerase passes could inhibit initiation of transcription at *p3*. It is therefore possible that the positive regulatory function of the ORF and the negative regulatory function of the antisense transcript are in direct mutual opposition at the level of their own transcription.

1.3 TRANSCRIPTIONAL ANTITERMINATION

The term antitermination seems to be a rather loosely-applied one, and at present appears to have been accepted to refer to any mechanism by which transcription of a gene is modulated by varying the degree of readthrough of terminator sequences between the gene's promoter and the coding region. A distinction is often drawn between processive and non-processive modes of antitermination, with some authors referring only to the former as antitermination and the latter as attenuation (Greenblatt et al., 1993; Landick & Turnbough, 1992; Roberts, 1992).

1.3.1 Processive antitermination: The λ paradigm

Broadly speaking, processive antitermination refers to antitermination brought about by the modification of the elongating RNA polymerase complex so that it is able to read through terminator sequences at which the unmodified complex would have been stopped (Greenblatt et al., 1993). The best characterised examples of this type of regulation are those of the N and Q proteins of phage λ, which have been comprehensively reviewed (Greenblatt et al., 1993; Roberts, 1992; Friedman & Court, 1995; Das, 1993; Weisberg & Gottesman, 1999).

Phage λ N Protein

The transcription of early λ phage genes proceeds from two divergent promoters, pL1 and pR1. Transcripts initiated at pL1 read through the gene encoding N, after which 50% of them are terminated at a factor-independent terminator. Of all of the transcripts initiated at pR1, 50% are terminated at a rho-dependent terminator structure, designated tR1, and located downstream of the *cro* gene. The remainder of the transcripts extend through the O and P genes, and are terminated in the *nin* region, which contains a collection of rho-dependent and factor-independent terminators.

N facilitates readthrough of these terminators by facilitating the formation of a complex consisting of N and the *E. coli* proteins NusA, NusB, NusD (rho termination factor), NusE (ribosomal protein S10) and NusG. This complex associates with the elongating polymerase and modifies it to a termination-resistant form. The essential cis-acting component of this system is the *nut* (**n**-**ut**ilisation) site (Barik et al., 1987), which consists of two regions, *boxA*

It has been demonstrated that the minimum requirements for a moderate level of antitermination with susta.ned processivity consisted of a QUT site and Q. In addition, the host protein NusA appeared to have a stimulatory effect (Goliger & Roberts, 1987), possibly through the mechanism described above. As is the case with N, analogous systems have been found in other bacteriophages (Weisberg & Gottesman, 1999).

1.3.2 Nonprocessive antitermination

Nonprocessive antitermination (also known as attenuation) refers to the process by which a decision is taken to terminate transcription, or to allow RNA polymerase to continue when it reaches a terminator sequence, not by altering the elongating complex itself, but rather by allowing or not allowing the terminator to form. The terminator is usually factor-independent, and the process by which the structure of the RNA leader is determined varies from system to system. In this section, emphasis will be placed on regulatory mechanisms which involve the modulation of transcription termination by the action of accessory factors (proteins and in one instance, tRNA) which interact with the nascent mRNA and influence its secondary structure, either facilitating or preventing termination.

1.3.2.1 Translation of a regulatory peptide

E. coli trp **operon**

The classic example of transcription attenuation is that of the *E. coli trp* operon, however, similar mechanisms have been identified in several other amino acid biosynthetic and aminoacyl-tRNA sythetase operons in *E. coli*, as well as several other bacteria. The *trp* operon encodes five genes necessary for the biosynthesis of tryptophan, and has been extensively studied. A detailed model has been developed for trancriptional regulation in this operon (Reviewed in Landick & Turnbough, 1992; Landick et al., 1996). The mRNA leader of this operon, which is 141 nucleotides in length, possesses several notable features which are essential for attenuation: Between the transcriptional start and the first gene of the operon, there is a short open reading frame encoding a polypeptide of 14 amino acids, two of which are tryptophan. In addition, the leader mRNA is capable of forming two secondary structures. The first structure consists of two hairpin-loops, the first of which characterises a site at which the elongating RNA polymerase pauses briefly, and the second is a rho-

and *boxB*. As the *nut* site is transcribed, RNA in the BOXB site (transcript of the *boxB* site) forms a stem-loop structure to which N binds (Chattopadhyay et al., 1995), after which the *E. coli* protein NusA attaches to N. It has been shown that readthrough of nearby terminators requires only these factors, however, in order for maximal processivity of antitermination, the whole complex is required (De Vito & Das, 1994). NusB and NusE (S10) interact to form a complex (Mason et al., 1992), a process which appears to require the BOXA site (Nodwell & Greenblatt, 1993). NusA has been shown to couple translation to transcription by reducing the rate of RNA polymerase elongation (Landick & Turnbough, 1992). Rho must translocate from its newly transcribed binding site to interact with the RNA polymerase, and it has been proposed that a ribosome closely trailing RNA polymerase could prevent this from occurring (Friedman & Court, 1995). NusG and NusD (Rho) appear to stabilize each other's binding (Li et al., 1993), as well as the whole termination complex (Burns & Richardson, 1995).

It is thought that BOXA provides a point of control, by forming a site at which an inhibitor of antitermination binds to counteract the effect of NusB. In support of this, it was observed that deletion of BOXA *in vivo* removed the need for NusB, but did not otherwise affect antitermination (Patterson et al., 1994).

Phage λ Q Protein

Q is responsible for the induction of the late phage genes by facilitating readthrough of terminators downstream of the gene encoding Q (Herskowitz & Signer, 1970). Like the N protein, Q acts to modify the RNA polymerase to a termination-resistant form. There are however, several differences between N and Q in the way they achieve this. Instead of binding to a site on the nascent RNA transcript, Q recognises a specific DNA sequence, *qut*, located between the -10 and -35 sites of pR1 (Yarnell & Roberts, 1992). The ability of Q to modify RNA polymerase is dependent on the presence of signals on the untranscribed DNA strand located at positions +2 and +6 relative to the start of transcription which cause the polymerase to pause at position +16 or +17 (Yarnell & Roberts, 1992). It appears that antitermination is a result of reduction in polymerase pausing at termination signals.

independent transcriptional terminator. The second structure consists of a single hairpin-loop, and its formation sequesters a region necessary for the downstream formation of a rho-independent terminator structure. For this reason it is termed an antiterminator. The pause hairpin also sequesters the 5' stem of the antiterminator, and its formation prevents antiterminator formation. For this reason it could also be referred to as an anti-antiterminator.

Under the model for transcriptional attenuation in this operon, after RNA polymerase has initiated transcription, and has progressed into the coding region of the leader ORF, it encounters the first hairpin-loop, which causes it to pause long enough for a ribosome to initiate translation of the ORF. Once the ribosome has reached the paused polymerase, the polymerase is released from its paused state and continues elongation. This pausing of the polymerase is essential for correct regulation, as it is the relative rates of transcription and translation in this region which determine whether the operon is expressed. In the presence of a low concentration of tryptophan, there is a shortage of charged tryptophanyl-tRNA and this causes the ribosome to stall at one of the two tryptophan codons in the leader ORF. In this stalled position, the ribosome prevents the formation of the pause, or anti-antiterminator hairpin, and thus enables antiterminator formation. This in turn prevents the terminator from forming, and the RNA polymerase transcribes the operon. When tryptophan is plentiful, the ribosome does not pause at the tryptophan codons, and follows the polymerase closely through to the stop codon of the ORF. At this point, if the ribosome remains bound at the stop codon, it blocks the formation of the antiterminator, and the terminator forms, preventing transcription. If, however, the ribosome dissociates before the polymerase has transcribed the terminator, the leader spontaneously adopts, with equal likelihood, either the anti-antiterminator or the antiterminator. This results in termination or readthrough, respectively. This randomly determined decision between termination and readthrough under conditions of tryptophan excess allows for a basal level of expression under non-inducing conditions.

E. coli pyrBI operon

Like the *trp* operon of *E. coli*, transcription of the *pyrBI* operon, which encodes two genes involved in the synthesis of pyrimidines in *E. coli*, relies on relative rates of transcription and translation of a regulatory peptide in the mRNA leader to determine whether or not the rest of

the operon is transcribed (Reviewed in Landick & Turnbough, 1992; Landick et al., 1996). There are, however subtle differences in the mechanisms by which this is achieved.

Unlike the *E. coli trp* operon, there is no alternative RNA secondary structure whose formation prohibits terminator formation. Although once again the relative positioning of the ribosome and the RNA polymerase determines whether or not the terminator is allowed to form, the determining factor in this case is the polymerase rather than the ribosome. Under the proposed model, the concentration of available UTP determines how quickly the polymerase is able to transcribe the ORF encoding the leader peptide, which contains several U-rich regions as well as a secondary structure which causes UTP-sensitive pausing of the elongating polymerase. Under conditions of high UTP concentration, the polymerase can transcribe the terminator before the trailing ribosome can reach it, and transcription is terminated. When UTP is limiting, the polymerase pauses at these U-rich regions, allowing the translating ribosome to catch up to, and follow the polymerase closely as it transcribes the rest of the ORF. As the terminator is transcribed, the ribosome prevents it from forming its secondary structure, and therefore permits readthrough.

1.3.2.2 Control of termination by tRNA

The best characterised example of this type of termination modulation is that of the *B. subtilis tyrS* gene, which encodes tyrosyl-tRNA synthetase, the enzyme responsible for charging tRNA with tyrosine (Reviewed in Henkin, 1994; Henkin, 1996). This regulatory system has been proposed to be responsible for maintaining the correct ratio of charged to uncharged tRNAs, and facilitates the expression of specific aminoacyl-tRNA synthetases in response to limitation of their cognate amino acids, and not to general amino acid starvation (Henkin et al., 1992). Features central to this system have been found in at least 25 genes encompassing 8 genera of Gram-positive bacteria, based on features of their respective mRNA leader sequences (Henkin, 1996). The basic principles of the system are outlined in Figure 1.4.

In contrast to the *trp* operon of *E. coli*, there is no evidence for the translation of a leader peptide being involved in regulation (Grandoni et al., 1993; Grundy & Henkin, 1993). It was found that the leader sequence, in addition to being able to form a factor-independent terminator, is also capable of forming a secondary structure, known as an antiterminator.

Formation of the antiterminator is energetically less favourable than that of the terminator, and it sequesters regions necessary for terminator formation. In addition to these structures, there are three other hairpin-loops which are conserved amongst all genes thought to be regulated by this mechanism (Grundy & Henkin, 1993; Grundy & Henkin, 1994).

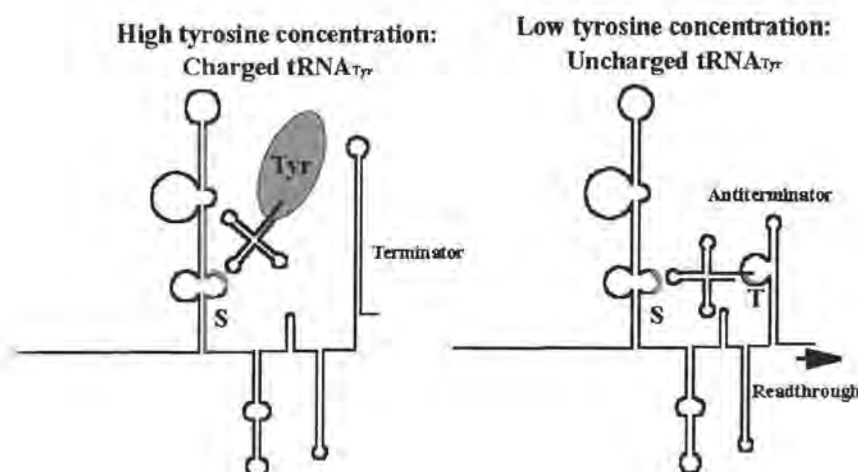


Figure 1.4: Control of transcription of the *B. subtilis tyr* operon by tRNA. Positions of the S and T box sequences are marked in grey (From Henkin, 1994).

An important feature of the system is the presence of a motif which forms a side-bulge near the base of the first stem-loop. Within this motif is the specifier region, which contains a triplet sequence corresponding to the codon of the tRNA whose charging it regulates. By altering the specifier from UAC (tyrosine) to UUC (phenylalanine), the ability of tyrosine limitation to induce expression was abolished, and instead expression was induced in response to phenylalanine limitation. However, this response was lower than that of the wild type to tyrosine, indicating the involvement of other factors (Grundy & Henkin, 1993).

Several lines of evidence have served to support the role of the ratio of charged to uncharged tRNA as the signal of amino acid availability to which the system responds (Grundy & Henkin, 1993; Grundy et al., 1994; Garrity & Zahler, 1994). A mechanism has been proposed whereby uncharged tRNA interacting with both the specifier and the antiterminator

is able, by stabilising the latter, to effect antitermination. This mechanism involves the T-box, a highly conserved 14-nucleotide motif which forms part of the 5' stem of the antiterminator. The central region of the T-box is variable, and when in the antiterminator configuration, this part of the T-box protrudes in a side-bulge. In this model, the acceptor stem of the tRNA, which in the case of tyrosyl-tRNA carries a CCA triplet, interacts with a UGG sequence in the variable region of the extruded T-box, stabilising the leader in the antiterminator formation. Furthermore, the base immediately 3' in the T-box is complementary to the discriminator base of the cognate tRNA in all genes studied so far. The discriminator is involved in determining the identity of the tRNA and is important in the charging reaction, and it has been proposed that this confers further specificity in the acceptor-T-box interaction (Grundy et al., 1994). This dual interaction between the tRNA and the leader sequence allows for further refinement of the model to incorporate regulation in response to the ratio of charged to uncharged tRNA. Because only uncharged tRNA can stabilise the antiterminator, competition between charged and uncharged tRNA for the specifier is alone sufficient to provide a measure of the relative amount of uncharged tRNA present, and to regulate the response accordingly.

It was observed that when the specifier region of the *tyrS* leader was mutated to a threonine codon, tyrosine-specific induction was lost. However, there was no concomittant increase in threonine-specific induction, as would have been expected. In the light of findings outlined above, an experiment was conducted in which the T-box of this mutant construct was mutated to be complementary to the threonyl-tRNA acceptor stem. However, although tyrosine-specific induction was decreased further, there was still no evidence of threonine-specific induction. Only when a mutant tyrosyl-tRNA with both threonine codon and threonyl-tRNA acceptor stem (i.e. corresponding to the specifier and T-box of the mutated *tyrS*) was expressed, could *lacZ* expression be restored. Thus, it appears that characteristics particular to tyrosyl-tRNA, other than the codon and acceptor stem, are involved in the tyrosyl-tRNA-*tyrS* leader interaction (Henkin, 1994).

Although the model outlined above is sufficient to account for response to limitation of specific amino acids, there is evidence for other factors which may provide additional points of control. For example, several primary sequence elements which have not been shown to

play a role in tRNA-mediated antitermination have been found to be highly conserved amongst the genes thought to use this mechanism. Furthermore, mutations in these regions have been shown to affect expression (Grandoni et al., 1993). It is uncertain whether this is reflective of their involvement in antitermination, or whether they modulate expression through other means (Henkin, 1994).

1.3.2.3 Control of termination by RNA-binding proteins

This category of regulatory mechanisms can be divided into several broad groups. In the first to be discussed here, the preferred RNA structure is the antiterminator, whose formation is prevented by the action of an mRNA binding protein. It is arguable whether this category can be described as antitermination in the strictest sense of the word, as the "active process" serves to facilitate termination. The regulatory mechanism does, however, involve a choice between two alternative RNA structures which is made by an RNA-binding protein, and for this reason it is included in this section. The second broad category includes regulatory systems involving RNA-binding proteins which act by binding RNA and preventing termination. At this point, further subdivisions can be made on the basis of how this is achieved.' Proteins can bind in such a way as to block the formation of the terminator directly, or they can stabilise secondary structures which themselves sequester regions of RNA necessary for terminator formation. Another criterion for classification of these systems is the mechanism by which the activity of the antiterminator proteins is regulated. Some of the best characterised antitermination systems are regulated by the phosphoenolpyruvate-dependent transport system (PTS) and are involved in the uptake and metabolism of sugars, for example, the *E. coli bgl* operon. Others are thought to be regulated by direct protein-protein interactions between the antiterminator and a regulatory factor, for example the *Pseudomonas aeruginosa ami* operon. In others, for example the *Klebsiella oxytoca nas* operon, it appears that the antiterminator protein is itself able to respond to chemical cues and regulate transcription. In this section I will describe examples of each of these broad categories of antitermination systems.

B. subtilis trp operon

The mechanisms governing the control of transcription of the *trp* operon of *B. subtilis* have been extensively studied and have been the subject of several recent reviews (Gollnick, 1994;

Henkin, 1996; Babitzke, 1997). The *trpEDCFBA* operon contains several of the genes involved in the synthesis of tryptophan from chorismic acid (Henner & Yanofsky, 1993). Transcription of the *trp* operon is initiated at a promoter located 203 nucleotides upstream of *trpE*, and appears to be constitutive (Gollnick, 1994). The mRNA leader sequence is capable of forming one of two mutually exclusive secondary structures, one of which is a rho-independent terminator which prevents transcription of the operon. The other is the antiterminator, which is energetically more favourable (Shimotsu et al., 1996) and prevents the formation of the terminator, allowing transcription of the operon (See Fig. 1.5; Merino et al., 1995; Babitzke & Yanofsky, 1993).

The *mtrB* gene was determined to encode the negative regulator of the *trp* operon, TRAP (*trp* RNA-binding attenuation protein) (Gollnick et al., 1990; Gollnick et al., 1994). The *mtrB* gene forms part of the *mtrAB* operon with *mtrA*, which encodes a GTP hydrolase involved in folic acid biosynthesis (Gollnick et al., 1990). TRAP binds to a region in the leader consisting of 11 (G/U)AG trinucleotide repeats separated by spacers comprising two or three nucleotides. It has been determined that the optimum sequence for TRAP binding would consist of 11 GAG repeats, each separated by two U residues (Babitzke et al., 1995), however this optimum arrangement is not observed in the native sequence. Binding of TRAP to the *trp* leader interferes with the formation of the antiterminator, and results in terminator formation (Merino et al., 1995; Shimotsu et al., 1996; Babitzke & Yanofsky, 1993).

TRAP has been found to associate into an 11-subunit ring-shaped structure, which has been termed the β-wheel (Anston et al., 1995). Each subunit binds a trinucleotide repeat with a three amino acid KKR motif (Yang et al., 1997). This motif consists of Lys-37 of one subunit and Lys-56 and Arg-58 of the adjacent subunit, and located at the outer edge of the wheel (Yang et al., 1997). Electron microscopy of the RNA-TRAP complex shows two 11-mers associated with each RNA molecule, which enters and exits the complex at the same subunit (Babitzke et al., 1995). This has resulted in a model of TRAP binding being proposed in which a complex of two β-wheels winds the mRNA leader around itself (Babitzke et al., 1994; Antson et al., 1995; Baumann et al., 1996). RNA binding of TRAP is modulated by the presence of tryptophan, which binds to the subunits in a highly co-operative

manner (Babitzke & Yanofsky, 1995). The tryptophan molecule binds at a hydrophobic region formed by two adjacent subunits (Antson et al., 1997).

In addition to regulating transcription of the *trp* operon, it has been observed that TRAP bound to the nascent *trp* transcript results in a secondary structure being formed which sequesters the *trpE* ribosome binding site in a hairpin loop (Merino et al., 1995). This indicates an additional level of control of expression with which *trpE* can be regulated specifically.

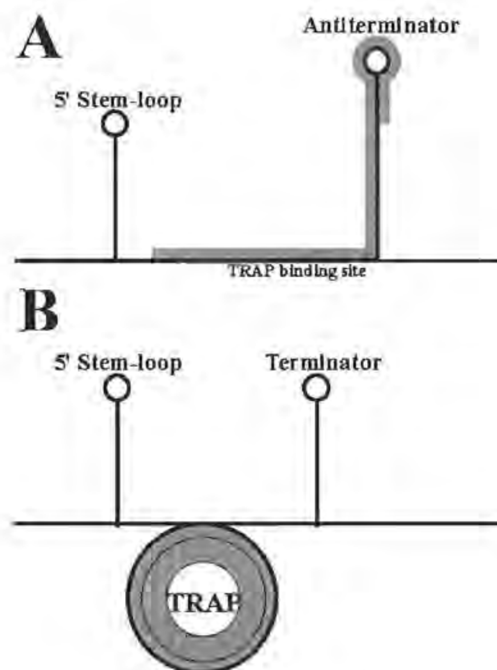


Figure 1.5: Regulation of *trpE* transcription in *B. subtilis*: TRAP binds to a region (marked in grey) consisting of 11 trinucleotide repeats, each separated by two or three nucleotides. TRAP binding prevents antiterminator formation, thereby causing termination of transcription. The TRAP binding region is marked in grey (Adapted from Babitzke, 1997).

The *trpG* gene encodes a subunit common to enzymes involved in both folic acid and tryptophan biosynthesis, and its translation, but not transcription, has been shown to be regulated by tryptophan availability (Yang et al., 1995). TRAP is responsible for regulating expression of *trpG* through a similar, but not identical, mechanism to that of *trpE*. In this case, rather than inducing a secondary structure in the RNA, TRAP binds directly to the

mRNA leader in the region of the ribosome binding site (Babitzke at al., 1994; Du et al., 1997). Here, however, the TRAP binding site consists of only nine trinucleotide repeats, and the spacers are less optimal than in the *trpE* binding site. It has been postulated that this suboptimal arrangement is to allow basal *trpG* expression for folate biosynthesis in the absence of tryptophan (Babitzke, 1997).

TRAP exhibits remarkable versatility in being able to regulate different genes in an operon in different ways at different points using a single mechanism of RNA binding. It appears that this system is shared by other members of the genus *Bacillus*, as well as other Gram-positive bacteria: Similar trinucleotide repeats and RNA secondary structures have been found in the *trp* leaders of *B. caldotenax* (Merino et al., 1995) and *B. pumilis* (Hoffman & Gollnick, 1995). *B. pumilis* has been found to possess a protein which shares 77% amino acid identity with *mtrB* and can bind *B. subtilis* and *B. pumilis trpE* mRNA in a tryptophan-dependent manner (Hoffman & Gollnick, 1995). Clostridium *thermocellum* possesses a *trp* operon leader with a similar arrangement of trinucleotide repeats (Merino et al., 1995).

***B. subtilis pyr* operon**

The enzymes responsible for pyrimidine biosynthesis in *B. subtilis* are encoded by the *pyr* operon, which consists of 10 genes. Transcription of the operon is regulated by the availability of uridine (Potvin et al., 1975), and all the genes are transcribed on a single polycistronic mRNA (Potvin et al., 1975, Lerner et al., 1987, Quinn et al., 1991). The first gene to be transcribed is *pyrR*, which encodes the negative regulator of transcription of the operon. Between *pyrR* and *pyrP*, which encodes a membrane-bound uracil permease, is a 173 nucleotide untranslated region. Downstream of *pyrP* is another untranslated region of 145 nucleotides, followed by *pyrB*, which encodes an aspartate carbamylase. This gene, and the remaining seven genes comprising the operon encode the enzymes responsible for uridine monophosphate (UMP) biosynthesis. The open reading frames of all of the latter genes overlap by between 1 and 32 nucleotides (Quinn et al., 1991; Fig. 1.6).

Untranslated RNA sequences found in the intergenic untranslated regions of the operon, as well as in the 151 nucleotide leader sequence have been found to be crucial to regulation of translation of the operon (Turner et al., 1994; Lu et al., 1995). Although there are differences

in primary sequence between the three regions, all three are capable of forming three different secondary structures: As with the *B. subtilis trp* operon, the formation of a factor-independent terminator is usually prevented by an antiterminator. Antiterminator formation is in turn prevented by another stem-loop structure, the anti-antiterminator, however, this requires an accessory factor as it is also energetically less stable than the antiterminator. It has been shown that these structures are essential for UMP-dependent regulation of transcription of the operon (Turner et al., 1994; Lu et al., 1996).

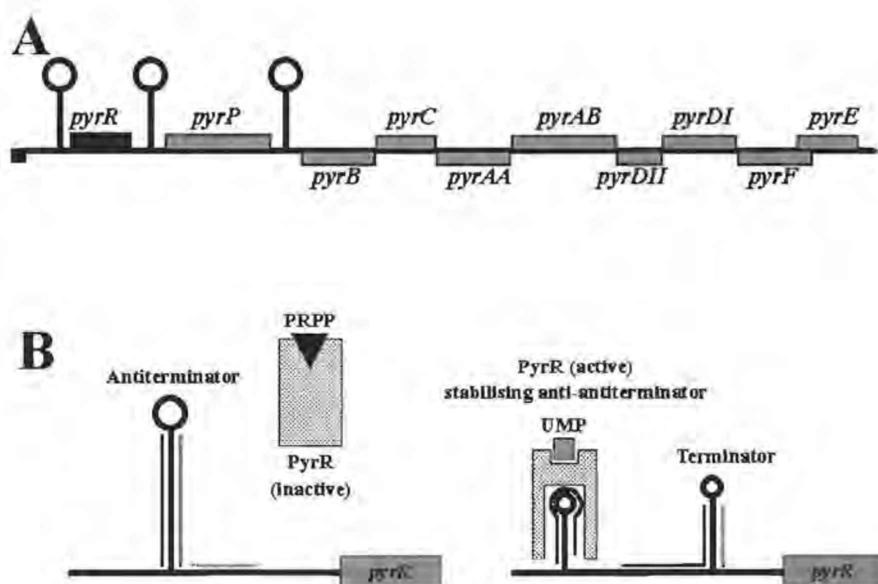


Figure 1.6: (A) Arrangement of the *B. subtilis pyr* opeon. Functions of the genes are described in the text. (B) Model of regulation of the *B. subtilis pyr* operon by PyrR in response to PRPP and UMP. Black lines indicate regions involved in terminator formation, whereas grey lines represent regions involved in formation of the antiterminator (adapted from Switzer et al., 1999).

Overexpression of the *pyr* mRNA leader resulted in loss of repression of *pyr* expression, which was restored by simultaneous overexpression of *pyrR* and the *pyr* leader. This led to the proposal that a titratable factor associated with the mRNA leader to effect repression of the operon (Antson et al., 1995). Similar effects were observed for all three of the regulatory regions, and the amount of derepression was proportional to the degree to which the leader was overexpressed (Lu & Switzer, 1996b). An in-frame deletion of *pyrR* resulted in loss of regulation, which was restored by supplying *pyrR in trans* (Turner et al., 1994). *In vitro*

experiments showed that purified PyrR was able to facilitate antitermination (Lu & Switzer, 1996) and bind specifically to *pyrR* leader mRNA (Turner et al., 1998).

The evidence presented above supports a model for the control of *pyr* transcription in which, under conditions of excess pyrimidines, the PyrR protein is bound by UMP and binds to the anti-antiterminator. PyrR binding stabilizes this structure and prevents the formation of the antiterminator. The terminator structure forms, and transcription is terminated at each of the three regulatory regions. In the absence of UMP, PyrR is inactive and does not bind to the *pyr* leader, resulting in antiterminator formation and transcription of the operon. It has been proposed that PRPP, which is a substrate in pyrimidine biosynthesis, antagonises the effect of UMP, and inctivates PyrR under conditions of high substrate availability (Switzer et al., 1999).

Studies of completely sequenced bacterial genomes has revealed the presence of elements of similar systems in *Bacillus caldolyticus, Lactobacillus plantarum, Enterococcus faecalis, Steptococcus pyogenes,* and *C. acetobutylicum* (Switzer et al., 1999). In *Deinococcus radiodurans* and *Thermus* sp. ZO5, there is evidence of a similar system, but which also shows similarities to the mechanism of control of the *E. coli trp* operon. In these organisms, it is proposed that translation of a leader peptide and polymerase pausing influences the relative positions of the ribosome and polymerase, which in turn causes the prevention or formation of a terminator (Switzer et al., 1999). Searches of genomes of other organisms, for example *E. coli* and *Helicobacter pylori*, amongst others, did not yield any evidence of similar systems, which indicates that although this mechanism is widespread, it is not universal.

***E. coli bgl* operon**

The previous two examples represent instances in which RNA-binding proteins facilitate the formation of a terminator. The rest of the examples in this section deal with situations where the binding of proteins have the opposite effect. Perhaps the best-studied example of this type of mechanism is the *bgl* operon of *E. coli*. The operon consists of four genes (Fig. 1.7) which are involved in the uptake and utilisation of aromatic β-glucosides. *bglG* encodes an antiterminator protein, *bglF* encodes a transport protein and *bglB* encodes a phopho-β-

glucosidase. Flanking *bglG* are factor-independent terminator stem-loop structures, which are overlapped by RAT (ribonucleic antiterminator) sites, which are sequences to which the regulator protein BglG binds. RAT sequences are capable of forming secondary structures which sequester regions necessary for terminator formation, however, unlike the aniterminator sequence presented above, the RAT is energetically less favourable than the terminator, and requires an accessory protein factor to stabilise it (Bramley & Kornberg, 1987; Mahadevan et al., 1987; Schnetz et al., 1987; Houman et al., 1990). When active, BglG is able to fulfil this function and induce transcription of the operon. The activity of BglG is dependent on its ability to exist as a dimer (Amster-Choder & Wright, 1992). Dimerisation is dependent on the phosphorylation state of the protein (Amster-Choder et al., 1989; Amster-Choder & Wright, 1990): Phosphorylation of BglG prevents its dimerization and thus inactivates it (Amster-Choder & Wright, 1992).

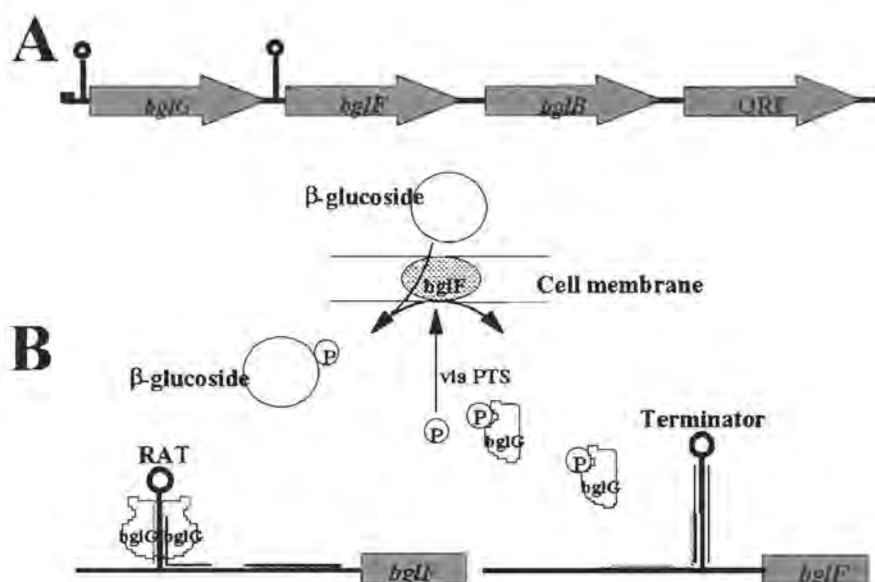


Figure 1.7: (A) Arrangement of the *E. coli bgl* operon. Functions of the genes are described in the text. (B) Model of regulation of *bgl* transcription by BglG. Black lines indicate regions involved in terminator formation, whereas grey lines represent regions involved in formation of the antiterminator (adapted from Amster-Choder & Wright, 1992).

Transcription of the *bgl* operon is regulated in response to the availability of β-glucosides by components of the PEP-dependent phosphotransferase system (PTS). When β-glucosides are

present, the phosphate from PEP is transferred through the system to BglF, which is the sugar-specific component of the PTS, and is homologous to EII^{Glc}. From here, it is transferred to the sugar, a step which is essential for it to be taken up into the cell by the *bglF* gene product. In the absence of β-glucosides, however, the phosphate is transferred instead to BglG, inactivating it and preventing it from facilitating transcription of the operon (Amster-Choder & Wright, 1990; Schnetz & Rak, 1990).

***B. subtilis sacPA* and *sacB* operons**

Sucrose utilization in *B. subtilis* is regulated by the *sacPA* and *sacB* operons. The *sacPA* operon is induced in the presence of 1mM sucrose, while the *sacB* operon is induced by a sucrose concentration of 30mM.

The *sacPA* operon consists of four genes: *sacP* encodes Enzyme II^{Scr}, a component of the PTS system and involved in the uptake of sucrose (Fouet et al., 1987). *sacA* encodes phosphosucrase (Fouet et al., 1986). *sacT* encodes a regulator, and there is an ORF encoding a protein of unknown function which is transcribed in the opposite orientation to the rest of the operon. The *sacB* regulon consists of three genes: *sacB* encodes an extracellular levansucrase (Steinmetz et al., 1985), while *sacX* encodes a protein homologous to SacB and *sacY* encodes a regulator.

Transcription of both the *sacB* and *sacPA* operons is regulated by antitermination, effected by the SacY and SacT proteins, respectively. These proteins are thought to operate by binding to the leader mRNA of their respective operons at RAT sequences and preventing termination at factor-independent terminators upstream of the coding sequences. Both of these proteins have been shown to bind RNA *in vitro* (Arnaud et al., 1996). In addition, SacT has been shown directly to prevent termination *in vitro* (Arnaud et al., 1996).

Regulation of SacY activity *in vitro* is facilitated by phosphorylation at at least three of the four histidine residues which are conserved amongst different proteins in this family, and which were identified by site-directed mutagenesis (Tortosa et al., 1997). *In vivo*, however, only one of these has been shown to be necessary for antitermination (Tortosa et al., 1997).

Phosphorylation has been shown to be performed directly by the PTS components EII and HPr *in vitro*, as well as via SacX (Tortosa et al., 1997). It is thought that, although the PTS is able to phosphorylate SacY directly, SacX serves to modulate this reaction by binding to SacY and facilitating its phosphorylation at higher efficiency. This proposal was based on the observation that in the presence of SacX, SacY was more readily phosphorylated (Tortosa et al., 1997). SacX is itself controlled by phosphorylation by an as yet unknown sucrose-specific component of the PTS (Crutz et al., 1990, Saier et al., 1993).

SacT activity is also controlled by phosphorylation, which is thought to occur at two different sites, the first of which is phosphorylated by the PTS and must be phosphorylated for the protein to be active (possibly in RNA binding). Phosphorylation of SacT at its second site inhibits dimerisation (Arnaud et al., 1992), on which its antitermination activity is thought to depend, by analogy with the mechanism of action of the *E. coli* BglG protein (see above). The protein responsible for phosphorylating SacT has not yet been identified, however it is thought to be a component of the PTS. SacP, which is the most obvious candidate by virtue of its similarity to SacX, is, paradoxically, not necessary for SacT activity. There is a degree of cross-talk between SacT and SacX under certain conditions (Steinmetz et al., 1989), however specificity of induction is achieved by differences in RAT sequences (Aymerich & Steinmetz, 1992).

In *B. subtilis*, several other operons involved in the uptake and metabolism of various compounds have been shown to be regulated by similar mechanisms, for example the *lic, bgl* and *glp* regulons (Rutberg, 1997).

Pseudomonas aeruginosa amiE operon

The amidase operon of *P. aeruginosa* enables the bacterium to catabolise short-chain aliphatic amides (Farin & Clarke, 1978). *amiE* encodes the amidase enzyme, while *amiB* and *amiS* encode proteins thought to be involved in the transport of amides. *amiC* encodes a protein which regulates the function of the antiterminator encoded by *amiR*. Although the amidase is able to catabolize a wide variety of amides, its expression is specifically induced by acetamide and propionamide, while it is repressed by butyramide (Brammar & Clarke, 1964).

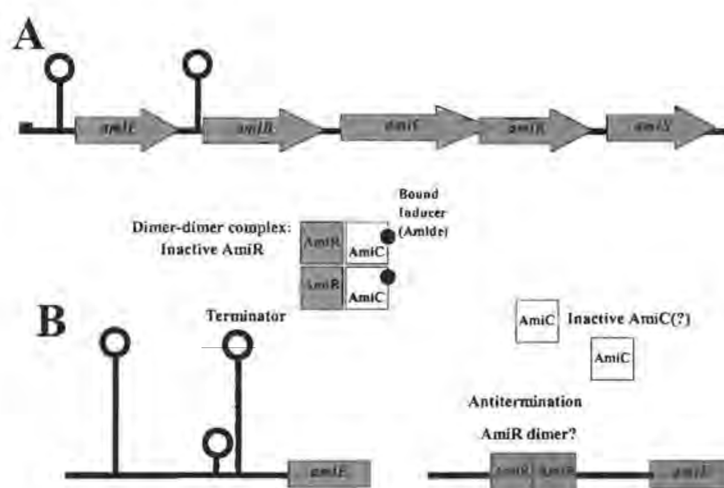


Figure 1.8: (A) The *Pseudomonas aeruginosa ami* operon. Functions of the gene products are described in the text. (B) Proposed model of regulation of transcription of the *ami* operon Hairpin loop structures thought to be important in transcriptional regulation are shown(Adapted from Rutberg, 1997; Wilson & Drew, 1996).

The arrangement of the operon is shown in Figure 1.8A. Transcription is primarily from the promoter upstream of *amiE*, with expression of the downstream genes being due to readthrough of the intrinsic terminator downstream of *amiE*. The terminator attenuates transcription of the rest of the operon by approximately 50% relative to *amiE*. There is also a second promoter located between *amiB* and *amiC* which may be responsible for constitutive low-level synthesis of the regulatory proteins, which is a prerequisite for inducibility (Wilson & Drew, 1995).

Sequence analysis of the region upstream of *amiE* showed the presence of a putative factor-independent terminator located between the promoter and the translational start of the gene. Deletions within this region, which disrupted terminator stability, resulted in high-level constitutive amidase activity independent of the presence of *amiR*. This led to the proposal of an antitermination model of control facilitated by the product of *amiR* (Drew & Lowe, 1989). The RNA binding function of AmiR has been supported by the observation that overexpression of the *amiE* mRNA leader results in the antitermination function of AmiR being titrated out (Wilson et al., 1996). Although no binding site for AmiR has yet been

identified, there are sequences on the leader which show similarity to RAT's (Drew & Lowe, 1989).

The gene *amiC*, encoding the negative regulator of AmiR and hence AmiE activity, was identified upstream of, and overlapping the transcriptional start of *amiR*. Disruption of *amiC* in the presence of *amiR* led to constitutive amidase expression (Wilson & Drew, 1991). Using three-plasmid complementation systems, it was found that AmiC and AmiR interact stoichiometrically, which suggests a direct interaction between the proteins (Wilson et al., 1993). AmiC has been determined to bind amides and is found in the cytoplasm, despite being a member of a family of membrane-bound binding proteins (Wilson et al., 1993).

From this evidence the following model of control of AmiR activity is deduced (fig 1.8B): In the absence of amides, AmiR and AmiC form a complex which prevents AmiR from binding to the transcript, possibly by sterically blocking the AmiR active site. When amides are present, however, they bind to AmiC, causing either a conformational change in, or dissociation of the complex, allowing AmiR to effect antitermination (Wilson et al., 1996). The simplest explanation for the behaviour of AmiR in the absence of AmiE is that of complex association and dissociation.

The crystal structure of AmiC has been resolved and no conformational changes appear to occur in response to different amide concentrations. Presence of the *amiE* inducer acetamide in high concentrations appears to promote trimerization, as does the repressor butyramide, although to a lesser extent (Chamberlain et al., 1997). The significance of this observation is unclear, but implies that the ability of AmiC to inactivate AmiR may be modulated by its trimerization state, which in turn is determined by the presence or absence of amides.

Klebsiella oxytoca M5a1 *nasFECBCA* operon

K. oxytoca (formerly *K. pneumoniae*) M5a1 is able to utilize nitrate and nitrite as sole sources of nitrogen. All of the components of the system by which nitrate and nitrite are taken up into the cell and reduced to ammonium are encoded in the *nasF* operon, shown in Figure 1.9A (Lin et al., 1994): *nasF, E* and *D* encode the components of a nitrate and nitrite transport system (Wu & Stewart, 1998), while *nasC* and *nasA* encode the two subunits of an assimilatory nitrate reductase (Lin et al., 1993; Lin et al., 1994). *nasB* encodes an assimilatory nitrite reductase (Lin et al., 1993).

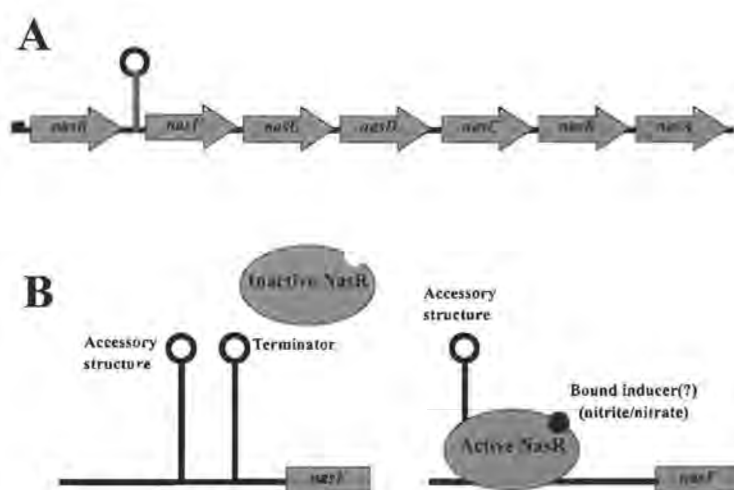


Figure 1.9: (A) Arrangement of the *Klebsiella oxytoca nas* operon. (B) Model of regulation of *nas* operon expression by NasR in response to nitrite and nitrate (Adapted from Chai & Stewart, 1998).

Initiation of transcription of the *nasF* operon from a promoter upstream of *nasF* is regulated by the Ntr global nitrogen regulatory system and is thus activated under conditions of nitrogen limitation (Cali et al., 1989) by the binding of NtrC (Lin & Stewart, 1996). The gene *nasR* was identified upstream of the *nasF* operon, and encodes a positive regulator of transcription (Goldman et al., 1994). The 44 kDa NasR protein was found to share amino acid sequence identity with the carboxyl terminus of AmiR of *P. aeruginosa* (see above).

The following evidence was found for an antitermination mechanism of nitrate and nitrate-specific regulation of transcription (see Fig. 1.9B): Replacement of the promoter with the *tac*

promoter did not alter nitrate/nitrite induction, although Ntr control was abolished (Lin & Stewart 1996). Analysis of the leader sequence of the *nasF* mRNA revealed the presence of a factor-independent terminator-like structure, which, if disrupted, resulted in constitutive expression (Lin & Stewart, 1996). Another hairpin-loop structure upstream of the terminator was found to be essential for transcription of the operon. These results were confirmed in *in vitro* studies, in which it was shown that NasR binds specifically to the leader, and facilitates readthrough of the terminator in response to the presence of nitrate and nitrite (Chai & Stewart, 1998). It also appears that the upstream hairpin loop structure is essential for NasR to bind to the leader and effect antitermination (Chai & Stewart, 1998).

***Azotobacter vinelandii nasAB* operon**

Although it has not been established conclusively whether control of this operon involves an antitermination mechanism, it has been included in this section because of the degree of amino acid sequence identity between the amino acid sequence of the NasT protein and that of the carboxy terminal of NasR of *K. oxytoca* (Lin & Stewart, 1998). The amino acid sequence of *nasT* also shares 22% sequence identity (39 identical residues out of 174) with *glnR* of *C. acetobutylicum* P262 (See chapter 2). Like *K. oxytoca*, *A. vinelandii* is able to utilise nitrate as a sole source of nitrogen, which involves the products of the *nasA* and *nasB* genes, which encode a nitrite and a nitrate reductase, respectively (Ramos et al., 1993).

A positive regulator essential for induction of the *nasAB*, *nasT* was identified and found to encode a protein, NasT, which shares amino acid sequence identity with response regulators of two-component signalling systems (Guttierez et al., 1995), but it has been pointed out, however, that the highly conserved lysyl residue essential for response regulator function is not present (Lin & Stewart, 1998). Additionally, the carboxyl terminus of NasT shares amino acid sequence identity with the carboxyl termini of NasR of *K. pneumoniae* and AmiC of *P. aeruginosa* which are both known to function as antiterminators (Lin & Stewart, 1998).

The product of the *nasS* gene, NasT, has been shown to regulate negatively the transcription of *nasAB*, and shares amino acid sequence similarity with proteins thought to bind nitrate in the periplasm (Guttierez et al., 1995). It has been speculated that this protein may have a function analogous to that of AmiC in *P. aeruginosa*, and that a mechanism of regulation by

antitermination, similar to the AmiC/AmiR system of *P. aeruginosa* is at work in *A. vinelandii* (Lin & Stewart, 1998).

It can be seen that there is considerable diversity within even this limited subset of examples of transcription antitermination mechanisms, both with respect to the structures formed by the mRNA leader and the proteins involved. These systems represent a set of elegant variations on the common theme of regulating gene expression through modulation of RNA secondary structure.

1.4 AIMS OF THIS DISSERTATION

The aim of this work consists primarily of gathering evidence in support of the proposed role of the *glnR* gene product as an antiterminator and regulator of *glnA* transcription. The only evidence available to date is circumstantial, and consists of (i) the presence of terminator-like structures in the leader sequence of *glnA*, and (ii) the similarity shared by the deduced amino acid sequence of the *C. acetobutylicum* NCP 262 *glnR* with those of the known antiterminators AmiR of *P. aeruginosa* (30% similarity over 60 residues of the carboxy terminus) and NasR of *K. oxytoca* (38% identity over 73 residues; see chapter 2). More direct and convincing evidence is needed to substantiate models previously proposed, which involve *glnR* in the regulation of nitrogen metabolism in *C. acetobutylicum* P262 (Woods & Reid, 1995). In order to do this, a suitable system must be developed to study the effects of the presence and absence of the *glnR*. As it is not possible to transform *C. acetobutylicum* P262 reliably with plasmid DNA, an alternative host is necessary. Previous work demonstrated that the cloned *glnA* was regulated in *E. coli* YMC11 in response to nitrogen availability, although this has been disputed. In this work, an attempt to demonstrate a regulatory role for *glnR* in an *E. coli*-based system will be described. Additionally, the evaluation of the closely related and better characterised *C. beijerinckii* NCIMB 8052 as an alternative host for these studies will be described.

In order to aid investigation of promoter activity, specifically the *C. beijerinckii* NCIMB 8052 *glnA* promoter, a reporter system was necessary. Attempts to develop such a system

based on the *eglA* gene cloned from *C. acetobutylicum* P262 will be described, as well as experiments in which it was used to investigate activity of the *C. beijerinckii* NCIMB 8052 *scr* and *glnA* promoters.

CHAPTER 2

STUDIES ON PHYSIOLOGICAL AND GENETIC ASPECTS OF NITROGEN METABOLISM REGULATION IN *Clostridium acetobutylicum* NCP 262 AND *Clostridium beijerinckii* NCIMB 8052

CONTENTS

2.1 INTRODUCTION **40**

2.2 MATERIALS AND METHODS **42**

2.2.1 Strains and plasmids 42
2.2.2 Regulation of the cloned *C. acetobutylicum* NCP 262 *glnA* gene in *E. coli* YMC11 43
2.2.3 Sequencing of the region upstream of *glnA* in *C. acetobutylicum* NCP 262 44
2.2.4 Growth of *C. beijerinckii* NCIMB 8052 using various nitrogen sources 45
2.2.5 Regulation of GS activity in *C. beijerinckii* NCIMB 8052 45
2.2.6 Detection of *glnA*, *glnR* and *gltA* homologues in *C. beijerinckii* NCIMB 8052 45
2.2.7 Cloning of the *C. beijerinckii* NCIMB 8052 nitrogen region 46
2.2.8 Complementation of glutamine auxotrophy in *E. coli* YMC11 47
2.2.9 Sequencing of the *C. beijerinckii* NCIMB 8052 nitrogen region 47
2.2.10 Mutagenesis of the *glnR* gene in *C. beijerinckii* NCIMB 8052 47
2.2.11 Determination of the start of transcription of *glnA* of *C. beijerinckii* NCIMB 8052 49

2.3 RESULTS AND DISCUSSION **49**

2.3.1 Effects of *glnR* on growth and GS activity in *E. coli* YMC11 49
2.3.2 Sequencing of the region upstream of *glnA* in *C. acetobutylicum* NCP 262 52
2.3.3 Growth of *C. beijerinckii* NCIMB 8052 using various nitrogen sources 57
2.3.4 Regulation of GS in *C. beijerinckii* NCIMB 8052 60
2.3.5 Detection of the *C. beijerinckii* NCIMB 8052 nitrogen region 61
2.3.6 Cloning of the *C. beijerinckii* NCIMB 8052 nitrogen region 63
2.3.7 Complementation of glutamine auxotrophy in *E. coli* YMC11 64
2.3.8 Sequencing of the *C. beijerinckii* NCIMB 8052 nitrogen region 66
2.3.9 Mutation of *glnR* in *C. beijerinckii* NCIMB 8052 68
2.3.10 Determination of the start of transcription of *glnA* of *C. beijerinckii* NCIMB 8052 69

2.4 CONCLUSIONS **73**

SUMMARY

An ORF encoding a putative antiterminator (*glnR*) was identified downstream of *glnA* on a fragment of DNA cloned from *C. acetobutylicum* NCP 262. To assess the effect of *glnR* expression on growth and GS activity, the *glnAntrBC*⁻ *E. coli* mutant YMC11 was transformed with plasmids carrying the cloned *glnA* gene with either the entire or truncated *glnR* gene. It was found that the presence of the *glnR* gene did not have any marked effect on either growth or GS activity, and it was proposed that this may have been due to the unsuitability of *E. coli* as a host organism for the reconstitution of the *C. acetobutylicum glnA* regulatory system. In an attempt to locate a cognate protein kinase for the GlnR protein, a region of approximately 1.3 kb upstream of the *C. acetobutylicum* NCP 262 was sequenced. An incomplete open reading frame was identified, which shared sequence identity with aspartokinases and bifunctional aspartokinase/homoserine dehydrogenases, which are involved in the biosynthesis of lysine, methionine and threonine. It appears the this ORF marks the 5' boundary of the *C. acetobutylicum* NCP 262 *glnA* operon. It is possible that the *glnA*, *glnR*, *gltA* and *gltB* genes are part of a larger cluster of amino acid biosynthesis genes. The closely related *C. beijerinckii* NCIMB 8052 was assessed for its suitability as an alternative host. *C. beijerinckii* NCIMB 8052 was found to be similar to *C. acetobutylicum* NCP 262 in its preference for organic nitrogen. Also, GS activity in *C. beijerinckii* NCIMB 8052 was found to be repressed at high concentrations, and induced at low concentrations of organic nitrogen, as has been observed for *C. acetobutylicum* NCP 262. Southern hybridisation analysis of *C. acetobutylicum* NCP 262 and *C. beijerinckii* NCIMB 8052 was performed using probes for the *C. acetobutylicum* NCP 262 *glnA*, *glnR* and *gltA* genes. Homologues of these genes were identified in *C. beijerinckii* NCIMB 8052, and their arrangement was found to be similar to that in *C. acetobutylicum* NCP 262. A 3.8-kb *Eco*RI fragment of *C. beijerinckii* NCIMB 8052 was cloned and found to complement glutamine auxotrophy in *E. coli* YMC11, demonstrating a functional *glnA* gene. The clone was sequenced completely, and was found to encode one incomplete and two complete open reading frames, bearing sequence identities of 87.8%, 86.5% and 86.8% to *glnA*, *glnR* and *gltA* of *C. acetobutylicum* NCP 262, respectively. Factor-independent terminator-like structures were found in the mRNA leader sequence of the *C. beijerinckii* NCIMB 8052 *glnA* transcript and were similar to those which had been identified previously in *C. acetobutylicum* NCP 262. It was concluded that physiologically and genetically, *C. beijerinckii* NCIMB 8052

and *C. acetobutylicum* NCP 262 are sufficiently similar with respect to nitrogen metabolism to warrant the use of the former as a model organism for the study of nitrogen metabolism in the latter. Attempts to mutate *glnR* in *C. beijerinckii* NCIMB 8052 using a mutagenesis vector derived from the *C. beijerinckii* NCIMB 8052 *glnR* internal sequence were unsuccessful. Primer extension reactions were performed using total RNA extracted from cultures grown under conditions of limiting and excess nitrogen. Four transcriptional start sites were identified, two of which corresponded approximately to the transcriptional start sites identified in *C. acetobutylicum* NCP 262, and which were adjacent to potential promoter sequences. Transcription from all four start sites appeared to be elevated under conditions of nitrogen limitation.

2.1 INTRODUCTION

A region of the *C. acetobutylicum* NCP 262 chromosome, encompassing several important genes involved in nitrogen metabolism, including *glnA* and *glnR*, had previously been cloned in *E. coli* and the *C. acetobutylicum* NCP 262 clone was able to complement glutamine auxotrophy in the *E. coli glnAntrBC* deletion mutant YMC11 (Usdin et al., 1986). It was also reported that the activity of the cloned GS was regulated in response to nitrogen availability (Usdin et al., 1986; Janssen et al., 1988). However, subsequent work using the cloned genes in *E. coli* could not conclusively link regulation of the cloned *glnA* gene with nitrogen availability, owing to wide variations in growth rates and GS activities under different nitrogen conditions (Fierro-Monti, PhD thesis). Studies on the cloned DNA in *E. coli* YMC11 also helped to characterise a mechanism of control of GS expression by antisense RNA (Janssen et al., 1990; Fierro-Monti et al., 1992).

Sequencing downstream of the *glnA* conding region revealed the presence of an open reading frame encoding a putative antiterminator protein. This, together with the presence of a putative terminator structure between the upstream promoter p_1 and the translational start of *glnA* (see chapter 1) suggested a transcriptional antitermination mechanism of control of *glnA* expression (Fierro-Monti, PhD thesis; Woods & Reid, 1995).

C. acetobutylicum NCP 262 is not amenable to many routine genetic manipulations, although some success has been acnieved in the generation of mutants through the use of conjugative transposons (Babb et al., 1990). Alteration of the genotype of this strain has also been achieved through protoplast transformation techniques (Reid et al., 1983), however, other methods of transformation, such as electroporation, have so far proven unsuccessful. Furthermore, plasmid replication in this strain has so far not been observed. *C. beijerinckii* NCIMB 8052 has been shown to be relatively closely related to *C. acetobutylicum* NCP 262, and the two strains share a 16S ribosomal RNA (rRNA) gene sequence similarity of 97.6%. This represents a closer evolutionary relationship than that which exists between *C. acetobutylicum* NCP 262 and *C. acetobutylicum* ATCC 824, another well-characterised strain. In this case, 16S rRNA sequence similarity is 88.9% (Keis et al., 1995). For this reason, *C. beijerinckii* NCIMB 8052 was chosen over *C. acetobutylicum* ATCC 824 as a subject for these studies. *C. beijerinckii* NCIMB 8052 has been relatively well characterised and many tools are available for its manipulation. Transformation of this strain by electroporation (Lee et al., 1992; Oultram et al., 1988) as well as conjugative transfer (Williams et al., 1990a; Williams et al., 1990b) has been achieved, and several cloning vectors have been developed, based on the pAMβ1 and pIM13 replicons (Oultram et al., 1988; Swinfield et al., 1990). *E. coli-Clostridium* (Lee et al., 1992), as well as *B. subtilis-Clostridium* (Mermelstein et al., 1992) shuttle vectors are available. In addition, a system for targeted mutagenesis has been developed for this strain, as described below. In order to determine its suitability as a platform for these studies, it was necessary to determine whether the mechanisms controlling nitrogen metabolism are sufficiently similar at the physiological and genetic levels to those of *C. acetobutylicum* NCP 262.

Targeted mutagenesis of genes using homologous recombination has been achieved in *C. beijerinckii* NCIMB 8052 (Wilkinson & Young, 1994). This process involves the introduction of a plasmid into *C. beijerinckii*, and its subsequent integration into the chromosome at a locus determined by the sequence carried by the plasmid. Homologous recombination through a single crossover event between a chromosomal gene and a portion of its internal sequence cloned into the plasmid results in integration of the entire plasmid. This causes disruption of the gene in question, and also allows markers carried by the plasmid to be expressed and used for selection for the integration event.

Primer extension analysis of the *C. acetobutylicum* NCP 262 *glnA* gene in both *C. acetobutylicum* NCP 262 and *E. coli* YMC11 has revealed the presence of two transcriptional start points, t_1 and t_2, which corresponded with the positions of the proposed promoters p_1 and p_2. These promoters had been identified by sequence inspection and comparison with known promoters in Gram negative bacteria (Janssen et al., 1990). In *E. coli*, additional transcriptional start sites were identified for the cloned *C. acetobutylicum* NCP 262 *glnA* gene. Transcriptional start site t_{1a} was located 8 nucleotides downstream of t_1, while t_{2a} was found 1 nucleotide downstream of t_2 (Janssen et al., 1990). As a defined growth medium had not yet been developed for *C. acetobutylicum* NCP 262 at that time, it was not possible to investigate whether transcription from the two promoters was differentially regulated in response to nitrogen availability in this organism.

Here, an attempt to reconstitute the proposed antitermination control mechanism in *E. coli* YMC11 is described. The use of *C. beijerinckii* NCIMB 8052 as a model organism for the study of nitrogen metabolism in *C. acetobutylicum* NCP 262 was also investigated. The *C. beijerinckii* NCIMB 8052 homologues of the *C. acetobutylicum* NCP 262 *glnA* and *glnR* genes, as well as part of the *gltA* homologue, were cloned and sequenced. In addition, the construction of a vector for the creation of a knockout *C. beijerinkii* NCIMB 8052 *glnR* mutant and attempts to obtain the mutant are described.

2.2 MATERIALS AND METHODS

2.2.1 Strains and plasmids

All bacterial strains and plasmids used in these experiments are listed in Table 2.1. All *E. coli* cultures were maintained on 2xYT agar (A.2.1.4), LA (A.2.1.2) or M9 minimal medium agar (A.2.2.2) with the appropriate antibiotic selection and amino acid supplements, where applicable. For growth in liquid culture, 2xYT (A.2.1.3), LB (A.2.1.1) or M9 minimal medium (A.2.2.2) with appropriate additives were used. *C. beijerinckii* NCIMB 8052 and *C. acetobutylicum* NCP 262 were routinely cultured on CBM agar (A.2.1.6) and in CBM (A.2.1.5) or GSMM (A.2.2.1) liquid media. Anaerobic conditions were achieved using a Forma Scientific anaerobic chamber.

Table 2.1 Strains and plasmids used in this work

Name	Relevant Characteristics	Reference
Bacterial strains		
C. acetobutylicum NCP 262	Wild-type	Jones et al. (1982)
C. beijerinckii NCIMB 8052	Wild-type	NCIMB
E. coli YMC11	Δ*lacU*169, Δ(*glnG-glnA*)	Chen et al. (1982)
E. coli JM109	Δ(*lac-proAB*) *lacI*^{q} Δ(*lacZ*)*M15 recA1*	Yanisch-Perron et al., (1985)
E. coli CA448 (formerly HB101(R702)[λcI857])	*λcI857F hsd20*($r_B^-m_B^-$)*endoI recA13 supE44 ara-14 proA2 galK2 rpsL20* (Sm^R) *xyl-5 mtl-1*	Williams et al., 1990a
Plasmids		
pHZ200	Ap^R *glnA*$^+$	Usdin et al. (1988)
pHZ202	Ap^R *glnA*$^+$ *glnR*$^+$	Usdin et al. (1988)
pHZ206	Ap^R *glnA*$^+$ *glnR*$^-$	Usdin et al. (1988)
pHZ207	Internal *glnR* fragment	Brown, unpublished
pHZ208	Internal *glnA* fragment	Brown, unpublished
pBluescriptSK+	Ap^R, T_3 and T_7 polymerase promoters	Stratagene, Ja Holla, California
pMTL30	Tra^- Mob^+ Ap^R Em^R *lacZ'*	Williams et al. (1990a)
pCTC1	Tra^- Mob^+ Ap^R Em^R; pAMβ1 replicon	Williams et al. (1990b)
pCBN1	*C. beijerinckii* NCIMB 8052 *glnA* region, cloned in pSK	This work
pIO2	Internal *C. beijerinckii* NCIMB 8052 *glnR* fragment cloned into pMTL30	This work

2.2.2 Regulation of the cloned *C. acetobutylicum* NCP 262 *glnA* gene in *E. coli* YMC11

E. coli YMC11 cells were transformed with pHZ202 and pHZ206 (see Fig. 2.1). Cultures of each of the transformants, as well as of untransformed YMC11 cells, were grown in 5ml volumes of 2xYT broth overnight to stationary phase and were diluted 40-fold in 200ml of 2xYT broth, either with or without glutamine supplemented to a final concentration of 0.15%. Growth of the cultures in these media was evaluated and glutamine synthetase activity was assayed at early exponential (OD_{600} = 0.6) and late exponential phase (OD_{600} = 1.5) using the γ-glutamyl transferase assay (A.1.6.1). As untransformed *E. coli* YMC11 was used as a negative control in these experiments, no antibiotic selection was used in any of the cultures.

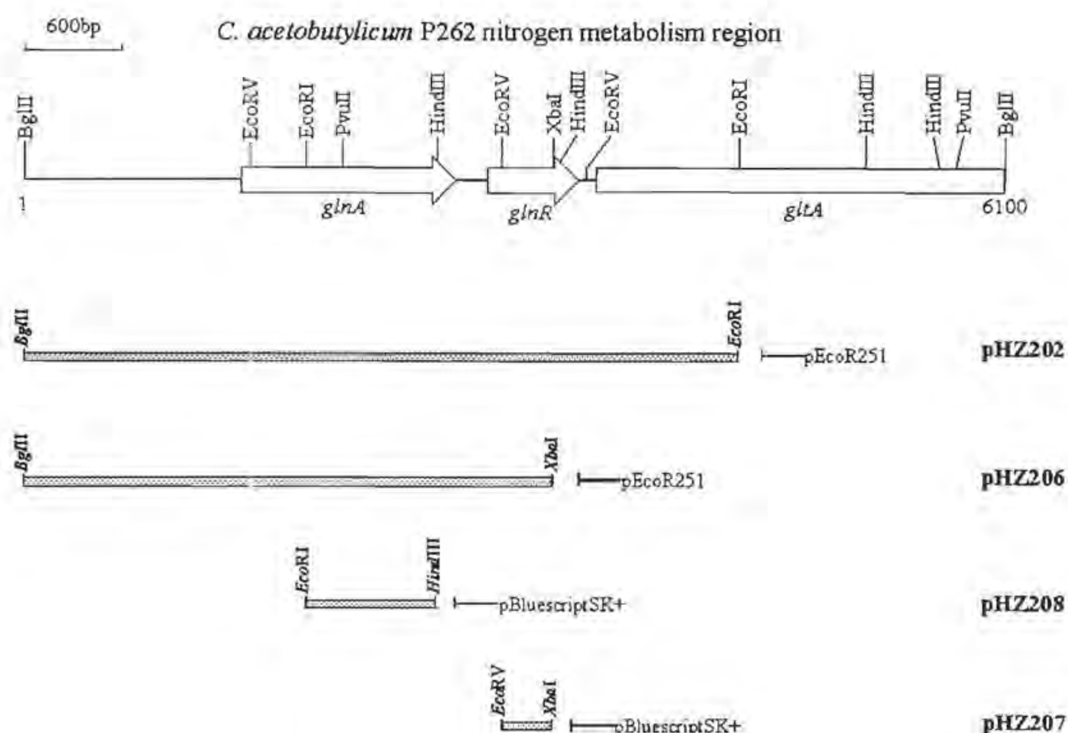


Figure 2.1: Partial restriction map of the *C. acetobutylicum* NCP 262 nitrogen metabolism region cloned on pHZ200, illustrating the regions used for experiments on *glnA* expression in *E. coli* YMC11 (pHZ202 and pHZ206), and construction of probes used in Southern hybridisation experiments (pHZ207 and pHZ208). Restriction sites shown are those used to excise DNA for probe preparation (*Eco*RI, *Eco*RV, *Hin*dIII and *Xba*I) and for digestion of chromosomal DNA for Southern hybridisation (*Eco*RI, *Bgl*II and *Pvu*II).

2.2.3 Sequencing of the region upstream of the *glnA* gene in *C. acetobutylicum* NCP 262.

Using the technique of Henikoff shortening (A.1.2.6), a series of 5' and 3' deletions were made of plasmid pC16 (Stutz, unpublished data), which encodes a region spanning 24 nucleotides of the 5' terminus of the *C. acetobutylicum* NCP 262 *glnA* gene and 1.3 kb of the region upstream of it. For the creation of 5' deletions, the plasmid was digested with *Kpn*I and *Hin*dIII. *Kpn*I was used to provide a 3' single-stranded overhang to prevent exonuclease III digestion into the vector, while *Hin*dIII was used to create a 5' single stranded overhang from which exonuclease III was able to digest into the insert. For the generation of 3' deletions, 3' and 5' single-stranded overhangs were generated using *Sac*I and *Bam*HI, respectively. After religation and transformation into *E. coli* JM109, plasmids were extracted and digested with *Pvu*II and clones carrying inserts of appropriate size were selected for sequencing.

2.2.4 Growth of *C. beijerinckii* NCIMB 8052 using various nitrogen sources

C. beijerinckii NCIMB 8C52 spores were used to inoculate 200ml volumes of CBM, as well as GSMM with the different combinations of organic and inorganic nitrogen sources. Cultures were grown anaerobically at 37^0C. Organic nitrogen was supplied in the form of casamino acids (Difco), with a concentration of 0.2% representing nitrogen excess and 0.05% and 0.025% representing nitrogen limitation. Inorganic nitrogen was supplied in the form of either ammonium acetate or ammonium sulphate, also at concentrations of 0.2% and 0.025% representing nitrogen excess and limitation, respectively. Each of these nitrogen sources was supplied separately and in combination. In addition, organic nitrogen sources were supplemented with glutamine at a concentration of 0.015%. The ability of each nitrogen source to sustain growth was assessed.

2.2.5 Regulation of GS activity in *C. beijerinckii* NCIMB 8052

C. beijerinckii NCIMB 8052 cultures were inoculated with spores and incubated anaerobically at 37^0C with occasional agitation. Conditions of nitrogen excess were represented by either complete medium (CBM) or minimal medium (GSMM) supplemented with 0.2% casamino acids, or 0.2% casamino acids plus 0.015% glutamine. Conditions of nitrogen limitation were represented by GSMM supplemented with 0.05% casamino acids, or 0.05% casamino acids plus 0.015% glutamine. Glutamine synthetase activity of cultures grown in each of these media during early, mid and late exponential phase growth (OD_{600} = 0.3, 0.5 and 1.0 respectively) was assayed.

2.2.6 Detection of *glnA*, *glnR* and *gltA* homologues in *C. beijerinckii* NCIMB 8052.

The plasmid pHZ208 (Brown, unpublished data), a derivative of the plasmid pHZ200 which carries the cloned *C. acetobutylicum* NCP 262 *glnA*, *glnR* genes and flanking sequences, carries a 1.5 kb *Eco*RI/*Hin*dIII internal fragment of *glnA* which was excised by digestion with *Eco*RI and *Hin*dIII (Fig. 2.1). pHZ207 (Brown, unpublished data), another pHZ200 derivative, carries a 325bp *Eco*RI/*Hin*dIII internal fragment of *glnR*. This fragment was excised using *Eco*RV and *Hin*dIII (Fig. 2.1). Both fragments were purified by electrophoresis in 1% low-melting-point agarose (Sea-Plaque) and extracted using a Gene-Clean kit as instructed by the manufacturers. The probes were labelled using the DIG non-radioactive system. The random-primed labelling reaction was used, as directed by the manufacturer.

C. beijerinckii NCIMB 8052 and *C. acetobutylicum* NCP 262 chromosomal DNA was extracted (A.1.1.6) and digested with *Bgl*II, *Eco*RI and *Pvu*II. The digested DNA was resolved on a 0.8% agarose gel. Southern transfer of the digested DNA, hybridisation and detection of the probe were all performed following the instructions supplied with the DIG nucleic acid labelling and detection system. Prehybridisation and hybridisation was carried out at 37°C in "DIG Easy Hyb" solution (Roche).

2.2.7 Cloning of the *C. beijerinckii* NCIMB 8052 nitrogen region

Chromosomal DNA was digested to completion with *Eco*RI. To identify the position of the band of interest, 30μg of the DNA, as well as of the appropriate controls, was electrophoresed on a 0.8% agarose gel. Another sample comprising 70μg of the digested DNA was run on the same gel. After the gel had been run, it was cut to separate the two sets of samples. The first set was used for Southern hybridisation, as described previously in this chapter. The probe used was that for the *glnA* gene (see above). Using the position of the band to which the probe hybridised as a guide, the corresponding band was excised from the half of the gel carrying the second sample. The DNA was extracted from the agarose by electroelution for 4 hours at 50 volts in 1xTBE (A.1.2.8) and cloned into the *Eco*RI site of pBluescriptSK+. The resulting population of plasmids was used to transform *E. coli* JM109 (A.1.5.1). Transformants were plated out onto 2xYT agar containing 100 μg/ml ampicillin, 0.004% X-gal and 0.5mM IPTG. A total of 900 recombinant transformants were streaked onto plates and incubated at 37°C overnight. Colony hybridisation was performed, using the DIG non-radioactive system according to the protocol recommended by the manufacturers. Again, the probe used was that for the *glnA* gene, and prehybridisation and hybridisation was performed in "DIG Easy Hyb" at 37°C.

DNA was extracted from four selected clones, digested with *Eco*RI, and subjected to Southern hybridisation using the *glnA* probe to determine whether they carried the correct insert. Additionally, the sequences of the 5' and 3' termini of the insert of one clone identified on the basis of colony and Southern hybridisation were determined using the Sanger di-deoxy chain termination reaction and an ALFexpress automated DNA sequencer (Pharmacia). Sequences were compared to the sequence of the *C. acetobutylicum* NCP 262 nitrogen region, using the FASTA program of the GCG v9.1 package (Genetics Computer Group, Madison, Wisconsin).

2.2.8 Complementation of glutamine auxotrophy in *E. coli* YMC11

The *E. coli* mutant YMC11 was transformed with pCBN1, which comprises the *C. beijerinckii* NCIMB 8052 DNA cloned into pBluescriptSK+ and which was found to hybridise with the *C. acetobutylicum* NCP 262 *glnA* probe. A single colony was selected from the pool of transformants and was streaked onto M9 minimal medium agar supplemented with glutamine at a concentration of 0.2%, as well as onto plates without glutamine. As a negative control, YMC11 cells transformed with non-recombinant pBluescriptSK+ were subjected to the same treatment. The plates were incubated at 37^{0}C and observed over the following 48 hours for the presence or absence of growth. Minipreps (A.1.1.1) were used to verify that the transformants carried the correct plasmids.

2.2.9 Sequencing of the *C. beijerinckii* NCIMB 8052 nitrogen region

In addition to sequencing the 5' and 3' termini of pCBN1, a series of deletion derivatives were created in pBluescriptSK+ and sequenced (Fig. 2.2). Sequencing was performed using an ALFexpress automated DNA sequencer (Pharmacia). Universal sequencing primers for pBluescriptSK+ were used. Sequence analysis was performed using the GCG v9.1, DNAman v4.0 (Lynnon Biosoft, 1997), and Genedoc (Nicholas & Nicholas, 1997) software packages. RNA secondary structures were predicted using the RNAstructure v2.0 program (Matthews et al., in press).

2.2.10 Mutagenesis of the *glnR* gene in *C. beijerinkii* NCIMB 8052

The mutagenesis vector pIO2 was constructed in the following manner: The *Pvu*II/*Eco*RI fragment of pCBN1 was cloned into the *Eco*RI/*Sma*I site of pBluescriptSK+ to produce the plasmid pCBN31. The *Xba*I fragment of pCBN31 was then excised and cloned into the *Xba*I site of pBluescriptSK+ to create pIO1. The *Xba*I/*Ssp*I fragment of pIO1 was cloned into the *Xba*I/*Sma*I site of pMTL30 (Chambers et al., 1988) to create the mutagenesis vector pIO2. The length of sequence available for recombination was 363bp.

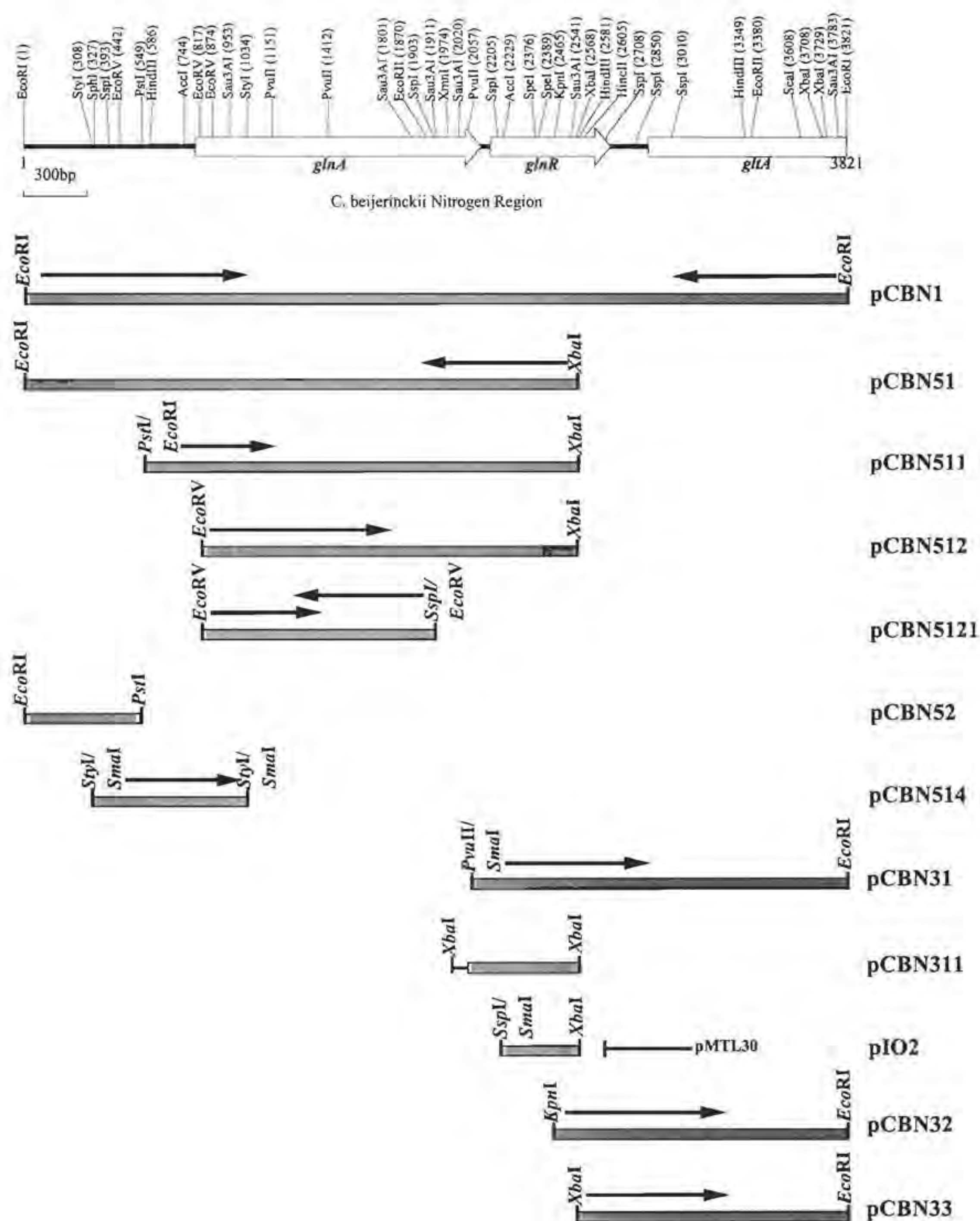


Figure 2.2: Map of the region of the *C. beijerinckii* NCIMB 8052 carried on pCBN1, illustrating subclones used for sequencing the clone and for the construction of the mutagenesis vector pIO2. Black arrows indicate the direction and extent of sequencing of each subclone. Unless otherwise indicated, all subclones were constructed in pBluescriptSK+.

E. coli strain CA448, which was to be the donor strain used in this experiment, was transformed with pIO2. The plasmid was transferred by conjugation (A.1.5.4) into *C. beijerinkii* NCIMB 8052, as were the plasmids pCTC1 and pMTL30, which were used as positive and negative controls respectively. Selection for transconjugants was performed on CBM agar plates containing erythromycin at a concentration of 20μg/ml, as well as glutamine at a concentration of 0.2%.

2.2.11 Determination of the start of transcription of *glnA* of *C. beijerinckii* NCIMB 8052

Cultures of *C. beijerinckii* NCIMB 8052 were grown under conditions of nitrogen limitation and excess, as described previously. At mid-exponential phase (OD_{600} = 0.5), RNA was extracted either using the procedure described by Aiba et al. (1981; A.1.1.7) or, using Tri-zol reagent as described by the manufacturers (Promega). GS activity of each of the cultures at the time of harvest was assayed (A.1.6.1).

Using the programs Prime, of the GCG v9.1 package and DNAman v4.0, an oligonucleotide primer was designed to anneal to nucleotides 37 to 54 of the *glnA* transcript (Fig 2.13). The sequence of the primer was as follows: 5'-TTAACTCCATTTTCCTTTAC-3', and the T_m of the DNA:RNA hybrid was calculated to be 33.2^{0}C in the absence of formamide. The 3' terminus of the oligonucleotide was labelled with cyanidin for detection. Primer extension reactions were conducted on 100μg of total RNA from each of the cultures as described (A.1.1.7). Resulting cDNA was run on an AlfExpress automated sequencer, along with the products of sequencing reactions performed using the same primer, and pCBN1 as a template.

2.3 RESULTS AND DISCUSSION

2.3.1 Effect of *glnR* expression on growth and GS activity in *E. coli* YMC11

The model of control proposed for the *C. acetobutylicum* NCP 262 *glnA* gene allows certain hypotheses to be generated. One of these is that as an antiterminator, GlnR would be expected to be a positive regulator of *glnA* expression and therefore GS activity. A second is that the presence of a transcriptional terminator upstream of the *glnA* structural gene would be

expected to prevent transcription of *glnA* in the absence of other factors. The absence of the *glnR* gene would therefore be expected to result in low levels of GS activity. To test this, the *E. coli glnAntrBC⁻* mutant YMC11 was transformed with two different constructs. The plasmid pHZ202 carries the *C. acetobutylicum* NCP 262 *glnA* structural gene, as well as upstream DNA carrying its promoter and other control elements. In addition to this, it also carries the downstream *glnR* gene. The plasmid pHZ206 also carries the entire *glnA* gene and its upstream sequence, but a truncated copy of the *glnR* gene, which would be expected to be inactive. Cultures of YMC11 transformed with pHZ202 and pHZ206 were grown in 2xYT medium and GS activity was assayed at mid and late exponential phase.

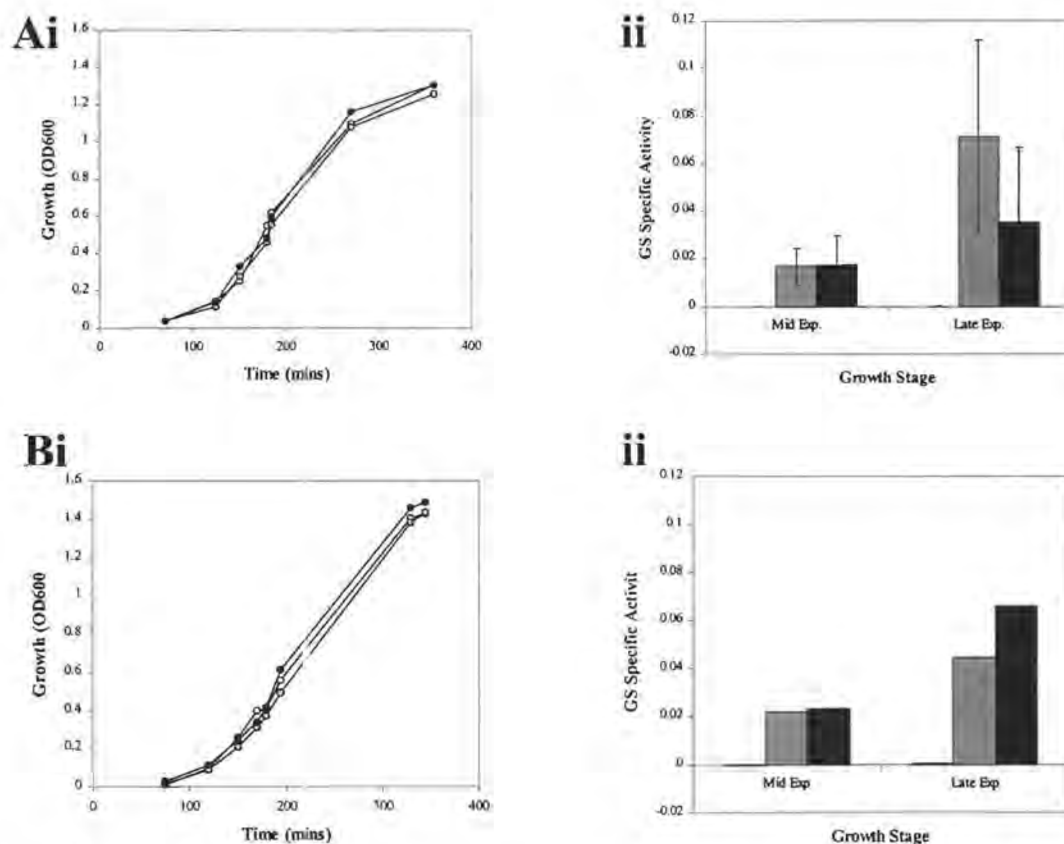


Figure 2.3: The effect of *glnR* expression on growth of and expression of the cloned *C. acetobutylicum* NCP 262 *glnA* in *E. coli* YMC11. Growth in 2xYT supplemented with 0.015% glutamine (Ai) and 2xYT (Bi). Open circles: YMC11; grey circles: YMC11 transformed with pHZ202; black circles: YMC11 transformed with pHZ206. GS activity in cells grown in 2xYT with 0.015% glutamine (Aii) and 2xYT (Bii). Open columns: YMC11; Grey columns: YMC11 transformed with pHZ202; black columns: YMC11 transformed with pHZ206. GS activities expressed as μmol. γ-glutamate released/min/mg protein. Results in A represent an average of three independent experiments, otherwise figures represent results of single experiments.

The effect on culture growth in complete medium of the two constructs was minimal, in both the presence and absence of glutamine (Figure 2.3 Ai and Bi). This indicates that the amount of glutamine present in the complete medium is sufficient to sustain high-level growth of the glutamine auxotroph YMC11. Differences in growth rate due to an uninducible GS (in the absence of *glnR*) would therefore not be expected.

GS activity of transformants harbouring pHZ202 and pHZ206 were assayed and compared (Fig. 2.3 Aii and Bii). During exponential growth in medium with and without glutamine supplementation, there was no noticeable difference in the GS activities of the cultures harbouring the two constructs. However, in late exponential phase, there appeared to be a relative increase in GS activity in pHZ202-containing cells grown in the presence of 0.15% glutamine. In apparent contradiction to this, however, was the observation that in cultures grown without glutamine supplementation, although there appeared to be no difference in growth rates, the GS activities of cells containing pHZ206 were higher than in those harbouring pHZ202. These contradictory observations may have been due to the wide variation observed in GS activity measured in late exponential phase, rather than indicative of regulation of the expression of *glnA*. From these results, there is no evidence to support the role of *glnR* as a positive regulator of the activity of the cloned *C. acetobutylicum* NCP 262 *glnA* gene in *E. coli* YMC11.

An important factor to be considered is that the genes being investigated are being expressed in a distantly-related organism. Although *glnA* is expressed at high levels, which would imply that the promoter is readily recognised by *E. coli* transcription factors, it cannot be assumed that other aspects of regulation are as efficient in the heterologous host, and integration of control mechanisms of the cloned gene with those of the host may not be possible. It was shown that the 6.5-kb fragment of *C. acetobutylicum* NCP 262 DNA was unable to regulate *hut* activity in *E. coli* (Usdin et al., 1986), so the cloned DNA probably does not carry any genes which could function analogously to *ntrB* and *C* in *E. coli*. This implies that if the genes present on the cloned DNA were regulating the cloned GS, as was concluded earlier (Usdin et al., 1986), they were doing so through mechanisms different to those in *E. coli*.

In the system being used to study the regulation of the cloned *glnA* gene, only two components of what may be a more complex system have been included. The amino terminus of the deduced GlnR protein shares sequence identity with response regulators of two-component regulatory systems (Fierro-Monti, PhD thesis; Woods & Reid, 1995). This implies that *glnR* may require a histidine protein kinase for activation. It was shown that *nasT* of *Azotobacter vinelandii*, to which the *C. acetobutylicum* NCP 262 *glnR* is most similar, is itself regulated by *nasS*, although this regulation is negative and does not appear to involve phosphorylation (Guttierez et al., 1995). Host factors in *E. coli* may be unable to fulfil such roles, and this is especially pertinent considering the fact that the regulatory mechanisms of *glnA* expression in the two organisms appear to be fundamentally different.

Initial evidence that the cloned *glnA* gene was subject to regulation in response to nitrogen availability was obtained using a 6.5-kb fragment of DNA (Usdin et al., 1986). This included extensive sequence both up- and downstream of the *glnA* gene. As the constructs used in this study carried only the *glnA* and *glnR* genes, it is possible that as yet unidentified genes, which are essential components of the system, were present on the initial clone, and have been omitted in this study. Also of concern is that the regulation of the cloned *glnA* gene by nitrogen has itself been subject to question, and repetition of this result has so far proved impossible (Fierro-Monti, PhD thesis). It is possible that the perceived nitrogen regulation of the cloned *glnA* gene was in fact due to spurious observations resulting from the wide variation observed in GS activity of *E. coli* YMC11 cells carrying the cloned *C. acetobutylicum* NCP 262 *glnA* gene.

2.3.2 Sequencing of the region upstream of *glnA* in *C. acetobutylicum* NCP 262.

Plasmid pC16 (Stutz, unpublished data) contains a 1.3-kb insert comprising 24 nucleotides of the 5' terminus of *glnA*, and approximately 1.29 kb of the region upstream of *glnA*. The 5' terminus of this clone had previously been sequenced (Stutz, unpublished data), and to facilitate complete sequencing of the clone, the generation of 5' and 3' deletion derivatives was attempted. Although the creation of 3' deletions proved to be unsuccessful and was therefore abandoned, three clones were generated, representing deletions of 0.25, 0.51 and 1.13 kb from the 5' terminus of the clone (Fig. 2.4). These deletions were designated pUG4,

pUG6 and pUG7 respectively, and were sequenced in one direction only, using the universal SK forward sequencing primer.

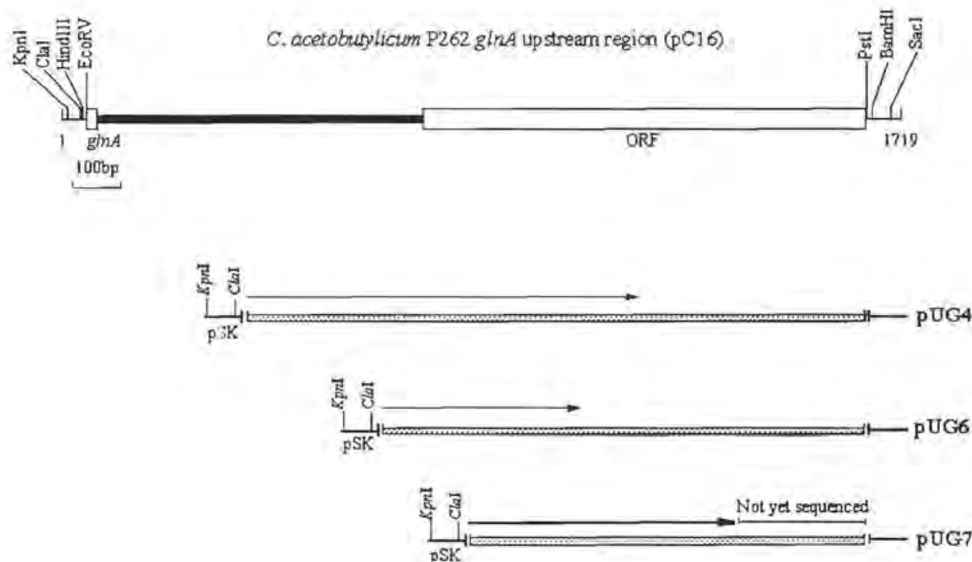


Figure 2.4: Partial map of the region upstream of *glnA* in *C. acetobutylicum* NCP 262, carried by pC16, showing the relative positions and orientations of *glnA* and the ORF proposed to encode an aspartokinase. Also shown are the 5' deletions generated for sequencing and the extent to which the clone was sequenced (arrows).

The resulting sequence data was combined with that previously obtained, and the sequence analysed using the GCG v9.1 package. The nucleotide and deduced amino acid sequences carried by the clone are presented in appendix B.1. A truncated ORF was identified 640bp upstream of the start codon of *glnA*, and transcribed in the opposite direction. The 680 nucleotide sequence was translated and a database search conducted using the blastp program (Antschul et al., 1997) at the BLAST utility of the NCBI. The deduced amino acid sequence was found to share significant identity with aspartokinases and bifunctional aspartokinase/homoserine dehydrogenases (see Table 2.2), which are involved in the biosynthesis of lysine. An alignment of the amino acid sequence of the ORF with the amino acid sequences of the five bacterial proteins to which it showed the highest identity is shown in Figure 2.5.

Table 2.2: Amino acid sequences found to be most similar to that encoded by the ORF upstream of *glnA* in *C. acetobutylicum* NCP 262.

Sequence	Organism	Identity	Similarity	Score	E
Homologue of aspartokinase II α and β subunits	*Bacillus subtilis*	36	53	124	$4.e^{-28'}$
Precursor of monofunctional aspartokinase	*Glycine max*	28	46	65.6	$3.e^{-10'}$
Lysine-sensitive aspartokinase III	*Escherichia coli*	28	43	65.6	$3.e^{-10'}$
Apartokinase III	*Chlamydia pneumoniae*	27	43	64	$9.e^{-10'}$
Aspartate kinase III	*Escherichia coli*	27	42	63.6	$1.e^{-9'}$
Aspartokinase/homoserine dehydrogenase	*Glycine max*	25	42	63.2	$1.e^{-9'}$
Aspartate kinase	*Arabidopsis thaliana*	26	45	62.8	$2.e^{-9'}$
Precursor of bufunctional aspartokinase/homoserine dehydrogenase	*Arabidopsis thaliana*	27	43	60.9	$7.e^{-0'}$
Precursor of chloroplast aspartokinase/homoserine dehydrogenase	*Daucus carota*	25	51	53.9	$2.e^{-8'}$
Aspartokinase (LysC)	*Rickettsia prowazekii*	24	41	58.9	$3.e^{-8'}$

Aspartokinases are responsible for the catalysis of the the ATP-dependent phosphorylation of aspartate to produce β-aspartyl phosphate, which is first step of the shared pathway in the biosynthesis of lysine, threonine and methionine. In both *E. coli* and *B. subtilis*, in which the pathway has been best characterised, there are three different isofunctional enzymes which catalyse the same reaction, but whose activities are regulated by different intermediates in the pathway. However, there are important differences between the organisms with respect to both the structures of the enzymes and the ligands to which each respond (Patte, 1996).

In *E. coli*, aspartokinase I activity, encoded by the *thrA* gene, is inhibited competitively by threonine and isoleucine, and has been found to be located on the same polypeptide as a homoserine dehydrogenase activity. The bifunctional enzyme exists as a tetramer, with the two enzyme activities located in distinct domains on each subunit. Homoserine dehydrogenase I activity is also inhibited by threonine and isoleucine through an allosteric mechanism. Homoserine dehydrogenase catalyses the first step in the threonine biosynthesis

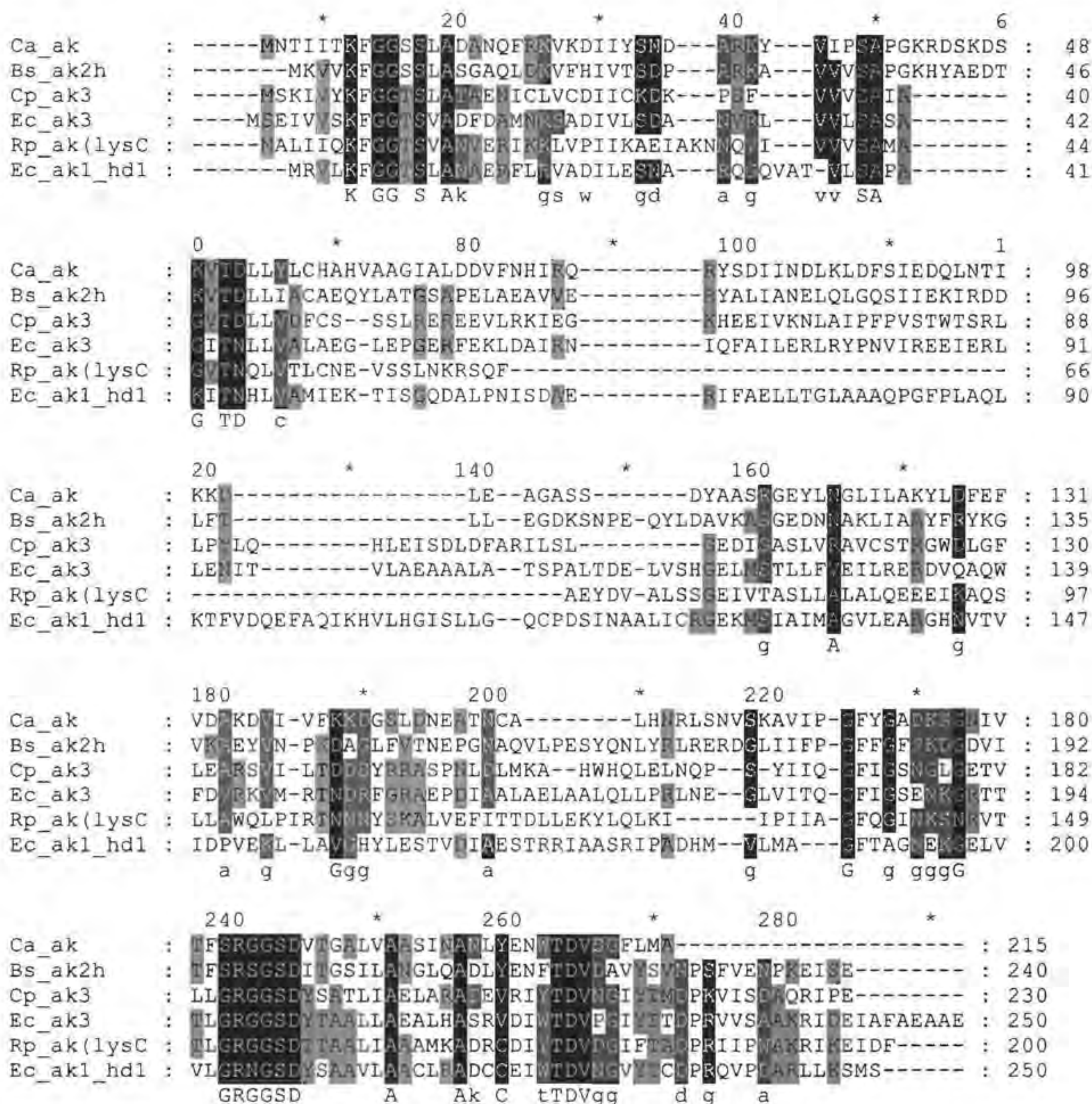


Figure 2.5: Multiple sequence alignment of amino acid sequence of proposed aspartokinase (Ca_ak) encoded by the incomplete ORF upstream of *glnA* in *C. acetobutylicum* NCP 262. Bs_ak2h: *B. subtilis* aspartokinase II homologue; Cp_akIII: *Chlamymia pneumoniae*; Ec_akIII: *E. coli* aspartokinase III; Rp_ak: *Rickettsia prowazekii*; Ec_ak1/hd1: *E. coli* bifunctional aspartokinaseI/homoserine dehydrogenase I. Conserved and similar resides are indicated by shading, and consensus sequences are indicated below the alignment.

pathway, which branches from the shared pathway of threonine, lysine and methionine biosynthesis, a reaction which entails the NADPH-dependent conversion of aspartate semialdehyde to homoserine. A second aspartokinase activity (aspartokinase II) is also located on the same polypeptide as a homoserine dehydrogenase activity, and this bifunctional

enzyme is encoded by the *metL* gene. This enzyme is not subject to inhibition by any ligand, but its expression is repressed by methionine through the action of the *metJ* transcriptional repressor. The enzyme exists as a dimer. The third aspartokinase activity (aspartokinase III) is encoded by the *lysC* gene, and its activity is inhibited by lysine leucine, isoleucine and phenylalanine. This is monofunctional enzyme also exists as a dimer (Patte, 1996).

In *B. subtilis*, all three aspartokinases are monofunctional enzymes, and there is only one, monofunctional, homoserine dehydrogenase (Paulus, 1993). Aspartokinase I is encoded by the *dapG* gene, and its activity is repressed by diaminopimelate, but not by any other amino acids. This is indicative of a role in providing intermediates for the synthesis of *meso*-diaminopimelate, which is an important for the production of cell wall and spore cortex peptidoglycan, as well as dipicolinate, which is an important constituent of the spore. This is important, as diaminopimelate cannot be obtained from protein turnover, which is the source of most amino acids in stationary phase and sporulation. Aspartokinase II activity, encoded by the *lysC* gene, is inhibited by lysine. Enzyme activity is also inhibited under conditions of glucose starvation and at stationary phase, circumstances which also repress its expression. This may reflect a role in providing amino acids for protein synthesis. Aspartokinase II consists of two different polypeptides, the α and β subunits. The α subunit is encoded by the *lysC* gene, while the β subunit is produced by in-frame initiation of translation within the *lysC* mRNA. The β subunit thus corresponds to the C terminal of the α subunit (Paulus, 1993). Aspartokinase III activity is inhibited synergistically by threonine and lysine, which also probably reflects a role in providing intermediates for protein synthesis (Paulus, 1993).

Aspartokinase enzymes consist of two distinct domains, while bifunctional aspartokinase/ homoserine dehydrogenases consist of three. The aspartokinase activity is located in the first domain, which generally extends from the N terminus to amino acid 290. The second starts at residue 310 and extends to residue 460. This domain is responsible for the ligand-responsive inhibition of the aspartokinase activity, and in the case of bifunctional enzymes, the homoserine dehydrogenase activity as well. In monofunctional enzymes, the end of this domain marks the C terminus of the polypeptide. The homoserine dehydrogenase activity resides in the third domain, which extends from residue 470 to the C terminus of bifunctional

enzymes, while in monofunctional homoserine dehydrogenases, a fourth domain at the C terminus is necessary for ligand-responsive regulation (Patte, 1996).

The high amino acid sequence identity shared by the ORF upstream of *glnA* and the N termini of these enzymes suggests that this ORF also encodes an aspartokinase or bifunctional aspartokinase/homoserine dehydrogenase. As the sequence of the ORF from *C. acetobutylicum* NCP 262 is incomplete, and spans only the N-terminus of the aspartokinase domain, it is impossible to tell whether it encodes a monofunctional or bifunctional enzyme. However, the fact that *B. subtilis* possesses no bifunctional aspartokinase/homoserine dehydrogenases suggests that the ORF is likely to encode a monofunctional enzyme. It is tempting to speculate, that this ORF, as well as the adjacent *glnA*, *glnR*, *gltA* and *gltB* genes form part of a cluster of amino acid biosynthesis genes. It would also appear that the divergently transcribed ORF marks the 5' boundary of the *C. acetobutylicum* NCP 262 *glnA* operon. This complicates the search for an accessory factor of *glnR*, responsible for regulating its function, and necessitates the employment of alternative genetic strategies to achieve this end.

2.3.3 Growth of *C. beijerinckii* NCIMB 8052 using various nitrogen sources

In order to evaluate the suitability of *C. beijerinckii* NCIMB 8052 for use as a model for the study of GS regulation in *C. acetobutylicum* NCP 262, it was necessary to establish whether the mechanisms of nitrogen assimilation used by the two strains were similar, at the physiological as well as the genetic level. The first aspect of nitrogen metabolism to be studied in *C. beijerinckii* NCIMB 8052 was whether organic or inorganic nitrogen sources were favoured. Cultures of *C. beijerinckii* NCIMB 8052 were grown in complete media as well as in minimal media supplemented with organic nitrogen in the form of casamino acids at a concentration of 0.05%, representing conditions of nitrogen limitation, and 0.2%, representative of conditions of nitrogen excess. These media were used both with and without glutamine supplemented at a concentration of 0.015%. Cultures were also grown in minimal medium containing inorganic nitrogen in the form of ammonium sulphate or ammonium acetate. Once again, a concentration of 0.2% was chosen to represent nitrogen excess, while nitrogen limitation was represented by a concentration of 0.025%. In addition, the effect of using both inorganic and organic nitrogen sources was tested by supplementing minimal

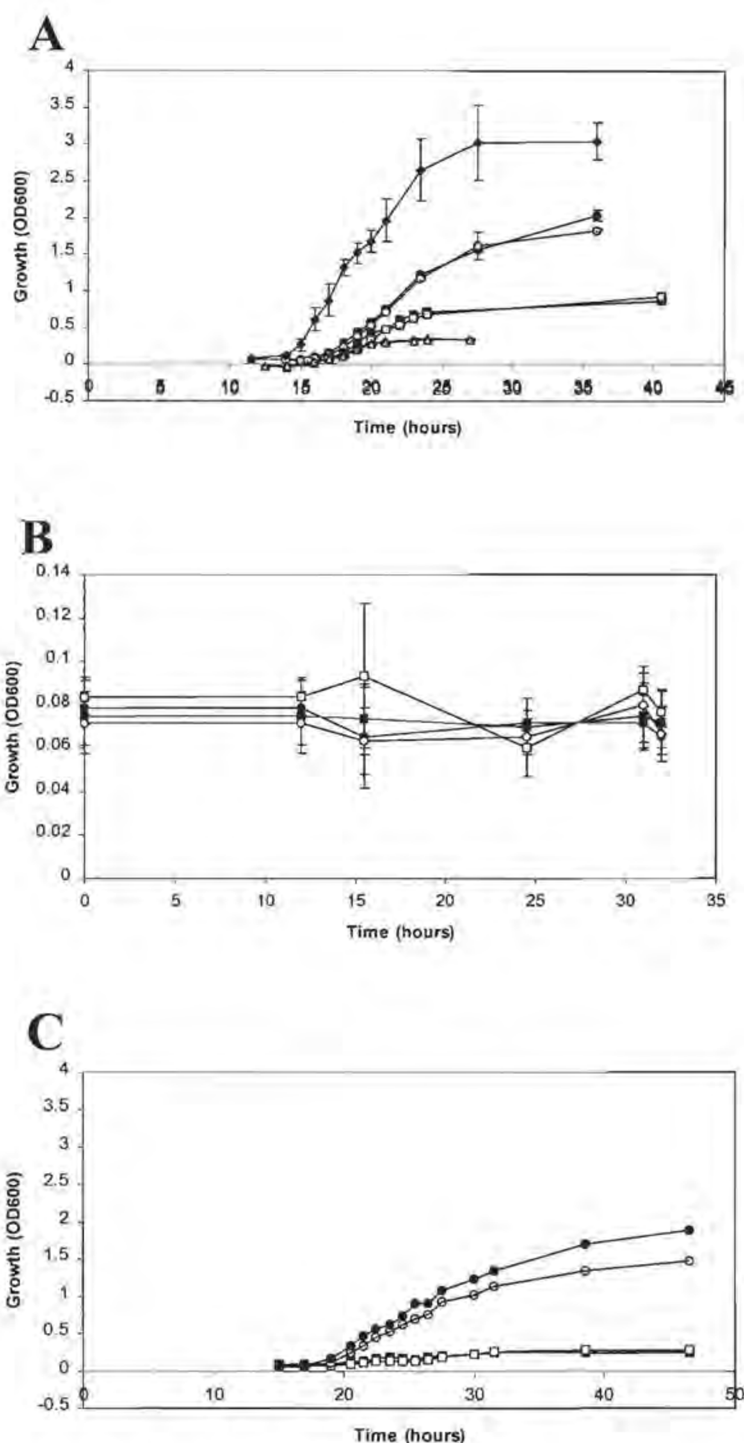


Figure 2.6: Growth of *C. beijerinkii* NCIMB 8052 under different conditions of nitrogen availability: (**A**) Complete medium (CBM; black diamonds) and minimal medium (GSMM) supplemented with different concentrations of organic nitrogen in the form of casamino acids (Caa): Black circles: 0.2% Caa; grey circles: 0.2% Caa + 0.015% gln; black squares: 0.05% Caa; grey squares: 0.05% Caa + 0.015% gln; black triangles: 0.025% Caa; grey triangles: 0.025% Caa + 0.015% gln. (**B**) GSMM supplemented with inorganic nitrogen sources: Black circles: 0.2% ammonium acetate; white circles: 0.025% ammonium acetate; black squares: 0.2% ammonium sulphate; white squares: 0.025% ammonium sulphate. (**C**) GSMM supplemented with organic and inorganic nitrogen sources: black circles: 0.2% Caa + 0.025% ammonium acetate; white circles: 0.2% Caa + 0.025% ammonium sulphate; black squares: 0.025% Caa + 0.015% ammonium acetate; white squares: 0.025% Caa + 0.025% ammonium sulphate. All values represent averages for three independent experiments.

medium containing excess and limiting concentrations of casamino acids with ammonium sulphate or ammonium acetate at concentrations of 0.025%. The ability of these media to sustain growth of *C. beijerinckii* NCIMB 8052 is shown in Figure 2.6.

These results show that the level of growth obtained by the cultures is highly dependent on the concentration of casamino acids. It also shows that, as is the case with *C. acetobutylicum* NCP 262 (Brown, unpublished results), organic nitrogen in the form of casamino acids is preferred as a nitrogen source over inorganic nitrogen in the form of either ammonium acetate or ammonium sulphate. Indeed, it appears that inorganic nitrogen has an inhibitory effect on growth of *C. beijerinckii* NCIMB 8052 (Fig 2.6C). It was previously observed that in *C. acetobutylicum* NCP 262, the preferred nitrogen source was casamino acids, and that inorganic nitrogen had an inhibitory effect on growth (Brown, unpublished results). *C. beijerinckii* NCIMB 8052, therefore, is similar to *C. acetobutylicum* NCP 262 in its preference for casamino acids as a nitrogen source.

The preference of these organisms for organic nitrogen is interesting, as it represents a divergence from the more common situation amongst bacteria of a preference for ammonium (Merrick & Edwards, 1995), and may represent an adaptation of their physiology to their environment. Casamino acids consist of a complex mixture of amino acids and peptides obtained through the hydrolysis of casein. It would be interesting, therefore, to investigate which of its constituents in particular are utilized, and whether the pathways through which they are assimilated and metabolised are similar to those known to be used for the metabolism of amino acids by other bacteria when inorganic nitrogen is scarce.

This observation is somewhat reflective of the preference of *B. subtilis* for glutamine as a nitrogen source, however this organism is also readily able to utilise ammonium (Fisher, 1999). It is apparent that nitrogen regulation in *B. subtilis* is considerably different to that of enteric bacteria, and the reasons for this may be related to the differences in physiology between *B. subtilis* and the Enterobacteriaceae.

2.3.4 Regulation of GS activity in *C. beijerinckii* NCIMB 8052

The regulation of GS in *C. beijerinckii* NCIMB 8052 was investigated by assaying GS activity of cultures grown in the following media: Complete medium; nitrogen-limiting minimal medium with and without glutamine; nitrogen-rich minimal medium, also with and without glutamine. GS activity was measured at early, mid and late exponential growth phases. The results are shown in Figure 2.7.

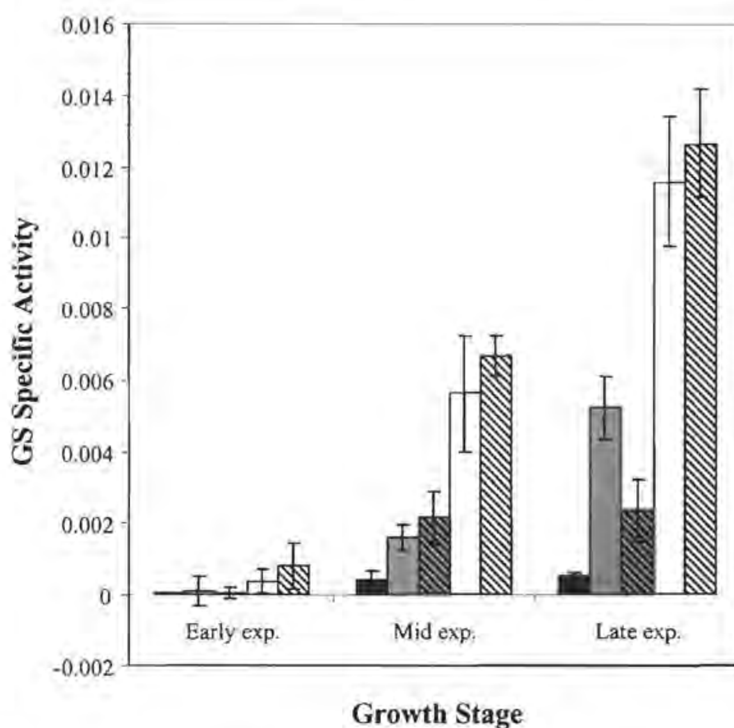


Figure 2.7: Regulation of GS activity in *C. beijerinckii* NCIMB 8052 in response to the availability of organic nitrogen in the form of casamino acids. Black columns: complex medium; grey columns: 0.2% casamino acids; grey striped columns: 0.2% casamino acids supplemented with 0.015% glutamine; White columns: 0.05% casamino acids; white striped columns: 0.05% casamino acids supplemented with 0.015% glutamine. Samples were taken at early mid and late exponential phase. GS specific activity is expressed as μM γ-glutamate released/min/mg protein. Results represent the averages of three independent experiments.

Strong repression of GS was observed in complete media, with repression lifted slightly in late exponential growth, presumably as a result of nitrogen depletion of the media. Repression, although not as strong as that observed in complete medium, occurred in nitrogen-rich minimal medium, and this was also lifted during late exponential growth. In nitrogen-limiting minimal medium, GS activity was as much as three-fold higher in mid-

exponential phase, and five-fold higher in late exponential phase, than in nitrogen-rich media. It is probable that even higher GS activities could have been measured in minimal medium containing 0.025% casamino acids. This medium was, however, unable to sustain growth to absorbances higher than 0.5, which is several fold lower than the growth observed in complete and nitrogen-rich minimal media. Thus, in order to allow more comparable growth, a concentration of 0.05% of casamino acids was used.

In mid-exponential phase, the inclusion of 0.015% glutamine in the medium resulted in slightly higher levels of GS activity under conditions of both nitrogen limitation and excess. At early and late exponential phase, the same was true for nitrogen-limited growth, but not for growth in nitrogen-rich media. It is uncertain whether these differences are physiologically relevant, and it cannot be concluded that the presence of glutamine has a significant effect on GS induction.

In *C. acetobutylicum* NCP 262, it was observed that the greatest induction of GS occurred in minimal medium supplemented with 0.025% casamino acids and 0.015% glutamine, and repression occurred in complete medium and in minimal medium supplemented with 0.2% casamino acids (Brown, unpublished data). With respect to the regulation of GS activity, therefore, it appears that there are similarities between *C. acetobutylicum* NCP 262 and *C. beijerinckii* NCIMB 8052.

2.3.5 Detection of *C. beijerinckii* NCIMB 8052 nitrogen region

C. acetobutylicum NCP 262 and *C. beijerinckii* NCIMB 8052 chromosomal DNA was digested with the restriction endonucleases *Bgl*II, *Eco*RI and *Pvu*II. Results of Southern hybridisation using internal fragments of the *glnA*, *glnR* and *gltA* genes from *C. acetobutylicum* NCP 262 are shown in Figure 2.8. The sizes of *C. acetobutylicum* NCP 262 DNA fragments to which the probes were expected to hybridise were calculated from a restriction map generated using previously obtained sequence and restriction mapping data for this organism (Fig. 2.1). In *C. beijerinckii* NCIMB 8052, all three probes appeared to hybridise to the same *Bgl*II fragment, estimated to be between 5 and 8 kb in size. Similarly, the probes for both *glnA* and *glnR* hybridised to the same 3.8-kb *Eco*RI fragment. Digestion with *Pvu*II yielded a different hybridisation pattern for the *glnA* and *glnR* probes. The *glnA*

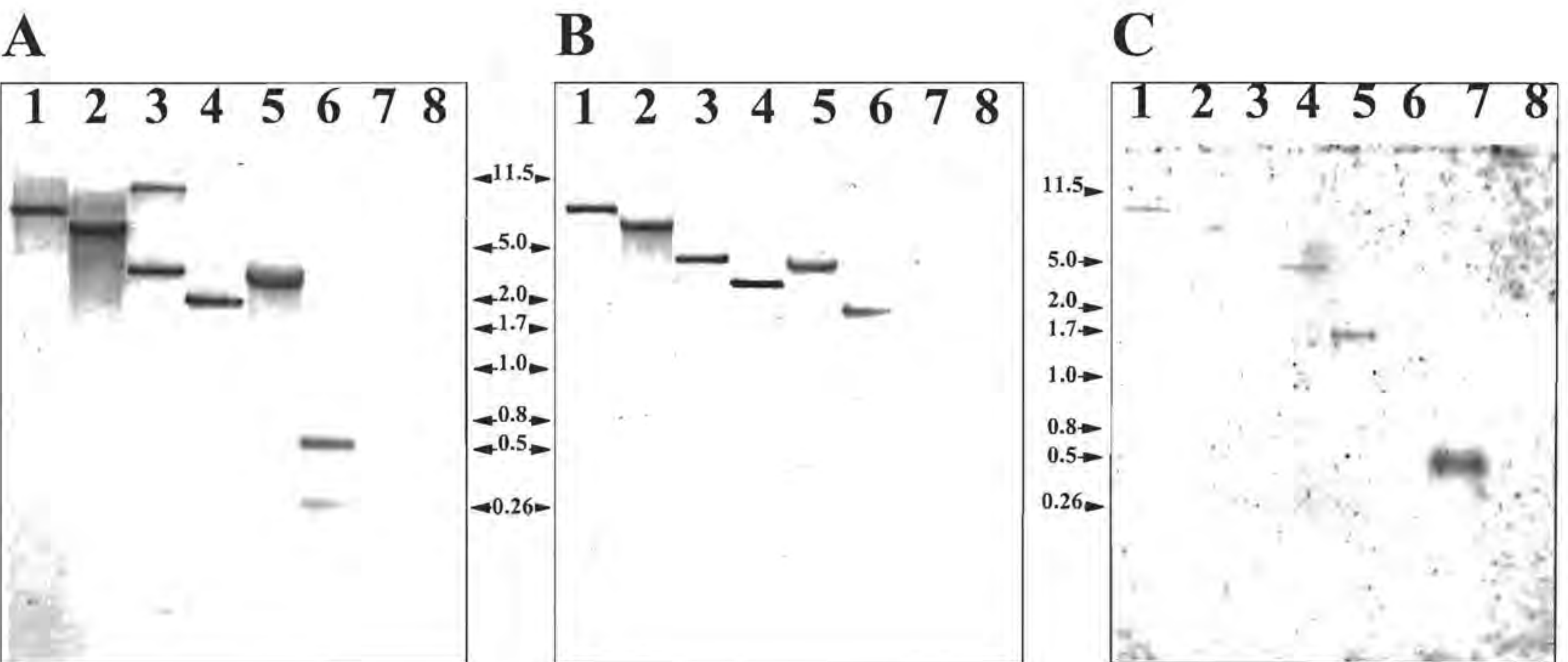


Figure 2.8: Southern hybridisation analysis of *Clostridium acetobutylicum* NCP 262 and *Clostridium beijerinckii* NCIMB 8052. In (A), the probe used was specific for *glnA* while in (B), the probe was specific for *glnR*. For both (A) and (B): Lane 1: *C. acetobutylicum* NCP 262, *Bgl*II; Lane 2: *C. beijerinckii* NCIMB 8052, *Bgl*II; Lane 3: *C. acetobutylicum* NCP 262, *Pvu*II; Lane 4: *C. acetobutylicum* NCP 262, *Eco*RI; Lane 5: *C. beijerinckii* NCIMB 8052, *Eco*RI; Lane 6: *C. beijerinckii* NCIMB 8052, *Pvu*II; Lane 7: Lambda marker; Lane 8: *E. coli*, *Eco*RI. The probe used in (C) was specific for *gltA*. Lane 1: *C. acetobutylicum* NCP 262, *Bgl*II; Lane 2: *C. beijerinckii* NCIMB 8052, *Bgl*II; Lane 4: *C. acetobutylicum* NCP 262, *Pvu*II; Lane 5: *C. beijerinckii* NCIMB 8052, *Pvu*II; Lane 7: Lambda marker; Lane 8: *E. coli*, *Eco*RI.

probe hybridised to a 260bp and a 600bp band, while the *glnR* probe hybridised to a single band of approximately 1.7 kb. The *gltA* probe, however, appeared to hybridise to the same 1.7-kb fragment. The fact that many of the probes hybridised to the same restriction fragments of *C. beijerinckii* NCIMB 8052 indicates that they are located close together. Thus, although the restriction pattern observed for *C. beijerinckii* NCIMB 8052 was different to that of *C. acetobutylicum* NCP 262, it indicated a close linkage, and therefore a similar arrangement of the respective genes on the chromosomes of the two species.

The low hybridisation temperature used (37^{0}C) would be expected to allow a degree of cross-hybridisation between dissimilar sequences. Despite this, no hybridisation was observed between the probe and *Eco*RI-digested *E. coli* DNA. In the case of *gltA*, however, the probe hybridised to the λ DNA marker. This may have been due to contamination of the probe with pBluescript vector DNA. Because both the *glnA* and *glnR* probes appeared to hybridise to a single 3.8-kb *Eco*RI fragment of *C. beijerinckii* NCIMB 8052 DNA, this enzyme was chosen for cloning of the region.

2.3.6 Cloning of the *C. beijerinckii* NCIMB 8052 nitrogen region

Of the 900 colonies streaked out and used for colony hybridisation, only one carried a plasmid to which the *glnA* probe (see previous section) hybridised strongly, although weak hybridisation to DNA from several other clones was observed. The clone showing strong hybridisation, as well as three other clones were selected and designated clones 1-4. DNA was extracted from each, as well as two of their respective neighbouring colonies, digested with *Eco*RI, and subjected to agarose gel electrophoresis and Southern hybridisation (Fig. 2.9). Although the probe hybridised to the 3.0-kb vector band, as well as to a band of approximately 1.6 kb, it was observed that especially strong hybridisation occurred between the 3.8-kb band released by digestion of clone 3 with *Eco*RI.

The 5' terminus of clone 3 was sequenced to determine whether it was indeed carrying the correct fragment. Comparison of these sequences to the sequence of the *C. acetobutylicum* NCP 262 nitrogen region revealed a 67.1% nucleotide sequence identity between the 5' terminus of clone 3 and the 5' region of the nitrogen region of *C. acetobutylicum* NCP 262. It

was concluded that clone 3 did indeed carry the correct DNA fragment and it was designated pCBN1.

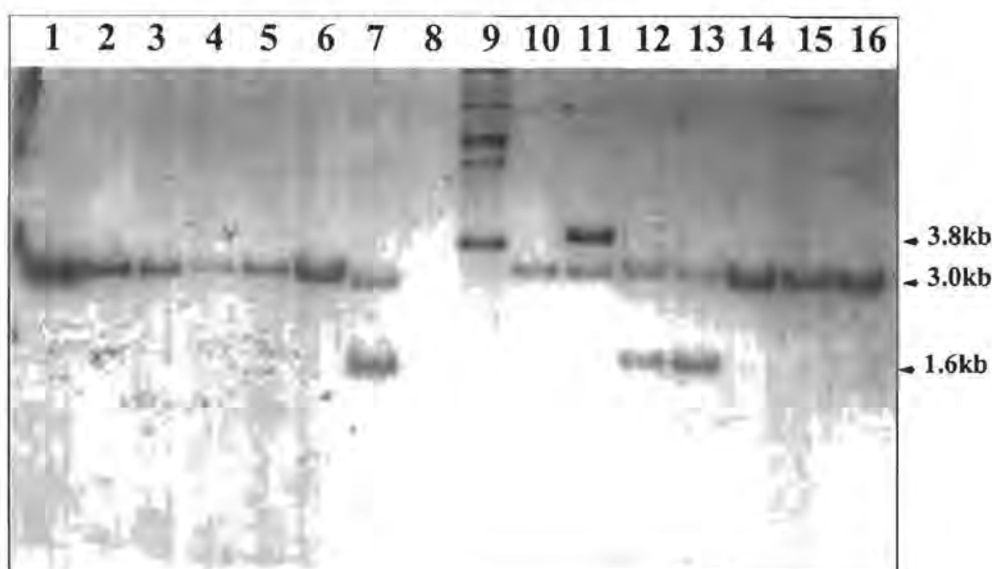


Figure 2.9: Southern hybridisation of clones identified by colony blotting. Lane 1: pSK; Lane 2: Clone 1; Lane 3: Clone 1_{ni}; Lane 4: Clone 1_{nii}; Lane 5: Clone 2; Lane 6: Clone 2_{ni}; Lane 7: Clone 2_{nii}; Lane 8: Lambda Marker; Lane 9: pHZ208; Lane 10: pSK; Lane 11: Clone 3; Lane 12: Clone 3_{ni}; Lane 13: Clone 3_{nii}; Lane 14: Clone 4; Lane 15: Clone 4_{ni}; Lane 16: Clone 4_{nii}; Lane 17: Lambda Marker. All plasmids were digested with *Eco*RI and probed with the *glnA* probe used previously. Subscripts refer to neighbour colonies to those showing signals in the colony blot experiment.

2.3.7 Complementation of glutamine auxotrophy in *E. coli* YMC11

The *E. coli glnAntrBC*⁻ mutant strain YMC11 was transformed with pCBN1 and pBluescriptSK+ and grown on minimal medium agar in the presence and absence of glutamine (Fig. 2.10). After two days' incubation at 37^{0}C, it was observed that while YMC11(pSK) was unable to grow in the absence of glutamine, glutamine auxotrophy was complemented by the presence of pCBN1. This serves to indicate that this plasmid carries a gene encoding a functional GS that is efficiently expressed in *E. coli*. A simple experiment to confirm this was conducted. Cultures of *E. coli* YMC11 transformed with either pSK or pCBN1 were grown in complete medium and assayed at mid-exponential phase for GS

activity. While GS activity was undetectable in cells harbouring pSK, high levels of GS activity (0.063μmol γ-glutamate released/min/μg protein) were observed in the cells carrying pCBN1 illustrating that complementation of glutamine auxotrophy in this strain was due to the presence of a functional GS.

Figure 2.10: Complementation of glutamine auxotrophy in *E. coli* YMC11 by pCBN1. *E. coli* YMC11 transformed with either pSK or pCBN1 was streaked out onto M9 minimal medium agar containing ampicillin at a concentration of 100μg/ml, with and without glutamine at a concentration of 0.2%.

2.3.8 Sequencing of the *C. beijerinckii* NCIMB 8052 nitrogen region

The entire nucleotide and deduced amino acid sequences of the region of the *C. beijerinckii* NCIMB 8052 DNA carried by pCBN1 are shown in Appendix B.2.

Open Reading Frames

Sequence analysis of the *Eco*RI insert carried by pCBN1 showed the presence of two complete and one incomplete open reading frames, all of which read in the same direction. The first ORF, beginning 794bp from the 5' terminus of the insert, was found to share 87.8% nucleotide sequence identity with the *C. acetobutylicum* NCP 262 *glnA* gene. It was concluded that this ORF encoded the *C. beijerinckii* NCIMB 8052 homologue of *glnA*, and was therefore designated *glnA*. Downstream of *glnA*, a second ORF was identified which shared 86.5% nucleotide sequence identity with the *glnR* gene of *C. acetobutylicum* NCP 262, and was therefore designated *glnR*. Alignment of the amino acid sequence of *glnR* of *C. beijerinckii* NCIMB 8052 with that of *C. acetobutylicum* NCP 262, as well as two other genes is shown in Figure 2.11. A third, incomplete, ORF was found downstream of *glnR*. The portion of sequence present was found to share 86.8% nucleotide sequence identity with the corresponding portion of *gltA* of *C. acetobutylicum* NCP 262, and was named *gltA*. The truncated ORF comprises 920bp, encoding 306 amino acids (as well as two bases of an incomplete codon at the 3' terminus).

Control regions

Overall, the region comprising 300 bases upstream of the translational start of *glnA* in the two strains shares a nucleotide sequence identity of 64%, which is markedly lower than the 86.5% to 87.8% identity observed for the coding regions of the genes studied. Similar -10 and -35 regions are present at approximately the same distances upstream of the translational start of the two genes. Additionally, there is a region between promoter p_1 and the translational start of the *C. beijerinckii* NCIMB 8052 *glnA* gene which is able to form a terminator-like structure, as is the case in *C. acetobutylicum* NCP 262. Using the program RNAstructure v2.0 (Matthews et al., in press), the free energy of this structure was calculated to be -20.4 kcal/mol. This figure is lower than that predicted for the same structure in *C. acetobutylicum* NCP 262 (-16.9 kcal/mol) and compares favourably with the free energy of -25.8 kcal/mol of the terminator structure known to be involved in antitermination control of the *nasF* operon in

K. oxytoca M5a1 (Chai & Stewart, 1999). Features of the promoter region of *C. beijerinckii* NCIMB 8052 will be dealt with in greater detail in section 2.3.10.

```
                    *        20         *        40         *        60
cb_glnr : ----------------------MAQDGR                                   :  40
ca_glnr : --------------------MVQSEKGK                                   :  42
kp_nasr : MNNMAGNTPEVVDWFRRARRLQKQQLHQ                                   :  62
av_nast : ---------------------------M                                   :  27
                                         na   Gg g       G     A C gGr

                  *        80         *       100         *       120
cb_glnr :                        -------------------------------------- :  63
ca_glnr :                        -------------------------------------- :  65
kp_nasr :                        PARDAASSALCWRIACAVWYLPQLAALRKACATGRLPP : 124
av_nast :                        ---------------------------------LIDTE :  55

                *       140         *       160         *       180
cb_glnr : -------------------------------------------------------------- :   -
ca_glnr : -------------------------------------------------------------- :   -
kp_nasr : KRPPGSLAVIRHLLNRTAAQRQHRRSANRRPHGCALQLYAGKRAGRTGARAGRAGFARGQFS : 186
av_nast : SPGRDVMEQVVLVSRDRPRPIVMFTDEHDPQVMRQAIQAGVSAY------------------ :  99

              *       200         *       220         *       240
cb_glnr : -------------------------------------------------------------- :   -
ca_glnr : -------------------------------------------------------------- :   -
kp_nasr : DELRQQLVDRIDGQQPCFDSFQALAQPPQTALFAEQCQASLEIEQLRRVACTRQPPADEGET : 248
av_nast : -------------------------------------------------------------- :   -

            *       260         *       280         *       300         *
cb_glnr : ---------------                        FIEE                    : 111
ca_glnr : ---------------                        FIEE                    : 113
kp_nasr : ALRWFCAQTQRLEQLRGVEELLIVDLLNAADALLEGEEPEAH                    : 310
av_nast : ------------------------------------------                    : 113
                          gv     r     a a          r   t  gtG  ar  t

                320         *       340         *       360         *
cb_glnr :                                                                : 173
ca_glnr :                                                                : 175
kp_nasr :                                                                : 372
av_nast :                                                                : 175

              380         *
cb_glnr :                      : 186
ca_glnr :                      : 188
kp_nasr :                      : 393
av_nast :                      : 192
```

Figure 2.11: Multiple sequence alignment of the deduced amino acid sequences of *glnR* of *C. beijerinckii* NCIMB 8052 (cb_glnr) with those of *glnR* of *C. acetobutylicum* NCP 262 (ca_glnr), *nasR* of *Klebsiella oxytoca* M5a1 (kp_nasr) and *nasT* of *Azotobacter vinelandii* (av_nast).

2.3.9 Mutation of *glnR* in *C. beijerinckii* NCIMB 8052

The mutagenesis vector pIO2 was used to transform the *E. coli* strain CA448. The vector was then transferred into *C. beijerinckii* NCIMB 8052 by conjugation. The plasmid is unable to replicate in *Clostridium* species, and will therefore only persist if it integrates into the chromosome. As it carries an erythromycin resistance gene, transconjugants possessing a plasmid which has integrated into their chromosomes can therefore be selected on erythromycin at a concentration of 10μg/ml. As a positive control for the conjugation experiment, the *E. coli-Clostridium* shuttle vector pCTC1 was used. This plasmid is able to replicate in *Clostridium*, and its persistence in *C. beijerinckii* NCIMB 8052 is therefore not dependent on integration. The plasmid pMTL30, which cannot be maintained in *Clostridium*, was used as a negative control. Numbers of erythromycin-resistant *C. beijerinckii* NCIMB 8052 colonies obtained using this pCTC1 were used as an estimate of the number of conjugation events for pIO2.

Table 2.2: Results of a typical conjugation experiment

Plasmid	Donor count (cfu/ml)	Transconjugant count (cfu/ml)	Transfer efficiency (#transconjugants recovered/donor)
pCTC1	$4.35x10^8$	$3.64x10^8$	0.84
pIO2	$3.18x10^9$	0	ND
pMTL30	$1.65x10^8$	0	ND

Although a large number of conjugation events occurred for pIO2, as estimated by the conjugation efficiency of transfer of pCTC1, no erythromycin-resistant colonies were observed in the case of pIO2. As there was no direct evidence for pIO2 being transferred into *C. beijerinckii* NCIMB 8052, the possibility that this did not in fact occur cannot be ruled out. Efficient transfer between *E. coli* strains CA448 and JM109 was observed, however (data not shown), so this is unlikely.

It is possible that by inactivating *glnR*, *glnA* would be rendered uninducible, resulting in a mutant with a glutamine auxotrophic phenotype, thus making its recovery difficult. To preclude this, however, the medium on which transconjugants were selected was supplemented with glutamine at a concentration of 0.2%. It is conceivable, however, that the uptake of glutamine may require a transport mechanism which may also depend on *glnR*, and

would therefore also be rendered uninducible by its inactivation. Alternatively, *glnR* may be an essential gene with functions beyond merely regulating *glnA* expression, and its inactivation may have been lethal. However, its arrangement in relation to *glnA* argues against a more global role. Yet another possibility is that antibiotic resistance conferred by one or several integrations of pIO2 is not strong enough to allow differentiation between mutants and breakthrough growth of unmutated cells, which was observed three to four days after plating onto selective media.

2.3.10 Determination of the start of transcription of *glnA* of *C. beijerinckii* NCIMB 8052

In order to determine the start of transcription of *glnA* in *C. beijerinckii* NCIMB 8052 under conditions of nitrogen excess and limitation, primer extension experiments were performed using RNA extracted from cultures grown under these conditions. In order to ensure correct regulation of GS activity in the cultures to be assayed, GS assays were performed on them. In cultures grown on 0.2% casamino acids (high nitrogen), GS activity was 0.00063μmol γ-glutamate released/min/μg protein, while in cultures grown on 0.05% casamino acids (low nitrogen), it was measured to be 0.0083μmol γ-glutamate released/min/μg protein. These values were similar to those observed previously (see Fig 2.7), demonstrating that GS activities were appropriate for the experiment.

Results of the primer extension experiment are shown in Figures 2.12 and 2.13. Under conditions of nitrogen excess as well as nitrogen limitation, there appeared to be four transcriptional start sites, which were designated t_1 to t_4, in order of decreasing amount of product obtained. At nucleotides -91 and -37 relative to the *glnA* start codon were start sites t_4 and t_3, respectively, which corresponded roughly to the positions the previously identified promoters *p1* and *p2* in *C. acetobutylicum* NCP 262 (Janssen et al., 1990). In this case, putative promoter elements had been identified by sequence inspection, and had been confirmed by results of primer extension experiments on cultures grown in CBM (Janssen et al., 1990). Transcription from these start sites was also observed for the cloned *C. acetobutylicum* NCP 262 *glnA* gene in *E. coli* YMC11 (Janssen et al., 1990), as well as from other sites less than 10 nucleotides away. It is possible that major transcriptional start sites

Figure 2.12: Graphical representation of the results of the primer extension experiment. The upper graph represents the cDNA generated from cultures grown in 0.2%casamino acids (red curve) and 0.05% casamino acids (blue curve). Also shown are the products of the sequencing reactions as described in the text, using dideoxythymidine (black curve) and dideoxyadenine (green curve). The lower graph shows the results of the sequencing reaction described in the text. The regions of sequence flanking the transcriptional start sites are presented as large boxed text with the transcriptional start sites underlined, with the corresponding regions in the sequence output data, shown below the curves, also boxed. As it was generated using a reverse primer, the reverse complementary sequence is shown here.

corresponding to t_1 and t_2 in *C. beijerinckii* NCIMB 8052 may be present in *C. acetobutylicum* NCP 262. However, they may not have been detected in previous primer extension experiments (Janssen et al., 1990), owing to lower reverse transcriptase processivity, or secondary structure in the leader which could have inhibited the reverse transcriptase.

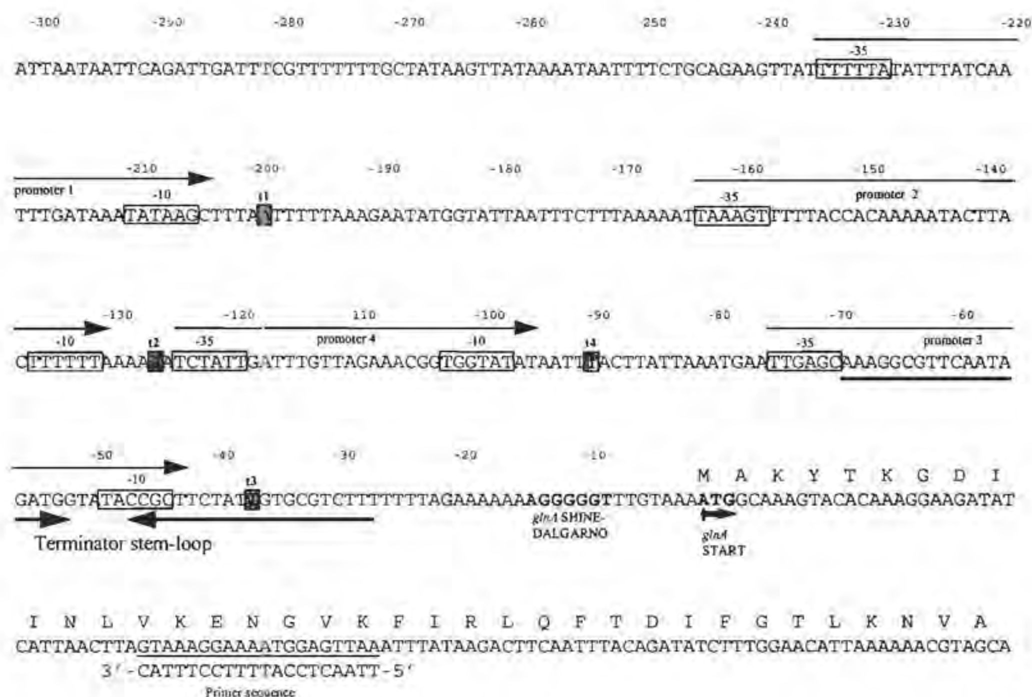


Figure 2.13: The 5' terminus and region upstream of *glnA* in *C. beijerinckii* NCIMB 8052. The positions of the four transcriptional start points are shown, as well as putative promoter regions (putative -10 and –35 sites are boxed and labelled). Also shown is the stem of the putative terminator structure (bold arrows) and the position of the reverse primer.

Under conditions of nitrogen limitation, more primer extension product was obtained from transcripts from all start sites than under nitrogen-excess conditions, indicating that transcription from all four promoters is increased under conditions of nitrogen limitation. It is possible that one or more of these start sites are artefacts resulting from altered processivity of the reverse transcriptase due to the extensive secondary structure of the region (see next section) . These results must therefore be confirmed using an alternative technique. Reporter assays using different portions of the promoter would serve to confirm their roles in transcription initiation and regulation.

A

Terminator
ΔG=-40.4 kCal/mol

Antiterminator
ΔG=-27.2 kCal/mol

B

Terminator
ΔG=-24.5 kCal/mol

Antiterminator
ΔG=-12.0 kCal/mol

C

Terminator
ΔG=-20.4 kCal/mol

Antiterminator
ΔG=-6.2 kCal/mol

Figure 2.14: Secondary structures formed by *glnA* transcripts t_1 (A), t_2 (B) and t_3 (C). Black lines indicate regions involved in terminator formation, while grey lines indicate regons involved in antiterminator formation. Free energies are those for the entire structure as shown, except (A). In (A), the boxed structures represent features which differ between terminator and antiterminator. The unboxed structures remain the same for terminator and antiterminator configuration.

Regulatory features of the *glnA* leader sequence

The apparent presence of four transcriptional start points upstream of *glnA* presents several problems for the generation of a model of regulation. The antitermination control model outlined in chapter 1 (Fierro-Monti, Ph.D thesis; Woods & Reid, 1995) is largely dependent on the structures formed by mRNA resulting from transcription from p_1, the more distal of two putative promoters identified upstream of *glnA* on the basis of sequence inspection and the results of primer extension experiments (Janssen et al., 1990).

Secondary structures formed by transcripts initiated at each of the four start sites were predicted using the RNAstructure v2 program. The proposed secondary structures are shown in Figure 2.14. Still present in transcripts t_1, t_2 and t_4 is the terminator structure, however sequences upstream of it, which are capable of forming additional secondary structures, are variable. Possible antiterminator structures, which all sequester residues necessary for terminator formation, are present in all three leaders. These structures are the same for t_1 and t_3, but t_2 forms a different structure. These different antiterminator structures may be involved in the recognition and binding of an antiterminator protein. It is possible, however, that primary, rather than secondary sequence features are recognised. Furthermore, it is also possible that the antiterminator protein may act by directly preventing terminator formation, rather than stabilising an alternative secondary structure. This would not be without precedent, as no antiterminator structure is present which would prevent terminator formation in the leader of the *K. oxytoca nas* operon (Chai & Stewart, 1998).

Also interesting is the apparent lack of consensus amongst the various putative promoter sequences. The presence of four different promoters introduces the possibility of multiple levels of regulation, with transcription from different promoters being determined by different σ factors in response to different conditions.

2.4 CONCLUSIONS

Although the use of *E. coli* YMC11 as a heterologous host proved to be valuable in studies on the regulation of *C. acetobutylicum* NCP 262 *glnA* expression by an antisense RNA

(Janssen et al., 1990, Fierro-Monti et al., 1992), it appears that this approach is unsuitable for the study of *glnR*. While GS activity appeared to be higher when *glnR* was present, this was only observed in cultures grown in complete media supplemented with glutamine at late exponential phase. In cultures grown in complete media both with and without glutamine, the presence of *glnR* had no effect on GS activity in mid-exponential phase. Where glutamine was absent, the effect seen at late exponential phase was reversed, with the presence of *glnR* apparently resulting in lower GS levels. The large standard deviations resulting from wide fluctuations in GS activity, under the conditions of these experiments, make a positive regulatory role of *glnR* difficult to demonstrate. The inability of *glnR* to affect the activity of GS in *E. coli* YMC11 may be as a result of the absence in *E. coli* of accessory factors necessary for the regulation of *glnR* itself, or some other limitation of the system.

The suitability of the closely related *C. beijerinckii* NCIMB 8052 as an alternative host was assessed, with respect to the regulation of its GS activity in response to nitrogen availability. Several parallels were observed in the nitrogen metabolism of *C. acetobutylicum* NCP 262 and *C. beijerinckii* NCIMB 8052. These included a preference for organic nitrogen, in the form of casamino acids, over inorganic nitrogen, the repression of GS by high concentrations of nitrogen, and the induction of GS under conditions of nitrogen limitation. At the physiological level, therefore, it would appear that the uptake of nitrogen, as well as its regulation, are similar in *C. acetobutylicum* NCP 262 and *C. beijerinckii* NCIMB 8052. Mechanisms controlling nitrogen metabolism in *C. beijerinckii* NCIMB 8052 may therefore be reflective of those of *C. acetobutylicum* NCP 262, and further study to establish the suitability of *C. beijerinckii* NCIMB 8052 as a model organism for these investigations is justified. The preference of these two organisms for organic nitrogen is interesting and further study of this aspect of their metabolism may yield some interesting insights into the mechanisms they use to assimilate and utilise nitrogen.

Southern hybridisation analysis indicated a large degree of similarity between *C. acetobutylicum* NCP 262 and *C. beijerinckii* NCIMB 8052 with respect to the sequence and relative arrangement of the *glnA*, *glnR* and *gltA* genes. The nitrogen region of *C. beijerinckii* NCIMB 8052 was isolated by cloning a 3.8-kb *Eco*RI fragment into pBluescript. The clone, designated pCBN1, was sequenced and found to carry homologues of *glnA*, *glnR*, and *gltA*.

The arrangement of the genes was indeed almost identical to that of *C. acetobutylicum* NCP 262 , except for some small differences in the spacing between them. Furthermore, the sequence identity at the nucleic acid level within the coding regions of the genes was found to be very high, ranging between 86.5% and 87.8%. This serves to confirm previous findings that the two strains are closely related (Keis et al., 1995). Complementation of glutamine auxotrophy in *E. coli* YMC11 demonstrated the presence of a functional *glnA* gene on the cloned DNA in pCBN1, and provided direct evidence that the correct region had been cloned.

The incomplete sequence of the *C. acetobutylicum* ATCC 824 genome is available on the world wide web. Interestingly, homology searches using the *C. beijerinckii* NCIMB 8052 *glnA* and *glnR* genes against this sequence did not identify any sequences sharing significant similarity with these two genes. *C. acetobutylicum* ATCC 824 is not as closely related to *C. acetobutylicum* NCP 262 as is *C. beijerinckii* NCIMB 8052, so a greater degree of genetic divergence would be expected. However, the identity at the amino acid level between *glnA* of *C. beijerinckii* NCIMB 8052 and that of the even more distantly related *B. subtilis* is higher than any identity shared with any available sequences of *C. acetobutylicum* ATCC 824. It is very probable that this lack of identity may be as a result of errors or gaps in the published sequence of the genome of *C. acetobutylicum* ATCC 824. A Southern hybridisation experiment might identify *glnA* and *glnR* homologues in this organism.

The first step in providing direct evidence for a regulatory role for *glnR* would be to create a mutant deficient for that gene and to establish the impact on the regulation of *glnA*. Under the proposed model, this would result in an uninducible *glnA* phenotype, or one displaying low levels of GS activity. However, attempts to create such a mutant were unsuccessful; despite a large number of conjugation events, as estimated by the conjugation frequency of pCTC1 (a replicative plasmid used as a positive control).

Primer extension experiments were conducted in order to determine the start of transcription of the *C. beijerinckii* NCIMB 8052. Four transcriptional start points were identified, two of which corresponded roughly with the two transcriptional starts identified for the *C. acetobutylicum* NCP 262 *glnA* gene. The other two start points were located upstream of these, and appeared to account for the majority of *glnA* transcription. Similar sequence motifs

were found to be present in the leader region of *glnA* of *C. acetobutylicum* NCP 262 and *C. beijerinckii* NCIMB 8052. These motifs are able to form factor-independent terminator structures in both strains, and this indicates that the antitermination model of control of *glnA* expression may also apply to *C. beijerinckii* NCIMB 8052.

It should be noted that in this experiment, inhibition of reverse transcriptase by secondary structures may have resulted in production of spurious cDNA's, and it is possible that one or more of the transcriptional start sites observed here may be artefacts resulting from this. For this reason, it is necessary to evaluate the roles of different portions of the promoter region using a different method, for example, a reporter plasmid system.

CHAPTER 3

DEVELOPMENT OF A REPORTER SYSTEM FOR *Clostridium beijerinckii* NCIMB 8052

CONTENTS

3.1 INTRODUCTION **78**
3.2 MATERIALS AND METHODS **81**
3.2.1 Strains and plasmids 81
3.2.2 Construction of plasmids 82
3.2.3 Selection of clones expressing *eglA* 84
3.2.4 Sequencing promoter/*eglA* fusions 84
3.2.5 Assaying of endoglucanase activity 84
3.3 RESULTS AND DISCUSSION **85**
3.3.1 Construction of promoterless reporter and control plasmids 85
3.3.2 Sequences of promoter/*eglA* fusions 85
3.3.3 Electroporation of *C. beijerinckii* NCIMB 8052 88
3.3.4 Plate assays for endoglucanase activity 89
3.3.5 Determination of temperature and pH optima of the cloned *eglA* in *C. beijerinckii* NCIMB 8052 92
3.3.6 Localisation of enzyme activity 93
3.3.7 Growth and reporter activity in different carbohydrate sources 95
3.3.8 Activity of the *glnA* promoter/reporter fusion under different conditions of nitroge availability 98
3.3.9 General comments on and suggestions for improvement of the reporter system 103
3.4 CONCLUSIONS **105**

SUMMARY

An endoglucanase gene (*eglA*) cloned from *C. acetobutylicum* NCP 262 was selected for use in the development of a reporter system for *C. beijerinckii* NCIMB 8052. Preliminary investigations indicated that the enzyme was efficiently secreted into the culture medium and did not remain associated with the cell. Optimum pH and temperature for the enzyme activity were found to be between 5.0 and 7.0, and 45^{0}C, respectively. Under the control of its own promoter, the enzyme was found to be repressed by cellobiose and, less markedly, by fructose. No significant induction by any of the other carbohydrate sources used was

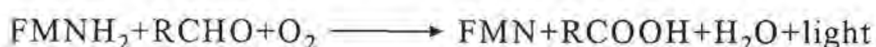
observed. Expression of *eglA* did not appear to affect growth or sporulation of *C. beijerinckii* NCIMB 8052. Transcriptional fusions between *eglA* and the *glnA* and *scr* promoter regions of *C. beijerinckii* NCIMB 8052 were constructed and reporter activity was assayed under various conditions. No activity could be measured in liquid culture of cells carrying the *scr* promoter-reporter fusion, indicating either that regions crucial for transcription had been omitted from the fusion, or that the presence of the fusion had disrupted normal *scr* regulation. In cultures of cells harbouring the *glnA* promoter-reporter fusion, high levels of reporter activity were observed, although the expected pattern of regulation was absent.

3.1 INTRODUCTION

Few reporter systems have been developed for use in bacteria of the genus *Clostridium*. In the study of promoter function in *Clostridium perfringens*, two strategies have been employed in the construction of reporter systems. The first involves the use of the chloramphenicol acetyltransferase (*catP*) gene (Matsushita et al., 1994; Bullifent et al, 1995) initially cloned from the *C. perfringens* plasmid pIP401 (Steffen et al., 1993), while the second entails the use of the genes encoding the luciferase enzyme of *Vibrio fischeri* (Phillips-Jones, 1993).

Using a construct in which the *catP* gene was flanked by two *rrnBT*$_1$ terminator sequences, Bullifent et al. (1995) showed that CAT could be expressed under the control of the *C. perfringens plc* promoter. Furthermore, CAT expression and hence activity could be correlated with alpha-toxin production. This system represented an improvement over a previous CAT-based reporter system (Matsushita et al., 1994), which lacked terminator sequences flanking the *catP* gene, and which also contained the native *catP* ribosome-binding site and transcriptional start site.

The luciferase enzyme of *V. fischeri*, the two subunits of which are encoded by the *luxAB* genes, has been used widely as a reporter system in aerobic and facultative organisms (Stewart & Williams, 1992). Luciferase catalyses the following reaction:



$$FMNH_2 + RCHO + O_2 \xrightarrow{\text{luciferase}} FMN + RCOOH + H_2O + \text{light}$$

Where RCHO is an exogenously added long-chain aliphatic aldehyde.

This system has not been used in obligately anaerobic bacteria, the main reason for this being the requirement of oxygen for the reaction. However, expression of luciferase and light production has been achieved in the relatively aerotolerant *C. perfringens* (Phillips-Jones, 1993). Production of light in *C. perfringens*, in which expression of *luxAB* was under the control of the *C. perfringens* alpha-toxin promoter, was shown to be reflective of luciferase expression. Furthermore, the amount of light produced was comparable to that observed in *E. coli*, although the kinetics of the reaction were different in the two organisms. In *E. coli*, light production increased steadily to a maximum over approximately five minutes and then decayed slowly, while in *C. perfringens* it increased rapidly and then decayed exponentially. It was proposed that the rapid decay in *C. perfringens* was due to exhaustion of $FMNH_2$, rather than cell death due to exposure to oxygen, as the cells remained viable subsequent to the assay (Phillips-Jones, 1993). It would appear that the use of luciferase as a reporter in *C. perfringens*, as well as other aerotolerant clostridial strains, is feasible.

In *C. beijerinckii* NCIMB 8052, both the *xylE* (xylanase) from *Pseudomonas* and the *cat* (chloramphenicl acetyltransferase) gene from the *Staphylococcus* plasmid pC194 were used to assess the activity of the *Clostridium pasteurianum* ferredoxin gene promoter in the development of an expression vector (Minton et al., 1993). While expression of *xylE* was relatively inefficient, high levels of *cat* expression were observed under appropriate conditions (Minton et al., 1993).

An endo-β-1,4-glucanase gene (*eglA*) cloned from *C. acetobutylicum* NCP 262 (Zappe et al., 1986) was used successfully as a reporter gene for investigating the regulation of *glnA* transcription in *Bacteroides fragilis* (Abratt et al., 1993). It was previously reported that a sub-strain of *C. acetobutylicum* NCP 262, *C. acetobutylicum* P270, possesses a weak

endoglucanase activity, induced by a small unidentified molecule present in molasses medium (Allcock & Woods, 1981). The enzyme activity had a pH optimum of 4.5 and an optimum temperature of 39^{0}C. In contrast to this, the enzyme encoded by the cloned *eglA* expressed in both *E. coli* and *B. fragilis* had a pH optimum of between 5.0 and 7.0, and a temperature optimum of 50^{0}C (Zappe et al., 1986). These discrepancies may have been due to, for example, host-specific factors such as post-translational modification, or the presence of two different endoglucanase activities in *C. acetobutylicum* NCP 262. Alternatively, it is possible that the altered physicochemical properties of the protein were due to the cloned gene being truncated. Analysis of the sequence of the cloned gene identified a putative signal sequence for secretion into the growth medium (Zappe et al., 1988; Zappe, PhD thesis). Consistent with this is the observation that in both *E. coli* and *B. fragilis*, the enzyme activity was found to be located predominantly in the periplasm (Zappe et al., 1986; Abratt et al., 1993).

Cellulose is an abundant and potentially rich source of carbohydrates, however, it is inaccessible to most living organisms, as few are able to digest it. Several species of bacteria and fungi, however, have evolved the ability to produce the enzymes needed to break down cellulose and utilize it. Cellulose digestion in *Clostridium thermocellum* and *Clostridium cellulovorans* has been extensively studied. The ability to digest crystalline cellulose, also known as true cellulase activity, depends on the presence of several different cellulolytic enzymes. Endo-β-1,4-glucanase enzymes cleave β-1,4-glucan bonds within the cellulose chain, while exo-1,4-β-D-glucanases, which can be either 1,4-β-D-glucan cellohydrolases or glucohydrolases, remove the respective sugars from the non-reducing end of the chain. 1,4-β-D glucosidases hydrolyse cellobiose or the non reducing ends of oligodextrins to glucose. Cellobiose and cellodextrin phosphorylases often form part of the enzyme complex and are responsible for the cleavage of oligosaccharides to yield glucose-6-phosphate (Saha et al., 1989).

Much of the cellulase activity of these two, as well as other bacterial species, is found in a large multienzyme complex called the cellulosome, which consists of a complex of enzymes which are responsible for the activities described above, and held together by a scaffold protein (reviewed in Felix & Ljungdahl, 1994). The cellulosome forms on the surface of the cell and anchors it to the surface of the substrate, which is usually crystalline cellulose. Large

aggregations of cellulosomes, called polycellulosomes, often form protuberances which project from the cell surface to the substrate. Other cellulases, however, exist independently of the cellulosome and act synergistically with it to enable the organism to utilize cellulose as a sole carbohydrate source (Doi et al., 1994).

In an attempt to expand its substrate range, *C. acetobutylicum* ATCC 824 was transformed with the *engB* gene, which encodes an endoglucanase from *C. cellulovorans* (Kim et al., 1994). Although this did not enable *C. acetobutylicum* to utilise crystalline cellulose as a sole source of carbon, an endoglucanase activity was purified from the culture supernatant which was weakly active against acid-swollen filter paper and semicrystalline cellulose. This activity was in addition to the endogenous *C. acetobutylicum* ATCC 824 endoglucanase which showed relatively weak activity against carboxymethyl cellulose (CMC). A similar experiment was performed using the *celE* and *celC* β-glucanases from *C. thermocellum*. Simultaneous expression of these two genes as a synthetic transcriptional fusion in *C. beijerinckii* NCIMB 8052 allowed the strain to utilize barley β-glucan as a sole carbon source (Minton et al., 1993).

This chapter describes the expression of an endoglucanase from *C. acetobutylicum* NCP 262 in *C. beijerinckii* NCIMB 8052. Attempts to create a reporter system for *C. beijerinckii* NCIMB 8052 using the cloned *C. acetobutylicum eglA*, and its use in investigating the activity of the *C. beijerinckii* NCIMB 8052 *glnA* and *scr* promoters, are also described.

3.2 MATERIALS AND METHODS

3.2.1 Strains and plasmids

All strains and plasmids used in this work are listed in Table 3.1. *E. coli* and *C. beijerinckii* NCIMB 8052 were maintained and cultured as described in section 2.2.1. *B. subtilis* IA46 was routinely cultured on LA or 2xYT agar plates, or in LB or 2xYT liquid medium, with antibiotic selection where appropriate. For long-term storage, nutrient agar (Difco) in a Bijou bottle was inoculated from a single colony and the bottle sealed tightly with Parafilm. Alternatively, spores were generated by streaking to single colonies on nutrient agar in a Petri

dish, which was also sealed and left for 14 days at room temperature. Again, antibiotic selection was used where appropriate.

Table 3.1 Strains and plasmids used in this work

Name	Relevant Characteristics	Reference
Bacterial strains		
B. subtilis IA46	*recA-4 thr-5 trpC-2*	Bacillus Genetic Stock Centre
C. acetobutylicum NCP262	wild-type	Jones et al. (1982)
C. beijerinckii NCIMB 8052	wild-type	NCIMB type strain
E. coli JM109	Δ(*lac-proAB*) *lacI*^{q} Δ(*lacZ*)*M15 recA1*	Yanisch-Perron et al., (1985)
Plasmids		
pHZ117	Ap^R EG^+; *eglA* in pUC19	Zappe (1988)
pHZ117Δ12	Ap^R, promoterless *eglA* in pUC19	Abratt et al. (1993)
pBluescriptSK+	Ap^R, T_3 and T_7 polymerase promoters	Stratagene, Ja Holla, California
pFNK1	MLS^R	Mermelstein et al. (1992)
pER1	Promoterless *eglA* from pHZ117Δ12 cloned in pFNK1	This work
pEC1	*eglA* from pHZ117 cloned in pFNK1	This work
pSP1	*C. beijerinckii* NCIMB 8052 *scr* promoter region cloned in pSK	This work
pISP1	pISP1/pFNK1 fusion	This work
pESP1	*Eco*RI/*Xho*I deletion of pISP1	This work
pEsk	Promoterless *eglA* cloned in pSK	This work
pEGPsk	*C. beijerinckii* NCIMB 8052 *glnA* promoter/*eglA* fusion cloned in pSK	This work
pEGP1	*C. beijerinckii* NCIMB 8052 *glnA* promoter/*eglA* fusion cloned in pFNK1	This work

3.2.2 Construction of plasmids

For the construction of the plasmid pER1 which carries the promoterless *eglA* cloned into the *Bacillus/Clostridium* shuttle vector pFNK1 (Mermelstein et al., 1992), the *Eco*RI/*Pst*I fragment of pHZ117Δ12 (Abratt et al., 1993) was excised and the ends blunted using T4 DNA polymerase (Roche; A.1.2.5). The fragment was gel-purified in 1% low melting point agarose and cloned into the *Pvu*II site of pFNK1. The *Eco*RI/*Pst*I fragment of pHZ117 (Abratt et al., 1993), which carries the coding region and functional promoter of *eglA*, was excised and cloned into pFNK1 as before. The resulting plasmid was designated pEC1.

pESP1 was constructed as follows: Oligonucleotide primers were designed to amplify a region consisting of 152 nucleotides encompassing the promoter region of the *C. beijerinckii* *scrA* operon (Reid et al., 1999). The sequence of the forward primer was 5'-AACTGCAGGGAGGCGTAATGG-3', while the sequence of the reverse primer was 5'-AAGTCGACAATTAACCTCGGAAACC-3'. Primers were designed to introduce a *Pst*I site and a *Sal*I site at the 5' and 3' ends of the promoter region respectively. The PCR reaction was performed using the following protocol: (1) 97^{0}C for 5 minutes; (2) 97^{0}C for 1 minute; (3) 40^{0}C for 1 minute; (4) 68^{0}C for 1 minute; (5) 68^{0}C for 5 minutes; steps 2 to 4 were repeated 30 times. The PCR product was digested with *Sal*I and *Pst*I and cloned into the *Sal*I and *Pst*I sites of pSK to create the plasmid pSP1. Both pSP1 and pER1 were digested with *Sma*I and ligated to produce pISP1. To delete the pSK sequence between the promoter and reporter, pISP1 was digested with *Eco*RI and *Xho*I and the cohesive ends were removed using T4 polymerase and ligated. The resulting plasmid, carrying the *scr* promoter-reporter fusion was named pESP1.

For the construction of pEGP1, the plasmid pHZ117Δ12, carrying the promoterless *eglA*, was digested with *Pst*I and the resulting cohesive ends removed with T4 DNA polymerase. The linearized plasmid was then cut with *Eco*RI, and the 1.4-kb fragment carrying the promoterless *eglA* was purified and cloned into the *Sma*I/*Eco*RI sites of pSK. The resulting plasmid pEsk was cut with *Eco*RI and the ends filled in with Klenow enzyme (A.1.2.4). The 726bp *Sty*I fragment of pNO1 was excised, purified, and cloned into the blunted *Eco*RI site of pEsk to create the plasmid pEGPsk. pEGPsk was digested with *Xho*I and *Xba*I to release a 2.2-kb fragment carrying the *C. beijerinckii* NCIMB 8052 *glnA* promoter fused to the *C. acetobutylicum* NCP 262 *eglA* structural gene. The cohesive ends of the fragment were filled in using Klenow enzyme and the fragment was purified on a 1.0% agarose gel and cloned into the *Pvu*II site of pFNK1.

In all of the above cases, pFNK1-based constructs were used to transform *B. subtilis* IA46, where they were amplified (A.1.5.2). Plasmids were harvested using the large-scale extraction protocol described in Appendix A.1.1.5 and introduced into *C. beijerinckii* NCIMB 8052 via electroporation (A.1.5.3). Plasmids were extracted from *C. beijerinckii*

transformants using a scaled-down DNA extraction protocol (A.1.1.6), and checked using restriction endonuclease digestion.

3.2.3 Selection of clones expressing *eglA*

A plate assay was used to identify *E. coli* and *C. beijerinckii* NCIMB 8052 clones expressing *eglA*. Cells were streaked or plated onto 2xYT agar containing 0.5% medium viscosity carboxymethyl cellulose (CMC), as well as 0.05% cysteine-HCl and 0.1% $NaHCO_3$ in the case of *C. beijerinckii* NCIMB 8052. *C. beijerinckii* NCIMB 8052 transformants were also plated onto CBM or GSMM containing 0.5% CMC. After incubation overnight, the plates were washed with sterile water to remove colonies, and stained with a 0.1% solution of Congo red for approximately 15 minutes. The Congo red was then discarded and the plates were destained with a 1M NaCl solution for 15 minutes. Regions of the agar in which the CMC had been digested were unable to retain the stain, while regions in which no endoglucanase activity was present remained a red colour.

3.2.4 Sequencing promoter/*eglA* fusion junctions

The programs Prime, of the GCG v9.1 package, and DNAman v2.0 were used to design an oligonucleotide for use as a primer to sequence from within the *eglA* coding region into the promoter region to which it had been fused. The sequence of the primer was as follows: 5'-CCACCAAAAGATGTTGTTGC-3'. The primer was designed to be complementary to nucleotides 115 to 134 of the *eglA* coding region. Sequencing was performed using the Sanger di-deoxy chain termination method on an AlfExpress automated DNA sequencer. DNA that had been obtained from *B. subtilis* IA46, and which was used to transform *C. beijerinckii* NCIMB 8052, was used as the template for sequencing reactions.

3.2.5 Assaying of endoglucanase activity

Endoglucanase activity was assayed using the p-nitrophenyl cellobiosidase assay of Deshpande et al. (1984), described in Appendix A.1.6.2. For the investigation of carbohydrate source on expression of *eglA* under the control of its own and *scr* promoters, the following procedure was followed: Cultures were grown in CBM with erythromycin at a concentration of 10µg/ml until an OD_{600} of 0.5 was reached. A 10ml aliquot of the culture was centrifuged anaerobically at 6000 rpm, washed in sterile distilled water, and inoculated

into 200ml of the appropriate culture medium. Samples of culture medium were taken at appropriate growth phases, as determined by spectrophotometric analysis. For studies on the *glnA* promoter, precultures were grown to mid-exponential phase ($OD_{600} = 0.5$) in the medium to be used for the assays. Aliquots of 10ml each were harvested as described above and used to inoculate 200ml of culture medium. Except where specified, assays were conducted on the culture supernatant.

3.3 RESULTS AND DISCUSSION

3.3.1 Construction of reporter vector and control plasmid

In order to construct a promoterless reporter vector, the promoterless *eglA* gene was excised from the plasmid pHZ117Δ12. Single stranded overhangs were removed, and the fragment was cloned into the *Pvu*II site of pFNK1. Plasmids were extracted and checked by restriction endonuclease digestion, to ensure that the fragment had been inserted in the correct orientation with respect to the restriction sites in pFNK1. The same was done for pEC1, using the *eglA* gene, with its own functional promoter, from pHZ117. The plasmids are shown in Figure 3.1.

In pER1, which comprises the promoterless *eglA* cloned into pFNK1, the *Pst*I, *Bam*HI and *Sma*I restriction sites are unique to the cloning site, and can therefore be used for cloning promoter fragments. Although there are two *Eco*RI sites in the cloning site, these can still be used for cloning, as they are located close together, and deletion of the intervening sequence will have no effect on the construct. The reporter gene possesses its own ribosomal binding site, and can therefore be used as a promoter probe, however when specific fusions are constructed, more reliable results would be likely to be obtained if care is taken to omit the ribosomal binding site from the promoter region to be cloned.

3.3.2 Sequences of promoter/*eglA* fusions

Sequences of the junction regions of the fusions between the *scr* and *glnA* promoters and the *eglA* coding region are shown in Figures 3.2 and 3.3 respectively.

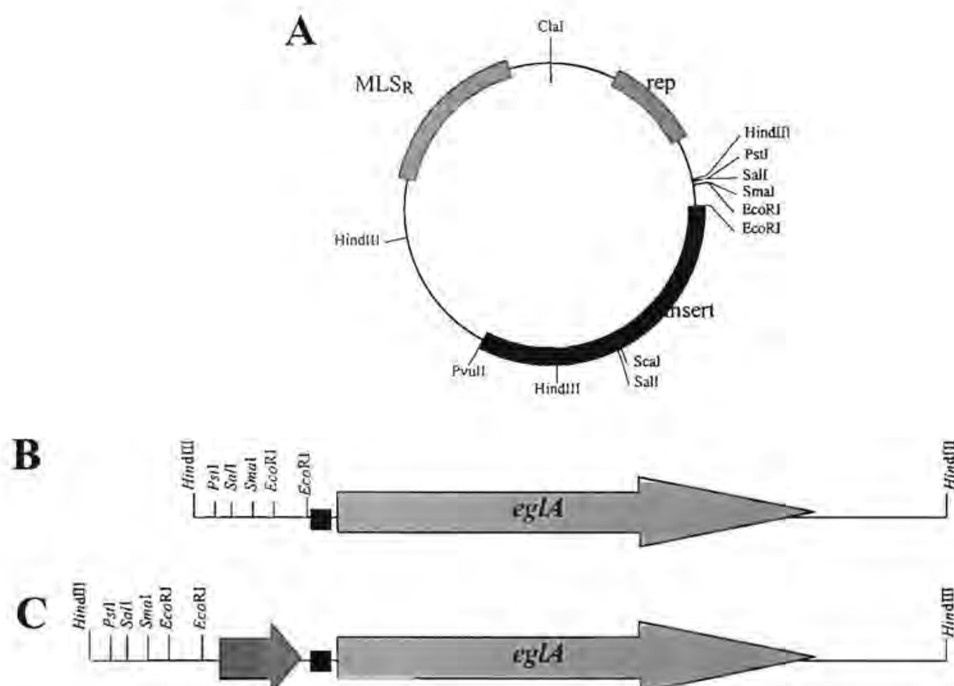


Figure 3.1: (A) Map of the backbone of the promoterless reporter vector pER1 and control plasmid pEC1. Approximate positions of restriction sites in the coding region of *eglA* are shown. (B) Map of the *eglA* region in the reporter vector, showing the relative positions of the promoterless endoglucanase (*eglA*) gene, and unique sites which could be used for cloning promoter fragments. (B) Map of the corresponding region of pEC1, showing the position of the *eglA* promoter. In (B) and (C), the *eglA* ribosomal binding site is shown as a black square.

pESP1: The *scr* operon promoter region to be fused to the reporter was initially obtained by PCR amplification and cloning of the product into pBluesciptSK+. Sequencing of the PCR product in pBluescriptSK+ revealed that 33 nucleotide residues at the 5' end had not been amplified (Fig. 3.2A). It is possible that at the relatively low annealing temperature (40^{0}C) used, mispriming may have occurred. The start of transcription of the *scr* operon has been identified (Reid et al., 1999), and is located 7 bases downstream of the -10 region of the putative promoter. A 12 nucleotide putative operator sequence for the binding of a catabolite repressor protein has been identified, starting one nucleotide downstream of the transcriptional start. All of these are included in the region fused to the reporter. The presence of 20 bases of pBluescript vector sequence between the promoter and start codon region is a potential source of problems, as it results in a longer mRNA leader sequence than that which occurs naturally.

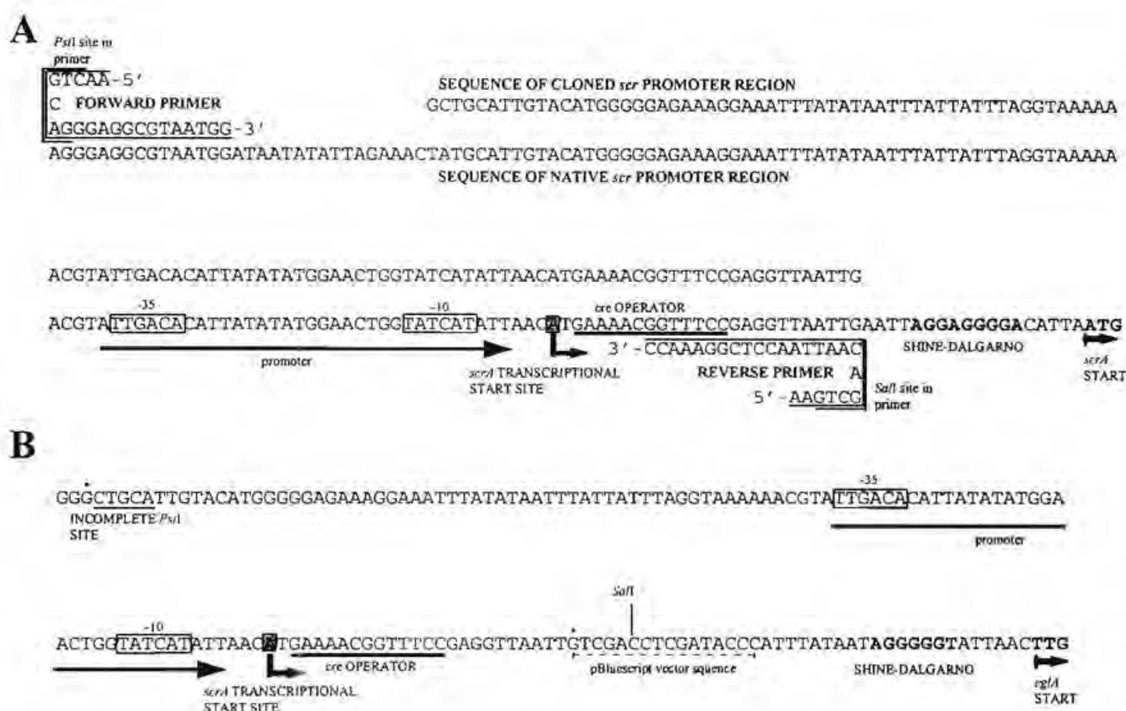


Figure 3.2: (A) PCR amplification and cloning of the *scr* promoter region. The native sequence and sequence of the cloned PCR product are shown. Also shown are the sequences and positions of the primers (underlined), including restriction sites (double underlined and labelled). The discrepancy between the target sequence and cloned sequence is shown. Other relevent features are labelled. (B) Sequence of the *scr* promoter-reporter fusion. Intervening pBluescript vector sequence is underlined with a dotted line and labelled. Asterisks mark the start and finish of the cloned promoter region.

pEGP1: The *Sty*I fragment carrying the promoter region of the *C. beijerinckii glnA* was cloned into the *Eco*RI site upstream of the reporter gene (see Fig. 3.3). The fusion carries 480bp of sequence upstream of the translational start of *glnA*, including all four transcriptional start sites. Nowever, it also carries the ribosomal binding site and 240bp of the *glnA* coding region. The presence of two ribosomal binding sites as well as part of the *glnA* coding region could be a potential source of problems in evaluating the function of the promoter. Inspection of the sequence at the junction of the promoter region and *eglA*, however, reveals that an amber (UAA) stop codon has been introduced at the 5' end of the truncated *glnA* gene. As there appeared to be no structures present which would be interpreted by the RNA polymerase

as a termination signal, it was hoped that the presence of the truncated *glnA* would not affect transcription from the *glnA* promoter.

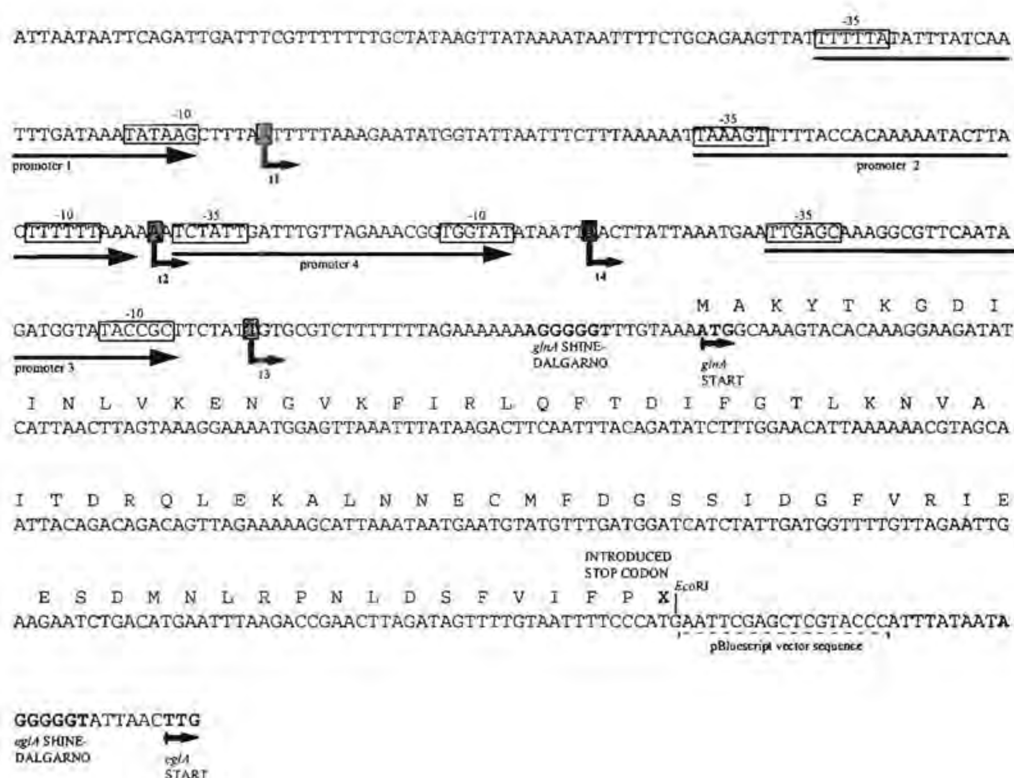


Figure 3.3: Sequence of the *C. beijerinckii* NCIMB 8052 *glnA* promoter region and the *eglA* reporter gene. The positions of Shine-Dalgarno sequences (bold), start codons (bold), transcriptional start sites (grey boxes), as determined in chapter 2, proposed promoter regions (underlined with bold arrows) and consensus sequences (boxed) are illustrated. Also illustrated is the amino acid sequence of the truncated 5' terminus of the *glnA* structural region. pBluescript vector sequence is underlined with a dotted line and labelled.

3.3.3 Electroporation of *C. acetobutylicum* NCIMB 8052

Table 3.2 shows the results of a typical electroporation of *C. beijerinckii* NCIMB 8052. Although the efficiency of transformation was orders of magnitude lower than the $4x10^2$ to $4x10^4$ transformants/μg DNA reported by Oultram et al. (1988), the results from this experiment were sufficiently reliable for the purposes of transforming cells with specific plasmids.

Table 3.2: Results of *C. beijerinckii* NCIMB 8052 electroporation

Plasmid	Amount DNA (μg)	Time constant*	Transformation Efficiency (# transformants/μg DNA)
pFNK1	16μg	3.7	8.78
pER1	3μg	4	5
pEC1	3μg	4	9.33
pESP1	3μg	3.9	2.33
pEGP1	4μg	1.9	2.25

*All time constants were obtained using identical experimental parameters (described in appendix A.1.5.3), with the exception of amounts of DNA used.

3.3.4 Plate assays for endoglucanase activity

Cultures of *C. beijerinckii* transformants carrying the plasmids pFNK1, pER1 and pEC1 were streaked out to single colonies on 2xYT or agar plates containing 0.5% CMC and 10μg/ml erythromycin and qualitatively assayed for CMCase activity. *C. beijerinckii* NCIMB 8052 transformants carrying the control plasmid pFNK1 (8052(pFNK1)) showed no zones of clearing around the colonies, while transformants carrying pEC1 (8052(pEC1)) showed substantial CMC digestion after overnight incubation, indicating an endoglucanase activity released into the medium. Transformants carrying pER1 (8052(pER1)) also produced zones of CMC digestion, although these were markedly smaller than those produced by 8052(pEC1) (Fig. 3.4). This indicates that although the promoter has been eliminated, there is still measurable expression of *eglA*, possibly due to transcription from an unidentified minor promoter downstream of the major promoter. A similar observation was made in *B. fragilis*, where a background level of endoglucanase activity was consistently measured in the presence of the promoterless gene (Abratt et al., 1993). An alternative explanation is that expression of the promoterless gene is being driven by a promoter in the vector. It is unlikely that the activity observed in 8052(pER1) is due to expression of an endogenous gene, as 8052(pFNK1) showed no activity in the plate assays.

Figure 3.4: Plate assays of endoglucanase activity in *C. beijerinckii* NCIMB 8052 transformed with pFNK1, pER1, pEC1 and pESP1 on CBM containing erythromycin at a concentration of 10μg/ml, and 1% glucose or 1% sucrose. Assays were conducted after incubation overnight.

C. beijerinckii NCIMB 8052 transformants harbouring pESP1 (8052(pESP1)), carrying the *scr* promoter-reporter fusion, were grown on CBM or 2xYT agar containing 0.5% CMC, with and without 0.2% sucrose. When grown on 2xYT, 8052(pESP1) colonies produced zones intermediate between those produced by 8052(pER1) and 8052(pEC1). Supplementing the media with 0.2% sucrose, however, did not appear to result in any noticeable increase in enzyme activity (Fig. 3.5). This is in apparent contrast to previous reports that the presence of sucrose in the media results in significant induction of transcription from this promoter (Reid et al., 1999). The significance of this result will be discussed in the context of results obtained from assays in liquid culture presented below.

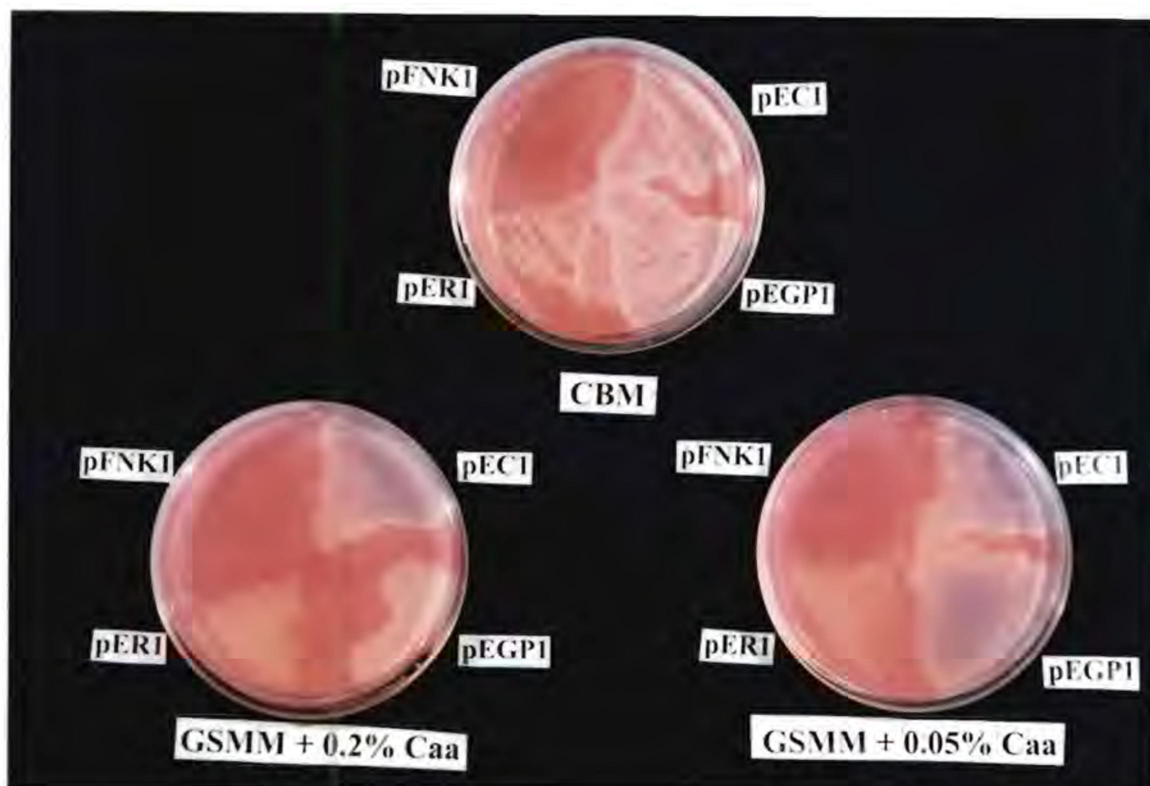


Figure 3.5: Plate assays of endoglucanase activity in *C. beijerinckii* NCIMB 8052 transformed with pFNK1, pER1, pEC1 and pEGP1 on CBM, or GSMM supplemented with 0.2% casamino acids or 0.05% casamino acids. All media was supplemented with erythromycin at a concentration of 10μg/ml. Assays were conducted after incubation overnight.

Transformants harbouring pEGP1 (*glnA* promoter-reporter fusion), pFNK1, pER1 and pEC1 were streaked to single colonies on the following media: CBM, GSMM containing 0.2% casamino acids, GSMM containing 0.025% casamino acids and GSMM containing 0.025% casamino acids with 0.015% glutamine (Fig 3.5). As was observed previously, after incubation at 37^{0}C overnight, 8052(pFNK1) produced no zones of CMC degradation, while 8052(pER1) produced small zones and pEC1 produced large zones on all the media. Although, in terms of the model of control proposed for the *Clostridium glnA*, 8052(pEGP1) was expected to produce more reporter activity under conditions of nitrogen limitation, it produced similar-sized zones on all the media. This could be interpreted as a lack of repression of the *glnA* promoter under conditions of nitrogen excess, however, the activity

observed was a result of accumulation of reporter activity over a considerable time, during which even nitrogen-rich media would become depleted. Again, the significance of these observations is uncertain, and assays in liquid culture would prove more informative.

3.3.5 Determination of temperature and pH optima of the cloned *eglA* in *C. beijerinckii* NCIMB 8052

Because there was a discrepancy between the pH and temperature optima reported for the endogenous endoglucanase activity of *C. acetobutylicum* NCP 262 and that of the cloned *C. acetobutylicum eglA* in *E. coli* and *B. fragilis*, the activity of the cloned enzyme expressed in a more closely related system was investigated. Cultures of *C. beijerinckii* NCIMB 8052 transformed with either pFNK1 or pEC1 were grown to stationary phase in CBM with antibiotic selection for the plasmids. Enzyme assays were performed on culture supernatant at a range of temperatures from 37^{0}C to 50^{0}C, and at a range of pH values between 5.0 and 7.0.

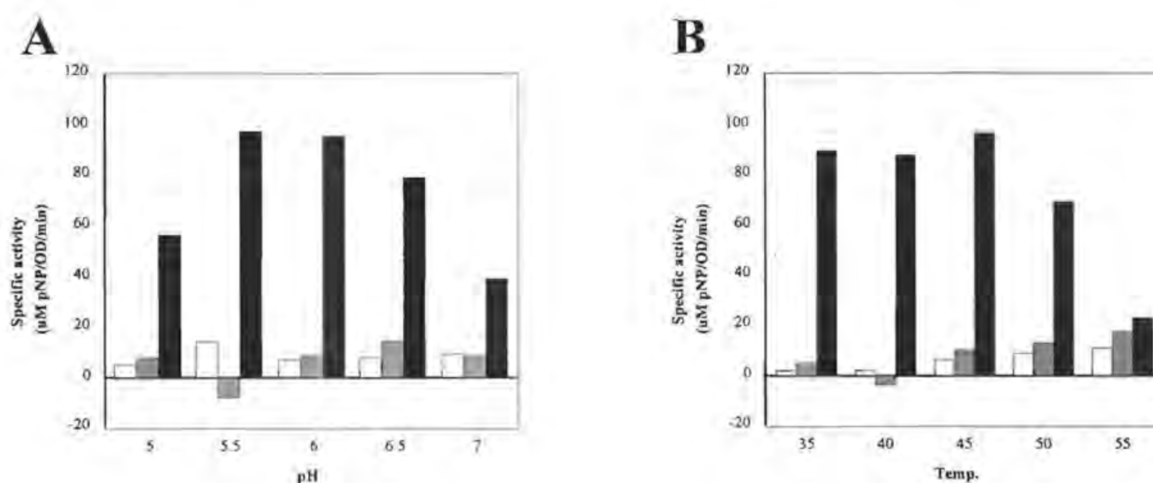


Figure 3.6: Determination of activity of the reporter enzyme at different temperature (A) and pH (B) values. (A) Assays conducted at pH 6.0. (B) Assays conducted at 45^{0}C. Open columns: 8052(pFNK1); grey filled columns: 8052(pER1); filled columns: 8052(pEC1). Reporter specific activity is expressed as µM p-nitrophenol released/min/culture OD unit/. Values represent the result of a typical experiment.

The enzyme was found to be most active at a temperature of 45^{0}C (Fig. 3.6(A)) and a pH of between 5.5 and 6.0 (Fig. 3.6(B)). This is consistent with the values observed for the cloned enzyme expressed in *E. coli* and *B. fragilis*. *C. acetobutylicum* NCP 262 and *C. beijerinckii* NCIMB 8052 are closely related, so it is unlikely, but not impossible, that different post-

transcriptional modifications of the enzyme could account for the differing physiological parameters of endoglucanase activity in these two organisms. There is no translational stop codon in the cloned *eglA* open reading frame, so the expressed protein, although enzymatically fully functional, could possess an altered carboxy terminus, which could in turn alter its stability and activity at different temperatures and pH values.

3.3.6 Localisation of enzyme activity

In order to determine the efficiency with which the product of the cloned *eglA* was exported from the cell, and whether it remains associated with it in any way, the following experiment was conducted: Mid- and late-exponential cultures of 8052(pFNK1) and 8052(pEC1) were harvested. Cells were lysed by sonication, and endoglucanase assays were conducted on the soluble and insoluble fractions of the cell lysate, as well as on the extracellular medium of the respective cultures. Although small amounts of activity could be detected in the soluble and insoluble fractions of the cell lysates of 8052(pFNK1), 8052(pER1) and 8052(pEC1), there was no significant difference between them, and the values themselves could be considered to be negligible (Fig. 3.7(B) and (C)). By contrast, although low endoglucanase activity could be measured in the culture medium of 8052(pFNK1) or 8052(pER1), that of 8052(pEC1) was shown to contain relatively high levels of activity (17.13μmol pNP/min/culture OD unit at mid-exponential phase and 30.52μmol pNP/min/culture OD unit at late exponential phase; Figure 3.7(A)).

Qualitatively, the results observed for the culture supernatant were reflective of those observed in the plate assays, with the exception of 8052(pER1) (promoterless *eglA*). Although this strain produced zones of CMC digestion on plates, no activity could be measured after growth in liquid medium. This is in apparent contrast to the observation that there was significant endoglucanase activity in colonies grown on CMC agar plates. There are at least three possible explanations for this. One is that the endoglucanase activity secreted into the liquid culture medium was diluted to undetectable levels. Another is that the presence of CMC in the solid medium was able to induce transcription from an unidentified promoter in the "promoterless" construct. The third is that some other factor related to differences in conditions between liquid and solid medium resulted in differential expression of the "promoterless" gene. These hypotheses were not tested, however, but whatever the

reason, it is apparent that plate assays are at best a crude qualitative measure of endoglucanase activity. From the results presented here, it can be concluded that the endoglucanase activity produced by the cloned *eglA* gene is exported from the cell into the medium, and does not remain associated with the cell in any way.

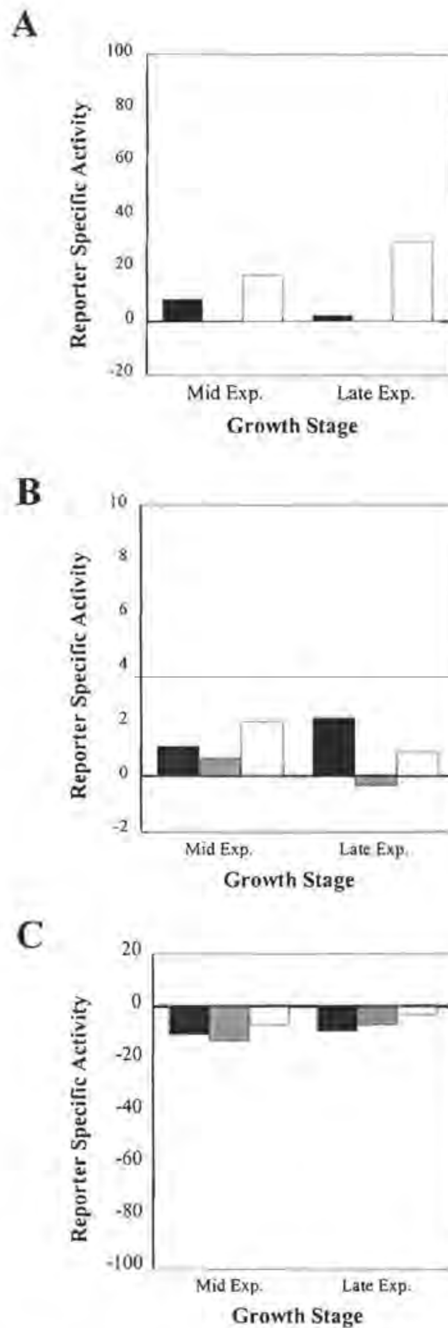


Figure 3.7: Localisation of reporter activity. Endoglucanase assays were conducted on medium (A), as well as soluble (B) and insoluble (C) fractions of cell lysate of mid- and late exponential phase cultures. Filled columns: 8052(pFNK1); grey filled columns: 8052(pER1); open columns: 8052(pEC1). Specific activity expressed as μmol p-nitrophenol released/ culture OD unit/min (A) or μmol p-nitrophenol released/μg protein (B) and (C). Values represent results of a typical experiment

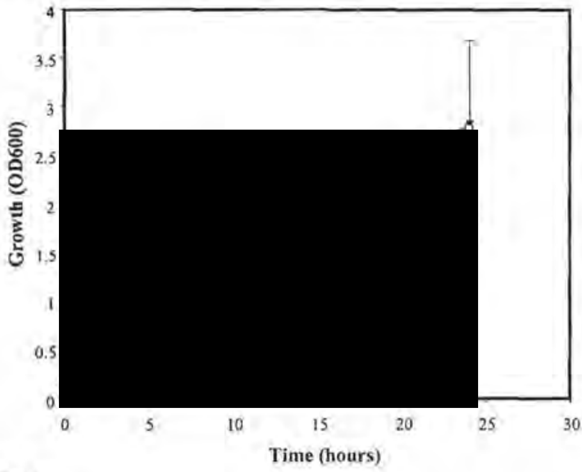
3.3.7 Growth and reporter activity on different carbohydrate sources

Expression of *eglA* in *C. acetobutylicum* NCP 262 was previously reported to be inducible rather than constitutive. As little other information is available regarding the regulation of *eglA* in *C. acetobutylicum* NCP 262 and as no endoglucanase activity has been reported in *C. beijerinckii* NCIMB 8052 (Minton et al., 1993), it is of interest to investigate how this enzyme is regulated in the heterologous host *C. beijerinckii* NCIMB 8052, prior to investigations of the activity of reporter/promoter fusions. To this end an experiment was conducted to test the effect of various carbohydrates on expression of the reporter under the control of both its own and *scrA* promoters. Cultures of *C. beijerinckii* NCIMB 8052 harbouring pFNK1, pER1, pEC1 and pESP1 were grown in CBM with glucose, sucrose, galactose, fructose or cellobiose as carbohydrate sources. Growth of the various strains was monitored in each medium, and endoglucanase activity was assayed in the culture supernatant at mid- (OD_{600} = 0.5) and late-exponential phases (OD_{600} = 1.0), as well as in stationary phase (OD_{600} = 2.0 - 3.0).

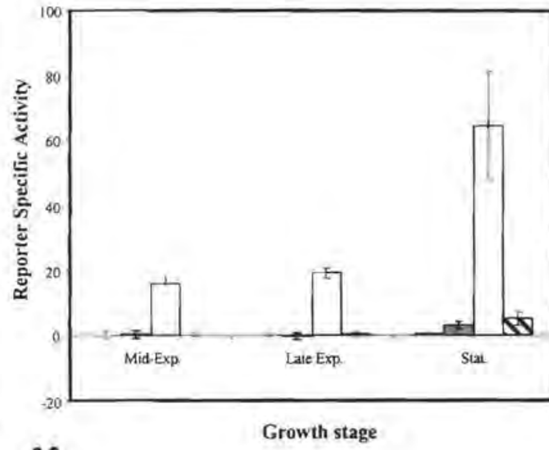
No large differences were observed between growth rates of strains harbouring the different constructs (Fig 3.8(Ai)-(Ei)), however a small depression in growth rates of 8052(pER1) and 8052(pEC1) relative to 8052(pFNK1) and 8052(pESP1) was observed during growth on all carbohydrate sources. When tested using plate assays, 8052(pER1), 8052(pEC1) as well as 8052(pESP1) were all shown to produce endoglucanase activity, but 8052(pESP1) showed a similar growth profile to that of 8052(pFNK1). This depression in growth rates, therefore, is probably not due to endoglucanase production per se.

In all cases, both 8052(pFNK1) nor 8052(pER1) produced very little, if any, reporter activity at any stage of growth (Figure 3.8(Aii)-(Eii)). This confirms that under the conditions of these experiments, there was no significant induction of any endogenous endoglucanase activity. The activity of the cloned *eglA* in *C. beijerinckii* NCIMB 8052 under the control of its own promoter was assayed in response to various carbohydrate sources. In all media, endoglucanase activity in mid-exponential phase was relatively low, varying between 6.8 and 15.7μM pNP/min/OD with progressively higher levels observed as growth progressed (up to 64.6μM pNP/min/OD in stationary phase), with the exception of cultures grown with fructose and cellobiose as carbohydrate sources (Fig. 3.8(Dii) and (Eii), respectively). This pattern could be interpreted as reflective of repression under conditions of carbohydrate excess, with

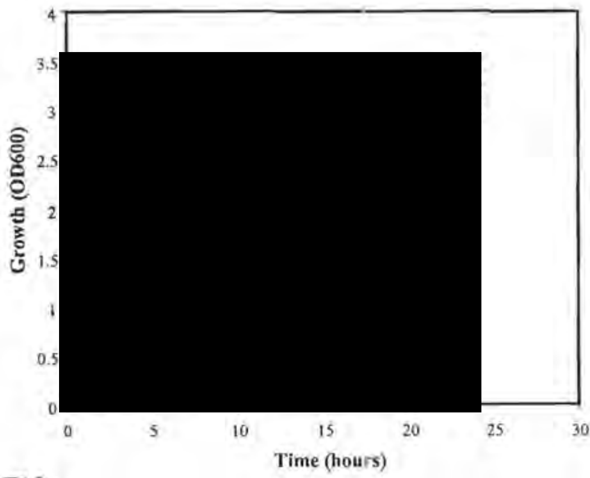
Ai



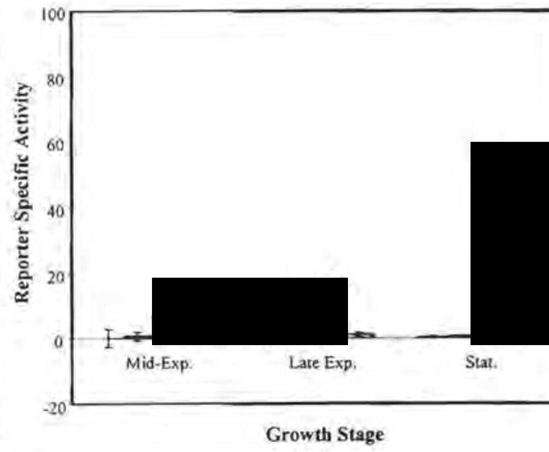
ii



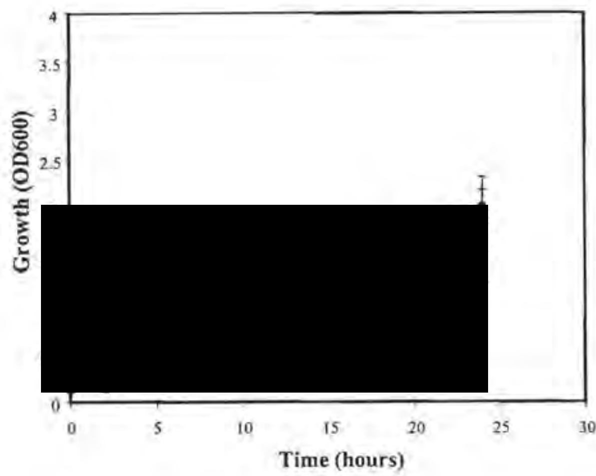
Bi



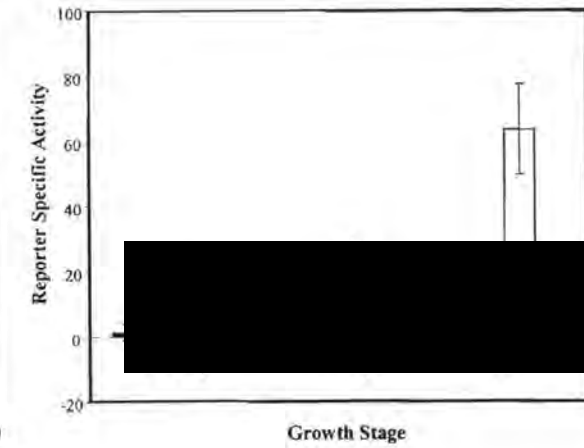
ii



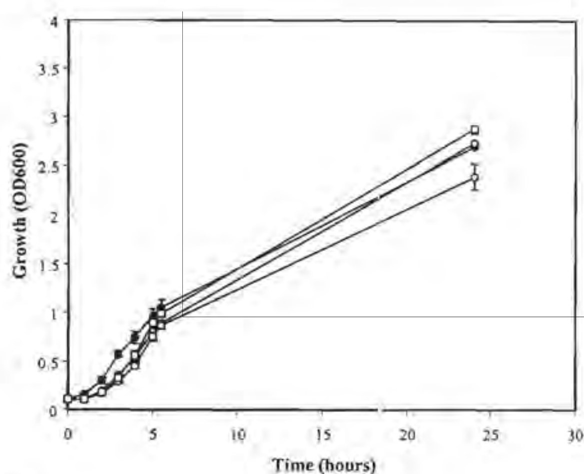
Ci

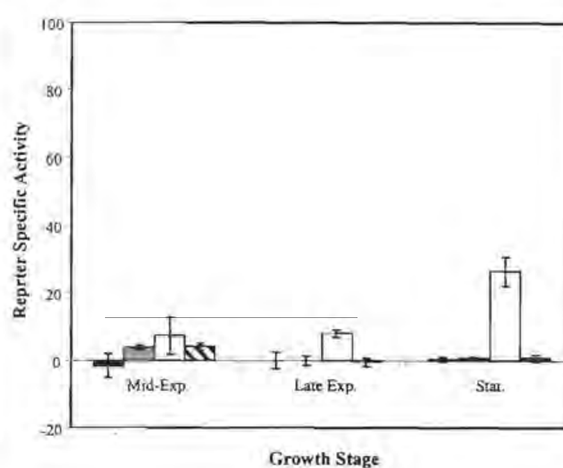


ii

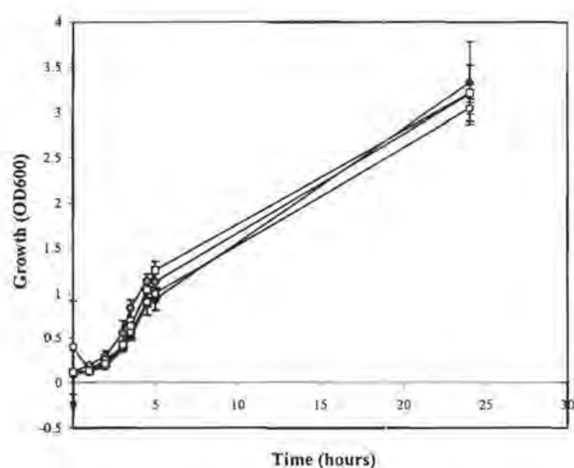


Di **ii**





Ei **ii**



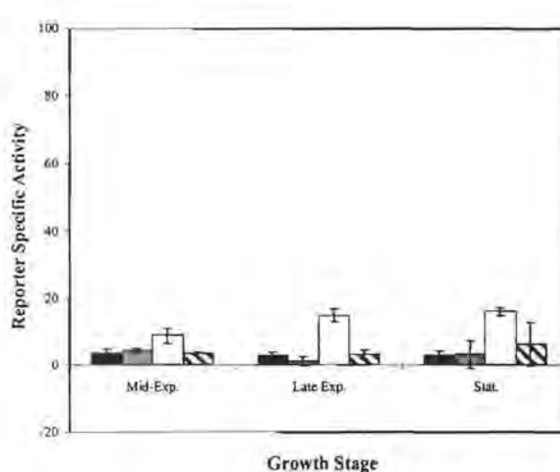


Figure 3.8: Growth (i) and reporter activities (ii) of cultures grown in different carbon sources: A: glucose; B: sucrose; C: galactose; D: fructose; E: cellobiose. (i): Absorbance at 600nm was used as the measure of cell density. Filled circles: 8052(pFNK1); grey filled circles: 8052(pER1); open circles: 8052(pEC1); open squares: 8052(pESP1). (ii): Reporter activity was measured in medium of cultures at mid- and late exponential phase, and in stationary phase. Specific activity is expressed as μmol p-nitrophenol released/min/culture OD unit. Black filled columns: 8052(pFNK1); grey filled columns: 8052(pER1); open columns: 8052(pEC1); oblique striped columns: 8052(pESP1). All values represent the average of three independent experiments.

the repression being lifted as easily metabolised carbohydrates were exhausted. Alternatively, this pattern could represent accumulation of activity as a result of low-level constitutive *eglA* expression. The enzyme was observed to be relatively stable under culture conditions (data not shown), and it is likely that the activity measured in stationary phase represented the sum total of all enzyme produced during growth, rather than that being produced at the time of the assay.

Growth on cellobiose and, less markedly, fructose, resulted in lower levels of endoglucanase activity (Fig 3.8 (Dii) and (Eii). As cellobiose is one of the major end products of cellulose digestion, and is readily utilised by *C. beijerinckii* NCIMB 8052, it possible that the observed repression of endoglucanase activity is due to a mechanism of end-product repression. In this case, whether this is at the level of transcription or translation, or even due to a post-translational mechanism, is uncertain, and this concern will be dealt with later. The significance of the apparent repression of endoglucanase activity in the presence of fructose is uncertain.

The *scrARBK* operon of *C. beijerinckii* NCIMB 8052 encodes four genes involved in the uptake and metabolism of sucrose. The genes are cotranscribed on a 5kb mRNA from the *scr* promoter, and transcription has been reported to be induced by the presence of sucrose, and repressed by glucose (Reid et al., 1999). In this experiment, no activity could be detected in the culture of 8052(pESP1) during exponential growth, and only very low levels were measured at stationary phase, irrespective of the carbohydrate source. It would appear that the region of the promoter which was cloned to form the fusion lacks a region, or regions, essential for transcription. This could also be attributed to the increased spacing between the promoter and translational start in the construct.

3.3.8 Activity of the *glnA* promoter/reporter fusion under different conditions of nitrogen availability

Because it had been established that GS activity in *C. beijerinckii* NCIMB 8052 is differentially regulated under different conditions of organic nitrogen availability (see chapter 2), the activity of the cloned *glnA* promoter region, as indicated by reporter activity, was assayed under the same conditions. Cultures of cells transformed with pFNK1, pER1, pEC1 and pEGP1 were grown in complex medium (CBM), as well as in minimal medium with excess (0.2%) and limiting (0.05%) concentrations of organic nitrogen. Culture growth was measured and reporter activity in the supernatant was assayed at early, mid and late exponential growth phases. To ascertain whether regulation of GS was affected in any way by the presence of the promoter-reporter fusion, GS activity of cultures of cells transformed with pEGP1 was compared with that of cells transformed with pEC1.

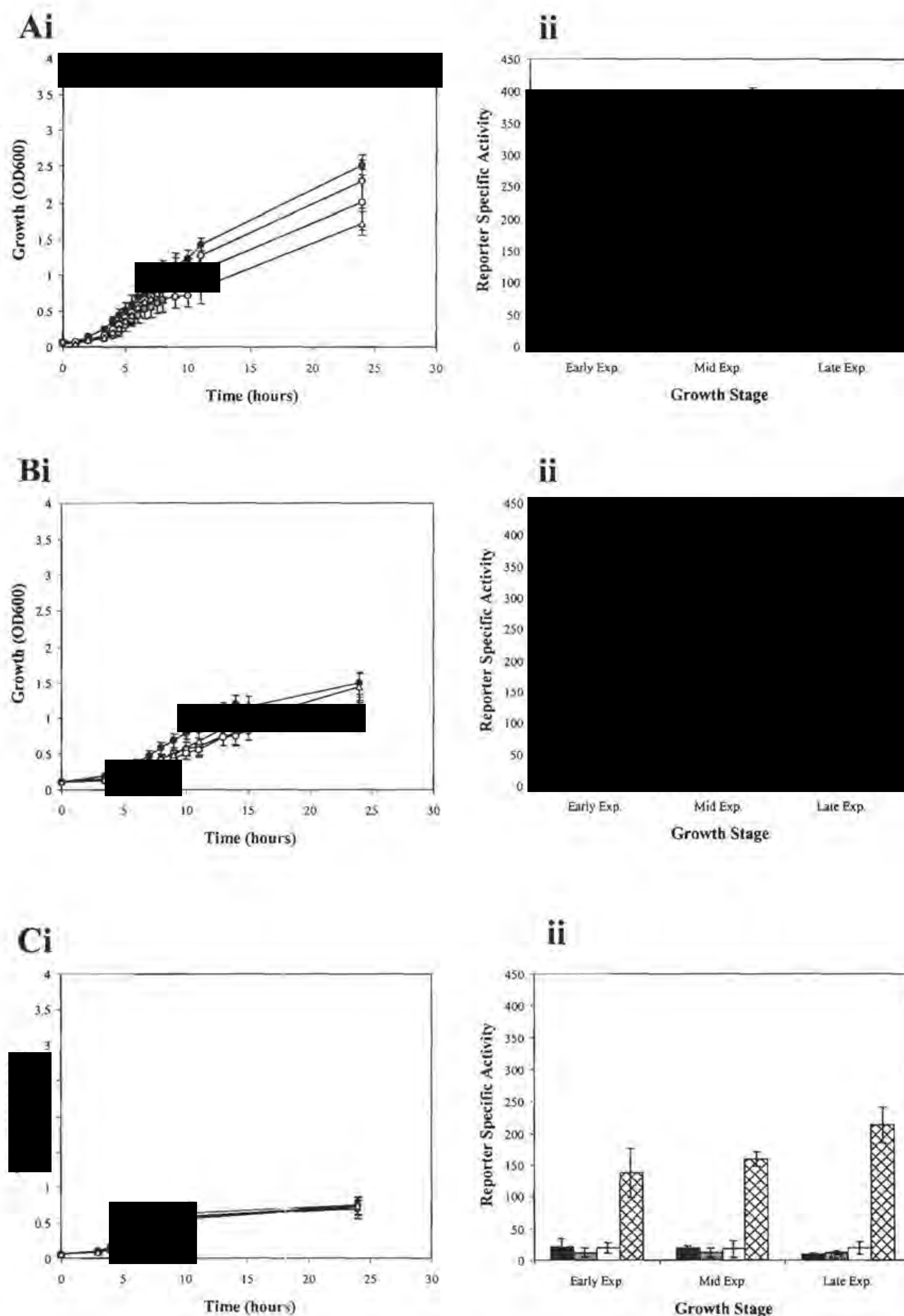


Figure 3.9: Growth (i) and reporter activity(ii) under different nitrogen conditions: (A) CBM; (B): GSMM+0.2% Casamino acids; (C): GSMM+0.05% Casamino acids. (i): Filled circles: 8052(pFNK1); Grey filled circles: 8052(pER1); Open circles: 8052(pEC1); Open triangles: 8052(pEGP1). Absorbance at 600nm was used as a measure of cell density (ii): Reorter activity was measured in the supernatant of cultures at early, mid and late exponential phase. Filled columns: 8052(pFNK1); Filled grey columns: 8052(pER1); Open columns: 8052(pEC1); Cross-hatched columns: 8052(pEGP1). Reporter activity is expressed as µmol p-nitrophenol released/min/cultire OD unit. All values represent the average of three independent experiments.

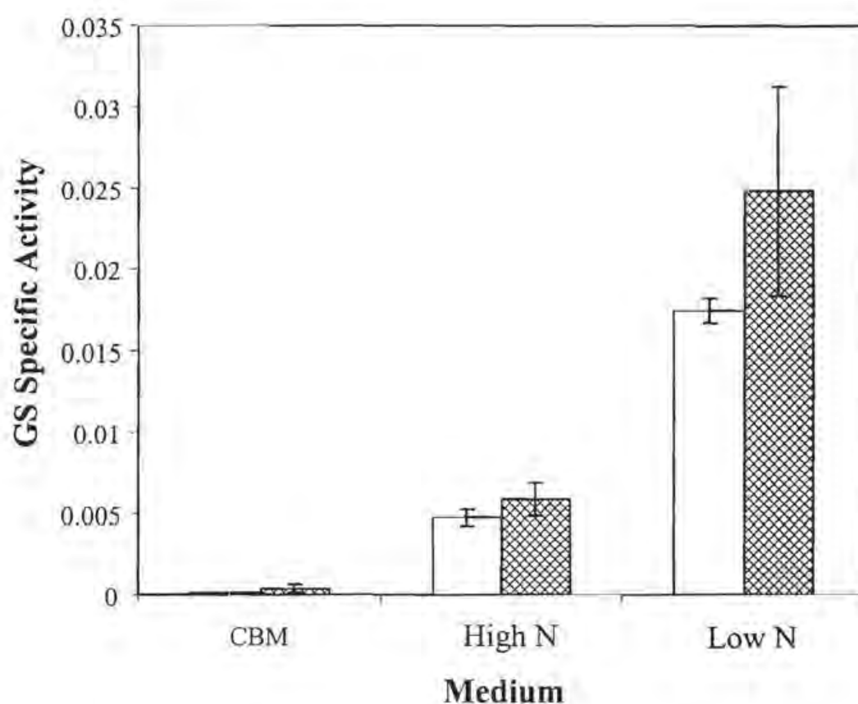


Figure 3.10: GS activity in 8052(pEC1) (open columns) and 8052(pEGP1) (cross-hatched columns). Samples were taken at mid-exponential phase from cultures used for reporter assays. Media used were: CBM; GSMM+0.2% Caa (high nitrogen); GSMM+0.05% Caa (low nitrogen). Specific activity is expressed as μmol γ-glutamate released/μg protein/min. All values represent the average of three independent experiments.

Growth of *C. beijerinckii* NCIMB 8052 (Fig 3.9) appeared to be unaffected by the presence or nature of the respective constructs, with the exception that growth of 8052(pER1) and 8052(pEC1) appeared to be slightly depressed, as was observed during previous experiments. Growth in minimal medium under conditions of different nitrogen concentrations was reflective of that of the untransformed wild-type strain (see chapter 2). GS activities of 8052(pEC1) and 8052(pEGP1) (Fig. 3.10) were essentially similar to those measured in untransformed cells, and were not significantly different from each other. The difference observed between the two strains grown under nitrogen-limiting conditions can be explained by the standard deviation observed for 8052(pEGP1). It would appear, therefore, that the presence of the *glnA* promoter, and indeed *eglA*, had no effect on the regulation of GS in the transformed strains.

Under all growth conditions, endoglucanase activity of 8052(pFNK1) and 8052(pER1) was marginally higher at mid-exponential phase than had previously been observed (activity at early exponential phase was not assayed previously). In minimal media, activity of 8052(pEC1) appeared lower than that in complex medium. These differences, however, are relatively small, and, especially in the light of the far higher activities observed for 8052(pEGP1), it could be argued that they merely reflect baseline fluctuations caused by inaccuracies in the experimental procedure.

The most notable observation here is that under the control of the *glnA* promoter, expression of *eglA*, as represented by endoglucanase activity, was far stronger than that observed for either the *scr* promoter (between 1000- and 100-fold) or the native *eglA* promoter (approximately 11-fold) at mid-exponential phase in complete media. This indicates that in this construct, the *glnA* promoter functions very strongly. However, the expected pattern of reporter expression for 8052(pEGP1) was not observed. During early exponential growth, cultures grown in nitrogen-limiting and nitrogen-rich media displayed similar reporter activities (130.17 and 138.17μM pNP/min/OD respectively). Cultures grown in complete medium produced almost two-fold higher values in early exponential growth than in those grown in minimal medium. In mid and late exponential phase, reporter activity was observed to increase in cultures grown in nitrogen-rich conditions, while those grown in nitrogen-limiting medium showed little increase in reporter activity above that observed in early exponential phase. These results are in direct contradiction with the reported pattern of GS activity; while GS activity is at its most repressed when cultures are grown in CBM, reporter activity was at its highest. GS assays conducted on 8052(pEC1) and 8052(pEGP1) confirm that this pattern was maintained during these experiments. It appears, therefore, that transcription from the cloned *glnA* promoter, while strong in minimal medium, is further induced under richer nutritional conditions.

The pattern of reporter expression observed here indicates that the reporter-promoter fusion is not being regulated co-ordinately with GS activity. Whether reporter activity is indicative of relatively high-level constitutive transcription driven by the cloned promoter fragment, or of induction by some factor associated with favourable growth and/or nutritional conditions, is unclear. Possible reasons for the absence of the expected pattern of regulation are numerous,

however the observation that GS is normally regulated eliminates several of these. The presence of the *glnA* promoter region and leader mRNA in multiple copies would be expected to titrate DNA-binding regulatory factors such as transcriptional repressors and enhancers, or RNA binding factors such as antiterminator proteins. It has been proposed that antisense RNA transcribed from a promoter within the *glnR* coding sequence negatively regulates *glnA* translation (see chapter 1). As the target for the antisense RNA binding site is present in multiple copies in this experiment, the antisense transcript would be titrated as well. The expected concomitant breakdown in the normal GS regulation was not observed, however. The result argues for the absence of a cis-acting factor on the cloned *glnA* promoter essential for normal regulation, although this would seem unlikely, as the fusion carries 486bp of sequence upstream of the *glnA* start codon, including all four of the proposed transcriptional start sites (see Fig. 3.3). The most likely explanation is that a cis-acting negative regulatory factor has been omitted from the construct, although this bodes ill for the antiterminator model of control, as the resultant large numbers of *glnA* leader transcript would be expected to titrate out the antiterminator and disrupt GS regulation.

An unlikely, but possible scenario is that the small portion of the *glnA* coding region is in some way directing transcription of the reporter independently of the promoter. An informative exercise would be to determine the start of transcription of the *glnA* promoter/reporter fusion to determine whether transcription is indeed being directed by the *glnA* promoter. It could be argued, however, that a more productive line of investigation would involve the cloning of different portions of the region upstream of *glnA* in front of the reporter to determine which regions are necessary for regulation.

One possibility that this experiment overlooks is the regulation of GS activity by post-translational modification, although it has been determined that this is unlikely (Woods & Reid, 1993). Nevertheless, this would account in part for the discrepancy between the regulation of the *glnA* promoter and GS activity. Quantitative Northern blot analysis of *glnA* transcription in 8052(pEGP1) might go some way to resolve this issue.

3.3.9 General comments on and suggestions for improvement of the reporter system

The *eglA* of *C. acetobutylicum* P262 is an attractive candidate for a reporter gene for *C. beijerinckii* NCIMB 8052 for several reasons. It has previously been used successfully as such in the anaerobic, Gram-negative organism *B. fragilis* (Abratt et al., 1993), and in addition has been shown to be expressed in *E. coli* (Zappe et al., 1986). The activity of the cloned enzyme in the heterologous hosts has been well characterised (Zappe et al., 1986; Abratt et al., 1993), and well-established, relatively easy and inexpensive colorimetric assays are available to measure it. Furthermore, qualitative (and to a limited degree, quantitative) assessment of activity can be performed using a plate assay, which is useful in identifying clones expressing the reporter gene. The gene was cloned from *C. acetobutylicum* P262, and for this reason, its codon composition would be expected to allow efficient expression in other clostridia. However, although it has been shown that the *C. acetobutylicum* P262 *eglA* is expressed and secreted efficiently by *C. beijerinckii* NCIMB 8052, the reporter system is far from perfect and the results presented here must be viewed in the light of the following points.

As the two strains are genetically very similar, it would be reasonable to assume that there is a homologue of *eglA* present in *C. beijerinckii* NCIMB 8052. The presence of an endogenous endoglucanase activity could complicate reporter assays, and for this reason pFNK1, with no reporter gene, was used as a negative control to eliminate this probability. Another potential problem arising from the presence of an *eglA* homologue is that recombination could occur between homologous sequences on the chromosome and the reporter plasmid. On occasions where plasmids were extracted from *C. beijerinckii* NCIMB 8052 and analysed by restriction endonuclease digestion, there appeared to be no structural rearrangements, which would be indicative of recombination, although this was not assayed quantitatively. Southern blot analysis would reveal whether a homologue is indeed present, however, it has been reported that *C. beijerinckii* possesses no detectable endoglucanase activity (Minton et al., 1993). It was previously reported that *C. acetobutylicum* ATCC 824 possessed a weak endoglucanase activity (Lee et al., 1985). A search of the incomplete genome sequence database of this organism using the tblastn program (Antshul et al., 1997) identified an open reading frame encoding an amino acid sequence which shares 51% identity with that of the *C. acetobutylicum* NCP 262 *eglA*.

In the experiments described here, the promoter-reporter fusion was carried on a multicopy plasmid. The vector pFNK1, on which the system was based, has a copy number of approximately 50 in *B. subtilis* (Monod et al., 1986), and its copy number in *C. beijerinckii* NCIMB 8052 would be expected to be similar. The presence of multiple copies is problematic in the investigation of promoter activity, particularly if titratable factors are involved, and could lead to spurious results. A solution would involve the integration of the promoter-reporter fusion into the chromosome as a single copy. This would approximate more closely the situation of the actual promoter being investigated.

The endoglucanase enzyme encoded by *eglA* is secreted into the medium. This is problematic for two reasons. First, a point of reference for determination of specific enzyme activity is difficult to obtain. Total protein content of the medium is unsuitable, as the nature and source of the proteins present is difficult, if not impossible to determine, as proteins may have been present as a constituent of the medium, or released as a result of cell lysis. Other enzymes may also have been secreted into the medium during culture growth, and proteins may have been removed from the medium by the cells. Quantitation of extracellular endoglucanase activity has traditionally involved the concentration and partial purification of the enzyme and determination of specific activity relative to the amount of protein present (Zappe et al., 1986; Kim et al., 1994). This approach is not practical for the purpose under discussion, for reasons primarily of convenience taking into account the number of samples involved. An alternative measure would be the number of viable cells present

Secondly, dilution of the enzyme in the medium could result in an undetectable level of activity. This was illustrated by the observation that in the case of pER1 and pESP1, visible zones of CMC degradation were observed when colonies were grown on plates, but no, or barely detectable, activity was measured in liquid culture assays. Although it could be argued that the vast differences observed between reporter activities in the culture supernatant of 8052-pEC1 and 8052-pEGP1 indicate that differential expression is easily detectable by the system as it stands, it is possible that it is not sufficiently sensitive to measure subtle differences and low-level regulation.

A possible suggestion for the improvement of the system is the disruption of the ability of the enzyme to be secreted. If the function of the enzyme is not dependent of its secretion, it may be possible to create a secretion-deficient mutant of EglA, which would remain within the cell. This would result in concentration of the enzyme activity, allowing detection of low activities, and increased reliability of readings. A signal sequence was identified in the N-terminal region of the polypeptide which is similar to signal sequences of secreted proteins in other Gram-positive bacteria (Zappe et al., 1988). A PCR-generated in-frame deletion of this region would be a possible means of generating a secretion-deficient mutant. As it is likely that this region is cleaved from the polypeptide to produce the mature protein (Zappe et al., 1988), it is likely that such a mutant would retain its enzymatic function.

3.4 CONCLUSIONS

To aid in the investigation of the regulation of expression of *glnA* in *C. beijerinckii* NCIMB 8052, a vector containing a promoterless reporter gene was developed for use in this organism. The endoglucanase (*eglA*) gene from *C. acetobutylicum* P262 was considered suitable as the basis for the system, and was cloned into the *Bacillus/Clostridium* shuttle vector pFNK1. Expression of *eglA* under its own promoter, in *C. beijerinckii* NCIMB 8052 was observed to remain at a relatively low level, and appeared to be repressed by cellobiose, the end-product of cellulose digestion, and, to a lesser extent, by fructose. It was concluded that the endoglucanase enzyme was secreted efficiently into the culture medium, and did not remain associated with the cell in any way. Enzyme activity was found to be relatively stable, with negligible loss of activity after storage at -20^{0}C, and pH and temperature optima were found to be in agreement with those previously reported for the cloned enzyme expressed in *E. coli* and *B. fragilis*.

In a preliminary attempt to use *eglA* as a reporter of *C. beijerinkii* NCIMB 8052 *glnA* and *scr* promoter activity, the respective promoters were cloned upstream of the *eglA* structural gene. While expression of the reporter under the control of these two promoters was not observed to follow patterns observed previously for these genes using other approaches, it was shown that in the case of the *glnA* promoter fusion, efficient endoglucanase expression and secretion

could be obtained under the control of a heterologous promoter. Further refinements are necessary to optimise the usefulness of this system.

CHAPTER 4

GENERAL CONCLUSIONS

In order to gain a better understanding of the processes governing nitrogen assimilation and metabolism in *C. acetobutylicum* NCP 262, an investigation was undertaken to characterise a gene, *glnR*, identified in this organism and thought to encode a transcriptional antiterminator protein, and regulator of *glnA* transcription. Initial attempts to reconstruct the regulatory system using cloned *glnA* and *glnR* genes in the heterologous host *E. coli* were unsuccessful. Attempts to locate accessory regulatory factors by sequencing the region upstream of *glnA* were also unsuccessful, although these investigations did identify an ORF which probably encodes an aspartokinase enzyme involved in the biosynthesis of threonine, methionine and lysine. It is therefore possible that the nitrogen assimilatory genes identified thus far form part of a larger cluster of amino acid biosynthesis genes.

C. acetobutylicum NCP 262 has so far proven to be recalcitrant to genetic manipulation. It was therefore deemed necessary to identify an organism which is more amenable to manipulation, as well as being sufficiently similar to *C. acetobutylicum* NCP 262, to serve as a model for nitrogen assimilation in this as well as, potentially, other solventogenic clostridia. *C. beijerinckii* NCIMB 8052 was chosen as a candidate for this role. It was shown to be similar to *C. acetobutylicum* NCP 262 with respect to its preferred nitrogen source, as well as regulation of GS activity in response to nitrogen availability. Homologues of genes involved in nitrogen assimilation in *C. acetobutylicum* NCP 262 were identified in *C. beijerinckii* NCIMB 8052 and were cloned and sequenced. A great deal of similarity was found between these genes in the two species with respect to sequence identity as well as their arrangement. It therefore appears that *C. beijerinckii* NCIMB 8052 is a suitable model organism for the study of nitrogen metabolism in *C. acetobutylicum* NCP 262. It is uncertain, however, how representative this model would be of other solventogenic clostridia, as analysis of the complete genome sequence of *C. acetobutylicum* ATCC 824 has failed to disclose any evidence of a similar nitrogen assimilation operon in this organism. It is possible that the

considerable diversity within this group of organisms may be reflected in this aspect of their biology.

An attempt to demonstrate a regulatory role for *glnR* by knocking out the gene in *C. beijerinckii* NCIMB 8052 proved unsuccessful. The start sites of *glnA* transcription in *C. beijerinckii* NCIMB 8052 under different conditions of nitrogen availability were identified. Transcription was shown to be initiated from four different points, suggesting four different promoters. As no consensus could be found between the proposed promoter regions, it is possible that each promoter responds to different factors, allowing increased flexibility of control. Transcription from all four start sites was found to be stronger under conditions of nitrogen limitation. It was also found the mRNA leader sequences in transcripts from three of the four promoters were able to form structures similar to those found in *C. acetobutylicum* NCP 262, and which are central to the antitermination model of control proposed for *glnA* in this species.

To broaden the range of tools available for the study of *C. beijerinckii* NCIMB 8052, a reporter system was developed, based on an endoglucanase gene (*eglA*) cloned form *C. acetobutylicum* NCP 262. Transcriptional fusions between the promoterless *eglA* and the promoters of the *C. beijerinckii* NCIMB 8052 sucrase operon (*scr*) and *glnA* promoter regions were constructed. Very low levels of *eglA* expression were observed in the case of the *scr* promoter, suggesting that the cloned promoter region excluded some factor essential for correct transcriptional regulation. Under the control of the *glnA* promoter, high levels of *eglA* expression were observed, however the pattern of regulation differed from what was expected. It is unclear whether this was due to a problem with the reporter system, the particular construct used, or the proposed model of control. Nevertheless, the reporter system shows great potential for use as a tool for investigating transcriptional regulation in *C. beijerinckii* NCIMB 8052, although some refinement is necessary.

Unfortunately, the primary aim of the study, to demonstrate a role for *glnR* in regulating *glnA* expression, was not achieved. However, both the establishment of *C. beijerinckii* NCIMB 8052 as a suitable alternative for study, and the preliminary work done in the development of

a reporter system for this organism may expand the range of tools available for studying both *C. acetobutylicum* NCP 262 and *C. beijerinckii* NCIMB 8052.

APPENDIX A

METHODS, MEDIA, BUFFERS AND SOLUTIONS

A.1 METHODS

A.1.1 NUCLEIC ACID EXTRACTION PROTOCOLS

A.1.1.1 Small-scale isolation of *E. coli* plasmid DNA (miniprep)

A 5ml culture was grown overnight in 2xYT (A.2.1.3) broth containing the relevant antibiotic(s). A 1.5ml volume of the culture was centrifuged in a 1.5ml microcentrifuge tube and resuspended in 200μl of solution I (A.3.1.1). After incubation at room temperature for 5 minutes, 400 μl of solution II (A.3.1.2) was added, and the mixture was vortexed and incubated on ice for 5 minutes. A 300μl volume of ice-cold solution III (A.3.1.3) was added and the sample was again vortexed briefly before being incubated on ice for a further 5 minutes. The sample was then centrifuged for 10 minutes at $4^{o}C$ to remove cellular debris and denatured chromosomal DNA. The supernatant was transferred to a clean microcentrifuge tube and an equal volume if isopropanol was added. DNA was harvested by centrifuging for 5 minutes, washed in 70% ethanol, dried, and resuspended in 50μl of TE buffer (A.3.19.4).

A.1.1.2 Alternative "quick" miniprep

A 5ml culture was grown up overnight in 2xYT (A.2.1.3) with the appropriate antibiotic(s). A 4ml aliquot of the culture was harvested in a 2ml microcentrifuge tube by two rounds of centrifugation and resuspended in 300μl TENS (A.3.2.1). Care was taken not to leave the sample in TENS for longer than 10 minutes. After the addition of 150μl of 3M sodium acetate (pH 5.4), the sample was vortexed briefly and centrifuged for 5 minutes at room temperature. The supernatant was transferred to a fresh microcentrifuge tube and 900μl of 100% ethanol was added. The sample was incubated for 20 minutes at $-20^{o}C$, centrifuged at $4^{o}C$ for 5 minutes, washed with 70% ethanol, dried and resuspended in 30μl TE buffer (A.3.19.4).

A.1.1.3 Large-scale isolation of plasmid DNA from *E. coli* (maxiprep)

Plasmid DNA was isolated and purified using a Nucleobond plasmid DNA extraction kit, following the manufacturer's instructions. The DNA concentration was determined spectrophotometrically using the relationship $A_{260} = 1$ for 50μg/ml of double-stranded DNA.

A.1.1.4 Small-scale isolation of plasmid DNA from *B. subtilis* (miniprep)

B. subtilis minipreps were performed according to Rodriguez & Tait (1987). A 1.5 ml culture was grown overnight in brain-heart infusion broth (Difco) containing the appropriate antibiotic and harvested by centrifugation for 1 minute in a 1.5ml microcentrifuge tube. The cells were washed in 1ml SET buffer (A.3.3.1) and then resuspended in 150μl SET, after which 20μl RNase buffer (A.3.3.2) and 50μl lysozyme (A.3.3.3) were added and the sample was vortexed briefly and placed on ice for 10 minutes. After this, 250μl sodium acetate buffer was added and the sample was inverted 15 times and incubated on ice for one hour to allow SDS and denatured chromosomal DNA to precipitate. The sample was centrifuged at $4^{O}C$ for 10 minutes and the supernatant was transferred to a fresh tube. One volume of isopropanol was added and the sample was centrifuged for 10 minutes at room temperature. The DNA pellet was washed with 70% ethanol and resuspended in 20μl of TE buffer (A.3.19.4).

A.1.1.5 Large-scale extraction of plasmid DNA from *B. subtilis* (maxiprep)

A 250ml culture was grown overnight in LB (A.2.1.1) with antibiotic selection and harvested by centrifugation in two Beckman GSA bottles at 7000 rpm for 10 minutes. The cells were washed in 50ml of TES buffer (pH 7.5) (A.3.4.1) and resuspended in 4ml lysozyme buffer (A.3.4.2). After this, 0.2ml lysozyme solution (10mg/ml in TES) was added and the sample was incubated for 30 minutes at $37^{O}C$. A volume (16ml) of SDS buffer (A.3.4.3) was added and the sample was mixed with a glass rod until it cleared, at which point 5ml of 5M sodium chloride was gently mixed in. The sample was incubated for 40 to 60 minutes at $-20^{O}C$ (or overnight on ice) and then centrifuged at 20 000 rpm for 30 minutes. The supernatant was decanted into a fresh tube and 5ml of a 50% solution of PEG 6000 was added. The sample was mixed gently and incubated on ice for at least an hour, but preferably overnight, before centrifugation at 10 000 rpm for 10 minutes. The supernatant was discarded and the pellet was resuspended in 2-3ml TES to which had been added 0.2ml of a 2mg/ml solution of RNase. After incubation at $65^{O}C$ for 30 minutes, 4.25ml of caesium chloride was added and

the tube was centrifuged for 5 minutes at 6000rpm to remove debris. To this was added 0.1ml of a 10mg/ml stock of ethidium bromide and the final volume was adjusted to 4.5ml. The refractive index was adjusted to 1.386 and the sample was loaded into a Beckman Quickseal tube and centrifuged overnight at 55 000 rpm. The band of plasmid DNA was removed from the tube and transferred to a microcentrifuge tube using a syringe and hypodermic needle. The ethidium bromide was removed by extracting 3 times with salt-saturated isopropanol. Two volumes of water were added and the DNA was precipitated by the addition of 1 total volume of isopropanol. The sample was left on ice for 10 minutes and centrifuged for 15 minutes. The DNA was resuspended in TE buffer (A.3.19.4) and the concentration measured spectrophotometrically.

A.1.1.6 Extraction of total genomic DNA from *C. acetobutylicum* NCP 262 and *C. beijerinckii* NCIMB 8052

DNA was extracted from *C. acetobutylicum* P262 and *C. beijerinckii* NCIMB 8052 following the method of Marmur (1961) as adapted by Zappe et al. (1986). One litre of CBM (A.2.1.5) in a Schott blue-top bottle was inoculated with spores and incubated at 37^{0}C until an OD_{600} of between 0.5 and 0.6 was reached. Cells were harvested by centrifugation for 20 minutes at 10 000 rpm in a Beckman GSA tube sealed well with Parafilm. The pellets were resuspended in a total volume of 20ml of CBM containing 1% sucrose, 12.5mM $MgCl_2$, 12.5mM $CaCl_2$ and 5mg/ml lysozyme, and transferred to a Beckman SS34 tube. The suspension was incubated anaerobically at 37^{0}C until approximately 90% of the cells had formed protoplasts, at which time 5ml of 0.5M EDTA was added and the suspension mixed by inversion for 5 seconds. Cells were lysed by the addition of 2ml of 25% SDS and the suspension was again mixed for 5 seconds. A 20ml volume of hot phenol (equilibrated to pH8.0 with 0.1M Tris-Cl; 50^{0}C) was added, mixed, and the tube centrifuged at 10 000 rpm for 10 minutes. The aqueous phase was recovered and extracted with an equal volume of chloroform:isoamyl alcohol (24:1). The tube was centrifuged at 15 000 rpm for 10 minutes and the top phase recovered and mixed with 0.2 ml of 5mg/ml RNase. The tube was incubated at room temperature for 30 minutes, after which 15ml of water-saturated diethyl ether was then added and the tube centrifuged at 15 000 rpm for 10 minutes. The top ether phase was discarded and the tube left to stand at 65^{0}C in the fume-hood until the remainder of the ether had evaporated. A one-tenth volume of 5M $NaClO_4$ was added, followed by one volume of isopropanol. The tube was centrifuged

at 15 000rpm for 15 minutes and the pellets washed in 70% ethanol. The pellets were resuspended in 0.5ml of TE and the DNA concentration determined spectrophotometrically.

A scaled-down version of this procedure was used to prepare small quantities of DNA for the screening of transformants for the presence of plasmids. All volumes were scaled down by a factor of 100, except for the final volume of TE in which the DNA was resuspended. It was found that if resuspended in 40µl, 3µl of the DNA was sufficient to visualise plasmids after restriction endonuclease digestion and electrophoresis.

A1.1.7 RNA extraction from *C. beijerinckii* NCIMB 8052

The following precautions were taken to avoid RNase contamination:

Milli-Q water was used exclusively.

All water and solutions (with the exception of those containing MOPS and Tris) were incubated overnight in a fume-hood with DEPC at a concentration of 0.1%, after which they were autoclaved.

All glass- and metalware was baked at 200^{0}C for at least 4 hours.

All plasticware was soaked in a 5% solution of Sodium Hypochlorite for at least 30 minutes, and then rinsed with DEPC-treated water.

Microfuge tubes and micropipette tips were used only from freshly opened bags.

Gloves were used at all times, and were changed regularly.

RNA was extracted from *C. beijerinckii* NCIMB 8052 following the protocol of Aiba et al. (1992). Spores were inoculated into 200ml of the culture medium and incubated until the required OD_{600} was reached. 25ml of culture was harvested by centrifugation at 5000 rpm for 8 minutes at 4^{0}C. The pellet was resuspended in 125µl of 0.3M sucrose/0.01M NaAc (pH 4.5), and transferred to a microfuge tube. 125µl of 0.01M sodium acetate (NaAc) (pH 4.5)/2% SDS was added and the tube was heated to 65^{0}C in a water bath for 1.5 minutes, and 250µl of hot phenol (65^{0}C, unequilibrated) was added to the tube, which was then vortexed and incubated for 3 minutes at 65^{0}C. The tube was then transferred to a -70^{0}C ethanol bath and incubated until the contents had frozen, and centrifuged for 5 minutes. The aqueous layer was recovered and extracted with phenol twice more. RNA was precipitated by adding 30µl (1/10 vol) of 5M NaAc and 900µl (3 vol) of ethanol, and the tube left at -70^{0}C for at least 5

minutes. At this point the tube may be stored overnight at -70°C. The tube was centrifuged for 10 minutes at 15 000 rpm at 4°C. The pellet was washed in 70% ethanol and centrifuged across the tube for a further 10 minutes. The RNA was resuspended in 180µl of DEPC-treated water and 20µl of 10X DNase buffer (A.3.5.1), 30U of RNase-free DNase was added and the reaction was incubated at room temperature for 30 minutes. To stop the reaction, 20µl of 250mM EDTA was added, and the mixture was then extracted by the addition of 220µl of phenol:chloroform (1:1). The tube was centrifuged and the aqueous layer was recovered, and further extracted with chloroform:isoamyl alcohol (24:1) to remove residual phenol. RNA was precipitated with 1/10 volume of 3M NaAc and 3 volumes of ethanol, and washed as above with 70% ethanol. The RNA was resuspended in 40µl DEPC-treated water and stored at -70°C. The concentration was determined spectrophotometrically using the relationship $A_{260} = 1$ for 40µg/ml of single-stranded RNA.

A.1.2 NUCLEIC ACID MANIPULATIONS

A.1.2.1 Restriction endonuclease digestion

DNA was digested according to the manufacturer's recommendations. Typically, a total reaction volume of 20µl was used, with approximately 3U of enzyme per 1µg of DNA. The restriction buffer supplied and recommended by the manufacturer was diluted 1/10, and the reaction was incubated at the recommended temperature for between 1 hour and overnight.

For double digestions using enzymes with compatible buffer requirements, reactions were simply incubated together. If the two restriction sites were close together, the DNA was divided into two reactions, one of which was digested with one enzyme, and one with the other for 3 hours. The two halves were then mixed and a second dose of both enzyme was added and the reactions incubated for a further 3 hours. If the buffer requirements of the two enzymes were incompatible, after digestion with the first enzyme, the DNA was subjected to phenol:chloroform extraction (A.3.7.2) and then digested with the second. DNA to be used for ligations was purified by phenol:chloroform extraction to remove the enzyme.

A.1.2.2 DNA ligation

DNA ligations were typically performed in a total volume of 20µl using T4 DNA polymerase (Roche). The supplied ligase buffer was diluted 1/10 as recommended. A total amount of 10pmol of DNA was used per reaction, with 1.0U of ligase and a vector:insert ratio of 1:4 for blunt-ended, and 0.1U and a vector:insert ratio of 4:1 for sticky-ended ligations. DNA concentration was calculated using the relationship of 1pmol = (0.662 x kb)µg. For blunt-ended ligations, reactions were incubated at 25°C overnight, while sticky-ended ligation reactions were incubated overnight at 16°C.

A1.2.3 Rapid DNA ligation in agarose

In this procedure (Struhl, 1985), DNA fragments were separated in 1.0% - 2.0% low-melting point (LMP) agarose made with, and run in TA buffer (see A.3.7.2). The required bands were visualized under long-wavelength UV, and excised in as small a volume of agarose as possible. The agarose was melted at 70°C for 5 minutes, and then equilibrated at 37°C. A quantity of 2µl agarose containing vector DNA and 8µl of insert DNA were added to 10 µl of a pre- warmed (37°C) ligation mixture consisting of ligase buffer, ligase and water made up to an appropriate concentration for a final volume of 20µl. The reaction was incubated for a minimum of 3 hours at room temperature. Before transformation, the ligation mixture was incubated at 70°C for 5 minutes and then equilibrated at 37°C before being added to the competent cell suspension.

A.1.2.4 Filling in of 5' overhangs using Klenow fragment

For the filling in of 5' single stranded overhangs created by restriction endonuclease digestion or PCR amplification, the following reaction was used (Sambrook et al., 1989): A quantity of 1µg of DNA was mixed with 2µl of 25mM dNTP mixture, 2µl of Klenow buffer supplied by the manufacturer (Roche), 1U of Klenow fragment and the mixture was made up to 20µl with water. The reaction was incubated at 37°C for 30 minutes, and the DNA was subjected to phenol:chloroform extraction (A.3.7.2) to remove the enzyme.

A.1.2.5 Removal of 3' overhangs using T4 DNA polymerase

The 3'-5' single-stranded exonuclease activity of T4 polymerase may be used to remove 3' single stranded overhangs produced by restriction endonuclease digestion (Sambrook et al.,

1989). 1μg of DNA was mixed with 2μl 0.1% BSA, 2μl restriction endonuclease buffer A (Roche), 2μl 25mM dNTP mixture, 1U of enzyme and the total volume made up to 20μl with water. The reaction was incubated for 30 minutes at 37°C. This reaction was also used for the filling in of 3' single-stranded overhangs. For both removal and filling in of overhangs, excess dNTP's were included in the reaction mixtures to inhibit double stranded exonuclease activity of the enzyme, as recommended by Sambrook et al. (1989).

A.1.2.6 Henikoff shortening of DNA

Terminal deletions of clones were produced following the method of Henikoff (1984). DNA was divided into two aliquots of 6μl each for restriciton digestion to produce the required single-stranded overhangs. One of the enzymes was used to cut each aliquot, and digestion was checked by transformation without ligation. The aliquots were mixed and digested with a twofold excess of both enzymes. Digestion was again checked by transformation, this time with a ligation step. Proteins were precipitated by the addition of 1/2 volume of 7.5M ammonium acetate (pH 7.5) and incubated at room temperature for 10-30 minutes before centrifugation at 15 000 rpm for 15 minutes. The supernatant was transferred to a new tube and the DNA precipitated by the addition of 2.5 volumes of 100% ethanol, incubation at room temperature for 10 minutes and centrifugation at 15 000 rpm for 15 minutes. The pellet was then washed in 70% ethanol, dried and resuspended in 100μl of exonuclease III buffer (Roche).

The mixture was equilibrated at 37°C for 5 minutes, and a 4.5μl sample was removed as a t_0 sample and placed in a tube containing 12.5μl of S1 nuclease mix (Roche) on ice. To the rest of the mixture was added 150U of exoIII and mixed well. At 20 second intervals subsequent to the addition of the exoIII, 4.5μl samples were transferred to tubes of S1 nuclease mix and mixed well. The S1 nuclease mixes were incubated for 30 minutes at room temperature, after which 1.75μl of S1 nuclease stop (A.3.6.1) was added to each of the tubes, which were then incubated at 70°C for 10 minutes. A 2μl aliquot was removed from every alternate sample and run on a gel to check digestion. In the meantime, 1.7μl of Klenow buffer (Roche) and 1U of Klenow fragment was added to each tube and equilibrated at room temperature for 3 minutes. After this, 1μl of 25mM dNTP mix was added to each tube and the reaction incubated at room temperature for 5 minutes. Samples of the mixtures of shortened DNA

were ligated and transformed into *E. coli*. Quantities of DNA to be ligated were determined empirically.

A.1.2.7 Phenol-chloroform cleanup of DNA

To remove contaminating proteins from a DNA solution, the volume of the solution was made up to 180µl with TE, and 20µl of phenol (A.3.19.3) was added and the mixture was emulsified by shaking. One volume of chloroform:isoamyl alcohol (24:1) was added and the mixture again emulsified. The mixture was centrifuged at 15 000 rpm for 5 minutes and the top, aqueous phase recovered. A one-tenth volume of 5M $NaClO_4$ was added, followed by an equal volume of isopropanol. The mixture was centrifuged at 15 000 rpm for 15 minutes, washed in 70% ethanol and dried for 5 minutes in a vacuum. The pellet was resuspended in the required volume of TE (A.3.19.4) or water.

A.1.3 ELECTROPHORESIS OF NUCLEIC ACIDS

A.1.3.1 Agarose gel electrophoresis of DNA

DNA was separated according to size using agarose gels ranging in density between 0.8% and 2.0% w/v agarose, depending on the size of fragments to be separated. Agarose was mixed with 1xTAE (A.3.7.1), or, if subsequent manipulations were to be performed in agarose (A.1.2.3), 1xTA (A.3.7.2). The mixture was heated to boiling point in a microwave oven, and then left to cool to approximately 50^0C. Ethidium bromide was added to a final concentration of 0.2µg/ml, before the gel was poured and left to solidify.

Sample buffer (A.6.7.3) was mixed with the DNA samples before loading, to a final concentration of 1x. Gels were run in tanks containing 1xTAE or 1xTA at between 15 and 120V, depending on the size of the tank. DNA bands were visualized on a UV transillurninator at 254nm, or 310nm if the DNA was to be recovered. Photographs were taken using a Pharmacia Biotech Gel Display System. Phage λ DNA digested with either *Pst*I or *Hind*III was run with the samples as a size standard. This was used to plot a standard curve of size vs. distance travelled to calculate sample DNA fragment sizes.

A.1.3.2 Denaturing RNA gel electrophoresis

The procedure used was that of Fourney et al. (1988). Tenfold concentrated MOPS (A.3.8.1) was diluted to a final concentration of 1x in water, and agarose was added to the solution to a concentration of 1.5%. The mixture was heated to boiling point in a microwave oven, and left to cool to approximately 70^0C before the addition of 37% formaldehyde to a final concentration of 18,87%. The mixture was poured into an RNase-free gel tray and allowed to set for at least one hour. The gel was then placed in an RNase-free electrophoresis tank containing 1xMOPS. The wells were flushed with 1xMOPS prior to loading of samples. For checking RNA concentrations, 10μg was loaded per lane, while for Northern hybridisation, 30μg per lane was loaded. Volumes used were between 5 and 10μl. If the sample volume was less than 5μl, the difference was made up using water. Prior to loading, 20μl of sample buffer (A.3.8.2) was added and the samples heated to 65^0C for 15 minutes and then flash-frozen and kept on ice until the gel was loaded. The gel was run at 50V for 4 hours. RNA was visualized on a 254nm ultraviolet transilluminator.

A.1.4 NUCLEIC ACID HYBRIDISATION

All capillary transfers, colony blots and nucleic acid hybridisations were carried out using the DIG non-radioactive nucleic acid detection kit, according to methods recommended by the manufacturer (Roche).

A.1.5 BACTERIAL TRANSFORMATION TECHNIQUES

A.1.5.1 $RbCl_2$ Competent cell transformation of *E. coli*

A 5ml volume of Ψ broth (A.2.2.2) was inoculated with a single colony of bacteria to be transformed, and incubated overnight at 37^0C with aeration. The entire 5ml was used to inoculate 200ml of Ψ broth pre-warmed to 37^0C. The culture was grown to mid-exponential phase (OD_{600} = 0.35), at which time the cells were transferred to a Beckman GSA tube and chilled on ice for 15 minutes. The cells were harvested by centrifugation at 2500 rpm for 15 minutes at 4^0C. The cells were washed in a total volume of 21ml of ice-cold TFB1 buffer

(A.3.12.1), transferred to a Beckman SS34 tube and incubated on ice for 90 minutes. The cells were centrifuged again at 2500 rpm for 15 minutes at 4^0C and gently resuspended in a total volume of 9ml TFB2 (A.3.12.2). The suspension was divided into 100μl aliquots, which were used for transformations or stored at -70^0C.

For the actual transformation, cells were thawed on ice , after which DNA was added and the mixture incubated on ice for 20 minutes. The cells were heat-shocked at 37^0C for 5 minutes, and returned to the ice for a further 2 minutes. A volume of 800μl of broth, pre-warmed to 37^0C, was added to the cells, which were incubated at 37^0C for 45 minutes to allow expression of selectable marker genes. The cells were finally plated onto LA (A.2.1.2) with appropriate selection and incubated at 37^0C until colonies were observed.

A.1.5.2 *B. subtilis* protoplasting and transformation

B. subtilis protoplasts were generated and transformed following the procedure of Chang & Cohen (1979). A volume of 5ml of LB (A.2.1.1) was inoculated with a single colony of bacteria, incubated overnight at 37^0C, and used to inoculate 50ml of LB. The culture was grown to an OD_{600} of between 0.4 and 0.5. The cells were harvested by centrifugation in a blue-top Sterilin tube at 4000 rpm for 15 minutes and then resuspended in 5ml of SMMP (A.3.13.2). Lysozyme was added to a final concentration of 1mg/ml, and the cells were incubated at 37^0C with gentle agitation until the majority of the cells had formed protoplasts. The protoplasts were harvested by centrifugation at 4000 rpm for 15 minutes, washed in 5ml of SMMP and resuspended in 5ml of SMMP. The protoplasts were divided into 500μl aliquots for immediate use or storage at -70^0C.

DNA was mixed with an equal volume of 2xSMM (A.3.13.2) in white-top Sterilin tubes. 500μl of protoplasts were added and mixed gently. Immediately after this, 1.5ml of 40% PEG 6000 (A.3.13.3) was added and mixed by gentle bubbling with a micropipette. Two minutes later, 5ml SMMP was added and mixed in by gentle inversion of the tube. The protoplasts were recovered by centrifugation at 4000 rpm for 15 minutes, resuspended in 1ml of SMMP, and incubated at 30^0C with gentle agitation for at least 90 minutes. The cells were plated out on fresh DM3 medium (A.2.2.3) with the appropriate antibiotic selection and incubated at 37^0C until colonies were visible (usually at least 48 hours).

A.1.5.3 *C. beijerinckii* NCIMB 8052 electroporation

A 200ml volume of CBM was inoculated with spores and incubated at 37^{0}C until an OD_{600} of 0.5 was reached. The culture was transferred to a Beckman GSA tube which was sealed with Parafilm and centrifuged for 9 minutes at 7000 rpm at 4^{0}C. The cells were washed first in 10ml and then in 20ml of ice-cold electroporation buffer (A.1.14.1). The cells were finally resuspended in a total volume of 2.5ml of ice-cold electroporation buffer and incubated on ice for 10 minutes. An aliquot of 300µl of cells was mixed with the DNA in an ice-cold 0.2 cm^3 electroporation cuvette, which was sealed with Parafilm and incubated on ice for 10 minutes. the cells were electroporated using a Bio-Rad electroporator set to the following parameters: 1.25kV, 400Ω, 2.5µF. A time constant of 4.0 was considered adequate. After electroporation the cells were placed on ice for 10 minutes before being mixed with 700µl of CBM (A.2.1.5) and removed from the cuvettes. The cells were incubated at 37^{0}C for approximately 90 minutes before being plated out on CBM agar (A.2.1.6) containing the appropriate antibiotic selection.

A.1.5.4 Conjugative transfer of plasmids from *E. coli* to *C. beijerinckii* NCIMB 8052

Plasmids were transferred into *C. beijerinckii* NCIMB 8052 using a procedure adapted from Williams et al. (1990b). *E. coli* CA448 which had been transformed with the plasmid to be transferred was grown to stationary phase in a 5ml culture of 2xYT with appropriate antibiotic selection, and diluted 1/50 in 2xYT (A.2.1.3) with antibiotic selection. The culture was grown overnight to stationary phase (OD_{600} = 4.0). A 100ml culture of *C. beijerinckii* NCIMB 8052 was grown to late exponential phase (OD_{600} = 1.2) in CBM (A.2.1.5). After the *E. coli* cultures had been left to become anaerobic for at least an hour, they were mixed with cultures of *C. beijerinckii* NCIMB 8052 in a sterile 10ml centrifuge in a donor-recipient ratio of 10:1, centrifuged at 6000rpm for 6 minutes, resuspended in 100µl of CBM, and plated onto CBM agar plates with no antibiotic selection. After incubation overnight at 30^{0}C, the cells were washed off the plates with 800ml of CBM. The cells were then plated onto CBM agar plates (130µl/plate) containing erythromycin at a concentration of 20µg/ml, and incubated at 42^{0}C until colonies were observed. Numbers of viable recipient cells were determined by serial dilution of the suspension washed off the plates, and plating onto CBM without antibiotic selection.

A.1.6 ENZYME ASSAYS

A.1.6.1 Glutamine synthetase

Glutamine synthetase activity was assayed using the γ-glutamyl transferase reaction (Shapiro & Stadtman, 1970). For *E. coli*, 4 ml of a 5ml overnight preculture was used to inoculate 200ml of the appropriate medium, while in the case of *C. beijerinckii* NCIMB 8052, 200ml of medium was inoculated with spores. The culture was grown up to the required OD_{600}, at which point 25ml was transferred to a centrifuge tube. To permeabilise the cells, 250μl of a 1% aqueous solution of CTAB was added and the tube was gently agitated for exactly 8 minutes before centrifugation at 6000 rpm (for *E. coli*) or 10 000 rpm (for *C. beijerinckii*) for 15 minutes. The cells were washed in 25ml of 0.8% NaCl solution and then resuspended in 500μl of ice-cold SB buffer (A.3.15.1) and incubated on ice. In the case of *C. beijerinckii*, everything up until this point was performed under strictly anaerobic conditions.

A 100μl aliquot of the cell suspension was mixed with either reaction (A..3.15.2) or blank mixture (A.3.15.3), in microfuge tubes, and equilibrated at 37°C for 5 minutes. To start the reaction, 50μl of 0.2M glutamine, which had also been equilibrated at 37°C for 5 minutes, was added and the reaction incubated at 37°C for 15 minutes. The reaction was stopped by the addition of 1ml of stop buffer (A.3.15.4), and the cells removed from the reaction mixture by centrifugation at 15 000rpm for 5 minutes. The absorbance at 540nm was read immediately.

The Bio-Rad protein micro-assay was used to determine the total protein concentration of the reaction mixture. The assay was performed on an empirically-determined dilution of an aliquot of the prepared cells.

A.1.6.2 Endo-1,4-β-glucanase

For assaying of intracellular endoglucanase activity, 100ml of culture was harvested by centrifugation at 7000rpm for 10 minutes at 4°C. The pellet was washed in 50ml and then resuspended in 5ml of ice-cold PC buffer (A.3.17.1). Cell lysis was achieved by sonication, and cell debris and protein precipitate was removed after the assay by centrifugation.

The cell lysate was equilibrated at 50^{0}C for five minutes, as was the reaction substrate (3.4mM p-NPC in PC buffer). A 125µl aliquot of the lysate was then mixed with an equal volume of substrate mixture and incubated for exactly 30 minutes, after which the reaction was stopped with 250µl of 14% Na_2CO_3. A 500µl volume of water was added and the absorbance at 405nm was measured. A substrate blank consisting of the substrate mixture with PC buffer only, and an enzyme blank consisting of lysate incubated with PC buffer only were incubated as for the assay mixtures. The absorbances were read and subtracted from the values obtained for the assay mixtures. A standard curve was constructed and used to determine the amount of p-nitrophenol liberated in the reaction. Protein concentrations were determined using the Bio-Rad protein microassay, which was performed on the cell lysate. Enzyme activity was expressed as micromoles of p-nitrophenol liberated per microgram of protein per minute.

For assaying of secreted endoglucanase activity in the culture medium, a 1ml sample of the medium was taken and the cells removed by centrifugation in a bench-top centrifuge for 10 minutes. The supernatant was used without any further modification. The assay was conducted as described above, with the following modifications: The substrate mixture consisted of 3.4mM p-NPC dissolved in 2xPC buffer, the substrate blank consisted of substrate mixture mixed with distilled water, and the enzyme blank consisted of culture supernatant mixed with 2xPC buffer. Enzyme activity was expressed as micromoles of p-nitrophenol released per culture OD unit per minute.

A.1.7 Primer extension

RNA was extracted from the relevant cultures, 100µg was precipitated, and resuspended in 100µl of HP buffer (A.3.19.1). Solution was transferred to a PCR tube and between 2 and 5 pmol of cyanidin-labelled primer was added. The mixture was heated at 95^{0}C for 10 minutes in a PCR machine, after which the primer was annealed at 37^{0}C overnight. The RNA with annealed primer was transferred to a standard 1.5ml microcentrifuge tube, and precipitated by the addition of 800µl of ice-cold ethanol, and incubation at -20^{0}C for 30 minutes. The tube was centrifuged for 12 minutes at 4^{0}C, washed with 70% ethanol, and centrifuged for 8 minutes at 4^{0}C. The pellet was dried , and dissolved in 20µl of RTB buffer (A.3.19.2), and the mixture warmed to 42^{0}C for 4 minutes. Reverse transcriptase (20U) was added, and the

reaction incubated at 45°C for 2 hours. The reaction was stopped by the addition of 1µl of 0.5M EDTA (pH8.0), and 1µl of a 10mg/ml stock of RNase was added. The mixture was incubated at 37°C for 30 minutes. The primer extension product was precipitated by the addition of 150µl TES (A.3.4.1) and 500µl of ethanol, and incubation at -20°C for 1 hour. The pellet was washed in 70% ethanol, dried and resuspended in 5µl TE, to which 8µl of ALF stop buffer (Pharmacia) was added. The mixture was then heated to 95°C for 5 minutes before being run on an ALF automated DNA sequencer (Pharmacia) along with the products of the appropriate sequencing reactions.

A.2 MEDIA

A.2.1 Standard growth media

A.2.1.1 Luria Broth (LB)

Ingredient	Final concentration	Per litre
Tryptone	1.0%	10g
Yeast Extract	0.5%	5g
NaCl	0.5%	5g

Make up in distilled water, autoclave.

A.2.1.2 Luria-Bertani Agar (LA)

Ingredient	Final concentration	Per litre
Tryptone	1.0%	10g
Yeast Extract	0.5%	5g
NaCl	0.5%	5g
Agar	1.3%	13g

Make up in distilled water, autoclave.

A.2.1.3 2X Yeast-tryptone broth (2xYT)

Ingredient	Final concentration	Per litre
Tryptone	1.6%	16g
Yeast Extract	1.0%	10g
NaCl	0.5%	5g

Make up in distilled water, autoclave.

A.2.1.4 2X Yeast-tryptone Agar (2xYT Agar)

Ingredient	Final concentration	Per litre
Tryptone	1.6%	16g
Yeast Extract	1.0%	10g
NaCl	0.5%	5g
Agar	1.3%	13g

Make up in distilled water, autoclave. Cool to 50^{0}C before adding antibiotic (if appropriate).

A.2.1.5 Clostridium Basal Medium (CBM)

Ingredient	Final concentration	Per litre
Glucose	1.0%	10g
Casamino Acids	0.4%	4g
Yeast Extract	0.4%	4g
$MgSO_4.7H_2O$	0.02%	1ml of 20% stock
$MnSO_4.4H_2O$	0.001%	1ml of 1% stock
$FeSO_4.7H_2O$	0.001%	1ml of 1% stock*
p-Aminobenzoic Acid	0.0001%	1ml of 0.1%
Biotin	$2x10^{-7}$%	1ml of 0.0002% stock
Thiamine-HCl	0.0001%	1ml of 0.1% stock
Cysteine-HCl	0.05%	0.5g
$NaHCO_3$	0.1%	1g
Resazurin	0.004%	2ml of 2% stock

All stocks autoclaved and stored at 4^{0}C, except * (Made up fresh).
Make up in distilled water, autoclave.

A.2.1.6 Clostridium Basal Medium Agar (CBM Agar)

Ingredient	Final concentration	Per litre
Glucose	1.0%	10g
Casamino Acids	0.4%	4g
Yeast Extract	0.4%	4g
$MgSO_4.7H_2O$	0.02%	1ml of 20% stock
$MnSO_4.4H_2O$	0.001%	1ml of 1% stock
$FeSO_4.7H_2O$	0.001%	1ml of 1% stock*
p-Aminobenzoic Acid	0.0001%	1ml of 0.1%
Biotin	$2x10^{-7}$%	1ml of 0.0002% stock
Thiamine-HCl	0.0001%	1ml of 0.1% stock
Cysteine-HCl	0.05%	0.5g
$NaHCO_3$	0.1%	1g
Resazurin	0.004%	2ml of 2% stock
Agar	1.5%	15g

All stocks autoclaved and stored at 4^{0}C, except * (Made up fresh).
Make up in distilled water, autoclave. Cool to 50^{0}C before adding antibiotic (if appropriate).

A.1.2.1.7 Ψ broth

Ingredient	Final concentration	Per litre
Tryptone	2%	20g
Yeast Extract	0.5%	5g
$MgSO_4.7H_2O$	0.4%	4g
KCl	10mM	0.75g

Autoclave

A.2.2 Minimal media

A.2.2.1 Glucose-Salts Minimal Medium (GSMM)

Ingredient	Final concentration	Per litre
Glucose	2.0%	20g
Casamino Acids	0.2% (N-excess); 0.025% (N-limiting)	2g 0.25g
Salts	1x	100ml of 10x stock[1]
$MgSO_4.7H_2O$	0.01%	0.5ml of 20% stock
$MnSO_4.4H_2O$	0.0005%	0.25ml of 1% stock
$FeSO_4.7H_2O$	0.0005%	0.25ml of 1% stock *
$Na_2MoO_4.2H_2O$	0.0005%	0.5m of 1% stock
$ZnSO_4$-$7H_2O$	0.0025%	0.25ml of 5% stock
p-Aminobenzoic Acid	0.00005%	0.5ml of 0.1% stock
Biotin	$1x10^{-7}$%	0.25ml of 0.02% stock
Thiamine-HCl	0.00005%	0.5ml of 0.1% stock
Cysteine-HCl	0.05%	0.5g
$NaHCO_3$	0.1%	1.0g
Vitamins	1x	0.1ml of 10 000x stock[2]
Resazurin	0.004%	2ml of 2% stock

[1] 10x Salts

Ingredient	Final concentration	Per litre
$CaCl_2$	0.02%	0.2g
K_2HPO_4	0.1%	1g
KH_2PO_4	0.1%	1g
NaCl	0.2%	2g

Autoclave 10x salts and GSMM separately.

[2] Vitamins

Ingredient	Final concentration	Per 100ml
Ca-D-pantothenate	0.02%	20mg
Nicotinamide	0.02%	20mg
Riboflavin	0.02%	20mg
Pyroxidine-HCl	0.02%	20mg
Folic Acid	0.00025%	0.25mg
Vitamin B_{12}	0.0001%	0.1mg

Filter-sterilize, add after autoclaving.
Make up in distilled water, autoclave. For solid media, add 1.5% agar.

A.2.2.2 M9 Minimal medium

Ingredient	Final concentration	Per litre
Glucose	0.2%	2g
M9 Salts	1x	100ml of 10x stock[1]
$MgSO_4$	1mM	1ml of 1M stock
$CaCl_2$	0.1mM	1m of 0.1M stock
Thiamine-HCl	0.001%	1ml of 0.1% stock
$(NH_4)_2SO_4$	0.1%	1g

[1] 10x M9 Salts

Ingredient	Final concentration	Per litre
Na_2HPO_4	6.0%	60g
KH_2PO_4	3.0%	30g
NaCl	0.5%	5g

Autoclave Salts separately. Amino acids supplemented at a concentration of 0.2%. For solid media, add 1.3% agar.

DM3 Medium

Ingredient	Final concentration	Per litre
Sodium succinate	0.5M	500ml of 1M stock
Casamino acids	0.5%	100ml of 5% stock
Yeast extract	0.5%	50ml of 10% stock
KH_2PO_4	0.15%	100ml of 1.5%/3.5% stock
K_2HPO_4	0.35%	
$MgCl_2$	0.02M	20ml of 1M stock
Glucose	0.5%	25ml of 20% stock
BSA	0.01%	5ml of 2% stock*
Agar	0.8%	8g in 200ml water

* Filter sterilize BSA, store at -20°C.
Autoclave all separately, add BSA once medium has cooled to 50°C.

A.3 BUFFERS AND SOLUTIONS

A.3.1 Small-scale *E. coli* plasmid isolation solutions

A.3.1.1 Solution I

Ingredient	Final concentration	Per 100ml
Tris-HCl pH8.0	0.25M	25ml of 1M stock
Glucose	0.5M	45.5ml of 20% stock
EDTA (pH8.0)	0.1M	20ml of 0.5M stock

Autoclave glucose separately; Make up to 100ml with distilled water.

A.3.1.2 Solution II

Ingredient	Final concentration	Per 100ml
NaOH	0.2N	2ml of 10N stock
SDS	1.0%	4ml of 25% stock

Make up to 100ml with distilled water; Make fresh every week.

A.3.1.3 Solution III

Ingredient	Final concentration	Per 100ml
K-Acetate	3M	147g
Acetic acid	2M	Add to pH 4.8

Dissolve K-Acetate in 250ml of water, adjust pH with glacial acetic acid and then make up to 100ml with water.

A.3.2 "Quick" miniprep solutions

A.3.2.1 Tris-EDTA-Sodium hydroxide-SDS (TENS) solution

Ingredient	Final concentration	Per 10ml
SDS	0.5%	200μl of 25% stock
NaOH	0.1N	100μl of 10N stock
Tris-HCl pH 8.0	10mM	100μl of 1M stock
EDTA (pH8.0)	0.1mM	20μl of 0.5M stock

Make up to 10ml with water; Make up fresh every week.

A.3.3 *B. subtilis* miniprep solutions

A.3.3.1 Sucrose-EDTA-Tris (SET) buffer

Ingredient	Final concentration	Per 100ml
Sucrose	20%	20g
Tris-HCl pH 7.6	50mM	5ml of 1M stock
EDTA (pH8.0)	50mM	1ml 0.5M stock

A.3.3.2 RNase buffer

Ingredient	Final concentration	Per 10ml
Rnase H	10mg/ml	100mg
NaAc	0.1M	
EDTA (pH8.0)	0.3M	6 ml of 0.5M stock

Adjust pH to 4.8 with glacial acetic acid, heat to 80^{0}C for 10 minutes, filter-sterilise and store 1ml aliquots at -20^{0}C.

A.3.3.3 Lysozyme buffer

Ingredient	Final concentration	Per 10ml
Lysozyme	5mg/ml	50mg
Tris-HCl pH 7.6	10mM	100µl of 1M stock
EDTA	1mM	0µl of 0.5M stock
NaCl	10mM	g

Make up to 10ml with water, filter-sterilise and store 1ml aliquots at -20^{0}C.

A.3.3.4 Lytic mixture

Ingredient	Final concentration	Per 20ml
SDS	1%	800µl of 25% stock
NaOH	0.2N	400µl of 10N stock

Add ingredients to sterile water; Make fresh every week.

A.3.3.5 Sodium acetate buffer

Ingredient	Final concentration	Per 100ml
NaAc	3M	

Adjust pH to 4.8 with glacial acetic acid, autoclave.

A.3.4 *B. subtilis* maxiprep solutions

A.3.4.1 Tris-EDTA-Sodium chloride (TES) buffer (10X stock)

Ingredient	Final concentration	Per 400ml
Tris-HCl pH 7.6	100mM	400 μl of 1M stock
NaCl	100mM	2.3376g
EDTA	10mM	8ml of 0.5M stock

Make up to 400ml with water, autoclave.

A.3.4.2 Lysozyme buffer

Ingredient	Final concentration	Per 200ml
Tris-HCl pH 8.0	30mM	16ml of 1M stock
EDTA	50mM	20ml of 0.5M stock
NaCl	50mM	2ml of 5M stock
Sucrose	25%	50g

Make up to 200ml with water, autoclave.

A.3.4.3 SDS buffer

Ingredient	Final concentration	Per 400ml
SDS	10%	40g
EDTA	50mM	40ml of 0.5M stock
Tris-HCl pH 8.0	100mM	4ml of 1M stock

Use sterile stocks and water.

A.3.5 RNA extraction solutions

A.3.5.1 10x DNase buffer

Ingredient	Final concentration	Per 50ml
NaAc	200mM	10ml of 1M stock
$MgCl_2$	100mM	5ml of 1M stock
NaCl	100mM	5ml of 1M stock

Make up in RNase-free bottle with RNase-free ingredients.

A.3.6 Heinikoff shortening solutions

A.3.6.1 S1 nuclease stop buffer

Ingredient	Final concentration	Per 100ml
Tris Base	0.3M	
EDTA	0.05M	10ml of 0.5M stock

Autoclave; Store 1ml aliquots at -20°C.

A.3.7 DNA electrophoresis solutions

A.3.6.1 50X Tris-acetate-EDTA (TAE) buffer

Ingredient	Final concentration	Per 1l
Tris base	1M	242g
Glacial acetic acid		57.1ml
EDTA	0.05M	100ml of 0.5M stock

Autoclave

A.3.7.2 50x Tris-Acetate (TA) buffer

Ingredient	Final concentration	Per 1l
Tris	1M	242g
Glacial acetic acid		57.1ml

Autoclave

A.3.7.3 Sample buffer

Ingredient	Final concentration	Per 100ml
Bromophenol blue	0.25%	0.25g
Sucrose	40%	40g
EDTA	20mM	4ml of 0.5M stock

Autoclave, store at -20°C.

A.3.8 RNA electrophoresis solutions

A.3.8.1 10x MOPS

Ingredient	Final concentration	Per 500ml
MOPS	200mM	20.93g
NaAc	50mM	3.402g
EDTA	10mM	1.8612g

Make up with RNase-free ingredients and water. Adjust pH to 7.0 with glacial acetic acid (RNase free).

A.3.8.2 Sample buffer

Ingredient	Per 1.5ml
Deionized formamide	750μl
Formaldehyde	240μl of 37% solution
MOPS	150μl of 10x stock
Glycerol	200μl of 50% stock
Ethidium Bromide	10μl of 10mg/ml stock
Bromophenol Blue	Add 10% stock dropwise until dark blue

Make up to 1.5ml with RNase-free water.

A.3.9 Southern hybridisation solutions

All solutions were prepared as instructed by the manufacturers of the DIG non-radioactive nucleic acid detection system (Roche)

A.3.10 Northern hybridisation solutions

All solutions were prepared as instructed by the manufacturers of the DIG non-radioactive nucleic acid detection system (Roche)

A.3.11 Colony hybridisation solutions

All solutions were prepared as instructed by the manufacturers of the DIG non-radioactive nucleic acid detection system (Roche)

A.3.12 *E. coli* transformation solutions

A.3.12.1 Transformation buffer 1 (TFB1)

Ingredient	Final concentration	Per 100ml
$RbCl_2$	100mM	5ml of 1M stock
$MnCl_2.4H_2O$	50mM	0.99g
KAc	30mM	0.294g
$CaCl_2$	10mM	1.34ml of 750mM stock
Glycerol	15%	30ml of 50% stock

Adjust pH to 5.8 with glacial acetic acid, make volume up to 100ml with water, filter-sterilise.

A.3.12.2 Transformation buffer 2 (TFB2)

Ingredient	Final concentration	Per 100ml
$RbCl_2$	10mM	1.0ml of 1M stock
$CaCl_2.2H_2O$	75mM	10ml of 750mM stock
Glycerol	15%	30ml of 50% stock
MOPS pH 7.0	10mM	10ml of 100mM stock

Make up to 100ml with water, filter-sterilize.

A.3.13 *B. subtilis* transformation solutions

A.3.13.1 2x Sucrose-Maleate-Magnesium chloride (SMM) buffer

Ingredient	Final concentration	Per 100ml
Sucrose	0.5M	34.2g
Maleic Acid	0.02M	0.4g
$MgCl_2$	0.02M	0.8g

Adjust to pH 6.5 with NaOH, autoclave.

A.3.13.3 SMMP

Ingredient	Final concentration	Per 100ml
SMM	1x	50ml of 2x stock
Penassay Broth (PAB)	2x	50ml of 4x stock

Make up fresh in a sterile container.

A.3.13.4 40% PEG 6000

Ingredient	Final concentration	Per 100ml
PEG 6000	40%	40g

Make up in 1x SMM, autoclave.

A.3.14 *C. beijerinckii* electroporation solutions

A.1.14.1 Electroporation buffer

Ingredient	Per 50ml
Sucrose	4.62g
$MgCl_2$	500µl of 0.1M stock
Phosphate Buffer	1.75ml of stock*

* Phosphate buffer

NaH_2PO_4	0.95ml of 0.2M stock
Na_2HPO_4	4.05ml of 0.2M stock

Make up to 10ml with water, autoclave.

Make up to 48.25ml with water, add phosphate buffer after autoclaving.

A.3.15 γ-glutamyl transferase assay solutions

A.3.15.1 SB buffer

Ingredient	Final concentration	Per 200ml
Imidazole pH 8.0	20mM	0.2723g
$MgCl_2.6H_2O$	10mM	0.4066g
3-Mercaptoethanol	2mM	28µl

Add 3-mercaptoethanol after autoclaving.

A.3.15.2 Reaction mixture

Ingredient	Per litre
Imidazole pH 7.15	1.5ml of 1M stock
Hydroxylamine	246µl of 0.8M stock
$MnCl_2$	30µl of 0.1M stock
K-arsenate pH 7.15	1ml of 0.28M stock
ADP (disodium salt)	100µl of 0.04M stock

Make up reaction rr ix fresh. Add components to 5.22ml of sterile water.

Filter-sterilize ADP and store 1ml aliquots at -20^{0}C.

A.3.15.3 Blank mixture

Ingredient	Final concentration	Per 8.1ml
Imidazole pH 7.15	1.6%	1.5ml of 1M stock
Hydroxylamine	1.0%	246µl of 0.8M stock
$MnCl_2$	0.5%	30µl of 0.1M stock

Make up fresh. Add components to 6.32ml of sterile water.

A.3.15.4 Stop buffer

Ingredient	Final concentration	100ml
$FeCl_3.6H_2O$	5.5%	5.5g
Trichloroacetic Acid	2%	2.0g
HCl		21ml of 36% solution

A.3.16 p-Nitrophenyl cellobiosidase assay solutions

A.3.16.1 Phosphate-citrate (PC) buffer

Ingredient	Final concentration	Per 100ml
K_2HPO_4	50mM	0.87g
Citric acid	12.5mM	0.26g

Adjust pH to 6.0 with concentrated phosphoric acid, make volume up to 100ml with water, autoclave.

A.3.17 Antibiotic stocks

A.3.17.1 Ampicillin

100mg/ml

Make up in distilled water, autoclave, store at -20^{0}C.

A.3.17.2 Erythromycin

10mg/ml

Make up in 100% ethanol, store at -20^{0}C.

A.3.17.3 Kanamycin

50mg/ml

Make up in water, filter-sterilize, store at -20^{0}C.

A.3.18 General Solutions and Buffers

A.3.18.1 (EDTA)

Ingredient	Final concentration	Per 500ml
EDTA	0.5M	93.05g
NaOH	Approximately 2%	Approximately 10g

Add EDTA to 400ml water. Slowly add NaOH until pH reaches 8.0 and EDTA dissolves. Make volume up to 500 ml with water, autoclave.

A.3.18.2 Ethidium bromide

Ingredient	Final concentration	Per 10ml
Ethidium bromide	10mg/ml	0.1g

Store in a light-tight container.

A.3.18.3 Phenol (equilibrated to pH 7.6)

Ingredient	Per 500ml
Phenol	500g
8-hydroxyquinoline	0.6g
NaOH	7.5ml of 2N stock
Tris-HCl pH 7.6	6ml of 1M stock

Add all of the above (except phenol) to a bottle of phenol with 130ml of water and store overnight at room temperature.

A.3.18.4 Tris-EDTA (TE) buffer

Ingredient	Final concentration	Per 100ml
Tris-HCl pH 7.6	10mM	1ml of 1M stock
EDTA pH 8.0	1mM	20µl of 0.5M stock

A.3.18.5 Tris-HCl

Ingredient	Final concentration	Per 100ml
Tris Base	1M	12.1g

Dissolve Tris in water, adjust pH with HCl, make volume up to 100ml with water, autoclave.

A.3.19 Primer extension solutions

A.3.19.1 HP buffer

Ingredient	Final concentration	Per 1ml
PIPES, pH 6,4	40mM	100µl of 0.4M stock
EDTA pH8.0	1mM	4µl of 250mM stock
NaCl	400mM	80µl of 5M stock
Formamide (deionised)	80%	800µl of stock

Use RNase-free ingredients, make up to 1ml with water.

A.3.19.2 RTB (Reverse Transcriptase Buffer)

Ingredient	Final concentration	Per 20µl
Reverse transciptase buffer (Promega)	1x	2µl of 10x stock
dNTP's	10mM	1 µl of stock
RNazin (Roche)	40U	4µl of stock
Actinomycin D	10µg/µl	2µl of 500mg/ml stock

Use RNase-free ingredients, make up to 20µl with water.

APPENDIX B

COMPLETE NUCLEOTIDE AND DEDUCED AMINO ACID SEQUENCES OF CLONES

B.1 PARTIAL NUCLEOTIDE AND AMINO ACID SEQUENCES OF THE INSERT IN pC16

```
1      ATCTTCCTTTGTGTATTTTGCCATTTTACAACTCCCCCTTTACATTTTTA    50

51     AAAAAAGCGCACAATAGACAAGTATAAACTTCTATCGAACGCCTTTGCTC    100

101    AATTCATATAATAACTAAATTATACTTCCCTTTTTTTTtAAGAAATCAAG    150

151    TGATTTTTTTAGTAAATAAATATTTTCGTACTAAAAACTTTACAATATAA    200

201    ATAATTTGATACGTTTTTCTTTAGATTCTAAAGTTATCAGGATTATTTTT    250

251    CCATTTTAAAAATTTATAGTCTCATTTTTTATTACAATATATATCCTTTA    300

301    TTATATAAATTTTATTTAACATTATACTATGCTCCATTCTGTAATTTTGT    350

351    ATATnATATTGTAATGAAAAAATTTTATAAATAATGCATAAATTTACAAA    400

401    AAAAACGAAgTCACGAGACGGCGTACCTCTATGACTTCGTTGTCAGCATA    450

451    TATTTATCTGTACTTTATTATAAAAGGTATGCTAAATACTGTCAACTATC    500

501    ATATCTCTTAGTTCAATTTTAAGATATGCACAAAAAAAAATTTTTATTCA    550

551    CAATATAATTCTACAATTAAATACTCTTTTATTATTAATAATTATCTTGA    600

601    ACATTTTTAGATAAATTGCTACAATCTTATATATAAAATTTATAATATAT    650

                           M  N  T  I  I  T  K  F  G  G  S  S
651    TGGGGGGCTATACTTATGAaCACAATTATAACAAAATTTGGTGGCAGCTC    700

        L  A  D  A  N  Q  F  R  K  V  K  D  I  I  Y  S  N
701    ATTAGCTGATGCCAATCAATTCAGAAAGGTTAAAGACATAATTTATTCTA    750

         D  A  R  K  Y  V  I  P  S  A  P  G  K  R  D  S
751    ATGATGCAAGAaAATATGTAATACCTTCTGCGCCTGGAAAAAGAGACTCT    800
```

```
        K  D  S  K  V  T  D  L  L  Y  L  C  H  A  H  V  A
801     AAAGATTCAAAAGTAACTGACTTATTATACCTTTGTCATGCTCATGTTGC     850

         A  G  I  A  L  D  D  V  F  N  H  I  R  Q  R  Y  S
851     TGCTGGAATCGCTTTAGATGATGTCTTTAATCACATAAGACAAAGATATT     900

          D  I  I  N  D  L  K  L  D  F  S  I  E  D  Q  L
901     CTGATATAATAAATGATTTAAAATTAGATTTTAGTATAGAAGATCAATTA     950

        N  T  I  K  K  D  L  E  A  G  A  S  S  D  Y  A  A
951     AATACAATAAAGAaAGACCTTGAAGCTGGTGCATCAAGTGATTACGCTGC    1000

         S  R  G  E  Y  L  N  G  L  I  L  A  K  Y  L  D  F
1001    AAGTAGAGGCGAATATTTAAATGGTTTAATTCTTGCGAAATATTTAGACT    1050

          E  F  V  D  A  K  D  V  I  V  F  K  K  D  G  S
1051    TTGAATTTGTTGATGCTAAAGATGTTATAGTTTTAAAAAGGATGGCTCA    1100

        L  D  N  E  A  T  N  C  A  L  H  N  R  L  S  N  V
1101    TTAGATAATGAAGCTACAAACTGCGCTCTACACAACAGATTATCTAATGT    1150

         S  K  A  V  I  P  G  F  Y  G  A  D  K  S  G  N  I
1151    TTCTAAGGCTGTTATTCCTGGATTTTATGGTGCTGATAAATCTGGTAACA    1200

          V  T  F  S  R  G  G  S  D  V  T  G  A  L  V  A
1201    TTGTTACATTCTCAAGAGGTGGTTCTGATGTTACTGGAGCATTAGTTGCT    1250

        A  S  I  N  A  N  L  Y  E  N  W  T  D  V  S  G  F
1251    GCAAGCATCAATGCAAATCTTTATGAAAATTGGACCGATGTTTCTGGTTT    1300

         L  M  A
1301    CTTAATGGCA
```

B.2 NUCLEOTIDE AND AMINO ACID SEQUENCES OF THE *Eco*RI INSERT IN pCBN1

```
1       GAATTCGATTTATTAGATGAACATATTTACAAGGTTGCTAAAACCTTGAG     50

51      TTAATGTCTTCACTAGAAATCATAAATATATTTGTAACCACAAGGGTCAT     100

101     AATATGGAGAAAACATTTGAAAATGAGCTGTTAAAGATTCTTAAGCAGAG     150

151     AACCTAAATCTTTGATTTAGTGAGAATCGCTTACTCAGCGAACGAACGTG     200

201     AGTCGAGTTTCTGCTAATGGTTGTTCCATTTTAGCTTGTGAAGTGAGAGT     250

251     CCAAAATTTTATTTTGGGTTCTCGAACTTCCTCAGCAAATGTAATGAGTC     300

301     GAGCTTCCCTAGGAAAGTTGGAGCATGCTAAAGTGGAACAACCTGCTGCT     350

351     TAGAACCTTCAGCGATATTTTCATAGTTTTTCGGAATCAAAATATTTATG     400

401     ATTTCGGGAAGTTACCACATAAATTTAGTTTTAAAGTTGGATATCTATAT     450

451     GTATTACAAATTTATATATACTTTGACAACGGAGTCATAGAATTAATAAT     500

501     TCAGATTGATTTCGTTTTTTTGCTATAAGTTATAAAATAATTTTCTGCAG     550

551     AAGTTATTTTTTATATTTATCAATTTGATAAATATAAGCTTTAATTTTTA     600

601     AAGAATATGGTATTAATTTCTTTAAAAATTAAAGTTTTTACCACAAAAAT     650

651     ACTTACTTTTTTAAAAAATCTATTGATTTGTTAGAAACGGTGGTATATAA     700

701     TTTACTTATTAAATGAATTGAGCAAAGGCGTTCAATAGATGGTATACCGC     750

                                                  M  A  K
751     TTCTATTGTGCGTCTTTTTTTAGAAAAAAAGGGGGTTTGTAAAATGGCAA     800

          Y  T  K  G  D  I  I  N  L  V  K  E  N  G  V  K
801     AGTACACAAAGGGAGATATCATTAACTTAGTAAAGGAAAATGGAGTTAAA     850

        F  I  R  L  Q  F  T  D  I  F  G  T  L  K  N  V  A
851     TTTATAAGACTTCAATTTACAGATATCTTTGGAACATTAAAAAACGTAGC     900

         I  T  D  R  Q  L  E  K  A  L  N  N  E  C  M  F  D
901     AATTACAGACAGACAGTTAGAAAAAGCATTAAATAATGAATGTATGTTTG     950

          G  S  S  I  D  G  F  V  R  I  E  E  S  D  M  N
951     ATGGATCATCTATTGATGGTTTTGTTAGAATTGAAGAATCTGACATGAAT     1000
```

```
        L  R  P  N  L  D  S  F  V  I  F  P  W  R  P  Q  Q
1001    TTAAGACCGAACTTAGATAGTTTTGTAATTTTCCCATGGAGACCGCAACA    1050

         G  K  V  A  R  L  I  C  D  V  Y  K  P  D  G  K  P
1051    AGGTAAGGTTGCAAGATTAATCTGTGACGTTTATAAACCAGATGGAAAGC    1100

          F  E  D  G  P  R  Y  I  L  K  R  A  I  A  D  A
1101    CATTTGAAGGTGATCCAAGATATATCTTAAAGAGAGCTATAGCAGATGCA    1150

        A  E  L  G  Y  T  M  N  V  G  P  E  C  E  F  F  L
1151    GCTGAACTTGGATATACAATGAATGTTGGACCTGAATGTGAATTTTTCTT    1200

         F  E  T  D  E  N  G  N  A  T  T  N  T  Q  D  K  G
1201    GTTTGAAACTGATGAAAATGGTAATGCAACAACAAACACTCAAGATAAAG    1250

          G  Y  F  D  L  A  P  T  D  L  G  E  N  A  R  R
1251    GTGGGTACTTTGATTTAGCACCTACAGATTTGGGAGAAAATGCAAGACGT    1300

        D  M  T  L  A  L  E  E  M  G  F  E  I  E  A  S  H
1301    GACATGACTTTAGCATTAGAAGAAATGGGATTTGAGATTGAAGCATCTCA    1350

         H  E  V  A  E  G  Q  N  E  I  D  F  K  Y  G  D  A
1351    CCATGAAGTTGCTGAAGGACAAAATGAAATTGACTTTAAATATGGTGATG    1400

          L  T  T  A  D  S  I  M  T  F  K  L  V  V  K  S
1401    CATTAACTACAGCTGATAGTATTATGACATTCAAGCTAGTTGTTAAGTCT    1450

        I  A  Q  R  G  H  L  H  A  S  F  M  P  K  P  I  F
1451    ATTGCACAAAGACACGGATTACATGCATCATTTATGCCAAAGCCAATTTT    1500

         G  I  N  G  S  G  M  H  V  N  M  S  L  F  K  D  G
1501    CGGAATTAATGGTTCAGGAATGCATGTTAATATGTCATTATTCAAGGATG    1550

          K  N  V  F  V  D  E  N  D  K  N  G  L  S  P  I
1551    GAAAGAATGTCTTTGTTGATGAAAATGATAAGAATGGCTTAAGTCCAATA    1600

        A  Y  N  F  I  A  G  L  L  K  N  I  K  G  L  A  A
1601    GCTTATAACTTTATAGCAGGATTACTAAAGAATATCAAAGGTCTTGCGGC    1650

         V  T  N  P  L  V  N  S  Y  K  R  L  V  P  G  Y  E
1651    TGTTACTAACCCGTTAGTTAATTCATACAAGAGATTAGTACCAGGATATG    1700

          A  P  V  Y  L  A  W  S  C  K  N  R  T  A  L  I
1701    AAGCACCAGTTTATTTAGCTTGGTCTTGTAAAAACAGAACAGCATTAATA    1750

        R  V  P  A  A  R  G  A  G  T  R  V  E  L  R  C  P
1751    AGAGTACCAGCAGCTAGAGGCGCAGGAACTAGAGTTGAGCTTAGATGTCC    1800

         D  P  S  S  N  P  Y  L  V  L  A  A  L  L  Q  A  G
1801    AGATCCAAGTTCTAACCCATATTTAGTATTAGCAGCTCTATTACAAGCAG    1850
```

L D G I K N N L Q P P A E V E A

1851 GACTAGATGGTATTAAAAACAATCTACAACCACCAGCAGAAGTTGAAGCA 1900

N I F A M T D E E R K A N G I D N

1901 AATATTTTTGCTATGACTGATGAGGAAAGAAAAGCTAACGGAATTGATAA 1950

L P N N L Y E A V Q F M R E S E L

1951 CTTACCTAATAACTTATATGAAGCAGTTCAATTCATGAGAGAAAGTGAGT 2000

A K S A L G D H V Y N N Y L E G

2001 TAGCTAAGTCAGCTTTAGGAGATCATGTATACAATAATTATCTTGAAGGT 2050

K A A E W D D Y R T K V H D W E L

2051 AAAGCAGCTGAGTGGGATGATTATAGAACTAAAGTACATGACTGGGAATT 2100

E N Y L N R Y *

2101 AGAAAACTATCTTAATAGATATTAATTTAGTAATACCAATTATTATTAAT 2150

M A Q D G R I I I A L S

2151 TAAGGTGGGGATTGCAATGGCTCAAGATGGAAGAATTATTATAGCTTTAA 2200

N I E T A K K L K S L L M Q E G

2201 GCAATATTGAGACAGCAAAAAAATTAAAGAGTTTACTAATGCAGGAAGGC 2250

Y E I I A L C A S G N E L I R L V

2251 TATGAAATCATAGCATTATGTGCTTCAGGAAATGAATTAATTAGATTAGT 2300

M Q H S P D L V L V G Y K F K D M

2301 TATGCAACATTCCCCAGACCTAGTCTTAGTAGGATATAAATTCAAGGATA 2350

S L L D V Y E T L V D V T S F L

2351 TGAGTTTACTAGATGTTTATGAAACACTAGTAGATGTAACTAGTTTTTTA 2400

A I V N E P Y K S F I E E D T D I

2410 GCTATCGTAAATGAGCCTTATAAATCATTTATAGAAGAAGATACTGATAT 2450

Y C I G T K I S N V L L T N A I N

2451 ATATTGTATTGGTACCAAAATCTCTAATGTTCTCTTAACTAATGCTATAA 2500

L I F Q S K K R I K K L R D Q V

2501 ATTTAATTTTTCAAAGCAAAAAGAGAATCAAGAAGCTAAGGGATCAAGTT 2550

E K L E H T L E D R K L I E K A K

2551 GAAAAGTTAGAACATACTCTAGAAGATAGAAAGCTTATCGAAAAGGCTAA 2600

G Q L M K T S G L T E N E A F R Y

2601 GGGTCAACTCATGAAAACTTCAGGACTTACGGAGAATGAAGCGTTTAGAT 2650

M Q K I S M D S G K R M K D I A

2651 ATATGCAGAAGATAAGTATGGATTCTGGAAAAGAATGAAAGATATTGCG 2700

```
        S  L  I  L  S  E  N  E  *
2701    AGTTTAATATTAAGTGAAAATGAATAACTTGTAAAATTTAGCAAATACGC    2750

2751    GAGTGATTATATTTCAAAAAAGGATGTATTTACTCGTTGTTTGCTACACA    2800

2801    TAAATGCAATATGATTGTTTTAAGGATAAATAGGTGCCTGATAGATAAAT    2850

2851    ATTTAGGAAGGTACAAATGAATATATTATTATCTCAAGGGGGGAATTTTA    2900

        M  E  R  N  N  P  N  A  Q  G  L  Y  N  P  C  F  E
2901    ATGGAAAGAAATAATCCGAATGCACAGGGGTTATATAATCCATGTTTTGA    2950

         H  D  A  C  G  I  G  T  I  V  N  I  D  G  E  K  S
2951    ACATGATGCTTGTGGAATTGGGACTATAGTAAACATTGATGGTGAGAAAT    3000

          H  E  I  L  S  D  C  L  T  I  L  E  K  L  E  H
3001    CTCATGAAATATTATCAGATTGTTTAACTATCCTAGAAAAGCTAGAGCAT    3050

        R  G  G  T  G  A  D  E  H  T  G  D  G  A  G  I  L
3051    AGAGGTGGTACAGGCGCTGATGAACATACAGGAGATGGAGCAGGAATATT    3100

         F  N  I  P  H  K  F  F  Q  E  E  L  K  S  K  G  M
3101    ATTTAATATACCACACAAATTCTTTCAGGAAGAATTGAAATCAAAAGGTA    3150

          T  L  G  N  E  G  D  Y  A  V  A  M  V  F  L  P
3151    TGACTCTCGGAAACGAAGGGGATTACGCTGTTGCTATGGTATTTTTACCA    3200

        Q  E  E  K  A  R  K  E  A  M  S  L  F  E  D  I  S
3201    CAAGAAGAAAAGGCAAGAAAAGAAGCTATGAGCCTTTTTGAAGATATATC    3250

         K  E  E  G  L  E  L  I  G  W  R  E  V  Q  T  K  P
3251    AAAGGAGGAAGGACTTGAACTAATTGGGTGGAGAGAAGTTCAAACAAAGC    3300

          S  I  L  G  K  A  S  L  E  A  M  P  A  I  M  Q
3301    CTTCAATACTTGGAAAAGCATCGCTAGAAGCAATGCCTGCAATAATGCAA    3350

        A  F  V  R  R  P  N  G  I  K  P  G  K  D  F  E  R
3351    GCTTTTGTAAGAAGACCAAATGGAATAAAACCAGGAAAAGATTTCGAAAG    3400

         N  L  Y  I  V  R  R  I  I  E  K  R  A  G  W  I  S
3401    AAATTTATATATTGTTAGAAGAATTATAGAAAAAAGAGCAGGATGGATAA    3450

          K  F  L  N  E  T  F  Y  I  A  S  F  S  S  K  T
3451    GTAAATTCCTAAATGAAACTTTCTACATAGCGTCTTTTTCATCAAAGACA    3500

        I  V  Y  K  G  M  L  L  S  T  Q  L  R  E  F  Y  K
3501    ATAGTATATAAAGGGATGTTACTATCTACACAGCTTAGAGAATTTTATAA    3550

         D  L  E  D  E  R  V  E  T  S  L  A  L  V  H  S  R
3551    AGATTTAGAAGATGAAAGAGTTGAAACATCTTTGGCATTAGTACATTCAA    3600
```

```
          Y  S  T  N  F  T  P  S  W  E  R  A  H  P  N  R
3601   GATATAGTACTAATACATTCCCAAGTTGGGAAAGAGCTCATCCTAATAGA   3650

       F  M  I  H  N  G  E  I  N  T  L  R  G  N  V  N  K
3651   TTTATGATTCACAATGGTGAAATTAATACACTTCGTGGAAATGTTAATAA   3700

        V  Y  S  R  E  T  N  V  K  S  R  A  L  G  K  D  L
3701   GGTTTATTCTAGAGAAACAAATGTTAAATCTAGAGCACTTGGAAAAGATT   3750

         N  R  V  L  P  I  I  N  K  E  G  S  D  S  A  I
3751   TAAACAGAGTGTTACCTATCATAAATAAAGAAGGATCAGACTCTGCGATA   3800

       F  D  N  N  L  E  F
3801   TTTGATAACAATTTAGAATTC   3821
```

REFERENCES

Abratt, V.R., H. Zappe & D.R. Woods. 1993. A reporter gene vector to investigate the regulation of glutamine synthetase in *Bacteroides fragilis* Bf1. *J. Gen. Microbiol.* **139**:59-65

Allcock, E.R. & D.R. Woods. 1981. Carboxymethyl cellulase and cellobiase production by *Clostridium acetobutylicum* in an industrial fermentation medium. *Appl. Environ. Microbiol.* **41**:539-541

Amster-Choder, O., F. Houman & A. Wright. 1989. Protein phosphorylation regulates transcription of β-glucoside utilisation operon in *E. coli*. Cell. **58**:847-855

Amster-Choder, O. & A. Wright. 1990. Regulation of activity of a transcriptional antiterminator in *E. coli* by phosphorylation *in vivo*. Science. **249**:540-542

Amster-Choder, O. & A. Wright. 1992. Modulation of the dimerization of a transcriptional antiterminator protein by phosphorylation. Science. **257**:1395-1398

Altschul, S. F., T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller & D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389-3402.

Antson, A.A., A.M. Brzozowski, E.J. Dodson, Z. Dautier, K.S. Wilson, T. Kurecki, J. Otridge & P. Gollnick. 1994. 11-fold symmetry of *trp* RNA-binding attenuation protein (TRAP) from *Bacillus subtilis* determined by X-ray analysis. J. Mol. Biol. **244**:1-5

Arnaud, M., P. Vary, M. Zagorec, M. Klier., M. Debarbouille, P. Postma & 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

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

Atkinson, M.R. & S.H. Fisher. 1991. Identification of genes and gene products whose expression is activated during nitrogen-limited growth in *Bacillus subtilis*. J. Bacteriol. **173**:23-27

Aymerich, S. & 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 USA. **89**:10410-10414

Babitzke, P. & C. Yanofsky. 1993. Reconstitution of the *Bacillus subtilis trp* attenuation *in vitro* with TRAP, the *trp* RNA-binding attenuation protein. Proc. Natl. Acad. Sci. USA. **90**:133-137

Babitzke, P., J.T. Stultz, S.J. Shire & C. Yanofsky. 1994. TRAP, the *trp* RNA-binding attenuation protein of *Bacillus subtilis*, is a multisubunit complex that appears to recognise G/UAG repeats in the *trpEDCFBA* and *trpG* transcripts. J. Biol. Chem. **269**:16597-16604

Babitzke, P., D.G. Bear & C. Yanofsky. 1995. TRAP, the *trp* RNA-binding attenuation protein of *Bacillus subtilis*, is a toroid-shaped molecule that binds transcripts containing GAG or UAG repeats separated by two nucleotides. Proc. Natl. Acad. Sci. USA. **92**:7916-7920

Babitzke, P. 1997. Regulation of tryptophan biosynthesis: Trp-ing the TRAP or how *Bacillus subtilis* reinvented the wheel. Mol. Microbiol. **26**:1-9

Barik, S., B. Ghosh, W.A. Whalen, D. Lazinski & A. Das. 1987. An antitermination protein engages the elongation transcription apparatus at a promoter-proximal recognition site. Cell. **50**:885-899

Baumann, C., J. Otridge & C. Yanofsky. 1996. Kinetic and thermodynamic analysis of the interaction between TRAP (*try* RNA-binding attenuation protein) of *Bacillus subtilis* and *trp* leader RNA. J. Biol. Chem. **271**:12269-12274

Bennett, G.N. & D.J. Petersen. 1993. Cloning and expression of *Clostridium acetobutylicum* genes involved in solvent production. In: Genetics and molecular biology of anaerobic bacteria, M. Sebald, ed., pp 317-343, Springer-Verlag, New York.

Bohannon, D.E. & A.L. Sonenshein. 1989. Positive regulation of glutamate biosynthesis in *Bacillus subtilis*. J. Bacteriol. **178**:4718-4727

Brammar, W.J. & P.H. Clarke. 1964. Induction and repression of *Pseudomonas aeruginosa* amidase. J. Gen. Microbiol. **37**:307-319

Brown, S.W. & A.L. Sonenshein. 1996. Autogenous regulation of the *Bacillus subtilis glnRA* operon. J. Bacteriol. **178**:2450-2454

Bullifent, H.L., A. Moir & R.W. Titball. 1995. The construction of a reporter system and use for the investigation of *Clostridium perfringens* gene expression. FEMS Lett. **131**:99-105

Burns, C.M. & J.P. Richardson. 1995. NusG is required to overcome a kinetic limitation to Rho function at an intragenic terminator. Proc. Natl. Acad. Sci. USA. **92**:4738-4742

Cali, B.M., J.L. Micca & V. Stewart. 1989. Genetic regulation of nitrate assimilation in *Klebsiella pneumoniae* M5a1. J. Bacteriol. **171**:2666-2672

Chai, W & V. Stewart. 1998. NasR, a novel RNA-binding protein, mediates nitrate-responsive transcription antitermination of the *Klebsiella oxytoca* M5a1 *nasF* operon leader in vitro. J. Mol. Biol. **283**:339-351

Chamberlain, D., B.P. O'Hara, S.A. Wilson, L.H. Pearl & S.J. Perkins. 1997. Oligomerization of the amide sensor protein AmiC by X-ray and neutron scattering and molecular modelling. Biochem. **36**:8020-8029

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

Chang, S. & S.N. Cohen. 1979. High-frequency transformation of *Bacillus subtilis* protoplasts by plasmid DNA. Mol. Gen. Genet. **168**:111-115

Chattopadhyay, S. J. Garcia-Mena, J. DeVito, K. Wolska & A. Das. 1995. Bipartite function of a small RNA hairpin in transcription antitermination in bacteriophage lambda. Proc. Natl. Acad. Sci. USA. **92**:4061-4065

Chen, Y.M., K. Backman & B Magasanik. 1982. Characterization of a gene, *glnL*, the product of which is involved in the regulation of nitrogen utilisation in *Escherichia coli*. J. Bacteriol. **150**:214-220

Chiurazzi, M., R. Meza, M. Lara, A. Lahm, R. Defez, M. Iaccarino & G. Espin. 1992. The *Rhizobium leguminosarum* biovar *phaseoli glnT* gene, encoding a glutamine synthetase III. Gene. **119**:1-8

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

Das, A. 1993. Control of transcription termination by RNA-binding proteins. Annu. Rev. Biochem. **62**:893-930

Dean, D.R., J.A. Hoch & A.I. Aronson. 1977. Alteration of *the Bacillus subtilis* glutamine synthetase results in overproduction of the enzyme. J. Bacteriol. **131**:981-987

Deshpande, M.V., K-E. Eriksson & L.G. Petterson. 1984. An assay for the selective determination of exo-1,4-glucanases in a mixture of cellulolytic enzymes. Ann. Biochem. **138**:481-487

DeVito, J. & A. Das. 1994. Control of transcription processivity in phage lambda: Nus factors strengthen the termination-resistant state of the RNA polymerase induced by N antiterminator. Proc. Natl. Acad. Sci. USA. **91**:8660-8664

Doi, R.H., M. Goldstein, S. Hashida, J-S. Park & M. Takagi. 1994. The *Clostridium cellulovorans* cellulosome. Crit. Rev. Microbiol. **20**:87-93

Drew, R. & N. Lowe. 1989. Positive control of *Pseudomonas aeruginosa* amidase synthesis is mediated by a transcription anti-termination mechanism. J. Gen. Microbiol. **135**:817-823

Du, H., R. Tarpey & P. Babitzke. 1997. The *trp* RNA-binding attenuation protein regulates TrpG synthesis by binding to the *trpG* ribosome binding site of *Bacillus subtilis*. J. Bacteriol. **179**:2582-2586

Farin, F, & P.H. Clarke. 1978. Positive regulation of amidase synthesis in *Pseudomonas aeruginosa*. J. Bacteriol. **135**:379-392

Felix, C.R. & L.G. Ljungdahl. 1993. The cellulosome: The exocellular organelle of *Clostridium*. Annu. Rev. Microbiol. **47**:791-819

Ferson, A.E., L.V. Wray, Jr. & S.H. Fisher. 1996. Expression of the *Bacillus subtilis gabP* gene is regulated independently in response to nitrogen and amino acid availability. Mol. Microbiol. **22**:693-701

Fierro-Monti, I.P., S.J. Reid & D.R. Woods. 1992. Differential expression of a *Clostridium acetobutylicum* antisense RNA; Implications for regulation of glutamine synthetase. J. Bacteriol. **174**:7642-7647

Fisher, S.H. 1999. Regulation of nitrogen metabolism in *Bacillus subtilis*: vive la difference! Mol. Microbiol. **32**:223-232

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

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

Friedman, D.I. & D.L. Court. 1995. Transcription and antitermination: the λ paradigm updated. Mol. Microbiol. **18**:191-200

Garcia-Dominguez, M., J.C. Reyes & F.J. Florencio. 1999. Glutamine synthetase inactivation by protein-protein interaction. Proc. Natl. Acad. Sci. USA. **96**:7161-7166

Garrity, D.B. & S.A. Zahler. 1994. Mutations in the gene for a tRNA that functions as a regulator of a transcriptional attenuator in *Bacillus subtilis*. Genetics. **137**:627-636

Goldman, B.S., J.T. Lin & V. Stewart. 1994. Identification and structure of the *nasR* gene encoding a nitrate- and nitrite-responsive positive regulator of *nasFEDCBA* (nitrate assimilation) operon expression in *Klebsiella pneumoniae* M5al

Goliger, J.A. & J.W. Roberts. 1987. Bacteriophage 82 gene Q and Q protein. Sequence, overproduction, and activity as a transcription antiterminator in vitro. J. Biol. Chem. **262**:11721-11725

Gollnick, P. 1994. Regulation of the *Bacillus subtilis trp* operon by an RNA-binding protein. Mol. Microbiol. **11**:991-997

Gollnick, P., S. Ishino, M.I. Kuroda, D.J. Henner & C. Yanofsky. 1990. The *mtr* locus is a two-gene operon required for transcription attenuation in the *trp* operon of *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA. **87**:8726-8730

Goodman, H.J.K. & D.R. Woods. 1993. Cloning and nucleotide sequence of the *Butyrivibrio fibrisolvens* gene encoding a type III glutamine synthetase. J. Gen. Microbiol. **139**:1487-1493

Grandoni, J.A., S.B. Fulmer, V. Brizzio, S.A. Zahler & J.M. Calvo. 1993. Regions of the *Bacillus subtilis ilv-leu* operon involved in regulation by leucine. J. Bacteriol. **175**:7581-7593

Greenblatt, J., J.R. Nodwell & S.W. Mason. 1993. Transcriptional antitermination. Nature. **364**:401-406

Grundy, F.J. & T.M. Henkin. 1993. tRNA as a positive regulator of transcription antitermination in *Bacillus subtilis*. Cell. **74**:475-482

Grundy, F.J., S.M. Rollins & T.M. Henkin. 1994. Interaction between the acceptor end of tRNA and the T box stimulates antitermination in the *Bacillus subtilis tyrS* gene-a new role for the discriminator base. *J. Bacteriol.* **176**:4518-4526

Gutowski, J.C. & H.J. Schreier. 1992. Interaction of the *Bacillus subtilis glnRA* repressor with operator and promoter sequences *in vivo*. J. Bacteriol. **174**:671-681

Guttierez, J-C., F. Ramos, L. Ortner & M. Tortolero. 1995. *nasST*, two genes involved in the induction of the assimilatory nitrite-nitrate reductase operon (*nasAB*) of *Azotobacter vinelandii*. Mol. Microbiol. **18**:579-591

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

Henkin, T.M., B.L. Glass & F.J. Grundy. 1992. Analysis of the *Bacillus subtilis tyrS* gene: conservation of a regulatory sequence in multiple tRNA synthetase genes. J. Bacteriol. **174**:1299-1306

Henkin, T.M. 1994. tRNA-directed transcription antitermination. Mol. Microbiol. **13**:381-387

Henkin, T.M. 1996. Control of transcription termination in prokaryotes. Ann. Rev. Genet. **30**:35-57

Herskowitz, I & E.R. Signer. 1970. A site essential for expression of all late genes in bacteriophage lambda. J. Mol. Biol. **47**:545-556

Hill. R., J. Parker, H. Goodman, D. Jones & D. Woods. 1989. Molecular analysis of a novel glutamine synthetase of the anaerobe *Bacteroides fragilis*. J. Gen. Microbiol. **135**:3271-3279

Hoffman, R.J. & P. Gollnick. 1995. The *mtrB* of *Bacillus pumilis* encodes a gene with sequence and function homology to the *trp* RNA-binding attenuation protein (TRAP) of *Bacillus subtilis*. J. Bacteriol. **177**:839-842

Houman, F. M.R. Diaz-Torres & A. Wright. 1990. Transcriptional antitermination in the *bgl* operon of *E. coli* is modulated by a specific RNA-binding protein. Cell. **62**:1153-1163

Janssen, P.J., W.A. Jones, D.T. Jones & D.R. Woods. 1988. Molecular analysis and regulation of the *glnA* gene of the gram-positive anaerobe *Clostridium acetobutylicum*. J. Bacteriol. **170**:400-408

Janssen, P.J., D.T. Jones & D.R. Woods. 1990. Studies on *Clostridium acetobutylicum* promoters and antisense RNA. Mol. Microbiol. **4**:1575-1583

Jones, D.T., A. van der Westhuizen, S. Long, E.R. Allcock, S.J. Reid & D.R. Woods. 1982. Solvent production and morphological changes in *Clostridium acetobutylicum*. Appl. Environ. Microbiol. **43**:1434-1439

Jones, D.T. & D.R. Woods. 1986. Acetone-butanol fermentation revisited. Microbiol. Rev. **50**:383-524

Jones, D.T. & D.R. Woods. 1989. Solvent production. In: Biotechnology handbooks, Vol. 3: Clostridia, N. P. Minton & D. J. Clarke, eds., pp 105-144. Plenum, New York.

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

Klingel, U., C.M. Miller, A.K. North, P.G. Stockley & S. Baumberg. 1995. A binding site for activation by the *Bacillus subtilis* AhrC protein, a repressor/activator of arginine metabolism. Mol. Gen. Genet. **248**:329-340

Landick, R. & C.L. Turnbough, Jr. 1992. Transcriptional attenuation. In: Transcriptional regulation, pp 407-446, Cold Spring Harbour Laboratory Press, New York.

Lee, S.F., C.W. Forsberg & L.N. Gibbons. 1985. Cellulolytic activity of *Clostridium acetobutylicum*. Appl. Environ. Microbiol. **50**:220-228

Lee, S.Y., G.N. Bennett & E.T. Papoutsakis. 1992. Construction of *Escherichia coli-Clostridium acetobutylicum* shuttle vectors and transformation of *Clostridium acetobutylicum* strains. Biotechnol. Lett. **14**:427-432

Lerner, C.G., B.T. Stevenson & R.L. Switzer. 1987. Structure of the *Bacillus subtilis* pyrimidine biosynthetic *pyr* gene cluster. J. Bacteriol. **169**:2202-

Li, J., S.W. Mason & J. Greenblatt. 1993. Elongation factor NusG interacts with termination factor rho to regulate termination and antitermination of transcription. Genes. Dev. **7**:161-172.

Lin, J.T., B.S. Goldman & V. Stewart. 1993. Structures of genes *nasA* and *nasB*, encoding assimilatory nitrate and nitrite reductase in *Klebsiella pneumoniae* M5a1. J. Bacteriol. **175**:2370-2378

Lin, J.T., B.S. Goldman & V. Stewart. 1994. The *nasFEDBCA* operon for nitrate and nitrite assimilation in *Klebsiella pneumoniae* M5a1. J. Bacteriol. **176**:2551-2559

Lin, J.T. & V. Stewart. 1996. Nitrate and nitrite-mediated transcription antitermination control of *nasF* (nitrate assimilation) in *Klebsiella pneumoniae* M5a1. J. Mol. Biol. **256**:423-425

Lin, J.T. & V. Stewart. 1998. Nitrate assimilation by bacteria. Adv. Microbial Physiol. **39**:1-30

Long, S., D.T. Jones & D.R. Woods. 1984. Initiation of solvent production, clostridial stage and endospore formation in *Clostridium acetobutylicum* P262. Appl. Microbiol. Biotechnol. **20**:256-261

Lu, Y., R.J. Turner & R.L. Switzer. 1995. Roles of the three transcriptional attenuators of the *Bacillus subtilis* pyrimidine biosynthetic operon in the regulation of its expression. J. Bacteriol. **177**:1315-

Lu, Y., R.J. Turner & R.L. Switzer. 1996. Function of RNA secondary structures in transcriptional attenuation of the *Bacillus subtilis pyr* operon. Proc. Natl. Acad. Sci. USA. **93**:14462-14467.

Lu, Y. & R.L. Switzer. 1996. Transciptional attenuation of the *Bacillus subtilis pyr* operon by the PyrR regulatory protein and uridine nucleotides *in vitro*. J. Bacteriol. **178**:7206-

Lu, Y. & R.L. Switzer. 1996b. Evidence that the *Bacillus subtilis* pyrimidine regulatory protein PyrR acts by binding to *pyr* mRNA at three sites *in vivo*. J. Bacteriol. **178**:5806-

Mahadevan, S., A.E. Reynolds & A. Wright. 1988. Positive and negative regulation of the *bgl* operon in *Escherichia coli*. J. Bacteriol. **169**:2570-2578

Matsushita, C., O. Matsushita, M. Kuyama & A. Okabe. 1994. A *Clostridium perfringens* vector for the selection of promoters. Plasmid. **31**:317-319

Mason, S.W., J. Li & J. Greenblatt. 1992. Host factor requirements for processive antitermination of transcription and suppression of pausing by the N protein of bacteriophage lambda. J. Biol. Chem. **267**:19418-19426

Merino, E., P. Babitzke & C. Yanofsky. 1995. *trp* RNA-binding attenuation protein (TRAP)-*trp* leader RNA interactions mediate translational as well as transcriptional regulation of the *Bacillus subtilis trp* operon. J. Bacteriol. **177**:6362-6370

Mermelstein, L.D., N.E. Welker, G.N. Bennett & E.T. Papoutsakis. 1992. Expression of cloned homologous fermentative genes in *Clostridium acetobutylicum* ATCC 824. Bio/Technol. **10**:190-195

Merrick, M.J. & R.A. Edwards. 1995. Nitrogen control in bacteria. Microbiol. Rev. **59**:604-622

Minton, N.P., T-J. Swinfield, J.K. Brehm, S.M. Whelan & J.D. Oultram. 1993. Vectors for use in *Clostridium acetobutylicum*. In: Genetics and molecular biology of anaerobic bacteria, M. Sebald, ed., pp. 120-140. Springer-Verlag, New York.

Monod, M., C. Denoya & D.A. Dubnau. 1986. Sequence and properties of pIM13, a macrolide-lincosamide-streptogramin B resistance plasmid from *Bacillus subtilis*. J. Bacteriol. **67**:138-147

Nakano, M.M., F. Yang, P. Hardin & P. Zuber. 1995. Nitrogen regulation of *nasA* and the *nasB* operon, which encode genes required for nitrate assimilation in *Bacillus subtilis*. J. Bacteriol. **177**:573-579

Nakano, M. M., T. Hoffman, Y. Zhu & D. Jahn. 1998. Nitrogen and oxygen regulation of *Bacillus subtilis nasDEF*, encoding NADH-dependent nitrate reductase by TnrA and ResD. J. Bacteriol. **180**:5344-5350

Nakano, Y. & K. Kimura. 1991. Purification and characterisation of a repressor for the *Bacillus cereus glnRA* operon J. Biochem (Tokyo). **109**:223-228

Nodwell, J.R. & J. Greenblatt. 1993. Recognition of boxA antiterminator RNA by the *E. coli* antitermination factors NusB and ribosomal protein S10. Cell. **72**:261-268

Oultram, J.D., M. Loughlin, T.J. Swinfield, J.K. Brehm, D.E. Thompson & N.P. Minton. 1988. Introduction of plasmids into whole cells of *Clostridium acetobutylicum* by electroporation. FEMS. Microbiol. Lett. **58**:83-88

Patte, J-C. 1996. Biosynthesis of threonine and lysine. In: *Escherichia coli* and *Salmonella*. Cellular and molecular biology. F. Niedhart, ed., pp528-541. American Society for Microbiology, Washington, D.C.

Patterson, T.A, Z. Zhang, T. Baker, L.L. Johnson, D.I. Freidman & D.L. Court. 1994. Bacteriophage lambda N-dependent transcription antitermination. Competition for an RNA site may regulate antitermination. J. Mol. Biol. **236**:217-228

Paulus, H. 1993. Biosynthesis of the aspartate family of amino acids. In: *Bacillus subtilis* and other Gram-positive bacteria. A.L. Sonenshein, J.A. Hoch & R. Losick, eds., pp237-267. American Society for Microbiology, Washington, D.C.

Phillips-Jones, M.K. 1993. Bioluminescence (*lux*) expression in the anaerobe *Clostridium perfringens*. FEMS Lett. **106**:265-270

Potvin, B.W., R.J. Kelleher & H. Gooder. 1975. Pyrimidine biosynthetic pathway of *Bacillus subtilis*. J. Bacteriol. **123**:604-

Ramos, F., G. Blanco, J.C. Guttierez, F. Luque & M. Tortolero. 1993. Identification of an operon involved in the assimilatory nitrate-reducing system of *Azotobacter vinelandii*. Mol. Microbiol. **8**:1145-1153

Reid, S.J., Rafudeen, M.S & Leat, N.G. 1999. The genes controlling sucrose utilisation in *Clostridium beijerinckii* NCIMB 8052 constitute an operon. Microbiol. **145**:1461-1472

Reyes, J.C. & F.J. Florencio. 1995. A novel mechanism of glutamine synthetase inactivation by ammonium in the cyanobacterium *Synechocystis* sp. PCC 6803. Involvement of an inactivating protein. FEBS. Lett. **367**:45-48

Roberts, J. 1992. Antitermination and the control of transcription elongation. In: Transcriptional regulation, pp 389-406, Cold Spring Harbour Laboratory Press, New York.

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

Saha, B.D., R. Lamed & G.J. Zeikus. 1989. Clostridial enzymes. In: Clostridia, N. P. Minton & D. J. Clarke, eds., pp 227-263. Plenum Press, New York.

Saier, Jr, M.H., M.J. Fagan, C. Hoischen & J. Reizer. 1993. Transport mechanisms. In: *Bacillus subtilis* and other Gram-positive bacteria. A.L. Sonenshein, J.A. Hoch & R. Losick, eds., pp13-156. American Society for Microbiology, Washington, D.C.

Sambrook, J., E.F. Fritsch & T. Maniatis. 1989. Molecular cloning: A laboratory manual. 2e. Cold Spring Harbour Laboratory Press.

Schnetz, K. & B. Rak. 1988. Regulation of the *bgl* operon in *Escherichia coli* by transcriptional antitermination. EMBO. J. **7**:3271-3277

Schnetz, K. & B. Rak. 1990. β-glucoside permease represses the *bgl* operon of *Escherichia coli* by phosphorylation of the antiterminator protein and also interacts with glucose-specific enzyme III, the key element in catabolite control. Proc. Natl. Acad. Sci. USA. **87**:5074-5078

Schnetz, K., C. Toloczyki & B. Rak. 1987. β-glucoside (*bgl*) operon of *Escherichia coli* K12: nucleotide sequence, genetic organization and possible evolutionary relationship to regulatory components of two *Bacillus subtilis* genes. J. Bacteriol. **169**:2579-2590

Shapiro, B.M. & E.R. Stadtman. 1970. Glutamine synthetase (*E. coli*). Methods Enzymol. **17**:910-922

Shatters, R., Y. Liu & M. Kahn. 1993. Isolation and characterization of a novel glutamine synthetase from *Rhizobium meliloti*. J. Biol. Chem. **268**:469-475

Schreier, H.J. 1993. Biosynthesis of glutamine and glutamate and the assimilation of ammonia. In: *Bacillus subtilis* and other Gram-positive bacteria: Biochemistry, Physiology, and Molecular Biology, A. L. Sonenshein, J. A. Hoch & R. Losick, eds., pp281-298. American Society for Microbiology Press, Washington, DC.

Schreier, H.J. & A.L. Sonenshein. 1986. Altered regulation of the *glnA* gene in glutamine synthetase mutants of *Bacillus subtilis*. J. Bacteriol. **167**:35-43

Schreier, H.J., S.H. Fisher & A.L. Sonenshein. 1985. Regulation of expression from the *glnA* promoter of *Bacillus subtilis* requires the *glnA* gene product. Proc. Natl. Acad. Sci. USA. **82**:3375-3379

Schreier, H.J., S.W. Brown, K.D. Hirschi, J.F. Nomellini & A.L. Sonenshein. 1989. Regulation of *Bacillus subtilis* glutamine synthetase gene expression by the product of the *glnR* gene. J. Mol. Biol. **210**:51-63

Serror, P. & A.L. Sonenshein. 1996. CodY is required for nutritional repression of *Bacillus subtilis* genetic competence. J. Bacteriol. **178**:5910-5915

Slack, F.J., P. Serror, E. Joyce & A.L. Sonenshein. 1995. A gene required for nutritional repression of the *Bacillus subtilis* dipeptide permease operon. Mol. Microbiol. **15**:689-702

Steffen, C. & M.K. Matzura. 1993. Nucleotide sequence analysis and expression studies of a chloramphenicol-acetyltransferase coding gene from *Clostridium perfringens*. Gene. **75**:349-354

Steinmetz, M., D. Le Coq, S. Aymerich, G. Gonzy-Treboul & 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

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

Swinfield, T-J., J.D. Oultram, D.E. Thompson, J.K. Brehm & N.P. Minton. 1990. Physical characterisation of the replication region of the *Streptococcus faecalis* plasmid pAMβ1. Gene. **87**:79-90

Switzer, R.J., R.J. Turner & Y. Lu. 1999 Regulation of the *Bacillus subtilis* pyrimidine biosynthetic operon by transcriptional attenuation: Control of gene expression by an mRNA – binding protein. Prog. Nucl. Acid Res. Mol. Biol. **62**:329-367

Tortosa, P. & D. Le Coq. 1995. A ribonucleic antiterminator sequence (RAT) and a distant palindrome are both involved in sucrose induction of the *Bacillus subtilis sacXY* regulatory operon. Microbiology. **141**:2921-2927

Wray, Jr., L.V. & S.H. Fisher. 1994. Analysis of the *Bacillus subtilis hut* operon indicates that histidine-dependent induction is mediated primarily by transcriptional antitermination and that amino acid repression is mediated by two mechanisms: regulation of transcription initiation and inhibition of histidine transport. J. Bacteriol. **176**:5466-5473.

Wray, Jr., L.V., A.E. Ferson, K. Rohrer & S.H. Fisher. 1996. TnrA, a transcription factor required for global nitrogen regulation in *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA. **93**:8841-8845

Wray, Jr., L.V., J.M. Zalieckas & S.H. Fisher. 1998. Mutational analysis of the TnrA-binding sites in the *Bacillus subtilis nrgAB* and *gabP* promoters. J. Bacteriol. **180**:2943-2949

Yang, M., A. De Saizieu, A.P.G.M VanLoon & P. Gollnick. 1995. Translation of *trpG* in *Bacillus subtilis* is regulated by the *trp* RNA-binding attenuation protein (TRAP). J. Bacteriol. **177**:4272-4278

Yang, M. X-P. Chen & P. Gollnick. 1997. Alanine-scanning mutagenesis of *Bacillus subtilis trp* RNA-binding attenuation protein (TRAP) reveals residues involved in tryptophan binding and RNA binding. J. Mol. Biol. **270**:696-710

Yarnell, W.S. & J.W. Roberts. 1992. The phage lambda gene Q transcription antiterminator binds DNA in the late gene promoter as it modifies RNA polymerase. Cell. **69**:1181-1189

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

Zappe, H. 1986. PhD thesis, Department of Microbiology, University of Cape Town, South Africa.

Zappe, H., D.T. Jones & D.R. Woods. 1986. Cloning and expression of *Clostridium acetobutylicum* endoglucanase, cellobiase and amino acid biosynthesis genes in *Escherichia coli*. J. Gen. Microbiol. **132**:1367-1372

Zappe, H., W.A. Jones, D.T. Jones & D.R. Woods. 1988. Structure of an endo-β-1,4-glucanase gene from *Clostridium acetobutylicum* P262 showing homology with endoglucanases from *Bacillus* species. Appl. Env. Microbiol. **54**:1289-1292

Turner, R.J., Y. Lu & R.L. Switzer. 1994. Regulation of the *Bacillus subtilis* pyrimidine biosynthetic (*pyr*) gene cluster by an autogenous transcriptional attenuation mechanism. J. Bacteriol. **176**:3708-

Turner, R.J., E.R. Bonner, G.K. Grabner & R.L. Switzer. 1998. Purification and characterisation of *Bacillus subtilis* PyrR, a bifunctional mRNA-binding protein/uracil phosphoribosyltransferase. J. Biol. Chem. **273**:5932-

Usdin, K.P., H. Zappe, D.T. Jones & D.R. Woods. 1986. Cloning, expression and purification of glutamine synthetase from *Clostridium acetobutylicum*. Appl. Env. Microbiol. **52**:413-419

Wang, L., R. Grau, M. Perego & J.A. Hoch. 1997. A novel histidine kinase inhibitor regulating development in *Bacillus subtilis*. Genes. Dev. **11**:2569-2579

Weisberg, R.A. & M.E. Gottesman. 1999. Processive antitermination. J. Bacteriol. **181**:359-367

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

Williams, D.R., D. I. Young, J.D. Oultram, N.P Minton & M Young. 1990a. Development and optimisation of conjugative plasmid transfer from *Escherichia coli* to *Clostridium acetobutylicum* NCIB 8052. In: Anaerobes in medicine and industry, S.P. Borriello & J.M. Hardie, eds., Wrightson Biomedical, Petersfield

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

Wilson, S. & R. Drew. 1991. Cloning and DNA sequence of *amiC*, a new gene regulating expression of the *Pseudomonas aeruginosa* aliphatic amidase, and purification of the *amiC* product. J. Bacteriol. **173**:4914-4921

Wilson, S.A., S.J. Wachira, R.E. Drew, D. Jones & L.H Pearl. 1993. Antitermination of amidase expression in *Pseudomonas aeruginosa* is controlled by a novel cytoplasmic amide-binding protein. EMBO J. **12**:3637-3642

Wilson, S.A. & R.E. Drew. 1995. Transcriptional analysis of the amidase operon from *Pseudomonas aeruginosa*. J. Bacteriol. **177**: 3052-3057

Wilson, S.A., S.J.M. Wachira, R.A. Norman, L.H. Pearl & R.E. Drew. 1996. Transcription antitermination regulation of the *Pseudomonas aeruginosa* amidase operon. EMBO J. **15**:5907-5916

Woods, D.R. & S.J. Reid. 1995. Regulation of nitrogen metabolism, starch utilisation and the β-*hbd-adh1* gene cluster in *Clostridium acetobutylicum*. FEMS Microbiol. Rev. **17**:299-306