Regulation and characterization of the nitrogen assimilatory gene cluster in *Clostridium acetobutylicum* P262

by

Helen Stutz

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Microbiology

UNIVERSITY OF CAPE TOWN

August 2000

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.

University of Cape Town

To my poppet, Carla

Acknowledgements

I am deeply grateful to all those people who have helped make this thesis possible. Special thanks go to my supervisor, Dr. Shez Reid for her sound guidance, enthusiasm and unfailing support through thick and thin. I am also grateful for the guidance received from Prof. D.R Woods who was co-supervisor for the first years of this study. Thanks to all the past and present members of the Anaerobe lab for their camaraderie, help and valuable discussions over the past years. Dr. Abratt was invaluable for her expertise in the smooth running of the anaerobic box, and generously lent her assistance with the GS assays when an extra pair of hands was needed. Dr. Helen Collett spent considerable effort proofreading the thesis and made valuable suggestions. Her moral support was very much appreciated. I also greatly appreciate the excellent technical and administrative assistance provided by Di James, Anne-Marie Clennel, Di de Vielliers, and Ernest Clark.

On a more personal note, I would like to extend my gratitude to my family: to my parents and parents-in–law for their encouragement, understanding and support in taking care of Carla; to my sisters for their companionship and all the hours of 'baby sitting', but most of all to Pete for generously sharing his computer knowledge, for compiling the references and for helping with all the final stages of preparation.

I acknowledge financial assistance from the Foundation for Research Development during the early part of this work.

TABLE OF CONTENTS

ABSTRACT ii
ABBREVIATIONS iv
CHAPTER 1 General introduction 1
CHAPTER 2 Cloning and molecular characterization of the genes encoding glutamate synthase (GOGAT) from *Clostridium acetobutylicum* P262 52
CHAPTER 3 Characterization of *gltX*, a second GOGAT β subunit-like gene in *C. acetobutylicum* P262 79
CHAPTER 4 Expression studies on the *C. acetobutylicum* P262 *gltA, gltB* and *gltX* genes in an *E. coli* GOGAT mutant 111
CHAPTER 5 Regulation of GS and GOGAT activity from *C. acetobutylicum* P262 128
CHAPTER 6 General conclusions 159
APPENDIX A *E. coli* strains, genotypes and references 169
APPENDIX B Plasmid vectors 170
APPENDIX C Multiple sequence alignment of GOGAT large subunit domains 175
APPENDIX D Sequence adjacent to the *C. acetobutylicum* P262 *gltX* gene 181
APPENDIX E GSMM minimal medium recipe 186
LITERATURE CITED 189

ABSTRACT

The solventogenic *Clostridium acetobutylicum* strain P262 was used for the commercial production of solvents in South Africa from 1945 to 1983 (Jones and Woods, 1986a). Our laboratory has focused on understanding two fundamental aspects of its physiology, namely nitrogen metabolism and electron transport pathways. The long term goal is the potential genetic modification of nitrogen utilization and/or electron distribution in the cell, to manipulate fermentation patterns for improved growth rate and solvent yields. The main aims of this research project were to: extend the sequence analysis of the *glnA* locus encoding the glutamine synthetase (GS) enzyme (Janssen *et al.*, 1990; Fierro-Monti *et al.*, 1992); to identify and characterize the locus encoding the structural genes for the glutamate synthase (GOGAT) enzyme; and to determine the mechanisms and growth conditions which regulate the activity of these two key enzymes of nitrogen assimilation. In addition, we were interested in characterizing clone pMET13C1, which was isolated by a selection system developed to clone genes involved in electron transport from *C. acetobutylicum* (Santangelo *et al.*, 1991).

Sequence analyses revealed that the region downstream of *glnA* and the putative regulatory gene, *glnR* (Reid and Woods, 1995), encoded the structural genes for the large (α) and small (β) subunits of GOGAT, respectively. This is the first report in which the genes encoding GS and GOGAT are genetically linked. The α subunit was designated *gltA*, and the downstream β subunit was designated *gltB*. All the likely cofactor and substrate binding sites identified in GOGAT enzymes (Vanoni and Curti, 1999) were conserved in the deduced *gltA* and *gltB* polypeptides. The identity of the *gltA* and *gltB* genes were functionally confirmed by complementation studies involving their co-expression from separate constructs in an *E. coli* glutamate auxotroph (strain MX3004), which restored the ability of this mutant to grow with ammonia as the sole source of nitrogen.

Physiological studies on the germination, growth and differentiation patterns of *C. acetobutylicum* P262 were assessed in relation to different nitrogen conditions. Significantly, organic nitrogen (casamino acids) was the preferred source of nitrogen, and not ammonia as

previously assumed. These studies led to the development of nitrogen limiting conditions (0.025% casamino acids + 0.15% glutamine) and nitrogen rich conditions (0.2% casamino acids) used for regulatory studies. GOGAT activity was optimised. It appeared to be sensitive to oxygen and specific for the co-enzyme NADH. Both GS and GOGAT activities were regulated by the nitrogen source in a similar way: induced in the nitrogen limiting conditions, and repressed in the nitrogen rich conditions. Northern blot analyses, in conjunction with the enzyme activity profiles and feedback inhibition studies, indicated that GS and GOGAT activities were regulated primarily at the level of transcription. Furthermore, *glnA* and *glnR*, and *gltA* and *gltB* are each transcribed as an operon under nitrogen limiting conditions. No assimilatory GDH activity could be detected. The implications of these results, as well as sequence features identified, are discussed in context with a proposed model for the regulation of GS and GOGAT activity in *C. acetobutylicum* P262.

Analysis of plasmid pMET13C1 identified a gene whose predicted ~46 kDa product was associated with an electron transport function. The deduced amino acid sequence was not typical of electron transport proteins, but rather shared striking homology to bacterial GOGAT β subunit polypeptides. This β subunit-like gene was thus designated *gltX*. We were, however, unable to relate it to GOGAT activity or nitrogen metabolism. Rather, it appears to belong to a novel family of FAD-dependent NAD(P)H oxidoreductases suggested by Vanoni and Curti (1999), and supported by an analyses of the evolutionary relationships of the GOGAT subunits/domains from various sources.

ABBREVIATIONS

A	adenosine
aa	amino acid(s)
ABE	acetone-butanol-ethanol
A_x	absorbance at x nm
ATCC	American Type Culture Collection
ATP	adenosine 5'-triphosphate
bp(s)	base pair(s)
C	cytidine
casamino acids	acid hydrolyzed casein
CBM	*Clostridium* Basal Medium
CFE	cell free extract
CoA	coenzyme A
DIG	digoxygenin
DNA	deoxyribonucleic acid
DSM	Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH
EDTA	ethylenediaminetetra-acetic acid
FAD	flavin adenine dinucleotide
Fd	ferredoxin
Fe	iron
FMN	flavin mononucleotide
g	gram or gravity, depending on context
G	guanosine
ΔG	Gibbs free-energy change
GDH	glutamate dehydrogenase
GOGAT	glutamate synthase
GS	glutamine synthetase
GSMM	glucose-mineral salts-biotin medium
h	hour(s)
H-T-H	helix-turn-helix motif
IPTG	isopropyl-β-D-thiogalactopyranoside
kb	kilobase pairs
kCal	kilocalories
kDa	kilodaltons
Km	Michaelis constant
l	litre(s)
LB	Luria-Bertani broth
LHS	left hand side
log	logarithmic
m	metre(s)
MIC	minimal inhibitory concentration
min	minute(s)

mol	mole(s)
mRNA	messenger RNA
MSG	monosodium glutamate
NADH	nicotinamide adenine dinucleotide (reduced)
NADPH	nicotinamide adenine dinucleotide phosphate (reduced)
NCP	National Chemical Products
NCIMB	National Collection of Industrial and Marine Bacteria
n	nano
nt	nucleotide
OD_x	optical density at x nm
ORF	open reading frame
ori	origin of replication
p	plasmid
PAGE	polyacrylamide gel electrophoresis
RBS	ribosome binding site
RHS	right hand side
RNA	ribonucleic acid
RNAse	ribonuclease
rRNA	ribosomal RNA
s	second(s)
SD	Shine-Dalgarno
SDS	sodium dodecyl sulfate
T	thymidine
TAE	Tris-acetate EDTA buffer
Tris	tris(hydroxymethyl)aminomethane
UV	ultraviolet (light)
WT	wild type strain
w/v	weight per volume (in g per 100ml)
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
YT	yeast tryptone medium
Δ	deletion
α	alpha
β	beta
γ	gamma
λ	lambda (phage)
μ	micro
σ	sigma

CHAPTER 1

General introduction

1.1 Introduction to *Clostridium acetobutylicum* P262 and the solventogenic clostridia..........2
1.1.1 *Clostridium acetobutylicum* research in context ---2
1.1.2. History of the ABE fermentation ---2
1.1.3 Taxonomy of the solventogenic clostridia--3
1.1.4 Growth, solventogenesis and differentiation of the solventogenic clostridia. --------------5
1.1.5 Nitrogen metabolism and solvent production.--8
1.1.6 Genetic manipulation of the solventogenic clostridia --8
1.2 General aspects of bacterial nitrogen metabolism 10
1.2.1 Nitrogen sources -- 10
1.2.2 The significance of glutamine and glutamate -- 11
1.2.3 Ammonia assimilation and biosynthesis of glutamine and glutamate -------------------- 11
1.2.4. Physiological significance of the GS-GOGAT and GDH pathways in ammonia assimilation -- 14
1.2.5 Characteristics of GS enzymes -- 15
1.3 Global nitrogen regulation in the enteric bacteria...16
1.3.1 The Ntr system -- 16
1.3.2 Expression of the *glnAntrBC* operon in *Escherichia. coli* (and related bacteria) -------- 18
1.3.3 The *nac* (nitrogen assimilation control) gene-- 19
1.4 Nitrogen control in Gram-positive bacteria..20
1.4.1 Nitrogen control and the regulation of GS activity in *Bacillus subtilis* ------------------- 20
1.4.2 Nitrogen control and the regulation of GS activity in *Streptomyces* spp. ----------------- 24
1.4.3 Structure and regulation of *C. acetobutylicum glnA* region ------------------------------- 25
1.5 Glutamate synthase (GOGAT) 30
1.5.1 General introduction -- 30
1.5.2 Classes of GOGAT --- 31
1.5.3 The role of GOGAT in plants-- 32
1.5.4 Enzyme model -- 33
1.5.5 Functional domains -- 35
1.5.6 Common model --- 37
1.5.7 Post-translational processing-- 37
1.5.8 GOGAT evolution -- 38
1.5.9 Regulation of GOGAT activity-- 39
1.6 Aims of this study 49

CHAPTER 1

General introduction

1.1 Introduction to *Clostridium acetobutylicum* P262 and the solventogenic clostridia

1.1.1 *Clostridium acetobutylicum* research in context

Acetone, butanol and isopropanol are important additives for the chemical industry besides their traditional uses as solvents. For decades these chemicals were produced by a fermentation process using strains of solventogenic clostridia, but are now manufactured more cost-effectively from petroleum. However, the search to strengthen petroleum importing nations, together with the increasing environmental awareness, has revived interest in the *Clostridium acetobutylicum* acetone-butanol (AB) fermentation process. The bioconversion of agricultural by-products into refined chemicals is an attractive alternative to the exploitation of fossil fuels. However, the economic viability of this process depends on considerably improved product yield and product selectivity. In the last two decades significant advances have been made towards understanding the biochemistry and genetics of solvent production. With a draft sequence of the *C. acetobutylicum* ATCC824 genome recently available (accession number AE001438), and the development of genetic tools for the clostridia, genetic manipulation of strains for industrial purposes is now becoming feasible.

1.1.2. History of the ABE fermentation

The outbreak of the First World War in 1914 resulted in a demand for acetone, a strategic material required for the manufacture of munitions. This lead to the development of the first large scale industrial fermentation, pioneered by Chaim Weizmann, for the production of acetone and butanol (Jones and Woods, 1986a; Jones and Keis, 1995). Acetone was produced from maize mash by a starch-utilizing strain of *C. acetobutylicum*, initially designated BY. After the war, the Commercial Solvents Corporation (CSC, USA) acquired the patent rights for the Weizmann acetone-butanol (AB) fermentation process. The early AB fermentations were all based on corn mash, a substrate rich in starch, however with the abundant and cheap supply

of molasses available in the 1930's, the process switched to a molasses-based AB fermentation. This necessitated the isolation of effective sugar-fermenting *C. acetobutylicum* strains (Jones and Keis, 1995). The expiry of the CSC patent in 1935 saw the establishment of fermentation plants worldwide and a renewed search for viable strains. The AB fermentation process served as the main supply of industrial solvents until it was superseded in the 1960's by the more economical synthesis from petrochemicals (Jones and Woods, 1986a). Today, the AB fermentation process is operated only in the People's Republic of China. However, with the revived interest in the production of solvents from renewable resources, and improved product recovery techniques, a pilot plant-scale study using potato as the raw material is now underway in Austria (Dürre, 1998).

1.1.3 Taxonomy of the solventogenic clostridia

The genus *Clostridium* comprises an extremely diverse group of Gram-positive, obligately anaerobic, endospore-forming bacilli (Cato and Stakebrand, 1989). The best known solvent producing bacterium is the Weizmann organism, *Clostridium acetobutylicum*. The majority of industrial solvent producing strains isolated subsequently were named haphazardly according to the specific patented process and were also referred to as *C. acetobutylicum* strains. However, research on a number of the most commonly studied solvent-producing strains, highlighted significant differences in their biochemical, physiological and genetic properties (Wilkinson and Young, 1993; Johnson and Chen, 1995). This prompted a detailed molecular study of the taxonomic relationships of solventogenic strains and related organisms (Keis *et al.*, 1995; Wilkinson *et al.*, 1995; Johnson *et al.*, 1997). This work revealed that these "*C. acetobutylicum*" strains belonged to four taxonomic groups (Fig. 1.1), each proposed to represent a different species (Keis *et al.*, 1995; Johnson *et al.*, 1997). Group I is represented by *C. acetobutylicum* ATCC 824, the *C. acetobutylicum* type strain, and includes strain DSM 792 (the DSM *C. acetobutylicum* type strain). Group II was represented by strains NRRL B643 and NCP 262, an NCP production strain used in South Africa from 1945 to 1983 (Jones and Woods, 1986a), and the laboratory strain used in this study. It was derived from *Clostridium saccharo-butyl-acetonicum-liquifaciens* (patented by CSC), one of the main industrial saccharolytic, solvent producing strains. It has been proposed that *C. acetobutylicum* NCP 262 and other members in this group should be referred to as *C. saccharobutylicum* in future, however this is not yet official (Shaheen *et al.*, 2000). For the purpose of this thesis we have retained the name *C. acetobutylicum* NCP 262, or just simply *C. acetobutylicum* P262. The

third group consisted of "C. *saccharoperbutylacetonicum*" strains represented by N1-4, while the *C. beijerinckii* strains, represented by strain NCIMB 8052, constituted Group IV.

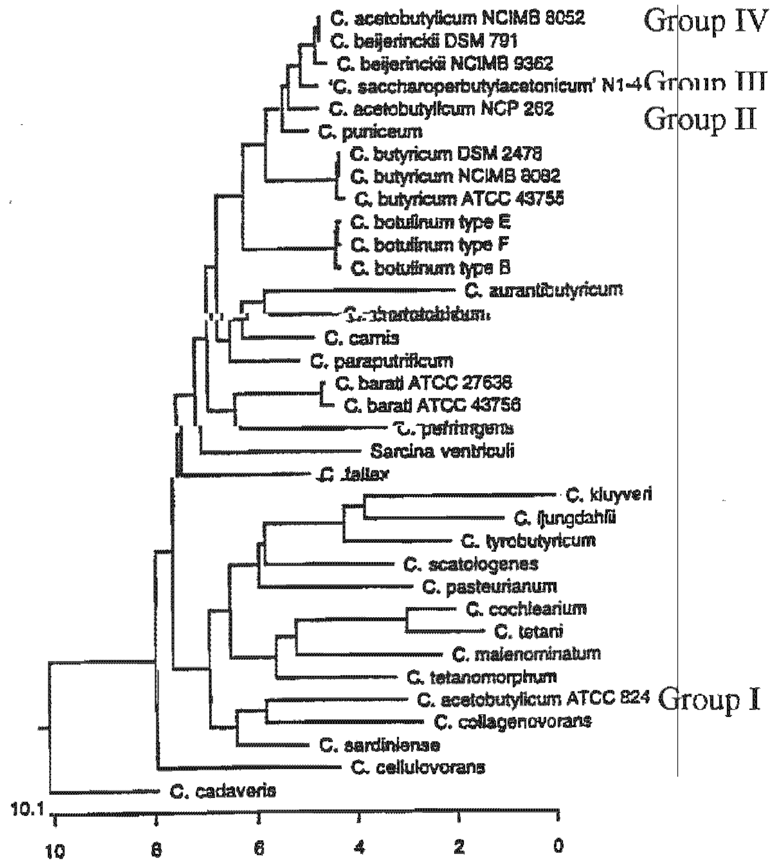


Fig. 1.1 Unrooted phylogenetic dendogram showing the taxonomic positions of: Group I representative *C. acetobutylicum* ATCC 824; Group II representative *C. acetobutylicum* NCP 262; Group III representative *C. saccharoperbutylacetonicum* N1-4 and Group IV representative *C. beijerinckii* NCIMB 8052. The dendogram was based on comparison of the 16S ribosomal RNA gene sequences (positions 100 to 1434 of the *E. coli* numbering system). The scale bar indicates evolutionary distance (taken from Keis *et al.* (1995)).

The historical background of the AB fermentation is reflected in these taxonomic relationships (Johnson and Chen, 1995; Jones and Keis, 1995), with a clear division between the starch-utilizing *C. acetobutylicum* strains (Group I) and the sugar utilizing strains (Group II-IV). This indicates that the transition from a starch-based fermentation process to a sugar-based fermentation process selected for organisms genetically distinct from, and only distantly related to, *C. acetobutylicum*.

1.1.4 Growth, solventogenesis and differentiation of the solventogenic clostridia.

Growth of the solventogenic clostridia, including *C. acetobutylicum* NCP 262, is associated with two phases (Jones and Woods, 1986a; Woods, 1995). During the exponential growth phase, typified by highly motile rod-shaped cells, the cells convert sugars to acids (predominantly acetate and butyrate) with the concomitant production of carbon dioxide and hydrogen. The acidogenic phase is followed by a metabolic switch to the solventogenic phase as the culture enters stationary phase metabolism. This is typified by nongrowing, nonmotile, clostridial (swollen cigar shaped) cells, which convert residual sugars and preformed acids into acetone, butanol and ethanol, via reduction reactions. The solventogenic phase is terminated by sporulation. The biochemical pathways of acidogenesis and solventogenesis in *C. acetobutylicum* are well formulated (Fig. 1.2) and many of the enzymes and/or corresponding genes have been isolated and characterized from various solvent producing strains, including NCP 262. These pathways have been reviewed by (Jones and Woods, 1986a; Rogers and Gottschalk, 1993; Dürre *et al.*, 1995; Mitchell, 1998). The basic features are outlined below.

The central pathway intermediates acetyl-CoA, acetoacetyl-CoA, and butyryl-CoA, are the key metabolites from which both solvents and acids are derived. During the acid producing phase, acetate and butyrate are derived from acetyl-CoA and butyryl-CoA respectively, via a parallel set of reactions (7, 8, 16, and 17; Fig. 1.2) yielding ATP. During solventogenic metabolism, acetate and butyrate are reassimilated, primarily via CoA transferase (13; Fig. 1.2), yielding predominantly central pathway intermediates, from which acetone, butanol and ethanol are derived in a ratio of ~6:3:1. Excess NAD(P)H generated by glycolysis serves as the reducing power. Small amounts of ethanol are produced constitutively throughout the fermentation (Gerischer and Dürre, 1992). In some strains acetone is further reduced to isopropanol (15; Fig. 1.2). Thus the complex network of metabolic pathways in the solventogenic clostridia results in the production of five major products from acetyl-CoA via three branch points in the fermentation pathway. Although this metabolic variety is essential to the physiology and adaptability of the organism, it would be commercially attractive if fermentation could be directed towards the production of just one or two of these products.

There has been much interest in determining the physiological factors that influence the onset of solventogenesis, and the mechanistic basis for the switch, as this is key for the optimization of solvent production. Some of the factors that trigger solvent production include a decreased pH, accumulation of undissociated acids, nutrient limitation and alteration of electron flow

(Mitchell, 1998), however the complete mechanism of regulation of solvent production still awaits elucidation. As discussed below, there is now evidence that the initiation of solvent formation is subject to control mechanisms similar to other stationary phase phenomena.

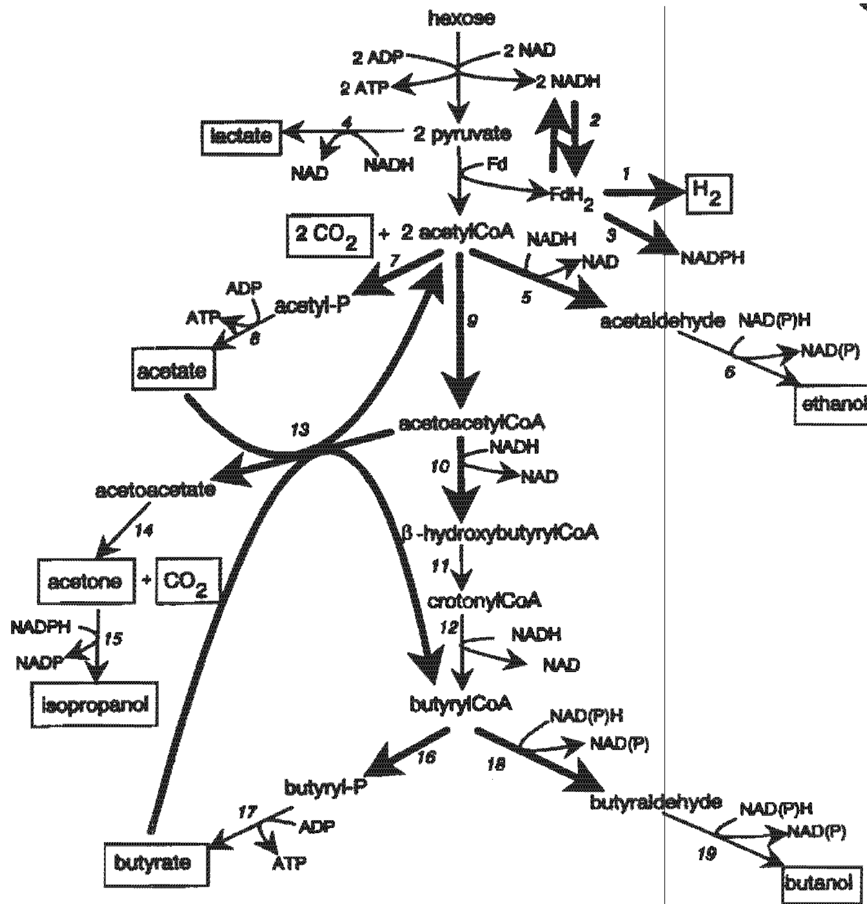


Fig. 1.2 The acetone-butanol fermentation pathways in the solventogenic clostridia.

Hexoses are metabolized via the Embden-Meyerhof pathway yielding pyruvate, which is then metabolized to acetyl-CoA by pyruvate-ferredoxin oxidoreductase, or to lactate by lactate dehydrogenase [4].

Numbers correspond to the appropriate enzyme names. [1] hydrogenase; [2] ferredoxin-NAD oxidoreductase; [3] ferredoxin-NADP reductase.

Central pathways are catalyzed by: [9] thiolase; [10] β-hydroxybutyryl-CoA dehydrogenase; [11] crotonase; [12] butyryl-CoA dehydrogenase.

Acidogenic pathways are catalyzed by: [7] phosphotransacetylase; [8] acetate kinase; [16] phosphotransbutyrylase; [17] butyrate kinase.

Solventogenic pathways are catalyzed by: [5] acetyldehyde dehydrogenase; [6] ethanol dehydrogenase; [13] acetoacetyl-CoA:acetate/butyrate:CoA transferase; [14] acetoacetate decarboxylase; [15] isopropanol dehydrogenase; [18] butyraldehyde dehydrogenase; [19] butanol dehydrogenase.

Bold arrows indicate enzyme reactions at key branch points, and end-products formed at various stages in the fermentation are boxed (Taken from Mitchell, 1998).

The transition from exponential growth to stationary phase is also the point at which sporulation is initiated and some strains have been shown to exhibit morphological and structural changes at this stage. The onset of solventogenesis in *C. acetobutylicum* P262 is strongly linked to endospore formation (Long *et al.*, 1984a; Long *et al.*, 1984b), however it is not a sporulation-specific event since asporogenous solvent-producing mutants, and sporulating asolventogenic mutants have been isolated (Jones *et al.*, 1982; Long *et al.*, 1984a; Babb *et al.*, 1993). It is possible that both processes are initiated in response to acid stress. It is thought that solventogenesis, provides a temporary mechanism to overcome the toxic effects of accumulated acid end products by recycling them to form alcohols, until the time-consuming process of sporulation is complete (Dürre *et al.*, 1995). However, the simultaneous loss of spore-forming potential and solvent-production in several *C. acetobutylicum* mutants and degenerate strains (Jones *et al.*, 1982; Long *et al.*, 1984a; Long *et al.*, 1984b; Woolley and Morris, 1990) indicates that solventogenesis and sporulation share common regulatory features.

Sporulation, viewed as an adaptive response which ensures survival of microorganisms under unfavorable conditions, is well characterized in *Bacillus subtilis* which is recognized as a model system for differentiation in bacteria. It involves 7 defined morphological stages (reviewed by Errington, 1993). Sporulation in *C. acetobutylicum* proceeds via a similar sequence of morphological changes (Long *et al.*, 1983), however, it is not triggered in response to nutrient deprivation as in *Bacillus*, but rather by cessation of growth in the presence of excess nutrients (carbon and nitrogen) or exposure to oxygen. Despite their different induction patterns, emerging evidence reveals basic similarities in the regulation of sporulation between *Bacillus* and *C. acetobutylicum.* For example, a number of sporulation specific sigma factor homologues have been identified in *C. acetobutylicum* (Sauer *et al.*, 1994; Wong *et al.*, 1995; Santangelo *et al.*, 1998). In *B. subtilis,* initiation of spore formation is controlled primarily by the SpoOA protein, a phosphorylation-activated transcription factor involved in the regulation of stationary phase metabolism (Brown *et al.*, 1994). Homologues have been identified in several other *Bacillus* and *Clostridium* species (Brown *et al.*, 1994), suggesting that it is a key regulator in endospore-forming bacteria. Insertional inactivation of the *C. beijerinckii spoOA* gene resulted in a strain unable to initiate solventogenesis or sporulation (Brown *et al.*, 1994). This provided the first direct evidence that its gene product probably encoded a regulator of both sporulation and solventogenesis (Wilkinson *et al.*, 1995). Furthermore, identification of OA box motifs in the upstream regulatory regions of several *Clostridial* genes concerned with fermentative

metabolism, whose expression is modulated at the onset of solventogenesis (Wilkinson *et al.*, 1995), strongly implicate the SpoA protein in regulation of the solventogenic response.

1.1.5 Nitrogen metabolism and solvent production.

When molasses, which is deficient in nitrogen, replaced the historical corn mash, the industrial process became more complex because nitrogenous compounds had to be added to the sugar based medium. Solvent yield was affected by the nitrogen source, however the effect could not be further analyzed in a complex medium. Since both acid production and the consumption of ammonia ions cause a decrease in pH of the medium, it had to be carefully maintained. The use of soluble, chemically defined media for laboratory purposes makes pH control an even more critical issue, because the buffering capacity of the medium is significantly reduced when ammonia salts or amino acids replace less metabolizable proteins and peptides.

Varying results have been reported on the effect nitrogen (ammonia) limitation has on solvent production. In chemically defined medium, under condition of high glucose and low ammonia concentration, *C. acetobutylicum* ATCC 824 produced a high concentration of solvents at pH5.0 (Monot *et al.*, 1982; Monot and Engasser, 1983). Similarly, Andersch *et al.* (1982) found that strain DSM1731 produced both acids and solvents at low pH and under ammonia limitation. The ratio between the carbon and nitrogen substrates, as well as the culture pH, has been reported to have a significant effect on solvent productivity (Roos *et al.*, 1985). On the other hand, *C. beijerinckii* NCIMB 8052 and the closely related *C. acetobutylicum* NCP 262 did not produce solvents in ammonia limited cultures (Gottschal and Morris, 1981; Long *et al.*, 1984a). This is consistent with the finding that *C. beijerinckii* required partially hydrolyzed proteins or amino acids for solvent production, presumably reflecting a preference for these amino acids as nitrogen sources. Thus a number of variables may combine to influence the effectiveness of solvent formation under ammonia limited conditions.

1.1.6 Genetic manipulation of the solventogenic clostridia

Until recently, the solventogenic clostridia have lacked well developed genetic systems. However, several plasmid vectors are now available, and methods have been developed for gene transfer, overexpression and inactivation.

A number of shuttle vectors (e.g. pMTL500E) have been constructed by combining Gram-negative and Gram-positive replicons from broad host range plasmids, with appropriate

antibiotic resistance markers, e.g. erythromycin, that are functional in Clostridia (Williams *et al.*, 1990; Minton *et al.*, 1993). The *B. subtilis/C. acetobutylicum* shuttle vector pFNK1, which has been designed to avoid the restriction system of *C. acetobutylicum* ATCC 824, has been successfully used for the overexpression of genes in strain ATCC824 (Mermelstein *et al.*, 1992). A system for complete *in vivo* protective methylation of *E. coli/C. acetobutylicum* shuttle vectors harboring the ColE1 origins of replication (susceptible to the restriction endonuclease *Cac*824I) have also been developed by including the ϕ3T I methyltransferase gene from *B. subtilis* phage ϕ3T (Mermelstein and Papoutsakis, 1993).

Clostridia do not appear to be naturally competent, however protoplasting and electroporation have both been successfully applied for plasmid transformation of several clostridial species including *C. acetobutylicum* ATCC 824 and derivatives of N1-4 (Mermelstein *et al.*, 1992; Reysset and Sebald, 1993). Recently electroporation and plasmid isolation procedures were optimised for *C. acetobutylicum* DSM 792 (Nakotte *et al.*, 1998).

An alternative and successfully applied technique is the conjugative mobilization of plasmids to *C. acetobutylicum*. This technique is based on the broad-host-range conjugation system encoded by IncP plasmids, and relies on an IncP helper plasmid to provide conjugation functions *in trans*. It has been developed to mobilize non conjugative shuttle vectors (Williams *et al.*, 1990), and adapted for the targeted integration of non-replicative mobilizable plasmids (e.g. pMTL30) into the *C. beijerinckii* genome (Wilkinson and Young, 1994). This is a revolutionary development in the field of clostridial genetics. Integration appears to occur via a Campbell-like recombination event between regions of clostridial target DNA cloned into the vector, and the homologous regions present on the clostridial genome. Unfortunately these approaches require a high transformation frequency to detect a low-frequency event such as recombination, which limits its use in strains such as *C. acetobutylicum* 262 in which acceptable transformation protocols and reliable genetic techniques have not yet been developed.

In addition, conjugative transposon mutagenesis is an effective way of producing mutations and identifying the inactivated gene by cloning of adjacent DNA, and can be exploited for those organisms which lack well-developed genetic systems such as *C. acetobutylicum* P262 (Babb *et al.*, 1993). While this approach has been used to generate an assortment of mutants, it is not

ideal as the transposon was found to form multiple insertions and in some cases promoted chromosomal rearrangements (Babb *et al.*, 1993).

The application of these new genetic techniques offers the opportunity to improve the fermentation process by genetic manipulation of metabolic pathways, and to increase our understanding of gene function. However, at present conjugative transposon mutagenesis is the only system for DNA delivery in strain P262.

1.2 General aspects of bacterial nitrogen metabolism

Bacterial nitrogen metabolism is a vast field of research reflected in the enormous amount of literature available. In preparing the following general discussion (Section 1.2) on the relevant aspects of this broad topic, the reviews of Reitzer and Magasanik (1987), Schreier (1993), Merrick and Edwards (1995) and Reitzer (1996) were extensively consulted.

1.2.1 Nitrogen sources

Bacteria can utilize a wide variety of nitrogen compounds as sources of cellular nitrogen. These range from simple inorganic compounds such as diatomic nitrogen and nitrate to more complex compounds including urea, amino acids, amino sugars and nucleosides. Endogenous ammonia is an absolute requirement for the biosynthesis of at least some essential nitrogen-containing compounds, thus the growth rate of bacteria on different nitrogen sources is generally limited by the rate at which ammonia ions can be generated from them. It is not surprising then that in the enteric (and many other) bacteria, ammonia supports the fastest cell growth rate, however this is not always the case. In *Bacillus subtilis* and *Corynebacterium callunae,* glutamine and glutamate are the preferred nitrogen sources, respectively (Atkinson and Fisher, 1991; Ertan, 1992). The expression of metabolic pathways involved in nitrogen metabolism, such as those required for nitrogen fixation, for the transport, catabolism and conversion of nitrogenous compounds to ammonia ions and other components, and for the assimilation of ammonia, are generally regulated in response to nitrogen availability such, that under conditions of nitrogen excess the cells utilize the preferred nitrogen source, and repress the mechanisms of utilization of alternative nitrogen sources. However, when ammonia ions becomes limiting, these alternative routes can be activated. The combination of pathways available and the regulatory systems controlling the preferential use of nitrogen compounds, depends on the particular

organism. Such a system of nitrogen control has only been studied comprehensively in the enteric bacteria, which will be discussed later.

1.2.2 The significance of glutamine and glutamate

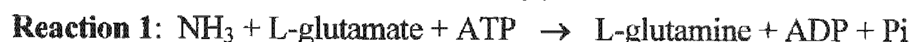
In most bacteria, the major route for the assimilation of ammonia ions is through the amino acids, glutamine and glutamate. These two compounds are key metabolites in bacterial metabolism, serving as principal precursors for nearly all other nitrogen containing compounds. The amidotransferases catalyze the transfer of the amide group of glutamine in the biosynthesis of a variety of compounds including amino sugars, purines, pyrimidines, *p*-aminobenzoate, tryptophan, histidine, asparagine and NAD. The transaminases catalyze the transfer of the amino group of glutamate in the synthesis of practically all other amino acids.

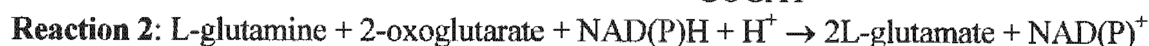
In addition to the biosynthesis of amino acids, glutamate plays an important role in a number of other functions, such as equalizing the external osmotic strength of the medium (Saroja and Gowrishankar, 1996), and influencing the UV and heat resistance of spores (Schreier, 1993). In *Bacillus* spp, the large glutamate pools provide carbon and nitrogen precursors, and possibly energy, during spore formation (Schreier, 1993). Glutamine is an important regulatory metabolite, signaling the intracellular nitrogen status in many bacteria.

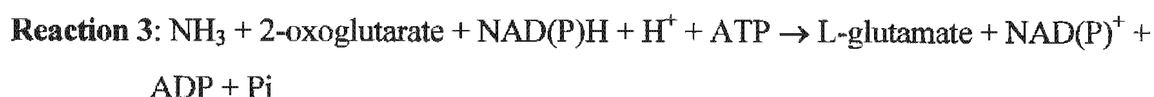
1.2.3 Ammonia assimilation and biosynthesis of glutamine and glutamate

Two major pathways facilitate the incorporation of ammonia into glutamine and glutamate. The most important pathway is the energy dependent glutamine synthetase-glutamate synthase (GS-GOGAT) pathway, which is ubiquitous in bacteria. The enzyme glutamine synthetase (GS: EC 6.3.1.2) catalyses the amidation of endogenous glutamate to form glutamine (Reaction 1), the only pathway for glutamine synthesis in the cell. It is worth mentioning here, however, that glutamine may in some instances (e.g. in *B.subtilis*) be produced from glutamate via a transamidation of glutamyl-charged $tRNA^{gln}$ (using asparagine, ammonia or glutamine as the nitrogen donor). Glutamate synthase (glutamine amide 2-oxoglutarate aminotransferase, for convenience designated GOGAT: EC 1.4.1.13) catalyzes the NAD(P)H-dependent transfer of the amide group from glutamine to the C(2) carbon of 2-oxoglutarate, to produce two molecules of glutamate (Reaction 2). The reactions can take place independently, but can also be understood as a cycle in which every turn results in the net synthesis of one molecule of glutamate from ammonia and 2-oxoglutarate (Reaction 3). The consumption of energy drives

these coupled reactions strongly in a forward direction, making this essentially a physiologically irreversible pathway.


Reaction 1: NH_3 + L-glutamate + ATP $\xrightarrow{GS}$ L-glutamine + ADP + Pi


Reaction 2: L-glutamine + 2-oxoglutarate + NAD(P)H + H^+ $\xrightarrow{GOGAT}$ 2L-glutamate + $NAD(P)^+$


Reaction 3: NH_3 + 2-oxoglutarate + NAD(P)H + H^+ + ATP $\rightarrow$ L-glutamate + $NAD(P)^+$ + ADP + Pi

The second route for ammonia assimilation is provided by glutamate dehydrogenase (GDH: EC 1.4.1.4). GDH catalyzes the reductive amination of 2-oxoglutarate to produce glutamate in a reversible reaction in the presence of NAD(P)H (Reaction 4):

Reaction 4: NH_3 + 2-oxoglutarate + NAD(P)H + H^+ $\xleftrightarrow{GDH}$ L-glutamate + $NAD(P)^+$

In the majority of microorganisms both the GDH and the GS-GOGAT pathways operate. The relative contribution of these pathways to ammonia assimilation and glutamate synthesis varies greatly depending on the organism and the nitrogen conditions (Section 1.2.4). However, in some microorganisms just one of the pathways can be detected. The assimilation of ammonia occurs solely by the coupled GS-GOGAT pathway in *B. subtilis* (Fisher and Sonenshein, 1991), *Clostridium pasteurianum* (Dainty, 1972), *C. thermoautotrophicum* (Bogdahn and Kleiner, 1986) and *Streptomyces clavuligerus* (Brana *et al.*, 1986) under all nitrogen conditions. On the other hand, GDH appears to be the sole enzyme in ammonia assimilation in the ruminal bacterium *Streptococcus bovis* (Chen and Russell, 1989), and is the primary enzyme involved in glutamate formation in *Bacteroides fragilis* (Yamamoto *et al.*, 1984) and in many yeast and fungi (Genetet *et al.*, 1984; Bogonez *et al.*, 1985; Kusnan *et al.*, 1987). The GDH pathway is also the prominent pathway of ammonia assimilation in some of the nitrogen fixing bacteria including *B. polymyxa* (Kanamori *et al.*, 1987a) and *B. macerans* (Kanamori *et al.*, 1987b) in both ammonia grown and nitrogen fixing cells. However, this is not a general characteristic of nitrogen fixing bacteria, since GDH activity is absent or barely detectable in other members of

this group under nitrogen limiting or nitrogen excess conditions, including *B. azotofixans* and the family Rhizobiacea, where the GS-GOGAT pathway predominates (Bravo and Mora, 1988; Kanamori *et al.*, 1988; Kanamori *et al.*, 1989). Interestingly, *S. cerevisiae* (Avendano *et al.*, 1997) and *B. fragilis* (Abrahams and Abratt, 1998) possess a second assimilatory GDH activity. In *B. fragilis* the GDH enzymes are differentially regulated with the dual cofactor NAD(P)H-dependent enzyme being induced by ammonia conditions, and the NADH-dependent enzyme being induced by organic nitrogen. Although GDH's are mostly associated with an anabolic role, they can also function in a catabolic capacity in some microorganisms. In *B. subtilis,* GDH only works catabolically to release ammonia (Belitsky and Sonenshein, 1998). NAD(P)H-dependent activity is generally associated with the anabolic function of the enzyme, while the NADH-dependent activity is generally associated with the catabolic function (Joe *et al.*, 1994). The loss of both GOGAT and GDH in enteric bacteria and *S. coelicolor,* but not either enzyme alone, or the loss of just GOGAT activity in *B. subtilis*, results in glutamate auxotrophy which can be relieved either by glutamate or a compound that can be readily metabolized to glutamate (Fisher, 1989).

Ammonia can also be assimilated via other routes, depending on the particular organism and physiological conditions. For example, in *E. coli* and *Klebsiella aerogenes,* alanine dehydrogenase (ADH; EC 1.4.1.1) may catalyze the reversible amination of pyruvate to form alanine, while asparagine synthetase (EC 6.3.1.1) catalyzes the ammonia-dependent amidation of aspartate to form asparagines (Reitzer, 1996). Ammonia-dependent asparagine synthetase has also been identified in some thermophilic clostridial strains (Bogdahn and Kleiner, 1986). In the presence of excess ammonia, ADH has an assimilatory function in *S. clavuligerus* (which does not possess GDH activity) (Aharonowitz and Friedrich, 1980; Brana *et al.*, 1986) and in the actinomycete *Norcadia mediterranei* (Mei and Jiao, 1988). However, in *S. coelicolor, C. thermoautotrophicum* and in *Bacillus* spp., ADH is involved in alanine catabolism, particularly during germination and sporulation (Bogdahn and Kleiner, 1986; Fisher, 1989; Schreier, 1993).

While the biosynthesis of glutamine occurs solely via the action of GS and absolutely requires ammonia, glutamate can be produced in additional ways to the GOGAT and GDH pathways, depending on the organism. Glutamate can potentially be produced directly from the degradation of a number of amino acids including glutamine, histidine, arginine and proline, or indirectly via a transamination reaction involving the transfer of the α-amino group from one amino acid e.g. aspartate to the keto group of 2-oxoglutarate. Glutamine can be degraded to

glutamate and ammonia by glutaminases (L-glutaminase aminohydrolase, EC: 3.5.1.2). If glutamate is a product of catabolism, ammonia is required only for glutamine synthesis, thus sparing the need for the energy-consuming synthesis of glutamate via the coupled GS-GOGAT pathway (Schreier, 1993). For nitrogen sources that generate ammonia but not glutamate (e.g. serine), GOGAT is necessary for glutamate synthesis.

1.2.4. Physiological significance of the GS-GOGAT and GDH pathways in ammonia assimilation

The presence of multiple pathways (GS-GOGAT and GDH) for the biosynthesis of glutamate in several microorganisms has been questioned. Because the GS-GOGAT pathway characteristically has a high affinity for ammonia (reflected in the low Km values of < ~0.2mM), and GDH enzymes generally have a lower affinity for ammonia (reflected in high Km values > ~3mM) (Merrick and Edwards, 1995), it was concluded that the role of the GS-GOGAT pathway was that of ammonia assimilation under ammonia limiting conditions. This was supported by an absolute requirement for GOGAT during nitrogen fixing conditions in a number of bacteria (Merrick and Edwards, 1995), and during nitrogen limited growth of *E. coli, Klebsiella aerogenes* and *S. coelicolor*, while GOGAT mutants grew like WT cells in ammonia sufficient conditions due to induced GDH activities (Tyler, 1978; Fisher, 1989; Reitzer, 1996). However, since GDH is often a dispensable enzyme, and since the GS-GOGAT pathway is energy dependent, it was suggested that the GDH pathway may be important for the assimilation of ammonia during growth of bacteria on low energy yielding, nitrogen rich substrates. Indeed, *E. coli* strains deficient in GDH activity were at a competitive disadvantage relative to WT strains in an energy limited environment (Helling, 1994).

However, while the proposed roles of the GDH and GS-GOGAT pathways under the different nitrogen conditions can be applied to some microorganisms (Helling, 1994; Valenzuela *et al.*, 1995), this model does not hold true for many other microorganisms studied. For example, some organisms, including many nitrogen fixing bacteria, have GDH's with low Km's for ammonia and can thus assimilate ammonia during ammonia limited conditions (e.g. during nitrogen fixation). In *Bacteroides* spp (Yamamoto *et al.*, 1984; Baggio and Morrison, 1996), *S. cereviseae* (Folch *et al.*, 1989) and *Neurospora crassa* (Lomnitz *et al.*, 1987), GDH is used to incorporate ammonia during either nitrogen limiting or nitrogen excess conditions, and in the fungus *A. nidulans,* the GDH and GS-GOGAT pathways operate concurrently (Kusnan *et al.*, 1987). The choice of pathway may depend on the fermentation pathways used by the organism,

and the generation of ATP and NAD(P)H (Kanamori *et al.*, 1989). For example, the assimilation of ammonia through the GDH pathway without further expenditure of energy may be more advantageous to those organisms having low energy yielding fermentation pathways.

For the purpose of this thesis, we will be focusing on the GS-GOGAT pathway in relation to nitrogen metabolism in bacteria. Since glutamine and glutamate are central to nitrogen metabolism, the regulation of GS and GOGAT activities are generally tightly controlled in response to the levels of nitrogen source available. Thus, cells growing with preferred nitrogen sources generally contain low levels of GS and GOGAT, whereas high levels of these enzymes are present during nitrogen limited growth to ensure adequate supplies of these key metabolites. Although the reactions catalyzed by the GS and GOGAT pathways are tightly coupled processes and are highly conserved in bacteria from diverse ecological niches, the genes encoding these enzymes are not linked, and no co-ordinate regulation has yet been documented. Furthermore, the structure and regulation of these key enzymes of nitrogen assimilation vary considerably, presumably reflecting different metabolic needs. Thus, we will now consider each enzyme separately.

1.2.5 Characteristics of GS enzymes

Three main classes of GS enzymes have been described in bacteria, and discussed in the review articles referred to at the beginning of this Section 1.2. GSI enzymes are the most typical type of bacterial GS, and have been extensively studied in the Enterobacteriaceae, but are also present in Gram-positive organisms *Bacillus, Clostridium,* and *Streptomyces* (Schreier, 1993). The GS enzyme is a dodecamer composed of 12 identical subunits, ranging between 44 and 59 kDa, and are organized in two superimposed hexagonal rings. GSII enzymes are typical of eukaryotes but are also found, together with GSI, in many symbiotic nitrogen fixing bacteria including *Agrobacterium*, *Rhizobium, Bradyrhizobium* and *Frankia* species (Woods and Reid, 1993), and also evidently in the free-living *Streptomyces,* including *S. coelicolor* (Behrmann *et al.*, 1990; Kumada *et al.*, 1990). The GSII enzyme is characteristically a thermolabile octomer of identical subunits arranged in two discs of four subunits (Carlson and Chelm, 1986). The smaller GSII subunit (~36 kDa) lacks the C-terminal portion of the GSI subunit which includes the adenylylation site involved in post-translational control of activity. Although the overall amino acid similarity between the GSI and GSII subunits is low (~15% identity), five regions associated with the GS active sites remain well conserved (Rawlings *et al.*, 1987; Janssen *et al.*, 1988). GSIII was first identified in the anaerobic rumen bacterium *Bacteroides fragilis* (Hill *et*

al., 1989; Southern *et al*., 1989). It is a hexamer of six large identical subunits (75 kDa) and is quite distinct from either GSI or GSII at both the amino acid and enzyme levels (Woods and Reid, 1993). Homologous genes have since been identified in other bacteria as well (Goodman and Woods, 1993; Reyes and Florencio, 1994). Additional GS isoenzymes have also been identified in species of *Rhizobium* having a much lower biosynthetic activity (Merrick and Edwards, 1995).

Very different regulatory mechanisms have evolved to control GS activities, both within and between each class. In most cases the structural genes are regulated in response to nitrogen, however, GSI type enzymes can be divided into two groups depending on whether they are regulated post-translationally by covalent modification, resulting in a reduction of GS activity. The subdivision is consistent with a phylogenetic analysis of GSI sequences (Tiboni *et al*., 1993) which revealed two main groups: one comprised of the cyanobacteria, proteobacteria and Gram-positive *S. coelicolor*, all of which are regulated by reversible subunit adenylylation, and the other group comprised of the GSI enzymes of archaea, and the Gram-positive *Clostridia* and *Bacillus* spp, none of which have been shown to be regulated by subunit adenylylation. The GSI and GSII activities are differentially regulated in some nitrogen fixing bacteria in response to nitrogen conditions (Woods and Reid, 1993). For example, in members of the family Rhizobiaceae, *glnII* (encoding GSII) is preferentially expressed during nitrogen-limited growth and nitrogen-fixation, while *glnA* is expressed constitutively and the corresponding GSI activity is controlled post-translationally (Carlson and Chelm, 1986). However, mutant strains lacking GSII activity exhibit no obvious phenotype and still successfully associate with plants (de Bruijn *et al*., 1989; Somerville *et al*., 1989). Thus the physiological and biochemical significance of these GS isoforms, or the existence of multiple forms within the same organism remains unclear.

1.3 Global nitrogen regulation in the enteric bacteria

1.3.1 The Ntr system

In the enteric bacteria, the transcriptional and post-translational regulation of GS activity is intricately controlled by a global regulatory network (the Ntr system) that regulates nitrogen metabolism in these organisms. Since the Ntr system serves as a well researched model

(detailed in the reviews cited) against which other bacterial systems can be assessed, the distinguishing features will be outlined.

The Ntr system consists of a complex regulon with its own sigma factor, sigma 54 (σ^{54}), also designated by σ^N, and a series of proteins able to recognize the nitrogen status of the cell. Its primary function is to regulate ammonia assimilation from alternate sources of nitrogen available in the environment when ammonia becomes growth rate limiting. Consequently, the Ntr regulatory system induces the expression of GS, and coordinates the expression of a number of other nitrogen utilization systems that generate ammonia. These include gene products required for the transport and catabolism of nitrogenous compounds such as the amino acids arginine, histidine, proline, tryptophane, asparagine and glutamine, the transport of ammonia, the degradation of urea by urease, the uptake and utilization of nitrate and nitrite by the assimilatory reductases (reviewed by Lin and Stewart, 1998), and proteins required for the regulation of nitrogen-fixation. In addition, the Ntr response activates genes that code for secondary transcriptional activators.

The activity of the Ntr response is controlled by a regulatory cascade (see Fig. 1.3), that senses and responds to the availability of fixed nitrogen and carbon in the cell, as reflected in the relative pool sizes of glutamine and 2-oxoglutarate. The intracellular nitrogen status is sensed and transduced by a pair of proteins UTase/UR and P_{II} such that when cells are nitrogen limited (low glutamine levels), UTase covalently modifies P_{II} by the addition of a UMP group on each subunit, but when nitrogen is not limiting, glutamine stimulates the UR activity to remove UMP from P_{II}-UMP, and prevents uridylylation of P_{II}. The overall uridylylation state of P_{II} in turn controls the activity of the primary Ntr regulators, NtrC and NtrB. They are members of the two component family of bacterial regulatory proteins in which NtrB is typical of sensory histidine kinases, and NtrC is typical of phosphorylatable response regulators. In the nonuridylylated form (indicative of nitrogen excess conditions), P_{II} interacts with NtrB to stimulate the dephosphorylation and consequent inactivation of NtrC. In the uridylylated state however, P_{II} does not interact with NtrB which, left to itself, is then able to catalyze the phosphorylation and hence activation of NtrC. The phosphorylated form of NtrC is required to activate transcription of the *glnA* gene and other Ntr responsive operons, through σ^{54} which directs RNA polymerase to –12/-24 promoters, and by binding to upstream enhancer-like sequences.

In addition, the catalytic activity of GS is regulated post-translationally by a reversible adenylylation/deadenylylation reaction, which is controlled primarily by the same regulators that control the phosphorylation of NtrC, UTase and PII (Fig. 1.3). In the uridylylated form (signaling nitrogen limitation), PII interacts with ATase to promote the deadenylylation and consequent activation of GS. In the unmodified form, PII stimulates the adenylylation of GS by ATase, resulting in its inactivation. In addition, adenylylation enhances the sensitivity of the GS enzyme to feedback inhibition by various end products of glutamine metabolism (Reitzer, 1996). Such a reversible covalent modification system enables cells to adapt much faster to changes in nitrogen availability than adjustments that may result from altered levels of gene expression, and it would protect them from metabolic imbalances that may occur during sudden changes in the availability of ammonia in the environment (Kustu *et al.*, 1984). Thus, PII is a pivotal protein in controlling the appropriate responses required under conditions of nitrogen excess (i.e. inactivation of GS and inhibition of *glnA* transcription), and nitrogen limiting conditions (i.e. activation of GS and transcription of *glnA*).

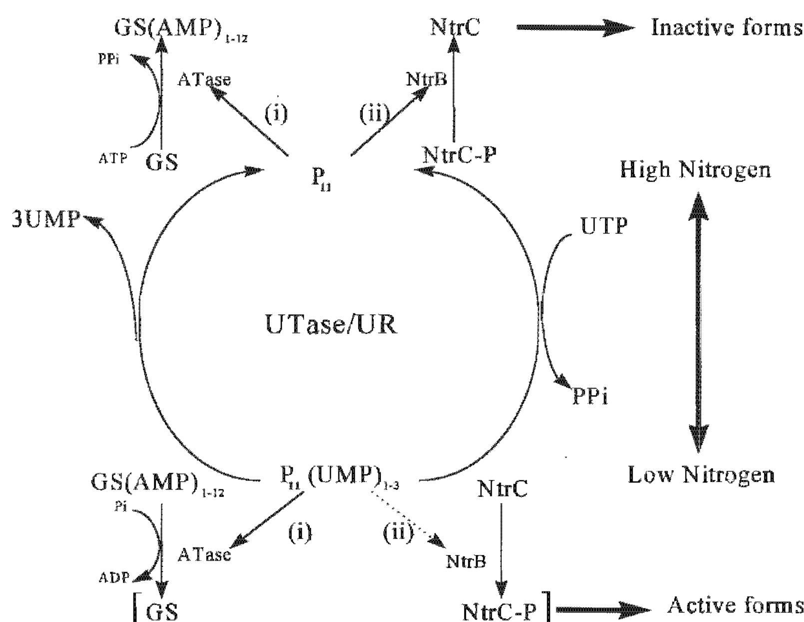


Fig. 1.3. Model illustrating regulation of GSI and NtrC in the enteric bacteria in response to nitrogen availability, as described in the text. Adapted from Merrick and Edwards (1995).

1.3.2 Expression of the *glnAntrBC* operon in *Escherichia coli* (and related bacteria)

As mentioned above, the Ntr system controls the increased expression of GS in enteric bacteria when nitrogen becomes limiting (Woods and Reid, 1993; Merrick and Edwards, 1995)

(Reitzer, 1996). In *E. coli,* the *glnA* gene, which forms an operon with the downstream genes encoding the Ntr regulators NtrB (*ntrB*) and NtrC (*ntrC*), can be expressed from either one of two tandem promoters, *glnA*p1 or *glnA*p2, depending on the nitrogen conditions (Fig. 1.4). Under nitrogen sufficient conditions, two minor consensus σ^{70}-dependent promoters, *glnA*p1 and *ntrB*p, ensure basal levels of the three products of the *glnAntrBC* operon. Transcription from *glnA*p1 is activated by a cyclic AMP-cyclic AMP receptor protein complex formed upstream of the promoter, and under these conditions most of the transcripts terminate at a Rho-independent terminator downstream of *glnA*. However, under nitrogen limiting conditions the enhancer protein NtrC, binds to two sites overlapping the –35 region and the transcriptional start site of *glnA*p1 respectively, repressing expression from this promoter, while activating transcription from the stronger *glnA*p2 promoter. Expression from *glnA*p2 is considerably elevated (approximately 14 fold) and is dependent on the phosphorylated form of NtrC, and RNA polymerase complexed with σ^{54}. Under these conditions the downstream *ntrBC* genes are expressed by read through resulting in a high level of NtrB and NtrC necessary to stimulate the transcription of other Ntr genes.

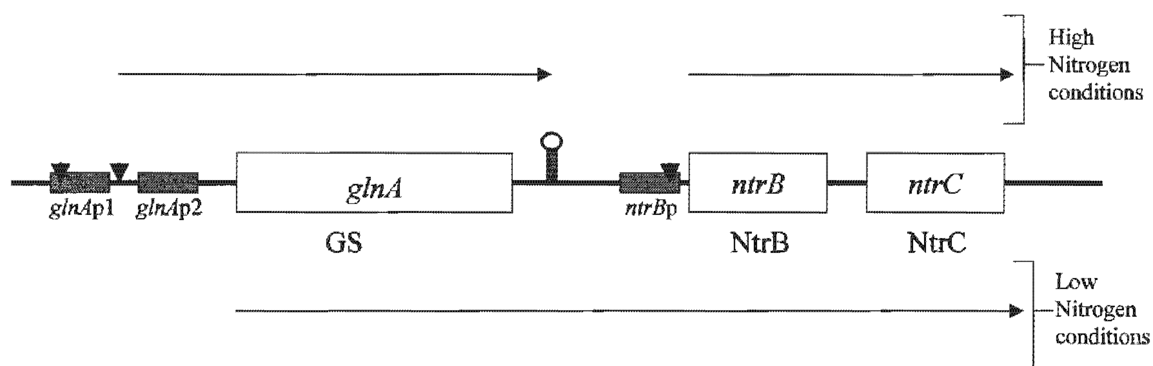


Fig. 1.4 Transcription of the *E. coli glnAntrBC* operon under high and low nitrogen conditions, as described in the text. Promoters are indicated by the hatched regions. NtrC binding sites are indicated by triangles.

1.3.3 The *nac* (nitrogen assimilation control) gene

Not all nitrogen-regulated genes are controlled directly by NtrC. In *K. aerogenes* a subset of genes in the Ntr regulon are under the control of a secondary transcriptional regulator, Nac, whose expression is dependent on the σ^{54}-RNA polymerase and NtrC (Macaluso *et al.*, 1990; Schwacha and Bender, 1993a; Schwacha and Bender, 1993b; Goss and Bender, 1995). Thus, under conditions of nitrogen limitation, Nac activates the expression of operons necessary for the utilization of histidine (*hut*), proline (*put*), and urea (*pur*), and represses the synthesis of

GDH (*gdh*) and GOGAT (*gltBDF*). Although *nac* is also present in *E. coli* and *K. pneumoniae* (Bender, 1991), *nac* regulation appears to be less significant in *E. coli*. The significance of Nac appears to be in coupling the expression of a number of otherwise σ^{70} dependent genes, to the σ^{54}-dependent *ntr* regulators (Bender, 1991).

Thus, Ntr control provides a highly flexible and coordinated response to nitrogen availability. Furthermore, the expression of various Ntr genes are modulated by additional factors as well, including the level of activator needed, the degree of phosphorylation and negative autogenous regulation of Nac and NtrC (Reitzer, 1996).

There is also growing evidence that the *ntr* system is more widely spread than just in the enteric bacteria. Regulatory circuits resembling the Ntr system appear to function in other Gram-negative bacteria (Merrick and Edwards, 1995), and a number of *ntr* homologues have been identified in many other bacterial genera. Of particular note is the ubiquity of the *glnB*-like genes encoding PII homologs, suggesting that this component of nitrogen regulation is highly conserved. In fact a second functional homologue has been identified in *E. coli* (encoded by *glnK*), which is itself regulated by the Ntr system (van Heeswijk *et al*., 1995; van Heeswijk *et al*., 1996), suggesting that nitrogen regulation in the enteric bacteria is even more sophisticated than originally thought.

1.4 Nitrogen control in Gram-positive bacteria

The regulation of nitrogen metabolism in Gram-positive bacteria has not been as well characterized as in the enteric bacteria. Evidence for a global regulatory network equivalent to the Ntr system is lacking, although as discussed below, some homologues appear to exist in *Bacillus subtilis*. There is also no indication of an alternative global regulatory system that characterizes this group, although there is still not much information available on nitrogen control of enzymes involved in the utilization of nitrogen sources. As discussed, the mechanism of GS modification by adenylylation plays a key role in nitrogen control in enteric bacteria, but apart from in *Streptomyces*, this does not occur in Gram-positive organisms, nor is there apparently any other post-translational control mechanisms.

1.4.1 Nitrogen control and the regulation of GS activity in *Bacillus subtilis*

In *B. subtilis,* amino acids (glutamine, arginine and possibly others) rather than ammonia are the preferred nitrogen source in that they support the fastest growth rate (Atkinson and Fisher, 1991). While the synthesis of amino acid degradative enzymes is substrate inducible, with the exception of asparaginase, their expression is generally not regulated in response to nitrogen availability (Atkinson and Fisher, 1991; Fisher, 1993; Schreier, 1993). However, nitrogen limitation stimulates the expression of other enzymes involved in nitrogen metabolism including urease, asparaginase, GSI and enzymes involved in nitrate assimilation (Atkinson and Fisher, 1991; Nakano *et al*., 1995). It was recognized that the GS protein played an important role in transmitting information about nitrogen availability to these enzyme systems, as well as in controlling its own synthesis (Schreier *et al.*, 1985; Schreier and Sonenshein, 1986). Only recently has a global system of nitrogen control begun to emerge in *B. subtilis,* which was the subject of a review by Fisher (1999). It involves at least three independent global regulatory proteins; GlnR, a member of the dicistronic *glnRA* operon (Fig. 1.5), TnrA and CodY. These regulators function under different nutritional conditions, and their activities support a system of control in which they direct the cell towards adaptive vegetative growth rather than towards sporulation, when nitrogen becomes limiting (Fisher, 1999).

In response to nitrogen limitation, TnrA activates the expression of asparaginase, γ-aminobutyric acid permease (*gabP*), urease (*ureABC*), the *nrgAB* operon encoding a putative ammonia permease (NrgA), a homologue of the *E. coli* PII protein (NrgB), and the nitrate assimilatory enzymes (*nasABCDEF*) (Wray, Jr. *et al*., 1996; Fisher, 1999). In addition, TnrA positively regulates its own synthesis, and represses the expression of *glnRA* and the genes encoding GOGAT (Wray, Jr. *et al.*, 1996; Fisher, 1999). In contrast, GlnR represses the expression of *glnRA, tnrA* and *ureABC* in cells growing with excess (preferred) nitrogen sources (Wray, Jr. *et al.*, 1997; Fisher, 1999). They are related proteins belonging to the MerR family of DNA-binding regulatory proteins. Although they differ in their C-terminal signal-transducing domains, their proposed N-terminal DNA-binding domains are almost identical. Both proteins bind to similar DNA sequences (TGTNAN$_7$TNACA), the GlnA/TnrA site (Fisher, 1999), under the different nutritional conditions, presumably accounting for the cross regulation observed in some instances. The fact that not all promoters are cross regulated by both proteins suggests that additional regulatory factors must be required. Since all the promoters positively regulated by TnrA contain a GlnR/TnrA site upstream of their –35 region,

it was suggested that TnrA may activate expression by facilitating the binding of RNA polymerase (Wray, Jr. *et al.*, 1997). Thus, GlnR and TnrA not only regulate their own synthesis, but also cross regulate each other's expression, and both negatively regulate GS activity (Schreier *et al.*, 1989; Wray, Jr. *et al.*, 1996).

The regulation of GS activity in *B. subtilis* has been reviewed by Schreier (1993) and Woods and Reid (1993). The activity of the GS enzyme is controlled primarily at the level of transcription in response to nitrogen availability (Fisher *et al.*, 1984; Schreier and Sonenshein, 1986; Strauch *et al.*, 1988; Schreier *et al*, 1989). The regulation of GS activity is schematically illustrated in Fig. 1.5. The *glnRA* operon is transcribed by the vegetative (σA-dependent) form of RNA polymerase (Schreier, 1993). During growth with excess nitrogen, the *glnR* product, GlnR, a small (135-amino acid) dimeric protein represses expression of the *glnRA* operon (Strauch *et al.*, 1988; Schreier *et al.*, 1989). Regulation of transcription involves the co-operative binding of GlnR dimers to each of two operators, *glnRA*o2 and *glnRA*o1 containing the GlnA/TnrA site (Gutowski and Schreier, 1992; Brown and Sonenshein, 1996). The upstream site, *glnRA*o1, is a symmetrical inverted repeat sequence whereas the *glnRA*o2 operator is only partially symmetrical. TnrA negatively regulates *glnRA* expression by binding to the *glnRA*o2 operator, which overlaps the –35 region of the promoter (Fisher, 1999). The binding of GlnR to the operators requires GS but the exact nature of the interaction between these two proteins is not understood. Mutational analysis indicated that the last seven residues of GlnR are part of a domain involved in sensing nitrogen conditions (Schreier, 1993) since removal of these sequences prevented derepression of GS expression under nitrogen limiting conditions.

The mechanisms by which nitrogen levels are sensed in *B. subtilis* is not known, but it does not appear to depend on the intracellular concentrations of glutamine (and 2-oxoglutarate) as in the enteric bacteria (Fisher, 1999; Hu *et al.*, 1999). Whatever the mechanism is, the GS protein is required for the transduction of the nitrogen signals to the GlnR and TnrA proteins (Schreier *et al.*, 1985; Schreier *et al.*, 1989; Wray, Jr. *et al.*, 1996; Fisher, 1999). Although the derepression of *glnA* under nitrogen limiting conditions may involve additional proteins (Gutowski and Schreier, 1992; Schreier, 1993), based on the patterns of GlnR and TnrA regulated gene expression, and the phenotypes of mutants altered in these regulators, Fisher (1999) proposes that, during nitrogen excess conditions, the nitrogen signal activates GlnR and inhibits TnrA activity. Thus, TnrA is only active during nitrogen-limited growth in which the excess nitrogen

regulatory signal is not present. In addition, the GlnR and TnrA proteins appear to have different sensitivities to the GS dependent nitrogen-regulatory signal, suggesting a flexible and fine-tuned system.

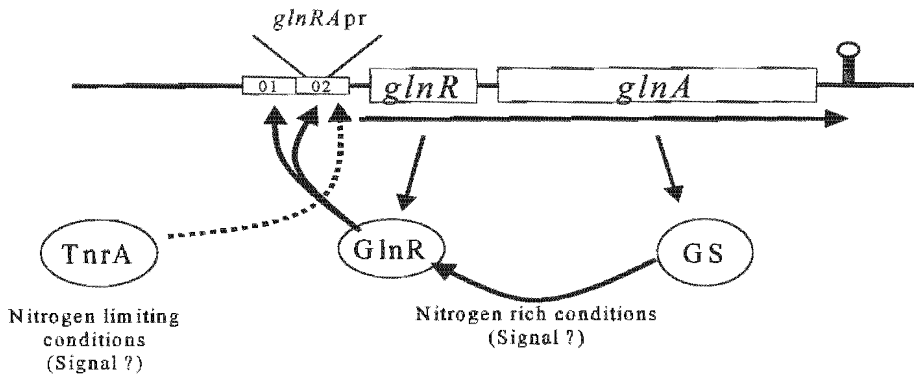


Fig. 1.5. Model for the control of the *B. subtilis glnRA* operon (adapted from Schreier, 1993). In response to an undefined signal of nitrogen excess conditions, the *glnR* gene product GlnR, in concert with the GS protein, binds to the two *glnRA* operators (*glnRA*o1 and *glnRA*o2) thereby repressing transcription from the *glnRA* promoter, *glnRA*pr. The operator sequences are present from –40 to –60 and –17 to –37 relative to the *glnR* start codon, respectively. Nitrogen limiting conditions activate the global regulator TnrA, which represses expression of the *glnRA* operon by binding to the *glnRA*o2 site which overlaps the –35 region of the promoter.

CodY-dependent regulation responds to the total nutritional status of the cell. It represses the expression of several genes involved in nitrogen metabolism, including the histidine degradative operon (*hut*), the dipeptide transport operon *dpp*, the isoleucine and valine degradative operon (*bkd*), *ureABC* and *gabP*, as well as genes involved in competence and acetate metabolism (Fisher *et al.*, 1996; Fisher, 1999). The highest levels of CodY-dependent repression occurred in cells growing rapidly in a medium rich in amino acids. While the signal regulating CodY activity is unknown, it is thought that the mechanism of transcriptional regulation involves binding of CodY to a structure formed by AT-rich DNA sequences (Wray, Jr. *et al.*, 1996). When carbon or nitrogen become limiting, CodY-dependent regulation is relieved. As pointed out by Fisher (1999), regulation of the utilization of carbon and nitrogen sources in response to growth rate helps ensure that nutrients are still available for adaptation to non-optimal growth conditions. This pattern of regulation was consistent with the observation that a hierarchy of amino acid utilization occurs in *B. subtilis* with histidine, isoleucine, threonine and valine not being used significantly until the onset of stationary phase (Liebs *et al.*, 1988).

The use of multiple systems and components to regulate gene products involved in nitrogen metabolism in response to different nutritional conditions, allows for non-coordinate gene expression, which may give the organism an adaptive advantage under diverse nutritional conditions. In addition it has been suggested by Fisher (1999) that, since the catabolism of amino acids plays an important role in the developmental life cycle of *B. subtilis*, providing the cell with energy during sporulation and germination, the regulation of amino acid degradative enzymes may have evolved to be present at high levels in sporulating cells and spores, rather than being preferentially expressed during nitrogen-limited growth as in the enteric bacteria. Although the genetic organization and GlnR-dependent regulation of the *B. subtilis glnRA* operon is preserved in *B. cereus* and *Staphylococcus aureus* (Nakano and Kimura, 1991; Gustafson *et al.*, 1994), as yet there is no evidence that a nitrogen-regulatory system analogous to that of *B. subtilis* is present in any other bacteria.

1.4.2 Nitrogen control and the regulation of GS activity in *Streptomyces* spp.

In contrast to other Gram-positive prokaryotes, *Streptomyces* species synthesize both GSI and GSII enzymes (Section 1.2.5 above), and the GSI activity, encoded by *glnA*, is regulated at both the transcriptional and post-translational levels in response to nitrogen availability (Paress and Streicher, 1985; Brana *et al*, 1986; Bascaran *et al.*, 1989a; Fisher and Wray, Jr., 1989; Hillemann *et al.*, 1993). An exception to this observation is *S. venezuelae* in which GSI is constitutively produced (Shapiro and Vining, 1983). At present nothing is known about the regulation of the *glnII* genes encoding GSII subunits in *Streptomyces* spp., and the respective roles of the two *gln* genes remains unclear. Neither *glnA* nor *glnII* single mutants appear to require glutamine for growth, and in *S. viridochromogenes*, no differential regulation of the expression of GSI and GSII in response to different nitrogen sources was observed, although GSI activity was always dominant (Hillemann *et al.*, 1993). This is very different to the differential expression of the two GS isoforms seen in the nitrogen fixing bacteria in which GSII is only expressed under conditions of nitrogen starvation (Edmands *et al.*, 1987).

In contrast to the *E. coli glnAntrBC* and *B. subtilis glnRA* operons, the *S. coelicolor glnA* transcription unit is monocistronic, and does not contain any potential regulatory genes (Fisher and Wray, Jr., 1989). Although the mechanisms regulating the expression of GSI in *Streptomyces* in response to nitrogen availability are not known, transcription of *glnA* in *S. coelicolor* requires a positive regulatory gene, *glnR,* and in view of the Gln- phenotype associated with *glnR* mutants, may also be required for transcription of *glnII* (Wray, Jr. *et al.*,

1991). The *glnR* gene encodes a 29-kDa polypeptide whose deduced amino acid sequence showed significant homology to other response regulator proteins that are known to act as transcriptional activators (Wray, Jr. and Fisher, 1993).

As indicated above, GSI modification by adenylylation in response to ammonia rich conditions, occurs in several *Streptomyces* species (Fisher and Wray, Jr., 1989). Recently, a *glnE* homologue encoding a putative adenylyltransferase, was characterized from *S. coelicolor* (Fink *et al.*, 1999). The *glnE,* and a *glnB* (P_{II}) homologue, were localized between the *glnA* and *glnII* genes (Fink *et al.*, 1999). This is different from the situation in *E. coli* (van Heeswijk *et al.*, 1993) where *glnE* is separate from the GS gene, although in *M. tuberculosis glnA*, *glnE* and *glnII* are adjacent (Fink *et al.*, 1999). A *glnE* mutant lost the ability for covalent modification of GSI by adenylylation, demonstrating for the first time that in *S. coelicolor* a GlnE-mediated GSI modification takes place. Furthermore, Southern blot hybridizations suggest that *glnE* is ubiquitous in *Streptomyces*. This is the first example of an Ntr-like component in the Gram-positive *Streptomyces*. Although it suggests a GS regulation system similar to that of *Enterobacteriaciae*, the pattern is unlike the enteric system since GS adenylylation is not coupled to GS synthesis in *S. coelicolor* (Fisher and Wray, Jr., 1989).

Investigation of several enzymes involved in assimilation of different nitrogen compounds in *S. clavuligerus* indicated that, although there did not appear to be uniform regulation of catabolic enzymes in this organism, at least four enzyme activities involved in nitrogen nutrition are controlled according to the nitrogen source (Bascaran *et al.*, 1989b). GS, urease, ornithine aminotransferase and arginase were subject to ammonia repression and induced by poor nitrogen sources, whereas enzymes involved in the catabolism of threonine, serine, proline and histidine were not (Bascaran *et al.*, 1989a; Bascaran *et al.*, 1989b). The characterization of *S. clavuligerus* mutants simultaneously deregulated for GS, urease and arginase activities, suggested that a system of nitrogen control may exist in the *Streptomyces* involved in the preferential utilization of some nitrogen sources, and suggested that GS played a key role. The lack of histidase repression in *S. clavuligerus* is also apparent in *S. coelicolor* (Kendrick and Wheelis, 1982) and *S. griseus* (Kroening and Kendrick, 1987), suggesting that nitrogen control is not involved in histidine utilization in the *Streptomyces.* In this regulatory aspect, *Streptomyces* differ from other histidine-utilizing bacteria.

1.4.3 Structure and regulation of *C. acetobutylicum glnA* region

The *C. acetobutylicum glnA* gene region was cloned by complementation of an *E. coli glnAntrBC* mutant, and the enzyme activity was not regulated by adenylylation (Usden *et al.*, 1986). Sequence analysis revealed that the structural gene encoded a GS protein of 444 amino acids, which was expressed from either of two promoter sequences, P1 and P2 (at positions –124 to –95 and –65 to –37) respectively upstream from the *glnA* start codon, which are regulated by nitrogen (Janssen *et al.*, 1988; Janssen *et al.*, 1990). In addition, an extensive inverted repeat sequence (158 nucleotides) was located immediately downstream of *glnA* suggesting a terminator-like element. Four regions exhibiting dyad symmetry are located upstream of *glnA*, three upstream of P1, and one overlapping P2 (Woods and Reid, 1995). Some of these features are summarized in Fig. 1.7.

Evidence presented by Janssen *et al.* (1990) and Fierro-Monti *et al.* (1992) strongly suggests the involvement of an antisense mRNA (AS-RNA) in the down-regulation of *glnA* gene expression. A functional promoter, P3, situated 227 to 257 nucleotides from the 3' end of the *glnA* gene, and oriented towards the *glnA* gene (Fig.1.7), has been shown to direct the synthesis of a short mRNA transcript in *C. acetobutylicum,* which is complementary to a 43 base region at the start of the *glnA* mRNA, including the ribosome binding site and the *glnA* initiation codon. An up-promoter mutation in P3 resulted in reduced levels of GS activity when the *glnA* region was expressed in *E. coli* (Janssen *et al.*, 1990). In addition, the production of this AS-RNA was differentially regulated by the nitrogen source, such that under nitrogen limiting conditions, which result in elevated GS activity, the expression of the AS-RNA was repressed while the expression of the *glnA* mRNA was induced; levels of *glnA* mRNA exceeded that of the AS-RNA by five fold. Under nitrogen rich conditions that repress GS activity, the situation was reversed; there was ~1.6-fold more AS-RNA transcripts over *glnA* mRNA transcripts. In addition the *glnA* mRNA was regulated in the same way as GS activity in the nitrogen limiting and nitrogen rich media, however, the relative levels of GS activity varied more than that of the corresponding mRNA in the different media, compatible with a post-transcriptional regulatory mechanism. Thus it was proposed that the AS-RNA hybridized to the sense mRNA, thereby inhibiting translation of nascent GS during nitrogen rich conditions (Fierro-Monti *et al.*, 1992).

Although all three promoters are typical of Gram-positive extended –10 and –35 promoter consensus sequences, their differential regulation by the nitrogen source suggested the involvement of additional regulatory elements. A more recent analysis of the region

downstream of *glnA* revealed an ORF of 566 nucleotides, present in the same orientation as the *glnA* gene (Fig. 1.7) (Woods and Reid, 1995). A corresponding polypeptide of the predicted size (21.4 kDa) was detected by *in-vitro* transcription-translation assays (Dr. L. Brown, personal communication). An intriguing feature was that the 5' start region of the ORF overlapped with the AS-RNA coding region, such that the *glnA* AS-RNA was complementary to the ribosome binding site and the first eight codons of the ORF mRNA. This implied that expression of the ORF must interfere with the expression of the AS-RNA and hence that of the *glnA* gene.

Based on amino acid sequence homologies, Woods and Reid (1995) proposed that the ORF represented a response regulator protein. The deduced amino terminal domain showed sequence homology to the conserved amino terminal domains of known response regulatory proteins which are members of the family of two-component signal transduction systems. These proteins function in pairs to regulate transcription of genes in response to environmental changes such as nutrient limitations. The predicted carboxy terminus shared similarity with the carboxy terminal domain of the aliphatic amidase regulator, AmiR, of *Pseudomonas aeruginosa* (18 identical residues over a region of 60 residues; Fig. 1.6)(Lowe *et al.*, 1989). AmiR positively controls the expression of the amidase encoded *amiE* gene via an anti-termination mechanism by allowing RNA polymerase to read through the rho-independent terminator present in the region between the *amiE* promoter and the *amiE* structural gene (Wilson *et al.*, 1993; O'Hara *et al.*, 1999). The AmiR protein is an unusual member of the response regulator family since its antitermination activity is controlled by sequestration rather than by phosphorylation. Under repressing growth conditions, the activity of AmiR is silenced by the formation of a complex with the ligand-sensitive negative regulator AmiC. However, the presence of small inducer molecules results in ligand-induced release of the RNA binding positive regulator AmiR, and the expression of the full-length *amiE* transcripts (O'Hara *et al.*, 1999).

These homologies are supported by a more recent search of the data bases (Fig. 1.6). The ORF shared significant homology over its entire length with two probable response regulators, one from *S. coelicor* and the other from *M. tuberculosis*, as well as to the response regulator protein NasT from *Azotobacter vinelandii* (27% identity). NasT is required for the expression of the assimilatory nitrite-nitrate reductase operon (*nasAB*) (Gutierrez *et al.*, 1995). It forms an operon with the upstream negative regulator gene *nasS*. In addition, the carboxy terminal region of the

ORF showed significant homology to the carboxy terminal region of NasR (28 identical residues over a region of 70 residues), the nitrite-nitrate responsive positive regulator of the *nasFEDCBA* operon encoding the assimilatory nitrite and nitrate reductases of *Klebsiella oxytoca* (Goldman *et al.*, 1994). NasR, which is encoded immediately upstream of the *nasF* operon, operates independently of Ntr control. Interestingly, the expression of the *nasF* operon in *K. oxytoca* (M5al) is also controlled by a transcriptional antitermination mechanism, in which it is hypothesized that, in response to nitrate or nitrite, the NasR protein binds to a stem loop structure in the *nasF* leader region, thereby mediating transcriptional readthrough of an adjacent downstream factor-independent terminator (Chai and Stewart, 1998; Lin and Stewart, 1998; Chai and Stewart, 1999). Several distinct RNA binding protein-dependent transcriptional antitermination mechanisms have been described in various bacteria (Babitzke, 1997; Henkin, 1996; Switzer *et al.*, 1999; Weisberg and Gottesman, 1999).

Thus, based on these sequence data, and its proximity to the *glnA* gene, it seems likely that the ORF represents a response regulator protein in *C. acetobutylicum* that may positively control *glnA* expression via an antitermination mechanism. This is consistent with the presence of a sequence of dyad symmetry in the region of P2, between promoter P1 and the *glnA* structural gene with the potential to form a stem-loop structure (ΔG = -11 kcal/mol) and act as a factor-independent transcription terminator (Woods and Reid, 1995). It may also be significant that a second inverted repeat sequence, part of the complex palindromic sequence previously identified by (Janssen *et al.*, 1990), lies immediately upstream of P1. In addition, deletion of the region downstream of *glnA* resulted in very low levels of GS activity in *E. coli* (Janssen *et al.*, 1990). However, the way that the ORF product may sense nitrogen conditions is unclear, since it does not contain certain highly conserved residues (Fig. 1.6), including the aspartate residue which is the site of phosphorylation characteristic of this family of regulators (Stock *et al.*, 1989). Although the proposed function of the ORF has not been confirmed, we have designated it *glnR* for the purpose of this thesis.

```
                                          *
ScoelReg?   MTAPESPQPVDVPDDDQSHVPPLTTRVVIAEDEALIRLDLKEMLEEEGYSVVGEAGDGEE    60
MtubeReg?   ...........MTGPTTDADAAVPRRVLIAEDEALIRMDLAEMLREEGYEIVGEAGDGQE    49
AvineNasT   .......................MLRILLINDTPKKVGRLKSALVEAGFEVVDESGLTID    37
CacetGlnR   ..................MVQSEKGKLIIALSNVEIAKKLKTSLTQEGFDIIALCTSGNE    42
KpneuNasR   ............................................................     0
PaeruAmiR   .............  .............................................     0
                         *
ScoelReg?   AVELAREHRPDLVILDVKMPKMDGISAAEKIAEESIAPVLMLTAFSQRDLVERARDAGAM   120
MtubeReg?   AVELAELHKPDLVIMDVKMPRRDGIDAASEIASKRIAPIVVLTAFSQRDLVERARDAGAM   109
AvineNasT   LPVRVEAVRPDVILIDTESPGRDVMEQVVLVSRDRPRPIVMFTDEHDPQVMRQAIQAGVS    97
CacetGlnR   LIRLVMQYSPDLVLVGYKFKDMSLLDVYENLVDLTSFLAIVNEPYRSFIEEDTDIYCIGT   102
KpneuNasR   ............................................................     0
PaeruAmiR   ............................................................     0
               ●
ScoelReg?   AYLVKPFSKSDVVPAIEMAVSRFTELKALEKEVAD[illegible]SLR[illegible]ET[illegible]LVDRAKSVLQTEY[illegible]LT   180
MtubeReg?   AYLVKPFSISDLIPAIELAVSRFREITALEGEVAT[illegible]SER[illegible]ET[illegible]LVERAKGLLQTKH[illegible]MT   169
AvineNasT   AYIVEGIQAQRLQPILDVAMARFESDQALRAQLQAREAQ[illegible]AE[illegible]RVELAKGLLMKMKNCS   157
CacetGlnR   KISNVLLTN.....AIDLIFQSKRRIKKLKEQVEK[illegible]EHT[illegible]ED[illegible]LIEKAKGQLMSTS[illegible]LT   157
KpneuNasR   ..............LLPLVRQQAHELQQLSGQLAS[illegible]KDA[illegible]EE[illegible]LIEKAKSVLMTYQ[illegible]MQ    46
PaeruAmiR   ..............VLVSARRISEEMAKLKQKTEQ[illegible]QDR[illegible]AGQARINQAKVLLMQRH[illegible]WD    46

ScoelReg?   EPAAFRWIQKTSWDRRMSMQQVAEAVIQDAEEKKASKG                         218
MtubeReg?   EPDAFKWIQRAAWDRRTTMKRVAEVVLETLGTPKDT                           205
AvineNasT   EEEAYTLMRRQAWSRQQKLIQVAEQVIAMHDMLGS                            192
CacetGlnR   ENEAFRYMQKISWDSGKRMKDIASLILSEIQ                                188
KpneuNasR   EEQAWQALRKMAWDKNQRMVEIARALLTVKALWRVTPK                          84
PaeruAmiR   EREAHQHLSREAWKRREPILKIAQELLGNEPSA                               79
```

Fig. 1.6 Multiple sequence alignment of deduced protein sequences homologous to the *C. acetobutylicum* putative regulator GlnR (CacetGlnR, GenBank accession no. AAB47033). Sequences included, together with the GenBank protein accession numbers (in brackets), are: ScoelReg? (T35758), probable response regulator from *Streptomyces coelicolor*; MtubeReg? (H70558), probable response regulator from *Mycobacterium tuberculosis*; AvineNasT (S52249), response regulator NasT from *Azotobacter vinelandii*; KpneuNasR (A55859), NasR regulator from *Klebsiella pneumoniae*; PaeruAmiR (S03884), amidase regulator AmiR from *Pseudomonas aeruginosa*. NasR and AmiR sequences were included from positions 309 and 118, respectively. Residues that are conserved in 100%, >75% and >50% of the sequences aligned, are indicated by dark, medium and light shading, respectively. The conserved Asp (*) and Lys (•) residues proposed to be involved in the phosphorylation and signaling activities of the two component family of regulators, are also indicated. The multiple sequence alignment was created using the DNAMAN software package.

A model was proposed by Woods and Reid (1995) for the regulation of the *C. acetobutylicum* *glnA* gene by the *glnR* gene product and AS-RNA (Fig. 1.7). It is envisaged that under limiting nitrogen conditions a signal transduction mechanism induces expression of *glnR*, which acts as a transcriptional antiterminator by interacting with the terminator-like structure in the upstream region of the *glnA* gene, thereby promoting transcription of *glnA* mRNA. In response to a rich nitrogen signal P3 is activated resulting in the production of *glnA* AS-RNA which would reduce the production of GS and the putative regulatory protein by interacting with the complementary 5' regions of the respective mRNAs, thereby inhibiting transcription.

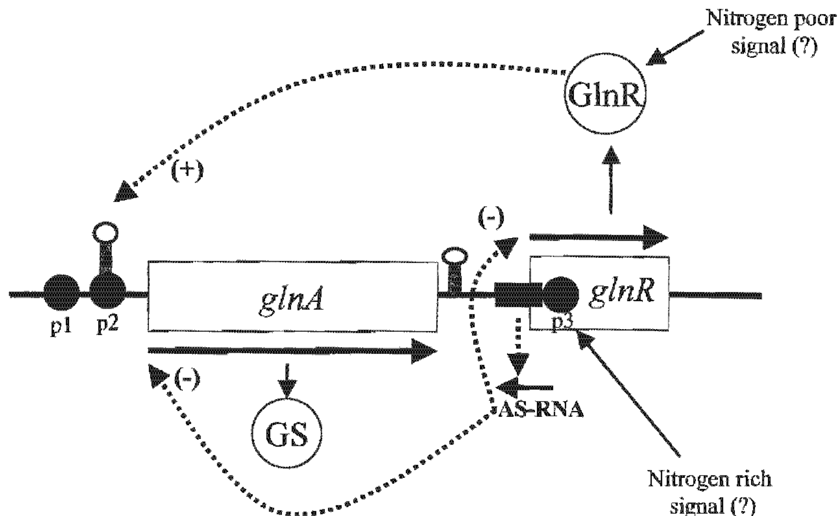


Fig. 1.7. Structure of the *C. acetobutylicum* P262 *glnA* gene region, and model for *glnA* regulation by the nitrogen content of the medium as discussed in the text. Arrows indicate direction of transcription and promoters are indicated by black dots. In response to nitrogen limiting conditions, an unknown signal activates the putative transcriptional anti-terminator GlnR, to bind to a region of dyad symmetry overlapping promoter p2, thereby activating transcription from the *glnA* promoter p1. In response to a rich nitrogen signal, transcription from the downstream promoter p3 is enhanced. It directs the synthesis of an antisense mRNA (AS-RNA) which binds to complementary sequences present in both the 5' regions of the *glnA* mRNA and *glnR* mRNA transcripts, thus reducing expression of the respective genes.

1.5 Glutamate synthase (GOGAT)

1.5.1 General introduction

Until the discovery of glutamate synthase (GOGAT) by Tempest *et al.* (1970), the only known route for the synthesis of glutamate from ammonia in bacteria was by the direct activity of GDH. GOGAT activity was first discovered in ammonia limited cultures (< 0.5 mM) of *Klebsiella aerogenes* (Meers *et al.*, 1970; Tempest *et al.*, 1970), in which GDH activity was barely detectable, while GS levels became notably elevated. Since then, the GOGAT pathway has been established in many groups of diverse bacteria, as well as in other microorganisms including yeast, fungi, algae, higher plants and recently in the silk worm. As outlined in Section 1.2.3, it forms the major pathway for ammonia assimilation in several microorganism and is characterized as catalyzing the synthesis of two molecules of glutamate from one molecule each of glutamine and 2-oxoglutarate during which one molecule of coenzyme (NAD(P)H) is oxidized. Thus, GOGAT catalyzes a reaction at the intersection of carbon and nitrogen metabolism. It is worth bearing in mind, that although this study focuses on the role of GS and

GOGAT in nitrogen metabolism, factors that affect the rate at which 2-oxoglutarate is synthesized would also be expected to affect the activity of the GOGAT pathway.

Numerous GOGAT structural genes have been cloned and sequenced, and a number of GOGAT enzymes have been purified to homogeneity from a variety of microorganisms including *E. coli* (Miller and Stadtman, 1972), *K. aerogenes* (Trotta *et al.*, 1974), *B. subtilis* (Matsuoka and Kimura, 1986), *B. megaterium* (Hemmila and Mantsala, 1978), *B. licheniformis* (Schreier and Bernlohr, 1984), *Azospirillum brasilense* (Ratti *et al.*, 1985), the anaerobic nitrogen fixing *C. pasteurianum* (Singhal *et al.*, 1989), *Nocardia medeterranei* (Mei and Jiao, 1988), *Saccharomyces cerevisiae* (Cogoni *et al.*, 1995), and *Medicago sativa* (Anderson *et al.*, 1989). Analyses of their combined structural, biochemical and sequence characteristics has lead to the identification of three classes of GOGAT.

1.5.2 Classes of GOGAT

The bacterial class of the enzyme is dependent on reduced pyridine nucleotides (NAD(P)H) for their reducing equivalents (NAD(P)H-GOGAT), and is composed of two dissimilar subunits, the large (α) subunit conserved around ~150 kDa, and the small (β) subunit conserved at ~50 kDa, that together form the active $\alpha\beta$ protomer (~200 kDa) (Vanoni and Curti, 1999). However, their oligomeric structures differ considerably. Some show a quaternary structure of the type $(\alpha\beta)_4$ as in *E. coli* and *A. brasilense,* while others are of the $\alpha\beta$ type and do not form an active higher MW species, thus accounting for the wide range of apparent molecular weights observed. The only controversial report is that of GOGAT purified from *C. pasteurianum* which apparently exists as a dimer of five distinctly different subunits (Singhal *et al.*, 1989). The second class of enzyme depends on reduced ferredoxin as its electron donor (Fd-GOGAT; EC 1.4.7.1), and are found in photosynthetic cyanobacteria (Marques *et al.*, 1992), and in the photosynthetic tissues of plants (Knaff *et al.*, 1991; Sakakibara *et al.*, 1991; Avila *et al.*, 1993). Molecular characterization of these enzymes has revealed that they appear to exist as homodimers composed of a single polypeptide chain varying in size from 125 to 180 kDa. The eukaryotic pyridine-dependent forms of the enzyme constitute the third class of GOGAT (NADH-GOGAT; EC 1.4.1.14), found in yeast, fungi and non photosynthetic tissues (nodules) of plants (Gregerson *et al.*, 1993). They are composed of a single high molecular weight polypeptide of approximately 200 kDa. *S. cerevisiae* GOGAT exists as a homotrimeric enzyme composed of three 199 kDa polypeptide monomers (Cogoni *et al.*, 1995), while in *Neurospora*

crassa and in plants the enzyme is composed of four identical monomers with an estimated molecular mass of ~220 kDa each (Hummelt and Mora, 1980; Anderson *et al.*, 1989).

In bacteria, the structural genes for the α and β subunit polypeptides are encoded by the *gltB* and *gltD* genes respectively, except in *B. subtilis* where the corresponding genes have rather been designated *gltA* and *gltB* (Bohannon *et al.*, 1985). In all cases so far the genes are genetically linked. While the α subunit is generally encoded upstream of the β subunit, in *A. brasilense* this gene order is uniquely reversed; the β subunit precedes the α subunit, and the two coding sequences are 141 nucleotides apart (Pelanda *et al.*, 1993). The gene encoding Fd-GOGAT was first cloned from maize (Sakakibara *et al.*, 1991), and later from cyanobacteria and plants (Temple *et al.*, 1998). These genes have been given various names including *gltF, glsF* and *gltS*, while the genes encoding the eukaryotic NADH-dependent enzyme have been referred to as *glsN*. Indeed, it would be beneficial to adopt a consistent nomenclature which would clearly distinguish between the different classes of enzyme, and avoid confusion.

Comparative analysis of the amino acid sequences of GOGAT from eukaryotes and bacteria has revealed highly conserved region, both within and between each class of enzyme (Sakakibara *et al.*, 1991; Gregerson *et al.*, 1993; Pelanda *et al.*, 1993; Valentin *et al.*, 1993). Specifically, the α subunit of bacterial GOGAT is very similar to the single polypeptide chain of Fd-dependent GOGAT, and to the amino terminal three quarters of eukaryotic NADH-dependent GOGAT, while the bacterial β subunit shares significant homology with the C-terminal one quarter of eukaryotic NADH-dependent GOGAT. The eukaryotic pyridine dependent form of the enzyme appears to derive from the fusion of polypeptides corresponding to the α and β subunits of bacterial GOGAT, linked by a short non-conserved polypeptide chain. The Fd-GOGATs lack a region corresponding to the bacterial β subunit.

1.5.3 The role of GOGAT in plants

The role of the pyridine-dependent plant enzyme seems to be similar to that of the bacterial enzyme being involved in the primary assimilation of ammonia derived from the action of symbiotic nitrogen fixing bacteria, and from the reduction of nitrates present in the soil (Temple *et al.*, 1998; Vanoni and Curti, 1999). In alfalfa nodules, a dramatic increase in NADH-GOGAT gene expression occurred with the onset of nitrogen fixation in the bacteroid (Gregerson *et al.*, 1993). It appears that the role of Fd-GOGAT, localized in chloroplasts in plants, is the secondary assimilation of ammonia derived either from the mobilization of

nitrogen storage compounds, or from photorespiration (Curti *et al.*, 1996). Although low constitutive levels of NADH-GOGAT have been detected in green tissues, Fd-GOGAT accounts for the major part of ammonia reassimilation in photosynthetic tissues, since Fd-*glt* mutants were also photorespiratory mutants. The control of GOGAT in plants is poorly understood, although Fd-GOGAT is induced by light.

1.5.4 Enzyme model

GOGAT is a complex iron-sulphur (Fe-S) flavoprotein, which was the subject of a recent review by Vanoni and Curti, (1999). This review was extensively consulted in preparing the following sections.

Early studies on the structural and biochemical properties of bacterial GOGATs purified from *E. coli* (Rendina and Orme-Johnson, 1978), *B. subtilis* (Matsuoka and Kimura, 1986) and *A. brasilense* (Vanoni *et al.*, 1991; Vanoni *et al.*, 1992) established that the $\alpha\beta$ protomer contains two flavin cofactors (one FAD and one FMN) and three distinct Fe-S centers: one $[3Fe\text{-}4S]^{0;1+}$ cluster (Center 1) and two $[4Fe\text{-}4S]^{1+;2+}$ clusters (Centers II and III). Based on kinetic and spectroscopic studies on these enzymes, a scheme for the enzyme reaction was proposed, which described three distinct catalytic sites carrying out partial reactions (Curti *et al.*, 1996).

Most of our biochemical knowledge of the enzyme comes from studies conducted on the enzyme from the Gram-negative nitrogen fixing bacterium *Azospirillum brasilense*. Recently, a series of very elegant studies on the kinetic, spectroscopic and biochemical properties of the *A. brasilense* holoenzyme, as well as the individual recombinant α and β subunits overproduced in *E. coli* (Vanoni *et al*, 1992; Pelanda *et al.*, 1993; Vanoni *et al.*, 1994a; Vanoni *et al.*, 1996; Vanoni *et al.*, 1998), have consolidated the mechanistic model, and established the localization and properties of the functional sites (Fig. 1.8). During catalysis, NAD(P)H binds to the first active subsite, Site 1 of the enzyme, which lies in the β subunit, and reduces Flavin 1, identified as FAD, located at this site. The electrons flow through a postulated intramolecular electron transfer chain, formed by some or all of the Fe-S centers, leading to reduction of the second enzyme flavin, identified as FMN. This is located in the α subunit at functional Site 2 of the enzyme. Here, the reduced FMN cofactor is in turn responsible for the reduction of an iminoglutarate intermediate, formed from the addition of the amide group of L-glutamine to the 2(C) carbon of 2-oxoglutarate. This site thus catalyzes a GDH-like reaction in which ammonia is derived from glutamine and reduced FMN is the electron donor. Steady state kinetic analysis

of the GOGAT reaction demonstrated that glutamine binding and hydrolysis takes place at another distinct site, Site 3, following 2-oxoglutarate binding to the reduced enzyme at Site 2. Essentially, a glutaminase reaction must take place at Site 3, prior to the addition of ammonia to 2-oxoglutarate to yield the postulated 2-iminoglutarate intermediate. Indeed, mutants have been described that exhibit a glutaminase activity uncoupled from reductive amination (Donald *et al.*, 1988), and glutaminase activity was found associated with the purified α subunit (Vanoni *et al.*, 1998). Thus, it was concluded that the α subunit contains the site of glutamine binding and hydrolysis, as well as the site where the actual glutamate synthesis takes place, and the β subunit acts as a FAD dependent NAD(P)H oxidoreductase, which serves to input electrons into the α subunit for reductive glutamate synthesis (Vanoni and Curti,, 1999), resulting in a tight coupling of glutaminase and glutamate dehydrogenase activities.

This model is consistent with the finding that FAD is absent from Fd-dependent GOGAT, and that they contained only one type of flavin, identified as FMN, and one type of Fe-S cluster, a $[3Fe\text{-}4S]^{0,+1}$ center (Knaff *et al.*, 1991; Marques *et al.*, 1992; Hirasawa *et al.*, 1996)

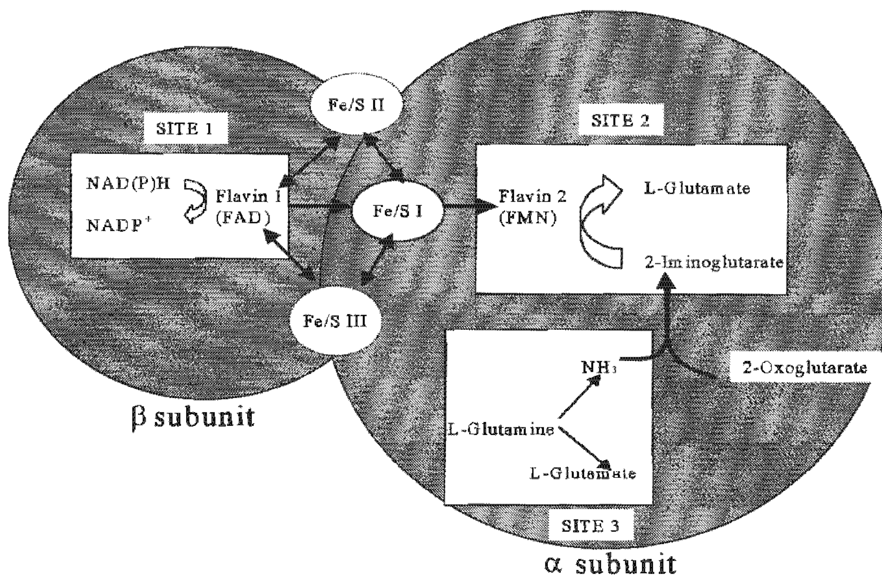


Fig. 1.8 Proposed model for the localization of cofactor and substrate binding sites, and the partial reactions that take place at the three subsites within bacterial glutamate synthases. Adapted from Vanoni and Curti (1999).

In the proposed model, communication between Site1 and Site2 is established by the intramolecular electron transfer chain. Reduction reactions established that Center 1, $[3Fe\text{-}4S]^{0,+1}$ center, and Center 2, a $[4Fe\text{-}4S]^{+1,+2}$ center, most likely participate in the intramolecular

electron transfer process from FAD to FMN, however the role of Center 3, the second [4Fe-4S]$^{+1,+2}$ center, is uncertain since its redox potential seems too low for it to be involved in the electron transfer process (Vanoni *et al.*, 1992). Furthermore, while Center 1 was localized in the α subunit (Vanoni *et al.*, 1998), the localization of Center 2 and 3 could not be assigned to either of the individual bacterial subunits. The study of recombinant GOGAT from *A. brasilense* indicated that correct folding of the two subunits to form the holoenzyme was needed to assemble the two 4Fe-4S clusters, suggesting that the α subunit may provide the correct environment for cluster formation and/or ligands to the clusters, and that these active centers are probably at the interface between the subunits (Vanoni and Curti, 1999). In addition it was concluded that the β subunit domain not only provides reducing equivalents for the overall oxidoreduction reaction, but is also required to determine some of the properties of the holoenzyme: i.e. electronic communication between the 3Fe-4S center and the FMN cofactor within the α subunit, and the tight coupling of ammonia release from glutamine to its addition to 2-oxoglutarate and reduction of the iminoglutarate intermediate to form L-glutamate. Vanoni and Curti, (1999) have suggested that the formation of Fe-S Centers II and III may trigger the conformational changes required to ensure coupling of the glutaminase and glutamate synthase activities and electronic communication between the redox centers.

Characteristically, GOGAT enzymes are very specific for their substrates and cofactors. Most microbes possess a NADPH-dependent activity, although NADH-dependent activity has been reported (Singhal *et al.*, 1989), and *Corynaebacterium callunae* possesses an enzyme with dual coenzyme specificity, although NADH was less effective (25% of NADPH activity) (Ertan, 1992). Reports exist on the ability of various bacterial GOGAT enzymes, to catalyze glutamate synthesis from 2-oxoglutarate and free ammonia, however, the ammonia dependent activity is usually only 2 to 5% of the glutamine-dependent activity, and it does not appear to be physiologically significant (Schreier, 1993). It has been suggested that it may reflect the primitive aminase activity before modification by association of a glutamine amidotransferase function.

1.5.5 Functional domains

The proposed catalytic model is consistent with the identification of strictly conserved domains within the α subunit regions of GOGATs from all three classes. These could be assigned to functional domains for the binding of FMN, for the 3Fe-4S center formation, and an N-terminal domain for glutamine binding and activation (Fig. 1.9), which shared sequence similarity with

the PurF-type class of glutamine amidotransferases whose prototype is the bacterial phosphoribosyl pyrophosphate amidotransferase (PRPP-AT). The N-terminal Cys1 residue of PRPP-AT is involved in the nucleophilic attack of the carboximade of glutamine to initiate amide transfer (Isupov *et al.*, 1996; Kim *et al.*, 1996). The finding that a conserved cysteine residue forms the amino terminal amino acid in all mature Fd and Pyridine-dependent GOGATs so far characterized (see Section 1.5.7 below), and that it is essential for glutamine hydrolysis (Vanoni *et al.*, 1994a), suggests that the GOGAT enzymes use a similar catalytic mechanism to the PurF-type amidotransferases to effect ammonia release from glutamine (Vanoni and Curti, 1999).

Similarly, regions conserved within the bacterial β subunit and the corresponding region of the eukaryotic pyridine-dependent enzyme, could be assigned to NAD(P)H and FAD binding sites. This was based on comparisons of the deduced amino acid sequences of GOGAT with well characterized enzymes that use the same substrates and cofactors (Gosset *et al.*, 1989; Pelanda *et al.*, 1993; Curti *et al.*, 1996; Filetici *et al.*, 1996). Although the isolated β subunit gave no signal that could be assigned to Fe-S clusters (Vanoni *et al.*, 1996), it has been suggested that the two conserved cysteine rich clusters found in β subunits, although not typical of well characterized Fe-S clusters, probably do represent sites for metal binding, but, as discussed earlier, the α subunit may be required for their formation. These functional domains, which are described in detail in Chapter 2, were supported by a series of limited proteolysis studies on the *A. brasilense* enzyme (Vanoni *et al.*, 1994b).

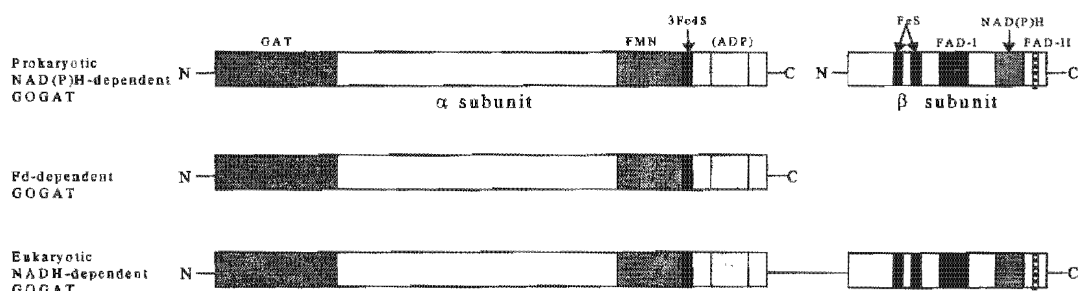


Fig. 1.9 Localization of likely functional domains within GOGATS. GAT, PurF-type aminotransferase domain; FMN, FMN binding domain; 3Fe4S, cystein rich region for the formation of a 3Fe-4S cluster; (ADP) possible adenylate binding domaine of unknown function; FeS, cystein rich regions that could be involved in the formation of the two 4Fe-4S clusters; FAD-I and NAD(P)H, adenylate binding folds for the binding of the respective flavin and pyridine nucleotides; FAD-II, a second FAD binding consensus sequence. Scales are approximate. Taken from Vanoni and Curti (1999).

1.5.6 Common model

The structural, biochemical and cofactor properties shared between GOGATs from various sources, suggest a common function (and presumably catalytic mechanism) in which they all share a common polypeptide harboring the glutamine amidotransferase and glutamate synthase activities (the bacterial α subunit, the single Fd-dependent polypeptide, or the corresponding part of eukaryotic pyridine-dependent GOGAT), but differ in their electron donors; reducing equivalents are either provided by reduced ferredoxin in photosynthetic tissues, or by a FAD-dependent NAD(P)H oxidoreductase in the case of the pyridine nucleotide dependent enzymes, which in eukaryotes is fused to the C-terminus of the glutamate synthase.

1.5.7 Post-translational processing

N-terminal sequencing of the GOGAT subunits has revealed that, in all GOGATs so far studied, post-translational processing occurs such that the mature N-terminus of the α subunit domain is created by proteolytic cleavage of a presequence (e.g. 53 amino acids long in yeast (Cogoni *et al.*, 1995) and 36 amino acids long in both *E. coli* and *A. brasilense* (Oliver *et al.*, 1987; Pelanda *et al.*, 1993) to expose a strictly conserved cysteine residue as the first amino acid. It is uncertain what the role of this presequence is. It is not characteristic of typical *E. coli* post-translationally processed proteins (Oliver *et al.*, 1985). It has been suggested that it may function as a transit peptide signal, targeting the protein to the mitochondria in yeast (Filetici *et al.*, 1996), or, since both Fd- and pyridine-dependent enzymes are nuclear-encoded chloroplast or plastid located proteins, the transit peptides may target the protein to these organelles in plants (Suzuki and Rothstein, 1997). However, they do not share common characteristics of the mitochondrial or chloroplast import signals. Another view is that since the conserved Cys-1 residue is within the glutamine amidotransferase domain and is essential for glutamine hydrolysis and catalysis (Vanoni *et al.*, 1994a), the presequence may be necessary in order to have a reserve of inactive GOGAT in the cell that could be quickly activated by the cleavage of the presequence and the consequent exposure of the amino-terminal cysteine residue. Nothing is known about this process, and it has not been established whether it is due to an autocatalytic property of the enzyme itself, or whether it requires a specific maturation enzyme. In addition, the initial methionine residue of the β subunit is removed to produce the mature polypeptide (Oliver *et al.*, 1987; Pelanda *et al.*, 1993).

1.5.8 GOGAT evolution

It appears that during the course of evolution, bacteria recruited an FAD-dependent NAD(P)H oxidoreductase to input electrons into the glutamate synthase protein. However, clues on the origins of the two subunits may come from studies on the Archaea bacteria. Analysis of the sequence of the archae, *Archaeoglobus fulgidus* (Klenk *et al.*, 1997) and *Methanococcus jannaschii* (Bult *et al.*, 1996) revealed the presence of genes encoding putative GltB-like proteins (estimated to be 55.45 kDa). Although there is no biochemical evidence for the function of these GltB-like proteins, based on sequence conservation of certain domains including the FMN binding site, and the presence of cystein rich regions for Fe-S center formation, it has been proposed that they may represent fully functional GOGAT enzymes comprising the minimal features needed for glutamate synthesis (Vanoni and Curti, 1999). Similarly, Jongsareejit *et al.* (1997) have characterized a 481 amino acid protein, GltA, from the hyperthermophilic archaebacterium *Pyrococcus* sp. KOD1, which is very similar to the bacterial β subunits (36.3% and 30.2% amino acid identity to and *A. brasilense* and *E. coli* enzymes respectively). Furthermore, they report that it is a fully functional GOGAT in the absence of a α subunit, a result which is surprising in the light of the known properties of bacterial GOGATs. These findings led Jongsareejit *et al.* (1997) to suggest that the archae GltB-like proteins and *Pyrococcus* KOD1 GltA may represent ancestral prototypes of the α and β subunit domains of GOGAT respectively. The finding that in some bacteria, α and β subunit homologues do not lie adjacent to each other on the chromosome e.g. in the cyanobacterium *Synechocystis* spp PCC1803 (Navarro *et al.*, 1995) and in *Aquifex aeolicus* (Deckert *et al.*, 1998), may be further evidence for the separate evolution of the two subunits. In fact, the recent emergence of a number of GOGAT β subunit-like proteins that are not associated with an α subunit domain, as well as β subunit-like domains found within other proteins, has also led to the suggestion that the β subunit of GOGAT is derived from a much larger family of FAD dependent NADPH oxidoreductases discussed in Chapter 3 (Vanoni and Curti, 1999).

Despite the very different physiological roles of the Fd-dependent and pyridine-dependent forms of the enzyme, the significant homology that they share suggests that they are evolutionary related proteins. The fact that the pyridine-dependent form of GOGAT is present in such a diverse number of organisms, suggests that it represents the ancestral form of the enzyme, required for basic ammonia assimilation. The apparent fusion of the two subunits can be regarded as characteristic of the evolution of eukaryotes. On the other hand, since Fd-

GOGAT is only present in plants and the cyanobacteria, it may have originated from the α subunit during the evolution of cyanobacteria in which the gene for the small subunit may have got lost. In higher plants both forms of GOGAT are localized in the plastid, although the proteins are encoded in the nuclear genome (Somerville and Ogren, 1982). It is generally accepted that chlorophytic plastid genomes are the remnants of an endosymbiotic ancestral organism believed to have been of cyanobacterial origin. The majority of genes required for plastid metabolism have either been transferred to the nuclear genome, or their function taken over by eukaryotic nuclear genes. Therefore, it seems possible that the plant Fd-GOGAT has an endosymbiotic cyanobacterial origin. This is supported by the discovery of a plastid encoded Fd-GOGAT homologue (sharing 55.1% similarity to maize) in the multicellular red algae (Valentin *et al.*, 1993). Furthermore, these authors indicate that in the chromophytic algae, proposed to have originated by secondary endosymbiosis of the red algae, the Fd-GOGAT encoding genes have been transferred to the nuclear genome, thus paralleling the scenario in chlorophytes. Hence, Fd-GOGAT may represent another example of a conserved gene transferred from a cyanobacteria-like organisms to plants, in this case to minimize the loss of nitrogen due to photorespiration.

1.5.9 Regulation of GOGAT activity

Despite the central role GOGAT plays in nitrogen metabolism, regulation of expression, translation and post-translational processes remain poorly understood. In general, GOGAT activity was found to be repressed by glutamate, or a good source of glutamate, in most bacteria. There is no evidence of regulation at the post-translational level, and generally it has been found to be relatively insensitive to regulation by feedback inhibition (Schreier, 1993; Merrick and Edwards, 1995). Specifically, L-glutamate had no significant inhibitory effect on any GOGATs so far examined. Regulation of GOGAT activity has been shown to occur primarily at the level of enzyme synthesis in several microorganisms including *E. coli* (Castano *et al.*, 1988), *K. aerogenes* (Senior, 1975) *S. typhimurium* (Brenchley *et al.*, 1975) *R. phaseoli* (Bravo and Mora, 1988), *S. coelicolor* (Fisher, 1989), *B. subtilis* (Bohannon *et al.*, 1985), suggesting a regulatory role for this enzyme in biosynthetic pathways involving glutamate. The genetic control of GOGAT expression has only been studied in *E. coli* and *B. subtilis* to date and will form the following discussion.

1.5.9.1 GOGAT in *E. coli*

In *E. coli,* the two GOGAT structural genes (*gltB* and *gltD*) form an operon with a third gene *gltF* (Castano *et al.*, 1988), suggested to encode a putative regulatory protein (26.3 kDa) (Castano *et al.*, 1992). The *gltBDF* genes are cotranscribed on a single major 7.5 kb mRNA (Castano *et al.*, 1988; Oliver *et al.*, 1987), and *gltB* mutants have been shown to exert polar effects over the downstream genes of the operon (Bolivar *et al.*, 1992). However, weak promoter activity was detected in the region preceding the *gltF* initiation site (Grassl *et al.*, 1999). In addition the long intercistronic region between *gltD* and *gltF* (561 bases) contains two potential sequences of imperfect symmetry that could function as transcriptional termination signals (Bolivar *et al.*, 1992). However, as yet it has not been established whether, under particular metabolic conditions, any of the transcripts initiated at the *gltB* promoter terminate downstream from *gltD.* Interestingly only 12 nucleotides separate the termination codon of the large (α) subunit from the translational start point of the small (β) subunit. This results in the RBS of the β subunit overlapping with the termination codon of the α subunit, and suggests that both genes are translationally coupled. Such close proximity is thought to ensure the co-ordinated expression of genes with related functions.

Expression of the *gltBDF* operon is not regulated by the global Ntr system (Reitzer and Magasanik, 1987). Unlike other Ntr enzymes, GOGAT expression is strongly repressed by growth in glutamate, yet derepressed to the same extent by growth in either a nitrogen limiting source such as glutamine or in excess ammonia (Castano *et al.*, 1988). Instead, the expression of the *gltBDF* operon in *E. coli* is positively regulated by the leucine-responsive global regulatory protein, (Lrp) (Ernsting *et al.*, 1993).

1.5.9.1.1 Lrp (leucine responsive regulatory protein)

Lrp has recently emerged as a global transcriptional regulatory protein in *E. coli* affecting the expression of numerous unrelated genes and operons, either positively or negatively, in response to leucine (reviewed by Calvo and Matthews (1994), and Newman *et al.* (1996). Part of the regulon consists of proteins with prominent roles in the regulation of nitrogen metabolism: Lrp positively regulates *glnA* and *gltBDF* genes, as well as genes involved in amino acid biosynthesis. Pyridine nucleotide transhydrogenase (*pnt*) expression is also positively regulated in keeping with the demand for NADPH during the biosynthesis of amino acids and the assimilation of ammonia. On the other hand Lrp negatively regulates many operons involved in amino acid catabolism and nutrient transport.

Lrp regulates its target genes by binding directly to specific sites in the promoter regions, and its coregulator leucine merely modulates the activation or repression of these target genes (Calvo and Matthews, 1994). A model, proposed by Ernsting *et al.* (1993) and supported by Borst *et al.* (1996), to explain the broad range of sensitivities to leucine observed in Lrp regulated operons, suggests that leucine affects the affinity of Lrp binding to its target DNA, without apparently affecting the sequence specificity of binding (Cui *et al.*, 1995). The extent of this effect is a function of both the intrinsic affinity of the specific target gene for Lrp, with leucine insensitive genes, e.g. the *gltBDF* operon, exhibiting a higher affinity for Lrp than leucine sensitive genes, as well as the effective intracellular concentration of Lrp.

The *lrp* gene has been cloned and sequenced (Austin *et al.*, 1989; Willins *et al.*, 1991; Platko and Calvo, 1993). Mutational analysis indicated that the protein contained three functional domains: a DNA-binding domain within the N-terminal part of the protein which contained a predicted helix-turn-helix (H-T-H) motif, a transcriptional activation domain within the middle part, and a leucine response domain which mapped to the C-terminal one third of the protein. In solution it exists as a dimer composed of two identical subunits of predicted size (18.8 kDa).

1.5.9.1.2 Regulation of the *gltBDF* operon in *E. coli*

The expression of the *gltBDF* operon is dependent on, and highly sensitive to Lrp levels (Ernsting *et al.*, 1992), with GOGAT activity being undetectable in an *lrp* strain. On the other hand, expression of the reporter gene construct *gltB*::*lacZ*, was 44-fold higher in an *lrp*+ strain compared to an isogenic *lrp* strain grown in minimal medium. The operon is relatively insensitive to leucine with the addition of saturating concentrations of L-leucine to the growth medium resulted in only a 2.2-fold decrease in transcriptional activation (Ernsting *et al.*, 1993).

Details of the finer genetic control of the *glt* operon is provided by the recent studies of Wiese *et al.* (1997), which form much of the following discussion. Transcription of the *gltBDF* operon is regulated by Lrp binding specifically to at least three sites upstream, spanning the region from -140 to -260 bp relative to the start of transcription. Maximum transcriptional activation requires that Lrp binds to all three sites and that these sites are properly phased with respect to one another and with the start of transcription. The binding of Lrp to flanking sites 1 and 3 was found to be highly co-operative with Lrp bound to site 2, which showed the highest affinity for

Lrp. A similar pattern is observed in other Lrp activated operons (Gally *et al.*, 1994; Gazeau *et al.*, 1994).

Analysis of the six Lrp binding sites in the *ilvIH* operons from *E. coli* and *S. typhimurium* revealed that they are organized either as palindromic motifs or as half motifs (Wang *et al.*, 1993), from which consensus sequences have been derived: 5'-AgaATTTTATtcT-3'and 5'-TTATtcT-3' or 5'-AAATtcT-3') (Wang and Calvo, 1993; Cui *et al.*, 1995). The palindromic nature of the consensus sequence, and the stoichiometry of binding, suggests that an Lrp dimer binds to the DNA in such a way that each monomer binds to a half site (Cui *et al.*, 1995; Cui *et al.*, 1996).

The central high affinity Lrp binding site of the *gltBDF* promoter region, site 2, shows clear palindromic symmetry with a very good match to the consensus sequences. Site 3 is also palindromic, while site 1 appears to only involve a half site interaction. Thus, the consensus-like palindromic features of site 2 appear to be required to establish strong initial binding, facilitating Lrp dimers to bind co-operatively to the flanking sites. The centers of site 1 and 2 are separated by 31 bp, effectively positioning the sites on the same face of the helix, which presumably facilitates co-operative protein-protein interactions. The proposed nucleoprotein activation complex so formed could then efficiently deliver Lrp to a critical proximal binding site where it could directly interact with RNA polymerase, hence regulating transcription.

Although regulation of the *gltBDF* operon by Lrp was consistent with the model in that mutations affecting the affinity of Lrp for its target sites significantly decrease *gltBDF* expression (Wiese *et al.*, 1997), this model was not sufficient to describe *in vivo* regulation. Leucine decreased the affinity of Lrp for *gltBDF* DNA, however it unexpectedly increased the amount of transcription from Lrp-DNA complexes that did form. This led to a refinement to the model which suggest that it is the difference between the positive effect leucine has on transcriptional activation by whatever Lrp is bound, and the negative effect it has on binding affinity, that determines the overall leucine sensitivity for a given operon. For the *gltBDF* operon therefore, the 2.2-fold decrease in activation observed in the presence of leucine, is the cumulative result of an ~8.8-fold decrease in the affinity of Lrp for DNA, and a ~3.6-fold increase in the activation efficiency of bound Lrp. Since homologues of Lrp are conserved in a number of other members of the family *Enterobactereacea* (Calvo and Matthews, 1994), it is

likely that the mechanism by which *gltBDF* is regulated in *E. coli* will serve as a model for Gram-negative bacteria in general.

1.5.9.1.3 Relationship between Lrp, the GOGAT operon and the Ntr network

The discovery that GOGAT mutants are unable to utilize any nitrogen sources that require induction of the Ntr response, and are unable to induce GS and other nitrogen regulated proteins in response to nitrogen limitation, led to the conclusion that expression of the *gltBDF* operon was required for induction of the Ntr response (Castano *et al.*, 1988; Reitzer, 1996). This may also explain the dependence of the Nif phenotype (i.e. the ability to fix nitrogen in the free living state) on GOGAT expression in many bacteria (Bani *et al.*, 1980; Donald *et al.*, 1988), since nitrogen fixation is controlled by the Ntr system in enteric bacteria.

One theory suggests that GOGAT activity is essential for induction of the Ntr response because it prevents the accumulation of intracellular glutamine. A second widely accepted hypothesis suggested that the *gltF* gene encoded a regulator, GltF, which was required for induction of the Ntr system, possibly by representing an alternative route for the phosphorylation of NtrC or NtrB, since it shared similarity with protein kinases. However, the exact role of *gltF* remained unclear (Castano *et al.*, 1988; Castano *et al.*, 1992). Recent studies on the role of the *gltF* gene product, argues strongly in favour of the first theory (Grassl *et al.*, 1999). First, a non-polar *gltF* insertion mutant was shown to have no detectable phenotype with respect to amino acid or ammonia transport, or utilization of alternative nitrogen sources. Second, the GltF protein, which is preceded by a signal peptide, is exported into the periplasmic space, an unlikely location for the proposed regulator. These results led Grassl *et al.* (1999) to conclude that GltF is not involved in regulation of the nitrogen regulon genes as previously thought.

Lrp- strains were also shown to be functionally Ntr- (Ernsting *et al.*, 1992). However, unlike the direct regulation of the *gltBDF* operon, regulation of GS and other Ntr enzymes by Lrp was indirect, and depended on a functional signal transduction pathway. Because GOGAT activity is undetectable in the absence of a functional Lrp protein, it appears that the effect of Lrp on the Ntr regulon is due to its direct effect on *gltBDF* transcription (Calvo and Matthews, 1994). Thus our understanding of Ntr regulation is broadened by the involvement of Lrp and the GOGAT operon. Newman *et al.* (1992) conclude that Lrp is a sensor of general nutritional conditions. Therefore, its levels will be high in a nutritionally poor environment, resulting in

high expression of the *gltBDF* operon and induction of the Ntr regulon in response to the availability of ammonia.

1.5.9.1.4 Additional aspects of GOGAT regulation

GOGAT levels are also controlled by other factors. Nac, which is itself controlled by the Ntr system, is involved in the negative regulation of GOGAT expression in some Gram-negative bacteria in the absence of ammonia (Bender *et al.*, 1983; Macaluso *et al.*, 1990). A cyclic AMP receptor protein (CRP) binding site overlaps the –35 RNA polymerase binding site of *gltBDF* operon in *E. coli* (Oliver *et al.*, 1987), and since GOGAT expression is affected by the cyclic AMP (cAMP) receptor protein-cAMP complex, it has been suggested that CRP dependent repression may prevent GOGAT from draining 2-oxoglutarate from the citric acid cycle when carbon or energy is limiting (Reitzer, 1996). However, in none of these cases is the mechanism of regulation understood. Clearly, the relationship between the *glt* operon expression, GOGAT activity, and nitrogen metabolism is complex in the enteric bacteria.

1.5.9.2 Regulation of GOGAT in *B. subtilis*

Not surprisingly, the mechanism of regulation of GOGAT activity in *B. subtilis* is very different from the enteric bacteria, and GS and GOGAT do not appear to be regulated in a co-ordinated manner in this bacterium either, since L-glutamine, ammonia, and glutamate affect them quite differently (Pan and Coote, 1979). Glutamate (or good sources of glutamate) represses GOGAT activity but induces GS activity. GS activity is repressed by L-glutamine and ammonia, while L-glutamine only partially repressed GOGAT activity and ammonia completely derepressed GOGAT activity. The ratio of GOGAT activity for cells grown in glutamate, glutamine, and ammonia as the sole nitrogen source are approximately 1:3:7 respectively (Pan and Coote, 1979; Schreier *et al.*, 1982). Bohannon *et al.* (1985) demonstrated that this nitrogen-source-dependent regulation of GOGAT activity was exerted primarily at the level of transcription. Subsequently, it was shown that GOGAT gene expression (*gltA* and *gltB*) is dependent on a positive regulatory protein, GltC (Bohannon and Sonenshein, 1989), unrelated to *E. coli* Lrp or GltF.

1.5.9.2.1 GltC belongs to the LysR family of bacterial regulators

Sequence analysis revealed that GltC belongs to the large LysR family of positive DNA-binding transcriptional activators (Bohannon and Sonenshein, 1989), of which *E. coli* LysR serves as the archetype (Henikoff *et al.*, 1988). The features that characterize this family of regulators, reviewed by Schell (1993), are conserved in GltC: the gene encoding the regulator,

gltC, is linked to and transcribed divergently from its target gene, *gltA*, from overlapping promoters; GltC negatively autoregulates its own expression (Bohannon and Sonenshein, 1989), and it is a small polypeptide (calculated molecular mass of 34 kDa) containing the highly conserved amino terminal domain (~65 residues) which includes the characteristic H-T-H motif (positions 18 to 37) associated with DNA binding activity (Belitsky *et al.*, 1995). The less well conserved C-terminal regions of the LysR-type regulators contain isolated domains involved in coinducer recognition and response, and have been suggested to play a role in subunit interactions and/or protein-DNA interactions necessary for transcription activation. Studies on the mechanism of activation of target genes by LysR-type regulators has been shown to depend on them binding to specific DNA regions of ~15 bp partially dyadic sequences, which contain the proposed LysR consensus binding motif, T-N_{11}-A (Goethals *et al.*, 1992). The primary recognition site is centered at ~-65 with respect to the transcription start sites of the target genes.

1.5.9.2.2 Regulation of the *gltAB* and *gltC* genes by GltC

The transcriptional starts of the two divergent genes, *gltA* and *gltC*, are separated by approximately 51/52 bases, and as illustrated in Fig. 1.10, the -10 and -35 regions of the two promoters overlap extensively (Bohannon and Sonenshein, 1989). Two 15 bp dyad symmetry sequences, BoxI and BoxII, conforming to the LysR consensus site, are present in the intergenic region immediately upstream of the *gltA* promoter. BoxI, centered at position -64 with respect to the start point for *gltA* transcription, is essential for both the GltC-mediated activation from the *gltA* promoter, as well as for repression from the *gltC* promoter (Belitsky *et al.*, 1995). BoxII, which is separated from BoxI by two DNA helical turns, and partially overlaps the -35 region of the *gltA* promoter, is also required for full activation of *gltA* transcription, but in contrast to BoxI, was not required for autorepression of the *gltC* gene (Belitsky *et al.*, 1995). The requirement for a second binding site situated closer to the promoter of the target gene is a common feature of LysR-type gene regulation, although the extent of the similarity between the two binding sites varies for different LysR-type proteins (Schell, 1993).

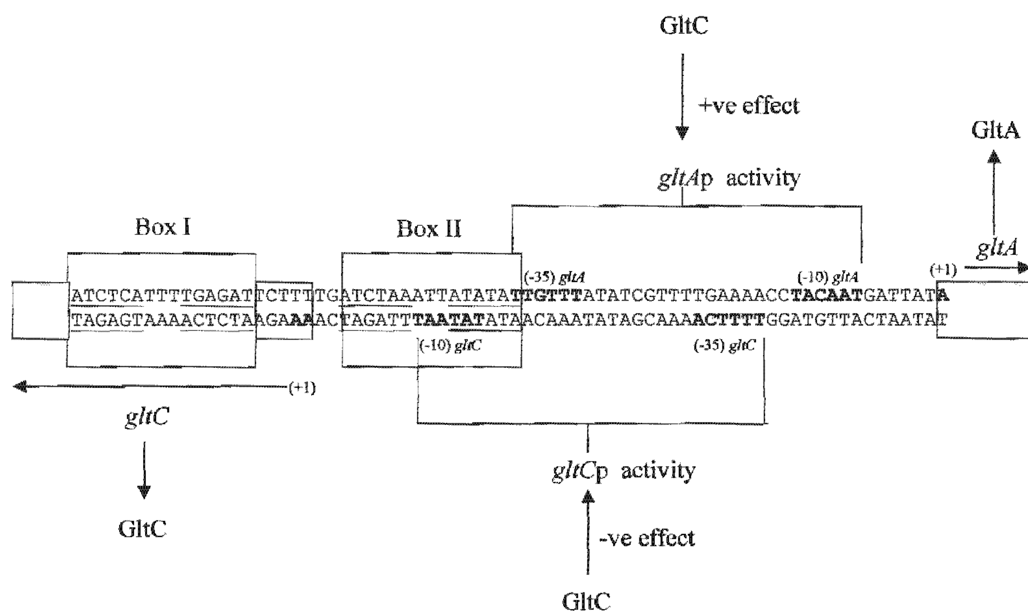


Fig. 1.10 Regulatory features spanning the *B. subtilis gltCA* intergenic region. The *gltC* and *gltA* -10 and -35 promoter sequences are indicated in boldface type, along with the transcription initiation sites (+1): there are two potential initiation sites for *gltC*. Arrows denote the direction of transcription of the respective genes. The underlined sequences highlight the regions of dyad symmetry within BoxI and BoxII. The binding of GltC, the product of *gltC*, to BoxI and BoxII is required for activation of transcription from the *gltA*p promoter (+ve effect), however this results in the repression of its own promoter (*gltC*p) activity (-ve effect). Adapted from Belitsky *et al.* (1995).

Mutations in the H-T-H motif of GltC resulted in glutamate auxotrophy (Schreier, 1993), confirming that this region was necessary for the expression of the GOGAT genes. Genetic evidence supports the theory that GltC is the regulatory protein directly involved in interactions with the BoxI-BoxII regulatory regions, rather than having some modulating effect on another protein. Elaborate mutational studies by (Belitsky and Sonenshein, 1995) showed that mutations in *gltC* could compensate for defects in *gltA* expression caused by mutations in BoxI and/or BoxII, and furthermore that different *gltC* mutants affected the expression of altered versions of the *gltC-gltA* regulatory region to different extents.

1.5.9.2.3 Proposed model for the regulation of GOGAT activity

The recent advances in the genetic regulation of GOGAT activity by Belitsky *et al.* (1995) and Belitsky and Sonenshein (1995), has led them to propose a model for GOGAT regulation in *B. subtilis*. Since the expression of *gltC* is not significantly regulated by the nitrogen source, it was proposed that GltC binds to BoxI under all growth conditions, leading to low but constant intracellular concentrations of *gltC* message. Such coinducer independent binding to the BoxI-

like primary binding site is characteristic of LysR-type proteins (Schell, 1993). Since BoxI is located just downstream of the *gltC* transcriptional start site, and since GltC-mediated autorepression is dependent on GltC being in a certain rotational orientation relative to RNA polymerase, as well as in close proximity to the *gltC* promoter, it is predicted that GltC interferes with its own transcription initiation.

It is envisaged that, under conditions of glutamate limitation, inducer molecules stimulate the binding of GltC to BoxII, the activation site. Since both Boxes have similar sequences of dyad symmetry, and because their relative spatial orientation is crucial for *gltA* activation (altering the spacing to a non integral number of helical turns reduced nitrogen dependent activation of *gltA*), it was proposed that each site binds a dimer of GltC, and that the positive regulation of *gltA* is dependent on co-operative dimer-dimer interactions. It is hypothesized that the interaction of inducer with regulator confers a conformational change upon the protein-DNA complex, which enhances the association of RNA polymerase with the promoter of the target gene, resulting in increased levels of transcription of the target genes (Fisher and Long, 1993; Schell, 1993). This dependence on a positive regulator may compensate for the particularly poor -35 *gltA* promoter sequence (Moran, Jr. *et al.*, 1982). This model suggests that the function of the nitrogen source dependent signal for *gltA* transcription is to promote GltC dimer-dimer interactions, stabilizing GltC binding to the activation site. This was supported by mutations, both in the *gltCA* intergenic region and in the GltC protein, near a region identified in LysR-type proteins to be involved in effector recognition and response (between positions 95 to ~173; Schell, 1993), which alleviate the requirement for a nitrogen source-dependent signal in *gltA* activation by apparently creating stronger dimer-dimer interactions.

The metabolic signals that inform GltC of the glutamate status inside the cell have not been identified, however it has been observed that the intracellular concentration of glutamine is involved in controlling GOGAT synthesis in *B. subtilis* (Deshpande and Kane, 1980; Deshpande *et al.*, 1981; Fisher and Sonenshein, 1984). Interestingly, there is no correlation between levels of glutamine and GOGAT activity in cultures of *B. licheniformis* (Schreier *et al.*, 1982).

1.5.9.2.4 Additional regulatory aspects

The gene specific regulator, GltC, is not the only factor controlling the expression of the *B. subtilis* GOGAT genes. GOGAT synthesis may also be regulated by a second LysR-type

protein, GltR, encoded by *gltR*, which was isolated and characterized following observations that the low level of *gltA* expression seen in a *gltC* mutant, was further down regulated under nonactivating conditions (Belitsky and Sonenshein, 1997). A single mutation in *gltR* (at position 219) allowed the protein to activate transcription of *gltAB* in the absence of GltC. Furthermore, this mutant GltR-dependent activation of *gltA* required the BoxI and BoxII sequences in the *gltCA* regulatory region. The significance of GltR is unclear since it is neither necessary nor inhibitory for *gltA* expression. However, as discussed by Belitsky and Sonenshein (1997), it may require specific conditions to trigger its conversion from an inactive to an active conformation. Since the GltR mutation lay within a region which produces constitutively active proteins in other LysR family members, it may have enabled the protein to gain its transcription-activating function in the absence of the otherwise required effector molecules.

Regulation of the *gltAB* operon also appears to be linked to the global nitrogen regulatory system involved in GS regulation described in Section 1.4.1. TnrA negatively regulates GOGAT expression during nitrogen limited growth (Fisher, 1999). The mechanism by which TnrA represses *gltAB* expression is not understood, but it does not involve GltC. Furthermore, since no obvious TnrA binding sites could be identified in the *gltAB* promoter region, TnrA regulation of the *gltAB* operon is probably mediated indirectly (Fisher, 1999). Furthermore, it has recently been established that expression of the *gltAB* operon in *B. subtilis* is induced by glucose, and that this induction depends on a functional catabolite control protein, CcpA (Faires *et al.*, 1999). CcpA is a positive transcriptional activator of the *ack* gene encoding acetate kinase (Grundy *et al.*, 1993). However, since a catabolite responsive element (*cre* sequence) could not be identified in the *gltAB* upstream region, it suggests that other factors controlled by CcpA are involved in this regulation. Since the CcpA protein is central to the regulation of carbon metabolism in Gram-positive bacteria, these findings provide a direct link between carbon and nitrogen metabolism, and suggest that the role of the CcpA protein in metabolic regulation may be conserved in other Gram-positive bacteria as well. Interestingly, an Lrp-like protein has recently been associated with branched chain amino acid transport in *B. subtilis*, but so far no *lrp*-like homologues identified in *B. subtilis* have been shown to be involved in GOGAT regulation (Belitsky *et al.*, 1997).

Given the pivotal role GOGAT plays in linking carbon and nitrogen metabolism, it is not surprising that a number of global regulators affect the steady state level of GOGAT, and

suggest that the regulation of GOGAT activity in *B. subtilis*, and possibly other Gram-positive organisms, is more complex in its response to the nitrogen source than originally thought.

From the above discussion it is clear that the mechanisms of nitrogen regulation present in those Gram-positive organisms so far studied are unique and distinct from the Gram-negative enteric Ntr system, and that there is no unifying theme in the arrangement or regulation of genes encoding GS and GOGAT.

1.6 Aims of this study

Although solventogenic clostridial strains were used extensively in fermentations for the commercial production of solvents, fundamental aspects of their physiology remain poorly understood. There is sparse information on the pathways involved in and the growth conditions regulating the incorporation of ammonia into glutamine and glutamate, key metabolites from which nearly all other cellular nitrogen containing compounds are derived (Reitzer, 1996). Since genetic engineering can be used to improve growth rate and solvent yields in fermentations, understanding the gene structure and regulation of enzymes involved in nitrogen metabolism in *C. acetobutylicum* has fundamental and practical importance. This, together with the obvious lack of information available on the mechanisms of nitrogen control in Gram-positive organisms, provided the motivation for this study. The *C. acetobutylicum* P262 *glnA* gene was previously sequenced, however the regulation of GS activity was only assessed under a limited set of conditions; (Janssen *et al.*, 1988; Janssen *et al.*, 1990; Fierro-Monti *et al.*, 1992). Our aim was to extend the genetic characterization of the GS locus, to identify and characterize the GOGAT locus, and to determine the conditions and factors regulating these enzyme activities in this Gram-positive spore forming anaerobe. Furthermore, these studies may contribute to our understanding of the conditions controlling cell differentiation.

In anaerobic microorganisms, the electron transport proteins that provide the source of electrons for the reductive activation of the pro-drug metronidazole are involved in pyruvate metabolism (Johnson, 1993). Previously in our laboratory, an *E. coli* F19 *recA*, nitrate reductase-deficient mutant was developed as a selection system to clone genes from *C. acetobutylicum* P262 which activate metronidazole and render the *E. coli* host strain sensitive to the drug (Santangelo *et al.*, 1991). Our second aim was to characterize clone pMET13C1, which rendered *E. coli* F19 highly sensitive to metronidazole, to increase our understanding of

C. acetobutylicum electron/carbon flow, gene regulation and general physiology in this organism, with the potential for future manipulation of fermentation patterns via changes in electron distribution.

CHAPTER 2

Cloning and molecular characterization of the genes encoding glutamate synthase (GOGAT) from *Clostridium acetobutylicum* P262

2.1 Summary **52**
2.2 Introduction **53**
2.3 Materials and methods **53**
2.3.1 Bacterial strains, plasmids, and culture conditions ---- 53
2.3.2 DNA extractions and general DNA manipulations ---- 54
2.3.3 Construction of size-selected genebanks of *C. acetobutylicum* ---- 54
2.3.4 Colony hybridization ---- 54
2.3.5 Southern hybridization ---- 55
2.3.6 Nucleotide sequencing ---- 55
2.3.7 Sequence analysis ---- 55
2.4 Results and discussion **56**
2.4.1 Sequencing of the regions flanking the *C. acetobutylicum* P262 *glnA* and *glnR* genes.- 56
2.4.2 Isolation of the chromosomal region extending downstream of *gln*R ---- 57
2.4.3 Nucleotide sequence analysis of the 7.95 kb region downstream of *glnR*. ---- 60
2.4.4 Protein analysis of the deduced GltA and GltB products ---- 66
2.4.5 Sequence identification of functional domains in the deduced *C. acetobutylicum* GltA and GltB proteins ---- 69
2.5 Conclusion **76**

CHAPTER 2

Cloning and molecular characterization of the genes encoding glutamate synthase (GOGAT) from *Clostridium acetobutylicum* P262

2.1 Summary

The regions flanking the *C. acetobutylicum* P262 *glnA* and *glnR* genes were analyzed. A truncated ORF of 215 amino acid residues was located 640 bp upstream, and in the opposite direction to *glnA*. It shared significant homology with the amino terminal regions of aspartokinases. The 7.95 kb DNA region extending downstream from *glnR*, was cloned and sequenced, yielding two complete ORFs in the same orientation as *glnA* and *glnR*. The first ORF (4554 bp) was separated by 108 bp from *glnR*, and coded for a putative protein of 168.5 kDa. The second ORF (1473 bp) coded for a putative protein of 53.8 kDa. Sequence comparisons of their deduced amino acid products identified them as the structural genes coding for the large (α) and small (β) subunits of GOGAT respectively. The α subunit gene was designated *gltA*, and the β subunit gene was designated *gltB*. The *gltB* initiation codon was located 12 bp downstream from the *gltA* termination codon. A third truncated ORF of 280 residues was located 877 bp downstream from *gltB*, and shared homology to the amino terminal regions of genes encoding isocitrate dehydrogenases. All three genes were preceded by a conserved clostridial RBS, while putative promoter sequences were identified upstream of the *icd* and *gltA* genes. Two inverted repeat sequences, with the potential to form transcriptional terminators, were identified. The first one (ΔG = -12.1 kcal/mol) was located downstream of the *glnR* gene, positioned between the putative *gltA* promoter sequence and the *gltA* structural gene. The second (ΔG = -23.7 kcal/mol) was located downstream of the *gltB* gene. Likely binding domains for flavin cofactors (FAD and FMN) and NADH, a domain for glutamine binding and activation, and cysteine clusters for iron-sulfur center formation were identified in the deduced *gltA* and *gltB* amino acid sequences.

2.2 Introduction

Plasmid pHZ200 was originally isolated from a *C. acetobutylicum* P262 genomic library by its ability to complement the GS-deficient *E. coli* strain YMC11 (Usden *et al.*, 1986). The *glnA* gene, encoding the structural gene for GS contained on this 6.1 kb insert, has been the focus of previous research by our laboratory (Janssen *et al.*, 1988; Janssen *et al.*, 1990; Fierro-Monti *et al.*, 1992)(Section 1.4.3). More recently, a second gene, *glnR*, associated with a 30 kDa protein, was identified immediately downstream of *glnA,* and has been proposed to code for a putative response regulator protein that may regulate *glnA* transcription via an antitermination mechanism (Woods and Reid, 1995). However, the ~1.4 and ~2.4 kb regions that lie upstream and downstream of *glnA* and *glnR* on pHZ200 respectively, have not been investigated. Since genes encoding sensor and response regulators are generally located adjacent to one another (Stock *et al.*, 1989), and since a GS regulator is encoded upstream of the *glnA* gene in *B. subtilis* (Section 1.4.1), it was decided to initiate our study by characterizing this GS locus more extensively with the aim of expanding our knowledge on nitrogen metabolism in this Gram-positive bacterium.

2.3 Materials and methods

2.3.1 Bacterial strains, plasmids, and culture conditions

The wild type *C. acetobutylicum* strain P262 was originally obtained from National Chemical Products (NCP), Germiston, South Africa, and has been described by Jones *et al.* (1982). A laboratory spore stock was maintained aerobically in sterile distilled water at 4 °C. Spores were activated by heat shock treatment at 70 °C for 3 min, cooled on ice for 1 min, and used as the primary inoculum. *C. acetobutylicum* was cultured in buffered Clostridium basal medium (CBM) broth, as described by Allcock *et al.* (1982), at 34 °C in an anaerobic glove cabinet (Forma Scientific Inc.) containing a gas atmosphere of 5% H_2, 10% CO_2 and 85% N_2.

E. coli strains JM105 and JM109 (Yanisch-Perron *et al.*, 1985) were used interchangeably as the cloning hosts. The Bluescript pSK vector (Stratagene) was used for all subcloning and sequencing of *C. acetobutylicum* genomic DNA fragments. Fig. 2.1 describes subclones derived from plasmid pHZ200 (Usden *et al.*, 1986), as well as additional plasmids generated in

this study. *E. coli* clones were routinely grown at 37 °C in 2xYT medium (Messing, 1983) containing ampicillin (100 μg/ml) for plasmid selection, and where necessary solidified with agar (1.5% w/v). Further details of the strains and vectors used are supplied in Appendices A and B, respectively.

2.3.2 DNA extractions and general DNA manipulations

C. acetobutylicum genomic DNA was prepared from one liter of CBM grown culture, according to the method of Marmur (1961), as modified by Zappe *et al.* (1986) to overcome the high nuclease activity exhibited by this organism. RNA was removed by treatment with ribonuclease A (Sigma), and the precipitated chromosomal DNA pellet was resuspended in distilled water and stored at 4 °C. Large scale plasmid DNA preparations were isolated from *E. coli* using the Nucleobond® AX kit (Macherey-Nagel), while the alkali-hydrolysis method of Ish-Horowicz and Burke (1981) was used for small scale (miniprep) plasmid isolations.

Restriction endonuclease digestions were performed according to the manufacturer's recommendations. All cloning procedures and other routine DNA manipulations were as described by Sambrook *et al.* (1989). Selected DNA fragments for subcloning were gel purified using the Geneclean® III kit (Bio 101, Inc). All gel electrophoresis of DNA was conducted in 0.8% w/v agarose gels using a Tris-Acetate EDTA buffer.

2.3.3 Construction of size-selected genebanks of *C. acetobutylicum*

C. acetobutyilicum chromosomal DNA (50 μg) was digested to completion with the appropriate restriction enzymes and the resulting fragments fractionated by gel electrophoresis. DNA fragments of the required size, determined by Southern blot analyses, were recovered from the gel using the Geneclean procedure, and ligated into the appropriately digested pSK vector. The ligated species were transformed into *E. coli,* and transformants harbouring recombinant plasmids were selected by colony hybridization.

2.3.4 Colony hybridization

E. coli colonies harboring recombinant plasmids with *C. acetobutylicum* insert DNA, were duplicated onto a master plate and a positively charged Hybond N^+ membrane overlaid on 2xYT solid media containing ampicillin. After overnight incubation the resulting colonies were lysed, their DNA fixed to the membrane filters, and the cell debris washed off according to

Sambrook *et al.* (1989). This was followed by hybridization and detection as described for Southern hybridization.

2.3.5 Southern hybridization

Southern hybridization was performed using the Boeringer Mannheim non-radioactive DIG DNA labeling and detection kit, according to their instructions. Plasmid and chromosomal DNA were digested to completion with the appropriate restriction endonucleases, and fractionated by gel electrophoresis. The DNA was transferred by capillary action, in 0.4 M NaOH, onto a Hybond N^+ nylon membrane and fixed accordingly. DNA fragments selected as probes were gel purified, and labeled with Digoxigenin-11-dUTP by the random primed method. Hybridization was performed using 20 ng/ml of denatured DIG-labeled probes at 68 °C overnight. Membranes were washed under stringent conditions, followed by chemiluminescent detection with Lumigen™ PPD.

2.3.6 Nucleotide sequencing

The complete sequence of 1308 bp upstream of *glnA*, and 7950 bp downstream of *glnR* was assembled from the sequences obtained from clones pHS1, pHS2, pHS4 and pHS5 (Fig. 2.1). To achieve this, nested deletions were generated of the two larger clones, pHS4 and pHS5, using the Exonuclease III degradation technique of Henikoff (1984). Details of the strategy are outlined in the legend to Fig. 2.1. The sequence downstream of *glnR* was submitted to GenBank and assigned the accession number AF082880.

Double-stranded plasmid DNA templates were sequenced by the dideoxy-chain termination method of (Sanger *et al.*, 1977) using the Auto read sequencing kit (Pharmacia) and an OmniGene thermocycler (Hyband). Depending on the strand to be sequenced, fluorescently labeled M13 1212 forward or 1201 reverse primers were used, and the sequence was resolved using an Alfexpress™ automated DNA sequencer (Pharmacia).

2.3.7 Sequence analysis

The sequence data were analyzed on a VAX 6000-330 computer using the Genetics Computer Group (GCG) sequence analysis software package (Devereux *et al.*, 1984). In particular, the BESTFIT component was used to calculate amino acid similarities and identities between protein sequences, and the FOLD component was used to determine the minimum free energy of stem-loop structures. Amino acid homology searches were carried out against the cumulative non-redundant updated protein database at NCBI using the BLAST algorithm of (Altschul *et*

al., 1997). DNAMAN (version 4.13) software was used to produce the multiple sequence alignments (gap penalty of 3, length penalty weight of 0.1).

2.4 Results and discussion

2.4.1 Sequencing of the regions flanking the *C. acetobutylicum* P262 *glnA* and *glnR* genes.

A comprehensive restriction endonuclease map of plasmid pHZ200 (Usden *et al*., 1986) was generated, and the ~1.4 kb *Pst*I-*Eco*RV fragment upstream of the *glnA* gene was subcloned, yielding plasmid pHS1 (Fig. 2.1). Sequence analysis of this fragment revealed a truncated open reading frame of 215 amino acids transcribed in the opposite direction to *glnA*, with the putative initiation codon separated by 640 bp from the *glnA* start codon (Fig. 2.2). Amino acid homology searches of this truncated open reading frame revealed high homology to the amino-terminal domain of aspartokinases.

Aspartokinases catalyze the first committing step in the utilization of L-aspartate for the biosynthesis of the amino acids diaminopimelate, lysine, threonin and methionine (reviewed by Paulus (1993)). Three isozymes of aspartokinase have been identified in *B. subtilis* which differ in their feedback control mechanisms suggesting functional specialization. Our truncated ORF showed the highest identity to the aspartokinase II isozyme from *B. subtilis* (38% identity), encoded by the *lysC* operon. In *B. subtilis,* both the activity and synthesis of aspartokinase II are regulated by lysine, suggesting that its primary function is to provide precursors for the synthesis of lysine. The extensive 330 bp untranslated *lysC* leader region plays an important role in the regulation of aspartokinase II synthesis by lysine, and is characterized by containing 5 inverted repeats. It may be significant that three inverted repeats were identified upstream of the putative aspartokinase gene (Fig. 2.2). A unique feature of the aspartate pathway in *B. subtilis* is that it also plays a critical role in sporulation during which it provides dipicolinate from an intermediate in lysine biosynthesis, a major constituent of bacterial heat resistant endospores. The finding of a putative aspartokinase gene adjacent to *glnA* is intriguing when we consider that nitrogen levels are also critical for sporulation of *C. acetobutylicum.*

The 0.86 kb *Eco*RV-*Xba*I fragment immediately downstream of *glnR* was cloned, generating plasmid pHS2 (Fig. 2.1), and sequenced in both directions. The entire region yielded an open reading frame transcribed in the same direction as *glnR*. The deduced amino acid product

showed striking homology to the large (α) subunit domains of glutamate synthases (GOGAT). The highest similarity (49% identity over 261 residues) was to the Fd-dependent GOGAT from the cyanobacteria *Synechocystis* (residues 28 to 286). In addition, sequence analysis of the last 335 bp of the pHZ200 insert, obtained from clone pHS3 (Fig. 2.1), confirmed that these homologies extended to the end of the pHZ200 insert. These preliminary DNA sequence data indicated that the ~2.4 kb region downstream of *glnR* encoded part of the *C. acetobutylicum* GOGAT α subunit. Since the genes encoding the α and β subunits of GOGAT have been found to be genetically linked in bacteria, with the α subunit generally encoded upstream of the β subunit, we employed the chromosome walking technique to isolate the region extending downstream of the pHZ200 insert.

2.4.2 Isolation of the chromosomal region extending downstream of *gln*R

Since the molecular mass estimates for the bacterial purified GOGAT protomer are in the range of 200 kDa, the region coding for the two subunits should be at least 6 kb in length. Isolation of the 4 kb chromosomal region extending downstream of the pHZ200 insert, was achieved in two stages. First, probe 1 (Fig. 2.1), derived from plasmid pHZ200, was used to identify and clone a ~3.8 kb *Xba*I fragment of *C. acetobutylicum* P262 genomic DNA (Fig. 2.3) by colony hybridization of a size selected *Xba*I genebank. This fragment contained an additional ~2.2 kb of downstream sequence. Of the 550 recombinants screened, three were positive. All three carried identical inserts, consistent with them overlapping with the 1.6 kb downstream region of pHZ200. Southern hybridization with one plasmid, designated pHS4 (Fig. 2.1), confirmed that the insert originated from *C. acetobutylicum* (results not shown), and was subsequently subjected to sequence analysis. This clone extended to the end of the α subunit region of bacterial GOGATs.

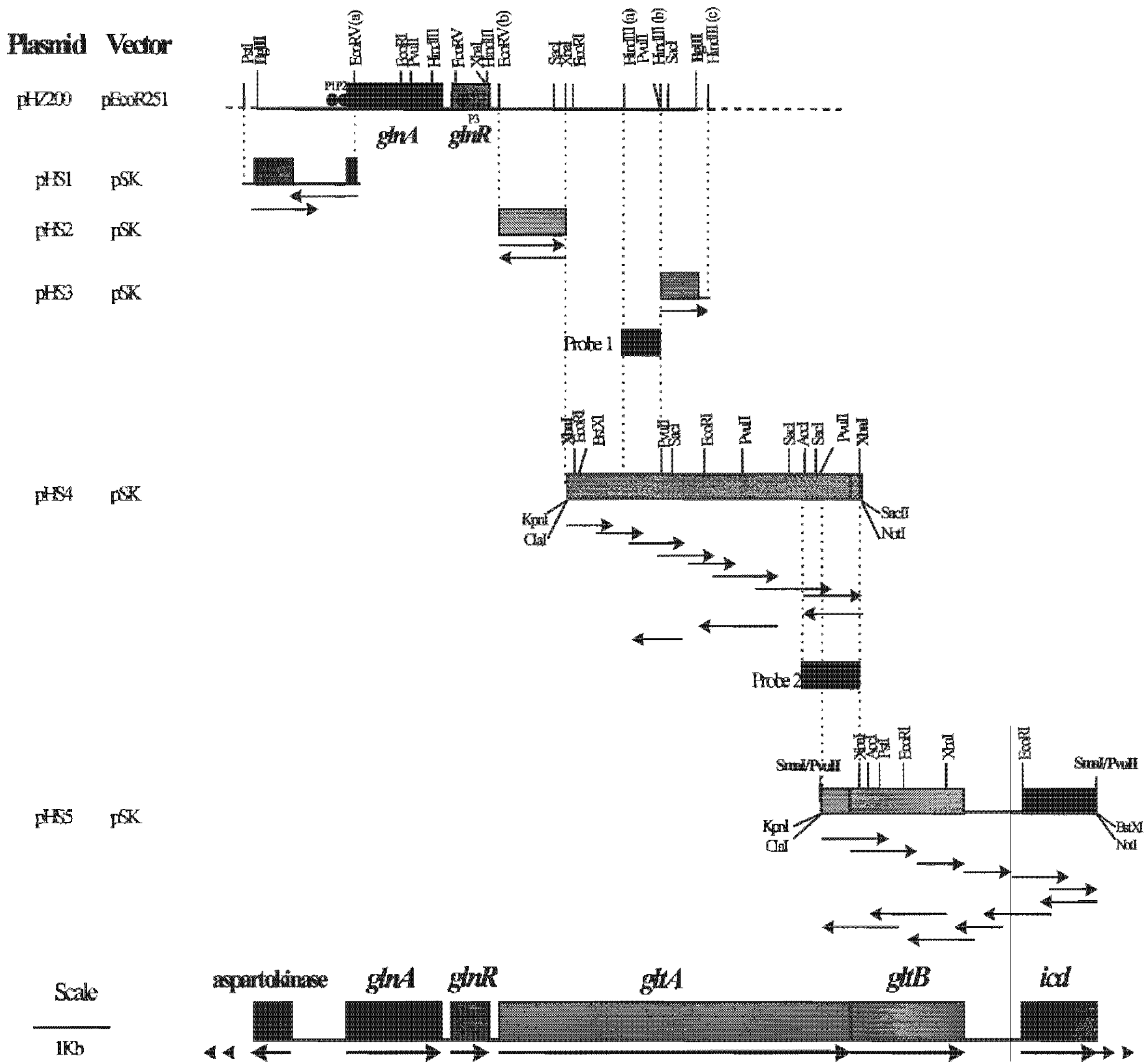


Fig. 2.1 Strategy for cloning and sequencing the region upstream of *glnA,* and downstream of *glnR* from *C. acetobutylicum* P262. Plasmid pHZ200 (Usden *et al.*, 1986) contains a 6.1 kb insert of chromosomal DNA cloned into the unique *Bgl*II site of pEcoR251. Plasmids pHS1, pHS2 and pHS3 were derived by cloning the 1.3 kb *Pst*I-*Eco*RV(a) fragment, the 0.86 kb *Eco*RV(b)-*Xba*I fragment, and the 0.56 kb *Hind*III(b)-*Hind*III(c) fragments of pHZ200, into the corresponding unique sites in the Bluescript vector pSK, respectively. Plasmids pHS4 and pHS5 were isolated consecutively by chromosome walking. Probe 1, a 0.38 kb *Hind*III fragment of pHZ200, was used to isolate pHS4, a Bluescript clone containing a 3.86 kb *Xba*I fragment. Probe 2, a 0.67 kb *Acc*I-*Xba*I fragment, derived from pHS4, was used to isolate pHS5, a Bluescript clone containing a 3.5 kb *Pvu*II fragment cloned into the unique *Sma*I vector site. The dashed horizontal lines represent vector sequence. Restriction enzyme sites are included on the appropriate plasmid maps. The unique 3' and 5' overhang restriction sites, used to generate nested deletions clones of pHS4 and pHS5 for sequencing, are included at the vector-insert junction. The thin arrows indicate the polarities and extent of sequencing of the clones. The relative positions of the identified genes, and their transcriptional polarities, are indicated below.

```
1      ATCTTCCTTTGTGTATTTTGCCATTTTACAACTCCCCCTTTACATTTTTAAAAAAAGCGCACAATAGACAAGTATAAACTTCTATCGAAC
       TAGAAGGAAACACATAAAACGGTAAAATGTTGAGGGGGAAATGTAAAAATTTTTTCGCGTGTTATCTGTTCATATTTGAAGATAGCTTG
        D  E  K  T  Y  K  A  M  <-GS   SD                        -10 -->         P2  <--  -35

91     GCCTTTGCTCAATTCATATAATAACTAAATTATACTTCCCTTTTTTTAAGAAATCAAGTGATTTTTTTAGTAAATAAATATTTTCGTAC
       CGGAAACGAGTTAAGTATATTATTGATTTAATATGAAGGGAAAAAAAATTCTTTAGTTCACTAAAAAAATCATTTATTTATAAAAGCATG
                                    -10          P1        -35

181    TAAAAACTTTACAATATAAATAATTTGATACGTTTTTCTTTAGATTCTAAAGTTATCAGGATTATTTTTCCATTTTAAAAATTTATAGTC

271    TCATTTTTTATTACAATATATATCCTTTATTATATAAATTTTATTTAACATTATACTATGCTCCATTCTGTAATTTTGTATATAATATTG

361    TAATGAAAAAATTTTATAAATAATGCATAAATTTACAAAAAAAACGAAGTCACGAGACGGCGTACCTCTATGACTTCGTTGTCAGCATAT

451    ATTTATCTGTACTTTATTATAAAAGGTATGCTAAATACTGTCAACTATCATATCTCTTAGTTCAATTTTAAGATATGCACAAAAAAAAAT
                                                             -35                    -10
541    TTTTATTCACAATATAATTCTACAATTAAATACTCTTTTATTATTAATAATTATCTTGACATTTTTAGATAAATTGCTACAATCTTATAT
                        SD
631    ATAAAATTTATAATATATTGGGGGCTATACTTATGAACACAATTATAACAAAATTTGGTGGCAGCTCATTAGCTGATGCCAATCAATTCA
1                 Aspartokinase-->M  N  T  I  I  T  K  F  G  G  S  S  L  A  D  A  N  Q  F

721    GAAAGGTTAAAGACATAATTTATTCTAATGATGCAAGAAAATATGTAATACCTTCTGCGCCTGGAAAAAGAGACTCTAAAGATTCAAAAG
20     R  K  V  K  D  I  I  Y  S  N  D  A  R  K  Y  V  I  P  S  A  P  G  K  R  D  S  K  D  S  K

811     TAACTGACTTATTATACCTTTGTCATGCTCATGTTGCTGCTGGAATCGCTTTAGATGATGTCTTTAATCACATAAGACAAAGATATTCTG
50     V  T  D  L  L  Y  L  C  H  A  H  V  A  A  G  I  A  L  D  D  V  F  N  H  I  R  Q  R  Y  S

901    ATATAATAAATGATTTAAAATTAGATTTTAGTATAGAAGATCAATTAAATACAATAAAGAAAGACCTTGAAGCTGGTGCATCAAGTGATT
80     D  I  I  N  D  L  K  L  D  F  S  I  E  D  Q  L  N  T  I  K  K  D  L  E  A  G  A  S  S  D

991    ACGCTGCAAGTAGAGGCGAATATTTAAATGGTTTAATTCTTGCGAAATATTTAGACTTTGAATTTGTTGATGCTAAAGATGTTATAGTTT
110    Y  A  A  S  R  G  E  Y  L  N  G  L  I  L  A  K  Y  L  D  F  E  F  V  D  A  K  D  V  I  V

1081   TTAAAAAGGATGGCTCATTAATAATGAATGCTACAAACTGCGCTCTACACAACAGATTATCTAATGTTTCTAAGGCTGTTATTCCTGGAT
140    F  K  K  D  G  S  L  I  M  N  A  T  N  C  A  L  H  N  R  L  S  N  V  S  K  A  V  I  P  G

1171   TTTATGGTGCTGATAAATCTGGTAACATTGTTACATTCTCAAGAGGTGGTTCTGATGTTACTGGAGCATTAGTTGCTGCAAGCATCAATG
170    F  Y  G  A  D  K  S  G  N  I  V  T  F  S  R  G  G  S  D  V  T  G  A  L  V  A  A  S  I  N

1261   CAAATCTTTATGAAAATTGGACCGATGTTTCTGGTTTCTTAATGGCA
200    A  N  L  Y  E  N  W  T  D  V  S  G  F  L  M  A
```

Fig. 2.2 Nucleotide and deduced amino acid (single letter code) sequence of the 1283 bp region upstream of the *C. acetobutylicum glnA* gene encoding GS. Putative promoter sequences (-35 and -10 regions), and a Shine-Dalgarno sequence (SD), are indicated in bold type and overlined for the putative truncated aspartokinase encoding gene. Inverted repeats are indicated by converging arrows. Also included are some of the *glnA* regulatory features (Janssen *et al.*, 1990); two promoter sequences (P1 and P2) and the Shine-Dalgarno sequence (SD).

A similar strategy was undertaken to further extend the region downstream from the pHS4 insert. Using probe 2 (Fig.2.1), a 3.5 kb *Pvu*II fragment (Fig. 2.3) was isolated from a size-selected genebank. This fragment was calculated to include an additional ~3 kb of downstream sequence. The detection of two fragments for the *Pvu*II digest (Fig. 2.3) is consistent with this restriction site occurring within the probe fragment, with the smaller band (1.13 kb) corresponding to the size predicted from the restriction map of pHS4. One clone containing the DNA of interest, was recovered from 300 recombinants screened, and the origin of the insert confirmed. This plasmid was designated pHS5 (Fig. 2.1), and subjected to sequence analysis.

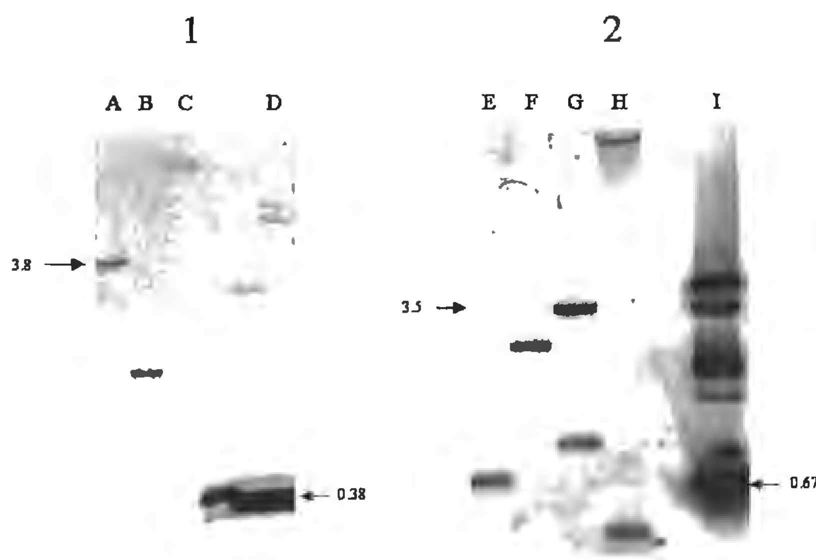


Fig.2.3 Southern blot analysis of digested *C. acetobutylicum* P262 chromosomal DNA using Probe 1 (panel 1) and Probe 2 (panel 2) (Probe 1 and 2 are described in Fig. 2.1). Lanes A, B, C, E, F,G and H represent 50 µg each of chromosomal DNA digested with *Xba*I, *Eco*RI, *Eco*RV, *Acc*I, *Eco*RI, *Pvu*II, and *Sac*I, respectively. Lane D represents 1 µg of *Hind*III-digested pHZ200, to release the 0.38 kb Probe 1 fragment. Lane I represents 1 µg of a shortened construct of pHS4 (containing the downstream 1.05 kb region), digested with *Acc*I and *Xba*I (incomplete) to release the 0.67 kb Probe 2 fragment. Sizes (kb) of the relevant chromosomal (left hand side) and plasmid (right hand side) bands are indicated.

2.4.3 Nucleotide sequence analysis of the 7.95 kb region downstream of *glnR*.

Sequence analysis of the 7.950 kb region extending downstream of *glnR* (from the *Eco*RV site of plasmid pHZ200), and presented in Fig. 2.4, revealed the presence of two major ORFs of 4554 bp (nt positions 195-4748) and 1473 bp (nt positions 4761-6233), which were present in the same orientation as *glnA* and *glnR*. Sequence comparisons of their deduced amino acid products identified them as the structural genes coding for the α and β subunits of GOGAT, respectively (Section 2.4.4 below). The overall base composition of the two genes (67.7% A+T for the α subunit and 64.6 % A+T for the β subunit) reflect the low G+C content characteristic of this bacterium, while the calculated codon usage confirms previous reports that codons containing an A or a U at the third position are preferred (Young *et al.*, 1989). The two open reading frames are separated by 12 bp and their arrangement is typical of that found in bacteria, in which the α subunit precedes the β subunit. Since *C. acetobutylicum* is more similar to *B. subtilis* than the enteric bacteria, we have chosen to designate the α subunit gene *gltA,* and the β

subunit gene *gltB,* as in *B. subtilis,* and will refer to their respective putative protein products as GltA and GltB.

Downstream from *gltB*, and separated by an intergenic region of 877 bases, a truncated ORF was identified extending to the end of the sequence. Its deduced product showed striking homology to the amino terminal regions of various isocitrate dehydrogenase (ICD) proteins, particularly from eukaryotes: 53% identity to *Schizosaccharomyces pombe* ICDH over 284 residues extending from nt 7111-7950 (Fig. 2.4). This is significant since isocitrate dehydrogenase catalyses the oxidative decarboxylation of isocitrate to form 2-oxoglutarate (and CO_2), the GOGAT substrate at the junction of carbon and nitrogen metabolism. No ORFs with homology to *gltF* of *E. coli* (Castano *et al.*, 1992), nor to any other sequences present in the databases, could be assigned to the large intergenic region downstream of *gltB,* however, three sets of short inverted repeats were identified. It remains to be established whether they are significant. Certainly an understanding of how the putative *icd* gene is regulated is sure to provide valuable insights into how carbon and nitrogen metabolism are interlinked in this Gram-positive bacterium.

Although transcription initiation sites have not been determined, possible promoter regions were identified upstream of the *gltA* and *icd* genes (Fig. 2.4), which showed homology to the clostridial extended consensus promoter sequences (Young *et al.*, 1989). The first promoter (TTGCTA-N_{18}-TACAAT), present from -140 to -111 upstream of the proposed *gltA* initiation codon, overlapped the carboxy terminal 10 residues of *glnR.* This suggests that the expression of *glnR* may compete with the expression of *gltA.* However, it is conceivable that the putative *gltAB* promoter may, in fact, not be recognizable if its transcription requires an alternative sigma factor, whose promoter recognition sequences have not yet been identified. Alternatively, the promoter may not be obvious if it is a weak promoter and requires a positive activator for expression. A putative promoter could not be identified upstream of the *gltB* gene. Two major inverted repeat sequences, with the potential to form stable stem-loop structures, were identified (Fig. 2.4). One of these structures (ΔG = -23.7 kcal/mol), was located 18 to 62 bp downstream of the *gltB* termination codon, and is consistent with the proposal that it acts as a transcriptional terminator. The possibility that the *gltA* and *gltB* genes are transcribed as an operon is investigated in Chapter 5. The second region of dyad symmetry (ΔG = -12.1 kcal/mol) was located between the putative *gltA* promoter and the *gltA* structural gene,

suggesting an operator-like structure. While it may act as a rho-independent terminator for the *glnR* gene, we speculate that it could play a role in influencing the transcription of the *glt* genes. As discussed in Chapter 1 (Section 1.4.3), a region of dyad symmetry was identified between promoter 1 and the *glnA* initiation codon. It was proposed that under certain metabolic conditions, the putative response regulator, GlnR, positively controls *glnA* expression via an antitermination mechanism (Woods and Reid, 1995). It is tempting to speculate that a similar mechanism may be involved in regulating the expression of the *glt* genes, especially since analysis of the sequence upstream of *gltA,* and within the *glnR* coding region, did not reveal any other features that could be attributed to transcriptional regulation of the *glt* genes. No sequence motifs characteristic of Lrp or LysR-type binding sites, (Section 1.5.9.1 and 1.5.9.2 respectively) could be identified, and there were no obviouse *cre* (catabolite responsive elements) or cAMP consensus binding sequences.

The assumption that the translation initiation sites for the *gltA, gltB* and truncated *icd* genes, are at the methionine residues located at nt positions 195, 4761 and 7111 respectively, is strongly supported by two findings: first, all three are preceded by conserved clostridial RBS (Young *et al.*, 1989), and second, the deduced amino acid sequences align with residues conserved in corresponding proteins (Section 2.4.4). The putative RBS are identical for all three genes, viz. 5'-AGGGGG, and are present 7, 8 and 7 bp upstream of the putative *gltA, gltB* and *icd* initiation codons, respectively (Fig. 2.4). This is identical to the sequence identified upstream of the *C. acetobutylicum* endo-β-1,4-glucanase gene (Zappe *et al.*, 1988). These observations imply that the TAG termination codon for the *gltA* gene overlaps the RBS for the *gltB* gene, and suggest that the two subunits are translationally coupled. This would ensure the production of equimolar amounts of both subunits. The distance between the termination codon for *gltA* and the translational start codon for *gltB* is 12 nucleotides. As pointed out for the *E. coli* enzyme (Oliver *et al.*, 1987), one could speculate that if a point mutation abolished this stop codon, it might be possible to obtain a fusion protein of approximately 217 kDa. This is interesting from an evolutionary point of view, since the eukaryotic pyridine-dependent GOGAT enzyme consists of a single polypeptide of approximately 220 kDa (Hummelt and Mora, 1980; Filetici *et al.*, 1996).

GlnR (continued)

```
                                                                         -35                   -10
1      CATTTAGATATATGCAGAAAATAAGCATGGATTCGGGGAAAAGAATGAAAGATATTGCTAGTTTAATATTAAGTGAAATACAATAAATTA
        F  R  Y  M  Q  K  I  S  M  D  S  G  K  R  M  K  D  I  A  S  L  I  L  S  E  I  Q  *
                                                EcoRV
91     TCAATTACAGCAGAATGCGAGTAATTATATCAAAAAAATAGATATATATTACTCGTTATTTTGCTATAAACAAATATATTATTAAATTTC
        SD
181    AAGGGGGAATTACAATGGAAAATAATATTCCGAATGCTCAAGGGTTATATGATCCATCTTTTGAGCATGATGCTTGTGGAATCGGAACTA
           GltA --> M  E  N  N  I  P  N  A  Q  G  L  Y  D  P  S  F  E  H  D  A  C  G  I  G  T

271    TAGTAAATATAGATGGTGAAAAATCACATGAAATTGTATCAGATGGTTTAACAATCCTAGAAAAATTAGAACATAGAGGCGGTACGGGAG
26      I  V  N  I  D  G  E  K  S  H  E  I  V  S  D  G  L  T  I  L  E  K  L  E  H  R  G  G  T  G

361    CTGATGAAAATACTGGAGATGGCGCTGGTATATTATTTAACATACCACATAAGTTCTTTAAGGAAGAGCTACAATCAAAAGGTATAACAC
56      A  D  E  N  T  G  D  G  A  G  I  L  F  N  I  P  H  K  F  F  K  E  E  L  Q  S  K  G  I  T

451    TTGAAGAAGAGGGAGACTATGCAGTTGCAATGATGTTCTTACCTCAAGATGAAAAGGCTAGAAAAGAAGCTGTCAGCCTTTTTGAAGATA
86      L  E  E  E  G  D  Y  A  V  A  M  M  F  L  P  Q  D  E  K  A  R  K  E  A  V  S  L  F  E  D

541    TATCAAAGGAAGAAGGACTTGAACTTATTGGGTGGAGAGAAGTTCAAACAAACCCTTCAATACTTGGAAAAGCATCTTTAGAAGCTATGC
116     I  S  K  E  E  G  L  E  L  I  G  W  R  E  V  Q  T  N  P  S  I  L  G  K  A  S  L  E  A  M

631    CATATATTATGCAGGCTTTTGTAAAGAGACCTAATGGCACAAAAAAAGAAAAAGATTTTGAAAGAAAATTATATATTGTTAGAAGAAACA
146     P  Y  I  M  Q  A  F  V  K  R  P  N  G  T  K  K  E  K  D  F  E  R  K  L  Y  I  V  R  R  N

721    TAGAAAAAAGAGCAGCGTGGATAAGCAAATTCTTAAATGAAACTTTTTACATAGCATCTTTTTCGTCTAAGACAATTGTATATAAGGGAA
176     I  E  K  R  A  A  W  I  S  K  F  L  N  E  T  F  Y  I  A  S  F  S  S  K  T  I  V  Y  K  G

811    TGCTTTTATCTACTCAATTAAGAGTATTTTACAAGGATTTAGAAGATGAAAGAGTTGAAACCTCTCTTGCATTAGTTCATTCAAGATATA
206     M  L  L  S  T  Q  L  R  V  F  Y  K  D  L  E  D  E  R  V  E  T  S  L  A  L  V  H  S  R  Y

901    GTACGAATACATTCCCAAGTTGGGAAAGAGCTCATCCAAATAGATTTTTGGTTCATAATGGCGAAATTAATACTTTAAGAGGAAATGTTA
236     S  T  N  T  F  P  S  W  E  R  A  H  P  N  R  F  L  V  H  N  G  E  I  N  T  L  R  G  N  V

991    ACAAGGTTTCTAGAGTGCTTGGAAAAGACGTAATTAGAGTATTACCTATCATAAATAAAGAAGGATCAGACTCAGCAATTTTTGATAATA
266     N  K  V  S  R  V  L  G  K  D  V  I  R  V  L  P  I  I  N  K  E  G  S  D  S  A  I  F  D  N

1081   ATTTAGAATTCTTATATATGAATGGTATGGATTTACCTAGAGCTGTAATGATGGCAATTCCAGAACCATGGTATAAGAGTAAAACTATGA
296     N  L  E  F  L  Y  M  N  G  M  D  L  P  R  A  V  M  M  A  I  P  E  P  W  Y  K  S  K  T  M

1171   GCAAAGAAAAAAGAGATTTTTATGAATATAACGCTACACTAATGGAACCATGGGACGGTCCAGCAGCTATAGTATTTACAGATGGTGAAA
326     S  K  E  K  R  D  F  Y  E  Y  N  A  T  L  M  E  P  W  D  G  P  A  A  I  V  F  T  D  G  E

1261   GAGTTGGTGCAGTTTTAGATAGAAATGGTTTAAGACCATCAAGATATTATATAACTAAAGATAGAAGACTTATACTTTCATCAGAAGTTG
356     R  V  G  A  V  L  D  R  N  G  L  R  P  S  R  Y  Y  I  T  K  D  R  R  L  I  L  S  S  E  V

1351   GAGCTTTAGATGTACCAGCAGAAATTGTTGAAAAGAAAGATAGATTAATGCCAGGTAGAATGTTACTTGTAGATACAGTTAAAAAGGAAG
386     G  A  L  D  V  P  A  E  I  V  E  K  K  D  R  L  M  P  G  R  M  L  L  V  D  T  V  K  K  E

1441   TTGTAAGTGATGAAGAACTAAAGGATACTTATGCAAAAGAAAATCCTTATGGTGAATGGTTAGAACAAAATTTAGTTACTTTAGATAAGA
416     V  V  S  D  E  E  L  K  D  T  Y  A  K  E  N  P  Y  G  E  W  L  E  Q  N  L  V  T  L  D  K

1531   TGACAGACAGCAAAACATTCACAATTGAATATGATAAAGACACTAGAAAGAGATTAGAAAAAACTTTTGGATATACTTATGAAGAAGTAA
446     M  T  D  S  K  T  F  T  I  E  Y  D  K  D  T  R  K  R  L  E  K  T  F  G  Y  T  Y  E  E  V

1621   ACACAACAATGCTTCCAATGGCTAAGACTGGTGCAGACCCATTAGGCAGCATGGGGGTAGATACTCCGGTAGCAGTATTATCTAAACAAT
476     N  T  T  M  L  P  M  A  K  T  G  A  D  P  L  G  S  M  G  V  D  T  P  V  A  V  L  S  K  Q

1711   CTCAACCATTATTCAATTACTTTAAGCAATTATTTGCGCAAGTAACTAATCCGCCTATAGATGCGATTAGAGAAGAAATAGTTACAAGCG
506     S  Q  P  L  F  N  Y  F  K  Q  L  F  A  Q  V  T  N  P  P  I  D  A  I  R  E  E  I  V  T  S

1801   TCTTAATGTTTATCTTGGGCCCAGAAGGGAATATTTTAGAAGATAAGTCTTCAAATTGTGAATTAATTAAGCTTGATTCACCAATAATCA
536     V  L  M  F  I  L  G  P  E  G  N  I  L  E  D  K  S  S  N  C  E  L  I  K  L  D  S  P  I  I

1891   ATAATGAAGAATTAGCGAAAATAAAAGAATACAATAAAGAAAATTTAAAACCAAAGGTAATAGATATTGTATTTGATAAAGGCGGATCTT
566     N  N  E  E  L  A  K  I  K  E  Y  N  K  E  N  L  K  P  K  V  I  D  I  V  F  D  K  G  G  S

1981   TAGAAGATGCATTACATGAAGTATTTGAAAAAGCACAAGAAGCATATGAAAAAGGATATACTATATTAATATTATCTGATAGAAATGTAT
596     L  E  D  A  L  H  E  V  F  E  K  A  Q  E  A  Y  E  K  G  Y  T  I  L  I  L  S  D  R  N  V

2071   GCGAAGCAAAGGTTCCAATTCCTTCATTACTTGCAGTATCTGCACTTCATCAATATCTAGTTCAAAAAGGAACAAGAACTTCAGTTGCAA
626     C  E  A  K  V  P  I  P  S  L  L  A  V  S  A  L  H  Q  Y  L  V  Q  K  G  T  R  T  S  V  A

2161   TAATATTAGAAAGTGGAGAACCAAGAGAAGTACATCATTTTGCTACTTTAATTGGATTTGGTGCATCAGCTGTAAATCCATATATGGCTT
656     I  I  L  E  S  G  E  P  R  E  V  H  H  F  A  T  L  I  G  F  G  A  S  A  V  N  P  Y  M  A
```

```
2251   ATGAAGCTTTAAGAGGATTAAGGGAAGAAGGATTACTTGAATTAGATTATGATAAGGCTGTTTATAATTATAACAAAGCAGTTCTTAAGG
686    Y  E  A  L  R  G  L  R  E  E  G  L  L  E  L  D  Y  D  K  A  V  Y  N  Y  N  K  A  V  L  K

2341   GAATAATAAAAATACTTTCAAAGATGGGTATTTCAACAATACAATCTTACCAAGGAGCTCAAATATTTGAGGCTATAGGTATAGGTAAAG
716    G  I  I  K  I  L  S  K  M  G  I  S  T  I  Q  S  Y  Q  G  A  Q  I  F  E  A  I  G  I  G  K

2431   AAGTTATTGAAAAATATTTTACTAATACCGTAAGCAGAATTGGTGGAATAGGATTAAAAGAAATTCAAAGAGAAGCAGAAATAAATCACG
746    E  V  I  E  K  Y  F  T  N  T  V  S  R  I  G  G  I  G  L  K  E  I  Q  R  E  A  E  I  N  H

2521   AAAAAGGATTTAATGATAAGACTTATGCTGCTGATTTTACATTAGATTCACCAGGATATGAAAAACTTAGATCTGGTGAGAATGGAGCAG
776    E  K  G  F  N  D  K  T  Y  A  A  D  F  T  L  D  S  P  G  Y  E  K  L  R  S  G  E  N  G  A

2611   AAGAGCATTTATATAATCCTTTAACAATTCATAAACTTCAAGAATCTACAAAGACAGGAAATTATGAATTATTTAAAGAATACACTTCTT
806    E  E  H  L  Y  N  P  L  T  I  H  K  L  Q  E  S  T  K  T  G  N  Y  E  L  F  K  E  Y  T  S

2701   TAATTGATAAAGAAGAAGCTGAAATAAATTTAAGAGGATTATTAGAATTCAACTATAATTCTAAAGAAATTCCAATAGAGGAAGTTGAAT
836    L  I  D  K  E  E  A  E  I  N  L  R  G  L  L  E  F  N  Y  N  S  K  E  I  P  I  E  E  V  E

2791   CAGTTTCAGAGATAGTTAAGAGATTTAAAACAGGAGCTATGTCTTATGGATCAATTTCTAAAGAAGCTCATGAAGCTTTAGCTATTGCAA
866    S  V  S  E  I  V  K  R  F  K  T  G  A  M  S  Y  G  S  I  S  K  E  A  H  E  A  L  A  I  A

2881   TGAACAGAATCGGTGGTAAATCTAACACTGGTGAAGGTGGAGAAGATAAAGAAAGATGGACTTTAGATGCAAATGGAGATTCAAGAAGAT
896    M  N  R  I  G  G  K  S  N  T  G  E  G  G  E  D  K  E  R  W  T  L  D  A  N  G  D  S  R  R

2971   CTTCAATAAAACAAATTGCATCTGGTAGATTTGGAGTAACTTCAGAATATCTAGTTAATGCAGATGAACTTCAAATTAAATTAGCACAAG
926    S  S  I  K  Q  I  A  S  G  R  F  G  V  T  S  E  Y  L  V  N  A  D  E  L  Q  I  K  L  A  Q

3061   GAGCAAAACCAGGAGAAGGTGGTCAATTACCTGCAACTAAGGTTTATCCATGGATAGCTAAGACTAGACATTCAACTACAGGGGTTGGGT
956    G  A  K  P  G  E  G  G  Q  L  P  A  T  K  V  Y  P  W  I  A  K  T  R  H  S  T  T  G  V  G

3151   TAATATCGCCACCACCACATCATGATATTTACTCAATAGAAGATTTAGCACAATTAATATATGATTTAAAGAATGCTAATACAGGTGCTA
986    L  I  S  P  P  P  H  H  D  I  Y  S  I  E  D  L  A  Q  L  I  Y  D  L  K  N  A  N  T  G  A

3241   GAGTATCTGTTAAGCTAGTTTCTGAATGTGGAGTTGGTACTGTTGCAGCTGGAGTTGCTAAAGGTGGAGCAGAAGTTATATTAATTTCAG
1016   R  V  S  V  K  L  V  S  E  C  G  V  G  T  V  A  A  G  V  A  K  G  G  A  E  V  I  L  I  S

3331   GATATGATGGAGGGACAGGTGCATCACCTAAGAACTCTATTAAGAATGCAGGACTTCCTTGGGAACTTGGACTTGCAGAAGCACATCAAA
1046   G  Y  D  G  G  T  G  A  S  P  K  N  S  I  K  N  A  G  L  P  W  E  L  G  L  A  E  A  H  Q

3421   CATTACTTCTTAATGATTTGAGAGAAAGAGTTAGAGTTGAAGTTGATGGTAAGCTTATGAGTGGTCGTGATGTTGCTGTTGCTGCACTTC
1076   T  L  L  L  N  D  L  R  E  R  V  R  V  E  V  D  G  K  L  M  S  G  R  D  V  A  V  A  A  L

3511   TTGGAGCAGAAGAATTTGGATTTGCAACTGCACCATTAGTAACTTTAGGCTGTGTAATGATGAGAGTTTGTAACTTAGACACTTGCCCAG
1106   L  G  A  E  E  F  G  F  A  T  A  P  L  V  T  L  G  C  V  M  M  R  V  C  N  L  D  T  C  P

3601   TTGGTGTTGCGACTCAAAATGAAGAATTAAGAAAAAGATTTAAAGGAAACCAGAATATGTTGTTAACTTTATGTATTTATAGCACAAGAA
1136   V  G  V  A  T  Q  N  E  E  L  R  K  R  F  K  G  N  Q  N  M  L  L  T  L  C  I  Y  S  T  R

3691   TTAGAGAAATCATGGCTAAGCTTGGATTTAGAAAGCTTGATGAAATGATTGGTAGAGTAGATAAACTTAAACAAAAAGAAAATATTCATG
1166   I  R  E  I  M  A  K  L  G  F  R  K  L  D  E  M  I  G  R  V  D  K  L  K  Q  K  E  N  I  H

3781   GTTGGAAAGCTAAGAACGTTGATTTAAGTGCGGTGCTTTATACTCCAGACAAATATAAAGGTAAAGTAGTTAAATTTGATGAAACTAAGA
1196   G  W  K  A  K  N  V  D  L  S  A  V  L  Y  T  P  D  K  Y  K  G  K  V  V  K  F  D  E  T  K

3871   AATATGACTTTAAATTAAACAAAGTAATAGATGAAAAAATATTCTTAGATAAATTTAAAGATGCCATAGAAAATAAAATAAAGACTAATT
1226   K  Y  D  F  K  L  N  K  V  I  D  E  K  I  F  L  D  K  F  K  D  A  I  E  N  K  I  K  T  N

3961   TTGAAATTGATGTAACTAATACTGACAGAGCTCTTGGAACTATACTAGGATCAGAAATAACTAGAGTTAATGGAACTGACGGATTACCAG
1256   F  E  I  D  V  T  N  T  D  R  A  L  G  T  I  L  G  S  E  I  T  R  V  N  G  T  D  G  L  P

4051   AAGATACTATAAGTATAAAATGTAATGGTGCGGCAGGACAAAGTTTTGGAGCATTTATTCCAAAAGGATTAACTTTTGAAGTTGAAGGAG
1286   E  D  T  I  S  I  K  C  N  G  A  A  G  Q  S  F  G  A  F  I  P  K  G  L  T  F  E  V  E  G

4141   ATGCTAATGATTATTTTGGTAAGGGATTATCAGGTGGTAAGCTTATAGTATACCCTCCAAAGAAATCTACATTTATTGCAGAAGATAACA
1316   D  A  N  D  Y  F  G  K  G  L  S  G  G  K  L  I  V  Y  P  P  K  K  S  T  F  I  A  E  D  N

4231   TCTTAATTGGAAATGTTGCTTTATACGGTGCAACTTCAGGAAAAGTATTTATAAACGGTATTGCTGGGGAAAGATTCTGCGTTAGAAATT
1346   I  L  I  G  N  V  A  L  Y  G  A  T  S  G  K  V  F  I  N  G  I  A  G  E  R  F  C  V  R  N

4321   CAGGTGCAACTGCTGTTGTTGAAGGAGTAGGAGCTCATGGATTAGAATATATGACTGGTGGTAAAGTTGTTGTTCTAGGTAAAACAGGAA
1376   S  G  A  T  A  V  V  E  G  V  G  A  H  G  L  E  Y  M  T  G  G  K  V  V  V  L  G  K  T  G

4411   TTAATTTTGCAGCTGGTATGAGTGGTGGAGTTGCTTACATTTATGAGGAAGATCCAAACTTCAGAATAAACTTAAATGAAGAAATGATTT
1406   I  N  F  A  A  G  M  S  G  G  V  A  Y  I  Y  E  E  D  P  N  F  R  I  N  L  N  E  E  M  I

4501   TACTAGAAGAGTTAAACATAGATGATGAAGAAGAATTGAAAGCTCTAATTGAAGAACATGTAAAAGTTACTGGCTCTCCTAAAGCAAATA
```

```
1436     L  L  E  E  L  N  I  D  D  E  E  E  L  K  A  L  I  E  E  H  V  K  V  T  G  S  P  K  A  N

4591     AAATATTATATAACTTTGAAACAGAAAAAGTTAAGTTCCATAAAATAATTCCTAAGGATTATAAGAAGGTATTAGAAACTGTTGAAAAGT
1466     K  I  L  Y  N  F  E  T  E  K  V  K  F  H  K  I  I  P  K  D  Y  K  K  V  L  E  T  V  E  K
                                                                                       SD
4681     ATAAGAATCTTGGATCTGATGAAGAAGAAGCGTTAATTAAGACTTTCCAAGAAATTAAAGGAATATAGGGGGCTAGAAGAATGGGTAAAC
1496/1   Y  K  N  L  G  S  D  E  E  E  A  L  I  K  T  F  Q  E  I  K  G  I  *   GltB →   M  G  K

4771     CAACTGGATTTTTAGAGTATGATAGAGAAGTAGGCAGAAATAGAGAACCTAAAGAAAGATTGAAAGATTATAAAGAGTTTCATCAAAGGC
4        P  T  G  F  L  E  Y  D  R  E  V  G  R  N  R  E  P  K  E  R  L  K  D  Y  K  E  F  H  Q  R

4861     TTCCTCTAGAAAAGCAATGCATTCAGGGCGCAAGATGTATGGATTGTGGTGTACCATTTTGCCAAGCTGGAGTATTATTTTCAGGTATGG
34       L  P  L  E  K  Q  C  I  Q  G  A  R  C  M  D  C  G  V  P  F  C  Q  A  G  V  L  F  S  G  M

4951     TATCAGGATGTCCGCTTCACAATTTAATTCCTGAATGGAACGATTTAGTATACAGAGGAAAATGGGAATTAGCTTATGAAAGATTAAATA
64       V  S  G  C  P  L  H  N  L  I  P  E  W  N  D  L  V  Y  R  G  K  W  E  L  A  Y  E  R  L  N

5041     AGACTAATCCATTCCCAGAATTTACAGGAAGAGTATGTCCAGCACCTTGTGAAGCTGGATGTACTGCAGGATTGAATGGTCCAGCAATAA
94       K  T  N  P  F  P  E  F  T  G  R  V  C  P  A  P  C  E  A  G  C  T  A  G  L  N  G  P  A  I

5131     CTATTAAGGAAAATGAAAGATCTATAATAGATAATGCATTTGAAAATGGATTTGTTAAAGCAAATACACCTCTTAAGAGAACAGGAAAGA
124      T  I  K  E  N  E  R  S  I  I  D  N  A  F  E  N  G  F  V  K  A  N  T  P  L  K  R  T  G  K

5221     AAGTTGCAGTAATTGGTTCTGGTCCATCGGGATTAGCGGTTGCTAATACGTTAAATAAATGTGGTCATAAGGTTACAGTATTTGAAAGAA
154      K  V  A  V  I  G  S  G  P  S  G  L  A  V  A  N  T  L  N  K  C  G  H  K  V  T  V  F  E  R

5311     GTGATAGACCAGGTGGACTATTAATGTATGGAATTCCTAATATGAAACTTGATAAAGAAGTTATTTTAAGAAGAGTTCACCTTATGGCGG
184      S  D  R  P  G  G  L  L  M  Y  G  I  P  N  M  K  L  D  K  E  V  I  L  R  R  V  H  L  M  A

5401     AAGAAGGCATAAAATTTGTAAACAATGCAAATGTTGGTGACAACTATGATGCAAAGGAAATTTTAGATGAATTTGATGCAGTAGTTCTTG
214      E  E  G  I  K  F  V  N  N  A  N  V  G  D  N  Y  D  A  K  E  I  L  D  E  F  D  A  V  V  L

5491     CGACAGGAGCATCTCAGCCTAGAGATCTTCAAGCTGAAGGAAGAGATAATGTTAATGGTATCCATTTTGCTGTAGATTTCTTAAAAGCTA
244      A  T  G  A  S  Q  P  R  D  L  Q  A  E  G  R  D  N  V  N  G  I  H  F  A  V  D  F  L  K  A

5581     ATACTAAGAGTCTTTTAGATTCAGATCATGAAGATAATAATTACATTTCTGCAAAAGATAAAAATGTAATTATTATTGGAGGTGGAGACA
274      N  T  K  S  L  L  D  S  D  H  E  D  N  N  Y  I  S  A  K  D  K  N  V  I  I  I  G  G  G  D

5671     CAGGAACAGACTGCGTTGCAACATCTCTTAGACATGGATGTAAATCAGTAGTTCAATTTGAAATAATGGGTGAACCAGCTCACGAAAGAA
304      T  G  T  D  C  V  A  T  S  L  R  H  G  C  K  S  V  V  Q  F  E  I  M  G  E  P  A  H  E  R

5761     CTGAAAGCAACCCCATGCCAGAATGGCCTAAAGTTCTTAAGGTTGATTATGGCCAAGAAGAATTTATAGAACTATATGGAAAAGATCCTA
334      T  E  S  N  P  M  P  E  W  P  K  V  L  K  V  D  Y  G  Q  E  E  F  I  E  L  Y  G  K  D  P

5851     GAGAATATTTAACAACAGTTACAGCAGTTCATTCAGATAAAGATGGAAATCTAGAAAGCGTAGATACTGTAAAAGTTGAATGGAAAAAAG
364      R  E  Y  L  T  T  V  T  A  V  H  S  D  K  D  G  N  L  E  S  V  D  T  V  K  V  E  W  K  K

5941     GTGACAATGGAAGAATGTTCCCAGCTCATGTTGAAGGTTCGGAAAAGAACTGGCCAGCAGAACTTATATTACTTGCAATGGGATTCACAG
394      G  D  N  G  R  M  F  P  A  H  V  E  G  S  E  K  N  W  P  A  E  L  I  L  L  A  M  G  F  T

6031     GTTCAGAAGATTATCTTAAAAATGCCTTTGGTATTGAATCAGATGAAAGAAGCAATGTAAAAGCTGGTAATGTTGACTTTATGACTAATG
424      G  S  E  D  Y  L  K  N  A  F  G  I  E  S  D  E  R  S  N  V  K  A  G  N  V  D  F  M  T  N

6121     TACCGAAAGTATTTGCAACAGGTGATGCAAGACGTGGTCAATCATTAGTTGTAACAGCTATAAGTGAAGGTATAGCAGCAGGAATAGCAG
454      V  P  K  V  F  A  T  G  D  A  R  R  G  Q  S  L  V  V  T  A  I  S  E  G  I  A  A  G  I  A

6211     TTGATAAATTCTTAAGAGCTTAAGGCGCATTCAGAAAAATAAAGAAAAAGTGGTGTTGAAAAATTTTTAATACCACTTTTTCTTTTATAT
484      V  D  K  F  L  R  A  *

6301     TGAATAGTATTACATATATTCAAAGGATAATAAATTAAGGAACAATCAACATTGTACATTAAGGCAGTGTTAATGTATAATAGAATAAGG

6391     TACATGAAATTAATGAACGCAGAGGTTGATTATGATGAATAAAGAATTTAAAAGATATATATATATAGCAGGAACAATTTGTGTAATATT

6481     TCTAGCAGGTGCTTTATTTTTGCATCTTCTTCCGTTCCTTTTAATGCTTGGAGTAATAGCTTATATTGGAATAAGAATTAACAGGTTTAT

6571     TAAGAAGAAGAAAGCTGAAAAAAGTAATAGCCAATCTCAATATAATCATAATTATAAAGATGAAAATAATTATGAGACTTCACCAGAAGA

6661     ATATACAACCGGAGAAGTTATTGATGTCGAGTATGAGGACGTAGATAATAAATAAAATAACACATAAGATTTAATATGAATTTAAACTCA

6751     GTTCAATTGACTGAGTTTTTTTGTTGATTAATTAATCTACTAACTGTATATATGGTGCAATATGATTTATAAATTTATTTAGTTATACGG

6841     TACATAGTTTTAATGTAGATTTCAATTTTTAACATATATATATTAATAAAAAATGCGATTTGGGTAATTTGAGATTTTGCAATAAAAATA

6931     TAAGAGTGTTGCAGTAGTTGAGCTTATTTTGGCGTTTAAATGAAAATAAAACTATAGAATTCTTTGATTTTGCAAGAGTGAAAATATAAT
                        -35                    -10                                       SD
7021     CATGCAAAAAAAGTTGATAAAACTTAGACGGTACTGTATAATCATTACTATAAGTTTAAAATTTAAATAGCAATTTAAGGGGGCAAAAAA
```

```
7111        ATGGAAAAAATAAAAATGTCTAATCCAATAGTTGAAATGGATGGAGATGAAATGACAAGAATAGTTTGGGGCGAAATAAAAAAGGAATTA
1   ICD---> M  E  K  I  K  M  S  N  P  I  V  E  M  D  G  D  E  M  T  R  I  V  W  G  E  I  K  K  E  L

7201        TTAAATCCTTTTATTGAATTAAATACAGAATATTATGATCTTGGATTAGAGCATAGAAATAATACTAATGATCAAGTTACTGTTGATGCA
31          L  N  P  F  I  E  L  N  T  E  Y  Y  D  L  G  L  E  H  R  N  N  T  N  D  Q  V  T  V  D  A

7291        GCGGAAGCTATTAAAAAGCATAAAGTAGGAGTAAAGTGTGCAACTATTACACCTAATGGGGCAAGAATGAAAGAATATAATCTAAAAGAA
61          A  E  A  I  K  K  H  K  V  G  V  K  C  A  T  I  T  P  N  G  A  R  M  K  E  Y  N  L  K  E

7381        ATGTGGAAAAGTCCTAATGGTACTATTAGAGCGATATTAGATGGTACTGTTTTTAGAATGCCAATAACTGTGGACTGCGTAAAGCCATAT
91          M  W  K  S  P  N  G  T  I  R  A  I  L  D  G  T  V  F  R  M  P  I  T  V  D  C  V  K  P  Y

7471        GTTAGATCATGGGAAAGTCCAATTACAATAGCAAGACATGCTTATGGGGATATATATAAGGCTGTAGAAATGAAGGTAGAAGGAAAAAGT
121         V  R  S  W  E  S  P  I  T  I  A  R  H  A  Y  G  D  I  Y  K  A  V  E  M  K  V  E  G  K  S

7561         AAATGTGAACTTGTACTTACTACTGAAAACGGTGAAGAAAAAAGGGAATTAATACACAACTTTTCAGATGATGGTGTTGTTATGGGGATG
151         K  C  E  L  V  L  T  T  E  N  G  E  E  K  R  E  L  I  H  N  F  S  D  D  G  V  V  M  G  M

7651        CATAATTTAAACAAATCCATAGAAAGCTTTGCAAGAAGCTGTTTTAGTTTTGCATTAGATGTTAAGCAGGATTTATGGTTTGCAAGTAAA
181         H  N  L  N  K  S  I  E  S  F  A  R  S  C  F  S  F  A  L  D  V  K  Q  D  L  W  F  A  S  K

7741        GATACAATATCAAAAAAATATGATCATACTTTTAAAGATGTATTTCAAGAATTATATGATAATGAATATAAAACAAAATTTGAAGAGGCA
211         D  T  I  S  K  K  Y  D  H  T  F  K  D  V  F  Q  E  L  Y  D  N  E  Y  K  T  K  F  E  E  A

7831        GGAATTAAATATATATATACATTAATAGACGCTGCCGTAGCGAATATTATGAAATGTAAGGGTGGAATAATCTGGGCTTGTAAAAATTAT
241         G  I  K  Y  I  Y  T  L  I  D  A  A  V  A  N  I  M  K  C  K  G  G  I  I  W  A  C  K  N  Y

7921        GATGGTGATGTAATGAGTGATATGGTAGCA
271         D  G  D  V  M  S  D  M  V  A
```

Fig. 2.4 Nucleotide and deduced amino acid (single letter code) sequence of the *C. acetobutylicum gltA*, *gltB* and truncated *icd* genes. Putative promoters (-35 and -10 regions) and ribosome-binding sites (SD) are indicated in bold type and overlined. Stop codons are marked with an asterisk. Converging arrows highlight inverted repeat sequences. For completeness, 133 bp of sequence including the end of the *glnR* gene (Woods and Reid, 1995), has been included upstream of the *Eco*RV restriction site, which marked the start of the region sequenced in this project.

2.4.4 Protein analysis of the deduced GltA and GltB products

The *gltA* and *gltB* genes code for proteins of 1517 amino acid residues (GltA) and 490 amino acid residues (GltB), with predicted molecular weights of 168.5 kDa and 53.8 kDa, respectively. Comparative sequence analyses of the deduced GltA and GltB proteins revealed striking homology to GOGAT proteins from both prokaryotic and eukaryotic organisms (Table 2.1 and 2.2), and clearly showed that both subunits shared the extensive, highly conserved domains found in pyridine- or Fd-dependent forms of the enzyme (Appendix C and Fig 2.6). Specifically, GltA was more similar to the *B. subtilis* α subunit protein (GltA) (56.4% identity), and to the cyanobacterial proteins; the deduced Fd-dependent GltB polypeptide from *Synechocystis* (56.0% identity), and the NADH-dependent α subunit (GltB) from *Plectonema boryanum* (55.2% identity), than to the α subunit homologue identified in the genome sequence of *C. acetobutylicum* ATCC824 (51.9% identity). Interestingly, a second Fd-dependent form of the enzyme has been identified in *Synechocystis* (*gltS*) (Navarro *et al.*, 1995), but homology to this deduced protein, and to the Fd-protein (GlsF) from *P. boryanum,* is noticeably lower

(45.7% identity). For the remainder, the *C. acetobutylicum* P262 α subunit generally appears more homologous to the fused NADH polypeptide from yeast and plants, and to the Fd-dependent forms of the enzyme, than to the other bacterial α subunits (Table 2.1). Overall, these results suggest that the α subunit domain is well conserved along the evolutionary scale, irrespective of whether it exists as a fused polypeptide (as in fungi and plants), or whether it is dependent on pyridine or ferredoxin as its electron donor. Indeed, Navarro *et al.* (1995) found that the deduced GltB protein from *Synechocystis,* was more similar to alfalfa NADH-GOGAT than to corresponding Fd-GOGATs from higher plants, the red alga *Antithamnion*, or to NADPH-dependent enzymes from bacteria.

Homologies to the *C. acetobutylicum* β subunit, GltB, were less conserved, ranging from 56.6% identity to 35.1% identity (Table 2.2). Again, β subunit proteins from *P. boryanum* and *B. subtilis* showed the highest homology (56.6% and 53.4% identity, respectively), together with the β subunit domain from *M. sativa* (53.9% identity). The *C. acetobutylicum* ATCC824 β subunit homologue shared 51.2% identity. Furthermore, the fused polypeptide from yeast was more homologous to GltB than the remainder of the bacterial β subunits. As expected, no regions of similarity were found between GltB and the single polypeptide chain of Fd-dependent enzymes. Interestingly, GltB showed the highest homology scores (58.8% identity) to the deduced GltD polypeptide identified in *Synechocystis* (accession number dbj:D90900), yet the gene encoding this protein (*gltD*) does not lie adjacent to a α subunit gene, criteria we used to define GOGAT α and β subunits. This raises the question of whether GOGAT activity can be constituted from the separate expression of the two subunit genes. The *Synechocystis* *gltD* gene belongs to an interesting class of β subunit-like genes, which includes the gene encoding the *Pyrococcus* KOD1 GOGAT protein (GltA), that will be discussed in detail in Chapter 3.

TABLE 2.1 Comparison of the deduced *C. acetobutylicum* P262 GltA protein sequence to known pyridine (Py)- and ferredoxin (Fd)-dependent GOGAT sequences. Sequences were compared between the conserved amino-terminal cystein residue, and the carboxy-terminal proline residue, corresponding to Cys21 and Pro1483 of the *C. acetobutylicum* GltA polypeptide (Appendix C).

Organism	**Name**	**Class of Protein and designated Name**	**Abbre-viation**	**Accession number**	***C. acetobutylicum* GltA**	
		Py-dependent α subunit			**%ID**	**%SIM**
Bacteria	*Bacillus subtilis*	GltA	Bsub-L	emb:Z99113	56.4	65.7
Bacteria	*C. acetobutylicum* ATCC824	(GltA)[a]	Ca824-L	gb:AE001438[b]	51.9	60.5
Bacteria	*Pseudomonas aeruginosa*	GltB[a]		gb:U81261	45.5	55.0
Bacteria	*Azospirillum brasilense*	GltB	Abra-L	sp:Q05755	45.0	54.4
Bacteria	*Rhodobacter sphaeroides*	GltB		emb:Y12482	44.8	54.1
Bacteria	*Escherichia coli*	GltB	Ecol-L	sp:PO9831	44.7	55.2
Cyanobacteria	*Plectonema boryanum*	GltB	Pbor-L	dbj:D85230	55.2	65.3
		Fd-dependent enzymes				
Cyanobacteria	*Synechocystis* sp. PCC6803	GltB	Syne-F	sp:P55037	56.0	65.8
Cyanobacteria	*Synechocystis* sp. PCC6803	GltS		emb:X92480	45.7	55.2
Cyanobacteria	*Plectonema boryanum*	GlsF		dbj:D85735	45.7	55.9
Eukaryota	*Antithamnion*	GlsF		sp:Q06434	44.7	54.3
Eukaryota	*Zea mays*		Zmay-F	sp:P23225	47.0	56.4
Eukaryota	*Arabidopsis thaliana*			emb:Y09667	46.1	55.5
		Py-dependent α subunit domain				
Eukaryota	*Saccharomyces cerevisiae*	GLT1	Scer-L'	sp:Q12680	47.9	58.9
Eukaryota	*Medicago sativa*		Msat-L'	sp:Q03460	51.9	61.0

The sequences can be retrieved from the GenBank (gb), EMBL (emb), SwissProt (sp) and DDBJ (dbj) databases using the accession numbers supplied.

a: The identity of these deduced proteins is based on sequence homology alone.

b: This GenBank accession number references the unfinished genome sequence of *C. acetobutylicum* ATCC824.

TABLE 2.2 Comparison of the deduced *C. acetobutylicum* P262 GltB protein sequence to known pyridine (Py)-dependent GOGAT sequences.

Organism	Name	Class of Protein and designated Name	Abbre-viation	Accession number	*C. acetobutylicum* GltB	
		Py-dependent β subunit			**%ID**	**%SIM**
Bacteria	*Bacillus subtilis*	GltB	Bsub-S	emb:Z99113	53.4	62.3
Bacteria	*C. acetobutylicum* ATCC824	(GltB)[a]	Ca824-S	gb:AE001438[b]	51.2	35.1
Bacteria	*Azospirillum brasilense*	GltD	Abra-S	sp:Q05756	39.7	49.3
Bacteria	*Escherichia coli*	GltD	Ecol-S	sp:PO9832	38.4	48.3
Bacteria	*Pseudomonas aeruginosa*	GltD[a]		gb:U81261	37.8	48.1
Bacteria	*Rhodobacter sphaeroides*	GltD		emb:Y12481	36.8	48.0
Bacteria	*Thiobacillus ferrooxidans*	GltD	Tfer-S	pir:JC5184	35.1	47.4
Cyanobacteria	*Plectonema boryanum*	GltD	Pbor-S	dbj:D85230	56.6	65.9
		Py-dependent β subunit domain				
Eukaryota	*Saccharomyces cerevisiae*	GLT1	Scer-S'	sp:Q12680	47.6	56.6
Eukaryota	*Medicago sativa*		Msat-S'	sp:Q03460	53.9	63.9

The sequences can be retrieved from the GenBank (gb), EMBL (emb), SwissProt (sp), DDBJ (dbj) and PIR (pir) databases using the accession numbers supplied.

a: The identity of these deduced proteins is based on sequence homology alone.

b: This GenBank accession number references the unfinished genome sequence of *C. acetobutylicum* ATCC824.

2.4.5 Sequence identification of functional domains in the deduced *C. acetobutylicum* GltA and GltB proteins

2.4.5.1 Glutamine amidotransferase domain

Sequence analyses revealed that the N-terminal ~390 residues of the mature bacterial α subunit of GOGAT was similar to the glutamine amidotransferase (GAT) domain of Pur-F type amidotransferases (provided allowances are made for a ~120 residue insert in GOGAT sequences) (Vanoni and Curti, 1999). Indeed, the N-terminal domain of the deduced *C. acetobutylicum* GltA protein contains all eleven invariant amino acids identified by Mei and Zalkin (1989) and Kim *et al.* (1996) to be involved in glutamine binding and catalysis in PurF-type glutamine amidotransferases (Fig. 2.5a). These residues, which were strictly conserved in all the GOGAT sequences aligned, correspond to positions Cys21, Gly22, Arg51, Gly52, Asp62, Arg 234, His254, Asn255, Gly256, Asp289, and Gly365 in the *C. acetobutylicum* GltA

protein (Fig. 2.5a). In addition, residues involved in the formation of the glutamine binding site in the *E. coli* PRPP-AT (Vanoni and Curti, 1999), were also conserved in GltA. These analyses substantiate the classification of the GOGAT class of enzymes as PurF-type amidotransferases, and imply that the first 20 residues of the *C. acetobutylicum* GltA are post-translationally removed to expose the conserved Cys21 as the N-terminal nucleophile of the mature α subunit peptide. Although autocatalytic excision of a propeptide to expose the N-terminal nucleophile has been proposed to occur in Pur-F-type hydrolases (Vanoni and Curti, 1999), nothing is known about post-translational processing in GOGAT enzymes.

2.4.5.2 FMN-binding domain

As discussed in Chapter 1, the FMN binding domain was shown to reside in the bacterial α subunit. Analysis of bacterial and plant GOGATs (Sakakibara *et al.*, 1991; Gregerson *et al.*, 1993; Pelanda *et al.*, 1993) revealed that they contained a central region within the α subunit domain which was very similar to the carboxy-terminal part of the FMN-binding domain of the flavocytochrome b2 family, rather than to bacterial flavodoxins. Fig. 2.5b shows that this region is also conserved in GltA (residues 1065 to 1119). In addition, Asp-419 of yeast flavocytochrome b2, which interacts with the ribityl side chain of FMN (Pelanda *et al.*, 1993), is conserved in *C. acetobutylicum* GltA (Asp-1091). These features, together with the glycine-rich nature of the extended highly conserved region (see also Appendix C) which is usually indicative of a characteristic conformation of the protein, supports its involvement in FMN binding as proposed for other GOGATs (Vanoni and Curti, 1999).

A

Maturation site

```
Pbor-L    ......MIMTRYGLPAKQG   QF       VFIVQMK QP HS  QQALTI ANLE    ACGAETN     66
Syne-F    FCTVTSPMNSSHLAPQVQQ   QN       VFIVQMK KV HD  EQ  LQM VNLE    CQCEPH      85
Bsub-L    ........MTYNQMPKAQG  R EF  H   P  LYAHLK KQTHD  KQ  LKM CQLD    QGSDPD GD   64
Ca262-L   .........MENNIPNAQG  SF  HD      TIVNID EK HE  SDG TI  EKLE    GTGAEEN     63
Msat-L'   KLRVAVKSSFSAVPDKPHG  AY  K  S    FVAELN QS RK  TD AL EE VRET     ACGCEAM     144
Scer-L'   NDEHHLHKSWANVIPDKRG  D  H  A     FVANKH EQ HK  TDARYL  VNET     AVSSDGN.    95
Ca824-L   .......MTRNIGYPEKQG  AY K  N     FITSIK EK HD  KKGIEI  VNLT     VGADTK GD   65
Abra-L    GEQFVADFRANAAALTTANA  ED         FIAAID KPRRS   EKGIEAL KAVW     VDADGK     79
Zmay-F    VARREAPPAPQKPTQQAA  NHILS  P     FVANLKNNS FD  RDALMA GCHE     CGADSDS     140
Ecol-L    VRSGSEVGFPQSLGEVHDM  DKSL  N   F LIAHIE EP HK   RTAIHA  ARHQ     ILADGK     85

Pbor-L     ILMQV  G  FQKVT.ATLNITLPDA.GQYG  V  SSPD.RAQREAG...RRIFEQVVAE   QKV  D     132
Syne-F      LIQV  K  IQKIAGAE.GITIP.APGQYA  NI GSPDPLARAEA....RQK  DIVAQ  LKV   D     151
Bsub-L     LLVQI DA  FRKEC...KNINLPEK.ERYG     FSQKEDERKKIE...KQINALIEQ.  QVV   T     128
Ca262-L    ILFNI  K  FKEELQS.KGITLEEE.GDYA      LPQDEKARKEA....VSLFEDISKE  LEL   E     129
Msat-L'    ILVALP G  YQ....EVVDFQLPPQ.GNYA   HF LPKSDSRRKE....SKNIFTKVAESL HKV   S     207
Scer-L'    ILLGI  E  MKREFKLDLDLDIPEN.GKYA   NV FKKNEKNNKKNLIKCQKIFEDLAASFNLSV   .     165
Ca824-L    INLQI DE  FRINCDN.LGIELPES.GKYA   HI LPKETALSYQCEG....ILERAVEE  QKV   V     131
Abra-L     IHVAVP K  YKDHVKV..IGHRAPDNK.LA   QV LPRISLDAQEACR...CIVETEILAF  YYIY   Q     145
Zmay-F     LLSAV  D  FDDWASK..QGLALFDRRNT    HV LP.QDEKSMEEAK...AATEKVFVDE  LEV    P     206
Ecol-L    C LLLQK DR  FRIVAQE..RGWRLAKNY..   L LN.KDPELAAAAR...RIVEEELQRETLSIV   D     149

Pbor-L      TDNST   T KA   F QVVFIQRSADLAD....DLA  RK    ..KRSHSAIRQ...TGIDPF  PA     195
Syne-F      QNEP    T IA   F Q  IARPEGLTD....DLD  RK    ..KLTHGAIRS...PKIDTY  VA     214
Bsub-L      VVGK  TV  QK  C F R  IGASSDLKD....NLS  RK    ..KQAENWG.....VTEGLD  FA     189
Ca262-L     QINPSI  KASLEAM Y  A  VKRPNGTKK....EK   RK     RNIEKRAAWIS...KFLNET  IA     194
Msat-L'     TNTG  KS QLT  VIE VFLTPSSDSK......VD  Q     KLSHVSITSALNLQSDGITD   IC     273
Scer-L'   TSRRFYY   V LSR  T  L PLLVPLYDEKQPEFNETK RTQ    KEASLQIGLEN........W  VC     229
Ca824-L     RENRS   T KG  V K  IGSNCENQT......D  N      KRAESEVKRLVER...GSEY   IC     194
Abra-L      INVDI   K NATR E E  ITVGNNKGVSDE.....QP LD     RRIEKAVKGE......QIND   IC     206
Zmay-F      FNVSV  RN KETH   Q IFVKVAKEDNAD.....D  E    S  KLIERAAKSFS.....WADEL  FC     268
Ecol-L      MEGV   I LSSL RIE IFVNAPAGWRPR.....D  RR   IA RRIEKRLEAD........KD   VC     208

Pbor-L      CR LV  G  MPL VKE. Y P  SDPDLQ AL  VHS F       S ERSH  Y Y AHNGE NTLR  I    266
Syne-F      AR LV      TTA VGQ. Y P  HDPDME AL  VHS F       S ERSH  Y Y AHNGE NT R  V    285
Bsub-L      SQ IV  G L TPE VDA. F S  LQDEAFV AF  VHS F       FT ER  H NFY VHNGE NTLR  I    260
Ca262-L     SK IV  G  LST LRV. F K  EDERVETSL   VHS Y       FS ER  H N F VHNGE NTLR  V    265
Msat-L'     SR VI     TPA LGEYY  A  GNERFT YM   IHS F       S DR   F   GHNGE NTLR  V    345
Scer-L'     NNT  IV   Q TPA VYN. Y P  TNAHFK HM  VHS F       S DR   PL W AHNGE NTLR  K    300
Ca824-L     SR IV  G L LAD IKS. F M  NDINFK AI  VHQ Y       FT DL   F F  AHNGE NTIR  R    265
Abra-L      ARSI   G  LAE LTT. F P  LLDERFE DF  YHQ Y       FT PL   F M AHNGE NTVK  V    277
Zmay-F      SR IV     RSEVLGQ. F L  QNELYK PF  YH  F       S R PL   M L GHNGE NTIQ  L    339
Ecol-L      NLVNI  G LCEPT LPR. F L  ADLRLE AIC F Q  F       V R PL   F Y AHNGE NTIT  R    279

Pbor-L      HA Q...SLFESDLFGDDLRKAQ VINIDG    TIP  NAL L TLS  .RSLPH V   P PUTAH...    331
Syne-F      HQ Q...ALFESSLFGEDMAKVQ VINIDG    TIP  NAL   YLA  .RSLPH   N  P PUSAH...    350
Bsub-L      HRA E...QQFVSESFGEDLNKIL  ILNADG   SIL  NAF   VMA  .RKPAHTA L  P PUTEN...  325
Ca262-L   KVSR...........VLGKDVIRVL  IINKE    AIP  NL   YMN  .MDLPR    A  P PNYKS...  322
Msat-L'     IKA EGLLKCKELGLSENDLKKFL  IVDAN    GCP  GVL F LHS  .KSLPE V  I  P AVQND...  413
Scer-L'     HRS EGVMNSATFKD...ELDKLY  IIEEG    AAL  NVL L TIN TLSLPE      P AYHKD...  366
Ca824-L     NA EG...VLKSDVFGKKISDLF  IVNPK    TSL  NTF L VAD  .RPLAQ    I  P AVENN...  330
Abra-L      HKA E...TRMEHPAFGTHMQDLK VIGVGL    GSL  TVF V VRA  .RTAPMVK L V  ALTSSQTT   345
Zmay-F      HRS E...TTLKSPVWRGREHEIC  FGDPK    ANL  STA L LRS  .RSPAE L L  P AYKNHPTL   407
Ecol-L    QHARA ....TYKFQTPLIPDLHDAA  FVNETG    SS   N L L LAG  .MDIIR  R L  P ANQ...NN  343

Pbor-L    .ES SPER   K HSCL   WDGP SIA    RMM  V  PNGLP  S  YV   DLVIMA  A VLPIEPE     402
Syne-F    .ES SQEK   K HSCL   WDGP SIA    KMM  V  PNGLP  S  YV   DLVIMA  A VLPIEPE     421
Bsub-L    .TH SKEKA  E HSSL        TAIS    KQI  I  PNGL  A  YV   DYIIFS   VIEVEQE     396
Ca262-L   .KT SKEKRD  E NATL     F  AIV    ERV  V  PNGLP  S  YI   RRLILS   ALDVPAE     393
Msat-L'   .KN DPQRK   E YSAL  E WDGP LIS    HYL  T  PNGLP  G  YV  HSGRVIMA    VDIPPE   484
Scer-L'   ...DSDLK    D AACL  F WDGP LLT    RYC  I  PNGLP  C  YI  S DRVICA    VIPIENS  435
Ca824-L   .ES ETWKR   E QGTL  E WDGP AVA    QQV  V   NGLP  A  LI    NIAVLA  A VLKFEPE  401
Abra-L    P.DN...HK  IQ CNSV  E WDGP ALAE   RWVV G  PNGLP  M  TI  T GLIIGG  T  VKIDET   413
Zmay-F    S.IKYPEVTD  D YKGQ  E WDGP LLL S  RTV  T  PNGLP  A  WR S DFVYVA  V  VIPEDES  478
Ecol-L    P.D DPELR   D NSMH      GIVES  RFA CN    NGLP  A  VI    KLITCA    IWDYQPD    414
```

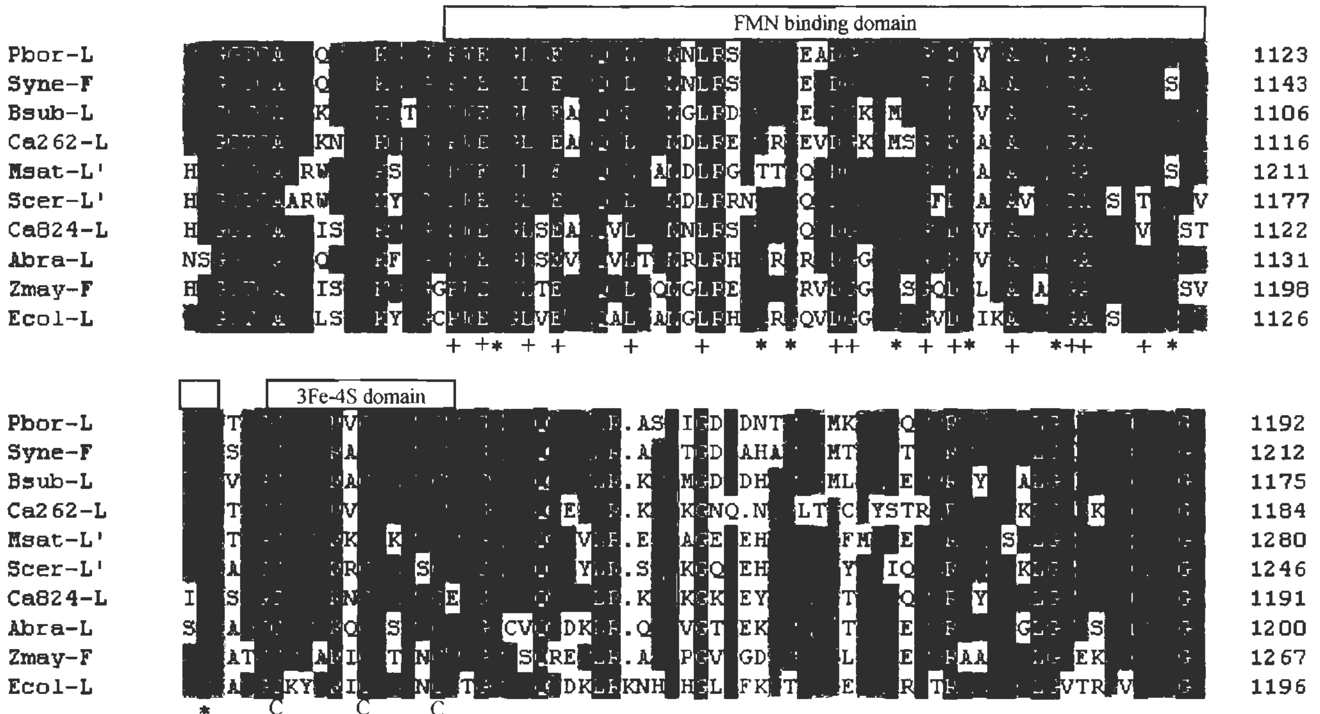


Fig. 2.5 Alignment of the deduced *C. acetobutylicum* P262 GltA protein (Ca262-L) with Fd-dependent GOGATs (-F), bacterial α subunits (-L), and eukaryotic α subunit domains (-L'), to illustrate the glutamine amidotransferase domain (boxed region in A), and the FMN and Fe-S binding domains (B). A full description of the abbreviated protein sequence names, together with their accession numbers, are supplied in Table 2.1. The black, gray and light shading represent residues conserved in 100%, > 75% and >50% of the sequences aligned, respectively. Residues conserved in purF-type glutamine amidotransferases are indicated by Δ Mei and Zalkin (1989) and • Kim *et al.* (1996) below the sequence. Residues which in PRPP-AT are involved in the formation of the glutamine binding site (Vanoni and Curti, 1999) are indicated by - above the sequences. The symbols + and * highlight identical and conserved substitutions (in at least 8 of the sequences aligned), respectively, corresponding to residues conserved in the FMN-binding domains of flavocytochrome b2 proteins from the fungi *Saccharomyces cerevisiae* (SwissProt accession number P00175), *Hansenula anomala* (SwissProt accession number P09437) and *Rhodotorula graminis* (EMBL accession number AJ001431). The conserved cysteine residues proposed to form the 3Fe-4S cluster are indicated with a C

2.4.5.3 FAD- and NAD(P)H-binding domains

Vanoni *et al.* (1996) and demonstrated conclusively that the binding sites for NADPH and FAD are within the β subunit of bacterial GOGAT, and that FAD is the flavin co-factor which receives electrons directly from NADPH during catalysis.

The adenylate portion of FAD or NAD(P)H is known to interact with a glycine-rich region, part of a β-α-β secondary structure [GXGXX(G/A/P)] (Wierenga *et al.*, 1985; Karplus *et al.*, 1991). Two such glycine-rich motifs were found in the *C. acetobutylicum* GltB protein (Fig. 2.6): residues 159-164 and 300-305. The first region satisfies all 11 positions of the consensus sequence defined by Wierenga *et al.* (1985) for the binding of FAD or NAD. The second region was not as well conserved (6 out of 11), however, the five conserved residues; Gly-300, Gly-302, Gly-305, Gly-316 and Glu-324, were identified by Scrutton *et al.* (1990) as distinctive features of NADH-binding domains. This is consistent with NADH being the required coenzyme for GOGAT activity in *C. acetobutylicum* (Chapter 5). Furthermore, these five residues are conserved in *M. sativa* and *S. cerevisiae*-both of which use the same coenzyme. On the other hand, *E. coli*, *T. ferrooxidans* and *A. brasilense*, which all use the NADPH cofactor, contain features that confer specificity for this reducing coenzyme: the third glycine residue is substituted by alanine, and a positively charged arginine residue, possibly involved in the binding of the phosphate group, is conserved at the C-terminus of the domain (Fig. 2.6). These features, together with recent site-directed mutagenesis of the *A. brasilense* β subunit region corresponding to the second glycine rich region of GltB which implicated it in NADPH binding (Morandi *et al.*, 2000), support the interaction of the first glycine-rich region of *C. acetobutyicum* GltA with FAD, and the second glycine-rich region with NADH. In addition, a region in the carboxy-terminus, amino acids 452 to 462, match all 7 positions of a second FAD consensus sequence defined by Eggink *et al.* (1990).

```
                                                              4Fe-4S cluster domain
Pbor-S    ..MGKPTIFIEYLIELPSELAPLDEIRNWDIFHLSMPEE.NLRTIGAFKRDYCIPVGHTGTLISGMASGCPI   69
Msat-S'   TDAVEHRVFVAYEEGVQYRDPNVILNDWNIVMMETKPGPLLKTSAAFTQDYCTPVCHQ......ENSGCPL   66
Bsub-S    ..MGKPTIRMEIKIEKPAERDPLTILKDWKIYSAPFSEE.ASKRIGAFVDDYCTPVCQIGADINGFTSGCPI   69
Ca262-S   ..MGKPTIFLEYDIEVGRNREPKEILKDYKIFHQRLPLE.KQCIRGAFKRDYCVPVCQAGVLFSGMVSGCPL   69
Scer-S'   ERIEETRVFMIHKIRHETHRDPRTIVNDWKIFTNPITKK.DAKYITAFNRDYCTPVCLS.......DTGCPL   64
Ca824-S   ..MGKVTVFKEYDIEESPSRPIDEIIKDYKIVHMGLDKE.KLKIIGAFTRDYCTPVCSWG.........CPL   60
Ecol-S    .MSQNVYQFIDLQIVDPPKKPLKIIKIEFVIIYEPFSEG.QAKAIADRVLSHCNAVCEWK.........CPV   61
Tfer-S    ...MSHFTFIEDPIQDPKKLDVEEIREAFRIIYQSFDLG.SAQLIADRVLHRCNAVCEWK.........CPV   59
Abra-S    MANQMMLIFVHTAQRMPDKRPAAEIRQDFAIIYARFSDE.RANEIANRVSDCGVAVCQVH.........CPV   62
                                                  C   C     C                 C

                                  4Fe-4S cluster domain
Pbor-S    NYLIPEWNDIIYRGLWKEILDRLHKTINFLEFTGRVCPA..APCEGSCVVG.IHNPPVTIKNIEYSIAEKAFN  138
Msat-S'   GMKIPEFNEVVYQNRWQEILERLLETNNFPEFTGRVCPA..APCEASCVVG.IIENPVSIKNIECAIIDKAFE  135
Bsub-S    YMLIPEWNDVVYRGRWKEILERLLKTNNFPEFTGRVCPA..APCEGSCTLA.ISDPAVSIKNIERTIIDKAFE  138
Ca262-S   HMLIPEWNDVVYRGKWELIYERLNKTNPFPEFTGRVCPA..APCEGSCTAG.LNGPAVTIKENERSIIDNAFE  138
Scer-S'   SMIIPKFNELFKNQKWLILDKLLETNNFPEFTGRVCPA..APCEAACTLG.IIEDPVGIKSVERIIIDNAFK  133
Ca824-S   GMLIPDFNDVVKGEWKKIYERISLTSCVPEFTGLICPA..ALCEGSCTLS.YNSDAVSIKEIELGIIEEAFK  129
Ecol-S    HNYIPNWLKLANEGRIFEAAELSHQTNTLPEVCGRVCPQDRLCEGSCTLN.DEFGAVTIGNIERYINDKAFE  132
Tfer-S    HNYIPNWLQLIVENRLEEAATLSHQTNTLPEICGRICPQDRLCEGACTLN.ATGGAVTIGNLERFITEEALA  130
Abra-S    SMNIPDWLKLTSEGRLEEAYEVSQATNNLPEICGRICPQDRLCEGNCVIEQSTHGAVTIGSVEKYINDTAWD  134
                                    C     C        C     C

                         FAD binding domain 1
Pbor-S    EGIKPEPIAKRTGKKIAIVGSGPAGLSAAAQLIKVGHWVTVFERADRPGGLLMYGIPNMKLDKQEVVMRRL  210
Msat-S'   EGMIPRPIVKRTGRKIAIVGSGPSGLAAAEQLIKMGHIVTVFERADRIGGLLMYGIPNMKTDKVDIVQRRV  207
Bsub-S    NGIQPRIIKKRTGKKIAIVGSGPAGLASADQLIQAGHSVTVFERADRAGGLLTYGIPNMKLEKG.IVERRI  209
Ca262-S   NGVKANTILKRTGKKVAIVGSGPSGLAVANTLIKCGHKVTVFERSDRPGGLLMYGIPNMKLDKE.VILRRV  209
Scer-S'   EGIKPCPISTRTGFTVAVIGSGPAGLACADMLIRAGHTVTVFERSDRCGGLLMYGIPNMKLDKA.IVQRRI  204
Ca824-S   NGVKPKIIKVRTGRKVAVIGSGPAGLSAAEELISVGHSVTVFERADKVGGLLRYGIPDFKLEKH.VIDRRI  200
Ecol-S    MGWRPDMSGVKQTGKKVAIIGAGPAGLACADVLTRNGVKAVVFDRHPEIGGLLTFGIPAFKLEKE.VMTRRR  203
Tfer-S    RGQPDMPDILPNGRRVAVVGGGPAGLGCADILARAGTEVVVFDRYPAVGGLLTFGIPSFKLEKA.VVARRA  201
Abra-S    QGVKPRTISRELSLSVAVIGAGPAGLAAAEELRAKGYEVHVFDRYDRMGGLVVYGIPGFKLEKS.VVERRV  205
                       ++ + + +   +    +  +       + + +

Pbor-S    NILEAEGIKVVCGTEIGKD..LPAEDLLRDYDAVVLCTGATKPRDLPIEGRQ.LKGIHFAMDFLTANTQAVL  279
Msat-S'   NLMAEEGINVVVGANIELPLYSLERLREENDAVVLAVGATKPRDLPVPGRE.LSGIHFAMEFLHANTKSLL  278
Bsub-S    KLLTQEGIDVTTGTEIGVD..ITADELKEQFDAIILCTGAQKQRDLLIEGRD.SKGVHYAMDYLTLATRSYL  278
Ca262-S   HLMAAEGIKVNTANVVGKD..YDAKELLDEFDAVVLATGASQPRDLQAEGRDNVNGIHFAVDFLKANTKSLL  279
Scer-S'   DLLSAEGIDVTTGTEIGKT..ISMDELKNKHDAVVYAIGSTIPRDLPIKGRE.LKNIDFAMQLLESNTKALL  273
Ca824-S   DVMEKSGIEKTSTNVVGF...VSAEELNDFDAVVLLTGGSTIPRDLKVEGRENIKGVHFAVDYLKQQNMRNA  270
Ecol-S    EIFTGMGIEFKLNTEVGR...VQLDDLLSDYDAVFLGVGTYQSMRGGLENED.ADGVYAALPFLIANTKQL  271
Tfer-S    NLLQEMGIRFQLNTEVGR...ISLGKLLEDFDAVFLGMGTYTAMTGDFPGEK.LPGVLKALSYLIANTRHL  269
Abra-S    KLLADAGIEIHPGFEVGR...ASLPEERRKHVDLLATGVYKAREIKAPGSG.LGNIVAALDYLTTSNKVSL  274

                        NAD(P)H binding domain
Pbor-S    DQTEP.....AITAQGKDVVVIGGGDTGTDCVGTSIRHGCNGVVQLEILPQ...PPEEAAPNNPWPEYPKVY  343
Msat-S'   DSNLQD..GNYISAKGKNVIVIGGGDTGTDCIGTSIRHGCTAVNLELLPQ...PPPTRAPGNPWPQWPRIF  345
Bsub-S    DSNFKD..KQFIDAKGKDVIVIGGGDTGADCVATALRQKAKSVHQFGKHPK...LPPARTNDNMWPEQPHVF  345
Ca262-S   DSDHED..NNYISAKGKNVIVIGGGDTGTDCVATSLRHGCKSVVQFEIMGE...PAHERTESNPWPEWPKVL  346
Scer-S'   NKDLEI..IREKIQ.GKKVIVIGGGDTGNDCLGTSVRHAASLNFELLPE...PPVERAKDNPWPQWPRVM  339
Ca824-S   GMEIKE..E.EITAKGKVVVVIGGGDTGSDCIGTAIRQGAKKVYQYEIMDK...PPAQRDETMPWPLFPRVF  336
Ecol-S    MGFGETRDEPFVSMEGRRVVVLGGGDTAMDCVRTSVRQGAKHVTCAYRRDEENMPGSRRE...........  331
Tfer-S    MQLPDP.TLPYISLSGRRVVVLGGGDTAMDCLRTAIRQGARQVTCVYRRDEENMPGSRRE...........  328
Abra-S    GDTVEAYENGSLNAAGRHVVVLGGGDTAMDCVRTAIRQGATSVKCLYRRDRKNMPGSQRE...........  334
                         ++ + + +  +    +    +        + + #↑

Pbor-S    KMDYGQEEAAAKFGGDPRAYLRTATHFEGDEQGNVKAVHTVEVQWER.NEQGQFIPKQIPGTEKIIPAQLVL  414
Msat-S'   RMDYGHEEAETKFGKDPRTYEVLTKRFVGDENGVVKGLEVVRVCWER.DETGKFQFKEIEGSEEIIEADLVL  416
Bsub-S    TLEYAYEEREAKFGRDPREYSIQTTKMVADKNGKLKELHTIQEKVE.NEEGKYEFREIPGTEKVWPAQLVF  416
Ca262-S   KMDYGQEEFIELYGKDPREYLTTVTAVHGDKDGNLESVDTVKVEWKK.GDNGRMFPAHVEGSEKNWPAELIL  417
Scer-S'   RMDYGHAEVKEHYGRDPREYCILSKEFIGNDEGEVTAIRTVRVEWKK.SQSGVWQNVEIPNSEEIFEADIIL  410
Ca824-S   KTTTSHEEG......CERLFGVSTKKLEG.KDGKLELLKGVQVKWER.DENGKMSMKEIEGSEFEKKVDLIL  400
Ecol-S    .VKNARREG......VEFKFNVQPLGIEVNGNGKVSGVKMVRTEWGEPDAKGRRRAEIVAGSEHIVPADAVI  396
Tfer-S    .VANARREG......VHFLFHRQPVEVVGDAAGSPGRASGGDPSGEPDARGRRHPEIIPGSEQVLDADVVI  393
Abra-S    .VAHAEEEG......VEFIWQAAPEGFTGDT..VVTGVRAVRIEHLGVAEATGRQTPQVIEGSEFTVQADLVI  397
```

```
                                                   FAD binding domain 2
Pbor-S      L  LPQP.LLDAMG..LDRDASNIKSEYGR T IPK..     C     N  G REC     481
Msat-S'     L  LP AT.IAEKLG..VERDNSNFKADYGR S VDG..     C     S  Q AQV     483
Bsub-S      I  E TQP.LLKQFG..VNSVNN.KISAAYGD Q NIDG..    A     N  EV REV    482
Ca262-S     L  T SDY.LKNAFG..IESDESNVKAGNVD M NVPK..  T  A  T  S  IA IAV    484
Scer-S'     LS V P LI.NGNDNE..VKKTR.GTIATLDDSSY IDGGKT C C    Q  KC ASV    478
Ca824-S     I  VHP HKGIVEDLQ..LKLDS.GNVFTDEN M REN..    M     H  QS KEI     467
Ecol-S      M F RPH MEWLAKHSVELDSQGRIIAPEGSDNA Q NPK..  IV SD T  A  K DGI   466
Tfer-S      I F DPSPAPWFAEHHITTDSRGRISTN....AQ Q NPR..  M SD T  D  R QGI    459
Abra-S      K  EP..EDLPNAFDEPELKVTRGTLLVDHRTKM NMDG..  IV A    R  D RGH     465
                                               *        **** **

Pbor-S      L MGETNLP                                                    492
Msat-S'     S TNEDHGIDGNQDEFVKRQQDLNKKHSKHTVM                            518
Bsub-S      R MGSSVLP                                                    493
Ca262-S     K RA                                                         490
Scer-S'     K MDGTTYLPSNGGIVQRDYKLLKELASQV                               510
Ca824-S     K MGETSLRG                                                   479
Ecol-S      MN EV                                                        472
Tfer-S      LD GV                                                        465
Abra-S      PRLRQGEGRGTGCRGG                                             481
```

Fig. 2.6 Alignment of the deduced *C. acetobutylicum* P262 GltB protein (Ca262-S) with GOGAT β subunits (-S), and β subunit domains (-S'), from bacteria and eukaryotes, respectively. A full description of the abbreviated protein sequences presented, together with their accession numbers, are supplied in Table 2.2. The *M. sativa* (Msat-S') and *S. cerevisiae* (Scer-S') fused polypeptides were included from residue positions 1676 and 1635, respectively. Identified cofactor-binding domains are labeled above the relevant regions. The pattern of conserved cysteine residues are indicated with a C. The regions that match the consensus sequence for the formation of an adenylate binding fold are marked with +. The symbol # indicates the proposed acidic residue which is absent in NADPH-dependent enzymes, and the arrow indicates the proposed basic residue which may favor NADPH binding (Vanoni and Curti, 1999). Residues that match the second FAD consensus sequence defined by Eggink *et al.* (1990) are marked with a *. The black, gray and light shading represent residues conserved in 100%, > 75% and >50% of the sequences aligned, respectively.

2.4.5.4 Iron-Sulfur centers

Bacterial GOGATs contain three distinct iron-sulfur (Fe-S) centers, one 3Fe-4S center located within the α subunit, and two 4Fe-4S centers (Section 1.5.4). Analysis of both the deduced GltA and GltB protein sequences showed that only three regions contained clusters of cysteine residues that could be involved in the formation of Fe-S centers: one region in GltA (residues 1123-1134), and two regions at the N-terminus of GltB (residues 46-68 and 106-114). The cysteine cluster found in GltA, CX_5CX_4C, is invariant in all the GOGAT sequences aligned (Fig. 2.5b). As pointed out by Pelanda *et al.* (1993), the cysteine spacing is similar to other enzymes containing the 3Fe-4S cluster, supporting the assignment of these cysteines to the $[3Fe\text{-}4S]^{1+,0}$ center. In contrast, the two cysteine-rich regions in GltB are not strictly conserved (Fig. 2.6). In fact, depending on the organism, the first cysteine-rich region can be defined as

$CX_2CX_4CX_{12/6/5/3}CP$, while the second cluster can be divided into patterns $CX_3CPX_4CX_3C$ or CPX_2CX_3C, none of which are typical of cysteine clusters involved in the formation of 4Fe-4S centers. Surprisingly, analysis of the properties of the recombinant β subunit of *A. brasilense* GOGAT failed to detect the presence of Fe-S clusters (Vanoni *et al.*, 1996). As suggested by Vanoni and Curti (1999), the cysteine clusters present in the β subunits are probably involved in the formation of the two 4Fe-4S clusters (Centers 2 and 3), however these clusters may be formed at the interface of the two subunits and require the correct interaction of the two subunits. This model could account for the variability seen within the two cysteine clusters. Interestingly, the pattern of the second cysteine cluster domain correlates with whether the particular protein is specific for NADH or NADPH, and with the different patterns of identities and similarities observed in the region following the NAD(P)H binding domain. Although the significance of these observations are unclear, it suggests a division of the β subunit into two subclasses: one comprising proteins from the first six organisms *C. acetobutylicum*, *P. boryanum*, *B. subtilis*, *M. sativa*, and *Saccharomyces*, and the other class including the remainder of the proteins from *E. coli*, *T. ferrooxidans*, and *A. brasilense*.

In addition, all the other highly conserved regions of unknown function identified in GOGAT proteins (Vanoni and Curti, 1999) are also conserved in the putative GltA protein (Appendix C). Presumably some of these regions are important to ensure interdomain coupling and efficient electron transfer for catalysis.

2.5 Conclusion

The cloning and sequence analysis of the regions flanking the *C. acetobutylicum* P262 *glnA* and *glnR* genes, enabled the identification and characterization of the *glt* gene cluster. Immediately downstream of *glnR* lie the structural genes encoding the α (*gltA*) and β (*gltB*) subunits of GOGAT. Their arrangement is typical of that found in bacteria in which the large subunit precedes the small, however unique with respect to their arrangement in the chromosome; this is the first report in which the structural genes coding for the two key enzymes of ammonia assimilation, GS and GOGAT, are tightly linked (separated by a putative regulator). Indeed, even in *C. acetobutylicum* ATCC824, the α subunit gene, which is similarly separated by 12 bp from the downstream β subunit gene, is preceded by an ORF coding for a probable ATP-dependent helicase (accession no. gb:AE001438). Only ORFs with homology to hypothetical

proteins could be identified in the ~2.2 kb region downstream of the β subunit. The sequence analyses also suggest a novel mechanism of regulation of *glt* gene expression in this Gram-positive anaerobic bacterium, and were essential for providing a background against which a physiological analysis of nitrogen metabolism could be conducted.

Since the proposed functional domains that have been described for other GOGAT enzymes are also conserved in the deduced amino acid sequences of the *C. acetobutylicum gltA* and ***gltB*** genes, we conclude that these genes must represent the structural genes for the α and β subunits of GOGAT, respectively. These sequence analyses also support the model proposed by Vanoni and Curti, (1999) for the *A. brasilense* GOGAT enzyme (Section 1.5.4).

CHAPTER 3

Characterization of *gltX*, a second GOGAT β subunit-like gene in *C. acetobutylicum* P262

3.1 Summary 79
3.2 Intruduction 80
3.3 Materials and methods 81
3.3.1 Bacterial strains and plasmids -------- 81
3.3.2 Media and growth conditions -------- 81
3.3.3 DNA extractions and general DNA manipulations -------- 81
3.3.4 Determination of metronidazole MIC's -------- 82
3.3.5 Metronidazole reduction assay -------- 82
3.3.6 *In vitro* transcription-translation -------- 82
3.3.7 RNA extractions and Northern hybridizations -------- 83
3.3.8 Construction of size-selected genebanks -------- 83
3.3.9 Colony hybridization -------- 83
3.3.10 Southern hybridization -------- 83
3.3.11 Southern hybridization using a synthetic oligonucleotide and the Tetramethylammonium chloride (TMA) procedure -------- 84
3.3.12 Nucleotide sequencing and analysis -------- 85
3.4 Results and discussion 85
3.4.1 Localization and sequencing of the region on plasmid pMET13C1 responsible for sensitivity to metronidazole in *E. coli* F19 -------- 85
3.4.2 Identification of the ORF*A* gene product -------- 88
3.4.3 Metronidazole reduction -------- 88
3.4.4 Nucleotide sequence analysis of ORF*A* -------- 90
3.4.5 Deduced amino acid sequence of ORF*A* -------- 92
3.4.6 Identification of conserved domains within *gltX* -------- 94
3.4.7 *gltX* homology in *C. beijerinckii* -------- 97
3.4.8 Cloning and sequencing of the chromosomal DNA regions flanking *gltX* -------- 99
3.4.9 Identification of the genes flanking *gltX* -------- 101
3.4.10 Determination of *gltX* transcript size -------- 103
3.4.11 Probing for a second GOGAT α subunit homologue -------- 104
3.4.12 *gltX* in context -------- 106
3.5 Conclusion 109

CHAPTER 3

Characterization of *gltX*, a second GOGAT β subunit-like gene in *C. acetobutylicum* P262

3.1 Summary

Plasmid pMET13C1 was isolated due to its ability to make *E. coli* strain F19 (Santangelo *et al.*, 1991) susceptible to the drug metronidazole. It contained a 4.8 kb insert of *C. acetobutylicum* P262 chromosomal DNA cloned into the vector pEcoR251. The region responsible for metronidazole sensitivity was localized on a 1.5 kb segment, sequencing of which revealed an ORF (ORF*A*) of 1245 bp. A corresponding protein product of the predicted size (~46 kDa) was identified, and correlated with an increased capacity to reduce the drug to its active bactericidal form. Amino acid homology studies revealed that ORF*A* coded for a protein with striking homology to GOGAT β subunit domains, and was thus designated *gltX*. All the functional domains proposed for GOGAT β subunits (discussed in Chapter 2) were conserved. However, *gltX* differs from bacterial GOGAT subunits at its amino terminus, and by the absence of certain conserved residues, features that are conserved in a highly homologous gene identified in the *C. acetobutylicum* ATCC824 genome (77.5% identity), and apparently also in *C. beijerinckii*. The DNA regions flanking *gltX* were cloned, sequenced and putative gene products identified. A second GOGAT α subunit homologue did not lie adjacent to *gltX*, which was independently transcribed, nor apparently anywhere else on the chromosome.

3.2 Introduction

Metronidazole (5-nitroimidazole) is a broad spectrum antimicrobial drug. It is a pro-drug which is activated by chemical or enzymatic reduction of the 5-nitro group to generate the toxic intermediate(s) which cause cell death mainly via DNA damage (Edwards, 1993; Dachs *et al.*, 1995). Metronidazole is selectively toxic to anaerobic organisms since it requires the activity of electron transport proteins with sufficiently negative redox potential for drug activation (Edwards, 1993; Johnson, 1993).

Clostridial electron transport proteins which have been reported to provide the source of electrons for the reductive activation of metronidazole include pyruvate-Fd-oxidoreductase (PFOR) (Narikawa, 1986), reduced ferredoxin (Fd) (Lindmark and Muller, 1976), and hydrogenase (Church *et al.*, 1988; Church *et al.*, 1995). These proteins and enzymes are involved in the phosphoroclastic breakdown of pyruvate to acetyl CoA - an important pathway for carbohydrate metabolism in anaerobic organisms (Johnson, 1993). Flavodoxin, a functional counterpart of Fd, has also been reported to transfer electrons to metronidazole (Chen and Blanchard, 1979). As part of a study on the molecular characterization of genes from *C. acetobutylicum* involved in the electron transfer system which are able to activate the drug metronidazole in *E. coli* (Santangelo *et al.*, 1991), our laboratory has been investigating genes, isolated by screening of the *C. acetobutylicum* gene bank, which make *E. coli* sensitive to metronidazole.

The aim of this work was to characterize plasmid pMET13C1, which rendered *E. coli* highly sensitive to metronidazole, for two reasons. First, the study of electron transport genes is important for the understanding and future manipulations of solvent pathways in this industrially important *C. acetobutylicum* strain (Jones and Woods, 1986). Second, since metronidazole susceptibility is often associated with altered carbon metabolism, it may contribute to our understanding of this process. In *Helicobacter pylori*, resistance to metronidazole correlates with repression of the activity of various Krebs cycle enzymes including isocitrate dehydrogenase (Hoffman *et al.*, 1996). Genes that affect the rate at which 2-oxoglutarate is synthesized will also affect the activity of the GS/GOGAT pathway.

3.3 Materials and methods

3.3.1 Bacterial strains and plasmids

The nitrate reductase deficient *E. coli* strain F19 was derived from the *recA E. coli* strain CC118 (*araD*139Δ(*ara, leu*)7697Δ*lac*X74 *phoA*Δ20 *galE galK thi rpsE rpoB argE*am *recA*1) (Manoil and Beckwith, 1985) by transposon mutagenesis (Santangelo *et al.*, 1991). This mutant is highly sensitive to the reduced toxic intermediates of metronidazole due to a deficient DNA repair system, yet it has increased tolerance to the drug (20 μg/ml in CC118 versus >35 μg/ml in F19) because it lacks nitrate reductase activity involved in the reduction of metronidazole to its toxic derivatives in *E. coli*. Thus F19 was used as a screening system for the isolation of genes on recombinant plasmids which activate metronidazole and render F19 sensitive to the antibiotic. Plasmid pMET13C1 was isolated using this system (Santangelo *et al.*, 1991). It contained a 4.8 kb *Sau*3A1 fragment of *C. acetobutylicum* P262 chromosomal DNA cloned into the positive selection vector pEcoR251.

C. acetobutylicum P262 wild type (Jones *et al.*, 1982), *C. beijerinkii* NCIMB 8052 (laboratory stock) and *E. coli* JM105 were used as the source of chromosomal DNA. RNA was extracted from *C. acetobutylicum* P262 and *E. coli* F19 harbouring the relevant plasmid constructs. *E. coli* strains LK111 (Zabeau and Stanley, 1982), JM105 and JM109 (Yanisch-Perron *et al.*, 1985) were used interchangeably as the cloning hosts. The origins of the plasmid constructs generated in this study are illustrated in Fig. 3.1 and 3.7.

3.3.2 Media and growth conditions

Section 2.3.1 describes the conditions and media used to grow *C. acetobutylicum* P262, *C. beijerinckii* 8052, and *E. coli* clones. Anaerobic growth of *E. coli* was carried out in an anaerobic glove cabinet (Forma Scientific). LB, supplemented with 0.2% (w/v) $NaNO_3$ and 0.1% (w/v) glucose to support anaerobic respiration of *E. coli,* was used for MIC determinations

3.3.3 DNA extractions and general DNA manipulations

E. coli genomic DNA was prepared according to Sambrook *et al.* (1989). The preparation of *Clostridium* genomic DNA, extractions of *E. coli* plasmids, as well as other general DNA manipulations are as described in Section 2.3.2.

3.3.4 Determination of metronidazole MIC's

The minimal inhibitory concentration (MIC) of metronidazole (Sigma) was determined for *E. coli* F19 harbouring various plasmid constructs under strict anaerobic conditions. *E. coli* transformants (5 ml) were grown aerobically overnight. The cultures were then processed in the anaerobic glove cabinet. Each culture was washed and diluted 10^{-4} in Ringers solution. A 100 μl of each sample dilution was plated in duplicate onto agar medium containing increasing concentrations of metronidazole, typically 0, 5, 10, 15 and 20 μg/ml. The MIC was defined as the lowest concentration of the drug that inhibited growth after overnight incubation at 34 °C. All solutions and media were equilibrated anaerobically for at least 24 hours prior to use.

3.3.5 Metronidazole reduction assay

Cell-free extracts were prepared in an anaerobic glove cabinet as follows. *E. coli* F19 transformants were harvested from overnight cultures (200ml), washed, and resuspended in 10 ml of 50 mM Tris-Cl (pH 8.0). Cells were lysed using the French-press method detailed in Chapter 5 (Section 5.3.3), centrifuged at 10 000 x g for 10 min, and the supernatant retained. Protein concentrations were determined using the Bio-Rad method.

Metronidazole reduction assays were adapted from the hydrogenase-linked assay for ferredoxin and flavodoxin from *C. pasteurianum* as described by Chen and Blanchard (1979). Assays were conducted at room temperature in the anaerobic glove cabinet. The reaction mixture comprised 50 mM Tris-Cl (pH 8.0), 0.5 mM metronidazole and cell-free extracts containing 2 mg/ml protein. The reactions were initiated by addition of the cell-free extract. Metronidazole irreversibly loses its absorption peak at 320 nm upon reduction, and so the rate of reduction was measured spectrophotometrically. Because of the strong absorption of metronidazole at 320 nm, the assays were monitored at 360 nm as specified by Chen and Blanchard (1979) for metronidazole concentrations of 0.5 mM. An anaerobic atmosphere was essential for this assay as the slightest exposure to oxygen resulted in loss of metronidazole-reducing activity.

3.3.6 *In vitro* transcription-translation

[^{35}S] methionine-labelled proteins were produced from shortened constructs of pSKMETS by *in vitro* cell-free transcription-translation using the prokaryotic DNA-directed translation kit from Promega. The resulting translation products were separated on a SDS polyacrylamide gel (4% stacking and 15% resolving phases) according to Laemmli (1970) and visualized by

autoradiography. Molecular weight markers obtained from Pharmacia were stained with PAGE 83 blue (BDH) prior to autoradiography.

3.3.7 RNA extractions and Northern hybridizations

Strict RNAse-free conditions were employed during all RNA handling procedures as detailed in Chapter 5. Total mRNA was isolated using the hot phenol extraction protocol described by (Aiba *et al.*, 1981). *C. acetobutylicum* RNA was extracted from cells grown in CBM media at various growth stages as described in the text. *E. coli* cultures were extracted from YT media after overnight growth. RNA was stored in water at - 70 °C in the presence of RNAse inhibitor (Sigma).

Northern blots were performed using the non-radioactive digoxigenin (DIG) Labeling and Detection kit according to the manufacturer's instructions (Boehringer Mannheim). RNA was separated by electrophoresis in 1.5% denaturing formaldehyde agarose gels (Fourney *et al.*, 1988), transferred by capillary action to Hybond N^+ nylon membranes and fixed. The ~850 bp RNA probe was synthesized by *in vitro* transcription of the gel-purified 1.42 kb *Ssp*I fragment of construct pSKMETS31 (Fig. 3.1), using T7 polymerase to produce a run-off transcript antisense to the ORF*A* mRNA (illustrated in Fig. 3.10). The probe includes 64 bp upstream of the presumptive ATG start codon. Hybridization was performed at 65 °C using a probe concentration of approximately 100 ng/ml. RNA markers were obtained from Gibco BRL.

3.3.8 Construction of size-selected genebanks

Size-selected genebanks of *C. acetobutylicum* P262 chromosomal DNA were generated in the Bluescript vectors as described in Chapter 2, Section 2.3.3.

3.3.9 Colony hybridization

Colonies harbouring recombinant plasmids with *C. acetobutylicum* insert DNA were processed for hybridization as described in Section 2.3.4.

3.3.10 Southern hybridization

Routine Southern hybridization was performed as described in Section 2.3.5.

3.3.11 Southern hybridization using a synthetic oligonucleotide and the Tetramethylammonium chloride (TMA) procedure

A region of low codon redundancy, ProProHisHisAspIleTyrSer, found strictly conserved in all bacterial and eukaryotic GOGAT α subunit domains (corresponding to amino acid positions 990-997 of *C. acetobutylicum* GltA (Appendix C)), was used to design a degenerate 23-mer oligonucleotide probe pool. Based on the codon preference usage of *C. acetobutylicum* (Young *et al.*, 1989), which strongly favors A or T in the third position, eight 23-mer oligonucleotides CCX•CCX•CAT•CAT•GAT•ATX•TAT•TC (where X indicates an equal mixture of A and T) were synthesized by D. Botes (Department of Biochemistry, University of Cape Town), and radioactively end-labeled using [γ-32P]ATP and bacteriophage T4 polynucleotide kinase.

The Tetramethylammonium chloride (TMA) oligonucleotide hybridization procedure (Wood *et al.*, 1985; DiLella and Woo, 1987) was used to control the stringency of the hybridizations, and increase the overall sensitivity of the probes. It is often unsuitable to use the empirically determined Td (temperature of dissociation) for oligonucleotide probes rich in AT content as the resulting background is high. TMA eliminates the influence of base bonding strengths by causing a stronger interaction of AT bp comparable to GC bp such that at 3 M Me_4NCl, the melting temperature of AT bp is shifted to that of GC bp (Wood *et al.*, 1985). The resulting Td value becomes a function of probe length only. For the degenerate 23-mer oligonucleotide the Td was 66 °C in 3M Me_4NCl solution.

E. coli and *C. acetobutylicum* chromosomal DNA was appropriately digested (see Fig. 3.11) and the fragments separated electrophoretically. Gels were loaded in triplicate. The DNA fragments were capillary transferred in 0.4 M NaOH onto $HybondN^+$ membranes as outlined for Southern blotting. The subsequent hybridization and TMA wash conditions were according to the protocol of Wood *et al.* (1985). The membranes were hybridized overnight under non-stringent conditions (37 °C) with the pool of radiolabeled oligonucleotide probes (each at approximately 2×10^6 cpm probe/ml of solution). Each of the membranes were washed in 3 M Me_4NCl solution at different temperatures; 62 °C (recommended wash stringency is 2-4 °C below the Td in 3M Me_4NCl), 56 °C, and 48 °C. Bound probe was detected after seven days of autoradiograph exposure at -70 °C.

3.3.12 Nucleotide sequencing and analysis

The ~3.5 kb *Eco*RI-*Hind*III DNA insert of plasmid pSKMETS, was sequenced in stages in both directions by generating overlapping deletion clones by exonuclease III digestion of appropriately restricted DNA fragments, according to the method of Henikoff (1984)(Fig. 3.1). Sequencing of the 5.4 kb *Cla*I-*EcoR*I(a) region of *C. acetobutylicum* chromosomal DNA, extending upstream of pSKMETS, was similarly achieved in two stages (the sequencing strategy is shown in Fig. 3.7). First, the 1.4 kb *Eco*RI DNA insert of plasmid pSKMETSUP2 was sequenced on one strand. Second, the DNA sequence of the 4 kb *Cla*I-*Eco*RI(b) fragment of plasmid pSKMETSUP3 was obtained in both directions by creating a deletion construct ΔpSKMETSUP3, and generating exonuclease III shortenings from the gel purified ~6.7 kb *Eco*RI(b)-*Pst*I-restricted fragment, followed by subcloning of the *Cla*I-*Eco*RI(b) fragment in pSK and generating nested deletion clones from *Cla*I-*Apa*I-restricted DNA.

DNA templates were manually sequenced by the dideoxy-chain termination method of Sanger *et al.* (1977) using [^{35}S]dATP and the Sequenase® version 2 kit (Biochemical Corp.). A 16-mer oligo primer (5'-TGGGAGATACAAGATG-3') was synthesized by D. Botes (Department of Biochemistry, University of Cape Town) to sequence over the *Eco*RI(b) junction (Fig.3.7). Sequence analysis was as described in Section 2.3.7, except that the Genedoc program was used to format the multiple sequence alignments generated in GCG.

3.4 Results and discussion

3.4.1 Localization and sequencing of the region on plasmid pMET13C1 responsible for sensitivity to metronidazole in *E. coli* F19

Plasmid pMET13C1 was isolated using the *E. coli* F19 screening system developed for the isolation of electron transport genes that activate metronidazole and render F19 sensitive to the drug (Santangelo *et al*., 1991). The minimum concentration of metronidazole that inhibited growth of *E. coli* F19 transformed with pMET13C1 under anaerobic conditions was established to be <5 μg/ml (Table 3.1). This groups it with the most metronidazole-sensitive class of recombinant plasmids isolated from screening of the *C. acetobutylicum* gene bank. A restriction map was generated for the plasmid (Fig. 3.1), and MIC assays performed on various subclones (not shown) indicated that the region associated with metronidazole sensitivity was contained on the 2.1 kb *Eco*RI-*Hinc*II fragment (Fig. 3.1).

The ~3.4 kb *Eco*RI-*Hind*III fragment was cloned in opposite orientations into the pBluescript vectors pSK and pKS generating plasmids pSKMETS (Fig. 3.1) and pKSMETS. Both constructs displayed an MIC to metronidazole of 5 μg/ml, equivalent to that of the parent plasmid pMET13CI (Table 3.1). Since the orientation of cloning did not affect the level of metronidazole susceptibility, it suggested that the putative gene(s) responsible for this phenotype were being expressed in *E. coli* from their own promoters. By performing MIC assays on *E. coli* F19 harbouring nested deletion clones of pSKMETS (summarized in Fig. 3.1) a ~1.5 kb region, located between deletion clones pSKMETS8 and pSKMETS98, was identified as the minimum region responsible for metronidazole susceptibility.

Table 3.1 Metronidazole minimum inhibitory concentrations (MIC) of *E. coli* strains harbouring various plasmid constructs.

E. coli strain	Plasmid	MIC[b] (μg/ml)
CC118		20
F19		35
F19	pEcoR251C[a]	35
F19	pMET13C1	<5
F19	pSK	35
F19	pSKMETS	<5
F19	pKSMETS	<5
AN3001		150
AN3001	pSK	150
AN3001	pSKMETS	100

E. coli strain genotypes and references are presented in Appendix A. Note: strain AN3001 is not *rec*A - and therefore can tolerate higher concentrations of metronidazole.

a: Plasmid pEcoR251C contains a ~0.4 kb fragment of non-coding *C. longisporum* chromosomal DNA cloned into the *Bgl*II site of vector pEcoR251, and served as a vector control.

b: MIC determinations are described in Section 3.3.4.

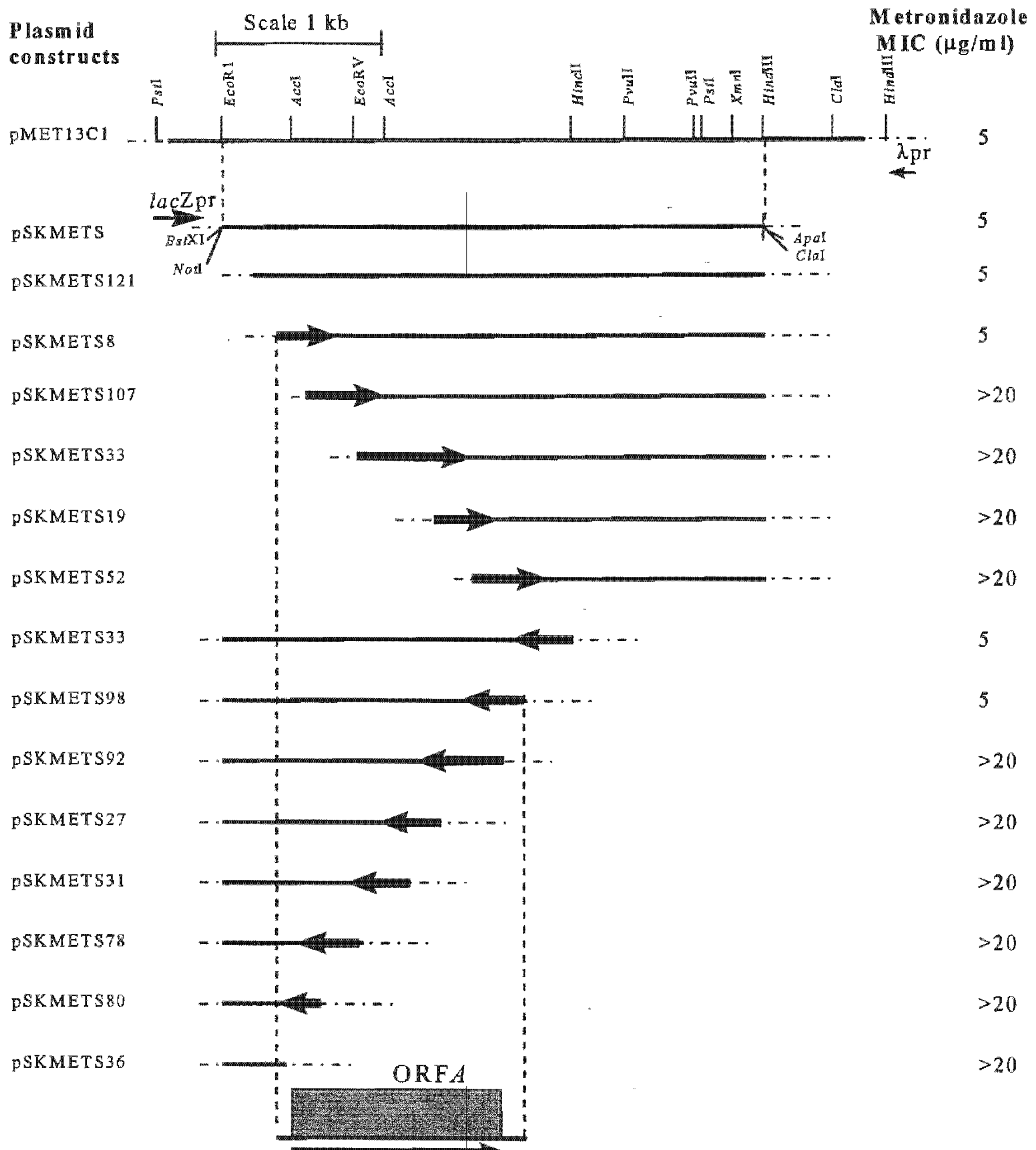


Fig. 3.1 Localization of the *C. acetobutylicum* chromosomal DNA region (thick lines) associated with metronidazole sensitivity in *E. coli* F19. Plasmid pSKMETS was derived by subcloning the 3.47 kb *Eco*RI-*Hind*III fragment from pMET13C1 into the corresponding sites in the Bluescript vector pSK. Nested deletion clones of pSKMETS, generated from both ends of the insert, are represented below. Their corresponding minimum inhibitory concentration (MIC) values for metronidazole are indicated on the RHS. The vertical dashed lines demarcate the minimum region associated with metronidazole sensitivity. Vector sequence is represented by the dashed horizontal lines. Arrows reflect the extent of sequencing of the respective clones. The position of the resulting open reading frame (ORF*A*) is shown at the bottom, together with the transcriptional polarity.

This region (1516 bp) was sequenced in both directions. Six-phase translation revealed one major open reading frame, designated ORF*A*, of 1245 bp, with a coding capacity of 414 amino acids. The loss of metronidazole sensitivity observed for constructs pSKMETS107 (in which the amino terminal 76 bp of ORF*A* are deleted), and pSKMETS92 (in which the carboxy terminal 39 bp are truncated), confirms a direct correlation between ORF*A* and metronidazole susceptibility. Furthermore, these results imply that the carboxy terminal 13 amino acids of ORF*A* are necessary for the metronidazole phenotype observed.

3.4.2 Identification of the ORF*A* gene product

An ORF*A* gene product was identified in a coupled transcription-translation assay (Fig. 3.2). A protein of the predicted size for the ORF*A* gene product (46.1 kDa) was synthesized *in vitro* from plasmid pSKMETS (lane c), pSKMETS8 (lane d), which contains 84 nucleotides upstream of ORF*A*, and pSKMETS98 (lane f), which contained an additional 101 bp of sequence downstream of ORF*A* (refer to Fig. 3.1 for plasmid constructs). No product could be identified for pSKMETS107 (lane e) in which the amino terminal 76 bp of ORF*A* are deleted. In contrast, plasmids pSKMETS92 (lane g), pSKMETS31 (lane h), pSKMETS78 (lane i) and pSKMETS80 (lane j), which contain the ORF*A* gene truncated at the 3' end to various extents, produced translational products consistent with these polypeptides being generated by transcriptional read-through into the Bluescript SK vector sequence. These results are consistent with ORF*A* producing a product responsible for the metronidazole phenotype observed.

3.4.3 Metronidazole reduction

Susceptibility to metronidazole in anaerobic bacteria correlates with the ability to reduce the drug to its active form. However, it could also be the result of increased permeability to the drug, or decreased tolerance to the drug's cytotoxic effects, as has been observed in DNA repair deficient strains of *E. coli* (Jackson *et al.*, 1984; Yeung *et al.*, 1984). Since it is generally assumed that metronidazole enters the cell by a simple passive diffusion process in which the rate of uptake is controlled by the rate of intracellular drug reduction (Muller, 1983), the other two mechanisms were investigated to account for susceptibility associated with ORF*A*.

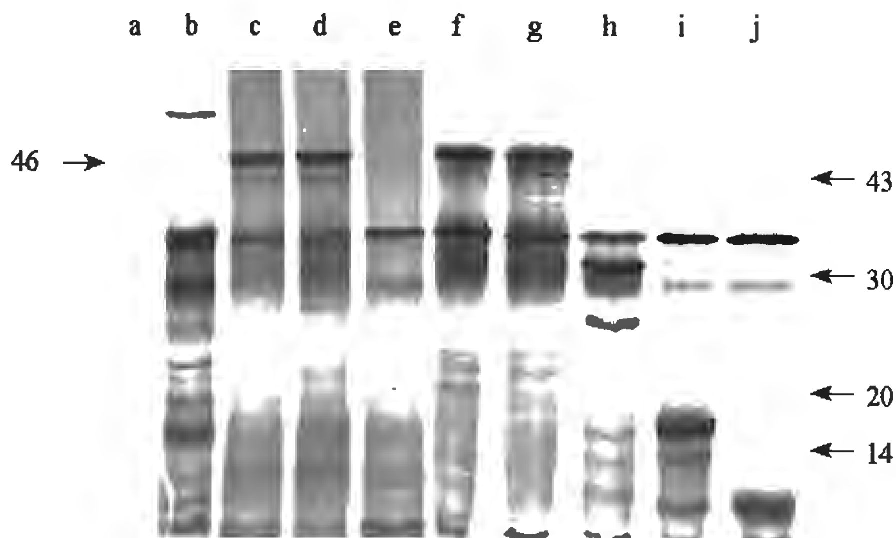


Fig. 3.2 Autoradiogram of [^{35}S] methionine-labeled proteins of Bluescript SK control (lane b), pSKMETS (lane c), pSKMETS8 (lane d), pSKMETS107 (lane e), pSKMETS98 (lane f), pSKMETS92 (lane g), pSKMETS31 (lane h), pSKMETS78 (lane i) and pSKMETS80 (lane j) synthesized in an *in vitro* coupled transcription-translation assay. Lane (a) represents a no DNA control. Protein molecular weight standards (kDa) are indicated along the RHS margin. The putative ORF*A* translation product (~46 kDa) is indicated in the LHS margin.

No differences could be observed in cell viability, morphology, or susceptibility to UV irradiation between *E. coli* F19 cultures harbouring plasmids pSKMETS or pSKMETS98, and the control vector pSK (results not shown). These studies suggested that ORF*A* is not associated with DNA damage, nor does it further reduce the efficiency of the compromised DNA repair mechanism.

The reducing activity of crude anaerobically prepared cell-free extracts of F19 transformed with plasmid pSKMETS98 and the control plasmid pSK, were compared. This experiment, based on the assay designed by Chen and Blanchard (1979), takes advantage of the fact that the reduction of metronidazole, which is irreversible because the reduced compound further decomposes, is indicated by a decrease in absorbance at 320 nm. Hence the rate of reduction can be continuously monitored with a spectrophotometer. Under these assay conditions, the rate at which metronidazole was reduced by the cell-free extract from the culture containing pSKMETS98 was approximately 4 times the rate at which the drug was reduced by the cell-free extract from F19 harbouring pSK only (Fig. 3.3). These results suggest that the metronidazole sensitive phenotype associated with the ORF*A* gene product is due to its ability to reductively activate the drug, either directly or indirectly, in the *E. coli* host. The *E. coli*

nitrate and chlorate reductase systems have been shown to reduce metronidazole (Yeung *et al.*, 1984)

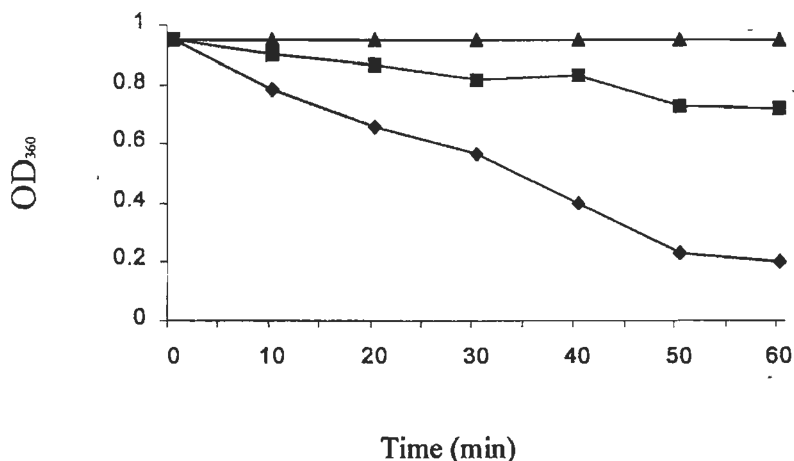


Fig. 3.3 Spectrophotometric analysis of metronidazole reduction by cell free extracts from F19 cultures harbouring plasmid pSKMETS98 (◊) and the control vector pSK (□), containing equivalent amounts of protein. The straight line (Δ) represents reaction mixture without cell free extract added.

3.4.4 Nucleotide sequence analysis of ORF*A*

The nucleotide and deduced amino acid sequences for ORF*A* and flanking regions are shown in Fig. 3.4. The gene encoded by ORF*A* presumably initiates at the ATG codon at nucleotide position 85, and terminates in TAA at nucleotide 1329. The ATG start is preceded by a putative Shine-Dalgarno (SD) sequence (5'-GAGGA-3') positioned 8 bp upstream, and matches the prototype *E. coli* binding site. A similar value (7 +/- 2 nucleotides) has been deduced for the optimal spacing in *E. coli* ribosome binding sites (Young *et al.*, 1989). As mentioned earlier, cloning of ORF*A* in both orientations with respect to vector DNA, indicated that ORF*A* is transcribed from a promoter present on the insert. A sequence similar to σ^{43} Gram-positive consensus-type promoters (TTGAAA[17 bp]TAGTAT) (Graves and Rabinowitz, 1986) is located 34 bp upstream of the presumptive start codon, andcontains a reasonable match to an extended -10 recognition sequence suggested for Clostridial promoters (Young *et al.*, 1989). However further transcriptional studies are necessary to determine promoter activity in both *E. coli* and *C. acetobutylicum*. An inverted repeat sequence with the potential to form a stem-loop structure (ΔG = -12.8 kcal/mol) is located 121 bp downstream of the TAA stop codon which can be implicated in factor-independent transcriptional termination.

```
                                                                                          ORFA
   1  ATTTGTTTTTTTAGGATAATATTGAAAAATAATAATAACTAAGTTAGTATACTATATATATATAAAGAAATGAGGAATTTTAAGATGACATGTAACGACA  100
                                                                                           M  T  C  N  D  I

 101  TACAATTATTAGATGAAGCAGATAGATGTTTATTGTGCAAGAAACCAAGATGTAAGGAAAATTGTCCAATACAAACATCTATTCCAGAGATTATATCTCT  200
         Q  L  L  D  E  A  D  R  C  L  L  C  K  K  P  R  C  K  E  N  C  P  I  Q  T  S  I  P  E  I  I  S  L

 201  TTATAAAGAAAATAGATTAAGAGAAGCAGGAGAAATTTTATTTAATAACAATCCACTTCCAGCAGTTTGTTCACTAGTATGTATTCATGAAGATCAATGT  300
        Y  K  E  N  R  L  R  E  A  G  E  I  L  F  N  N  N  P  L  P  A  V  C  S  L  V  C  I  H  E  D  Q  C

 301  GAAGGTAACTGTATAAAGGGAATTAAAGGGGAACCAATAAGATTTCATGATATAGAATATGAGATATCAGAGAAGTATTTAGAAGAGGTGAGATTTGAAA  400
       E  G  N  C  I  K  G  I  K  G  E  P  I  R  F  H  D  I  E  Y  E  I  S  E  K  Y  L  E  E  V  R  F  E  N

 401  ATTTACCAAAAGATAAAGATAGAATAGCAATAATTGGTGGAGGTCCGGCAGGAATAACAATTGCATTCATATTAGCTAATAAAGGTTATGATGTTACTAT  500
         L  P  K  D  K  D  R  I  A  I  I  G  G  G  P  A  G  I  T  I  A  F  I  L  A  N  K  G  Y  D  V  T  I

 501  TTTTGATGCACACGTGAAAATAGGTGGAGTACTTAGATATGGTATACCAGAATACAGATTGCCTAAAAAAATACTAGATACAATTGAAGATAGACTTGTT  600
        F  D  A  H  V  K  I  G  G  V  L  R  Y  G  I  P  E  Y  R  L  P  K  K  I  L  D  T  I  E  D  R  L  V

 601  GAATTAGGAATTAAAATTAGACCTAATACTCTTATTGGGCCAGTAATAACTCTTGATAGATTGTATGAGGATGGATATAAGGCTATATTTATTGGGACTG  700
       E  L  G  I  K  I  R  P  N  T  L  I  G  P  V  I  T  L  D  R  L  Y  E  D  G  Y  K  A  I  F  I  G  T  G

 701  GAGTATGGAATCCTAAGACTTTGAATATAAAGGGAGAAACTAGAGGAAATGTACATTTTGCTATAGATTATTTAAAGTCTCCAGAAGCGTATAGATTAGG  800
         V  W  N  P  K  T  L  N  I  K  G  E  T  R  G  N  V  H  F  A  I  D  Y  L  K  S  P  E  A  Y  R  L  G

 801  AAAAAAAGTTGCAGTTATAGGAGCTGGAAATGTTGCAATGGATGCAGCAAGAAGTGCAAGGAGAAATGGTGCTAAAGAAGTAACTGTACTTTATAGAAAA  900
        K  K  V  A  V  I  G  A  G  N  V  A  M  D  A  A  R  S  A  R  R  N  G  A  K  E  V  T  V  L  Y  R  K

 901  GGCTTTGATGAAATGAGTGCGACCAAACAAGAAATAATGGAAGCTAAGGAAGATGGAGTTATATTCAATTATTTCAAATCACCTATTGAAATAACAGAAA  1000
       G  F  D  E  M  S  A  T  K  Q  E  I  M  E  A  K  E  D  G  V  I  F  N  Y  F  K  S  P  I  E  I  T  E  K

1001  AAGGTATTAGATTATCATCTACAGAAAATGTAACAGATGAAAATGGTAAGATAAGAACTAAGGTTATTGAAGGAAAAGAGGAATTTTTTGAATGTGATTC  1100
         G  I  R  L  S  S  T  E  N  V  T  D  E  N  G  K  I  R  T  K  V  I  E  G  K  E  E  F  F  E  C  D  S

1101  AATAATAATTGCTGTAAGCCAAACTCCTAAAACAAATATAGTATCTAATACAAAAGAACTTAACACTAATAAATGGGGATTAATTATTACAGATGAAAAA  1200
        I  I  I  A  V  S  Q  T  P  K  T  N  I  V  S  N  T  K  E  L  N  T  N  K  W  G  L  I  I  T  D  E  K

1201  GGAAATACAACTAGAAAAGGTACATTCGCATCTGGAGATGTAGTCACAGGAGCCAAAACAGTTGTTGAAGCAGTAGTTCAAGCTAAAACTGTAGCAAATA  1300
       G  N  T  T  R  K  G  T  F  A  S  G  D  V  V  T  G  A  K  T  V  V  E  A  V  V  Q  A  K  T  V  A  N  T

1301  CTATAGAAGAATATTGTAAAAATAATTAAGAATTTGCGTAAATTGAATAAACTAAAAGTTCGAAATTAAAGTTTAGTTAGAGAAATATTTTACGAGAAA  1400
         I  E  E  Y  C  K  N  N  *

1401  AGTGTTGAATTTTTGATAAATTATGAATATATTTGCATTAATGTATTCATAATAAAACCTAGAAGTGTATTCTTCTAGGTTTTATTTATATACTACATTA  1500

1501  TTATATATTGTACTTC  1516
```

Fig. 3.4 Nucleotide and deduced amino acid sequence (single letter code) of the *C. acetobutylicum* P262 ORF*A* structural gene and flanking regions. The -35 and -10 regions of the putative promoter (underlined) are double underlined, and the putative SD sequence is indicated in boldface type. The stop codon is indicated by an asterisk. The terminal inverted repeat sequence is marked with converging arrows.

3.4.5 Deduced amino acid sequence of ORF*A*

The results of a search of the protein databases for sequences with homology to the 414 amino acid sequence predicted for ORF*A* were surprising: they revealed that the ORF*A* product shared a striking degree of homology over its entire length with numerous GOGAT β subunits and β subunit-like genes, ranging from approximately 30 to 44% identity (Table 3.2). The β subunit-like genes include several putative proteins which show a striking similarity to GOGAT β subunits, but which do not lie adjacent to a putative α subunit gene on the chromosome, and/or have not been identified biochemically. These are the criteria we used to define GOGAT β subunits consistent with Vanoni and Curti (1999).

We designated the gene encoded by ORF*A* *gltX*, to indicate its similarity to GOGAT encoding genes. The deduced *gltX* product, GltX, shared 32.5% identity (46.3% similarity) with the *C. acetobutylicum* P262 GOGAT β subunit protein, GltB. Although the two proteins shared similar homologies to a number of β subunits (for example those from *Pseudomonas*, *Thiobacillus* and *E. coli*), the homology trends appeared different. Whereas GltB shared the highest homology with the two cyanobacterial GOGAT genes (58.8 and 56.6% identity), GltX showed the highest degree of homology to proteins from the hyperthermophilic archaebacteria *Pyrococcus*; 41.5% identity (55.8% similarity) to the GOGAT GltA protein characterized from *Pyrococcus* sp KOD1, and approximately 42% identity to the two β subunit-like proteins identified at separate loci in the genomes of both *P. horikoshii* OT3, and *P. abysii* (Table 3.2). Furthermore, GltX showed higher homology to the single β subunit-like homologues present in the hyperthermophilic bacteria *Thermotoga maritima* (41.7% identity) and *Aquifex aeolicus* (39.1% identity), indicating that *gltX* appears most closely related to genes found in the hyperthermophilic archaebacteria.

Table 3.2 Percentage identity (ID) and similarity (SIM) of the deduced amino acid sequences of the *gltX* and *gltB* genes from *C. acetobutylicum* P262 to GOGAT β subunit-like sequences, and to known GOGAT β subunits.

Organism	Name (abbreviation)	Protein	Accession Number	Length (aa)	*C. acetobutylicum* P262			
					GltX (414 aa)		GltB (491 aa)	
		β subunit-like[a]			%ID	%SIM	%ID	%SIM
Archaea	*Pyrococcus horikoshii* OT3 (Phori)	GltX1	dbj:AP000004	472	43.4	56.4	30.0	41.0
Archaea	*Pyrococcus horikoshii* OT3	GltX2	dbj:AP000007	476	41.5	55.8	30.0	40.7
Archaea	*Pyrococcus abysii* (Pabys)	GltX1	emb:AJ248288	475	42.0	55.5	30.2	41.1
Archaea	*Pyrococcus abysii*	GltX2	emb:AJ248286	474	42.3	53.5	32.8	44.1
Archaea	*Pyrococcus* sp KOD1 (PKOD1)	GltA	dbj:D86223	481	41.4	55.8	33.9	46.0
Bacteria	*C. acetobutylicum* ATCC824 (Ca824)	GltX	(AE001438)[b]	411	77.5	85.2	32.2	44.9
Bacteria	*Thermotoga maritima* (Tmari)	GltX	gb:AE001806	468	41.7	53.4	33.8	42.8
Bacteria	*Aquifex aeolicus* (Aaeol)	GltX	gb:AE000770	476	39.1	50.3	31.6	41.2
Bacteria	*Treponema pallidum* (Tpall)	GltX	gb:AE001245	518	35.9	48.3	38.6	48.8
Bacteria	*Synechocystis* sp PCC6803 (Synec)	GltX	dbj:D90900	494	35.8	47.8	58.8	66.1
Bacteria	*Escherichia coli*	GltX	gb:AE000303	412	32.4	44.0	30.3	38.2
Bacteria	*Rhodobacter capsulatus*	GltX	gb:AF031406	443	27.9	38.1	33.3	40.0
		β subunit						
Bacteria	*Bacillus subtilis* (Bsubt)	GltB	emb:Z99113	493	37.8	48.8	53.4	62.3
Bacteria	*Plectonema boryanum* (Pbory)	GltD	dbj:D85230	492	37.3	48.3	56.6	65.9
Bacteria	*Pseudomonas aeruginosa*	GltD	gb:U81261	477	35.0	47.4	37.8	48.1
Bacteria	*Thiobacillus ferrooxidans* (Tferr)	GltD	pir:JC5184	465	34.1	47.9	35.1	47.4
Bacteria	*Azospirillum brasilense* (Abraz)	GltD	sp:Q05756	481	33.6	47.3	39.7	49.3
Bacteria	*Escherichia coli* (Ecoli)	GltD	sp:P09832	472	33.3	46.2	38.4	48.3
		β subunit domain						
Eukaryota	*Medicago sativa* (Msati)	GlsN	sp:Q03460	2194	36.8	48.6	53.9	63.9
Eukaryota	*Saccharomyces cerevisiae* (Scere)	GLT1	pir:S61041	2145	33.2	43.7	47.6	56.6

The sequences can be retrieved from the GenBank (gb), EMBL (emb), SwissProt (sp), DDBJ (dbj) and PIR (pir) databases using the accession numbers supplied.

a: Deduced protein sequences of genes that share significant homology to GOGAT β subunits, but which do not lie adjacent to an α subunit homologue, have been labeled GltX, with the exception of the *Pyrococcus* sp KOD1 GltA protein for which GOCAT activity has been demonstrated.

b: The accession number AE001438 refers to the unfinished genome sequence of *C. acetobutylicum* ATCC824.

3.4.6 Identification of conserved domains within *gltX*

Analysis of the deduced *gltX* amino acid sequence revealed that the proposed functional domains characteristic of GOGAT β subunits, and discussed in detail in Chapter 2 (Section 2.4.5), were also conserved within this β subunit-like protein (Fig. 3.5). Two N-terminal cysteine clusters $CX_2CX_4CX_3CP$ (residues 15-27) and $CX_3CX_5CX_3C$ (residues 62-76), matched the conserved patterns of cysteine-rich regions proposed to be involved in the formation of [4Fe-4S] clusters. Similarly, two regions matched the conserved sequence requirements for the formation of an ADP-binding fold for the binding of FAD (residues 113-141) and NAD(P)H (residues 241-271), respectively.

The second conserved ADP-binding fold has been confirmed to bind the cofactor NADPH in the *A. brasilense* GltD protein (Morandi *et al.*, 2000). Since GltX shares features with the *A. brasilense* GltD protein that have been proposed to confer specificity for the reducing coenzyme NADPH, rather than NADH (the presence of an A instead of a G in the last position of the motif GXGXX(G/A/P) and an arginine residue at position 271) (Pelanda *et al.*, 1993), it suggests that this may be a functional site for NADPH binding in GltX. Furthermore, a region in the COOH-terminus (residues 375-385), matches 6 out of 7 positions the second FAD binding consensus sequence, however, the mismatch, viz., a T in the second position of the motif ($TX_4I/VFAGGD$), also occurs in the *S. cerevisiae* sequence. Interestingly, in these pileups (Fig. 3.5), GltX groups with the *E. coli/A. brasilense/T. ferrooxidans* subclass of β subunit proteins proposed in Chapter2, as do the β subunit-like proteins.

```
                                                                      C  C     C              C
Pbory-B  : ~~MGKPTGFIEYLRELPSELAELD IRN DEFHL-SMPEENLRTQ A   D GIPFCHTGTLISGMASGCPINNLIPEW :   76
Synec-X  : ~~MGKPTGFLEYTREIPQELSEGD LRN DEFHV-TMPDKQVETQ A   D GTPFCHTGSLISGMASGCPINNLIPEF :   76
Msati-B' : TDAVKHRGFVAYERGVQYRDPNV LND NEVMMETKPGPLLKTQSA   D GTPFCHQEN------SGCPLGNKIPEF : 1748
Bsubt-B  : ~~MGKPTGFMEIKREKPAERDPLT LKD KEYSAP-FSEEASKRQ A  D GTPFCQIGADINGFTSGCPIYNLIPEW :   76
Cacet-B  : ~~MGKPTGFLEYDREVGRNRPKE LKD KEFH-QRLPLEKQCIQ A  D GVPFCQAGVLFSGMVSGCPLHNLIPEW :   76
Scere-B' : ERIEKTRGFMIHKRRHETHRDPRT VND KEF-TNPITKKDKKYETA  D GTPFCLSDT-------GCPLSNKIPKF : 1705
Ca824-X  : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MDNPNLLSEE N  L KNE--RCKANCPINTPIPEIISLYKEG :   42
Ca262-X  : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~M-TCNDIQLLDE D  L KKE--RCKENCPIQTSIPEIISLYKEN :   43
Ecoli-B  : ~MSQNVYQFIDLQRVDPPKKPLKI KIE V IYE-PFSEGQAKAQ D  S GNPY--CEWKCPVHNYIPNWLKLANEG :   75
Tferr-B  : ~~~MSHFSFIEDPRQDPKKLDVEE REA R IYQ-SFDLGSAQLQ D  H GNPY--CEWKCPVHNYIPNWLQLIVEN :   73
Abraz-B  : MANQRMLGFVHTAQRMPDKRPAAE RQD A IYA-RFSDERANEQ N SQ GVPF--CQVHCPVSNNIPDWLKLTSEG :   76
Pabys-X  : ~~~~~~MPKLIKERVPTPERSPVE VRD N VNL-GYTWELALRE    Q PKEYAPCIKGCPVNIDIPGFIRALREN :   72
PKOD1-X  : ~~~~MPRRKLIKERVPTPERPPEE IKD K VNL-GYTFELAVKE E  Q PYEYAPCIKGCPVHINIPGFISKLVEY :   74
Phori-X  : ~~~~~~MPKLIKERVPTPERPVEE IKD K VNL-GYTFELAKKE E  Q PPEYAPCIKGCPVHINIPAFIAKIKEG :   72
Tmari-X  : ~~~~~~~~~MKNRKTPMKEQSPES RRN E VAL-GYTLEEALEE Q  Q PTH--PCVSGCPVEIDIPGFIRKLRNG :   67
Tpall-X  : GKELKMKDRTQIPLQRMRMLPEKK SLLMQ AAL-GFTEQQALVESQ  N KTK--PCVKGCPVGVPIPEFIACVQRG :  121
Aaeol-X  : ~~~~MAKKIRYLDRNKEPTLPEEE VKT R FEL-GYSVNLALDEPQ  F KDAEQRCIKGCPVHVDIPGEIKKITEG :   74
                                                                      C  C       C   C

Pbory-B  : NDLIYRGLWKE LDR HKT NF EFT    AP--   S V --------GIHNPE  KN  YS A K WNE WIKP- :  144
Synec-X  : NDLVYRGLWPE LDR HKT NF EFT    AP--   S V ---- ---GINNPP  KN  YS I K WQE MVTP- :  144
Msati-B' : NELVYQNRWQE LER LET NF EFT    AP-- E S V --------GIIENP S KN  CA I K FE  MIP- : 1816
Bsubt-B  : NDLVYRGRWKE  ER LKT NF EFT    AP--   S T --------AISDPA S KN  T  I K FEN WIQP- :  144
Cacet-B  : NDLVYRGKWEL  ER NKT PF EFT    AP--   G TA--------GLNGPA I KEN RS I NAFEN EVKA- :  144
Scere-B' : NELLFKNQWKL LGK LEE NF EFT    AP--     T --------GIIEDE G KS  G I NAFKE WIKP- : 1773
Ca824-X  : K------I-ME GGI FNN P SVICSL  IHEDQ K N V--------RGIKSEE KFHE EE S--------KYL :   98
Ca262-X  : R------L-EE GGI FNN P AVCSL  IHEDQ   N I--------KGIKGEE RHD YE S--------KYL :   99
Ecoli-B  : R------I-FE AELSHQT T  EVC     QQRLIE S T-NDEF--------GA  N     N KAFEM RPDM :  139
Tferr-B  : R------L-EE ATRSHQT T  SIC     QQRL   A T-NATG--------GA  N     T EALAR HQPDM :  137
Abraz-B  : R------L-EE YEVSQAT NF EIC     QQRL   N V EQSTH--------GA  GS    N TAWEQ HVKPR :  141
Pabys-X  : RDNPRRAV-EE LRV WAC SL ATT    QEDQ   V V ----GKV--GD--F N        AIYARE  IDDEL :  142
PKOD1-X  : RDDPDKAV-EE LNV WAC SL ATT    QEDQ  MN V ----GKV--GD--K N        AIYARE  IDEEL :  144
Phori-X  : -------DI-EG LRI WND TL ATT    QEDQ   A V ----GKV--GD--F N        AIYARE  IEEEL :  136
Tmari-X  : K------L-EESYRE KSY NL AVC     QEVQ  SR V ----GKM--KDSEE N        AIWAAE-----L :  128
Tpall-X  : -------AF-EE VEV KTESL  AIC     HERQ QLQ T ----GRMFKDVSKA  SA       AIWERQH----- :  184
Aaeol-X  : -------DL-VG YKV TQSDIF SSC     QERQ   S I YYDTVRNRKNKGLP  SA       GIFIRIS IEIEE :  146
                                  (C)   C       C   C

                        ==========  FAD-I ==========
Pbory-B  : ----EPPAKRTGK       S     SAEAQ NKV HW    RADRF          NM   QE  MRRLNI EAE K :  219
Synec-X  : ----EPPAKRTGK       S     AAEAQ NKA HW    REDRF          NM   EE  LSR NI EAE T :  219
Msati-B' : ----RPPVKRTGK       S  S  AAEDQ NKA HI    RADRI          NM   VE  QRRVNL AEE N : 1891
Bsubt-B  : ----RIPKKRTGK       S     ASEDQ NQA HS    RADRA   T      NM  -G  ERRIKL TQE D :  218
Cacet-B  : ----NTPLKRTGK       S  S  AVENT NKC HK    RSDRI          NM  -E  LSRVHL AEE K :  218
Scere-B' : ----CPPSTRTGKT      S     ACEDM NRA HT    RSDRC          NM  A-  QRRIDL SAE D : 1847
Ca824-X  : KEARLKNVQKDKD             TVI VI ANK YN    AHDRI   R      E  T-K  DILEER IEV K :  176
Ca262-X  : EEVRFENLPKDKD             TIE FI ANK YD    AHVSI   R      E   -K  DTIEDR VE  K :  177
Ecoli-B  : -----SGVKQAG              CIDV TKN VKAV    GHPEI          A   -E  TPEREFTGI E :  212
Tferr-B  : -----PDILPNGK     SGGS    CIDI NPA TE V    YPAV    T      SF  -A  AFEAN QEL R :  210
Abraz-B  : -----TPSRELSLS    SGGS    AAEEE RAK YE H    YDRM    V      GF  -S  AEAVKK ADA I :  214
Pabys-X  : LMEEIRGIKENGR      SAGS    TCIAE AKM YE I    ALHEP   A      EF  -S  RREIKK KFI E :  220
PKOD1-X  : LFEIIPKIEKKDQ      SAGS    TAIGE AKL YD I    ALHEA          EF  -E  HKEIEK RKE K :  222
Phori-X  : LREFTEKCDGEKG             TCIGE AKM YK T    ALHRP   A      EF  -S  DHELKK KML E :  214
Tmari-X  : EEDVKPLAGSKKE      S      TAIAD AKM YH D    AFHRP   V      EF  -R  EEEVSY RKE N :  206
Tpall-X  : QITVPYCAPSEHE      S      AVISDTARA HS      ALHKE   T   I  EF  -E  VTEIET KKM T :  262
Aaeol-X  : EKE-----PPEEF      S      SCIHE ARK HE H    ALPKE   Y      NA  -S  EWEVKR EKE K :  219
                               ++ + + +  +    +  +     + + +

Pbory-B  : EVCSTE  KD--EPAEDLLRDFE  V CT ATKP D  E RQ- KE HF      TANEQAV DQT-EPA----ITAQ  :  290
Synec-X  : EVCSTE  KD--EPPETLLKDYE  V CT ATKP D  A RE- EE HF      TANT AI DKT-PGPNF--ISAKD :  292
Msati-B' : EVVNAN  LDPLYSLERLREENE  V AV ATKP D  P RE- SE HF      HANE SL DSNLQDGNY--ISAKE : 1967
Bsubt-B  : EVTNTE  VD--ETADELKEQFE  I CT AQKQ D  E RE- SKS HY      TLAE SY DSNFKDKQF--IDAKE :  292
Cacet-B  : EVNHAN  --DNYDAKEILDEFE  V VT ASQE D  A RE- EN HF      KANE SL DSDHEDNNY--ISAKD :  293
Scere-B' : EVTNTE  K--TSMDELKNKHNE  VYAI STIP D  P RE- KN DF  MQL ESNT AL NKDLEIIRE---KIQE : 1920
Ca824-X  : IRPSKV  DVKSDRELED-SYK  F ST VWNE T DVK -ETGN HF      KSPESYR-------- -------- :  238
Ca262-X  : IRPNKL  EVITDREYED-GYK  F ST VWNE T NIK -ETRGN HF     KSPEAYR---------------- :  239
Ecoli-B  : EKLNTE  RDVQDDELSDY--E  E EV TYQSMRGGLEN-EDADE YA  LP IANT -Q MGFGETRDEPFVSME :  287
Tferr-B  : EQLDTE  QDISGKELEDF--E  F EV TYTALTGDFP- EKHPE LK  SI IANT -H MQLPEP-TLPYISLS  :  284
Abraz-B  : EHPNFE  RDASPEERRKH--V  L AT VYKAD KAP--SG GN VA    TTSN VS GDTVEAYENGSLNAA  :  290
Pabys-X  : IKTDHI  RTVTPEELEEY--E  F ST AGTE LKEP--IN NS YS      ---T IN MKAYKFPEYDTPIVV :  293
PKOD1-X  : ILTDHV  RTVTEELLQEY--E  F SE AGTE  NAP--IN NE YT      ---T VN MKAYLFPEYDTPVYV :  295
Phori-X  : IRTDHV  RTVTDEELQEY--E  F ST AGTE  NEP--IL DR YS      ---E IN MKAYEFPEYDTPIAV :  287
Tmari-X  : EHLTEV  ETVKKEELSEY--E  F ST AGTE  GIP--TN NS YS      ---E VN MKAYLFPEYDTPIRV :  279
Tpall-X  : ERMEFL  ETAREQEFSQYGFE  F SI AGLE  NEE--EE CG FA      ---E AT MKAYETAHADTPVYA  :  337
Aaeol-X  : EFYGYL  KNITQEERREKY--E  E AV AGRG  GLEP--EH KE YS      -- M VN YKANEFPKKGVPVEL  :  291
```

Fig. 3.5 Comparison of the deduced amino acid sequence of the *C. acetobutylicum* P262 *gltX* gene (Ca262-X) with bacterial GOGAT β suuunits (-B), eukaryotic GOGAT β subunit domains (-B') and β subunit-like proteins (-X). A full description of the abbreviated protein sequences presented, together with their accession numbers, are supplied in Table 3.2. The Cys residues that are proposed to participate in the formation of 4Fe-4S clusters are indicated by a C above and/or below the sequences. The proposed FAD and NAD(P)H binding domains are described in the legend to Fig. 2.6. The black and gray backgrounds represent conserved residues in a minimum of 80 % and 50 % of the sequences aligned, respectively.

Thus, it seems reasonable to suppose that GltX also binds NADPH and FAD, and is involved in an electron transfer mechanism similar to GOGAT β subunits. This is supported by the results

that metronidazole is reduced by *E. coli* F19 cells specifically expressing the *gltX* gene, and the fact that loss of the C-terminal 13 amino acids of GltX (construct pSKMETS92 in Fig. 3.1), which disrupts the conserved C-terminal FAD binding region, correlates with loss of metronidazole sensitivity. It is not evident, however, whether GltX is capable of direct enzymatic reduction of metronidazole, as reported for pyruvate:Fd oxidoreductase (Narikawa, 1986), or whether it occurs indirectly via other low potential electron carriers. Whatever the mechanism, it does not appear to require a GOGAT α subunit, since a similar metronidazole sensitive phenotype was observed when *gltX* was introduced into *E. coli* strain AN3001, which is unable to synthesize either of the GOGAT subunits (Table 3.1).

The question arose of whether *gltX* represented a second GOGAT β subunit protein in *C. acetobutylicum* that may function under different growth and/or substrate conditions to GltB; e.g. GltB appears to be reduced by NADH, and GltX by NADPH. However, there are three noticeable features that distinguish it from the rest of the β subunit proteins aligned in Fig. 3.5. First, the protein is approximately 27-33 amino acids shorter at its amino terminus, resulting in the loss of three strictly conserved residues; two arginine residues and one glutamine residue. Second, some residues within the second cysteine cluster domain are not conserved. Third, *gltX* contains three gaps at positions 95, 237 and 303. Although they do not lie within highly conserved regions, it results in the absence of certain residues conserved in some of the subunits. It remains to be established whether these differences would distinguish it from functioning as a GOGAT β subunit. It is worth remembering here that the β subunit is important for modulating the properties of the α subunit and the holoenzyme. Whatever the biochemical function of this β subunit-like protein is, these uncharacteristic features are completely conserved in a homologous similar-sized gene (coding for a putative 411 amino acid protein) identified in the unfinished genome sequence of *C. acetobutylicum* ATCC824, which shares 77.5% identity (85.2% similarity) with the *gltX* gene (Fig. 3.5). Furthermore, its homologue has been identified in *C. beijerinckii*.

3.4.7 *gltX* homology in *C. beijerinckii*

Southern blot analysis indicated that the *gltX* gene from *C. acetobutylicum* P262 is highly conserved in the genome of *C. beijerinckii* NCIMB 8052 (Fig. 3.6), which is representative of a group of related strains widely used in molasses-based ABE fermentations (Keis *et al.*, 1995). As expected, a *gltX*-specific DIG-labeled probe detected *C. acetobutylicum* P262 chromosomal fragments overlapping with the *gltX* gene locus; a ~ 4.7 kb *Eco*RI fragment (lane 3) and a ~3.9

kb *Hind*III fragment (lane 6) (see Fig. 3.1 and 3.7 for restriction map details). No fragments corresponding to the homologous *gltB* gene locus (illustrated in Fig. 2.1) were detected using stringent hybridization (60 °C) conditions. Thus the single specific ~ 3.6 kb *Eco*RI (lane 4) and ~ 1.7 kb *Hind*III (lane 7) fragments of *C. beijerinckii* genomic DNA detected, must represent a gene highly homologous to *gltX* in this organism. (A locus encoding adjacent genes homologous to the GOGAT α and β subunits has been identified in *C. beijerinkii*, and is distinct from this *gltX*-like locus (K. Quixley, personal communication)). Clearly *gltX* is highly conserved within *C. acetobutylicum* ATCC824 and *C. beijerinkii* 8052, suggesting that it has a significant role to play in the Clostridia.

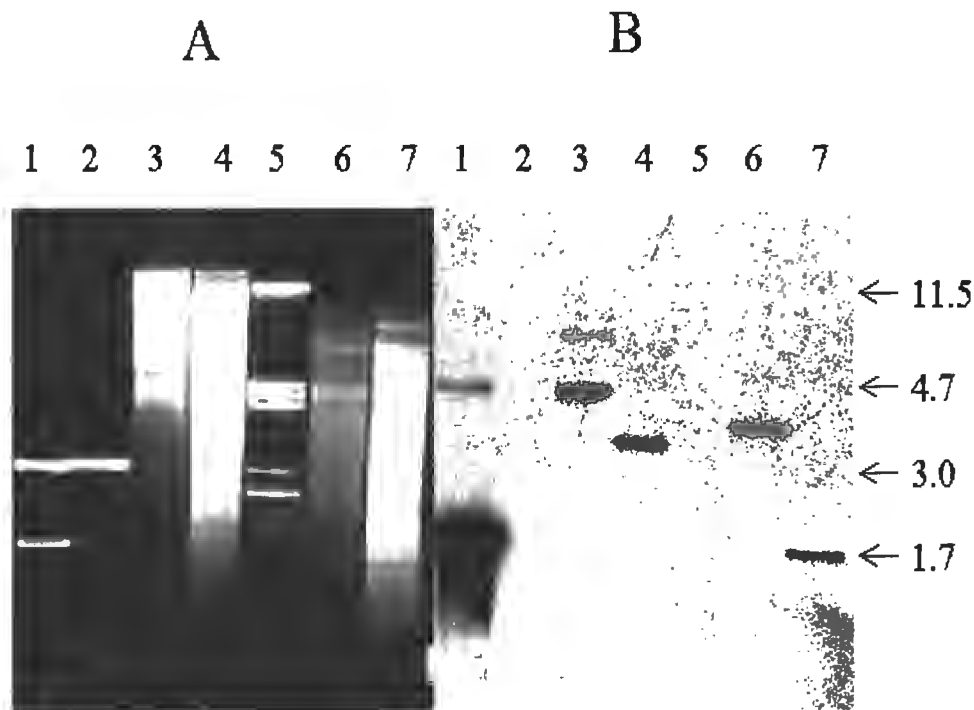


Fig. 3.6 Detection of *gltX* homology in the *C. beijerinckii* NCIMB 8052 genome by Southern hybridization. The agarose gel is presented in A, and the corresponding autoradiograph in B. Lanes 3 and 6 represent 30μg each of *C. acetobutylicum* P262 chromosomal DNA digested with *Eco*RI and *Hind*III, respectively. Similarly, lanes 4 and 7 represent 30 μg each of *C. beijerinckii* chromosomal DNA digested with *Eco*RI and *Hind*III, respectively. The DIG-labeled *gltX* probe was synthesized from plasmid construct pSKMETS92, and consisted of a 1.26 kb fragment extending from a *Ssp*I site positioned 60 bases upstream of the *gltX* initiation codon, to the *Kpn*I vector site bordering the downstream cloning junction. Lane 1 contains *Pst*I-*Kpn*I digested pSKMETS92 DNA as a probe control. Lane 2 contains vector pSK linearized with *Hind*III. Lane 5 contains *Pst*I-digested lambda DNA as size markers which are also indicated along the margin (in kb).

3.4.8 Cloning and sequencing of the chromosomal DNA regions flanking *gltX*

The existence of a second GOGAT β subunit-like gene raised the question of whether another α subunit homologue may also exist. Since in bacteria the genes encoding the α and β subunits are located adjacent to each other, we sought to identify the regions flanking *gltX* to define it in context with its neighbouring genes, and possibly gain further insight as to its function.

To isolate the flanking regions, we employed the chromosome walking technique discussed in Chapter 2, and graphically presented in Fig. 3.7. A ~9 kb *Pst*I-*Cla*I fragment which included approximately 6 kb of DNA upstream of *gltX*, and a ~1.4 kb *EcoR*I fragment which extended upstream from pSKMETS, were isolated by colony hybridization of appropriate size-selected genebanks of *C. acetobutylicum* chromosomal DNA, using the 0.52 kb *Pst*I-*Eco*RI DIG-labeled fragment of plasmid pSKMETSUP1. Two positive clones were recovered out of 550 recombinants screened from each of the genebanks. These clones were confirmed by Southern blot hybridization (Fig. 3.8); they all contained the common 1.4 kb *EcoR*I fragment. Two plasmids designated pSKMETSUP2 (harbouring the ~1.4 kb *EcoR*I fragment) and pSKMETSUP3 (harbouring the ~9 kb *Pst*I-*Cla*I fragment)(Fig. 3.7), were selected for subsequent sequencing analysis. The majority of the 5.4 kb region extending upstream of the pSKMETS insert was sequenced on both strands, however two small gaps (estimated to be ~150 bp each) remain near to the *Eco*RI(b) border. Furthermore, the 1.5 kb region downstream of *gltX*, and the 0.5 kb region immediately upstream of *gltX*, was sequenced in both directions, using deletion clones previously generated for pSKMETS (Chapter 2).

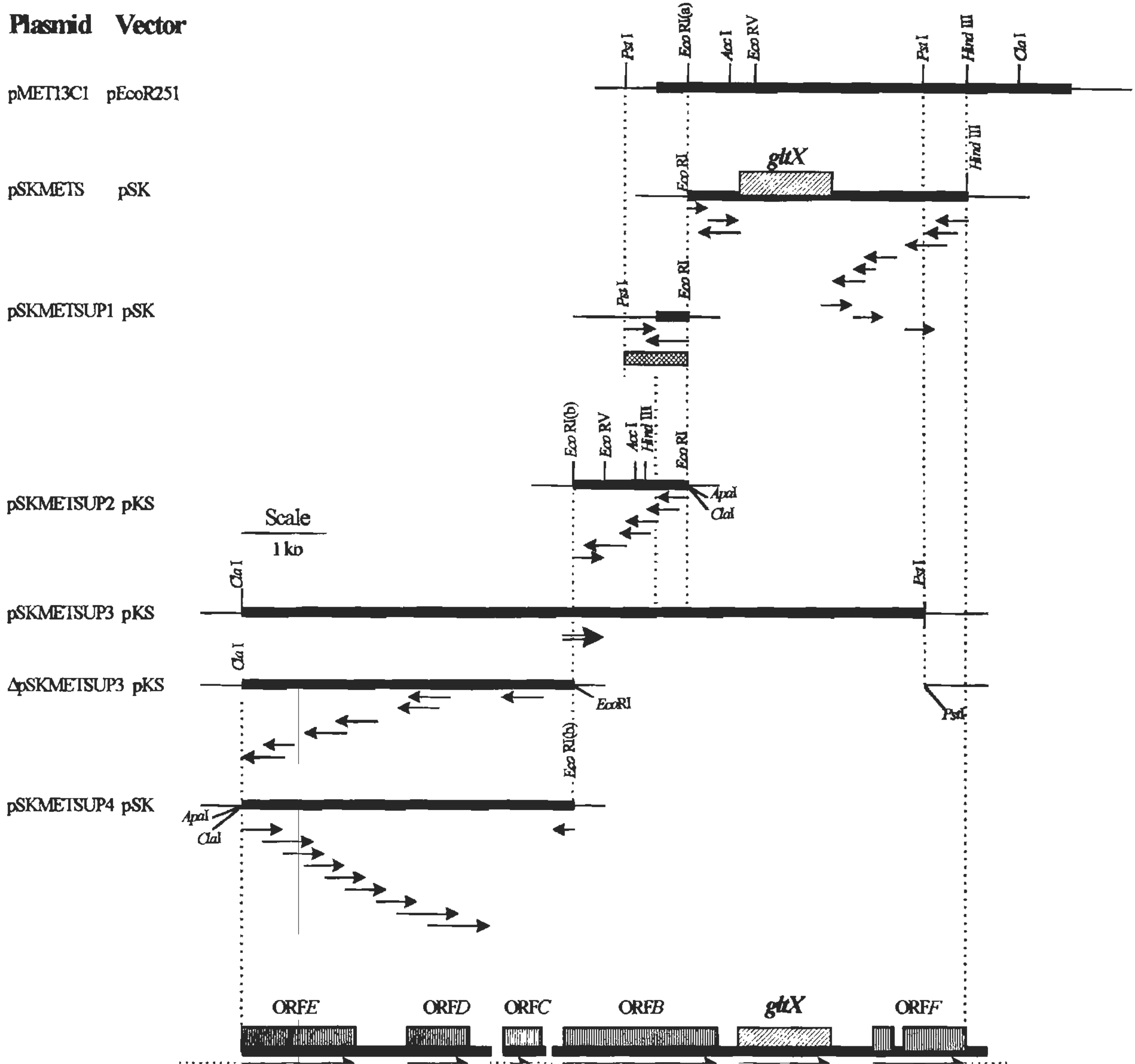


Fig.3.7 Strategy for cloning and sequencing the *C. acetobutylicum* chromosomal DNA regions flanking *gltX*. The 0.52 kb *Pst*I-*EcoR*I(a) fragment, containing the terminal 300 bp of the pMET13C1 insert DNA, and including 220 bp of vector sequence, was subcloned into the corresponding unique sites in the vector pSK to yield plasmid pSKMETSUP1, from which the probe (indicated by a hatched bar) was derived. This probe was used to isolate the two Bluescript clones, plasmid pSKMETSUP2 containing a 1.4 kb *Eco*RI fragment, and plasmid pSKMETSUP3 harbouring a 9 kb *Cla*I-*Pst*I fragment. For sequencing purposes, a deletion derivative of pSKMETSUP3 was created (ΔpSKMETSUP3), and the 4 kb *Cla*I-*Eco*RI(b) fragment of pSKMETSUP3 was subcloned into the corresponding sites in pSK to yield plasmid pSKMETSUP4. The thin horizontal lines represent vector sequence. Relevant restriction enzyme sites are indicated. The unique 3'(*Apa*I, *Pst*I) and 5'(*Cla*I, *Eco*RI) overhang restriction sites used to generate nested deletion clones for sequencing are included diagonally below the respective plasmids at the vector-insert junctions. The thin horizontal arrows indicate the direction and extent of sequencing obtained from the various deletion clones. The hollow arrow indicates the region sequenced using an oligonucleotide primer. The relative positions of identified ORFs are indicated below, together with their transcriptional polarities.

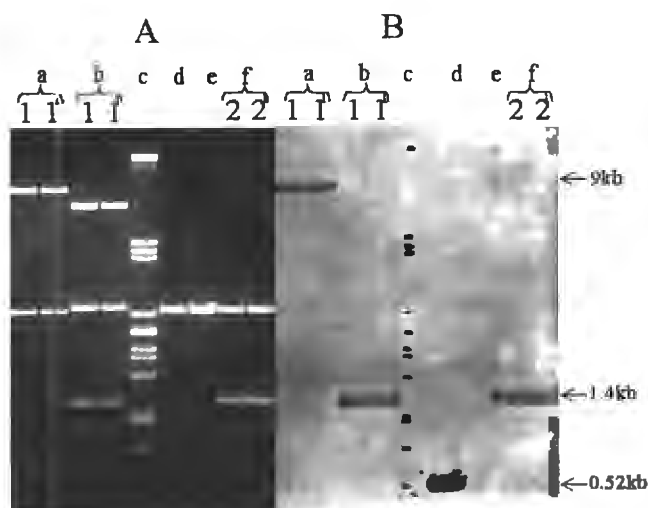


Fig. 3.8 Confirmation of clones harbouring *C. acetobutylicum* chromosomal DNA upstream of *gltX* by restriction digestion (A) and corresponding Southern blot analysis (B). Numbers 1 and 1' represent two positive clones identified from the *Pst*I-*Cla*I constructed genebank in vector pKS. They were digested with *Pst*I-*Cla*I (a), releasing a ~ 9 kb insert, and *Pst*I-*EcoR*I (b). Similarly, 2 and 2' represent two clones identified from the *EcoR*I constructed genebank. They were digested with *EcoR*I (f), releasing a ~1.4 kb insert. The 0.52 kb *Pst*I-*Eco*RI DIG-labeled fragment of plasmid pSKMETSUP1 (Fig. 3.7) served as the probe. Lane (c) contains *Pst*I-digested lambda DNA. Lane (d) represents pSKMETSUP1 digested with *Pst*I-*EcoR*I as a probe control. Lane (e) contains linearized vector (pSK) DNA.

Taken together, in the ~ 6 kb region upstream, and the ~ 1.5 kb region downstream of *gltX*, 2 complete ORFs (ORF*B* and ORF*D*), a section of an ORF (ORF*C*) and two truncated ORFs (ORF*E* and ORF*F*) were identified (Fig. 3.7). All the ORFs are transcribed in the same direction with respect to *gltX*. These predicted coding regions were consistent with TESTCODE (a function of GCG) analysis. (The TESTCODE program identifies protein coding sequences independently of the reading frame by plotting the nonrandomness of the composition at every third base). The nucleotide sequence and deduced amino acid sequences of these ORFs is included in Appendix D.

3.4.9 Identification of the genes flanking *gltX*

None of the identified ORFs flanking the *gltX* gene showed similarity to GOGAT α subunit genes, nor to genes related to nitrogen metabolism (summarized in Fig. 3.9). Immediately upstream of *gltX*, and separated by an intergenic region of 214 nucleotides, lies ORF*B* (1779 bp). Amino acid homology searches of the deduced 592 aa ORF*B* product revealed significant similarity to a hypothetical 612 aa protein, YitJ, from *B. subtilis* (30% identity) (accession no. CAA70665). A homologue (598 aa) was also detected in the unfinished *C. acetobutylicum*

ATCC 824 genome sequence (57.4% identity)(accession no. AE001438). Interestingly, ORF*B* appears to comprise two domains, both of which show similarity to enzymes from the vitamin B12-dependent methionine biosynthesis family. The amino terminal ~280 residues shared striking homology to the amino terminal regions of *metH* genes encoding methionine synthase (5-Methyltetrahydrofolate—Homocysteine methyltransferases); e.g. 32% identity to the terminal 280 amino acids from *Mycobacterium leprae* MetH (SwissProt accession no. Q49775). This region may be involved in binding either methyltetrahydrofolate or homocysteine (Drummond *et al.*, 1993). The latter half of ORF*B* consistently showed homology to MetF proteins (5,10-methylenetetrahydrofolate reductase) e.g. the carboxy terminal 278 amino acids shared 30% identity to the *metF* gene (296 aa) from *A. aeolicus* (accession number AE000740). MetF catalyses the synthesis of N-methyl tetrahydrofolate which is the donor of the methyl group of methionine. Situated 507 bp downstream of the *gltX* stop codon, lies the truncated ORF*F* (~1145 bp). According to amino acid homology searches this appears to code for the amino terminal region of an amino acid transport protein.

The *gltX* locus does not show any similarity to sequences present adjacent to β subunit or β subunit-like genes present in the databases, including that of the *gltA* gene from *Pyrococcus* sp KOD1 (Jongsareejit *et al.*, 1997) and the highly homologous *gltX*-like gene from *C. acetobutylicum* ATCC824 (Fig. 3.9). This suggests that the *gltX* gene from *C. acetobutylicum* P262 may not be associated with any other genes of related function.

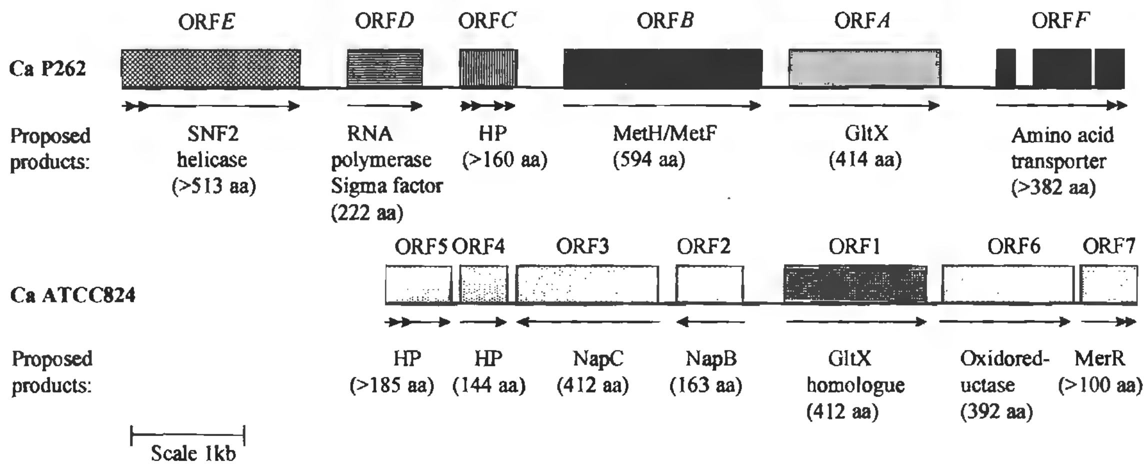


Fig. 3.9 Comparison of the *C. acetobutylicum* P262 *gltX* locus (Ca P262) to the regions flanking the homologous *gltX*-like gene in *C. acetobutylicum* ATCC 824 (Ca ATCC824). Proposed products were assigned to the various ORFs based on sequence similarity to proteins present in the data bases. *C. acetobutylicum* P262 ORF*B* and ORF*F* are discussed in the text. The ORF*D* gene product showed homology to RNA polymerase sigma factors (38% identity to a 251 aa protein from *B. subtilis*, accession no. CAB13218). The truncated ORF*E* product displayed strong homology with the conserved carboxy terminal domains of helicases of the SNF2 family (43.4% identity to the SNF2 protein from *B. cereus*, accession no. CAA67095). *C. acetobutylicum* ATCC 824 ORF1 shared 77.5% identity with the *gltX* gene product. ORF3 and ORF2 showed similarity to a tetracycline efflux protein NapC (54.8% identity) and a regulatory protein NapB (44.8% identity), respectively, identified in *Enterococcus hirae* (accession no. AJ000346). ORF6 shared 36% identity with an oxidoreductase of the aldo/keto reductase family from *Thermotoga maritima* (accession no. AE001775), while the truncated ORF7 shared 50% identity to the *B. subtilis* mercuric resistance operon regulatory protein MerR (accession no.X92868). HP indicates a hypothetical protein. Arrows show the direction of transcription and double arrow heads indicate where genes are truncated.

3.4.10 Determination of *gltX* transcript size

In order to determine whether *gltX* forms part of a larger operon, Northern blots were performed on total RNA isolated from various *C. acetobutylicum* P262 cultures, using a *gltX* specific DIG-labeled RNA probe (Fig. 3.10). A single specific hybridization signal of approximately 1.4 kb, was detected in two of the Clostridial samples (lanes 7 and 8), as well as in the *E. coli* sample harbouring plasmid pSKMETS (lane 4). These results are in agreement with the estimated transcript size of 1.45 kb, which includes the putative promoter and transcriptional termination sequences. Thus *gltX* is independently transcribed. This is consistent with the presence of an inverted repeat sequence with the potential to form a stem loop structure (ΔG = -13.1 kcal/mol) in the intergenic region between ORF*B* and *gltX* (Appendix D), and previous indications that the *gltX* gene is expressed from its own promoter. These results

also indicate that, under these growth conditions, *gltX* is preferentially expressed at late log growth phase in *C. acetobutylicum* (lane 8).

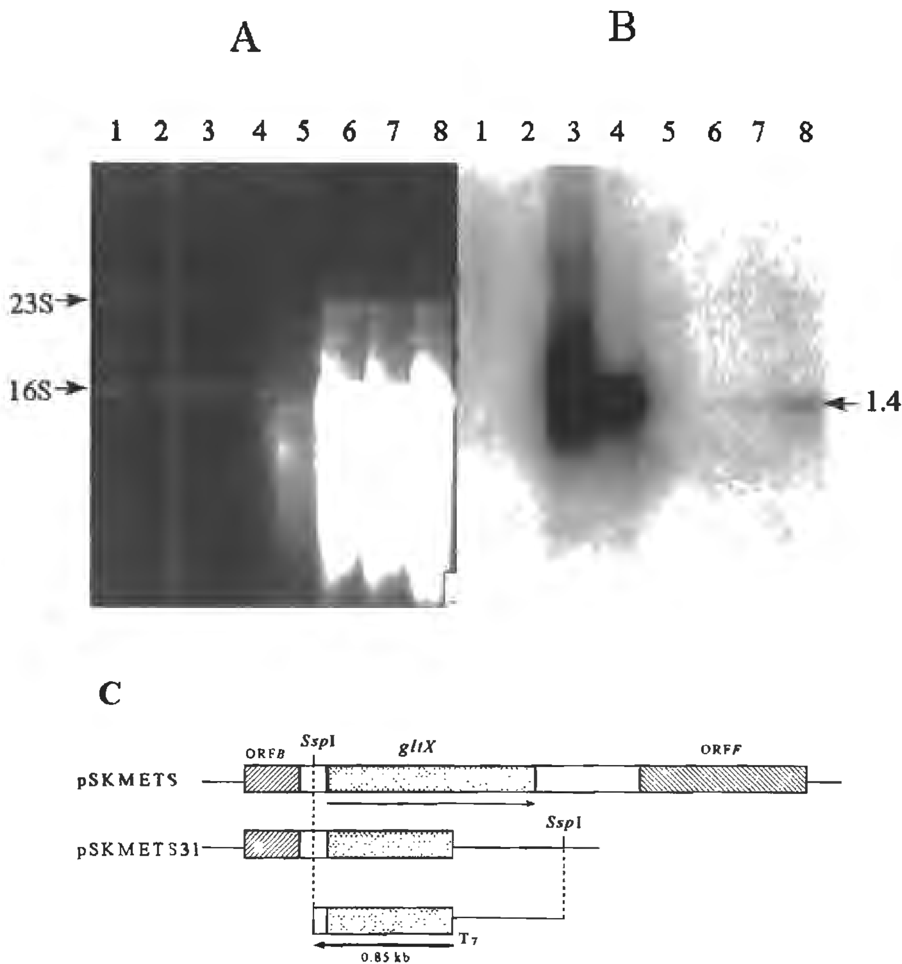


Fig. 3.10 Northern blot showing the size of the *gltX* transcript. The RNA gel is presented in A, and the corresponding autoradiograph in B. Lanes 1 to 4 represent 20 μg each of RNA extracted from overnight cultures of *E. coli* F19 (lane 1), and F19 harbouring plasmids pSK (lane 2), pSKMETS31, in which the terminal ~350 bp of *gltX* are deleted (lane 3), and pSKMETS (lane 4). Lanes 6 to 8 contain 50 μg each of total RNA extracted from *C. acetobutylicum* cultures grown to OD_{600} 0.35 (lane 6), OD_{600} 2 (lane 7) and OD_{600} 4 (lane 8). The RNA was hybridized with the 0.85 kb *gltX*-specific DIG-labeled RNA probe construct illustrated in C. The position and size (kb) of the *gltX* gene transcript, as well as the 23S and 16S ribosomal RNA bands, are indicated. Lane 5 contains RNA molecular weight markers (Boehringer Mannheim).

3.4.11 Probing for a second GOGAT α subunit homologue

In the hyperthermophilic bacterium *Aquifex aeolicus,* for which the whole genome has been sequenced (Deckert *et al*., 1998), the single GOGAT β subunit homologue, which shared striking homology to the *gltX* gene product (39.1% identity and 50.3% similarity), was not

located adjacent to a gene encoding a α subunit homologue. Similarly, a GOGAT β subunit-like gene is not located adjacent to either of the two Fd-dependent enzymes in *Synechocystis* (accession no. dbj:D90900), yet in the closely related *P. boryanum*, the α and β subunits of the pyridine dependent enzyme are encoded adjacent to each other, and separate from an Fd-dependent GOGAT (Okuhara *et al.*, 1999). In the light of the evolution of the bacterial enzyme, we were interested to know if a second α subunit homologue existed elsewhere on the *C. acetobutylicum* P262 chromosome.

A synthetic degenerate 23-mer oligonucleotide probe pool was synthesized to a strictly conserved region (ProProHisHisAspIleTyrSer) found in all GOGAT α subunit domains (corresponding to amino acid positions 990-997 of *C. acetobutylicum* GltA, see Appendix C), and used to probe *C. acetobutylicum* chromosomal DNA (Fig. 3.11) using the TMA method described in Section 3.3.11. As expected, the probe, which was specific for clostridial DNA, hybridized to chromosomal bands that correspond to regions spanning the *gltA* gene; a 2.6 kb *Eco*RI fragment (lane 2) and a 2.36 kb *Eco*RI-*Pst*I fragment (lane 3), diagramatically presented in Fig.3.11B. These results reflect a moderate TMA wash stringency of 56 °C. However, even at less stringent conditions (48 °C, results not shown), no other hybridization signals were detected, suggesting that it is very unlikely that a second GOGAT α subunit homologue exists elsewhere on the chromosome. This is consistent with the literature to date that the only bacteria which posses two glutamate synthase domains are the unicellular cyanobacteria (Navarro *et al.*, 1995).

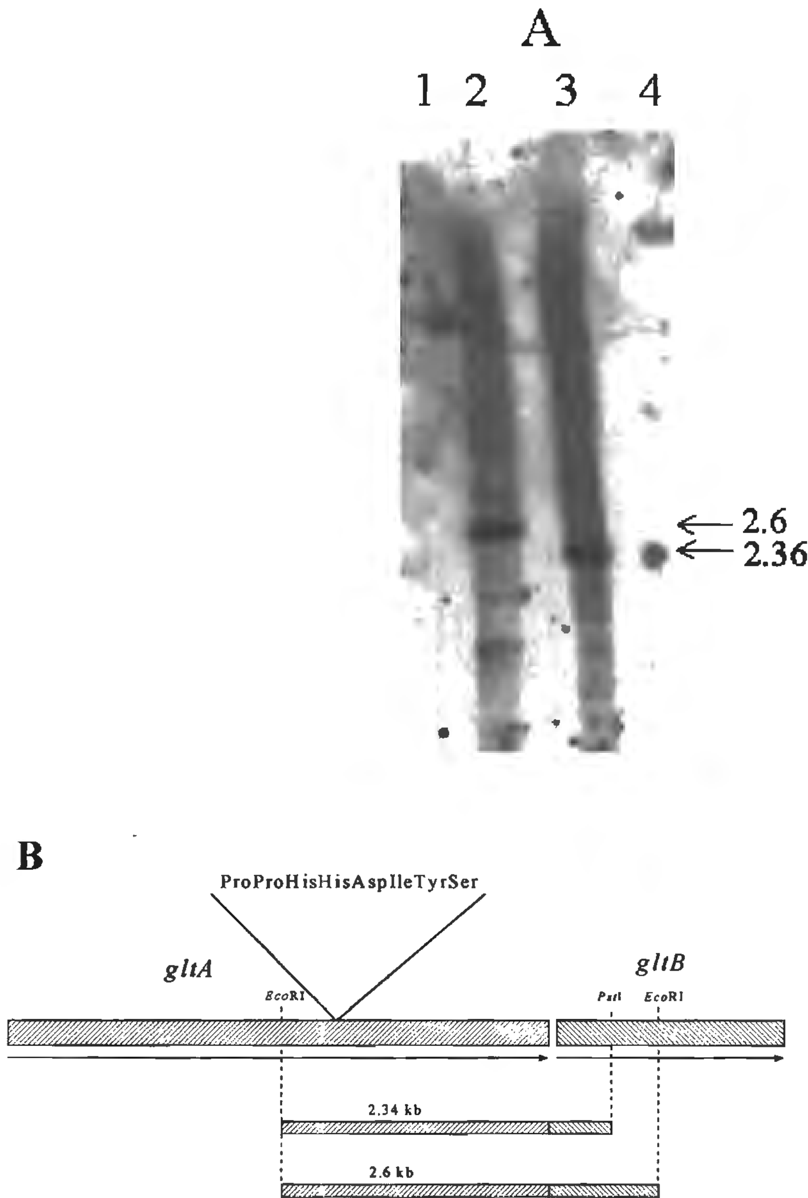


Fig. 3.11 Southern blot detection of GOGAT α subunit domains in the genome of *C. acetobutylicum* P262, using the tetramethylammonium chloride (TMA) method (A). Chromosomal DNA was digested with *Cla*I-*Eco*RI (lane 2), and *Eco*RI-*Pst*I (lane 3), and probed with a degenerate 23-mer oligonucleotide designed to a strictly conserved region found in all GOGAT α subunits (Section 3.3.11). Lane (1) contains *E. coli* chromosomal DNA digested with *Eco*RI. Lane (4) contains *Pst*I-digested lambda DNA. This blot represents a TMA stringency wash of 56 °C. B) Illustration of the origin of the probe, and the two hybridizing fragments.

3.4.12 *gltX* in context

Homology searches of both GltX and GltB revealed additional interesting results: the complete GOGAT β subunit domain is conserved, to varying degrees, within a diverse group of much larger proteins as well (Fig. 3.12). These include the formate dehydrogenase beta subunit (708 aa) from *Moorella thermoacetica,* the C-terminal two-thirds of the *E. coli aegA* gene product,

and the amino terminal regions of the highly conserved eukaryotic dihydropyrimidine dehydrogenase (DPD) enzyme, another complex iron-sulfur flavoprotein involved in pyrimidine base degradation. From the structural and functional properties of the DPD enzyme, it was concluded that the β subunit-like domain most likely serves to input electrons from NADPH into the dihydrorotate-like domain for uracil reduction (Vanoni and Curti, 1999). Characterization of the *aegA* gene (Cavicchioli *et al.*, 1996) revealed that it lies adjacent to, and in the opposite direction from *narQ,* which encodes a second nitrate/nitrite-responsive sensor-transmitter protein in *E. coli* (Chiang *et al.*, 1992). Although the function of the *aegA* gene is unclear, its expression and regulation suggest that it is involved in either fermentation or anaerobic respiration (Stewart, 1993; Gunsalus and Park, 1994), and the N-terminal domain contains four sets of cysteine clusters suggestive of iron-sulfur centers associated with oxidoreductase activities. Similar sequence analyses, and the finding that the putative β subunit-like proteins clustered with the β subunit-like domains found in larger proteins, rather than with GOGAT β subunits, led Vanoni and Curti (1999) to propose that the GOGAT β subunit is a member of a novel family of FAD and iron-sulfur containing oxidoreductases, which serve to transfer electrons between a reduced pyridine nucleotide and a second protein or protein domain. In the case of GOGAT, this would be the compatible α subunit. According to this hypothesis, GltX (and other β subunit-like proteins including those from *A. aeolicus* and *Pyrococcus* sp.) would represent members of this novel family of oxidoreductases, unrelated to GOGAT activity.

However, an alternative hypothesis that must be considered is that, since GltX is most closely related to β subunit-like genes found in the archaebacteria *Pyrococcus,* it may represent a functional homologue of the GltA protein (480 aa) from *Pyrococcus* sp KOD1. GltA has been reported to encode the smallest known fully functional GOGAT, capable of both glutamine-dependent and ammonia-dependent synthesis, in the absence of a α subunit homologue (Jongsareejit *et al.*, 1997). No α subunit homologue could be identified adjacent to *gltA*, nor anywhere else on the chromosome. This is consistent with there being no α subunit homologue in either *P. horikoshii* strain OT3, or *P. abysii,* for which the complete genomes have been sequenced (Kawarabayasi *et al.*, 1998; Heilig, 1999,unpublished). It is also interesting to note that in the syphilis spirochete *Treponema pallidum* (Fraser *et al.*, 1998), in which catabolic and biosynthetic activities are minimized, only a single GOGAT β subunit homologue can be identified which shows the highest homology (50% identity) to the KOD1 GltA protein. As

discussed in Chapter one, the studies on the *Pyrococcus* GltA protein led (Jongsareejit *et al.*, 1997) to propose that it represented an ancestral prototype of the GOGAT β subunit, however, as pointed out by Vanoni and Curti (1999), it is unclear as to how the purified GltA protein might function in substrate binding and catalysis. It will be interesting to see if the properties of GltA withstand these challenges, or alternatively, if GOGAT activity can be associated with any other independent β subunit-like proteins. However, if we accept that GltA does represent a functional GOGAT enzyme, it raises obvious questions about *gltX*. Could it represent such an ancestral archaebacterial-type GOGAT gene possessing GOGAT activity? There is evidence for lateral gene transfer between Archaea and bacteria from the genome sequence of the thermophilic *T. maritima* (Nelson *et al.*, 1999).

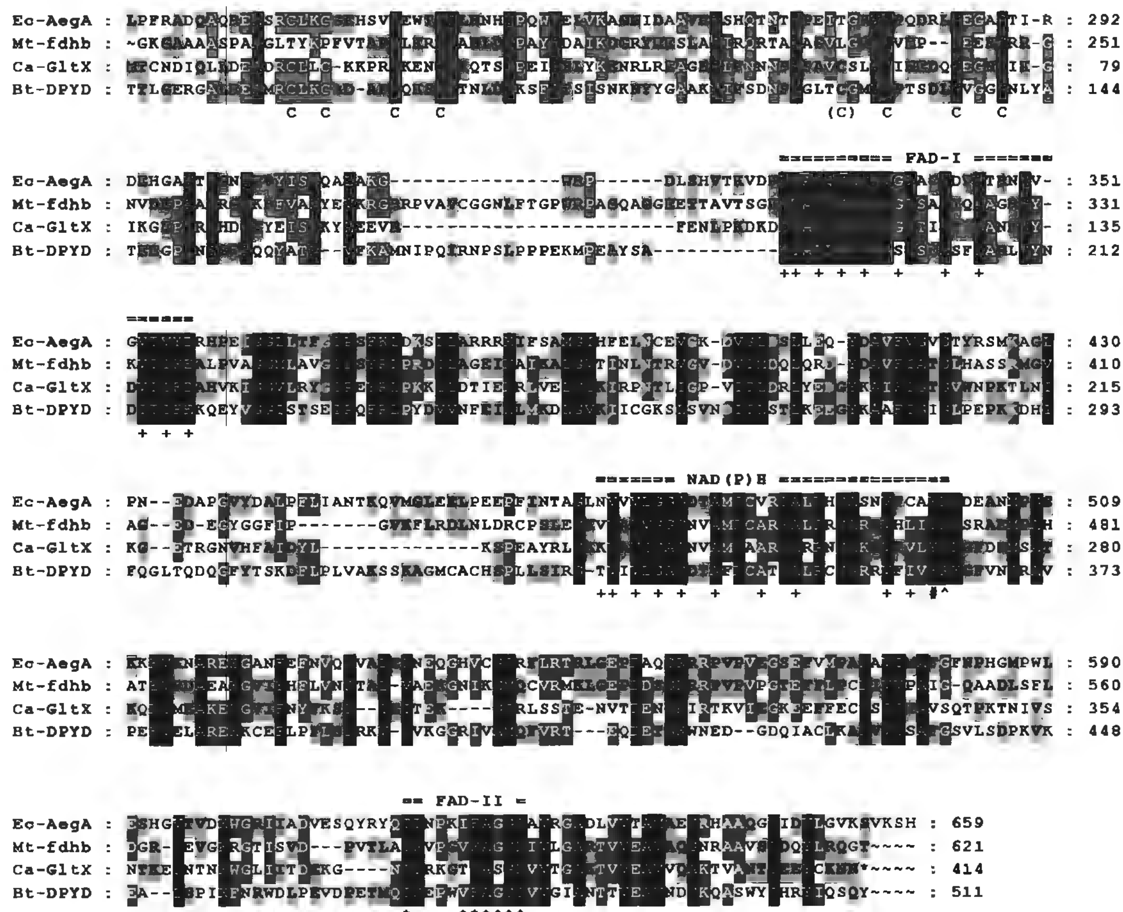


Fig. 3.12 Amino acid sequence alignment of *C. acetobutylicum* P262 GltX (Ca-GltX) with GOGAT β subunit-like domains within other proteins; *E. coli* AegA (Ec-AegA, accession no. sp:P37127), bovine dihydropyrimidine dehydrogenase (Bt-DPYD, accession no. sp:Q28007), and *Moorella thermoacetica* formate dehydrogenase beta subunit (Mt-fdhb, accession no. gb:U73807). Proposed GOGAT β subunit domains are annotated, and described in the legend to Fig. 3.5. Residues that are conserved in all four sequences are shaded against a black background. Residues conserved in three and two of the sequences aligned, are shaded against a dark grey and light grey background, respectively.

3.5 Conclusion

The *C. acetobutylicum* P262 *gltX* gene (1245 bp in length) encodes a product, GltX, in agreement with the predicted molecular weight of 46.1 kDa, which is responsible for the reductive activation of metronidazole. Although it is not apparent whether GltX is involved in the direct or indirect reduction of metronidazole, we conclude that since its deduced amino acid sequence showed significant homology to GOGAT β subunits, and contained the conserved domains proposed to be functional FAD and NADPH binding sites, GltX is probably involved in an electron transport mechanism similar to that proposed for GOGAT β subunits. Thus, this study serves to highlight additional interesting genes encoding proteins not previously considered to be involved in metronidazole reduction. An understanding of the mechanisms involved in activation of the drug is important for the development of improved analogues for use in medicine

The *gltX* gene, which appears conserved in the clostridia, is independently transcribed. It does not form part of a typical GOGAT operon as a second α subunit homologue does not lie adjacent to it (nor anywhere else on the chromosome). Since the deduced amino acid sequence of *gltX* is most closely related to isolated β subunit-like genes found in the archaebacteria *Pyrococcus,* it may represent an ancestral GOGAT enzyme similar to the GltA protein characterized from *Pyrococcus* sp KOD1. This possibility is investigated in the following chapter. Alternatively, GltX may represent a member of a much more general family of FAD-dependent NAD(P)H oxidoreductase unrelated to GOGAT activity.

CHAPTER 4

Expression studies on the *C. acetobutylicum* P262 *gltA, gltB* and *gltX* genes in an *E. coli* GOGAT mutant

4.1. Summary **111**
4.2. Introduction **112**
4.3. Materials and methods **112**
4.3.1. Bacterial strains and media -------- 112
4.3.2. General DNA manipulations -------- 113
4.3.3. Plasmid constructs -------- 113
4.3.4. GOGAT complementation assays -------- 116
4.3.5. RNA extractions and Northern blots -------- 116
4.3.6. Protein analysis of cell free extracts (CFE) -------- 117
4.4. Results and discussion **117**
4.4.1 Expression of the *C. acetobutylicum gltB* gene in *E. coli* MX3004 -------- 117
4.4.2. Complementation of *E. coli* MX3004 with both the *C. acetobutylicum gltA* and *gltB* genes together -------- 120
4.4.3. Expression of *C. acetobutylicum gltA* and *gltB* genes in *E. coli* MX3004 -------- 120
4.4.4. Detection of the *C. acetobutylicum glt* gene products in *E. coli* MX3004 -------- 121
4.4.5. *C. acetobutylicum gltX* gene expression -------- 123
4.4.6. Considerations for the production of functional enzyme -------- 123
4.5. Conclusion **125**

CHAPTER 4

Expression studies on the *C. acetobutylicum* P262 *gltA*, *gltB* and *gltX* genes in an *E. coli* GOGAT mutant

4.1. Summary

The *C. acetobutylicum* P262 *gltA* gene, identified in Chapter 2 as the GOGAT α subunit, was cloned into the pACYC184 vector yielding construct pHS9. Similarly, the *gltB* gene, identified as the GOGAT β subunit was cloned downstream of the lambda promoter in pEcoR251 yielding construct pHS7, and downstream of the *lac*Z promoter in the Bluescript vector yielding construct pHS6. Expression of *gltB,* from either pHS6 or pHS7, in an *E. coli* glutamate auxotroph (strain MX3004) that is a mutant in the GOGAT β subunit, failed to complement the *E. coli* α subunit, and restore growth of MX3004 on minimal medium containing ammonia as the sole source of nitrogen. However, the co-transformation of MX3004 with both pHS7 and pHS9 enabled the *E. coli* strain to grow, under anaerobic conditions, on the minimal medium, indicating that GOGAT activity had been restored. RNA and protein analysis confirmed the expression of the *C. acetobutylicum gltA* and *gltB* genes in *E. coli.* Expression of the *C. acetobutylicum gltX* gene, identified in Chapter 3 as a GOGAT β subunit-like gene, either alone (constructs pMET13C1 or pSKMETS) or in combination with pHS9, failed to restore the ability of MX3004 to grow using ammonia as the sole source of nitrogen. Nor could it complement a *B. subtilis* GOGAT β subunit mutant (strain 1A490), and restore GOGAT activity.

4.2. Introduction

The isolation and molecular characterization of the *gltA* and *gltB* genes, identified by sequence homology as the structural genes encoding the α and β subunits of GOGAT respectively, was described in Chapter 2. Similarly, the *gltX* gene, identified as another GOGAT β subunit-like gene, was described in Chapter 3. The aim of this Chapter was twofold: to functionally confirm the identity of the *gltA* and *gltB* genes, and to try and elucidate the role of the *gltX* gene, by complementation studies.

Since *C. acetobutylicum* P262 lacks a transformation system, the physiological characterization of these genes was limited to a heterologous host. There are several reports of the heterologous expression of GOGAT genes in *E. coli* glutamate auxotrophs which carry mutations in the *gltB* (α subunit) or *gltD* (β subunit) genes. (In *E. coli,* the *gltB* and *gltD* genes form an operon with a downstream gene, *gltF*). For example, *E. coli* GOGAT mutants have been complemented by the yeast glutamate synthase gene (Gonzalez *et al.*, 1992), the *gltBD* region of *A. sesbaniae* chromosomal DNA (Hilgert *et al.*, 1987; Donald *et al.*, 1988) and the *glt* locus from *Rhizobium meliloti* 1021 (Lewis *et al.*, 1990). Furthermore, the GOGAT β subunit from *Thiobacillus ferrooxidans* complemented the *E. coli* α subunit, and restored GOGAT activity in an *E. coli gltD* mutant strain (Deane and Rawlings, 1996). This chapter focuses on the expression of the *C. acetobutylicum* P262 *gltB, gltD* and *gltX* genes in an *E. coli* glutamate auxotroph (strain MX3004) that is a mutant in both the GOGAT β subunit gene (*gltD*), and the gene (*gdh*) encoding the GDH enzyme.

4.3. Materials and methods

4.3.1. Bacterial strains and media

E. coli strain MX3004 (*thi*-1, *gdh* 1, *pro*+Δ(*lac*U169)*hut*C, *gltD*227::MudIIPR13)(Castano *et al.*, 1992), is a mutant in both the GDH and GOGAT pathways and thus is a glutamate auxotroph (Berberich, 1972). It is unable to synthesize the GOGAT β subunit, and hence was used to test for plasmids that could complement for GOGAT activity. MX3004 transformants were screened for their ability to grow on NN minimal medium (Covarrubias *et al.*, 1980) solidified with agar (1.5% w/v) and containing ammonia (0.5 mM $(NH_4)_2SO_4$) as the sole

source of nitrogen, under both aerobic and anaerobic conditions. Control plates contained monosodium glutamate (MSG, 0.2% w/v, Merck). For anaerobic screening, NN minimal media was supplemented with 0.5% w/v glucose, and was carried out in an anaerobic glove cabinet (Forma Scientifica inc.) with a gas phase of 5% H_2, 10% CO_2 and 85% N_2. All media and solutions were equilibrated under anaerobic conditions prior to use. MX3004 cultures were grown with Cm (50 μg/ml) selection, and supplemented with ampicillin (100 μg/ml), and/or tetracycline (50 μg/ml) for the appropriate plasmid selection.

Similarly, *B. subtilis* 1A490 (*gltB*-, *leu*38, *metB*10) is a glutamate auxotroph due to a mutation in the β subunit (*B. subtilis* does not posses assimilatory GDH activity), and was also used in complementation studies. Transformants were screened on M9 minimal medium (Harwood and Cutting, 1990) containing chloramphenicol selection (15 μg/ml), 0.5 mM $(NH_4)_2SO_4$ as the nitrogen source, and supplemented with 5 mg/ml of the amino acids L-leucine, L-methionine, and in the control plates, MSG.

E. coli JM109 (Yanisch-Perron *et al.*, 1985) was used as the host for cloning purposes, and cultures were grown at 37 °C in 2xYT broth or on 2xYT agar (Messing, 1983), containing the appropriate antibiotic selection.

4.3.2. General DNA manipulations

Plasmid DNA was isolated from *E. coli* either by the alkaline hydrolysis method of Ish-Horowicz and Burke (1981), or using the Nucleobond® AX KIT (Macherey-Nagel). Routine DNA cloning and transformation procedures were as described by Sambrook *et al.* (1989). Restriction endonucleases, T_4 DNA ligase and Klenow were used according to the manufacturer's recommendations. DNA fragments were gel-fractionated (0.8% w/v agarose) using Tris-Acetate (EDTA) buffer, and selected fragments for subcloning were purified using the Gene Clean kit (Amersham). *B. subtilis* was transformed using the DM3 regeneration protocol (Harwood and Cutting, 1990).

4.3.3. Plasmid constructs

Plasmid constructs used in this study to test for GOGAT complementation in *E. coli* MX3004 are illustrated in Fig. 4.1. Plasmids pMEXP1 and pKS2, which have been described by Deane and Rawlings, (1996), both harbour the same 2.4 kb fragment of *T. ferrooxidans* ATCC 33020 DNA. This fragment includes the terminal 309 nucleotides of the GOGAT α subunit gene,

gltB, immediately upstream of the 1.694 kb β subunit gene, *gltD*. They are positioned downstream of the lambda promoter in the pEcoR251 vector construct pMEXP1, and downstream of the *lacZ* promoter in the Bluescript vector construct pKS2.

Similar plasmids were constructed containing the *C. acetobutylicum* β subunit gene, *gltB*. The *C. acetobutylicum glt* locus is presented schematically in Fig. 4.1(B). The 3.53 kb *Pvu*II restriction fragment was cloned into another blunt-ended restriction site, *Sma*I, within the multiple cloning cassette of the high copy number Bluescript vector pSK (Stratagene), to yield plasmid pHS6. This fragment included the terminal 326 nucleotides of the *gltA* gene upstream of the 1.47 kb *gltB* gene, and the orientation of cloning determined that they lay downstream of the *lacZ* promoter. Plasmid pHS5 (see Fig. 2.1) is identical to pHS6, except that the insert is present in the opposite orientation. Plasmid pHS7 was derived from pHS5 by subcloning the insert from the adjacent *Eco*RI and *Bam*HI vector sites, into the unique *Hind*III and *Bam*HI sites of the low copy number vector pEcoR251 (Zabeau and Stanley, 1982). The *Eco*RI and *Hind*III sites were made compatible by blunt-ending them. This directional cloning ensured that the *gltB* gene was positioned downstream of the lambda promoter.

Plasmid pHS9, harbouring the *C. acetobutylicum gltA* gene, was constructed in two stages. First, the purified 3.86 kb *Xba*I fragment was subcloned into the unique *Xba*I site of pHS2 (Fig. 2.1) to generate plasmid pHS8. Restriction digestion and sequence analysis confirmed the correct orientation of cloning, and that the *Xba*I junction site had been correctly reconstituted. The second stage involved subcloning the entire reconstructed *gltA* gene, i.e. the *Eco*RV-*Not*I fragment of pHS8, into the unique *Eco*RI site of pACYC184 (Rose, 1988), such that *gltA* was read in the same direction as the disrupted Cm gene. This was achieved by blunting the *Not*I and *Eco*RI restricted sites. The *Eco*RV restriction site, which lies 65 bp upstream of the predicted *gltA* translation initiation codon, disrupts the inverted repeat sequence (see Fig. 2.4) identified in Chapter 2 as a potential transcriptional terminator. Plasmid pACYC184, is able to co-exist with vectors that carry the ColE1 origin of replication such as pEcoR251.

Plasmids pMET13C1 and pSKMETS have been described in Chapter 3 and are illustrated in Fig. 3.1. Plasmid pCMX (obtained from Dr. V. Mittendorf, University of Cape Town) served as the pEcoR251 vector control. It harboured a ~0.5 kb insert of non coding *Clostridium longisporum* genomic DNA cloned into the unique *Bgl*II site of the *Eco*RI endonuclease gene.

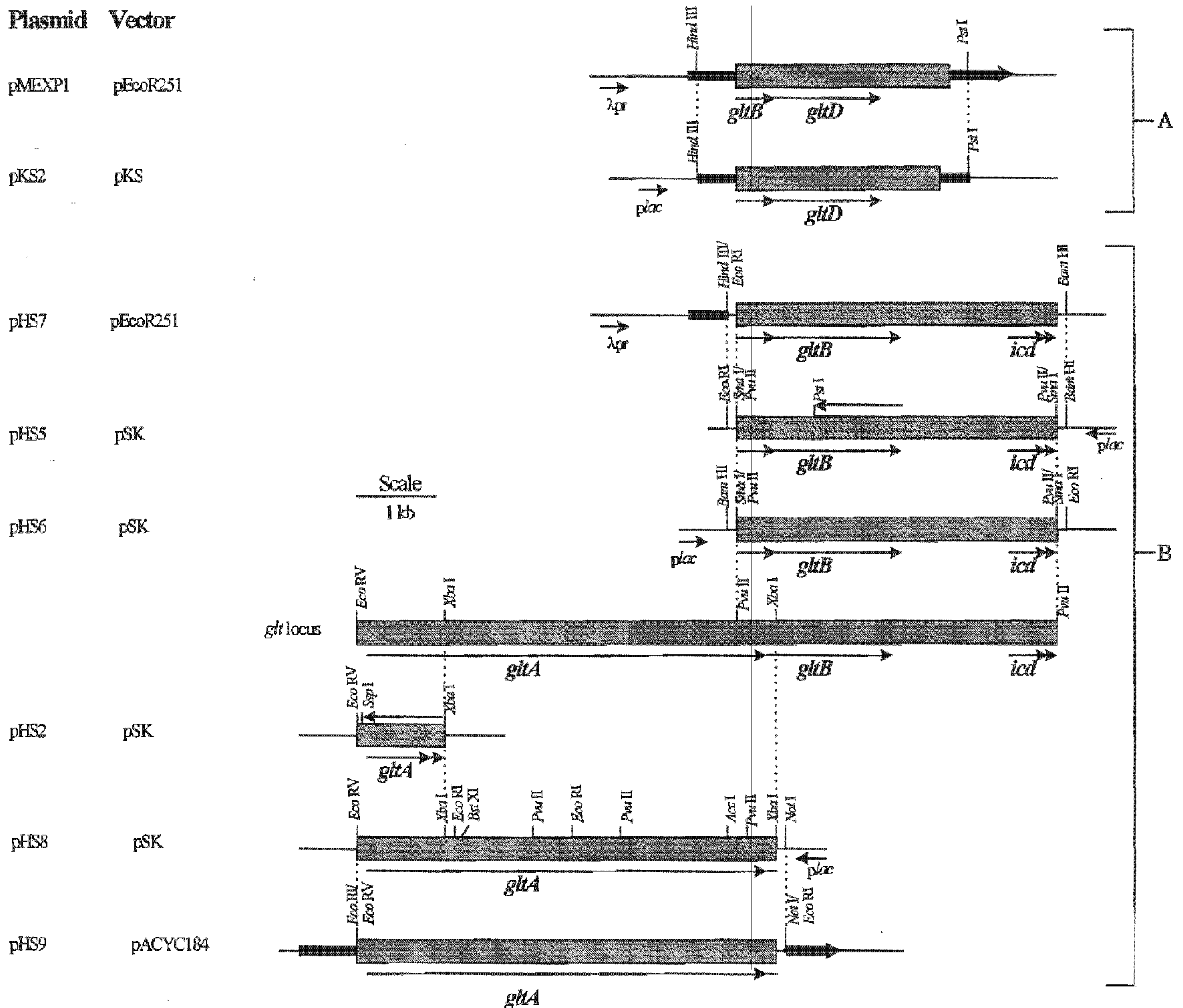


Fig.4.1. Illustration of the clones constructed to test for the complementation of GOGAT activity in *E. coli* MX3004. Restriction sites used for the respective cloning events are joined by dotted lines. Regions representing *T. ferrooxidans* (A) and *C. acetobutylicum* (B) chromosomal DNA are indicated by thick hatched bars. Vector DNA is represented by thin horizontal lines except where the *Eco*R1 endonuclease gene of vector pEcoR251, and the chloramphenicol resistance gene of vector pACYC184, are included (thick solid bars). The *lac*Z promoter is indicated by p*lac*, and the lambda promoter by λpr. Arrows below the hatched bars indicate the extent and direction in which the genes are encoded. Arrows above the hatched bars depict the length and direction in which RNA probes were synthesized for Northern blots. Double arrow heads indicate an incomplete coding sequence.

To investigate *B. subtilis* GOGAT complementation with the *gltX* gene, the 3.0 kb *Pst*I-*Eco*RI insert fragment from pMET13C1, was cloned into the corresponding sites in the *E. coli/B. subtilis* shuttle vector pEB1 (Anaerobe laboratory, University of Cape Town). In addition, the

3.4 kb *Eco*RI-*Hind*III fragment from pMET13C1 was cloned into the corresponding sites in the *Bacillus* chromosomal integration vector pDG364 (Harwood and Cutting, 1990). This vector is designed for the targeted integration of a single copy of the region/gene of interest, into the *amyE* locus. Transformants are identified as being chloramphenicol resistant, and having lost the ability to produce hallows on 2xYT media containing 1% starch. Vector details are supplied in Appendix B.

4.3.4. GOGAT complementation assays

E. coli MX3004, transformed with the relevant plasmid constructs, were cultured in 50 ml of 2xYT broth, containing the appropriate antibiotic selection, at 37 °C, to OD_{600} of ~1.5. Samples (5 ml) were harvested in duplicate from each culture for separate processing. The two samples were processed in the following way, one under aerobic and the other under anaerobic conditions: each pellet was washed twice with an equal volume of saline (0.15 M NaCl) and resuspended in 2ml of the solution. Dilutions of 10^{-3}, 10^{-4} and 10^{-5} were made in saline and 100µl of each were plated on NN minimal medium with and without glutamate, in duplicate. One set of plates were incubated at 37 °C, and the other set at 30 °C in the event of a less stable hybrid GOGAT protein being reconstituted. Selected *B. subtilis* 1A490 transformants were processed in a similar way, except that transformants were screened on M9 minimal media containing chloramphenicol.

4.3.5. RNA extractions and Northern blots

To prevent RNA degradation, strict RNAse-free conditions were employed during all RNA handling procedures: glassware was heat treated, only Milli-Q water (Millipore water purification system) and RNAse-free chemicals were used, and plastic-ware and suitable solutions were subjected to two cycles of autoclaving.

Total mRNA was isolated using the hot phenol extraction protocol described by Aiba *et al.* (1981). MX3004 cultures harbouring the relevant plasmid constructs were extracted after overnight growth at 37 °C in 2xYT medium containing the appropriate antibiotic selection. Purified mRNA was resuspended in water and stored at -70 °C in the presence of RNAse inhibitor (Sigma).

Northern blots were performed using the non-radioactive digoxigenin (DIG) Labeling and Detection Kit according to the manufacturer's protocol (Boehringer Mannheim). RNA was

separated by electrophoresis in 1.5% denaturing formaldehyde agarose gels, and Boehringer Mannheim RNA molecular weight markers were included. RNA was transferred to a positively charged nylon membrane (Boehringer Mannheim) by capillary blotting for 10 hours at room temperature, and was subsequently fixed onto the membrane by UV exposure (340 nm for 5 min).

DIG-labeled homologous RNA probes to *gltA* and *gltB* were synthesized from plasmid pHS2, and a shortened derivative of plasmid pHS5, respectively (illustrated in Fig. 4.1). The T_3 polymerase-specific promoter was used to direct the synthesis of both probes, and their lengths (798 bp for *gltA* and ~1.1 kb for *gltB*) were defined by linearization of the template DNA at the respective *Ssp*I and *Pst*I restriction sites. The template fragments were gel-purified prior to labeling. Probes were used at a concentration of 100 ng/ml, and hybridized at 68 °C for 12 hours, followed by stringent washing conditions. Probe bound to the membrane was detected using the chemiluminescent substrate AMPPD (Boehringer Mannheim).

4.3.6. Protein analysis of cell free extracts (CFE)

MX3004 transformants, harbouring the relevant plasmids, were grown overnight at 37 °C in 100 ml of LB, containing the appropriate antibiotic selection, and where indicated, 0.5 mM of IPTG. The cultures were harvested, washed with an equal volume of Tris-HCl buffer (0.1 M pH 8), and resuspended in 5ml of buffer, prior to sonication (Soniprep 150). The BIO-RAD protein assay was used for the quantitation of proteins in the CFE, with BSA fraction V (Boehringer Mannheim) serving as the standard. The proteins were resolved by SDS-PAGE on a 10% gel according to Laemmli (1970), and visualized by staining with PAGE Blue 83 (BDH). The molecular weight standards were obtained from Pharmacia.

4.4. Results and discussion

4.4.1 Expression of the *C. acetobutylicum gltB* gene in *E. coli* MX3004

E. coli MX3004 is a mutant in both the ammonia assimilatory pathways viz. the GDH and the GS-GOGAT pathway. The latter is due to insertional inactivation of the GOGAT β subunit. Hence it is unable to use ammonia as a nitrogen source, but grows like wild type *E. coli* when glutamate is supplied (Table 4.1). Since the *T. ferrooxidans* GOGAT β subunit gene, *gltD*, was shown to complement the MX3004 α subunit (Deane and Rawlings, 1996), we sought to

functionally confirm the identified *C. acetobutylicum glt* locus by similar complementation studies. Two plasmids carrying the *T. ferrooxidans gltD* gene, pKS2 and pMEXP1 (Deane and Rawlings, 1996) (Fig. 4.1A), served as our positive controls for heterologous complementation. Both were able to restore growth of MX3004 on NN minimal medium containing ammonia (Table 4.1), although as reported, complementation was less efficient with pKS2 than with pMEXP1. This difference is thought to be due to *gltD* being expressed from the stronger λ promoter in the case of pMEXP1, and the weaker *lacZ* promoter in pKS2, suggesting that high levels of *gltD* expression are required for efficient complementation.

Two plasmids, carrying the same 2.5 kb fragment of *C. acetobutylicum* chromosomal DNA containing the *gltB* gene, were constructed very similar to pKS2 and pMEXP1, respectively, as described in Section 4.3.3: in pHS6 the *gltB* gene was cloned downstream of the *lacZ* promoter of the Bluescript vector pSK, and in pHS7 the *gltB* gene was cloned downstream of the lambda promoter of pEcoR251. However, unlike the results obtained with pKS2 and pMEXP1, transformation of MX3004 with either of these plasmids, failed to restore growth of the *E. coli* mutant on minimal media containing ammonia. This result remained unchanged by incubation at lower temperatures (~ 18 °C and 30 °C), or under anaerobic conditions (Table 4.1).

An observation that emerged from these studies however, was that in the presence of glutamate, the size of the *E. coli* colonies harbouring the β subunit genes from either *T. ferrooxidans* (constructs pKS2 or pMEXP1) or *C. acetobutylicum* (constructs pHS6 or pHS7), were significantly smaller than the corresponding controls (MX3004 transformed with the relevant vector control constructs). Since the regions downstream of either the *T. ferrooxidans gltD* gene (~ 675 bp) or the *C. acetobutylicum gltB* gene (1.717 kb) are dissimilar, it is unlikely that they are exerting this growth phenotype. Rather, it suggests that the GOGAT β subunit, in the presence of glutamate, has an effect on *E. coli* metabolism, and supports the assumption that the *C. acetobutylicum* β subunit is being expressed in *E. coli.* This observation may be related to glutamate-dependent repression of the *glt* operon reported for *E. coli* (Castano *et al.*, 1992).

TABLE 4.1. The effect of various plasmid constructs on the ability of *E. coli* MX3004 to grow on NN minimal media containing ammonia as the sole source of nitrogen, in the presence and absence of glutamate, and under aerobic or anaerobic conditions.

MX3004 transformed with:	**AEROBIC**		**ANAEROBIC**	
	with glutamate	**without glutamate**	**with glutamate**	**without glutamate**
(JM109 untransformed)	+	+	+	+
MX3004 untransformed	+	—	+	—
pSK	+	—	+	—
pKS2	+	+	+	+
pHS6	+	—	+	—
pCMX	+	—	+	—
pMEXP1	+	+	+	+
pHS7	+	—	+	—
pACYC184	+	—	+	—
pHS7 and pHS9	+	—	+	+
pMET13C1	+	—	+	—
pSKMETS	+	—	+	—
pHS9 and pMET13C1	+	—	+	—

These results are representative of two independent experiments. Each construct was tested in duplicate at 30 °C and 37 °C under both aerobic and anaerobic conditions. Growth is indicated by the positive symbol +, which represents 200-300 colonies per plate, and corresponds to 100 µl of a 10^{-4} dilution of the original culture (Section 4.3.4). No growth is indicated by the symbol —. These trends were the same at both 30 °C and 37 °C incubation temperatures.

4.4.2. Complementation of *E. coli* MX3004 with both the *C. acetobutylicum gltA* and *gltB* genes together

If complementation of *E. coli* MX3004 with the *C. acetobutylicum gltB* gene failed to reconstitute GOGAT activity due to incompatibility of the subunits from these two heterologous bacteria, then expression of both homologous subunits of *C. acetobutylicum* would be expected to reconstitute GOGAT activity in MX3004. To test this assumption the *C. acetobutylicum gltA* gene was cloned into vector pACYC184 to yield plasmid pHS9 (Section 4.3.3). Indeed, co-transformation of MX3004 with both plasmids pHS7 and pHS9 enabled the

mutant to grow on the defined NN minimal medium containing ammonia as the sole source of nitrogen, but notably only under anaerobic conditions (Table 4.1). Growth was, however, delayed by approximately two days relative to the controls, and the colonies remained small, regardless of the incubation period or temperature (30 °C or 37 °C). While these results imply that functional recombinant GOGAT enzyme is produced in MX3004, under anaerobic conditions, from the separate expression of the *gltA* and *gltB* genes, they also indicate that these conditions are not optimal for the efficient production of mature enzyme. In *C. acetobutylicum, gltA* and *gltB* are co-transcribed (Chapter 5) and probably translationally coupled, however, attempts to clone the complete *glt* locus into a high expression vector was unsuccessful.

4.4.3. Expression of *C. acetobutylicum gltA* and *gltB* genes in *E. coli* MX3004

Northern blot analysis using homologous RNA probes (Section 4.3.5) confirmed that both the *gltA* and *gltB* genes, are transcribed in the *E. coli* MX3004 host (Fig. 4. 2). The *gltA* specific probe detected a transcript size of approximately 4.6 kb in lane D, which contained RNA isolated from the MX3004(pHS9) transformant. Since the *gltA* gene was calculated to be 4.554 kb long (Chapter 2), this result suggests that transcription of the *gltA* gene may be initiated from a promoter-like sequence in the AT-rich region preceding the start of the gene, rather than from the promoter for the chloramphenicol gene. However, the accurate sizing of the RNA transcript is limited by the quality of these RNA species in *E. coli*. We deduce that the non-specific signals detected in the control lanes (Panel I- lanes A, B and C) are due to heterologous detection of the equivalent conserved region of the *E. coli* α subunit that has got tangled in the 16S and 23S rRNA bands. The *gltA* probe was made to the N-terminal 12-810 nucleotides.

In contrast, the *gltB* probe, which did not overlap the region corresponding to the N-terminal 225 bases of the truncated MX3004 *gltD* gene, was very specific for the *C. acetobutylicum gltB* gene (Fig.4.2, Panel II). In the lanes representing RNA extracted from *E. coli* MX3004 transformed with either pHS7 (lane E), or pHS7 and pHS9 (lane F), a transcript of approximately 2.4 kb was detected. This transcript size is consistent with transcription initiating at the pEcoR251 λ promoter of plasmid pHS7, and reading through the terminal 326 bases of *gltA*, to a transcriptional terminator sequence identified downstream of the *gltB* gene (Section 2.4.3). The faint signal detected in the pEcoR251 vector control (lane B) may be due to contamination of the probe with vector sequence common to both pSK and pEcoR251.

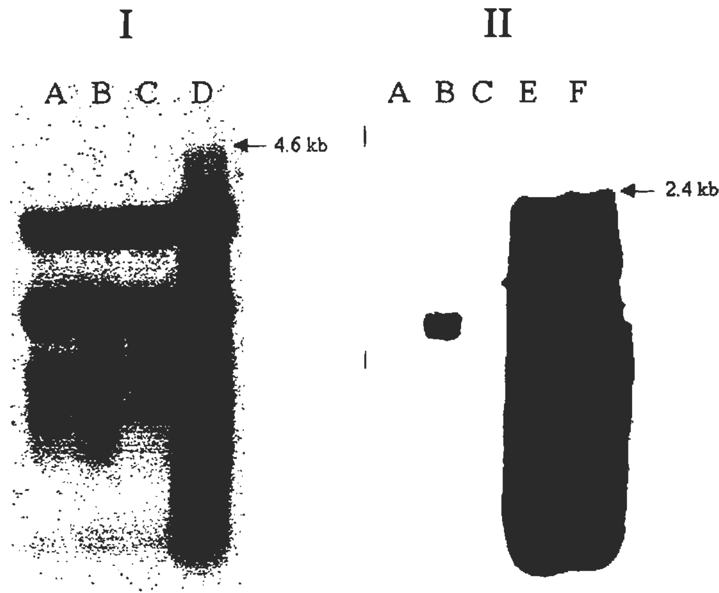


Fig. 4.2. Northern blot detection of *C. acetobutylicum gltA* (Panel I) and *gltB* (Panel II) transcription in *E. coli* MX3004 using *gltA* and *gltB* specific RNA probes respectively. Each lane contains 30 μg of the respective RNA samples. Lane A represents RNA isolated from MX3004. Lanes B, C, D, E, and F represent RNA isolated from MX3004 transformed with pCMX, pACYC184, pHS9, pHS7, and pHS9 + pHS7, respectively.

4.4.4. Detection of the *C. acetobutylicum glt* gene products in *E. coli* MX3004

Cell free extracts of MX3004 transformants were analyzed by SDS-PAGE to determine whether polypeptides corresponding to the *C. acetobutylicum* GOGAT α and β subunits could be detected (Fig. 4.3). Sequence data revealed that both *gltA* and *gltB* are preceded by the identical RBS, present 7 and 8 nucleotides upstream from the initiation codons respectively. This is similar to the value (7 +/- 2 nucleotides) that has been deduced for the optimal spacing in *E. coli* in which expression of clostridial genes has been documented (Cary *et al.*, 1988; Youngleson *et al.*, 1988; Young *et al.*, 1989). Indeed, a polypeptide of approximately 50 kDa, which is consistent with the calculated size of the GOGAT β subunit (53 kDa), was clearly identified in the protein products of MX3004 transformed with either pHS6 (lanes B), or pHS7 (lane E). Furthermore, protein bands of a similar size were detected in the positive controls (lanes C and F) which represented CFE from MX3004(pKS2) and MX3004(pMEXP1), respectively. No corresponding bands were detected in the negative controls (lanes A and G). These results show that the *C. acetobutylicum gltB* gene is efficiently translated in *E. coli* MX3004, and thus support the notion that the *C. acetobutylicum* β subunit cannot functionally complement the *E. coli* GOGAT α subunit.

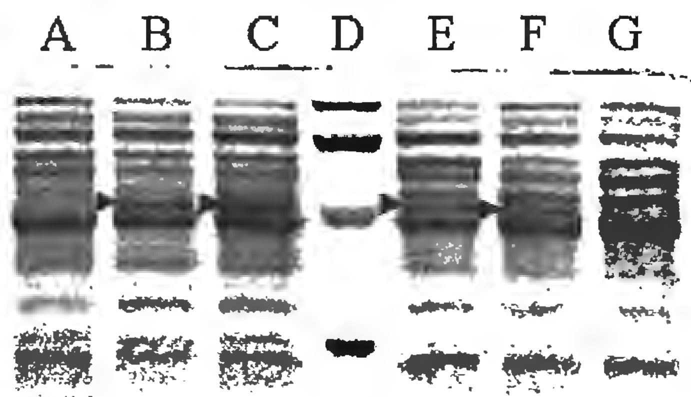


Fig. 4.3. Analysis of the protein products from MX3004 transformants by SDS-PAGE electrophoresis. Each lane represents 20 μg of protein from the CFE's of overnight cultures of MX3004 transformed with; pSK (lane A), pHS6 (lane B), pKS2 (lane C), pHS7 (lane E), pMEXP1 (lane F) and pCMX (lane G). IPTG (0.5 mM) was added to the growth medium of cultures harboring the Bluescript plasmids (pSK, pHS6 and pKS2) to induce expression from the p*lac* promoter. Lane D contains molecular weight standards (Pharmacia). Arrow heads indicate the position of a protein band of approximately 50 kDa.

No specific polypeptide corresponding to the size of the α subunit of GOGAT (168.5 kDa) could be detected in the protein products of MX3004 transformed with either of the constructs carrying the *gltA* gene; pHS8 or pHS9 (results not shown). This is not unexpected since the level of *gltA* expression from the low copy number pACYC184 clone, pHS9, or from the Bluescript clone pHS8, would not be expected to be as high as expression of the *gltB* gene from either the strong lambda promoter in pHS7, or the induced *lacZ* promoter in pHS6. In addition, undetectable levels of GltA could be attributed to protein instability, since the α subunit from *A. brasilense* was found to be less stable than the β subunit protein in *E. coli* (Curti *et al.*, 1996). Also, since *C. acetobutylicum* DNA has a low G + C ratio (~ 28-30%), the preferential use of rare codons in which A and U predominate, could constitute a translational barrier. Specifically, it has been suggested that the expression of large genes is more seriously impaired by unfavorable codon usage than that of average-sized (~30 to 40 kDa) clostridial genes (Eisel *et al.*, 1986; Garnier and Cole, 1986).

4.4.5. *C. acetobutylicum gltX* gene expression

In order to investigate the possibility that the *gltX* gene encodes an archaebacterial type GOGAT enzyme, or that it represents a second GOGAT β subunit, we performed similar complementation analysis involving *gltX*. In Chapter 3 we established that the *gltX* gene is

expressed in *E. coli* (strain F19) from either the pEcoR251 construct pMET13C1, or the Bluescript construct pSKMETS.

The *gltX* gene was introduced into *E. coli* MX3004 on plasmids pMET13C1 and pSKMETS, and *gltX* specific mRNA was detected by Northern dot blots (results not shown). However, neither MX3004(pMET13C1) nor MX3004(pSKMETS) transformants were able to grow on NN minimal medium containing ammonia as the sole source of nitrogen (Table 4.1). Furthermore, MX3004 co-transformed with both *gltX* (construct pMET13C1) and *gltA* (construct pHS9) was also unable to complement for GOGAT activity (Table 4.1).

Since *B. subtilis* is more closely related to *C. acetobutylicum*, we also investigated the expression of *gltX* in the *B. subtilis* β subunit mutant 1A490 (*gltB*-, *leu*38-, *metB*10-). As described in Section 4.3.3, the *gltX* gene was introduced into strain 1A490 on the shuttle vector pEB1, and by integrating it into the chromosome using the pDG364 integration vector (Fig. 4.4). Neither constructs restored the ability of *B. subtilis* 1A490 to grow on M9 minimal medium containing ammonia as the nitrogen source.

Although the expression of *gltX* was not demonstrated in *B.subtilis*, these results tentatively suggest that it is unlikely that *gltX* codes for an enzyme with GOGAT activity such as the *Pyrococcus* KOD1 GltA, or that it encodes a second functional GOGAT β subunit, since it was unable to complement either the *E. coli*, *B. subtilis* or *C. acetobutylicum* P262 α subunits.

4.4.6. Considerations for the production of functional enzyme

As discussed in Chapter 1, the β subunit not only serves to input electrons into the α subunit, but is also important for determining some of the properties of the α subunit and for the formation of the two 4Fe-4S centers, Center 2 and 3. Vanoni and Curti (1999) have suggested that the formation of these Fe-S centers occurs at the interface between the two subunits requiring their correct interactions and folding, and may be responsible for triggering the conformational changes required for the production of active holoenzyme. It is interesting to note that the cysteine cluster domains, proposed to be involved in the formation of these Fe-S centers (Section 2.4.5.4), are very similar in the β subunit polypeptides from *E. coli* and *T. ferrooxidans*, which both form NADPH-dependent GOGATs, however, differ distinctly in the *C. acetobutylicum* β subunit polypeptide, which specifies a NADH-dependent enzyme. If the

lack of heterologous complementation between the *E. coli* α subunit and the *C. acetobutylicum* β subunit is due to incompatible cysteine cluster arrangements, it further suggests that the *gltX* gene, which shares a very similar arrangement of cysteine clusters with the *E. coli* β subunit, does not represent an alternative functional GOGAT β subunit. It would be interesting to see if functional complementation could be achieved between the *B. subtilis* α subunit and the *C. acetobutylicum* β subunit, since the cysteine cluster domains are almost identical in the β subunit polypeptides from these two organisms.

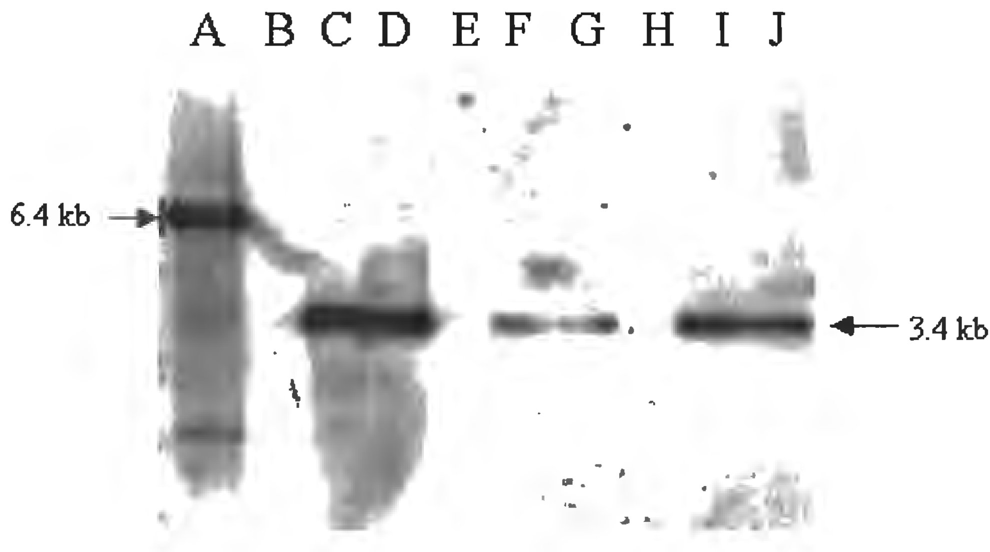


Fig. 4.4. Southern blot showing chromosomal insertion of the *gltX* gene in *B. subtilis* 1A490. Chromosomal DNA (40 μg) was prepared from the parental strain 1A490 (lane H) and from four chloramphenicol resistant transformants which had lost the ability to produce halos on media containing starch (lanes F, G, I and J), digested with *Eco*RI and *Hind*III, and probed with the DIG labeled *gltX* specific fragment described in the legend to Fig. 3.6. The positive controls included 100 ng each of: pSKMETS digested with *Hind*III (lane A), and two pDG364 constructs (lane C and D) containing the 3.4 kb *Eco*RI and *Hind*III insert derived from pMET13C1. Lane B contains *Eco*RI-*Hind*III digested pDG364 vector only. Size markers were run in lane E. The common 3.43 kb *Eco*RI-*Hind*III band containing the *gltX* gene region is indicated.

The low efficiency of GOGAT complementation achieved in *E. coli* MX3004, when both homologous subunits were introduced, is probably due to a number of factors. As already mentioned, levels of expression, protein instability and unfavorable codon usage in *E. coli* may contribute. In addition, the rate at which mature enzyme is produced from the separate expression of the *gltA* and *gltB* genes, may be limited by the rate of formation of the 4Fe-4S centers 2 and 3, which has been suggested to constitute the difficult step in protein assembly.

Indeed, Vanoni and Curti (1999) found that there was a correlation between the lack of activity of recombinant GOGAT with the absence of Fe-S content. Also, activation of GOGAT is thought to depend on the exposure of the catalytically critical Cys-1 residue following cleavage of the leader peptide (Chapter 2). While these results imply that the GltA propeptide is correctly processed in *E. coli*, at present nothing is known about this process. It is interesting to note however, that the amino acid sequence around the Cys1 residue is well conserved in all types of GOGAT enzymes (Appendix C), and thus may be sufficient to determine correct protein maturation in such heterologous hosts. Cofactor dependency may also be a consideration. Ideally, we would like to optimise the overexpression and purification of active recombinant holoenzyme, as reported for the *A. brasilense* GOGAT overproduced in *E. coli* (Stabile *et al.*, 2000), to conduct more detailed analyses, however, the finding that the enzyme appears to be sensitive to oxygen (see Chapter 5) imposes serious limitations to this work.

4.5. Conclusion

From these *E. coli* complementation studies, we conclude that the *C. acetobutylicum* P262 *gltA* and *gltB* genes are the structural genes encoding the α and β subunits of GOGAT respectively, and furthermore that the enzyme is sensitive to oxygen. These results imply that functionally mature *C. acetobutylicum* GOGAT enzyme can be produced in this heterologous host from the separate expression of the two subunit genes, although clearly the conditions we have used are not optimal.

In spite of the overall high degree of homology shared between bacterial GOGAT subunits, the *C. acetobutylicum* β subunit cannot complement the *E. coli* α subunit. This suggests that there are critical subunit interactions required to produce functional enzyme which are satisfied between the *E. coli* α and the *T. ferrooxidans* β subunits. On the other hand, the β subunit of *C. acetobutylicum* appears to share a common property with the β subunits from *E. coli* and *T. ferrooxidans,* since expression of these polypeptides in the presence of glutamate resulted in down-regulated *E. coli* growth.

The inability of the *gltX* gene to restore GOGAT activity in *E. coli* MX3004 or *B. subtilis* 1A490, as well as the failure to complement the *C. acetobutylicum* GOGAT α subunit in MX3004, strongly suggests that it does not encode an independent GOGAT enzyme, nor an

alternative functional form of the β subunit. The fact that it does not exert the down-regulated growth phenotype in *E. coli* in the presence of glutamate, further suggests it has a function different from that of GOGAT β subunits.

CHAPTER 5

Regulation of GS and GOGAT activity from *C. acetobutylicum* P262

5.1 Summary **128**
5.2 Introduction **129**
5.3 Materials and methods **130**
5.3.1 Bacterial strains ---------- 130
5.3.2 Media, growth conditions and culture analysis ---------- 130
5.3.3 Preparation of crude cell free extract (CFE) ---------- 131
5.3.4 Protein concentration determination ---------- 131
5.3.5 Enzyme assays ---------- 132
5.3.6 RNA methods ---------- 133
5.3.7 Software analysis of dot blots ---------- 135
5.4 Results and discussion **135**
5.4.1 Growth and differentiation in relation to nitrogen source ---------- 135
5.4.2 GS activity as an indicator of nitrogen metabolism ---------- 139
5.4.3 Characterization of GOGAT activity from *C. acetobutylicum* P262 ---------- 140
5.4.4 Comparison of GS and GOGAT activity levels ---------- 144
5.4.5 Feedback regulation of GS and GOGAT activity ---------- 146
5.4.6 GDH activity assays ---------- 149
5.4.7 Transcriptional regulation of the *glnA*, *glnR*, *gltA* and *gltB* genes. ---------- 150
5.4.8 Quantitative analysis of *glnA*, *glnR*, *gltA* and *gltB* gene expression ---------- 152
5.4.9 Expression of the *C. acetobutylicum gltX* gene ---------- 155
5.5 Conclusion **156**

CHAPTER 5

Regulation of GS and GOGAT activity from *C. acetobutylicum* P262

5.1 Summary

Spore germination, growth profiles and cell differentiation of *C. acetobutylicum* P262 cultures were assessed in relation to different combination of nitrogen sources. Organic nitrogen (in the form of casamino acids) was the preferred source of nitrogen, while ammonia appeared to retarded spore germination and cell development. These studies, in conjunction with GS indicator assays, led to the definition of nitrogen limiting conditions (0.025% casamino acids and 0.15% glutamine, also referred to as the inducing conditions), and nitrogen rich conditions (0.2% casamino acids, also referred to as the non-inducing conditions) which were used subsequently for enzyme regulatory studies.

GOGAT assay conditions were established and optimised. GOGAT activity was found to be specific for the cofactor NADH, and sensitive to oxygen. Both GS and GOGAT activities were regulated by the nitrogen source in a similar way, being most induced by the defined nitrogen limiting conditions, and repressed by the defined nitrogen rich conditions. At early exponential phase (OD_{600} 0.3), a 6.9 and 5.6 fold difference was recorded between the inducing and non-inducing nitrogen conditions for GS and GOGAT activities, respectively. No assimilatory GDH activity could be detected.

Quantitative Northern blot analyses and feedback inhibition studies, indicated that the GS and GOGAT activities were regulated by the nitrogen conditions primarily at the level of transcription. Furthermore, during growth in the nitrogen limiting conditions, *glnA* and *glnR*, and *gltA* and *gltB* were each transcribed as an operon. Expression of the second β subunit-like gene *gltX*, was not regulated by the nitrogen source.

5.2 Introduction

Although the mechanism of regulation of solvent production in *C. acetobutylicum* is not fully understood, studies showed that efficient sporulation was required for the maintenance of high levels of solvent production. In *C. acetobutylicum* P262, ammonia levels were reported to be important for solventogenesis and the initiation of sporulation (Long *et al.*, 1984; Jones and Woods, 1986). However, while ammonia is the preferred source of nitrogen for most microorganisms, *C acetobutylicum* P262 grew poorly when ammonia salts, either alone or in combination with glutamine and/or glutamate (at 0.1 or 0.2% w/v) served as the nitrogen source (maximum OD_{600} = 0.4) (Fierro-Monti *et al.*, 1992)(Dr L. Brown, unpublished results). On the other hand, *C. acetobutylicum* grows well in CBM media which is a rich source of organic nitrogen. In order to study the regulation of the key enzymes of nitrogen assimilation, GS, GOGAT and GDH, it was first necessary for us to establish what nitrogen conditions defined a nitrogen-rich versus a nitrogen-poor environment in relation to the different stages of cell differentiation. Marked differences have been observed in the utilization and assimilation of nitrogen sources in the few clostridial strains that have been studied (Kleiner, 1979; Bogdahn *et al.*, 1983; Bogdahn and Kleiner, 1986; Kanamori *et al.*, 1989; Amine *et al.*, 1990).

The regulation of bacterial GS and GOGAT enzymes has been discussed in Chapter 1. GS activity was not regulated by adenylylation in *C. acetobutylicum* P262 (Usden *et al.*, 1986), and in *B. subtilis,* GS was regulated primarily at the transcriptional level (Fisher *et al.*, 1984). In *C. acetobutylicum*, antisense RNA complementary to the 5' end of the *glnA* mRNA, has been implicated in the post-transcriptional regulation of the *glnA* gene (Fierro-Monti *et al.*, 1992). The mechanism by which GOGAT genes are regulated by nitrogen availability appears to differ from one microorganism to another, although in general, GOGAT activity has been found to be repressed by glutamate, or a good source of glutamate. In *B. subtilis*, expression of the *gltAB* operon is positively regulated by the divergently transcribed *gltC* product (Bohannon *et al.*, 1985; Bohannon and Sonenshein, 1989). In *C. acetobutylicum* P262, the structural genes encoding the GS and GOGAT enzymes are separated by a gene (*glnR*) coding for a putative response regulator protein, GlnR, which has been suggested to act as a positive regulator of *glnA* expression (Woods and Reid, 1995).

In addition, since the *C. acetobutylicum gltX* gene (described in Chapter 3) shared striking homology with GOGAT β subunits, we wished to establish whether its expression was regulated by nitrogen.

5.3 Materials and methods

5.3.1 Bacterial strains

C. acetobutylicum P262 wild type strain (Jones *et al.*, 1982) was used in these growth and enzyme activity studies, and served as the source of total RNA. *Bacteroides fragilis* strain Bf1 (Abrahams and Abratt, 1998) served as the positive control for the GDH assays.

5.3.2 Media, growth conditions and culture analysis

Growth of *C. acetobutylicum* P262 was initiated from a spore stock maintained aerobically at 4°C in sterile water. Spores were activated by heat shock treatment (5 min at 70 °C), chilled on ice for 2 min, and allowed to de-oxygenate in the anaerobic glove cabinet for approximately 5 min prior to being used as the primary inoculum. All cultures were grown anaerobically at 37 °C in either buffered CBM (Allcock *et al.*, 1982), or in glucose-mineral salts-biotin minimal medium (GSMM, see Appendix E) (Holdeman *et al.*, 1977) modified by containing various combinations and concentrations of carbon (1 or 2% glucose), organic nitrogen (casamino acids (DIFCO), glutamine and monosodium glutamate (MSG)) and/or inorganic nitrogen (predominantly NH_4OAc), as discussed in the text.

The growth profile, pH and cell morphology of cultures of *C. acetobutylicum* grown in these various modified GSMM media, was analysed in duplicate at various growth stages on both cell line and spore line cultures grown in 10 ml Hungate tubes. Spore line refers to cultures grown directly from 5 µl of heat shocked spores. The cell line cultures were initiated from cells germinated in 10 ml of CBM, grown to OD_{600} 0.3, resuspended in an equal volume of sterile distilled water and used at a 5% inoculum volume. At selected time intervals during the growth cycles, 100 µl aliquots were withdrawn, mixed with 900 µl of 0.4% formalin, and 5 µl of these suspensions counted microscopically using a Thoma counting chamber. Counts were done to differentiate between the percentage of phase-bright spores (endospores) and cells. The latter included a wide variety of morphology types; mobile and immobile rods, chains, granular and non granular cells, and phase-dark spores.

In addition to these physiological analyses, GS activity was determined on cultures of *C. acetobutylicum* grown with these various combinations of nitrogen sources (2% glucose), as a further indicator of nitrogen conditions. GSMM media containing 0.025% casamino acids and 0.15% glutamine as the sole nitrogen source was selected to represent nitrogen limiting conditions (also referred to as GSMM inducing media). GSMM media containing 0.2% casamino acids as the sole nitrogen source was selected to represent nitrogen rich conditions (or GSMM non-inducing media). These media (and modifications thereof as discussed in the text) were used for studying the regulation of GS, GOGAT and GDH activity.

B. fragilis Bf1 was grown anaerobically in 10 ml of BHI broth (Holdeman *et al.*, 1977) in Hungate tubes incubated overnight at 37°C in a shaking water bath.

5.3.3 Preparation of crude cell free extract (CFE)

Crude CFE was prepared from cultures of *C. acetobutylicum* P262 and *B. fragilis* under strictly anaerobic conditions. All solutions and equipment used were pre-equilibrated in the anaerobic glove cabinet. At selected growth stages, approximately 300 mg of cells (equivalent to ~430 ml of culture at OD_{600} 0.3) were harvested from the relevant media, and washed twice in 60 ml 20mM KH_2PO_4-K_2HPO_4 buffer at pH 6.5 (except during enzyme pH studies when different pH buffers were used) in sealed centrifuge tubes. The pellets were resuspended in 6 ml of the buffer, and immediately disrupted in a French pressure cell. The lysate was collected under an anaerobic atmosphere in a sealed Hungate tube on ice, and transferred back to the glove cabinet for processing. The lysate was clarified (15 min spin in an Eppendorf micro centrifuge) at room temperature. The efficiency of lysis was routinely greater than 95% (as estimated microscopically), and was independent of the growth stage. This consistently yielded approximately 2.5 mg of protein per ml of CFE. If not used immediately for enzyme studies, the CFE was stored anaerobically at -70 °C.

5.3.4 Protein concentration determination

Protein concentrations were determined by the Bio-Rad micro protein assay system as suggested by the supplier (Hercules), using bovine serum albumin (fraction V) as the standard. Assays were always performed in duplicate. Protein samples were diluted in KH_2PO_4-K_2HPO_4 buffer pH 7, such that the absorbance did not exceed 0.55, since this corresponds to the upper limit of linearity for this method.

5.3.5 Enzyme assays

GOGAT and GDH assays were performed in the anaerobic glove cabinet (Forma Scientific Inc.). All solutions and equipment required for these experiments were equilibrated in the cabinet prior to use. Due to the restrictions imposed by working in the glove cabinet, the number of samples which could be processed was limited.

5.3.5.1 GOGAT assay

CFE prepared from selected cultures of *C. acetobutylicum* were assayed for GOGAT activity by spectrophotometrically measuring the rate of NADPH or NADH oxidation according to the procedure of (Meister, 1985). The peak absorbance of NAD(P)H is 340 nm. However, due to the requirement for anaerobic assay conditions (see Section 5.4.3.3), we were restricted to monitoring the reaction on a Spectronic 20 (Bausch and Lomb) spectrophotometer at 366 nm in the glove cabinet, which may account for the initial lag phase observed during the assays.

Due to these modified conditions and the use of crude enzyme preparations, Km values were not determined, however substrate, cofactor and enzyme concentrations were selected such that the rate of cofactor oxidation was always linear between OD_{366} 0.4 and 0.1, and proportional to the quantity of protein added. The standard reaction mixture contained 20 mM potassium phosphate (KH_2PO_4-K_2HPO_4) buffer, 5 mM each of the substrates L-glutamine and 2-oxoglutarate, 0.4 mM of cofactor (NADH or NADPH), and appropriate amounts of the enzyme source, CFE, in a final volume of 5 ml. The reaction was initiated by the addition of cofactor. Unless otherwise specified, the standard assay conditions employed were 37 °C, pH 6.5. Specific activity, calculated from the linear stage of the reaction, was expressed as μmoles of NAD(P)H oxidized (determined from a standard curve) per min per mg of protein. NADPH was naturally less stable than NADH.

For the pH optimum studies, all assay components were resuspended in the respective KH_2PO_4-K_2HPO_4 pH buffers. To study feedback inhibition of GOGAT activity, the CFEs were incubated in the presence of various concentrations of test compounds for 15 min at 37 °C, before the reaction components were added and the reaction initiated. For these experiments, CFE which had been stored at -70 °C was used under the standard reaction conditions.

5.3.5.2 GDH assays

Assimilatory GDH activity assays were conducted on *C. acetobutylicum* CFE essentially as described for the standard GOGAT assay, except that 100 mM NH_4Cl replaced glutamine, and 2-oxoglutarate was used at a 10 mM concentration. Both cofactors were tested over a range of pHs. CFE, prepared from overnight cultures of *B. fragilis* Bf1, served as the positive control.

5.3.5.3 GS assays

GS activities were measured by the γ-glutamyl transferase (GGT) assay described by Shapiro and Stadtman (1968). This assay has no requirement for ATP activation and is less subject to interference by contaminating enzymes than the biosynthetic reaction. The assays were performed on whole cell suspensions, from various cultures of *C. acetobutylicum,* made porous by pre-treatment with C-Tab (hexadecyltrimethyl-ammonium bromide): under anaerobic conditions, 25 ml of culture at the desired growth stage, was mixed with 0.01% C-Tab for 10 min, harvested in sealed centrifuge tubes, and washed in an equal volume of saline (0.8% NaCl). The pellets were resuspended in 0.5 ml of GGT resuspention buffer, and the assays performed immediately, in duplicate, under aerobic conditions.

Transferase activity was followed by measuring the formation of γ-glutamyl hydroxamate in the reaction mixture (0.5 ml): 0.1 ml cell suspension, 148 mM imidazole-HCl buffer pH 7.15, 19.5 mM hydroxylamine, 0.3 mM $MnCl_2$, 28 mM potassium arsenate, 0.4 mM NADP. The reaction was initiated by the addition of 50 µl of L-glutamine (0.2 M), and allowed to proceed for 15 min at 37 °C, after which it was stopped by the addition of 1 ml of GGT termination mix. The reaction mix was clarified (10 000 rpm for 1 min in an Eppendorf micro centrifuge), and the absorbance of the supernatant read immediately at 540 nm. The amount of γ-glutamyl hydroxamate formed was determined from a standard curve. Enzyme specific activity was expressed as µmoles of γ-glutamyl hydroxamate produced per min per mg of protein.

For the feedback inhibition studies, CFE served as the enzyme source, and was pre-incubated in the presence of the test substances for 15 min at 37 °C prior to the assay. Results were interpreted relative to a control.

5.3.6 RNA methods

To prevent RNA degradation, strict RNAse-free conditions were employed during all RNA handling procedures. All glassware was heat treated. Only Milli-Q water (Millipore water

purification system) and RNAse-free chemicals were used. Plastic ware and suitable solutions were subjected to two cycles of autoclaving.

5.3.6.1 RNA extractions

C. acetobutylicum total mRNA was extracted from cultures grown in the GSMM inducing, GSMM non-inducing and CBM media at various growth stages as described in the text, using the hot phenol extraction protocol described by Aiba *et al.* (1981). RNA was stored in water at -70 °C in the presence of RNAse inhibitor (Sigma).

5.3.6.2 Northern blots

Northern blots were performed using the non-radioactive digoxigenin (DIG) Labeling and Detection kit according to the manufacturer's protocol (Boehringer Mannheim). RNA was separated by electrophoresis in 1.5% denaturing formaldehyde agarose gels according to Fourney *et al.* (1988). RNA molecular weight markers (Gibco BRL) were included. The RNA was transferred to a positively charged nylon membrane (Boehringer Mannheim) by capillary blotting for a duration of 10 hours at room temperature. For the dot blots, RNA was transferred directly onto the membrane using a dot blot suction apparatus. A 50 µl volume, containing the relevant concentration of RNA, was loaded per well. RNA was fixed onto the membranes by UV exposure. Probes were hybridised at 68 °C at a concentration of approximately 100 ng/ml, followed by stringent washing conditions. Probe bound to the membrane was detected by the chemiluminescent reaction using the substrate AMPPD.

5.3.6.3 RNA probes

DIG labelled homologous RNA probes were synthesized, from template clones constructed in the Bluescript vector pSK, according to the protocol supplied by Boehringer Mannheim. Information about the respective probes is illustrated in Fig. 5.1. The *glnA*-specific probe was synthesized from a clone containing an internal 0.82 kb *Eco*R1-*Hind*III *glnA* fragment derived from pHZ200 (Fig. 2.1). Similarly, the *glnR* probe was synthesized from a clone harbouring the internal 0.31 kb *Eco*RV-*Xba*I fragment. Probes specific for the *gltA* and *gltB* genes were synthesized from plasmid pHS2 and a shortening of construct pHS5, respectively (Fig. 2.1). The orientation of the individual cloning events determined whether the T_7 or T_3 polymerase specific promoter was used to direct probe synthesis. The probe lengths were defined by linearization of the respective template DNAs at an appropriate restriction site downstream of the promoter. The template fragments were gel-purified using Gene Clean (Amersham) prior to

probe synthesis. The *gltX*-specific gene probe used in these studies is described in Section 3.3.7.

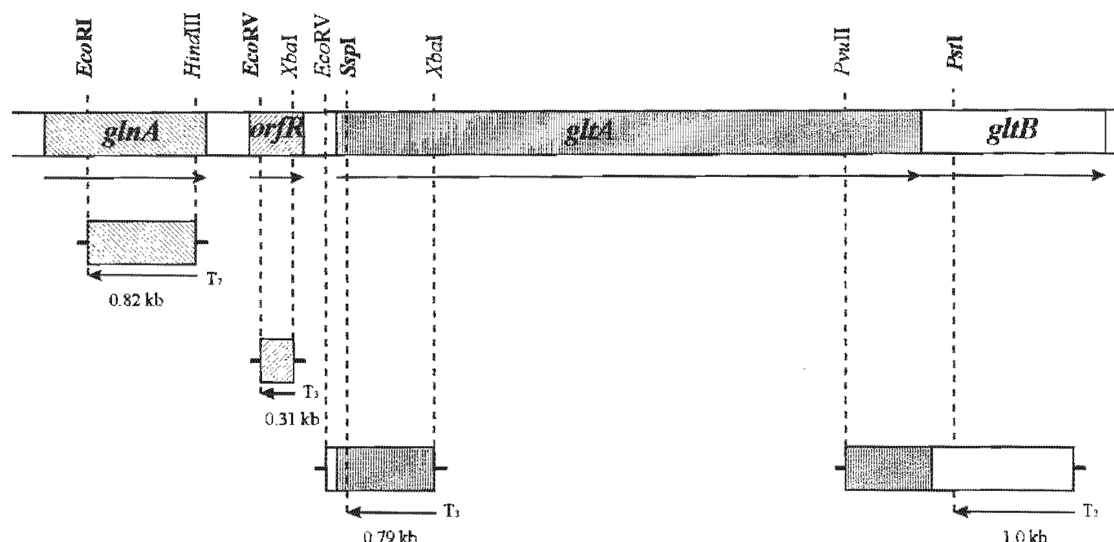


Fig. 5.1 Origin of the RNA probes synthesized to the *glnA*, *glnR*, *gltA* and *gltB* genes. The thin horizontal lines represent vector sequence (Bluescript pSK) of the respective template clones, and the arrows below each clone indicate the relative size of the RNA probes and the direction in which they were synthesized. The respective template linearization sites are indicated in boldface type.

5.3.7 Software analysis of dot blots

The software program Gel Trak (D. Maeder, Department of Biochemistry, University of Cape Town) was used to quantitate the relative intensities of signals obtained in the RNA dot blot experiments. It supports a densitometer which scans a defined set of data interpreting each signal relative to a specified maximum and minimum background intensity.

5.4 Results and discussion

5.4.1 Growth and differentiation in relation to nitrogen source

With the assistance of Dr. L. Brown (University of Cape Town post-doctorate) and Prof. D. Jones (University of Otago, New Zealand; personal communication), the growth and morphology of both spore and cell line cultures (see Section 5.3.2) of *C. acetobutylicum* P262, grown in a defined medium containing a variety of organic and/or inorganic nitrogen sources, were examined. Growth was found to be proportional to the concentration of organic nitrogen added in the form of casamino acids (Fig. 5.2A), with growth of the cell line culture containing 0.2% w/v casamino acids being almost equivalent to that of the CBM culture for the first ~18

hours. Further addition of the organic nitrogen sources glutamine or glutamate (at 0.2% w/v) made no significant difference (results not shown). In the absence of casamino acids no growth occurred initially, however, after 27 hours highly motile spindle shaped cells were detected in the cell line cultures, and by 90 hours growth was significant. This raised the possibility that *C. acetobutylicum* P262 could fix nitrogen from the atmosphere, consistent with the fact that organisms switch on their nitrogen fixing systems under nitrogen starved conditions. The ability to fix nitrogen has been reported in several clostridia (Merrick and Edwards, 1995), and a cluster of *nif* genes have been identified in the unfinished genome sequence of *C. acetobutylicum* ATCC824 (accession number AE001438).

When ammonia acetate (0.2% w/v NH_4OAc) was added to the media containing various concentrations of casamino acids, no significant differences were observed in the cell line cultures, however, growth of the equivalent spore line cultures were notably retarded during the first 24 hours (results not shown). Interestingly, NH_4OAc (present as the sole source of nitrogen), could not support outgrowth of the spore line, however, it supported steady growth of the cell line after an initial lag phase (Fig. 5.2A1), reaching an OD_{600} of 1.5 by 90 hours. Together these results indicated that organic nitrogen is essential for spore germination and is clearly the preferred source of nitrogen, and that ammonia appears to affect the germination of spores. It is worth noting here that partially hydrolyzed proteins or amino acids were required by *C. beijerinckii* for solvent production (Prescott and Dunn, 1959), and that glutamine is the preferred source of nitrogen for *B. subtilis* (Fisher and Sonenshein, 1991) and glutamate is the preferred source of nitrogen for *Corynebacterium callunae* (Ertan, 1992).

The retarding effect of ammonia acetate was more closely examined in the spore line cultures in the presence of 0.2% w/v casamino acids (Fig. 5.2B). Repression of growth was proportional to the NH_4OAc concentration added above 0.05% w/v. This effect was not related to acetate, and was also observed with alternate sources of ammonia including NH_4Cl, NH_4NO_3, and $(NH_4)_2SO_4$ at 0.4% w/v (results not shown). Morphological analyses confirmed that ammonia affected spore germination and early cell growth. However, a positive correlation was observed between the percentage sporulation and the concentration of ammonia; at 24 hours 66% spores were recorded in media containing 0.2% w/v ammonia, whereas in media containing no ammonia only 3% spores were recorded. This phenotype has been previously observed (Long *et al.*, 1984), which led to the conclusion that ammonia played an important role later on in the growth cycle for sporulation and the onset of solventogenesis. However, by increasing the

glucose concentration from a standard 1% to 2% w/v, the growth profiles were improved up to 40 hours, sporulation was delayed, and the trend was no longer proportional to the concentration of ammonia. The percentage sporulation recorded in cultures containing 0.1% w/v NH_4OAc or less (including no NH_4OAc) was between 60 and 80 percent. In all cases sporulation efficiency correlated with increased pH of the culture medium, consistent with the conversion of acids to solvents. These observations were however dependent on a minimum concentration of 0.2% w/v casamino acids; at lower concentrations the growth cycle became progressively retarded and spore development impeded. The addition of glutamine or glutamate to yield a final organic nitrogen concentration of 0.2% w/v, was unable to compensate. This indicated that growth and differentiation of *C. acetobutylicum* P262 was dependent on specific amino acids within the casamino acid cocktail, and not just on the total organic nitrogen content.

Taken together, these results suggest that ammonia is neither critical nor necessary for growth or sporulation as previously thought, and in fact may actually inhibit the uptake of casamino acids in the medium. Rather, carbon limitation was the cause of the low sporulation efficiency associated with low concentrations of ammonia. It is propose that ammonia affects the efficiency of spore germination and growth so that, at high concentrations of ammonia, carbon is less likely to be limiting, enabling more of the germinated cells to proceed to the sporulation stage. In addition, organic nitrogen limitation retarded growth and impeded sporulation. Thus it is reasonable to suppose that the organism regulates its metabolism according to the availability of carbon and nitrogen, and that high organic nitrogen levels represent nitrogen rich conditions, and ammonia and/or low organic nitrogen availability represent a nitrogen poor status.

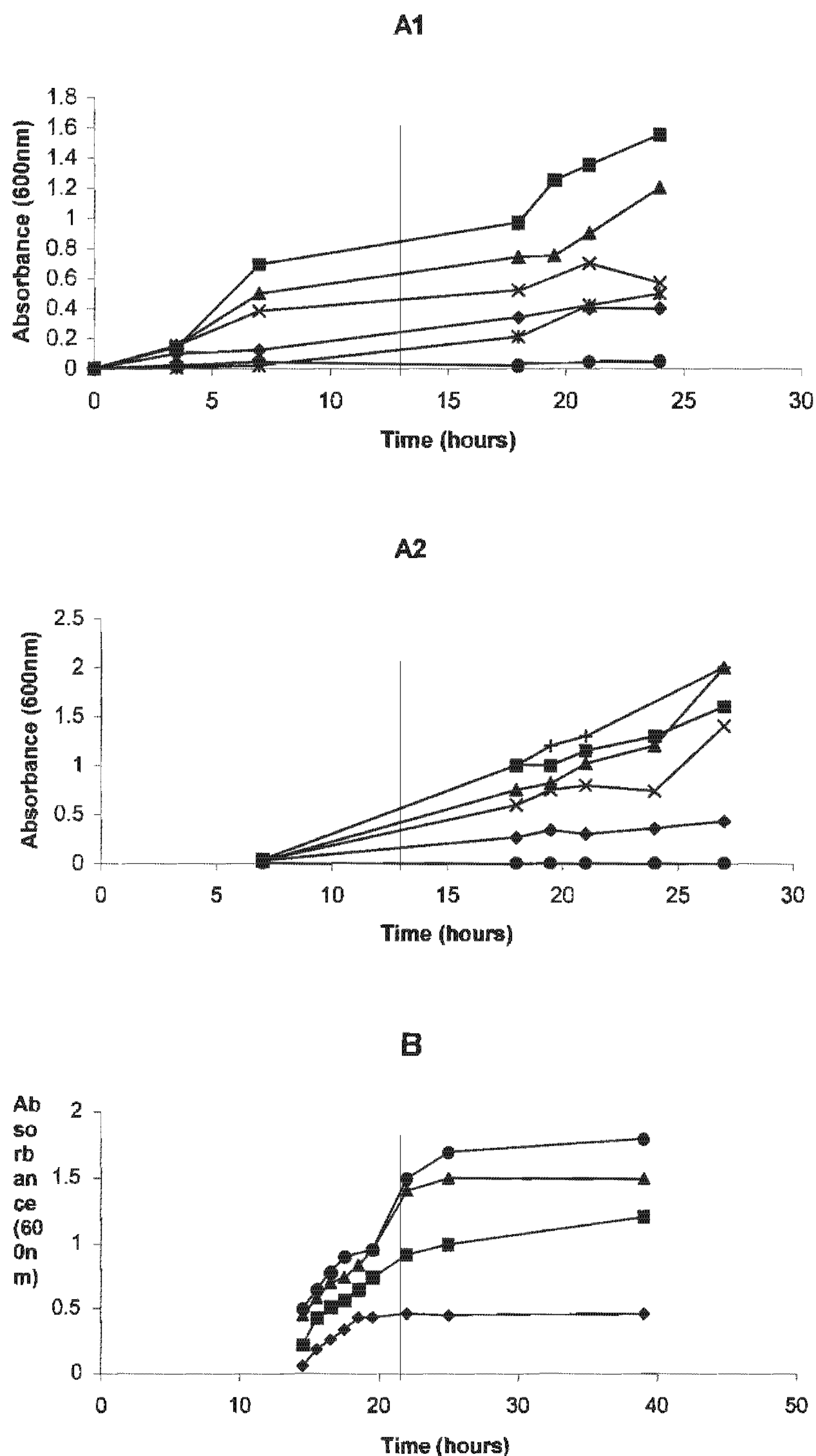


Fig. 5.2 The effect of carbon, inorganic and organic nitrogen sources on the growth of *C. acetobutylicum* P262 monitored at an absorbance of 600nm. (A) illustrates the growth profiles of cell line cultures (1) and spore germinated cultures (2) grown in GSMM media containing 1% w/v glucose and either 0.2% w/v ~~NH_4OAc (*)~~ or various concentrations of casamino acids (w/v): 0.2% (■), 0.1% (▲), 0.05% (×), 0.01% (♦) and no casamino acids (●). A CBM culture control was included in A2 (+). (B) illustrates growth profiles for spore germinated cultures in GSMM media containing 2% w/v glucose, 0.2% w/v casamino acids, and various concentrations of NH4OAc (w/v): 0.4% (●), 0.2% (■), 0.1% (▲) and 0.05% (♦).

5.4.2 GS activity as an indicator of nitrogen metabolism

In conjunction with these physiology results, GS activity assays, which have been optimised for *C. acetobutylicum* and can be conveniently performed aerobically, were used as an indicator of general nitrogen metabolism in order to select nitrogen-rich conditions (associated with repressed GS activity), and nitrogen-limiting conditions (associated with induced GS activity) for enzyme regulatory studies. Table 5.1 summarizes the GS activities measured from cultures grown with various combinations of nitrogen sources (2% w/v glucose). Only media that could support a steady state culture at OD_{600} ~0.3, equivalent to early exponential phase and typified by elongated highly motile non granular cells, were considered.

As expected, GS activity was inversely proportional to the concentration of casamino acids present in the medium. Neither ammonia nor glutamate appeared to significantly affect GS activity in any predictable way. However, GS levels were unexpectedly elevated in the presence of its product, glutamine, a result which is counter-intuitive to the role for enzyme in anabolism. It is possible that glutamine is being cleaved on its way into the cell by a membrane-bound glutaminase activity which could have the effect of raising the intracellular glutamate and ammonium concentration, which in term would result in inducing GS activity. Perhaps *C.acetobutylicum* can only assimilate ammonium as part of an organic compund. Thus, the highest levels of GS recorded were for cultures grown in 0.025% casamino acids containing 0.15% glutamine. These nitrogen conditions were selected to represent the nitrogen-limiting conditions (and referred to subsequently as the GSMM inducing media). Conversely, a distinct repression of GS activity was observed in cultures grown with a high level of casamino acids (0.2%). This media was thus selected to represent the nitrogen-rich conditions (referred to subsequently as the GSMM non-inducing media). These induced and non-induced trends were most pronounced at early exponential growth phase (OD_{600} 0.3), during which GS levels varied as much as 9.2 fold.

In contrast, the biosynthesis of GS from *C. acetobutylicum* ATCC 824 was regulated by the ammonia concentration in the culture medium, although, as in *C. acetobutylicum* P262, the addition of high concentrations of casamino acids (1% w/v) also repressed GS biosynthesis (Amine *et al*., 1990).

Table 5.1. GS activity trends determined for spore germinated cultures of *C. acetobutylicum* P262 grown in GSMM media containing 2% w/v glucose and various combinations of nitrogen sources, at early (OD_{600} 0.3) and late (OD_{600} 0.5) exponential growth phase. Each sample was assayed in duplicate. GS activity is expressed as µmoles of γ-glutamyl hydroxymate produced/min/mg of protein.

Nitrogen source	Growth stage (OD_{600})	Sample size	Average GS activity	Std deviation
CBM	0.3	4	0.088	0.013
0.2% casaa	0.3 0.5	8 3	0.125 0.058	0.016 0.021
0.2% casaa + 0.2% NH_4OAc	0.3 0.5	4 4	0.155 0.185	0.045 0.019
0.05% casaa	0.3 0.5	2 2	0.359 0.515	
0.05% casaa + 0.2% NH_4OAc	0.3	2	0.363	
0.05% casaa + 0.2% NH_4OAc +0.15% gln	0.3 0.5	3 2	0.540 0.38	0.074
0.05% casaa + 0.2% NH_4OAc + 0.15% MSG	0.3 0.5	2 2	0.322 0.41	
0.05% casaa + 0.15% gln	0.3 0.5	10 6	0.536 0.34	0.056 0.112
0.05% casaa + 0.15% MSG	0.3 0.5	2 2	0.323 0.544	
0.025% casaa + 0.15% gln	0.3	10	0.806	0.097

Abbreviations: glutamine (gln), monosodium glutamate (MSG), casamino acids (casaa) and ammonia acetate (NH_4OAc).

Percentages (%) imply w/v.

5.4.3 Characterization of GOGAT activity from *C. acetobutylicum* P262

Having established a pattern of nitrogen source dependent regulation of GS activity, we were interested to know whether the closely coupled and genetically linked GOGAT pathway was

regulated by similar conditions. To achieve this we first needed to develop a suitable GOGAT assay system for this obligate anaerobic bacterium (see Section 5.3.5.1), and define the optimum enzyme conditions. For this purpose we used CFE from mid exponential phase (OD_{600} 1.5) CBM cultures.

GOGAT activity was detected using a variation of the conditions described by (Meister, 1985) Under anaerobic conditions, 1.5 mg of mid exponential phase CFE protein was capable of oxidizing 98.04 μM of NADH cofactor in 9.2 min at 37 °C at pH 7.0 (Fig. 5.3). This converts to a specific activity of 17.43 μM of NADH oxidized per min per mg of protein. The exclusion of either one of the substrates in the reaction, resulted in no measurable NADH oxidation, confirming that the assay was specific for GOGAT activity. The enzyme was also specific for its substrates. Substitution of NADH with NADPH decreased the activity to almost zero, and ammonium chloride (100 mM) could not be substituted for glutamine. The enzyme activity remained stable when stored on ice for up to one hour under anaerobic conditions, and only lost approximately 10% activity after anaerobic storage of the CFE for 48 hours at -70 °C. To account for any non-specific coenzyme oxidation, assay results were interpreted relative to a control containing only CFE and coenzyme.

5.4.3.1 Optimum pH

The production of solvents by the clostridia is influenced by the pH, and the pH of the culture was shown to modify the level of some enzyme activities (O'Brien and Morris, 1971). In general GOGATs have been reported to act optimally in the neutral and/or basic range (Meers *et al.*, 1970; Meister, 1985). In *C. callunae,* both NADPH and NADH-dependent GOGAT activity has been reported which show different pH optimums (Ertan, 1992). Mid exponential phase cultures of *C. acetobutylicum* displayed a broad pH range for the NADH-dependent GOGAT activity in KH_2PO_4-K_2HPO_4 buffer, with a pH optimum between 6.3 and 6.5 (Fig. 5.4). There was only a difference of 14.6% between the two extremes tested, 5.8 and 7.9. A broad pH profile is consistent with the enzyme being able to function efficiently within a constantly fluctuating pH environment. No measurable NADPH-dependent GOGAT activity could be detected over this pH range, and it thus seems highly unlikely that this cofactor is involved in GOGAT activity. These results support the proposal stated herein, that *C. acetobutylicum* codes for a NADH-dependent enzyme based on analysis of the functional domains within the β subunit (Section 2.4.5.3).

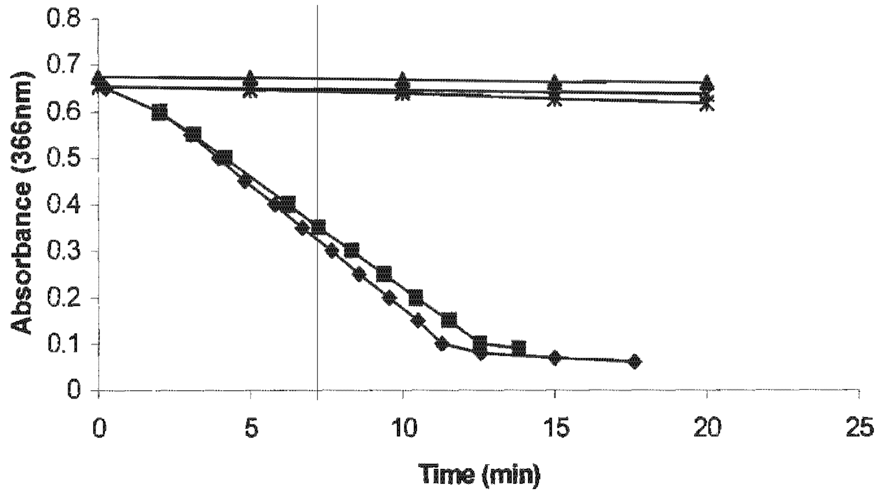


Fig. 5.3 *C. acetobutylicum* P262 GOGAT activity determined spectrophotometrically by the rate of cofactor oxidation. Represented is the complete system containing the NADH cofactor and freshly prepared CFE (♦), or CFE stored at -70 °C for 48 hours (■). Also included is the complete system with NADPH (*), or NADH minus either L-glutamine or 2-oxoglutarate (×). NADH and CFE only are represented by (▲). In all cases 1.5 mg of CFE protein was used.

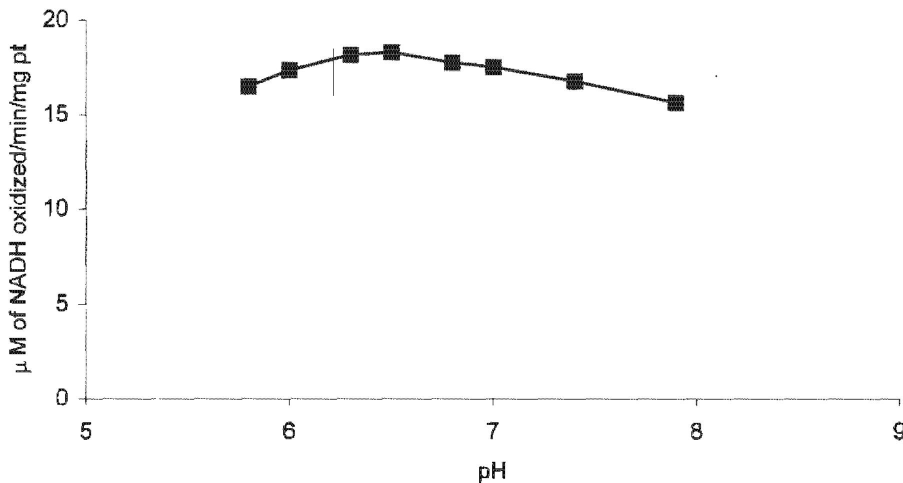


Fig. 5.4 pH profile of *C. acetobutylicum* P262 NADH-dependent GOGAT activity measured at 37°C. The results represent a trend from three experiments using mid exponential phase (OD_{600} 1.5) CFE prepared from CBM grown cultures.

5.4.3.2 Temperature optimum

The rate of the GOGAT reaction increased significantly with temperature up to 48 °C (Fig. 5.5), reflecting an increase of 2.5 units from 37 °C, and a thermostable enzyme. However, if the CFE was incubated for 10 min at 48 °C prior to initiation of the reaction, no activity could be detected initially, although after 60 minutes at room temperature (18 °C), GOGAT activity gradually increased (data not shown). This supports a subunit assembly model in which the disrupted subunits may be capable of reconstituting the active enzyme again under these conditions. When the reaction was initiated at 48 °C immediately, the activity was not disrupted after a 10 minute period (equivalent to approximately half way through the linear stage of the reaction). Presumably the enzyme complex was stabilized by the interaction with cofactor and substrates. No GOGAT activity could be detected at 52 °C.

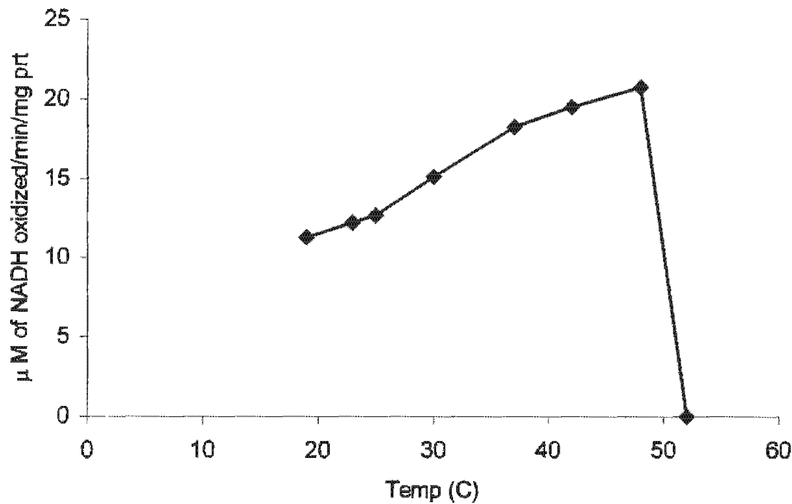


Fig. 5.5 *C. acetobutylicum* P262 GOGAT activity as a function of temperature. These results represent a trend from three experiments using CFE from mid exponential phase (OD_{600} 1.5) CBM grown cultures assayed at pH 6.5.

5.4.3.3 Oxygen sensitivity

Unlike *C. acetobutylicum* GS activity, GOGAT activity was very sensitive to oxygen (Table 5.2). A 10 minute exposure of the CFE to oxygen resulted in a 79% loss of activity. No GOGAT activity could be detected after a 30 min exposure to oxygen, and it was not reconstituted after two hours of anaerobic equilibration. Oxygen sensitivity may also account for the very low and unstable GOGAT activities reported from studies on *C. acetobutylicum*

ATCC 824 (Amine *et al.*, 1990). The implication of these results is that research on the enzyme will have to be restricted to experiments that can be performed in the anaerobic glove cabinet.

Table 5.2 Effect of oxygen on *C. acetobutylicum* P262 GOGAT activity determined on CFE from CBM grown cultures (OD_{600} 1.5) at 37 °C, pH 6.5.

Exposure of CFE to O_2	Anaerobic equilibration of CFE	GOGAT activity (μM of NADH oxidized /min/mg protein)	% Activity
0 min	by default	17.51	100
10 min	5 min	3.68	21
30 min	5 min	0	0
30 min	120 min	0	0
Aerobic assay	none	0	0

From these enzyme studies the following optimal conditions were selected to investigate GOGAT regulation; pH 6.5, 37 °C, NADH and strict anaerobic conditions. However, we cannot rule out the possibility that the nitrogen source may influence some of the properties.

5.4.4 Comparison of GS and GOGAT activity levels

Both GS and GOGAT activities were measured simultaneously on the same spore germinated cultures of *C. acetobutylicum* grown in 400 ml of the nitrogen limiting medium (GSMM inducing media; 0.025% casamino acids + 0.15% glutamine) and the nitrogen rich media (GSMM non-inducing media; 0.2% casamino acids), to determine their relative patterns of regulation.

As illustrated in Fig. 5.6, both enzyme activities were significantly induced and repressed by the same nitrogen conditions. At early exponential growth phase (OD_{600} 0.3), there was a 5.6 fold difference in GOGAT activity between the inducing (111.30 μM of NADH oxidized/min/mg prt.) and non-inducing growth media (19.92 μM of NADH oxidized/min/mg prt.). Similarly, a 6.9 fold increase was measured in GS activity between the inducing media (1.161 μM γ-glutamyl hydroxymate produced/min/mg prt.) compared to the non-inducing

growth media (0.167 μM γ-glutamyl hydroxymate produced /min/mg prt.). Both induced enzyme activity profiles decreased towards late exponential phase, however, whereas a gradual decrease was recorded for the GOGAT activity, the GS levels dropped a significant 3.5 fold between OD_{600} 0.6 and 0.8. The drop in enzyme activity towards stationary phase has also been observed in *B. subtilis* (Schreier, 1993), and suggests that neither enzyme is required at high levels during late stationary phase or sporulation. This may be explained by the fact that amino acids are produced by degradative pathways during sporulation in *Bacillus* spp, and suggests that the same may be true of *C. acetobutylicum*. Neither of the enzyme activity levels fluctuated much throughout growth in the non-inducing medium.

These results suggest that GS and GOGAT activities are regulated by similar nitrogen source availability, with the effect being most pronounced at early exponential phase. It would be interesting to determine whether glutamine also plays a specific role in the regulation of GOGAT activity as it does for GS activity. In *C. pasteurianum,* an increase in the glutamine pool levels correlated with increased GOGAT activity (Kleiner, 1979) while in *B. subtilis* an inverse relationship existed between GOGAT specific activity and the intracellular pool of glutamine (Deshpande *et al.*, 1981). In addition, it was investigated whether growth in glutamate affected GOGAT activity since it is known to repress the expression of the *glt* operon in *B. subtilis, E. coli* and *Saccharomyces* (Bohannon *et al.*, 1985; Castano *et al.*, 1988; Fisher, 1989; Valenzuela *et al.*, 1998). Growth in either the inducing or non-inducing media with various concentrations of glutamate (0.1%, 0.5% or 1% (53 mM)), had no effect on either GOGAT or GS activity, although since *C. acetobutylicum* is unable to grow with glutamate as the sole nitrogen source (results not shown), it was suspected that it cannot be taken up effectively by the cells.

The levels of enzyme activities measured are representative of the total amount of functional enzyme present in the cell at any given stage. However, these activities may not necessarily reflect the true amount of physiologically active enzyme present, since the assay procedure may dilute out the effect of any potential feedback regulators.

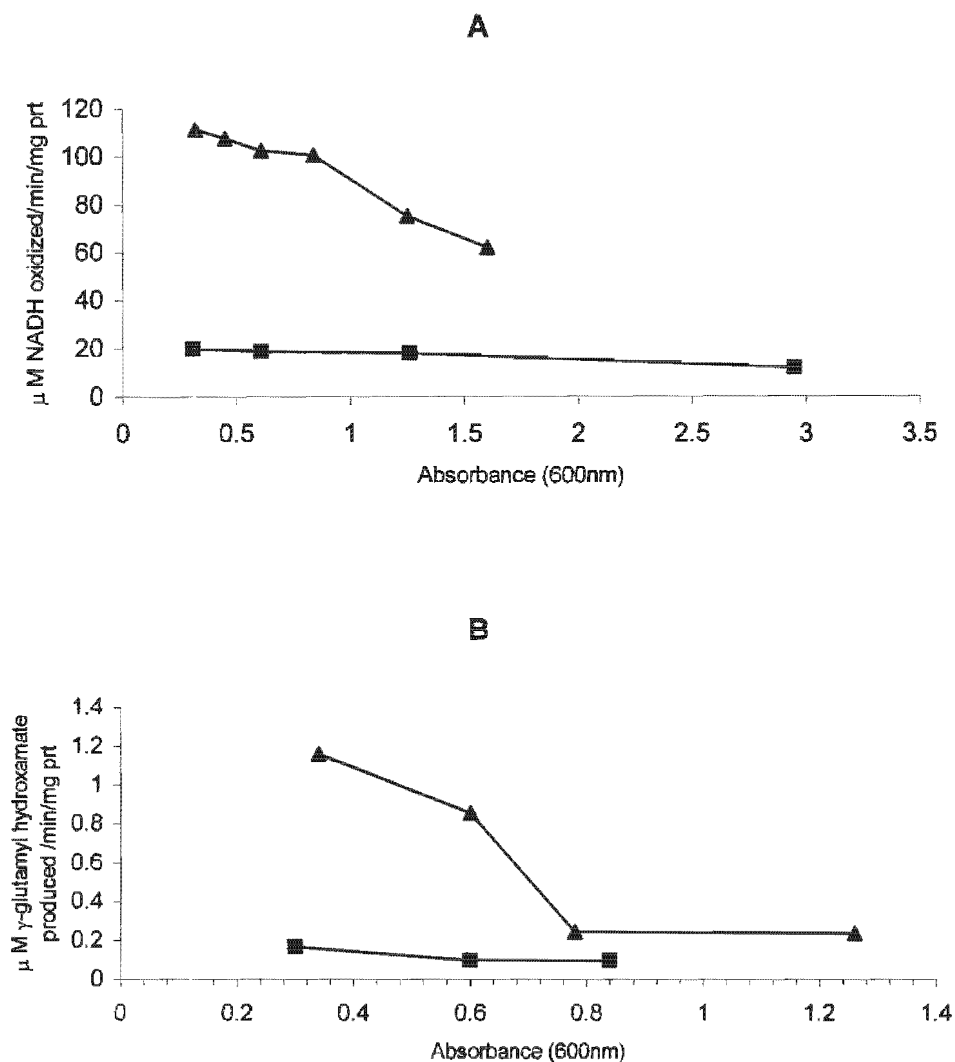


Fig. 5.6 *C. acetobutylicum* P262 GOGAT activity (A) and GS activity (B), measured from cultures grown in GSMM inducing media (▲) and GSMM non-inducing media (■), as a function of the growth stage. Each individual result is the average of at least three independent experiments in which the growth stages were selected within a maximum range of 0.05 absorbance units.

5.4.5 Feedback regulation of GS and GOGAT activity

Feedback inhibition is generally important for organisms with no post-translational modification systems. To investigate the role feedback inhibition may play in the regulation of *C. acetobutylicum* GS and GOGAT activities, CFE was prepared from cultures grown in GSMM inducing media at mid exponential growth (OD_{600} 0.4), and incubated in the presence of selected compounds for 15 min at 37 °C, prior to initiation of the respective reactions. The results are illustrated in Table 5.3 and 5.4 respectively.

A number of compounds have been shown to regulate GS activity by feedback inhibition. However, in *C. pasteurianum* GS displayed only weak feedback inhibition properties associated with a few amino acids (Schreier, 1993). A concentration of 0.25% w/v casamino acids (equivalent to the GSMM non-inducing growth conditions) exerted a partial inhibitory effect on the GS transferase activity of approximately 20%. This was increased to more than 30% by 0.5% w/v casamino acids. Neither L-glutamine, monosodium glutamate (MSG) or L-aspartic acid showed any effect on GS activity, although GS activity appeared to be sensitive to L-alanine, L-glycine and L-serine, with L-alanine exerting the largest effect: activity was inhibited by approximately 33% by 5 mM, and 68% by 15 mM. A mixture of these three amino acids appeared to have a cumulative effect, resulting in only 42.5% GS activity, and may suggest that these modifiers act independently. Although these test concentrations are probably not physiologically significant, these results give a general indication that GS activity is probably sensitive to a pattern of feedback regulation. Furthermore, the induced GS activity associated with cultures grown in the presence of glutamine is not due to glutamine mediated enzyme activation.

Table 5.3 Effect of various amino acids on *C. acetobutylicum* P262 GS activity. The 100% control value (no additives) was equivalent to 0.88 μM of γ-glutamyl hydroxamate produced/min/mg protein.

Amino acid tested	Concentration	% GS Activity
Casamino acids	0.5 % w/v	66.4 %
Casamino acids	0.25 % w/v	80.1 %
Casamino acids	0.025 % w/v	99.1 %
MSG	5 Mm	99.4 %
MSG	30 mM	93.2 %
L-glutamine	30 mM	96.0 %
L-alanine	15 mM	32.0 %
L-alanine	5 mM	67.4 %
L-aspartic acid	30 mM	92.5 %
L-aspartic acid	5 mM	94.8 %
L-glycine	15 mM	65.9 %
L-glycine	5 mM	86.2 %
L-serine	15 mM	65.3 %
L-serine	5 mM	83.4 %
L-ala, L-gly, L-ser	5 mM each	42.5 %

For *Bacillus* spp and most other bacteria, GOGAT has been shown to be relatively insensitive to regulation by feedback inhibition (Miller and Stadtman, 1972; Schreier and Bernlohr, 1984; Matsuoka and Kimura, 1986; Caballero *et al.*, 1989). A variety of amino acids, citric acid cycle intermediates, adenosine-containing nucleotides (Stabile *et al.*, 2000) and other metabolites have been examined for inhibitory capabilities and very few displayed significant inhibition at concentrations that were physiologically relevant. Apart from a few amino acids, L-malate was shown to be a potent effector in *B. subtilis*, yielding 70% inhibition at a 10 mM concentration (Schreier and Bernlohr, 1984). The coenzyme products $NAD(P)^+$ have been reported to show some measure of effective inhibition in *C. callunae, B. licheniformis* and *N. mediterranei* (Schreier and Bernlohr, 1984; Mei and Jiao, 1988; Ertan, 1992), although it is questionable whether these results were physiologically significant (Schreier and Bernlohr, 1984). As illustrated in Table 5.4, no significant regulation of GOGAT activity was observed by any of the compounds tested with the exception of glutamate. Although a marginal effect was exerted by L-serine, L-methionine and L-alanine at 30 mM concentrations, this was not considered to be physiologically significant. The only notable inhibition was caused by the product glutamate: concentrations of 15 and 5 mM resulted in approximately 20% and 15% activity inhibition, respectively. Glutamate has also been reported to inhibit *C. pasteurianum* enzyme activity. These results indicate that feedback inhibition is not likely to play a significant role in the regulation of GOGAT activity in *C. acetobutylicum*, although glutamate may be important.

Table 5.4 Effect of some compounds on *C. acetobutylicum* P262 GOGAT activity. The 100% control value (no additives) was 87.95 µM of NADH oxidized/min/mg of protein. These results reflect a trend from three independent experiments performed in duplicate.

Metabolite tested	Concentration	% GOGAT Activity
Casamino acids	0.5 % w/v	92.2 %
L-serine	30 mM	89.4 %
L-valine	30 mM	100 %
L-methionine	30 mM	90.2 %
L-alanine	30 mM	87.3 %
L-leucine	30 mM	98.9 %
MSG	15 mM	79.4 %
MSG	5 mM	85.6 %
NAD^+	1 mM	100.8 %
NAD^+	0.5 mM	102.1 %
Fumarate	30 mM	102 %
Malate	30 mM	101.7 %
$Mg2^+$	30 mM	96.2 %

5.4.6 GDH activity assays

While both the GOGAT and GDH pathways are found in most microorganisms, others including *B. subtilis*, and *C. pasteurianum* assimilate ammonia solely by the coupled GS-GOGAT pathway (Dainty, 1972; Deshpande and Kane, 1980). No assimilatory GDH activity could be detected in *C. acetobutylicum* cultures grown in either CBM, GSMM inducing or non-inducing media. Since it has been suggested that the GDH pathway may be important for the assimilation of ammonia during growth on low energy yielding nitrogen rich substrates, GDH assays were also conducted on cultures grown in the nitrogen rich medium (GSMM non-inducing media) restricted in the carbon source (0.5 and 0.25% w/v glucose), and supplemented with ammonia (100 mM NH_4Cl). Assays were performed at mid exponential growth phase with both coenzymes, NADH and NADPH, at 37 °C and varying the pH between 5.8 and 7.9. Still no significant cofactor oxidation could be detected. *B. fragilis* strain Bf1 provided a positive control. At 37 °C and pH 7.0, addition of 1 mg of CFE protein from overnight cultures grown in BHI broth (a rich organic nitrogen source), caused a rapid oxidation of NADH (>81.7 µM of

cofactor oxidized/min/mg prt.), while NADPH was oxidized at approximately a 15 fold lower rate (5.4 μM/min/mg prt.) (Fig. 5.7). From these results we predict that, like *C. acetobutylicum* ATCC 824 (Amine *et al.*, 1990) and *C. thermoautotrophicum* (Bogdahn and Kleiner, 1986), *C. acetobutylicum* P262 does not contain an assimilatory GDH activity, and that the energy dependent GS-GOGAT pathway is the primary route for ammonia assimilation. For obligate anaerobes such as *C. acetobutylicum*, development of high energy yielding fermentation pathways was probably crucial for their survival.

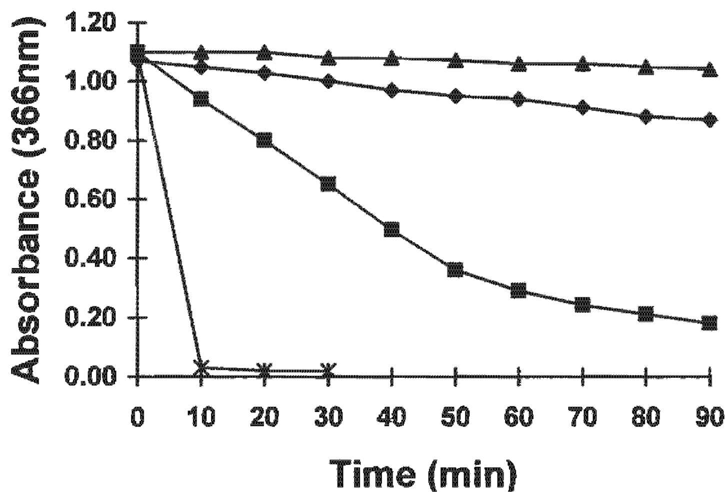


Fig. 5.7 GDH activity determined spectrophotometrically by the rate of cofactor oxidation from: overnight cultures of *B. fragilis* Bf1 grown in BHI and assayed using NADH (*) or NADPH (■) at 37 °C pH 7; or from mid exponential phase cultures of *C. acetobutylicum* P262 grown in modified GSMM non-inducing media (containing 0.05% glucose and 100 mM NH_4Cl) and assayed using NADH (▲) or NADPH (◆) at 37 °C, pH 6.5. In all cases 1 mg of crude CFE protein was used.

5.4.7 Transcriptional regulation of the *glnA*, *glnR*, *gltA* and *gltB* genes.

To investigate the regulation of GS and GOGAT activity at the level of transcription, total RNA was extracted from cultures of *C. acetobutylicum* grown in GSMM inducing and non-inducing media to early and late exponential phase, and hybridised with probes specific for *glnA*, *glnR*, *gltA* and *gltB* (Fig. 5.8). These Northern blots showed pronounced induction and repression in the expression of all four genes under the defined nitrogen limiting (GSMM inducing) and nitrogen rich (GSMM non-inducing) growth conditions, respectively.

Both the *glnA* and *glnR*-specific gene probes detected a single band of approximately 2.3 kb (panel A and B, respectively). Since the sizes of the *glnA* and *glnR* genes are 1.332 kb and

0.566 kb respectively, and are separated by an intergenic region of 200 bp, these results imply that under these nitrogen limiting conditions, *glnA* and *glnR* are transcribed as an operon, and the inverted repeat sequence identified downstream of *glnA* (Janssen *et al.*, 1988) does not function as a transcriptional terminator. It also follows that, under these nitrogen limiting conditions, expression of the regulatory *glnA* antisense RNA would be repressed. Similarly, these results confirm that the α and β subunits of GOGAT are co-transcribed. If the two subunits were independently transcribed, we would expect to detect a transcript of approximately 4.45 kb for the *gltA* probe (panel C), and a much smaller ~1.55 kb transcript for the *gltB* probe (panel D). Clearly neither are present, and a common ~6.2 kb hybridisation signal was detected for both the *gltA* and *gltB*-specific probes. This result agrees with the estimated transcript size of 6.24 kb which includes both genes, and the inverted repeat sequence downstream of *gltB*. The co-transcription of the *gltA* and *gltB* genes would ensure that equimolar amounts of the two subunits are synthesized. This is further suggested by the close proximity of the two genes (separated by a 12 bp intergenic region), since it has been observed that if the mRNA molecule is polycistronic and the AUG codon initiating the second polypeptide is not too far away from the stop codon of the first, the 70S ribosome does not always dissociate, but can reform an initiation complex with the second AUG codon (Goodrich *et al.*, 1996).

Several attempts failed to yield less degraded RNA preparations, in spite of procedures to minimize nuclease activity. Since a similar RNA extraction procedure was used for the detection of the distinct 1.4 kb *gltX* transcript (Fig. 3.10), one possibility is that these transcripts inherently have a short half life, making the detection of large transcripts more problematic. This would ensure that regulation at the level of transcription was very sensitive.

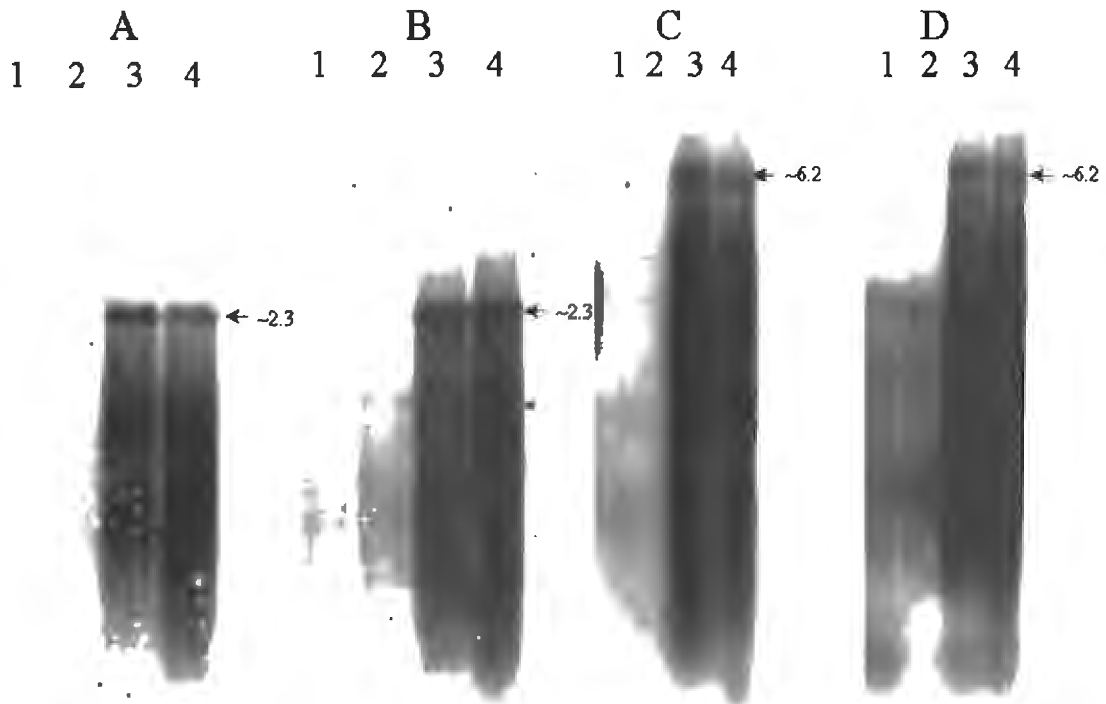


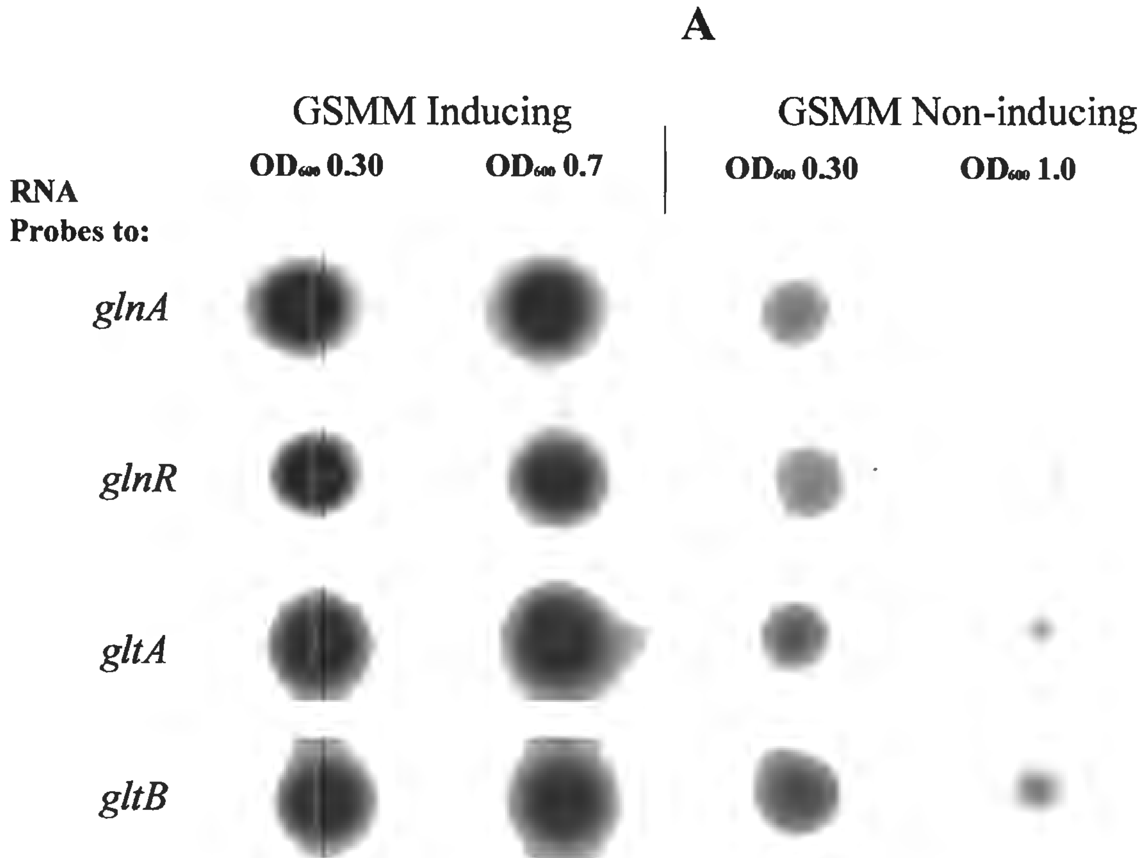
Fig. 5.8 Northern blots showing transcriptional regulation of *C. acetobutylicum glnA* (A), *glnR* (B), *gltA* (C) and *gltB* (D) genes. DIG-labeled probes specific for each of these four genes are described in Section 5.3.6.3. Lanes 1 and 2 contain 30 μg each of RNA extracted from GSMM non-inducing media at OD_{600} 0.30 and 1.0, respectively. Lanes 3 and 4 each contain 30 μg of RNA extracted from GSMM inducing media at OD_{600} 0.30 and 0.7, respectively. Arrows indicate transcript sizes (kb).

5.4.8 Quantitative analysis of *glnA*, *glnR*, *gltA* and *gltB* gene expression

To determine the relative levels of *glnA*, *glnR*, *gltA* and *gltB* gene expression in response to the defined nitrogen limiting and nitrogen rich conditions, the signal intensities generated from Northern dot blots were quantitated (Fig. 5.9). At early exponential phase (OD_{600} 0.30) there was approximately an 8.4 and 7.3 fold difference in the signals generated for the *glnA* and *glnR* genes respectively, between the nitrogen rich and the nitrogen poor growth conditions (illustrated in Fig 5.10). However, under nitrogen limiting conditions the two genes were co-transcribed. This discrepancy may suggest differential expression of the two genes which could not be detected by Northern blot analyses. By comparison, the difference in the expression levels recorded for the *gltA* and *gltB* genes between the two media was very similar (6.0 and 5.5 fold respectively)(Fig. 5.10). This supports their co-transcription under all conditions. By late exponential phase (OD_{600} 0.7) there was a drop in the induced expression levels of *glnA* and *glnR* of approximately 1.6 and 1.7 fold respectively, whereas expression of the *gltA* and *gltB*

genes remained relatively constant. By OD_{600} 1.0 the expression of all four genes decreased further in the non-induced media.

A



B

	GSMM Inducing		GSMM Non-inducing	
RNA probes to:	**OD_{600} 0.30**	**OD_{600} 0.7**	**OD_{600} 0.30**	**OD_{600} 1.0**
glnA	4.02	2.51	0.48	0.18
glnR	5.39	3.22	0.74	0.20
gltA	8.36	8.30	1.40	0.30
gltB	5.28	5.28	0.96	0.23

Fig. 5.9 Quantitative analysis of *glnA*, *glnR*, *gltA* and *gltB* gene expression by dot blot intensity comparison (A) and Gel Trak densitometer analysis (B). Dots represent 5 μg each of RNA extracted from cultures of *C. acetobutylicum* grown in GSMM inducing medium to OD_{600} 0.30 and 0.7, and in GSMM non-inducing media to OD_{600} 0.30 and 1.0, and hybridised with the respective gene-specific DIG-labeled RNA probes. Each probe specific set of data were generated in triplicate from which the average densitometer values were determined. The autoradiographs for some of the data sets were exposed for different lengths of time.

By comparing these results with the GS and GOGAT activity profiles, it was evident that the changes in the *glnA, gltA and gltB* mRNA levels in relation to the nitrogen source was reflected in the corresponding levels of GS and GOGAT activities (Fig. 5.10). This reconfirmed that the *gltA* and *gltB* genes code for the GOGAT enzyme, and led to the conclusion that both GS and GOGAT activity are regulated by the nitrogen source primarily at the level of transcription. There was no indication to suggest that GOGAT was regulated at the post-transcriptional or post-translational level by proteolytic cleavage of a pre-sequence as discussed in Chapter 2. However, these results do suggest that GS activity was regulated post-transcriptionally to some degree. Specifically, at late exponential phase (~OD_{600} 0.7), the drop in the level of GS activity in the inducing media (~3.5 fold) was not reflected in the corresponding levels of RNA (which decreased by 1.6 fold). This is unlikely to reflect feedback regulation since the enzyme preparation is significantly delimited, but may reflect a post-transcriptional regulation by antisense RNA, which would presumably be induced under these conditions. Antisense RNA was shown to be differentially regulated by nitrogen conditions in comparison with the *glnA* mRNA and GS activity (Fierro-Monti *et al.*, 1992).

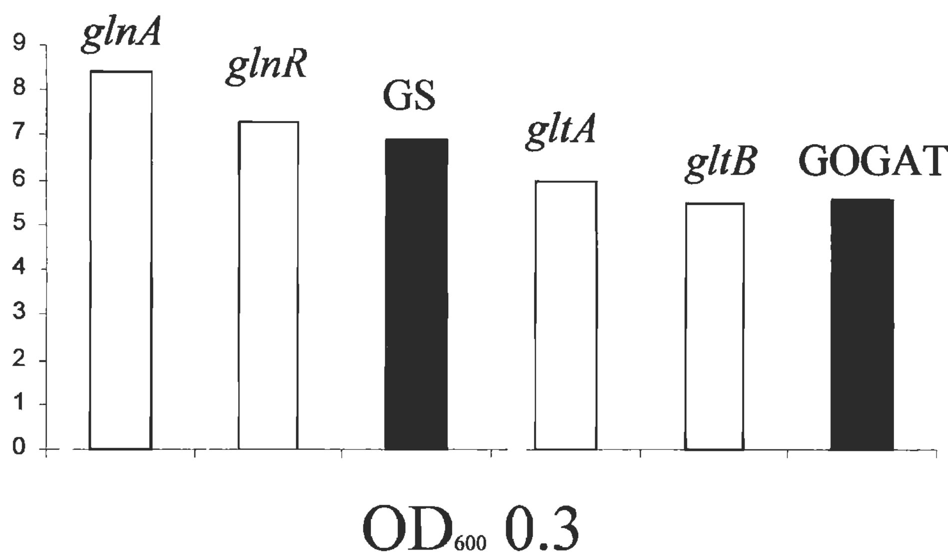


Fig 5.10 Graphic representation of the induction ratios observed for the *glnA*, *glnR*, *gltA* and *gltB* gene expression (open bars), and for GS and GOGAT enzyme activities (shaded bars). They were calculated by dividing the respective RNA intensity signals, or enzyme activities, detected in nitrogen limiting cultures (GSMM inducing media) by that detected in nitrogen rich cultures (GSMM non-inducing media), at early exponential phase (OD_{600} 0.3).

5.4.9 Expression of the *C. acetobutylicum gltX* gene

In Chapter 3 we established that the *gltX* gene (1245 bp) shared striking homology with GOGAT β subunits and was independently transcribed, preferentially at late log phase (OD_{600} 4.5), in CBM grown cultures. We were unable to associate it with GOGAT activity, however, we reasoned that if it was involved in nitrogen metabolism its expression would surely be influenced by the nitrogen growth conditions.

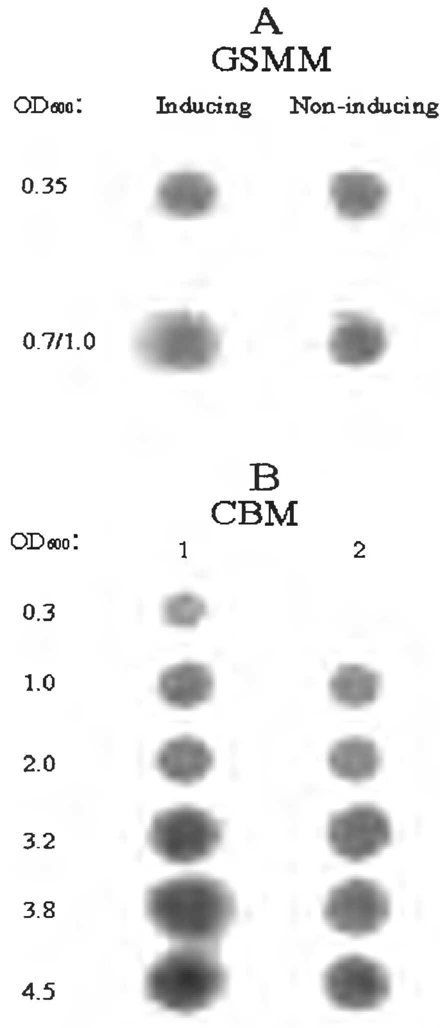


Fig. 5.11 Northern dot blot detection of *C. acetobutylicum gltX* gene expression. Panel A represents 10 µg of RNA per dot extracted from cultures grown in GSMM inducing media to OD_{600} 0.30 and 0.7, and from cultures grown in GSMM non-inducing media to OD_{600} 0.30 and 1.0. Panel B represents RNA extracted from CBM grown cultures at various growth stages as indicated. Lanes 1 and 2 reflect concentrations of 15 and 8 µg of RNA per dot, respectively. The DIG-labelled *gltX*-specific RNA probe used is described in Section 3.3.7.

Northern blots (results not shown) and dot blot experiments (Fig. 5.11) clearly showed that expression of *gltX* was not regulated by the nitrogen conditions. In both the GSMM inducing

and non-inducing media, at early and late exponential phase, *gltX* expression remained at a low constitutive level. The only difference observed was towards late log phase in CBM grown cultures, where there was a gradual increase in the level of *gltX* transcription. This confirmed previous results, and indicated that its product probably plays an enhanced role in stationary phase metabolism.

5.5 Conclusion

Nitrogen metabolism of *C. acetobutylicum* P262 shared features with other Clostridia, but also displayed marked differences. Significantly, organic nitrogen is critical for the germination of *C. acetobutylicum* P262 spores, and is clearly the preferred source of nitrogen for growth and differentiation. However, it remains to be elucidated which amino acids are important at what stage of growth, and this will be the focus of future research. In contrast inorganic nitrogen (ammonia) appeared to retard the efficiency of spore germination and early cell development, a perplexing result that cannot simply attribute to lack of systems for ammonia sensing and/or uptake. The nitrogen status of the cells appeared to be primarily determined by the levels of organic nitrogen available, which was reflected in the activity of GS.

Another significant conclusion was that both GS and GOGAT activities were regulated in a similar way by the same defined nitrogen limiting (0.025% w/v casamino acids + 0.15% w/v glutamine) and nitrogen rich (0.2% w/v casamino acids) conditions. Furthermore, these enzyme activities were regulated by the nitrogen source primarily at the level of transcription, and that under nitrogen-limiting conditions at least, *glnA* and *glnR*, and *gltA* and *gltB* were each transcribed as an operon. These results suggest that the genetically linked loci share a common mechanism of regulation, or at least components thereof, which sense and respond to the level of certain organic nitrogen metabolites. However, these conditions cannot be interpreted as the maximum induction or repression conditions for GOGAT activity, and a more elaborate study of the effect of different nitrogen sources on GOGAT activity should still be undertaken. For example, it will be important to determine whether glutamine has any effect on GOGAT activity. The intracellular glutamine concentration inversely controls GOGAT activity in *B. subtilis* (Deshpande *et al*., 1981).

It was concluded that the *C. acetobutylicum* P262 GOGAT enzyme is sensitive to oxygen, has a broad pH range, and, unlike most other GOGATs characterized, is specific for the cofactor

NADH. Furthermore, since assimilatory GDH activity was undetectable, the GS-GOGAT pathway appears to be the primary route for ammonia assimilation as in *C. acetobutylicum* ATCC 824 (Amine *et al.*, 1990).

Finally, it was concluded that *gltX* is not involved in nitrogen metabolism, but since its expression is increased towards late log growth phase, it may be important during solventogenesis and/or sporulation.

The results presented in Chapters 2 to 5 are consolidated in Chapter 6, and the implications of these findings are discussed in context with a proposed model for the regulation of GS and GOGAT activity in *C. acetobutylicum* P262, and proposed research for the future.

CHAPTER 6

General conclusions

6.1 The *gltA* and *gltB* genes and their deduced protein products 159

6.2 The *gltX* gene and deduced product GltX 162

6.3 Regulation of the GS-GOGAT pathway 163

6.4 Future research 167

CHAPTER 6

General conclusions

6.1 The *gltA* and *gltB* genes and their deduced protein products

Cloning and comparative sequence analysis of the regions flanking the *C. acetobutylicum* P262 *glnA* and *glnR* genes has enabled the characterization of this GS locus in context with its neighbouring genes on the chromosome. A truncated (215 aa) aspartokinase (equivalent to the aspartokinase II isozyme in *B. subtilis*) was located 640 bp upstream, and in the opposite orientation to *glnA*. A short distance (108 bp) downstream from *glnR,* and in the same orientation, lay the *glt* operon encoding the α and β subunits of GOGAT. The gene encoding the α subunit (*gltA*) was separated from the downstream β subunit gene (*gltB*) by 12 bp. A third truncated ORF of 280 residues, encoding an isocitrate dehydrogenase, was located 877 bp downstream from *gltB*.

Proposed GOGAT substrate and cofactor binding domains were conserved in the deduced GltA and GltB protein products. From these comparative analyses it was predicted that the first 20 residues of the α subunit are post-translationally cleaved to expose the conserved Cys-21 as the first amino acid residue of the mature α subunit, and concluded, both from sequence analysis and physiological evidence, that the enzyme is specific for the NADH cofactor. Furthermore, the highly conserved spatial localization of these functional domains in prokaryotic and eukaryotic enzymes, suggests the importance of the structural orientation of the domains, and of the GS-GOGAT pathway in nitrogen metabolism.

The identity of the *gltA* and *gltB* genes was functionally confirmed by complementation studies in which the independent expression of the two subunit genes enabled weak growth of an *E. coli* glutamate auxotroph under anaerobic conditions, using ammonia as the sole source of nitrogen. The significance of these results suggests that functionally mature *C. acetobutylicum* GOGAT enzyme can be produced in this heterologous host from the separate expression of the two subunit genes, and that the enzyme is sensitive to oxygen. To our knowledge this is the first

report in which the genes coding for the two key enzymes of nitrogen assimilation, GS and GOGAT, are tightly linked on the chromosome.

All GOGAT α subunit domains and β subunit domains share a basal identity of approximately 30%, indicating a common origin of the different classes of enzymes. However, close sequence inspection of the pyridine-dependent β subunit domains suggested a division into two subclasses which correlated with the patterns of cysteine clusters proposed to be involved in Fe-S binding, and with the cofactor binding specificity. This prompted a phylogenetic study. Indeed, two main branches of β subunit domains appear to have evolved (designated I and II in Fig. 6.1B) which were consistent with the designated subclasses. Furthermore, these two subclasses are also reflected in the phylogenetic analysis of the α subunit domains (designated I and II in Fig. 6.1A), suggesting that the two subunits have co-evolved within each group of organisms. These results confirm the amino acid homology studies which indicated that the deduced GltA and GltB proteins from *C. acetobutylicum* P262 were more closely related to the corresponding proteins from *B. subtilis* and the cyanobacteria, followed by the fused polypeptides from yeast and plants, than to the remainder of the bacterial proteins.

These phylogenetic divisions within the α and β subunits support the argument that the ability of the *T. ferrooxidans* β subunit to functionally complement the *E. coli* α subunit and restore growth of the *E. coli* glutamate auxotroph on minimal medium containing ammonia, yet the inability of the *C. acetobutylicum* β subunit to do likewise, is due to subunit compatibility and incompatibility, respectively. This also supports the proposal by Vanoni and Curti (1999), that the correct interaction of the two subunits is required for the fermentation of the Fe-S centres 2 and 3, which are involved in communicating the partial reactions that take place at the three active site sub centres.

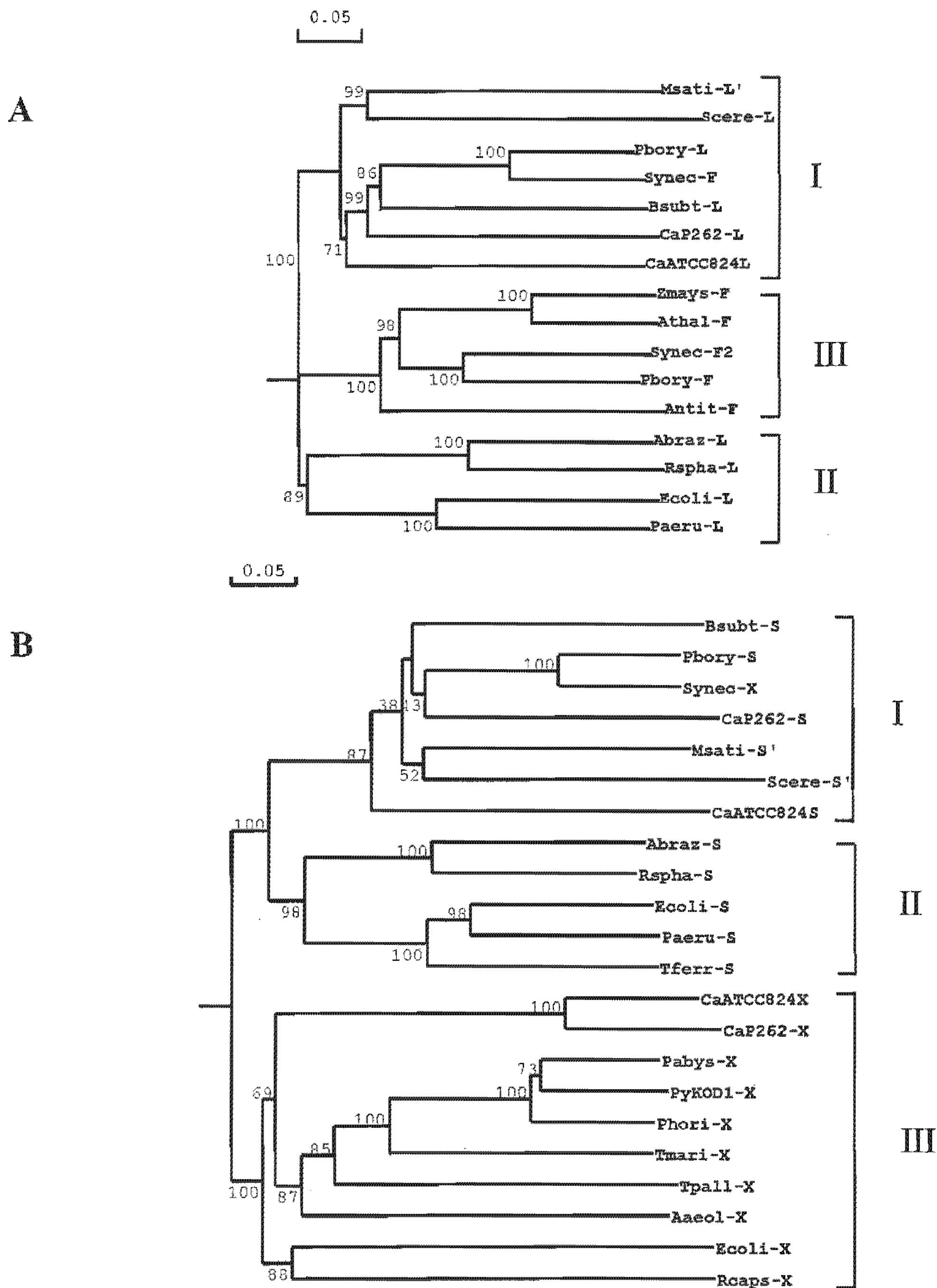


Fig. 6.1 Phylogenetic relationships between the GOGAT α (A) and β (B) (including β-like) subunit domains. A full description of the abbreviated organism names, together with the accession numbers used to generate these trees, can be found in Tables 2.1, 2.2 and 3.2. The different extensions refer to the following classes of proteins: -L, bacterial pyridine-dependent α subunit; -L', eukaryotic pyridine-dependent α subunit domain, -F, ferredoxin dependent; -S, bacterial pyridine-dependent β subunit; -S', eukaryotic β subunit domain; -X, β subunit-like proteins which does not lie adjacent to an α subunit homologue. Phylogenetic analysis was conducted using the DNAMAN optimal alignment function. Branch lengths are proportional to phylogenetic distance. Bootstrap values as a percentage of 1000 trials are included at the branch points. The main branches are indicated by the numerals I, II and III.

6.2 The *gltX* gene and deduced product GltX

The search for genes involved in electron transport in *C. acetobutylicum* resulted in the identification of a second, independently transcribed, β subunit-like gene designated *gltX*. The deduced product, GltX, (~46 kDa) shared striking homology with GOGAT β subunits, and contained all the functional domains characteristic of these polypeptides. In fact, sequence homology searches highlighted the emergence of a number of other genes coding for putative proteins similar to GOGAT β subunits, but which are not associated with a α subunit homologue. However, phylogenetic analysis clearly showed that these β subunit-like genes form an independent evolutionary branch (designated III in Fig. 6.1B). The one exception is the GltD protein from *Synechocystis* which is more homologous to the GOGAT β subunit polypeptides in branch I. In this context, it is also intriguing that *Synechocystis* possesses two Fd-dependent enzymes of which one is most related to the pyridine dependent enzymes in branch I of Fig. 6.1A. These results confirm that GltX is most similar to putative proteins from the hyperthermophilic archaebacteria, *Pyrococcus*.

It has been reported that the β subunit-like protein, GltA, from *Pyrococcus* KOD1, is a functional GOGAT in the absence of a α subunit homologue. In the light of the properties of GOGAT enzymes, this certainly represents a very unique enzyme. The results indicated that GltX was involved in an electron transport mechanism, presumably similar to that of GOGAT β subunits. However, it was concluded that it does not represent an independent GOGAT enzyme similar to the archaebacterial-type GltA protein, nor an alternate functional form of the β subunit, since its expression failed to complement GOGAT activity in β subunit mutants of either *E. coli* or *B. subtilis*, or to complement the *C. acetobutylicum* GOGAT α subunit. In addition, *C. acetobutylicum* CFE was incapable of oxidizing NADPH (for which GltX contains the consensus binding domain) in the GOGAT assay, further suggesting that glutamine and 2-oxoglutarate are not its substrates. Finally, its expression was not regulated by the nitrogen source. It will be interesting to see if GOGAT activity can be associated with any other β subunit-like proteins.

The presence of the GOGAT β subunit domain within a diverse group of much larger proteins, clearly quite different from GOGAT, was also highlighted. Thus, it was concluded that GltX represents a member of a much more general family of FAD-dependent NAD(P)H oxidoreductases, that has been proposed by Vanoni and Curti (1999), in which the β subunit

domain serves to mediate electron transfer between a reduced pyridine nucleotide and a second protein, or protein domain. Although the function of GltX remains unclear, the finding that its expression was induced during stationary phase growth, and that it is highly conserved in both *C. beijerinckii* and *C. acetobutylicum* ATCC 824, has led to the suggestion that it may be significant during stationary phase metabolism in the Clostridia.

6.3 Regulation of the GS-GOGAT pathway

From the physiological studies it was concluded that organic nitrogen is critical for the germination of *C. acetobutylicum* P262 spores, and is the preferred source of nitrogen for growth and differentiation - a result which was deduced to depend on the presence of specific amino acids rather than on the total organic nitrogen content. In contrast, ammonia appeared to delay spore germination and early cell development.

Regulatory studies showed that both the GS and GOGAT enzyme activities, and the expression of their corresponding genes (*glnA* and *gltA* and *gltB,* respectively), were regulated in a similar way, being induced by the defined nitrogen limiting conditions (0.025% casamino acids and 0.15% glutamine) and repressed by the selected nitrogen rich conditions (0.2% casamino acids). Under nitrogen limiting conditions, *glnA* and *glnR* were transcribed as an operon, and *gltA* and *gltB* were transcribed as an operon. By comparing the repression ratios of enzyme activities and corresponding mRNA levels in the different nitrogen media, it was concluded that these enzymes are regulated by the nitrogen source, primarily at the level of transcription, and that feedback inhibition was not likely to play a significant role. These results, in conjunction with the basic molecular analysis of the *glt* gene cluster which revealed 1) the lack of a complex regulatory region upstream of the *gltAB* operon, 2) its close proximity to the *glnR* gene (108 bp), itself a putative transcriptional regulator of *glnA*, 3) the identification of a putative *gltA* promoter overlapping the carboxy-terminal region of the *glnR* gene, and 4) the presence of an inverted repeat sequence with the potential to form a transcriptional terminator between the putative *gltA* promoter and structural gene, suggested that the GS and GOGAT enzymes share some common regulatory mechanism. This allowed for the development of hypotheses about nitrogen regulation in this Gram-positive organism, despite the limited evidence available.

Based on these results, a simple working model was proposed that is based on two nutritional conditions (Fig. 6.2). However, it is possible that other conditions might lead to a different relationship between *glnA/glnR* and *gltAB.* As discussed in Chapter 1, it has been proposed that the *glnR* product, GlnR, represents a response regulator component and may positively control *glnA* transcription via an anti-termination mechanism by binding to the putative termination structure present between promoter 1 and the *glnA* initiation codon (Woods and Reid, 1995), however the mechanism is uncertain. Anti-termination as a mechanism of regulation of this nitrogen assimilatory locus is also supported by the fact that the *glnA* and *glnR* genes are transcribed together under nitrogen limiting conditions, implying that, under these conditions, the inverted repeat sequence identified downstream of *glnA* (Janssen *et al.*, 1988) does not function as a transcriptional terminator. The most obvious explanation is that GlnR regulates transcriptional read through of this terminator-like structure, thus positively regulating its own expression during nitrogen limitation. Furthermore, the position of a putative transcription terminator between a likely *gltA* promoter sequence and the *gltA* structural gene, suggested that, in addition to acting as a transcriptional terminator for *glnR*, it may influence the expression of the *gltAB* operon. Thus, in the absence of further evidence, it is tentatively proposed that GlnR may also acts as a transcriptional anti-terminator for the expression of the *gltAB* operon. It is envisaged that under nitrogen limiting conditions, a signal transduction mechanism activates residual GlnR protein to promote the transcriptional induction of the *glnAR* and *gltAB* operons, by allowing RNA polymerase to read through the terminator-like structures. On the other hand, quantitative analysis suggested a low level of differential *glnA* and *glnR* mRNA expression, which would presumably ensure that sufficient *GlnR* protein was always available should nitrogen become limiting.

Under nitrogen rich conditions, *glnA* anti-sense RNA (AS-RNA) has been implicated in the negative regulation of GS expression by binding to a 43 bp complementary sequence spanning the Shine-Dalgarno and start codons of the *glnA* mRNA (Fierro-Monti *et al.*, 1992). In addition, since it is transcribed off the opposite DNA strand in the 5' start region of the *glnR* gene, it would reduce the production of GlnR. This is consistent with the model in which expression of AS-RNA would be preferentially achieved under nitrogen rich conditions, when *glnA* and *glnR* expression is repressed. The involvement of a post-transcriptional regulatory system is also compatible with the results that, at late exponential phase, the decrease in GS activity (3.5 fold) in nitrogen limiting media was not reflected in the decrease in mRNA levels (1.6 fold). Possibly, at this growth stage, conditions are starting to favour the expression of AS-RNA

resulting in the inhibition of *glnA* and *glnR* mRNA translation. In any case, it was concluded that AS-RNA only represents a fine tuning system of regulation. Although the AS-RNA also showed complementarity to a 20 bp region spanning the Shine-Dalgarno and ATG start codon of *gltA* (Fig. 6.3), no comment can be made on its possible involvement in the expression of the *gltAB* operon.

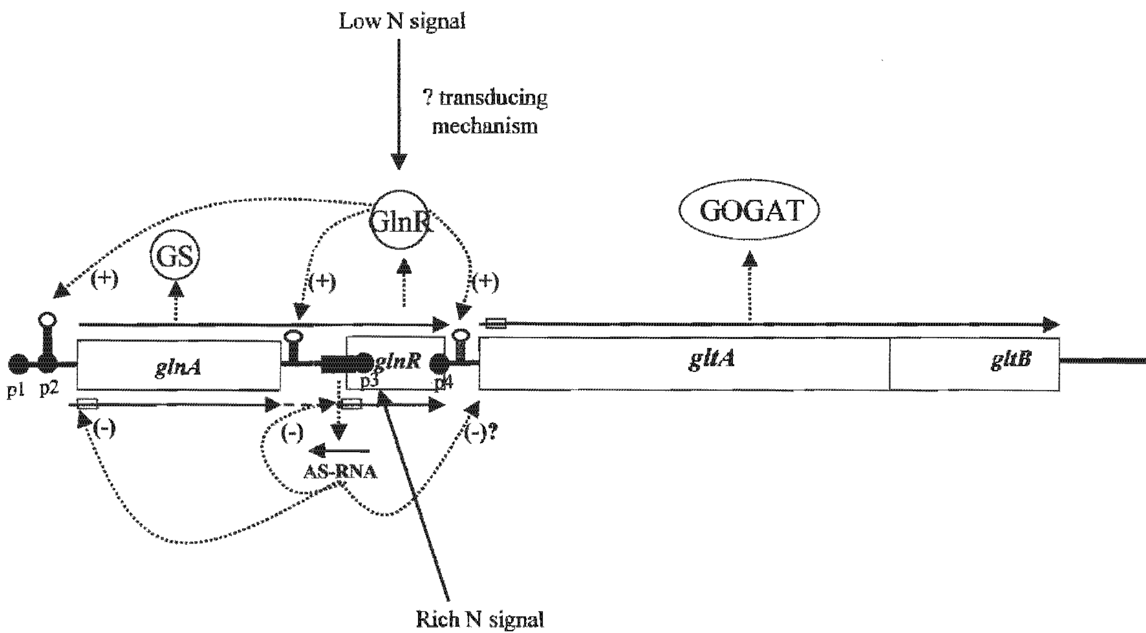


Fig. 6.2 Hypothetical model for the regulation of the *C. acetobutylicum* P262 *glnA, glnR, gltA* and *gltB* genes by nitrogen content of the medium (N). The two *glnA* promoter sequences are indicated by P1 and P2 The putative *gltAB* promoter is indicated by P4. Promoter P3 is situated in the 5' region of the *glnR* gene, and directs the synthesis of the antisense RNA (AS-RNA) which has the potential to bind the 5' regions (indicated with open boxes) of the *glnA, glnR* (and possibly *gltA*) mRNA transcripts, reducing expression of the respective genes in response to a nitrogen rich signal. (The respective transcript sizes have not yet been determined under nitrogen rich conditions). GlnR may act as a transcriptional anti-terminator by interacting with the terminator-like structure, in response to a low nitrogen signal transduction mechanism, thereby inducing expression of the *glnAR* and *gltAB* operons. This diagram is not drawn to scale.

Fig. 6.3. Hypothetical RNA-RNA hybrid formation between *gltA* mRNA and anti-sense RNA, blocking the putative Shine-Dalgarno sequence (SD) and the AUG start codon.

More work needs to be done to understand the preference for organic nitrogen, and explain the poor growth with ammonia. These studies showed that the presence of glutamine specifically induced *C. acetobutylicum* P262 GS activity. However, this effect was not due to glutamine mediated enzyme activation, and therefore probably acts at the transcriptional level. Thus glutamine, or a metabolite derived from it (e.g. glutamate), may represent one of the regulatory metabolites signalling nitrogen conditions, and the regulatory activity of other amino acids may partly be related to the rate at which they contribute to the intracellular glutamine (or glutamate) pool. It is conceivable that genes encoding amino acid transport and degradative pathways are co-ordinately regulated with the GS and GOGAT operons via a sensory transduction system which may cause the phosphorylation of positively acting transcriptional regulators such as GlnR. In *B. subtilis*, mutations in the CcpA gene affect the expression of GOGAT and catabolic operons (Faires *et al.*, 1999). Interestingly, a CcpA-like gene, *regA*, which is able to partly complement a *B. subtilis ccpA* mutant (Davison *et al.*, 1995), has been identified and sequenced from *C. acetobutylicum* P262. Also, in *R. meliloti*, disruption of the *glt* locus appeared to be linked to the degradation of certain amino acids (Lewis *et al.*, 1990). This mode of nitrogen regulation may be independent of regulation by inorganic nitrogen sources. However, it cannot be ruled out that intracellular ammonia levels may be a regulatory factor, but that ammonia may not be effectively transported into the cells.

Finally, the failure to detect assimilatory GDH activity, leads to the conclusion that the GS-GOGAT pathway is the primary route for ammonia assimilation. The arrangement of the genes, and the regulatory features observed for this pathway, are distinct from the Ntr system or any other bacterial system so far described, implying that *C. acetobutylicum* uses a novel mechanism of nitrogen regulation. This may reflect a more primitive origin, and the apparent co-ordinated regulation of GS and GOGAT, at least under these conditions, seems to suggest less metabolic flexibility. It remains to be seen whether this proposed model is unique to *C. acetobutylicum* strain P262.

6.4 Future research

These studies provoke a lot of questions for future research. Ultimately it would be nice to decipher the signal transduction system which communicates the cellular nitrogen status to the relevant transcriptional machinery. The isolation and characterization of *C. acetobutylicum* mutants modified in GS and/or GOGAT activity by gene disruption techniques would be expected to identify factors involved in their regulation. However, a major stumbling block in these, and related studies, is the lack of a transformation and plasmid integration system for *C. acetobutylicum* P262. This is an area of research which needs to be developed. However, since the closely related *C. beijerinckii* strain is relatively amenable to genetic manipulation (Wilkinson and Young, 1994), it may provide a suitable alternative host for functional studies.

Of primary interest is to establish the pattern of amino acid utilization, to determine the regulation of amino acid catabolic enzymes by the nitrogen source, and find out whether a relationship exists between amino acid catabolic pathways and GS and GOGAT activity. A modified control of catabolic enzymes would be expected in a GS or GOGAT mutant if a regulatory relationship exists with these enzymes. As evident from Chapter 1, there is no unifying control of amino acid catabolic pathways in the Gram-positive bacteria. In *B. subtilis,* the synthesis of amino acid degradative enzyme is generally not nitrogen regulated, although a hierarchy for amino acid utilization is present in cells growing with mixtures of amino acids (Fisher, 1999). Specifically, the importance of glutamine as a regulator of nitrogen metabolism should be investigated, and whether it is also involved in the regulation of GOGAT activity.

To support the regulatory model, it is important to investigate phosphorylation of GlnR, and demonstrate its binding to the proposed regions of dyad-symmetry upstream of the *glnA* and

gltA genes by gel shift experiments. The determination of the *gltA* transcription initiation sites would also help put the proposed transcriptional regulatory features in perspective. In addition, the development of a promoter probe system to quantitate the *glnA* and *gltA* promoter activities in relation to different nitrogen growth conditions would be useful. Deletion derivatives would enable the definition of the promoter regions more specifically, which could then serve as a foundation for base substitutions or deletions to determine the critical features of the ammonia assimilatory promoters in this organism.

In addition, it would be interesting to establish the effect of different carbon sources on the levels of GS and GOGAT enzymes, to determine whether carbon catabolic mechanisms are involved in their regulation. In this respect it is intended to investigate the role of the *C. acetobutylicum* P262 CcpA homologue, *regA*. Conversely, it will be interesting to know if the isocitrate dehydrogenase gene is regulated by the nitrogen conditions, as reported in the cyanobacteria (Muro-Pastor *et al*., 1996).

Finally, the present study has provided a better understanding of this Gram-positive solvent producing bacterium, and enabled the first steps to be made towards elucidating the mechanism of ammonia assimilation. These studies should provide a sound base from which to design future experiments aimed at better understanding the physiology of *C. acetobutylicum* P262, and may help towards more effective and economic control of the solvent fermentation process.

APPENDIX A

E. coli strains, genotypes and references

Strain	Genotype	Reference
JM105	*thi pps endA sbcB*15 *hspR*4 Δ(*lac-proAB*)[*F' traD*36 *proAB lacI*^{q}*Z*Δ*M*15]	Yanisch-Perron *et al.*, 1985
JM109	*recA*1 *endA*1 *gyrA*96 *thi hsdR*17 *supE*44 *relA*1 λ-Δ(*lac-proAB*)[*F' traD*36 *proAB lacI*^{q}*Z*Δ*M*15]	Yanisch-Perron *et al.*, 1985
LK111	*thr*1 *leuB6 thi*1 *supE*44 *tonA*21 r_k-m_k^+ *lacY*$^+$ *lacI*^{q}*Z*Δ*M*15	Zabeau and Stanley, 1982
CC118	*araD*139 Δ(*ara, leu*)7697 Δ*lac*X74 *phoA*Δ20 *galE galK thi rpsE rpoB argE*$_{am}$ *recA*1	Manoil and Beckwith, 1985
F19	*ntr phoA*Δ20, otherwise as CC118	Santangelo *et al.*, 1991
MX3004	*thi*-1 *gdh*1 *pro*$^+$Δ(*lac*U169)*hut*C *gltD*227::MudIIPR13	Castano *et al.*, 1992
AN3001	*gltB*31 *gdh*::Mud	Donald *et al.*, 1988

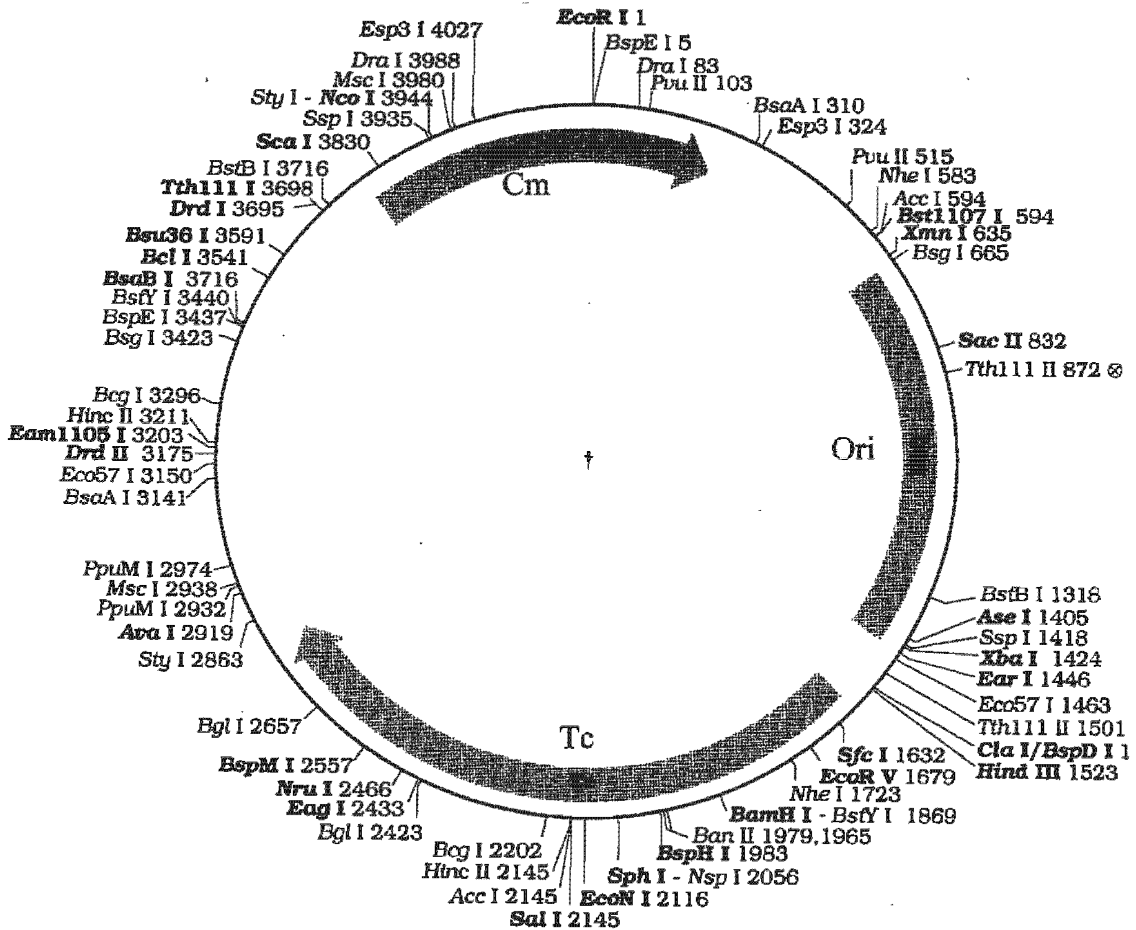
APPENDIX B

Plasmid vectors

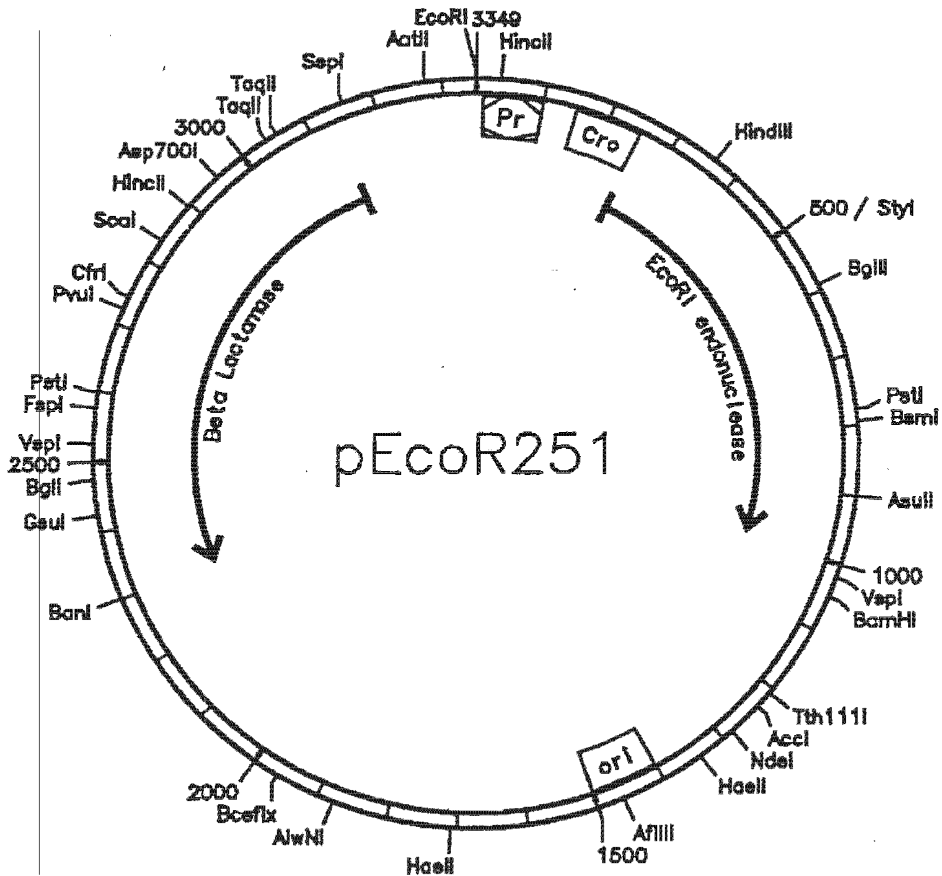
```
      ApaI
KpnI           XhoI             ClaI          EcoRV
GGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTC
   DraII             SalI          HindIII        EcoRI
                     AccI
                     HincII

 PstI            BamHI         XbaI    EagI        SacII    SacI
CTGCAGCCCGGGGGATCCACTAGTTCTAGAGCGGCCGCCACCGCGGTGGAGCTC
        SmaI           SpeI          NotI
                                                BstXI
```

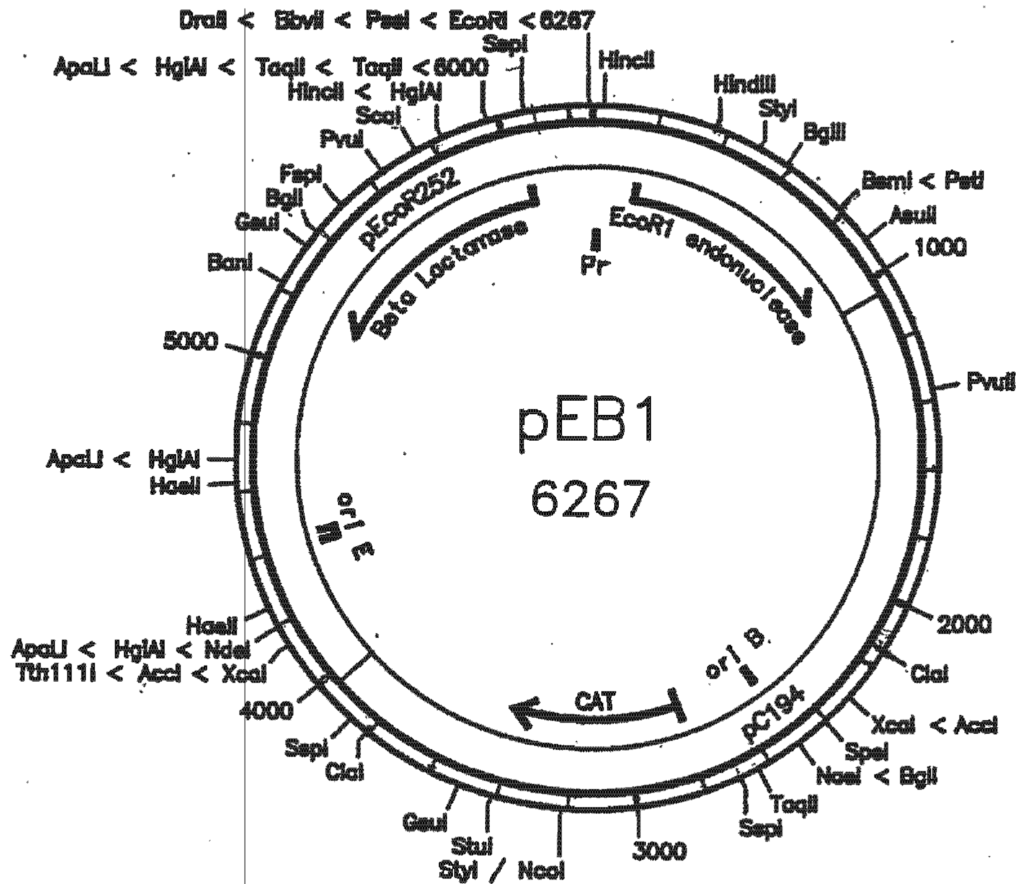
BlueScript SK (2959 bp; Stratagene, San Diego, CA). This high copy number *E. coli* cloning vector is specifically designed for exonuclease III shortening. The nucleotide sequence of the multiple cloning site polylinker (in the *lacZ* gene) is shown below the circular plasmid map. BlueScript KS differs from BlueScript SK in the orientation of the multiple cloning site polylinker.



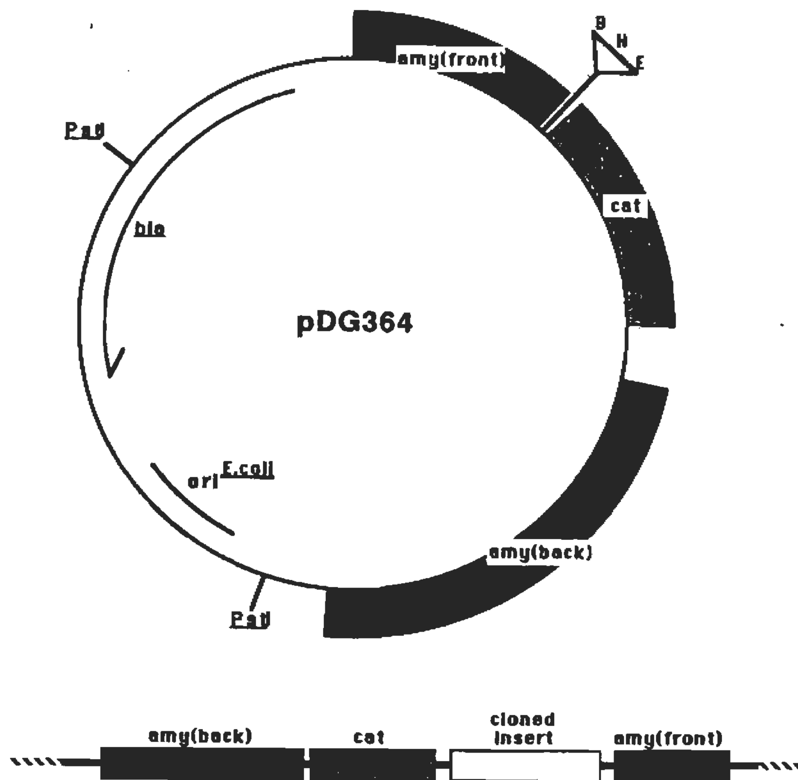
pACYC184 (4244 bp; Rose, 1988). This is a small low copy-number *E. coli* cloning vector. It carries the origin of replication (Ori) from plasmid p15A, which enables it to co-exist with vectors that carry the ColE1 origin of replication e.g. pEcoR251. It carries the chloramphenicol-resistance gene (Cm) from Tn9, and the tetracycline-resistance gene (Tc) from pSC101. Unique restriction sites are shown in bold type.



pEcoR251 (3349 bp; Zabeau and Stanley, 1982). This is a relatively low copy number positive selection vector, used routinely by insertional inactivation of the *Eco*RI endonuclease gene by cloning into either the *Bgl*II or *Hind*III site.



pEB1 (6267 bp; Anaerobe laboratory, University of Cape Town) is an *E. coli*/*B. subtilis* shuttle plasmid derived from the fusion of the positive selection vector pEcoR252 (identical to pEcoR251 except that it does not contain the second *Pst*I site in the beta lactamase gene), and pC194. It contains the chloramphenicol resistance marker (CAT) for selection in *B. subtilis*.



pDG364 (6257 bp; Harwood and Cutting, 1990). This plasmid was designed to integrate into the *amyE* locus of *B. subtilis*. It contains three unique restriction sites (E, *Eco*RI; H, *Hind*III; B, *Bam*HI) for cloning DNA fragments of interest, an ampicillin resistance gene (*bla*) and an *E. coli* origin of replication (ori) for propagation in *E. coli*, *Pst*I sites for linearization, and the *cat* gene for selection (Cm^R) in *B. subtilis*. Disruption of the *amy* gene will result if *cat* (and the adjacent cloned DNA) has integrated by a double-crossover recombination event (marker replacement) between the *amy*E locus and the homologous regions of DNA in *amy* (front) and *amy* (back). The resulting structure of an $AmyE^-$, Cm^R transformant is illustrated below the circular map.

APPENDIX C

```
Pbor-L    ................................................................       0
Syne-F    ...................................................MPCHEGLHPLVPN      13
Bsub-L    ................................................................       0
Ca262-L   ................................................................       0
Msat-L'   MSNSLSLTFTALNNPQINAISNPNARLRPLARVTRCSATCVERKRWLGTKLRSGGGLERIQLWESGGLGRLP      72
Scer-L'   ................................................MPVLKSDNFDPLEEAYEGGTIQNY      24
Ca824-L   ................................................................       0
Abra-L    .........................................................MTTELNQ       7
Zmay-F    ....MATLPRAAPPTPAALLPLPRAAPPLLLAGRAAAARRSRLRARGPSAAARRSWVVASAASSSSRAVVGG      68
Ecol-L    ...................................................MTRKPRRHALSVP      13

Pbor-L    ......MIMTRYGLPAKQGIDEQFWKADEVEFIVQMKRQPWHSIDQQALTIDANLEHPEACGAETNWGIG      66
Syne-F    FCTVTSPMNSSHLAPQVQGIDNQNWKAEGVEFIVQMKRKVWHDIMEQGLQMLVNLEHPEACGCEPNWGIG      85
Bsub-L    ........MTYNQMPKAQGIRVEFWKADGIDLYAHLKKKQTHDIMKQGLKMLCQLDHPEGQGSDPDWGIG      64
Ca262-L   .........MENNIPNAQGIDESFWKAEGIDTIVNIDKEKWHEIMSDGLTILEKLEHPEGTGADENWGIG      63
Msat-L'   KLRVAVKSSFSAVPDKPMGIDDAFWKKSEGVEFVAELNKQSWRKTWTDALEMLVRHTHPEACGCEANWGIG     144
Scer-L'   NDEHHLHKSWANVIPDKRGIRDYFWKADGVEFVANKHMEQWHKIMTDARYLLVNHTHPEAVSSDGM.GIG      95
Ca824-L   .......MTRNIGYPEKQGIDAAFWKNEGIEFITSIKKEKWHDIMKKGIEILVNLTHPEAVGADTKWGIG      65
Abra-L    GEQFVADFRANAAALTTANAINEDQWKADGIEFIAAIDRKPRRSVWEKGIEALKAVWHPEAVDADGKWGIG      79
Zmay-F    VARREAPPAPQKPTQQAADINHILSWKADGVEFVANLKNESWFDIMRDALMALGCHEHPEGCGADSDSGIG     140
Ecol-L    VRSGSEVGFPQSLGEVHDMNKDKSLWRENGEFELIAHIEWEPWHKVWRTAIHALARHQHPEAILADGKWGIG      85

Pbor-L    GILMQVFEGWFQKVT.ATLNWTLPDA.GQYGWLEVNSSPD.RAQREAG...RRIFEQVVAEEWQKVLEWED     132
Syne-F    GILIQVFEKWIQKIAGAE.GWTIP.APGQYAWLNIGSPDPLARAEA....RQKFNDIVAQEALKVLEWED     151
Bsub-L    GLLVQIFDAWFRKEC...KNWNLLEK.ERYGWLRWFSQKEDERKKIE...KQINALIEQ.EWQVVLEWET     128
Ca262-L   GILFNIFEKWFKEELQS.KGWTLEEE.GDYKWLELLPQDEKARKEA....VSLFEDISKEEWLELIEWEE     129
Msat-L'   GILVALFEGWYQ....EVVDWQLFPQ.GNTAWLNFLPKSDSRRKE....SKNIFTKVAESLWHKVLEWES     207
Scer-L'   GILLGIFHEWMKREFKLDLDWDLPEM.GKYAWLNVLFKKNEKNNKKNLIKCQKIFEDLAASFNLSVLEWE.     165
Ca824-L   GIHLQIFDEWFRINCDN.LGWELWES.GKYLWLNWLPKETALSYQCEG....ILERAVEEQWQKVLEWEV     131
Abra-L    GIHVAVFQKWFKDHVKV..IGHRLPDNK.LWWQQILPRISLDAQEACR...CIVETEILAFWYYIYEWEQ     145
Zmay-F    GLHSAVFWDWVDDWASK..QGLALFDRRNTGWLEWLP.QDEKSMEEAK...AATEKVFVDEWLEVLEWEP     206
Ecol-L    CGLLLQKFDRWFRIVAQE..RGWRLAKNY..AWLLLN.KDPELAAAAR...RIVEEELQRETLSIVEWED     149

Pbor-L    WHNSTLGETWKASEEFHQEVFIQRSADLAD....DLAFLRKIWVIR..KRSHSAIRQ...TGIDPFWWPA     195
Syne-F    WTQNEPWGETWIASEEFHQEVWIARPEGLTD....DLDFLRKIWVIR..KLTHGAIRS...PKIDTYWWVA     214
Bsub-L    WVNVGKWGTVWQKECRFVRRVFIGASSDLKD....NLSFVRKIWVIR..KQAENWG.....VTEGLDWWFA     189
Ca262-L   WQWPSIWGKASLEAMPYIHEAFVKRPNGTKK....EKDFLRKIWIVERNIEKRAAWIS...KFLNETWWIA     194
Msat-L'   WIENTGLGKSWQLTWPVIEEVFLTPSSDSK......VDLLRQIWILEKLSHVSITSALNLQSDGITDWWIG     273
Scer-L'   TSRRFYYLGDVWLSRWPTWLEPELVPLYDEKQPEFNETKFRTQIWLLEKEASLQIGLEN........WWVG     229
Ca824-L   WRENRSWGETWKGEEEVWKEIFIGSNCENQT......DFLRNIWIIEKRAESEVKRLVER...GSEYWWIG     194
Abra-L    WIHVDIWGEKWNATRHELEEIWVGNNKGVSDE.....QFWLDIWIIERRIEKAVKGE......QINDWWIG     206
Zmay-F    WFEVSVWGRNWKETMRNIQEIWVKVAKEDNAD.....DIWREIWISEKLIERAAKSFS.....WADEWWFG     268
Ecol-L    WGEGVLWGEIWLSGLERIEEIFVNAPAGWRPR.....DMWRRIWIAERRIEKRLEAD........KDWWVG     208
```

```
Pbor-L     S  CRTLVYKGMMMPLVKE.YYPLSDPDLQALGVHSFFSTNTFPSWERSHPYRYLAHNGEINTLRGNI    266
Syne-F     S  ARTLVYKGMTTALVGQ.YYPLHDPDMEALAVHSFFSTNTFPSWERSHPYRYLAHNGEINTMRGNV    285
Bsub-L     S  SQTIVYKGLTPELVDA.FYSLQDEAFVAFLVHSFFSTNTFPTWERAHPNRYVHNGEINTLRGNI    260
Ca262-L    S  SKTIVYKGMLSTLLRV.FYKLEDERVETSLAVHSFYSTNTFPSWERAHPNRFVHNGEINTLRGNV    265
Msat-L'    S  SRTVIYKGQTPALLGEYYYALGNERFTYMALIHSRFSTNTFPSWDRAQPFRVLGHNGEINTLRGNV    345
Scer-L'    S NNTTIVYKGQTPALVYN.YYPLTNAHFKHMLVHSRFSTNTFPSWDRAQPLRWLAHNGEINTLRGNK    300
Ca824-L    S  SRTIVYKGLLADLIKS.FYMLNDINFKAIAVHQRYSTNTFPTWDLAQPFRFLAHNGEINTIRGNR    265
Abra-L     S  ARSIIYKGMLAELLTT.FYPLLDERFEDFLYHQRYSTNTFPTWPLAQPFRMLAHNGEINTVKGNV    277
Zmay-F     S  SRTIVYKGMRSEVLGQ.FYLLQNELYKPFLYHRRFSTNTSFRWPLAQPMRLGHNGEINTIQGNL    339
Ecol-L     S  NLVNIYKGLCMPTLLPR.FYLLADLRLEAICFHQRFSTNTVPRWPLAQPFRYLAHNGEINTITGNR    279

Pbor-L     WHHAQQ...SLFESDLFGDDLRKAQIVINIDGSDSTIFDNALELTLSG.RSLPHAVMHIPPWTAH...   331
Syne-F     WHQAQQ...ALFESSLFGEDHAKVQIVINIDGSDSTIFDNALELYLAG.RSLPHAVMHIPPWSAH...   350
Bsub-L     WHRAAE...QQFVSESFGEDLNKILPILNADGSDSSILDNAFEFVMAG.RKPAHTAMLIPPWTEN...   325
Ca262-L    MKVSR...........VLGKDVIRVLPIINKEGSDSAIFDNNLEFYMNG.MDLPRAVMAIPPWYKS...  322
Msat-L'    WIKAEGLLKCKELGLSENDLKKFLPIVDANSSDSGCFDGVLEFLLHSG.KSLPEAVMMIPPAWQND...  413
Scer-L'    WMRSREGVMNSATFKD...ELDKLYPIIEEGGSDSAALDNVLELTINGTLSLPEAVMMVPPAYHKD...  366
Ca824-L    WHNAQEG...VLKSDVFGKKISDLFPIVNPKGSDSTSLDNTFELVADG.RPLAQAVMMLIPPAWENN... 330
Abra-L     WHKAQE...TRMEHPAFGTHMQDLKPVIGVGLSDSGSLDTVFEVVRAG.RTAPMVKMMLVPPALTSSQTT 345
Zmay-F     WMRSRE...TTLKSPVWRGREHEICPFGDPKASDSANLDSTAELLRSG.RSPAEALMMLVPPAYKNHPTL 407
Ecol-L     QWARAQ....TYKFQTPLIPDLHDAAPFVNETGSDSSSLDNMLELLLAGG.MDIIRAMRLLVPPAWQ...NN 343

Pbor-L     .ESPSPERKAFYKYHSCLMEPWDGPASIAFTDGRMMGAVLDRNGLRPSRYYVTKDDLVIMASEAGVLPIEPE 402
Syne-F     .ESPSQEKKAFYKYHSCLMEPWDGPASIAFTDGKMMGAVLDRNGLRPSRYYVTKDDLVIMASEAGVLPIEPE 421
Bsub-L     .THMSKEKRAFYEYHSSLMEPWDGPTAISFTDGKQIGAILDRNGLRPARYYVTKDDYIIFSSEVGVIEVEQE 396
Ca262-L    .KTLSKEKRDFYENATLMEPWDGPAAIVFTDGERVGAVLDRNGLRPSRYYITKRRLILSSEVGALDVPAE  393
Msat-L'    .KNMDPQRKAFYEYYSALMEPWDGPALISFTDGHYLGATLDRNGLRPGRFYVTHSGRVIMASEVGVVDIPPE 484
Scer-L'    ...DSDLKAMYDYAACLMEPWDGPALLTFTDGRYCGAILDRNGLRPCRYYITSDDRVICASEVGVIPIENS 435
Ca824-L    .ESLETWKRAFYEYQGTLMEPWDGPAVAFTDGQQVGAVLDRNGLRPARYLITKDNIAVLASEAGVLKFEPE 401
Abra-L     P.DN...HKAFIQYCNSVMEPWDGPALAMFSDGRWVVGALDRNGLRPMRYTITTDGLIIGGSETGMVKIDET 413
Zmay-F     S.IKYPEVTDFYDYYKGQMEPWDGPALLLFSDGRTVGATLDRNGLRPARYWRTSDDFVYVASEVGVIPMDES 478
Ecol-L     P.DMDPELRAFYDYNSMHMEPWDGPAGIVMSDGRFAACNLDRNGLRPARYVITKDKLITCASEVGIWDYQPD 414

Pbor-L     RVAQKGRLQPGRMFLVDMEQGRIVADEEIKHQIASAEPYQDWIDQHMVSLADLKDAP...PSSSDLNTV...  468
Syne-F     RIAKKGRLQPGRMFLVDMEQGRIIADEEIKQEIVSQHPYGEWLAANLKSLEQLP.SPGNVPG.TDAES....  487
Bsub-L     NVLYKNEGRLEPGKMLLIDLEEGRIISDEEVKTQIATEYPYQKWLEEELVQVNPDPESRE.....EEQFSD... 460
Ca262-L    IVEKKGRLMPGRMLLVDTVKKEVVSDEEVLDTYAKENPYGEWLEQNLVTLDKMTDSKT..FTIEYDKDT...  460
Msat-L'    DVCRKGRLNPGMMLLVDFEKQIVVNDDALKEQYSLARPYGDWLEKQKIELKDIIDSVHESDIVPPTISGVPP  556
Scer-L'    LVVQKGRLKPGDIIPSDTQLGEMVDTKKLKSQISKRQDYKSWLS..KVIKLDDLLSKTANLVPKEFISQDSL 505
Ca824-L    EVAYKGRLKPGKMFLIDTKEGRIIDDEEIKKSICLDKEPEKIVEKNKFTLDDFEATVGEEVVNDEVLKEK..  471
Abra-L     QVIEKGRLGPGEMIAVDLQSGELYRDRELKDHLATLKPWDKWVQNTTHLDELVKTASLKGEP.SDMDKAELR 484
Zmay-F     KVVMKGRLGPGMMITVDLQTGQVLENTEVKKTVASASPYGTWLQECTRLIKPVNFLSS.....TIMDNETVL 545
Ecol-L     EVVEKGRLGPGEMMVIDTRSGRILHSAETDDDLKSRHPYKEWMEKNVRRLVPFEDLPDEEVGSRELDDDTLA 486

Pbor-L     ................L.QRQLAFGYTFEELRLLITPMARDGVEAVGSMGNDAPLAVLSDRPKLLYDYFKQLF 524
Syne-F     ................LRQRQMAFGYTFEELRILLAPMGRDGVEAIGSMGNDAPLAVLSDKPKLLYNYFQQLF 544
Bsub-L     ................LLTRQKAFGYTYEDIQKYLIPIIKEGKDPLGSMGNDAPLAVLSDRAQSLFNYFKQLF 517
Ca262-L    ................RKRLEKTFGYTYEEVNTTMLPMAKTGAEPLGSMGVDTPLAVLSKQSQPLFNYFKQLF 517
Msat-L'    LSNDDVDMENMGIQGLLAPLKAFGYSVESLEILLLPMAKDGVEALGSMGNDTPLAVMSNREKLTFEYFKQMF  628
Scer-L'    LKVQSDPR..........LLGNDFGYTFEQVTFLLTPMALTGKEALGSMGNDAPLACLNENPVLLYDYFRQLF 567
Ca824-L    ..................QQAYGYTLEDLRVILGPMAATGKEPLGSMGNDAPLAVLNKSQLLFAYFKQLF  524
Abra-L     ..................RRQQAYGLTMEDHELILHPMVEDGKEAIGSMGDDSPLAILSDKYRGHHRYFRNF 539
Zmay-F     ..................RHQQAFGYSSEDVQMVIESMASQGKEPTFCMGDDIPLAVLSQRPHLLYDYFKQRF 600
Ecol-L     ..................SYQKQFNYSAEELDSVIRVLGENGQEAVGSMGDDTPFAVLSSQPRIIYDYFRQQF 541
```

```
Pbor-L      QVTNP I SI  EI   AETTI A. EK L KPLPESCHLIE      SNEDLAKLK..QVDEQGFKSTTIA     593
Syne-F      QVTNP I S   EI   AEFTI G.     DPRPKSCRLIE       TNEDLAKLK..ALDDDEFKSVTLD     613
Bsub-L      QVTNP I   FEQL   TMFWL A.     HPSERNVRRI  YT    SNEQFYALK..TIVHPDLKSQKID     586
Ca262-L     QVTNP I    EI   VLMFILGP     EDKSSNCELIK D     NNEELAKIK..EYNKENLKPKVID     587
Msat-L'     QVTNP I    KI   RCM  P.     TETTEEQCHRLS      STKEMEAIK..KMNYRGWRSKVID     697
Scer-L'     QVTNP I     AN    ECY  P.    EMHSSQCDRLL       HWNEFQALENIEAAYPSWSVAEID     638
Ca824-L     QVTNP I     EE   VNYD S.     NKDTEAVPFIE       TDLEMEKIKN..LRNKDFKTTTIP     593
Abra-L      S VTNP I S   RR   KTRL N.L   DEDETQTRLLQ ES    TTAEFRAMRDYMGDTAAEIDATFP     610
Zmay-F      QVTNP I     GL   EVNI K.R    EVGPENADQMA S      NEGELETLLNDSKLKPKVLSTYFD     671
Ecol-L      QVTNP I     AH   LATSI R.EM   CEAEGQAHRLS       LYSDFKQLTT.MKEEHYRADTLDI     611

Pbor-L      IVFDPKSGVK   SA DNL TQ DQ IEA INL      R..SVDAENA          AGV H   RTGA  RV  663
Syne-F      ILFDPNQ EA  KT  DNLFTE DQ ISQ ANL      R..QVSAEKA               N  S  KV  683
Bsub-L      VLFS.....E   R  KDMFTQ EK ISQ VS        ..KMNERLT          L Q  RK  L TKV  651
Ca262-L     IVFD...KGG   D  HEVFEK QE YEK YT        R..NVCEAKV  S         Q    K T  SV  654
Msat-L'     ITYSKERSTK   E  DRI TE HN  SE YTT        ..AFSKKHV SS      V  Q   KTLE  RV  767
Scer-L'     ITFDKSEGLLGYTDT DKITKL SE  DD KK      T  R..KMGANRVS SS      CI H    NKQ SQV 708
Ca824-L     ITFKYDT ID  KE  EKI ER SKR EE FN        ..HIDSYEA   S       V H    KEKT  KV  663
Abra-L      VDGGPE....  RD  RRIRQETED  RG ATH   T   ..EAMGPARA  A     AV        SNL  TFT  676
Zmay-F      IRKGLDG...  DKT QAL EE DA  RS SQLL     RSEAPEPTRP   I      AI  Q   QNG  MSA  740
Ecol-L      TFDVTKT...T  AT KEL DK EK  RS TV       ..NIAKDRL    AP      I      DQ L CDA  678

Pbor-L      GLVL  SG P       F           CG   IN     F TIESMIAQ.........GLLPNLEYAT CK  IKSATK  726
Syne-F      GLVL  SG P          V        CG   IN     F TLDGMIAE.........GLLVNVDHKT CK  IKAATK  746
Bsub-L      SIIV  SG A       F  A        AD   IN      YATYKQEIDE.........GRLD.ISYEEAVSK GKSITE  713
Ca262-L     AIIL  SG P          T         S   VN   Y  Y ALRGLREE.........GLLE.LDYDKAVY  NKAVLK  716
Msat-L'     ALM   SA P       CT           AD   IC      I AIWRLQVDGKIPP..KASG.DFNSKDELVKK FKASTY  836
Scer-L'     ALIL  T  A        CV         CD   WY      M TLVRMNREGLLRN..VNNDNDTLEEGQILE  KHAIDA  778
Ca824-L     SIV   TG A  TM  F  L          T            F SIRQIVRE.........KDIEVESQEKAIE  IYAINH  726
Abra-L      SLNVRTA GLDT YF   V          TTV           Q AIAER.........HRRGLFGSMPLEKGMA  KKAIDD  739
Zmay-F      SIVA TA CFST      F  C        S   VC       L TCRQWRLSNKTLNLMRNGKMPTVTIEQAQR  IKAVKS  812
Ecol-L      NIIV TASA DP       I  V        AT   IY   Y   Y TLGRLVDT.........HAIAKDYRTVML  RNGINK  740

Pbor-L      VV  VA    GIS  IQ           FE      NRSVVDRY TWTA   E VDLEVITQEALLRH.HQAFSERQPSG.  796
Syne-F      VI  VA    GIS  IQ           FE      NQSVIDEY CRTS   QESDLGVIAQEAILREQH.AFAPR.P.GD  815
Bsub-L      VV        GIS  NQ           FE      SRDVIDRY SGTA Q G IDLQTIAKEAQRRH....R.EAYQDDY  780
Ca262-L     II  L     GIS  IQ     Q     FE      GKEVIEKY TNTV   G IGLKEIQREAEINHEKGFNDKTYAAD.  787
Msat-L'     MM  LA    GIS  LA           FE      SSEVIEKC AGTP   E ATFEMLAQDALHLHELAFPSRIFSPGS  908
Scer-L'     IL        GIS  HA           FE      DNSIVDLC TGTS   R VTFEYLAQDAFSLHERGYPSRQTIS..  848
Ca824-L     IL  L      IS  LR           FE      SSKLVSEY EGTP   E IGIDEVAKEVLNRYKNAFNKIRKPVS.  797
Abra-L      LL        G   VIS       GN  E       SRALVAEH PANV   S IGLNGIQKKVLEQHATAYNEEVVA...  808
Zmay-F      LL  L     GI  LLS    C   Q  F  IY   GQEVVDLA CGSV   G LTLDELGRETLSFWVKAFSEDTAKR..  882
Ecol-L      LY        GIS  AA    CSK    FE      HDDVVGLC QGAV   G ASFEDFQQDLLNLSKRAWLARKPI...  809

Pbor-L      ...QT GV   E Q    ...       AH FS  Q      RAV EG   ELE Q AALV   QNQQHFT    G  Q KP.  861
Syne-F      ..LHT DV   E Q    ...   P   E  FS   Q     RA  E    EL  Q AALV   QNQKFFT    G  D QD.  881
Bsub-L      ..SKT EP   D Q    ...N      H AF   KT    T  WCARN  NL   Q TKAA  ER..IGF     N  A DGN  845
Ca262-L     ...FT DSP  YEKL SGEN       E  Y   L    KK  ESTR     EL   E TSLI  KEE.AEIN     G  E NYN  855
Msat-L'     AEAVA PNP  D H    ...       V  N  LA  AK   EAA TNSVDA   Q SKTI   LNKACNLRGL  KFKDAA  977
Scer-L'     .KSVN PE   E H    ...D      YK  N  TA  AS  DT  NK DVS  QL VKKEH  AIRDCTLRGL  ELDFEN  916
Ca824-L     ....V SV   Q S    ...  N    Y  F  D    YR  VSA    KM  KE SHII   QDKNLCT    G  E KD.  861
Abra-L      ..... PV   F R    ...       R  GW GGV      Q  TNDS TT  K  SEQV  KRP..PHQ     D  E RST  870
Zmay-F      ..... ENF  FIQS   ...P       Y  ANK EMSKL   K   EKRDNA TV  QQHLASRP..VNV     D  E KSD  944
Ecol-L      ......SQ   LLK V  ...H       Y  AY  DV   T   Q  QS   SD  QE AKLV  ERP..ATT     D  A TPG  870
```

```
Pbor-L    RQPI LEE EPIEA M FFK MSY SISKE HEA IAMN G S TGEGGE PD YTWSN........      925
Syne-F    RESI LEE EPIEA M FFK MSY SISKE HES IAMN G S TGEGGE PE FTWTN........      945
Bsub-L    RKPLKLEE ESAES V FFK MSF SLSKE HEA IAMN G S SGEGGE PK FVPD.........      908
Ca262-L   SKEI IEE ESVSE V FFK MSY SISKE HEA IAMN G S TGEGGE KE WTLD.........      918
Msat-L'   SK.V ISE EPASE V FFC MSY SISLE TA TAMNT G S TGEGGE PS MEPL.........     1039
Scer-L'   SVSI LEQ EPWTE A FFAS MSY SISME HST IAMN G S CGEGGE AE SAVQ.........      979
Ca824-L   LKPI LEE EPVNE L FFSS MSF SISKE HET IAMN G S SGEGGE NE YKAD.........      924
Abra-L    KAPV VDE ESITA R FFI P MSM ALSPE HGT NVAMN G S SGEGGE PA FRPD.........      933
Zmay-F    RAPI IGK ESATS VE FC MSL AISRET HEA IAMN G S SGEGGE PI WNPLTDVVDGYSP    1016
Ecol-L    ENAVNIAD EPASE F FD MSI ALSPE HEA EAMNS G NS SGEGGE PA YGTN.........      933

Pbor-L    .........EQ S..KN IKQVAS RFGV SL LSQ R I IKM AQGAKPGEGG LF R YPW AKV HS      987
Syne-F    .........DQ S..KN IKQVAS RFGV SL LSQ K I IKM AQGAKPGEGG LF K YPW AKV HS     1007
Bsub-L    .........EN D..RR IKQIAS FFGV KSH LVN D L IKM AQGAKPGEGG LF NK YPW DV GS      970
Ca262-L   .........AN S..RR SIKQIAS RFGV SE LVN D L IKL AQGAKPGEGG LF T YPW AKT HS      980
Msat-L'   .........AD SRNPKR IKQVAS RFGV SSY LTN D L IKM AQGAKPGEGG LF H IGD AIT NS     1103
Scer-L'   .........EN ..TMR IKQVAS RFGV SY LSD D I IKI AQGAKPGEGG LF H SKD AKT HS     1041
Ca824-L   .........AN LR..R IKQIAS RFGV AE LVN D L IKM AQGAKPGEGG LF R DVN AKV HS      986
Abra-L    .........KN N..WN IKQVAS RFGV AE LNQCR L IKV AQGAKPGEGG LF F TEM ARL HS      995
Zmay-F    TLPHLKGLQN T..AT IKQVAS RFGV PT LVN DQ I IKI AQGAKPGEGG LF K SAY ARL NS     1086
Ecol-L    ................KV RIKQVAS RFGV PA LVN DVI IKV AQGAKPGEGG LF D TPY AKL YS      990

Pbor-L    TPGV GLISPPPHHDIYSIEDL LIHDL NA RD ARIS VKLVS V G GTIAAGV S A HA VVLI GYD GGTG   1059
Syne-F    TPGV GLISPPPHHDIYSIEDL LIHDL NA RE ARIN VKLVS V G GTIAAGV A HA VVL GYD GGTG   1079
Bsub-L    TPGV GLISPPPHHDIYSIEDL LIHDL NA RD ARIS VKLVS KA GTIAAGV A TA VIVI GYD GGTG   1042
Ca262-L   TTGV GLISPPPHHDIYSIEDL LIYDL NA TGARV VKLVS C GTVAAGV A GA VIL GYD GGTG   1052
Msat-L'   TAGV GLISPPPHHDIYSIEDL LIHDL NA PAARIS VKLVS AG GVIASGVV HA HVL HD GGTG   1175
Scer-L'   TPNV GLISPPPHHDIYSIEDL K LIYDL CA PRAGIS VKLVS V G GIVASGV A KA HIL HD GGTG   1113
Ca824-L   TPGI DLISPPPHHDIYSIEDL LIFDL CV PSSRIS VKLVS V G GTVAAGV A HA SIL H GGTG   1058
Abra-L    TPGV MLISPPPHHDIYSIEDL LIYDL QI PDAKVT VKLVS RS G GTIAAGV A NA IILI NS GGTG   1067
Zmay-F    KPGV PLISPPPHHDIYSIEDL LIYDL HQI PKAKVS VKLVS A GTVASGVS K NA IIQ H GGTG   1158
Ecol-L    VPGV TLISPPPHHDIYSIEDL LIFDL QV PKAMIS VKLVS P V GTIATGVA YA LITI AGY GGTG   1062

Pbor-L    A QT KH GLPWE GLAETH TLVLNNL S IVVEA GQMK R VVM AL GA A AP VTL GCI   1131
Syne-F    A QT KH GLPWE GLAETH TLVLNNL S IVVET GQMK R VAIA L GA S AP VSL GCI   1151
Bsub-L    A KT KHT GLPWE GLAEAH TLMLNGL D VVLET GKLM R VVM AL GA A AP VVL GCV   1114
Ca262-L   A KN KN GLPWE GLAEAH TLLLNDL E VRVEV GKLMS R VAV AL GA A AP VTL GCV   1124
Msat-L'   A RWT S GLPWE GLAETH TLVANDL G TTLQT GQLK R VAIA L GA S AP ITL GCI   1247
Scer-L'   ARWT Y GLPWE GLAETH TLVLNDL RNVVVQT GQLR F IAV VL GA S T VP IAM GCV   1185
Ca824-L   A IS KH GIPWE GLSEAQ VLLLNNL S VVLQT GQLK R VVIA L GA V ASTI VSL GCV   1130
Abra-L    A QT F GLPWE GLSEVH VLTLNRL H VRLRT GGLK R IVI A GA G AS IAM GCI   1139
Zmay-F    A IS KH GGPWE GLTETN TLIQNGL E VVLRV GFRS Q VLI A A GA GSVA IAT GCV   1230
Ecol-L    A LS Y GCPWE GLVETQ ALVANGL H IRLQV GGLK G VIIK I GA S G GP VAL GCK   1134

Pbor-L    VC L TQP GI DPQLR.AS I D DNTV FMKFI Q V I AQL RTLN V T I EAKQ..   1200
Syne-F    A L TQP GI NPELR.AK TGD AHAV FMTFI T L V AQL RTIN V T I EPKK..   1220
Bsub-L    A L TQP GV NPELR.KK D DHIV YMLFI E VR Y AL KTFD I T V HASER.   1184
Ca262-L   V L TQP GV NEELR.KR KNQ.NLLLTLCIYSTRI I KL RKLD I V K KQKE..   1192
Msat-L'   K K TQP GI DPVLR.EK AE EHVI FFFMV E I SQL RTVN V S M EVDKEV   1318
Scer-L'   LPR L SQA GI DPYLR.SK KQ EHVI FFYYLIQ L KL RTID V S...EKLKKR   1253
Ca824-L   L N L T E GI DPELR.KR K K EYVI FLTFI Q V Y AQL RTIN V V K QAKN..   1199
Abra-L    V Q S TQP V VCV DDKLR.QK V T EKVV LFTFL E V ILAGL RSLN VI T L HQVSR.   1209
Zmay-F    A I T N P V S REELR.AR P V GDLV YFLFV E V AA QL EKLD II T L KPKHI.   1300
Ecol-L    YL I L N A TE V DDKLRKNH H L FKVT YFEFI R TR L QL VTRLV LI T L KELDG.   1205
```

```
Pbor-L     AVEH  AKGL   KI YQ DV.D.PS GL  QIP    GL DKS  LTVL   LCKD  I EG.  P KATLP    V   1269
Syne-F     AVAH  AKG    TI HQ EVGD..D GR  QIP    G QHS  ITQL   LCQP  IAKGE  .  TATLP  T I   1289
Bsub-L     AKEH  SQL    TL YQ EG.....  .RTFQSP    K DQS  ITTI  PAVQE  I SG.  EADISIE  N     1249
Ca262-L    NIHG  AKN    AV YT DKYKGKV  KFDE KKY  FK NKV  EKIF   KFKD  I NKI  .TNFEID   T     1263
Msat-L'    VKGNA LENJ   LL RP AELR.PEAAQ  CVQK   GLDMA   N..K ISLSN ALEKGLP  YIETP  C     1387
Scer-L'    DDVNA INI    PI TP HVIR.PG PTKF TKK   K HTR   N..K IDEAEVTLDRGLP NIDAS  I      1322
Ca824-L    AVSH  AKGI   KI YK DMPKR..  KP   VA     GLDRI   Y..K IQIAKDALNSK S VGNFE  K V   1267
Abra-L     ..GAEHLDD    NPR AQVDPGEN...AR   LQGR    PDT   ARIVA ARPLFE G.E  . QLAYNAR      1274
Zmay-F     ..SLV TQHI   YL SN GLPEWS..SSQIRSQ   TNGPV   ETILA PEIADAIENE E VSKAFQ  Y V     1368
Ecol-L     ..FTA QQK A  KL ET EPHPGK..AL    ENNPPFDNGL   AQLLQ AKPFVD RQS ..TFWFD  R       1271

Pbor-L     N VV TILGNEITKRHWN.  PE  TVHLHFQ                VPK VTLE  EG  NDYLGKG    KLIVY P   1340
Syne-F     N VV TIVGNEITKRHWE.  PE  TVHLHFQ                IPK MTLE  EG  NDYLGKG    KIIVY P   1360
Bsub-L     N VA TITGSEISNRYGEE  PE  TIKLHFT                VPK MTLY  DG  SNDY        KIIVKSS  1321
Ca262-L    D AL TILGSEITRVNGTD  PE  TISIKCN                IPK LTFE  EG   ND F       KLIVY P   1335
Msat-L'    M AV TNLSHEVTKRYNLA  PA  TIHIQFT                LCP ITLE  EG  SNDY        KVVVY P   1459
Scer-L'    D AL STLSYRVSKKFGED  PK  TVVVNIE                LAS ITFI  N    NDY        IIVIK P   1394
Ca824-L    D SV AMLSGKIAKIYGEN  PD  TIRFNFF                GMK MTIV  EG   NDY        KIVIKTP   1339
Abra-L     Q AI TRLSSMVTRKFGMF  QP  HITIRLR                AVQ IKLE  M    NDY        TIVVR T   1346
Zmay-F     D AVCGRVAGVIAKKYGDT  AG.QLNITFN               C LTP MNIR  V    NDY      A  ELVVV V   1439
Ecol-L     D SV ASLSGYIAQTHGDQ  AA PIKAYFN                NAG VELY  T    NDY      A  LIAIR P   1343

Pbor-L     AVSSFDPSE  I  I  VA    AT  EVYI  I   EPF      VNTW E V .DHG  EYMTG TV   L AT .   1410
Syne-F     KGSSFIASE  I  A  VCL   AT  EVYIS  V   EPF      VNTW E V .DHG  EYMTG KV   L QT .   1430
Bsub-L     EGFNSASDD  V  I  VA    AT  EAYIN  R   EPF      VNVW E I .DHG  EYMTG SV   L DV .   1391
Ca262-L    KKSTFIAED  I  I  VAL   AT  KVFIN  I   EPF      ATAW E V .AHGL EYMTG KV   L KT .   1405
Msat-L'    KGSNFDPKD  I  I  VAL   AT  EAYFN  M   EPF      ALAW E V .DHG  EYMTG TV   L KT .   1529
Scer-L'    KDSKFKSDE  V  V  TCF   AT  TAFIS      EPF      ATIV ERIKGNNAF EYMTG RA   L QM S   1466
Ca824-L    ERASYKQEE  V  A  TIL   AT  KLFIN  V   EPF      AYAWAE T .DHC  EYMTG IA   L ET .   1409
Abra-L     TSSPLETNK  T  I  TVL   AT  KLFAA  Q   EPF      ATVW E C .SNG  EYMTG TA   L RV .   1416
Zmay-F     DKTGFVPEDAT   V  TCL   AT   QVFVR  K   EPF      LCQAW E T .DHC  EYMTG CV   L KA .   1509
Ecol-L     VGSAFRSHEAS   I  TCL   AT  ERLYAA R   EPF      AITW E I .DNG  EYMTG IVC  L KT .   1413

Pbor-L     R  A          A  IYD AG.D  ATR  MQMVAI PFEDPED.IAIVRQ   Q  AN   H HKAKD I ENWEQA    1480
Syne-F     N             VA  IFD TG.D  ATR  SAMVGLEKLEDPEE.IKDLKE       N   D  AKGKAV ADWE.A   1499
Bsub-L     R  A          IA  YLTED VKR  KRK  LEMILFESLEDEKE.IQQIKA   R   TA    QKAEDL DQWE..   1460
Ca262-L    I  A          VA  IYK DP.N  RINL  EEMILLEELNID.D.EEELKA   E   KV   G PKANKI YMFET.   1473
Msat-L'    R  A          IA  VLDVDG.T  QSRC  LELVDLDKVEEEED.IITLRM   QQ  QRH   N LLAKEW VDFEN.   1598
Scer-L'    L  AFS  AT    IA  CLTSDYDD  VGKI  KDTVELESLCDPVE.IAFVEN   QE  WN    Q DLAARI GNFNH.   1536
Ca824-L    P  G          M   ILD D.T  NDKCKA TLEITSDYDEED.EKVLRG   EE  YN    N DKAKVI ENWG..   1477
Abra-L     D  A    T     A   VYDLDD.S  PLYI  DESVIFQRIEVGHY.ESQLRH  EE  VTE   Q RFAAEI NDWAR.   1485
Zmay-F     R  VA   T     L   ILD DD.T  VPKV  KEIVKMQRVNAPAG.QMQLKG   EAYWEK   G ERGIAI REWEA.   1578
Ecol-L     V       T      A  VLDESG.D  RKRV  PELWEVLSVDALAIHEEHLRG  TE  VQH  G QRGEEI ANWST.   1483

Pbor-L     RS..         D      QAL.RRAEESGLSGDD LTAAFEENARDVARVGGS                           1530
Syne-F     .SIP         D      QAI.KKALEAGLSGDD LNAAFEENAKDVARIGGS                           1550
Bsub-L     DSVK         I  N   ASI.EEQK AGLSDEE IMFAFEANTKPKQNTAASGQKQAVVQ                   1520
Ca262-L    .EKV    H     I  D   ETV.EKYKNLGSDEEE LIKTFQEIKGI                                 1517
Msat-L'    .LLP     F     E     ASM.KSDA SKDAVER AEDVDEQDDEAQAVEKDAFEELKKLATASLNEKPSEA      1668
Scer-L'    .YLKD VK   I  TD     LKE.KAEA KAKAKATSEYLKKFRSNQEVDDEVNTLLIANQKAKEQEKKKSITI       1606
Ca824-L    EYKT  K   I  TA  LI  EKS.KQKA AVNM                                               1507
Abra-L     .EVT  WQ   V  EMLNR  EVPVHLPK ISAE                                               1515
Zmay-F     .YLPL WQ   V PSEEDSPEACAEFERVLAKQATTQLSAK                                        1616
Ecol-L     .FAT  AL   K  SSDVKALLGHRSRS AELRVQAQ                                            1517

Msat-L'    PKRP                                                                              1672
Scer-L'    SNKATLKEPKVVDLEDAVPDSKQL                                                          1630
```

Appendix C

Alignment of the deduced *C. acetobutylicum* P262 GltA (Ca262-L) protein with Fd-dependent GOGAT enzymes (-F), and the α subunit (-L) and α subunit domains (-L') of bacterial and eukaryotic pyridine dependent GOGATs, respectively. A full description of the abbreviated protein sequences included, together with their accession numbers, are supplied in Table 2.1. The black, gray and light shaded regions represent residues conserved in 100%, > 75% and >50% of the sequences aligned, respectively.

APPENDIX D

Section A

```
     ORFE
   1 GTAAGTGATAAAAAAGATACATATGAATTAGAGTTGGGAAAGATATATTATTTAAATGAAAAATTAAATGGTGATATTAAGAAAATTGCAAATTCAAAAG   100
     V  S  D  K  K  D  T  Y  E  L  E  L  G  K  I  Y  Y  L  N  E  K  L  N  G  D  I  K  K  I  A  N  S  K  E

 101 AAACAATATCAAATGCATTAGAAAAGTTGAATGAAATTGATGAAAACAATAGTGAGATTCCTAAAAAATTTAATGGTAGGCTAAGGGATTATCAGATAAA   200
       T  I  S  N  A  L  E  K  L  N  E  I  D  E  N  N  S  E  I  P  K  K  F  N  G  R  L  R  D  Y  Q  I  K

 201 AGGGTATAATTGGTTTGAAAATTTAAGCTATTTAGGCCTTGGAGGAATCTTGGGCGATGAAATGGGGCTTGGAAAAACAATTCAAACTATTGTATTTTTA   300
      G  Y  N  W  F  E  N  L  S  Y  L  G  L  G  G  I  L  G  D  E  M  G  L  G  K  T  I  Q  T  I  V  F  L

 301 GCATCTAAGCAAGGAAAAACGATTCTCATATTATGCCCAACCTCTCTTATATATAATTGGAAAGAAGAGTTCAATAAGTTTGCTCCAAGTTTAAGCGTAG   400
     A  S  K  Q  G  K  T  I  L  I  L  C  P  T  S  L  I  Y  N  W  K  E  E  F  N  K  F  A  P  S  L  S  V  G

 401 GTATTGTACATGGAAATAAAAATGAAAGAAAAAAAGTATTGGATAATATTGAGGAATATGATGGTTTGTTAACCACTTATGGAACATTAAGAAATGATTG   500
       I  V  H  G  N  K  N  E  R  K  K  V  L  D  N  I  E  E  Y  D  G  L  L  T  T  Y  G  T  L  R  N  D  C

 501 TTTAGAATATGAGAATATTAAATTTGATTACTGTATTTTAGATGAAGGTCAAAATATAAATAATCCTAAAGCTGAAACTACAAAGATAGTAAAATATATA   600
      L  E  Y  E  N  I  K  F  D  Y  C  I  L  D  E  G  Q  N  I  N  N  P  K  A  E  T  T  K  I  V  K  Y  I

 601 AATTCAAAGAGTAGATTTATATTAACAGGACCTCCTATAGAAAATAATTTATTAGAATTATGGTCATTATTTGATTTTATAATGCCAGGATATCTTTATA   700
     N  S  K  S  R  F  I  L  T  G  P  P  I  E  N  N  L  L  E  L  W  S  L  F  D  F  I  M  P  G  Y  L  Y  S

 701 GTAAAGAAGAATTTTCAAATAAATTTATTTTTTTGGATAAGGAAAATTTAGATGATCTTAAAATTCTGATTAGACCATATATATTAAGAAGACTTAAAAA   800
       K  E  E  F  S  N  K  F  I  F  L  D  K  E  N  L  D  D  L  K  I  L  I  R  P  Y  I  L  R  R  L  K  K

 801 AGATGTAATAAAAGAATTGCCAGATAAAATAGAGAAGAAATTTTTTGTTGAATTATCCTTAGAACAAAAGAAATTATATAGAAGTTTTATAAAAGATGTT   900
      D  V  I  K  E  L  P  D  K  I  E  K  K  F  F  V  E  L  S  L  E  Q  K  K  L  Y  R  S  F  I  K  D  V

 901 CAAAGCAAATTACAAAATGCAGAAACACGAGGAAATAATATGACTATATTTTCTTATTTGACAAGACTTCGTCAAATTTGTTTAGATCCATCTATAATAA  1000
     Q  S  K  L  Q  N  A  E  T  R  G  N  N  M  T  I  F  S  Y  L  T  R  L  R  Q  I  C  L  D  P  S  I  I  I

1001 TTGATGATTACTTAGGGGAAAGTTCTAAATTAAATTTAGCAAAGGCCTTAATAATTAAAATATATATAGACACAAATTTAATATTCTCACAATTTACTAC  1100
       D  D  Y  L  G  E  S  S  K  L  N  L  A  K  A  L  I  I  K  I  Y  I  D  T  N  L  I  F  S  Q  F  T  T

1101 AGTATTAAAAAGAATGGAAAATCAGTTGAATCTTGTGGAAATTGATTATTGTTATATAGATGGAAGTACTTCCTCTAATGATAGAATTAAAATTGTTGAT  1200
      V  L  K  R  M  E  N  Q  L  N  L  V  E  I  D  Y  C  Y  I  D  G  S  T  S  S  N  D  R  I  K  I  V  D

1201 GAGTTTAATACAAAGATCAAAAAGAGTGTCTTTTTGATATCACTGTTAGCTGGAGGGACTGGACTTAATTTAACAGGCGCAGATATGGTAATTCATTTTG  1300
     E  F  N  T  K  I  K  K  S  V  F  L  I  S  L  L  A  G  G  T  G  L  N  L  T  G  A  D  M  V  I  H  F  D

1301 ATCCTTGGTGGAATCCATCAGTGGAAGAGCAAGCTACTGATAGAGGACATAGAATAGGACAAAAGCATATAGTTGAAGTAATTAAGCTAATAGCTAAAGA  1400
       P  W  W  N  P  S  V  E  E  Q  A  T  D  R  G  H  R  I  G  Q  K  H  I  V  E  V  I  K  L  I  A  K  D

1401 TACAATTGAAGAAAAAATAATGCTTCTTCAGGAAGATAAAAAAATACTTATAAATAATATAATAACTGAAGAGCTTGAAGATATTAATATTACAAGTGCA  1500
      T  I  E  E  K  I  M  L  L  Q  E  D  K  K  I  L  I  N  N  I  I  T  E  E  L  E  D  I  N  I  T  S  A

1501 TTAAAAAGTGAAGAATTGTTAGATTTACTTTTAAATTAATAAGGCACATGAAAGAAAACAGCTAGTCCAAAATAAGATTGGCAGATGGACTTGTCATTTT  1600
     L  K  S  E  E  L  L  D  L  L  L  N  *
```

```
1601  TGTGAATGTGGCTAATCTAAATTTATGTTAAGTTAGATTATAAAAAATAAAACATAATAATTAAAATATGGAGTATAGAATTTTTATAAATTAAAGAAA  1700

1701  TTCTATACTCCATATTTTAATTTTTAAATACGATAAATTACTATAAGAATATATAATAGCTACTTATAGTATAATATAACAGGACAACAATAATTATCGA  1800

1801  ATAAAATTGGAGCAATGCTTCTTTTTACACCAAGTCTTGTTACTTTATTAAGCAAATTATGATAATATATATGTATGTGTGGAATACATAAGTTAAGTTT  1900

1901  AAGAAAATTTTTCTATAAATTAATTATGTATTTATTAAAAACAGTACCAAAAATTATAAGTTTTTTTGTATATAATAATGACCGCGGAAGGGAGCGCAGA  2000

      ORFD
2001  TATGCTCGAATGTGTTTTAGTAGATTTAAATATTAATGAACACACAAGTGTTAATGAACTAATTGAAAATCATATGCCATTTATAATTAAAAGTATTTCT  2100
       M  L  E  C  V  L  V  D  L  N  I  N  E  H  T  S  V  N  E  L  I  E  N  H  M  P  F  I  I  K  S  I  S

2101  GACGTCACAGGTAGGTATGTATCTTGTGAAAATGATGAAGAATTAAGCGTTGGAATGCTTGGATTTAGTGAAGCTATTGAGAGATATGATAATGAAAAAG  2200
      D  V  T  G  R  Y  V  S  C  E  N  D  E  E  L  S  V  G  M  L  G  F  S  E  A  I  E  R  Y  D  N  E  K  G

2201  GTCACTTTTTGTCCTTTGCTAAGCTTGTTATAGGAAGTAGAATTAAGAATTATCTAAAAGCAGAAAATAGACATCAACATTCATCACTTGAAGAGTTGTT  2300
        H  F  L  S  F  A  K  L  V  I  G  S  R  I  K  N  Y  L  K  A  E  N  R  H  Q  H  S  S  L  E  E  L  L

2301  GGACAAGGGTTTAGAAATTAAAGATGAATATTTAGAGCAAAAAGAAGATAATAGCTTGCTGGTGGAAGAGATAAATAGATTAAAATCTGAAATAAGCTCC  2400
       D  K  G  L  E  I  K  D  E  Y  L  E  Q  K  E  D  N  S  L  L  V  E  E  I  N  R  L  K  S  E  I  S  S

2401  TTTGGTTTTACATTAGAGGATTTAGTGAATGAAGCTCCTAAACAACAGGCAACGAGACAAAATGCAATAAATTTATCAGAAAAAATAAGCAACGAAGAAG  2500
      F  G  F  T  L  E  D  L  V  N  E  A  P  K  Q  Q  A  T  R  Q  N  A  I  N  L  S  E  K  I  S  N  E  E  E

2501  AATTTACTTCTTTTATGTATTTGAAAAAAAGATTACCTATTAAGAGAATTGTTTTAAAATTCTCTGTTACGGAAAAGGTAATTAAAAGAAGTAAAAAATT  2600
        F  T  S  F  M  Y  L  K  K  R  L  P  I  K  R  I  V  L  K  F  S  V  T  E  K  V  I  K  R  S  K  K  F

2601  CATAATTTCTGTTGTAATAATACTTGATAAAAATCTTATTGCTTTGAAAAATTGGATCAGGAAGTAGGTGATAATAATGAATAAGGGAATAATAATGGAG  2700
       I  I  S  V  V  I  I  L  D  K  N  L  I  A  L  K  N  W  I  R  K  *

2701  ATTAATAAAAATTATGCCGTAGTGCTAAATCAACAGGGGTAATGGAAAATACAATCCAAAGAAACATGAAATTGGCAAAAATATTTTACTTTGAGGATCA  2800

2801  TATTTGGTCCAATTCACCTATAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  2900

                                                                                   ORFC
2901  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTCACCTTCCTTCCTTTAACTCGCGA  3000
                                                                                   H  L  P  S  F  N  S  R

3001  CAACATAAGTGTATGCAGTTGTTAGCCTTAGTATTAATCCAGTATTCAATTGAAGCAGTAGTAATCAAAGTTATTAAAGTTGATGGAATTAATGCTGATG  3100
      Q  H  K  C  M  Q  L  L  A  L  V  L  I  Q  Y  S  I  E  A  V  V  I  K  V  I  K  V  D  G  I  N  A  D  G

3101  GCCCAAGTATAGATTTTAGTGATGTTAAAGGTGAAAATATAGATGATGGAATTGAGAAAATAAAAGAGAAGCTAGTTGAAGAAAAATATTTAGACAATAA  3200
        P  S  I  D  F  S  D  V  K  G  E  N  I  D  D  G  I  E  K  I  K  E  K  L  V  E  E  K  Y  L  D  N  N

3201  TAAAGAAGTATTGGTTGCTTTTGCATTTGTTCAAAATGATGGGAATAGTAATTATGAAGAAGAAGTAAAAGATGCTATTCAATCAACATTTAAGTCAGAG  3300
       K  E  V  L  V  A  F  A  F  V  Q  N  D  G  N  S  N  Y  E  E  E  V  K  D  A  I  Q  S  T  F  K  S  E

3301  AAAATAACATATGTAAAAGCAGATAAAAATGCTGTTGATCAAGCTAAGACTGAGGGCATAAGTTTAGGTAGATATGAAGTAGCTATTAATGCTGATGAAG  3400
      K  I  T  Y  V  K  A  D  K  N  A  V  D  Q  A  K  T  E  G  I  S  L  G  R  Y  E  V  A  I  N  A  D  E  E

3401  AAACAAAGAGTAAAATAGATAAAGCACCAGTTAAAGATATTACTTCAATAATAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  3500
        T  K  S  K  I  D  K  A  P  V  K  D  I  T  S  I  I

3501  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  3600

3601  NNNNTTACTCATATCTATATAGAGGATCAAGATAAAAACTAGAGTTATGTGTNATNATATGCTTCCTTATAGAAACGGTTAAAGTAATAAAACTGTtTtT  3700
```

Appendix D

```
3701 ATAAAGtGtGCATATATTATGAATAAAAAAGTAACGACTATATTGAAAAGTcATAGTATAATACTATCAaGATATTTAAAAATAATTATAATGAAGTTAA 3800

                                                                                             .ORFB
3801 ATTATTAAAATTgGGGAGATACAAGATGATAAGAGaATATTTAAAAAATAATTATAATGAAGTTAAATTATTAAAATTGGGAGATACAAGATGATAAGAG 3900
                                                                                              M  I  R  E

3901 AATATTTAAAAAATAATGTATTAATATGTGATGGTGCAATGGGAACATATTATTCAGAATTAACAGGTAATGATATAACTTATTGTGAATTTGGGTATAT 4000
       Y  L  K  N  N  V  L  I  C  D  G  A  M  G  T  Y  Y  S  E  L  T  G  N  D  I  T  Y  C  E  F  G  Y  I

4001 AAATAATAAAGATATTATCAAAAGAATTCATGAAGAATATATTAATGCAGGGGCTAAGTTAATAAGGACTAATACTTTTTCAGCGAATAGATATGATTTA 4100
      N  N  K  D  I  I  K  R  I  H  E  E  Y  I  N  A  G  A  K  L  I  R  T  N  T  F  S  A  N  R  Y  D  L

4101 GGAATTTCATTTGATAAACTTAAAGACATAATAACTTTAGGTATAAATATTGCTAAAGAAGTAACAGAAAATACAGCTGTATTCATTGGAGCAAGTATTG 4200
     G  I  S  F  D  K  L  K  D  I  I  T  L  G  I  N  I  A  K  E  V  T  E  N  T  A  V  F  I  G  A  S  I  G

4201 GCCCTATAAGAGAAGAAAGCATAGATGAATGTGATAATGATATTTTAGATGAATATAAGTTTATTGTTGATTGTTTTATAGAAAACAACATAGATATCTT 4300
       P  I  R  E  E  S  I  D  E  C  D  N  D  I  L  D  E  Y  K  F  I  V  D  C  F  I  E  N  N  I  D  I  F

4301 TATTTTTGAAACATTGAGTAATTACAACTATTTAGAAGAAATATGTGGCTATATTAAATTGAAAAATCCTAATAGTTTCATATTATCTCAGTTCGCAGTG 4400
      I  F  E  T  L  S  N  Y  N  Y  L  E  E  I  C  G  Y  I  K  L  K  N  P  N  S  F  I  L  S  Q  F  A  V

4401 AACCAGGATGGGTTTACAAGAGATGGATTGAGTGTAACTAATATTAATAATGTTAAGGGAATAAATCAAATTGATGCATATGGATTTAATTGTGGATCTG 4500
     N  Q  D  G  F  T  R  D  G  L  S  V  T  N  I  N  N  V  K  G  I  N  Q  I  D  A  Y  G  F  N  C  G  S  G

4501 GACCAACTCATATATATGATATAATTAAAAAAATTAATATAGATGGTGATATAGTTTCAGCTCTTCCTAATGCTGGATATCCAGAAATAATTCATGAAAG 4600
       P  T  H  I  Y  D  I  I  K  K  I  N  I  D  G  D  I  V  S  A  L  P  N  A  G  Y  P  E  I  I  H  E  R

4601 AACTGTATATCCTAATAATCCAGTATACTTTGCTCAAAAAGTTAACAAAATAAAATCTTTAGGAACATCAATTATTGGTGGATGTTGTGGTACAAATCCT 4700
      T  V  Y  P  N  N  P  V  Y  F  A  Q  K  V  N  K  I  K  S  L  G  T  S  I  I  G  G  C  C  G  T  N  P

4701 GTTTATATAAAAGAATTAGCTAATTTAGTTAATAAGAATGTTAATGTTGTTAATACATTAACGAAGAATGAGGAAGAAACTAATAAATTTGAAAACAAGC 4800
     V  Y  I  K  E  L  A  N  L  V  N  K  N  V  N  V  V  N  T  L  T  K  N  E  E  E  T  N  K  F  E  N  K  Q

4801 AAGAGAATAGTTTTAAAAATAAGTTGGATAATAATGAGTTTGTAATTGCAATTGAGTTGTCATCACCAACAAATACTGATATATCAAAGCTTATGCAAGG 4900
       E  N  S  F  K  N  K  L  D  N  N  E  F  V  I  A  I  E  L  S  S  P  T  N  T  D  I  S  K  L  M  Q  G

4901 TGCTAAATTATGCAAAGAAAATAATATAGATTTGGTTACAATTCCAGATTCGCCGATGTCAAGGGTAAAAGCTGAATCTACAATTATATCCGCAAAAATT 5000
      A  K  L  C  K  E  N  N  I  D  L  V  T  I  P  D  S  P  M  S  R  V  K  A  E  S  T  I  I  S  A  K  I

5001 AAAAGAGAAATAGGGATTGAAGCTATGCCTCATATTTGTTGCAGAGATAAAAATATTAATGCTATAAGATCAGGTTTGATAGGTGCTCATATTGAAAATA 5100
     K  R  E  I  G  I  E  A  M  P  H  I  C  C  R  D  K  N  I  N  A  I  R  S  G  L  I  G  A  H  I  E  N  I

5101 TAAGAAATGTGCTTGCAATTACAGGAGATCCTATATCAGATGCTAGTAAAGTTGAAACAAAGAGTGTGTTTAATTTAAACTCTTTTAAATTAATAGAGCT 5200
       R  N  V  L  A  I  T  G  D  P  I  S  D  A  S  K  V  E  T  K  S  V  F  N  L  N  S  F  K  L  I  E  L

5201 TATAGATGATATGAATAGTGAAGTTTTTAATAATGATAATATTTTAATTGGTGGAGCATTGAATTTAAATGTATTAAATAAAGAAGTAGAATTTAATCGT 5300
      I  D  D  M  N  S  E  V  F  N  N  D  N  I  L  I  G  G  A  L  N  L  N  V  L  N  K  E  V  E  F  N  R

5301 ATGATGAAAAAGATAGAAAAGGGAGCAAATTTCTTTTTAACTCAACCTATATATGATGATGATGCTATAGAATTCTTAAAAAAGATAAAAGAAAGAACTA 5400
     M  M  K  K  I  E  K  G  A  N  F  F  L  T  Q  P  I  Y  D  D  D  A  I  E  F  L  K  K  I  K  E  R  T  N

5401 ACGTGAAGATACTTGCTGGATTGCTTCCAATTGTGAGTTATAGAAATGCAATGTTTTTAAATAACGAATTACCTGGTGTTACTATACCGGAAAAATATAT 5500
       V  K  I  L  A  G  L  L  P  I  V  S  Y  R  N  A  M  F  L  N  N  E  L  P  G  V  T  I  P  E  K  Y  I

5501 AAATATGTTTTCAGAGGATATGACTAAAGAAGAGGGACAGCAAGTTGGAATAGATATATCAGTGGAAATAGGAAGAAAATTAAAGGGGCTTTGCGATGGA 5600
      N  M  F  S  E  D  M  T  K  E  E  G  Q  Q  V  G  I  D  I  S  V  E  I  G  R  K  L  K  G  L  C  D  G
```

```
5601 TTGTATTTTGTAACGCCATTTAATAGAGTTAATATGCTTATAGAAATTATAAATAAGATAAAGGGATAAGATATTCTTTATATTTAAAAGAAGTGCATAA 5700
     L  Y  F  V  T  P  F  N  R  V  N |M  L  I  E  I  I  N  K  I  K  G  *

5701 AATGTGGTTGATTTTAAACATATCAACCACATTTTTATTTTGATATAAAAATAAGGAGTATGTGTATATATGTGGTAAAAATATGTTAAAAAAAAGACAA 5800
```

Section B

```
     g1tX
   1 TAAAAATAATTAAGAATTTGCGTAAATTGAATAAACTAAAAGTTCGAAATTAAAGTTTTAGTTAGAGAAATATTTTACGAGAAAAGTGTTGAATTTTTGA 100
     K  N  N  *

 101 TAAATTATGAATATATTTGCATTAATGTATTCATAATAAAACCTAGAAGTGTATTCTTCTAGGTTTTATTTATATACTACATTATTATATATTGTACTTC 200

 201 TATTAAATTGATTATTATAAATGATTGTGTAGAAAATATTTTTATTATATTTAATTAGCATTGGGCATAATATATACTATAAAAATGTAAAAATAACAAG 300

 301 CAATATGTAGAACGTTAGATAATATTTTTATATAAGGGACCAATTAAATACTAAAAAGATAAATAAAGTCAAATTTTTGGCTTAAAATACTTTTAAATAT 400

 401 AATTTCTTTACTTTATCAGACTGTTAACATCTTGCTAATTATTATTAAAATATGATAATATATTAATGAGTTTATTGTGTTTTTAAAANTTTTTCGGTAA 500

                                                                                    .ORFF
 501 ATATAATAGTATGCTATGTCATTAGCACATTAAAACGGAATTTTTAACGAATGGATTAAATAATCAAATAGGGAGGATTTTTATGAACATTTTTAAGAAA 600
                                                                                         M  N  I  F  K  K

 601 AAATCATTAGAACAGATGTTGCAAGGTGCACAAAAGACAGATTTAAAGAAAAATCTTAAAGCTAAAGATATAGCAGCATTTGGTATTGGAGCAGTTGTAG 700
     K  S  L  E  Q  M  L  Q  G  A  Q  K  T  D  L  K  K  N  L  K  A  K  D  I  A  A  F  G  I  G  A  V  V  G

 701 GTGTTGGTATCTTTGTCGCAACAGGAGAAGGGGCACATGCAGCTGGACCAGCAGTAATAGTTTCATTTATTATCGCAGGTATAATANNNNNNNNNNNNNN 800
      V  G  I  F  V  A  T  G  E  G  A  H  A  A  G  P  A  V  I  V  S  F  I  I  A  G  I  I

 801 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 900

 901 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1000

1001 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATATTAATAACTGCGGTAATAACAATAATGCTATATTATGGAATGAA 1100
                                                          I  L  I  T  A  V  I  T  I  M  L  Y  Y  G  M  K

1101 GGAAAGTGCAAAAGTTAATAACATAATTGTAGGTATAAAAATTTCTATAATCGTTATATTTGTAATACTTGGAGTAACACATATAAAGTCCACTAATTAT 1200
      E  S  A  K  V  N  N  I  I  V  G  I  K  I  S  I  I  V  I  F  V  I  L  G  V  T  H  I  K  S  T  N  Y

1201 AAACCTTTCGCTCCATTTGGATTTAATGGAATTTTTGCGGCAACAGCTGCTATATTCTTCTCATTTATAGGTTTTGATGCAATATCAACAGCTGCAGAAG 1300
     K  P  F  A  P  F  G  F  N  G  I  F  A  A  T  A  A  I  F  F  S  F  I  G  F  D  A  I  S  T  A  A  E  E

1301 AAGCTGAAAATCCTAAAAGAGATATTCCATTAGGAATTATAATTTGTTTAATAGCAGTTACAGTACTTTATGTTGCAGTTGCAGTTGTACTTACAGGAAT 1400
       A  E  N  P  K  R  D  I  P  L  G  I  I  I  C  L  I  A  V  T  V  L  Y  V  A  V  A  V  V  L  T  G  M

1401 GGTTCCTTTCCAAGAAATAATTTCTGAAAATGCAGTANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGAGCAGTA 1500
      V  P  F  Q  E  I  I  S  E  N  A  V                                                        G  A  V

1501 CTTGGAATGATTTCTACTATTATGGTAGTACTTTATGGTCAAGTTAGAATATTCATGGTTATGTCAAGAGATGGACTTTTACCAAAGATATTTTCAAAAG 1600
     L  G  M  I  S  T  I  M  V  V  L  Y  G  Q  V  R  I  F  M  V  M  S  R  D  G  L  L  P  K  I  F  S  K  V

1601 TACATCCTAAACACAAAACACCATACTTCTCAACTTTAATAACTGGTACTATAGCAGCAATAATAGCTGGATTTTTACCATTAGATATTATTGTTCAATT 1700
       H  P  K  H  K  T  P  Y  F  S  T  L  I  T  G  T  I  A  A  I  I  A  G  F  L  P  L  D  I  I  V  Q  F
```

```
1701  TTTAAGTATAGGAACATTATTAAGCTT  1727
        L  S  I  G  T  L  L  S
```

Nucleotide sequence of the ~5.8 kb *C. acetobutylicum* P262 chromosomal DNA region present upstream of the *gltX* gene (Section A; base no. 5800 corresponds to the first base in Fig. 3.4), and the ~1.7 kb region present downstream of the *gltX* gene (Section B) which includes the terminal three amino acids, K N N of *gltX*. The deduced amino acid sequences of ORF*B*, ORF*C*, ORF*D*, ORF*E* and ORF*F* are shown in the single letter code. The -35 and -10 regions of putative promoter sequences (underlined) are double underlined. Putative Shine-Dalgarno sequences are indicated in boldface type, and stop codons by an asterisk. Inverted repeats with the potential to form stable stem-loop structures are highlighted by converging arrows. Estimated gaps in the sequence are indicated by the symbol N.

APPENDIX E

Standard GSMM minimal medium recipe

The following GSMM base components were combined and made up to the appropriate volume with distilled water and autoclaved. Typically, *C. acetobutylicum* was cultured in 400 ml volumes of GSMM media for the study of glutamine synthetase and glutamate synthase activities at various growth stages.

GSMM base medium

D-glucose	2% w/v
L-cystein hydrochloride	0.05% w/v
Sodium hydrogen carbonate ($NaHCO_3$)	0.1% w/v
Resazurin	$4x10^{-5}$ w/v
Salt stock	25 fold dilution

Salt stock solution:

Calcium chloride ($CaCl_2$, anhydrous)	0.2 g/l
Magnesium sulfate ($MgSO_4.7H_2O$)	0.2 g/l
Potassium phosphate (K_2HPO_4)	1.0 g/l
Potassium phosphate (KH_2PO_4)	1.0 g/l
Sodium hydrogen carbonate ($NaHCO_3$)	10.0 g/l
Sodium chloride (NaCl)	2.0 g/l

The $CaCl_2$ and $MgSO_4$ was first dissolved in 300 ml of distilled water before the remainder of the salts were added, and the solution made up to 1.0 liter and autoclaved.

Once the GSMM base medium had equilibrated in the anaerobic glove cabinet, the following media stock components (made in distilled water) were either autoclaved or filter sterilized as appropriate, reduced, and added aseptically in the box to yield the indicated final concentrations.

GSMM additives

Magnesium sulfate ($MgSO_4.7H_2O$)	0.02% w/v
Manganese sulfate ($MnSO_4.H_2O$)	$5x10^{-4}$ w/v
Iron sulfate ($FeSO_4.7H_2O$)	$5x10^{-4}$ w/v
Zinc sulfate ($ZnSO_4.7H_2O$)	$2.5x10^{-4}$ w/v
Sodium molybdonate ($Na_2MoO_4.2H_2O$)	$2.5x10^{-4}$ w/v
p-aminobenzoic acid (p-ABA)	$5x10^{-5}$ w/v
Thiamin hydrochloride	$5x10^{-5}$ w/v
Biotin	$5x10^{-6}$ w/v
Vitamin B cocktail	4000 fold dilution

Vitamin B cocktail:

Components	**Stock solution**	**Quantity mixed**
Thiamin hydrogen chloride	10 mg/ml	2 ml
Calcium D-pantothenate	10 mg/ml	2 ml
Nicotinamide	10 mg/ml	2 ml
Riboflavin	10 mg/ml	2 ml
Pyridoxine hydrogen chloride	10 mg/ml	2 ml
p-Amino benzoic acid	1 mg/ml	1 ml
Biotin	1 mg/ml	0.25 ml
Folic acid	1 mg/ml	0.25 ml
Vitamin B_{12}	1 mg/ml	0.1 ml
Water		88.4 ml

All components were filter sterilized separately prior to mixing, and the cocktail was stored in the anaerobic box.

Finally, the appropriate volume of sterile anaerobic stock solutions of glutamine (freshly prepared) and/or casamino acids were added in the box to give the final desired nitrogen rich (GSMM non-inducing) or nitrogen poor (GSMM inducing) conditions.

Appendix E

GSMM non-inducing conditions:

Casamino acids (Difco)	0.2% w/v

GSMM inducing conditions:

Casamino acids (Difco)	0.025% w/v
L-glutamine (Sigma, stored at 4 °C)	0.15% w/v

In addition, this standard recipe was modified (as indicated in the text) when the effect of different concentrations of carbon and organic and/or inorganic nitrogen were investigated in relation to the growth and physiology of *C. acetobutylicum*.

LITERATURE CITED

Abrahams, G. L. and Abratt, V. R. (1998). The NADH-dependent glutamate dehydrogenase enzyme of *Bacteroides fragilis* Bf1 is induced by peptides in the growth medium. *Microbiology* **144 (Pt 6)**, 1659-1667.

Aharonowitz, Y. and Friedrich, C. G. (1980). Alanine dehydrogenase of the beta-lactam antibiotic producer *Streptomyces clavuligerus*. *Arch.Microbiol.* **125**, 137-142.

Aiba, H., Adhya, S., and de Crombrugghe, B. (1981). Evidence for two functional *gal* promoters in intact *Escherichia coli* cells. *J.Biol.Chem.* **256**, 11905-11910.

Allcock, E. R., Reid, S. J., Jones, D. T., and Woods, D. R. (1982). *Clostridium acetobutylicum* protoplast formation and regeneration. *Appl.Environ.Microbiol.* **43**, 719-721.

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

Amine, J., Marczak, R., Maazouzi, N., and Masion, E. (1990). Regulations of ammonium assimilations by *Clostridium acetobutylicum*. *J.Basic Microbiol.* **30**, 635-642.

Andersch, W., Bahl, H., and Gottschalk, G. (1982). Acetone-butanol production by *Clostridium acetobutylicum* in an ammonium-limited chemostat at low pH values. *Biotech.Lett.* **4**, 29-32.

Anderson, M. P., Vance, C. P., Heichel, G. H., and Miller, S. S. (1989). Purification and characterisation of NADH-glutamate synthase from alfalfa root nodules. *Plant Physiology* **90**, 351-358.

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

Austin, E. A., Andrews, J. C., and Short, S. A. (1989). Selection, characterization and cloning of *oppI*, a regulator of the *E. coli* oligopeptide permease operon. In *Abstracts Molecular Genetics of Bacteria Phages*. pp. 153-. Cold Spring Harbour Laboratory: Cold Spring Harbour, N.Y.

Avendano, A., Deluna, A., Olivera, H., Valenzuela, L., and Gonzalez, A. (1997). GDH3 encodes a glutamate dehydrogenase isozyme, a previously unrecognized route for glutamate biosynthesis in *Saccharomyces cerevisae*. *J.Bacteriol* **179**, 5594-5597.

Avila, C., Marquez, A. J., Pajuelo, P., Cannell, M. E., Wallsgrove, R. M., and Forde, B. G. (1993). Cloning and sequence analysis of a cDNA for barley ferredoxin-dependent glutamate synthase and molecular analysis of photorespiratory mutants deficient in the enzyme. *Planta* **189**, 475-483.

Babb, B. L., Collett, H. J., Reid, S. J., and Woods, D. R. (1993). Transposon mutagenesis of *Clostridium acetobutylicum* P262: isolation and characterization of solvent deficient and metronidazole resistant mutants. *FEMS Microbiol.Lett.* **114**, 343-348.

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

Baggio, L. and Morrison, M. (1996). The NAD(P)H-utilizing glutamate dehydrogenase of *Bacteroides thetaiotaomicron* belongs to enzyme family I, and its activity is affected by trans-acting gene(s) positioned downstream of *gdhA*. *J.Bacteriol* **178**, 7212-7220.

Bani, D., Barberio, C., Bazzicalupo, M., Favilli, F., Gallori, E., and Polsinelli, M. (1980). Isolation and characterization of glutamate synthase mutants of *Azospirillum brasilense*. *J.Gen.Microbiol.* **119**, 239-244.

Bascaran, V., Hardisson, C., and Brana, A. F. (1989a). Isolation and characterization of nitrogen-deregulated mutants of *Streptomyces clavuligerus*. *J.Gen.Microbiol.* **135 (Pt 9)**, 2475-2482.

Bascaran, V., Hardisson, C., and Brana, A. F. (1989b). Regulation of nitrogen catabolic enzymes in *Streptomyces clavuligerus*. *J.Gen.Microbiol.* **135 (Pt 9)**, 2465-2474.

Behrmann, I., Hillemann, D., Puhler, A., Strauch, E., and Wohlleben, W. (1990). Overexpression of a *Streptomyces viridochromogenes* gene (*glnII*) encoding a glutamine synthetase similar to those of eucaryotes confers resistance against the antibiotic phosphinothricyl-alanyl-alanine. *J.Bacteriol* **172**, 5326-5334.

Belitsky, B. R., Gustafsson, M. C., Sonenshein, A. L., and Von Wachenfeldt, C. (1997). An *lrp*-like gene of *Bacillus subtilis* involved in branched-chain amino acid transport. *J.Bacteriol* **179**, 5448-5457.

Belitsky, B. R., Janssen, P. J., and Sonenshein, A. L. (1995). Sites required for GltC-dependent regulation of *Bacillus subtilis* glutamate synthase expression. *J.Bacteriol* **177**, 5686-5695.

Belitsky, B. R. and Sonenshein, A. L. (1995). Mutations in GltC that increase *Bacillus subtilis* gltA expression. *J.Bacteriol* **177**, 5696-5700.

Belitsky, B. R. and Sonenshein, A. L. (1997). Altered transcription activation specificity of a mutant form of *Bacillus subtilis* GltR, a LysR family member. *J.Bacteriol* **179**, 1035-1043.

Belitsky, B. R. and Sonenshein, A. L. (1998). Role and regulation of *Bacillus subtilis* glutamate dehydrogenase genes. *J.Bacteriol* **180**, 6298-6305.

Bender, R. A. (1991). The role of the NAC protein in the nitrogen regulation of *Klebsiella aerogenes*. *Mol.Microbiol.* **5**, 2575-2580.

Bender, R. A., Snyder, P. M., Bueno, R., Quinto, M., and Magasanik, B. (1983). Nitrogen regulation system of *Klebsiella aerogenes*: the *nac* gene. *J.Bacteriol* **156**, 444-446.

Berberich, M. A. (1972). A glutamate-dependent phenotype in *E. coli* K12: the result of two mutations. *Biochem.Biophys.Res.Commun.* **47**, 1498-1503.

Bogdahn, M., Andreesen, J. R., and Kleiner, D. (1983). Pathways and regulation of N_2, ammonium and glutamate assimilation by *Clostridium pasteurianum. J.Bacteriol.* **110**, 103-109.

Bogdahn, M. and Kleiner, D. (1986). N_2 fixation and NH_4^+ assimilation in the thermophilic anaerobes *Clostridium thermosaccharolyticum* and *Clostridium thermoautotrophicum. Arch.Microbiol.* **144**, 102-104.

Bogonez, E., Satrustegui, J., and Machado, A. (1985). Regulation by ammonium of glutamate dehydrogenase ($NADP^+$) from *Saccharomyces cerevisae. J.Gen.Microbiol.* **131 (Pt 6)**, 1425-1432.

Bohannon, D. E., Rosenkrantz, M. S., and Sonenshein, A. L. (1985). Regulation of *Bacillus subtilis* glutamate synthase genes by the nitrogen source. *J.Bacteriol.* **163**, 957-964.

Bohannon, D. E. and Sonenshein, A. L. (1989). Positive regulation of glutamate biosynthesis in *Bacillus subtilis. J.Bacteriol.* **171**, 4718-4727.

Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heynker, H. L., Boyer, H. W., Crosa, J. H., and Falkow, S. (1992). Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. 1977 [classical article]. *Biotechnology* **24**, 153-171.

Borst, D. W., Blumenthal, R. M., and Matthews, R. G. (1996). Use of an in vivo titration method to study a global regulator: effect of varying Lrp levels on expression of *gltBDF* in *Escherichia coli. J.Bacteriol* **178**, 6904-6912.

Brana, A. F., Wolfe, S., and Demain, A. L. (1986). Relationship between nitrogen assimilation and cephalosporin synthesis in *Streptomyces clavuligerus. Arch.Microbiol.* **146**, 46-51.

Bravo, A. and Mora, J. (1988). Ammonium assimilation in *Rhizobium phaseoli* by the glutamine synthetase-glutamate synthase pathway. *J.Bacteriol* **170**, 980-984.

Brenchley, J. E., Baker, C. A., and Patil, L. G. (1975). Regulation of the ammonia assimilatory enzymes in *Salmonella typhimurium. J.Bacteriol* **124**, 182-189.

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

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

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., and Venter, J. C. (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii. Science* **273**, 1058-1073.

Caballero, F. J., Igeno, I., Cardenas, J., and Castillo, F. (1989). Regulation of reduced nitrogen assimilation in *Rhodobacter capsulatus* E1F1. *Arch.Microbiol.* **152**, 508-511.

Calogero, S., Gardan, R., Glaser, P., Schweizer, J., Rapoport, G., and Debarbouille, M. (1994). RocR, a novel regulatory protein controlling arginine utilization in *Bacillus subtilis*, belongs to the NtrC/NifA family of transcriptional activators. *J.Bacteriol* **176**, 1234-1241.

Calvo, J. M. and Matthews, R. G. (1994). The leucine-responsive regulatory protein, a global regulator of metabolism in *Escherichia coli. Microbiol.Rev.* **58**, 466-490.

Carlson, T. A. and Chelm, B. K. (1986). Apparent eukaryotic origins of glutamine synthetase II from the bacterium *Bradyrhizobium japonicum. Nature* **322**, 568-570.

Cary, J. W., Petersen, D. J., Papoutsakis, E. T., and Bennett, G. N. (1988). Cloning and expression of *Clostridium acetobutylicum* phosphotransbutyrylase and butyrate kinase genes in *Escherichia coli. J.Bacteriol.* **170**, 4613-4618.

Castano, I., Bastarrachea, F., and Covarrubias, A. A. (1988). *gltBDF* operon of *Escherichia coli. J.Bacteriol.* **170**, 821-827.

Castano, I., Flores, N., Valle, F., Covarrubias, A. A., and Bolivar, F. (1992). *gltF*, a member of the *gltBDF* operon of *Escherichia coli*, is involved in nitrogen-regulated gene expression. *Mol.Microbiol.* **6**, 2733-2741.

Cato, E. P. and Stakebrand, E. (1989). 'Taxonomy and phylogeny.' Plenum Publishing Co.: New York.

Cavicchioli, R., Kolesnikow, T., Chiang, R. C., and Gunsalus, R. P. (1996). Characterization of the *aegA* locus of *Escherichia coli*: control of gene expression in response to anaerobiosis and nitrate. *J.Bacteriol* **178**, 6968-6974.

Chai, W. and Stewart, V. (1998). NasR, a novel RNA-binding protein, mediates nitrate-responsive transcription antitermination of the *Klebsiella oxytoca* M5al *nasF* operon leader *in vitro. J.Mol.Biol.* **283**, 339-351.

Chai, W. and Stewart, V. (1999). RNA sequence requirements for NasR-mediated, nitrate-responsive transcription antitermination of the *Klebsiella oxytoca* M5al *nasF* operon leader. *J.Mol.Biol.* **292**, 203-216.

Chen, G. J. and Russell, J. B. (1989). Transport of glutamine by *Streptococcus bovis* and conversion of glutamine to pyroglutamic acid and ammonia. *J.Bacteriol* **171**, 2981-2985.

Chen, J. S. and Blanchard, D. K. (1979). A simple hydrogenase-linked assay for ferredoxin and flavodoxin. *Anal.Biochem.* **93**, 216-222.

Chiang, R. C., Cavicchioli, R., and Gunsalus, R. P. (1992). Identification and characterization of *narQ*, a second nitrate sensor for nitrate-dependent gene regulation in *Escherichia coli. Mol.Microbiol.* **6**, 1913-1923.

Church, D. L., Rabin, H. R., and Laishley, E. J. (1988). Role of hydrogenase 1 of *Clostridium pasteurianum* in the reduction of metronidazole. *Biochem.Pharmacol.* **37**, 1525-1534.

Church, D. L., Rabin, H. R., and Laishley, E. J. (1995). Reduction of metronidazole by hydrogenase from *Clostridia. Anaerobe* **1**, 81-92.

Cogoni, C., Valenzuela, L., Gonzalez-Halphen, D., Olivera, H., Macino, G., Ballario, P., and Gonzalez, A. (1995). *Saccharomyces cerevisae* has a single glutamate synthase gene coding for a plant-like high-molecular-weight polypeptide. *J.Bacteriol* **177**, 792-798.

Covarrubias, A. A., Sanchez-Pescador, R., Osorio, A., Bolivar, F., and Bastarrachea, F. (1980). ColE1 hybrid plasmids containing *Escherichia coli* genes involved in the biosynthesis of glutamate and glutamine. *Plasmid* **3**, 150-164.

Cui, Y., Midkiff, M. A., Wang, Q., and Calvo, J. M. (1996). The leucine-responsive regulatory protein (Lrp) from *Escherichia coli*. Stoichiometry and minimal requirements for binding to DNA. *J.Biol.Chem.* **271**, 6611-6617.

Cui, Y., Wang, Q., Stormo, G. D., and Calvo, J. M. (1995). A consensus sequence for binding of Lrp to DNA. *J.Bacteriol* **177**, 4872-4880.

Curti, B., Vanoni, M. A., Verzotti, E., and Zanetti, G. (1996). Glutamate synthase: a complex iron-sulphur flavoprotein. *Biochem.Soc.Trans.* **24**, 95-99.

Dachs, G. U., Abratt, V. R., and Woods, D. R. (1995). Mode of action of metronidazole and a *Bacteroides fragilis metA* resistance gene in *Escherichia coli*. *J.Antimicrob.Chemother.* **35**, 483-496.

Dainty, R. H. (1972). Glutamate biosynthesis in *Clostridium pasteurianum* and its significance in nitrogen metabolism. *Biochem.J.* **126**, 1055-1056.

Davison, S. P., Santangelo, J. D., Reid, S. J., and Woods, D. R. (1995). A *Clostridium acetobutylicum* regulator gene (*regA*) affecting amylase production in *Bacillus subtilis*. *Microbiology* **141 (Pt 4)**, 989-996.

de Bruijn, F. J., Rossbach, S., Schneider, M., Ratet, P., Messmer, S., Szeto, W. W., Ausubel, F. M., and Schell, J. (1989). *Rhizobium meliloti* 1021 has three differentially regulated loci involved in glutamine biosynthesis, none of which is essential for symbiotic nitrogen fixation. *J.Bacteriol* **171**, 1673-1682.

Deane, S. M. and Rawlings, D. E. (1996). Cloning and sequencing of the gene for the *Thiobacillus ferrooxidans* ATCC33020 glutamate synthase (GOGAT) small subunit and complementation of an *Escherichia coli gltD* mutant. *Gene* **177**, 261-263.

Debarbouille, M., Martin-Verstraete, I., Kunst, F., and Rapoport, G. (1991). The *Bacillus subtilis sigL* gene encodes an equivalent of sigma 54 from gram-negative bacteria. *Proc.Natl.Acad.Sci.U.S.A* **88**, 9092-9096.

Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Aujay, M., Huber, R., Feldman, R. A., Short, J. M., Olsen, G. J., and Swanson, R. V. (1998). The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392**, 353-358.

Deshpande, K. L. and Kane, J. F. (1980). Glutamate synthase from *Bacillus subtilis*: *in vitro* reconstitution of an active amidotransferase. *Biochem.Biophys.Res.Commun.* **93**, 308-314.

Deshpande, K. L., Katze, J. R., and Kane, J. F. (1981). Effect of glutamine on enzymes of nitrogen metabolism in *Bacillus subtilis*. *J.Bacteriol* **145**, 768-774.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387-395.

DiLella, A. G. and Woo, S. L. (1987). Hybridization of genomic DNA to oligonucleotide probes in the presence of tetramethylammonium chloride. *Methods Enzymol.* **152**, 447-451.

Donald, R. G., Lapointe, J., and Ludwig, R. A. (1988). Characterization of the *Azorhizobium sesbaniae* ORS571 genomic locus encoding NADPH-glutamate synthase. *J.Bacteriol.* **170**, 1197-1204.

Drummond, J. T., Huang, S., Blumenthal, R. M., and Matthews, R. G. (1993). Assignment of enzymatic function to specific protein regions of cobalamin-dependent methionine synthase from *Escherichia coli. Biochemistry* **32**, 9290-9295.

Dürre, P. (1998). New insights and novel developments in clostridial acetone/butanol/isopropanol fermentation. *Appl.Microbiol.Biotechnol.* **49**, 639-648.

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

Edmands, J., Noridge, N. A., and Berberich, M. A. (1987). The actinorhizal root-nodule symbiont *Frankia* sp. strain CpII has two glutamine synthetases. *Proc.Natl.Acad.Sci.U.S.A* **84**, 6126-6130.

Edwards, D. I. (1993) . Nitroimidazole drugs--action and resistance mechanisms. I. Mechanisms of action. *J.Antimicrob.Chemother.* **31**, 9-20.

Eggink, G., Engel, H., Vriend, G., Terpstra, P., and Witholt, B. (1990). Rubredoxin reductase of *Pseudomonas oleovorans*. Structural relationship to other flavoprotein oxidoreductases based on one NAD and two FAD fingerprints. *J.Mol.Biol.* **212**, 135-142.

Eisel, U., Jarausch, W., Goretzki, K., Henschen, A., Engels, J., Weller, U., Hudel, M., Habermann, E., and Niemann, H. (1986). Tetanus toxin: primary structure, expression in *E. coli*, and homology with botulinum toxins. *EMBO J.* **5**, 2495-2502.

Ernsting, B. R., Atkinson, M. R., Ninfa, A. J., and Matthews, R. G. (1992). Characterization of the regulon controlled by the leucine-responsive regulatory protein in *Escherichia coli. J.Bacteriol* **174**, 1109-1118.

Ernsting, B. R., Denninger, J. W., Blumenthal, R. M., and Matthews, R. G. (1993). Regulation of the *gltBDF* operon of *Escherichia coli*: how is a leucine- insensitive operon regulated by the leucine-responsive regulatory protein? *J.Bacteriol.* **175**, 7160-7169.

Errington, J. (1993). *Bacillus subtilis* sporulation: regulation of gene expression and control of morphogenesis. *Microbiol.Rev.* **57**, 1-33.

Ertan, H. (1992). Some properties of glutamate dehydrogenase, glutamine synthetase and glutamate synthase from *Corynebacterium callunae. Arch.Microbiol.* **158**, 35-41.

Faires, N., Tobisch, S., Bachem, S., Martin-Verstraete, I., Hecker, M., and Stulke, J. (1999). The catabolite control protein CcpA controls ammonium assimilation in *Bacillus subtilis. J.Molec.Microbiol.Biotechnol.* **1**, 1-8.

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

Filetici, P., Martegani, M. P., Valenzuela, L., Gonzalez, A., and Ballario, P. (1996). Sequence of the *GLT1* gene from *Saccharomyces cerevisae* reveals the domain structure of yeast glutamate synthase. *Yeast* **12**, 1359-1366.

Fink, D., Falke, D., Wohlleben, W., and Engels, A. (1999). Nitrogen metabolism in *Streptomyces coelicolor* A3(2): modification of glutamine synthetase I by an adenylyltransferase. *Microbiology* **145 (Pt 9)**, 2313-2322.

Fisher, R. F. and Long, S. R. (1993). Interactions of NodD at the nod Box: NodD binds to two distinct sites on the same face of the helix and induces a bend in the DNA. *J.Mol.Biol.* **233**, 336-348.

Fisher, S. H. (1989). Glutamate synthesis in *Streptomyces coelicolor*. *J.Bacteriol.* **171**, 2372-2377.

Fisher, S. H. (1993). Utilization of amino acids and other nitrogen-containing compounds. In *Bacillus subtilis and other Gram-positive Bacteria: Biochemistry, Physiology, and Molecular Biology*. pp. 221-8. American Society for Microbiology Press: Washington, DC.

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

Fisher, S. H., Rohrer, K., and Ferson, A. E. (1996). Role of CodY in regulation of the *Bacillus subtilis* hut operon. *J.Bacteriol* **178**, 3779-3784.

Fisher, S. H., Rosenkrantz, M. S., and Sonenshein, A. L. (1984). Glutamine synthetase gene of *Bacillus subtilis*. *Gene* **32**, 427-438.

Fisher, S. H. and Sonenshein, A. L. (1984). *Bacillus subtilis* glutamine synthetase mutants pleiotropically altered in glucose catabolite repression. *J.Bacteriol.* **157**, 612-621.

Fisher, S. H. and Sonenshein, A. L. (1991). Control of carbon and nitrogen metabolism in *Bacillus subtilis*. *Annu.Rev.Microbiol.* **45**, 107-135.

Fisher, S. H. and Wray, L. V., Jr. (1989). Regulation of glutamine synthetase in *Streptomyces coelicolor*. *J.Bacteriol* **171**, 2378-2383.

Folch, J. L., Antaramian, A., Rodriguez, L., Bravo, A., Brunner, A., and Gonzalez, A. (1989). Isolation and characterization of a *Saccharomyces cerevisae* mutant with impaired glutamate synthase activity. *J.Bacteriol* **171**, 6776-6781.

Fourney, R. M., Miyskoshi, J., Day, R. S., III, and Paterson, M. C. (1988). Northern blotting: efficient RNA staining and transfer. *Focus* **10**, 5-7.

**Fraser, C. M., Norris, S. J., Weinstock, G. M., White, O., Sutton, G. G., Dodson, R., Gwinn, M., Hickey, E. K., Clayton, R., Ketchum, K. A., Sodergren, E., Hardham, J. M., McLeod, M. P., Salzberg, S., Peterson, J., Khalak, H., Richardson, D., Howell, J. K., Chidambaram, M., Utterback, T.,

McDonald, L., Artiach, P., Bowman, C., Cotton, M. D., and Venter, J. C. (1998). Complete genome sequence of *Treponema pallidum,* the syphilis spirochete. *Science* **281**, 375-388.

Gally, D. L., Rucker, T. J., and Blomfield, I. C. (1994). The leucine-responsive regulatory protein binds to the *fim* switch to control phase variation of type 1 fimbrial expression in *Escherichia coli* K-12. *J.Bacteriol* **176**, 5665-5672.

Gardan, R., Rapoport, G., and Debarbouille, M. (1997). Role of the transcriptional activator RocR in the arginine-degradation pathway of *Bacillus subtilis. Mol.Microbiol.* **24**, 825-837.

Garnier, T. and Cole, S. T. (1986). Characterization of a bacteriocinogenic plasmid from *Clostridium perfringens* and molecular genetic analysis of the bacteriocin-encoding gene. *J.Bacteriol.* **168**, 1189-1196.

Gazeau, M., Delort, F., Fromant, M., Dessen, P., Blanquet, S., and Plateau, P. (1994). Structure-function relationship of the Lrp-binding region upstream of *lysU* in *Escherichia coli. J.Mol.Biol.* **241**, 378-389.

Genetet, I., Martin, F., and Stewart, G. R. (1984). Nitrogen assimilation in mycorrhizas. Ammonium assimilation in the N-starved ectomycorrhizal fungus *Cenococcum graniforme. Plant Physiology* **76**, 395-399.

Gerischer, U. and Dürre, P. (1992). mRNA analysis of the *adc* gene region of *Clostridium acetobutylicum* during the shift to solventogenesis. *J.Bacteriol.* **174**, 426-433.

Goethals, K., Mergaert, P., Gao, M., Geelen, D., Van Montagu, M., and Holsters, M. (1992). Identification of a new inducible nodulation gene in *Azorhizobium caulinodans. Mol.Plant Microbe Interact.* **5**, 405-411.

Goldman, B. S., Lin, J. T., and Stewart, V. (1994). Identification and structure of the *nasR* gene encoding a nitrate- and nitrite-responsive positive regulator of *nasFEDCBA* (Nitrate Assimilation) operon expression in *Klebsiella pneumonia* M5al. *J.Bacteriol* **176**, 5077-5085.

Gonzalez, A., Membrillo-Hernandez, J., Olivera, H., Aranda, C., Macino, G., and Ballario, P. (1992). Cloning of a yeast gene coding for the glutamate synthase small subunit (GUS2) by complementation of *Saccharomyces cerevisae* and *Escherichia coli* glutamate auxotrophs. *Mol.Microbiol.* **6**, 301-308.

Goodman, H. J. and Woods, D. R. (1993). Cloning and nucleotide sequence of the *Butyrivibrio fibrisolvens* gene encoding a type III glutamine synthetase. *J.Gen.Microbiol.* **139 (Pt 7)**, 1487-1493.

Goodrich, J. A., Cutler, G. and Tjian, R. (1996). Contacts in context: Promoter specificity and macromolecular interactions in transcription. *Celll.* **84**, 825-830.

Goss, T. J. and Bender, R. A. (1995). The nitrogen assimilation control protein, NAC, is a DNA binding transcription activator in *Klebsiella aerogenes. J.Bacteriol* **177**, 3546-3555.

Gosset, G., Merino, E., Recillas, F., Oliver, G., Becerril, B., and Bolivar, F. (1989). Amino acid sequence analysis of the glutamate synthase enzyme from *Escherichia coli* K-12. *Protein Seq.Data Anal* **2**, 9-16.

Gottschal, J. and Morris, J. G. (1981). Non-production of actone and butanol by *Clostridium acetobutylicum* during glucose- and ammonium-limitation in continuous culture. *Biotech.Lett.* **3**, 525-530.

Grassl, G., Bufe, B., Muller, B., Rosel, M., and Kleiner, D. (1999). Characterization of the *gltF* gene product of *Escherichia coli. FEMS Microbiol.Lett* **179**, 79-84.

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

Green, E. M. and Bennett, G. N. (1996). Inactivation of an aldehyde/alcohol dehydrogenase gene from *Clostridium acetobutylicum* ATCC 824. *Appl.Biochem.Biotechnol.* **57-58**, 213-221.

Green, E. M., Boynton, Z. L., Harris, L. M., Rudolph, F. B., Papoutsakis, E. T., and Bennett, G. N. (1996). Genetic manipulation of acid formation pathways by gene inactivation in *Clostridium acetobutylicum* ATCC 824. *Microbiology* **142 (Pt 8)**, 2079-2086.

Gregerson, R. G., Miller, S. S., Twary, S. N., Gantt, J. S., and Vance, C. P. (1993). Molecular characterization of NADH-dependent glutamate synthase from alfalfa nodules. *Plant Cell* **5**, 215-226.

Grundy, F. J., Waters, D. A., Allen, S. H., and Henkin, T. M. (1993). Regulation of the *Bacillus subtilis* acetate kinase gene by CcpA. *J.Bacteriol* **175**, 7348-7355.

Gunsalus, R. P. and Park, S. J. (1994). Aerobic-anaerobic gene regulation in *Escherichia coli*: control by the *ArcAB* and *Fnr* regulons. *Res.Microbiol.* **145**, 437-450.

Gustafson, J., Strassle, A., Hachler, H., Kayser, F. H., and Berger-Bachi, B. (1994). The *femC* locus of *Staphylococcus aureus* required for methicillin resistance includes the glutamine synthetase operon. *J.Bacteriol* **176**, 1460-1467.

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

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

Harwood, C. R. and Cutting, S. M. (1990). Molecular Biological Methods for *Bacillus*. Edited by C. R. Harwood and S. M. Cutting. John Wiley & Sons Ltd.: West Sussex PO19 1UD.

Helling, R. B. (1994) . Why does *Escherichia coli* have two primary pathways for synthesis of glutamate? *J.Bacteriol* **176**, 4664-4668.

Hemmila, I. A. and Mantsala, P. I. (1978). Purification and properties of glutamate synthase and glutamate dehydrogenase from *Bacillus megaterium*. *Biochem.J.* **173**, 45-52.

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

Henikoff, S., Haughn, G. W., Calvo, J. M., and Wallace, J. C. (1988). A large family of bacterial activator proteins. *Proc.Natl.Acad.Sci.U.S.A* **85**, 6602-6606.

Henkin, T. M. (1996). The role of CcpA transcriptional regulator in carbon metabolism in *Bacillus subtilis*. *FEMS Microbiol.Lett* **135**, 9-15.

Hilgert, I., Stoyanov, S., and Kristofova, H. (1987). Immunosuppression induced by monoclonal anti-Thy-1,2 antibodies, cyclosporine A, and lentil lectin in mice. *Transplant.Proc.* **19**, 1269.

Hill, R. T., Parker, J. R., Goodman, H. J., Jones, D. T., and Woods, D. R. (1989). Molecular analysis of a novel glutamine synthetase of the anaerobe *Bacteroides fragilis*. *J.Gen.Microbiol.* **135 (Pt 12)**, 3271-3279.

Hillemann, D., Dammann, T., Hillemann, A., and Wohlleben, W. (1993). Genetic and biochemical characterization of the two glutamine synthetases GSI and GSII of the phosphinothricyl-alanyl-alanine producer, *Streptomyces viridochromogenes* Tu494. *J.Gen.Microbiol.* **139 (Pt 8)**, 1773-1783.

Hirasawa, M., Hurley, J. K., Salamon, Z., Tollin, G., and Knaff, D. B. (1996). Oxidation-reduction and transient kinetic studies of spinach ferredoxin- dependent glutamate synthase. *Arch.Biochem.Biophys.* **330**, 209-215.

Hoffman, P. S., Goodwin, A., Johnsen, J., Magee, K., and Veldhuyzen van Zanten, S. J. (1996). Metabolic activities of metronidazole-sensitive and -resistant strains of *Helicobacter pylori*: repression of pyruvate oxidoreductase and expression of isocitrate lyase activity correlate with resistance. *J.Bacteriol.* **178**, 4822-4829.

Holdeman, L. V. E., Cato, E. P., and Moore, W. E. C. (1977). 'Anaerobe laboratory manual.' 4 Edn. Virginia Polytechnic Institute and State University: Blacksburg.

Hu, P., Leighton, T., Ishkhanova, G., and Kustu, S. (1999). Sensing of nitrogen limitation by *Bacillus subtilis*: comparison to enteric bacteria. *J.Bacteriol* **181**, 5042-5050.

Hummelt, G. and Mora, J. (1980). NADH-dependent glutamate synthase and nitrogen metabolism in *Neurospora crassa*. *Biochem.Biophys.Res.Commun.* **92**, 127-133.

Ish-Horowicz, D. and Burke, J. F. (1981). Rapid and efficient cosmid cloning. *Nucleic Acids Res.* **9**, 2989-2998.

Isupov, M. N., Obmolova, G., Butterworth, S., Badet-Denisot, M. A., Badet, B., Polikarpov, I., Littlechild, J. A., and Teplyakov, A. (1996). Substrate binding is required for assembly of the active conformation of the catalytic site in Ntn amidotransferases: evidence from the 1.8 A crystal structure of the glutaminase domain of glucosamine 6-phosphate synthase. *Structure.* **4**, 801-810.

Jackson, D., Salem, A., and Coombs, G. H. (1984). The *in-vitro* activity of metronidazole against strains of *Escherichia coli* with impaired DNA repair systems. *J.Antimicrob.Chemother.* **13**, 227-236.

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

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

Joe, A., Murray, C. S., and McBride, B. C. (1994). Nucleotide sequence of a *Porphyromonas gingivalis* gene encoding a surface-associated glutamate dehydrogenase and construction of a glutamate dehydrogenase-deficient isogenic mutant. *Infect.Immun.* **62**, 1358-1368.

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

Johnson, J. L., Toth, J., Santiwatanakul, S., and Chen, J. S. (1997). Cultures of "*Clostridium acetobutylicum*" from various collections comprise *Clostridium acetobutylicum, Clostridium beijerinckii,* and two other distinct types based on DNA-DNA reassociation. *Int.J.Syst.Bacteriol.* **47**, 420-424.

Johnson, P. J. (1993) . Metronidazole and drug resistance. *Parasitol Today* **9**, 183-186.

Jones, D. T. and Keis, S. (1995). Origins and relationships of industrial solvent-producing clostridial strains. *FEMS Microbiol.Rev.* **17**, 221-230.

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

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

Jones, D. T. and Woods, D. R. (1986b). Gene transfer, recombination and gene cloning in *Clostridium acetobutylicum. Microbiol.Sci.* **3**, 19-22.

Jongsareejit, B., Rahman, R. N., Fujiwara, S., and Imanaka, T. (1997). Gene cloning, sequencing and enzymatic properties of glutamate synthase from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1. *Mol.Gen.Genet.* **254**, 635-642.

Kanamori, K., Weiss, R. L., and Roberts, J. D. (1987a). Ammonia assimilation in *Bacillus polymyxa.* 15N NMR and enzymatic studies. *J.Biol.Chem.* **262**, 11038-11045.

Kanamori, K., Weiss, R. L., and Roberts, J. D. (1987b). Role of glutamate dehydrogenase in ammonia assimilation in nitrogen- fixing *Bacillus macerans. J.Bacteriol* **169**, 4692-4695.

Kanamori, K., Weiss, R. L., and Roberts, J. D. (1988). Glutamate biosynthesis in *Bacillus azotofixans.* 15N NMR and enzymatic studies. *J.Biol.Chem.* **263**, 2817-2823.

Kanamori, K., Weiss, R. L., and Roberts, J. D. (1989). Ammonia assimilation pathways in nitrogen-fixing *Clostridium kluyverii* and *Clostridium butyricum. J.Bacteriol* **171**, 2148-2154.

Karplus, P. A., Daniels, M. J., and Herriott, J. R. (1991). Atomic structure of ferredoxin-NADP+ reductase: prototype for a structurally novel flavoenzyme family. *Science* **251**, 60-66.

Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., and Kikuchi, H. (1998). Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. *DNA Res.* **5**, 55-76.

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

Kendrick, K. E. and Wheelis, M. L. (1982). Histidine dissimilation in *Streptomyces coelicolor. J.Gen.Microbiol.* **128 (Pt 9)**, 2029-2040.

Kim, J. H., Krahn, J. M., Tomchick, D. R., Smith, J. L., and Zalkin, H. (1996). Structure and function of the glutamine phosphoribosylpyrophosphate amidotransferase glutamine site and communication with the phosphoribosylpyrophosphate site. *J.Biol.Chem.* **271**, 15549-15557.

Kleiner, D. (1979). Regulation of ammonium uptake and metabolism by nitrogen fixing bacteria. iii. *Clostridium pasteurianum. Arch.Microbiol.* **120**, 263-270.

Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., Richardson, D. L., Kerlavage, A. R., Graham, D. E., Kyrpides, N. C., Fleischmann, R. D., Quackenbush, J., Lee, N. H., Sutton, G. G., Gill, S., Kirkness, E. F., Dougherty, B. A., McKenney, K., Adams, M. D., Loftus, B., and Venter, J. C. (1997). The complete genome sequence of the hyperthermophilic, sulphate- reducing archaeon *Archaeoglobus fulgidus* [published erratum appears in Nature 1998 Jul 2;394(6688):101]. *Nature* **390**, 364-370.

Knaff, D. B., Hirasawa, M., Ameyibor, E., Fu, W., and Johnson, M. K. (1991). Spectroscopic evidence for a [3Fe-4S] cluster in spinach glutamate synthase. *J.Biol.Chem.* **266**, 15080-15084.

Kroening, T. A. and Kendrick, K. E. (1987). *In vivo* regulation of histidine ammonia-lyase activity from *Streptomyces griseus. J.Bacteriol* **169**, 823-829.

Kumada, Y., Takano, E., Nagaoka, K., and Thompson, C. J. (1990). *Streptomyces hygroscopicus* has two glutamine synthetase genes. *J.Bacteriol* **172**, 5343-5351.

Kusnan, M. B., Berger, M. G., and Fock, H. P. (1987). The involvement of glutamine synthetase/glutamate synthase in ammonia assimilation by *Aspergillus nidulans. J.Gen.Microbiol.* **133 (Pt 5)**, 1235-1242.

Kustu, S., Hirschman, J., Burton, D., Jelesko, J., and Meeks, J. C. (1984). Covalent modification of bacterial glutamine synthetase: physiological significance. *Mol.Gen.Genet.* **197**, 309-317.

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

Landgraf, J. R., Wu, J., and Calvo, J. M. (1996). Effects of nutrition and growth rate on Lrp levels in *Escherichia coli. J.Bacteriol* **178**, 6930-6936.

Lewis, T. A., Gonzalez, R., and Botsford, J. L. (1990). *Rhizobium meliloti* glutamate synthase: cloning and initial characterization of the *glt* locus. *J.Bacteriol.* **172**, 2413-2420.

Liebs, P., Riedel, K., Graba, J. P., Schrapel, D., and Tischler, U. (1988). Formation of some extracellular enzymes during the exponential growth of *Bacillus subtilis. Folia Microbiol.(Praha)* **33**, 88-95.

Lin, J. T. and Stewart, V. (1998). Nitrate assimilation by bacteria. *Adv.Microb.Physiol* **39**, 1-30, 379.

Lindmark, D. G. and Muller, M. (1976). Antitrichomonad action, mutagenicity, and reduction of metronidazole and other nitroimidazoles. *Antimicrob.Agents Chemother.* **10**, 476-482.

Lomnitz, A., Calderon, J., Hernandez, G., and Mora, J. (1987). Functional Analysis of ammonium assimilation enzymes in *Neurospora crassa. J.Gen.Microbiol.* **133**, 2333-2340.

Long, S., Jones, D. T., and Woods, D. R. (1983). Sporulation of *Clostridium acetobutylicum* P262 in a defined medium. *Appl.Environ.Microbiol.* **45**, 1389-1393.

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

Long, S., Jones, D. T., and Woods, D. R. (1984b). The relationship between sporulation and solvent production in *Clostridium acetobutylicum* P262. *Biotechnol Lett* **6**, 529-534.

Lowe, N., Rice, P. M., and Drew, R. E. (1989). Nucleotide sequence of the aliphatic amidase regulator gene (*amiR*) of *Pseudomonas aeruginosa. FEBS Lett* **246**, 39-43.

Macaluso, A., Best, E. A., and Bender, R. A. (1990). Role of the *nac* gene product in the nitrogen regulation of some NTR- regulated operons of *Klebsiella aerogenes. J.Bacteriol* **172**, 7249-7255.

Manoil, C. and Beckwith, J. (1985). TnphoA: a transposon probe for protein export signals. *Proc.Natl.Acad.Sci.U.S.A* **82**, 8129-8133.

Marmur, J. (1961). A procedure for the isolation of deoxyribonucleic acid from microorganisms. *J.Mol.Biol.* **3**, 208-218.

Marques, S., Florencio, F. J., and Candau, P. (1992). Purification and characterization of the ferredoxin-glutamate synthase from the unicellular cyanobacterium *Synechococcus* sp. PCC 6301. *Eur.J.Biochem.* **206**, 69-77.

Matsuoka, K. and Kimura, K. (1986). Glutamate synthase from *Bacillus subtilis* PCI 219. *J.Biochem.(Tokyo)* **99**, 1087-1100.

Meers, J. L., Tempest, D. W., and Brown, C. M. (1970). 'Glutamine(amide):2-oxoglutarate amino transferase oxido-reductase (NADP); an enzyme involved in the synthesis of glutamate by some bacteria. *J.Gen.Microbiol.* **64**, 187-194.

Mei, B. and Zalkin, H. (1989). A cysteine-histidine-aspartate catalytic triad is involved in glutamine amide transfer function in PurF-type glutamine amidotransferases. *J.Biol.Chem.* **264**, 16613-16619.

Mei, B. G. and Jiao, R. S. (1988). Purification and properties of glutamate synthase from *Nocardia mediterranei. J.Bacteriol.* **170**, 1940-1944.

Meister, A. (1985). Glutamate synthase from *Escherichia coli, Klebsiella aerogenes*, and *Saccharomyces cerevisae. Methods Enzymol.* **113**, 327-337.

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

Mermelstein, L. D., Welker, N. E., Bennett, G. N., and Papoutsakis, E. T. (1992). Expression of cloned homologous fermentative genes in *Clostridium acetobutylicum* ATCC 824. *Biotechnology (N.Y.)* **10**, 190-195.

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

Messing, J. (1983). New M13 vectors for cloning. *Methods Enzymol.* **101**, 20-78.

Miller, R. E. and Stadtman, E. R. (1972). Glutamate synthase from *Escherichia coli.* An iron-sulfide flavoprotein. *J.Biol.Chem.* **247**, 7407-7419.

Minton, N. P., Brehm, J. K., Swinfield, T. J., Whelan, S. M., Mauchline, M. L., Bodsworth, N., and Oultram, J. D. (1993). Clostridial cloning vectors. In *The Clostridia and Biotechnology*. pp. 119-50. Butterworth-Heinemann: Stoneham, MA.

Mitchell, W. J. (1998). Physiology of carbohydrate to solvent conversion by clostridia. *Adv.Microb.Physiol* **39**, 31-130.

Monot, F. and Engasser, J. M. (1983). Production of acetone and butanol by batch and continuous culture of *Clostridium acetobutylicum* under nitrogen limitation. *Biotech.Lett.* **5**, 213-218.

Monot, F., Martin, J.-R., Petitdemange, H., and Gay, R. (1982). Acetone and butanol production by *Clostridium acetobutylicum* in a synthetic medium. *Appl.Environ.Microbiol.* **44**, 1318-1324.

Moran, C. P., Jr., Lang, N., LeGrice, S. F., Lee, G., Stephens, M., Sonenshein, A. L., Pero, J., and Losick, R. (1982). Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. *Mol.Gen.Genet.* **186**, 339-346.

Morandi, P., Valzasina, B., Colombo, C., Curti, B., and Vanoni, M. A. (2000). Glutamate synthase: identification of the NADPH-binding site by site- directed mutagenesis. *Biochemistry* **39**, 727-735.

Muller, M. (1983). Mode of action of metronidazole on anaerobic bacteria and protozoa. *Surgery* **93**, 165-171.

Muro-Pastor, M. I., Reyes, J. C., and Florencio, F. J. (1996). The NADP+-isocitrate dehydrogenase gene (*icd*) is nitrogen regulated in cyanobacteria. *J.Bacteriol* **178**, 4070-4076.

Nakano, M. M., Yang, F., Hardin, P., and Zuber, P. (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, Y. and Kimura, K. (1991). Purification and characterization of a repressor for the *Bacillus cereus glnRA* operon. *J.Biochem.(Tokyo)* **109**, 223-228.

Nakotte, S., Schaffer, S., Bohringer, M., and Dürre, P. (1998). Electroporation of, plasmid isolation from and plasmid conservation in *Clostridium acetobutylicum* DSM 792. *Appl.Microbiol.Biotechnol.* **50**, 564-567.

Narikawa, S. (1986). Distribution of metronidazole susceptibility factors in obligate anaerobes. *J.Antimicrob.Chemother.* **18**, 565-574.

Navarro, F., Chavez, S., Candau, P., and Florencio, F. J. (1995). Existence of two ferredoxin-glutamate synthases in the cyanobacterium *Synechocystis* sp. PCC 6803. Isolation and insertional inactivation of *gltB* and *gltS* genes. *Plant Mol.Biol.* **27**, 753-767.

Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Nelson, W. C., Ketchum, K. A., McDonald, L., Utterback, T. R., Malek, J. A., Linher, K. D., Garrett, M. M., Stewart, A. M., Cotton, M. D., Pratt, M. S., Phillips, C. A., Richardson, D., Heidelberg, J., Sutton, G. G., Fleischmann, R. D., Eisen, J. A., and Fraser, C. M. (1999). Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323-329.

Newman, E. B., D'Ari, R., and Lin, R. T. (1992). The leucine-Lrp regulon in *E. coli*: a global response in search of a raison d'etre. *Cell* **68**, 617-619.

Newman, E. B., Lin, R. T., and D'Ari, R. (1996). The leucine/Lrp regulon. In *Escherichia coli and Salmonella typhimurium: cellular and molecular biology.* pp. 1513-25. American Society for Microbiology: Washington D.C.

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

O'Hara, B. P., Norman, R. A., Wan, P. T., Roe, S. M., Barrett, T. E., Drew, R. E., and Pearl, L. H. (1999). Crystal structure and induction mechanism of AmiC-AmiR: a ligand- regulated transcription antitermination complex. *EMBO J.* **18**, 5175-5186.

Okuhara, H., Matsumura, T., Fujita, Y., and Hase, T. (1999). Cloning and inactivation of genes encoding ferredoxin- and NADH-dependent glutamate synthases in the cyanobacterium *Plectonema boryanum.* Imbalances in nitrogen and carbon assimilations caused by deficiency of the ferredoxin-dependent enzyme. *Plant Physiol* **120**, 33-42.

Oliver, G., Gosset, G., Sanchez-Pescador, R., Lozoya, E., Ku, L. M., Flores, N., Becerril, B., Valle, F., and Bolivar, F. (1987). Determination of the nucleotide sequence for the glutamate synthase structural genes of *Escherichia coli* K-12. *Gene* **60**, 1-11.

Oliver, G., Valle, F., Rosetti, F., Gomez-Pedrozo, M., Santamaria, P., Gosset, G., and Bolivar, F. (1985). A common precursor for the two subunits of the penicillin acylase from *Escherichia coli* ATCC11105. *Gene* **40**, 9-14.

Pan, F. L. and Coote, J. G. (1979). Glutamine synthetase and glutamate synthase activities during growth and sporulation in *Bacillus subtilis. J.Gen.Microbiol.* **112**, 373-377.

Paress, P. S. and Streicher, S. L. (1985). Glutamine synthetase of *Streptomyces cattleya*: purification and regulation of synthesis. *J.Gen.Microbiol.* **131 (Pt 8)**, 1903-1910.

Paulus, H. (1993). Biosynthesis of the asparate family of amino acids. In *Bacillus subtilis and other gram-positve bacteria: biochemistry, physiology, and molecular genetics.* American Society for Microbiology: Washington D.C.

Pelanda, R., Vanoni, M. A., Perego, M., Piubelli, L., Galizzi, A., Curti, B., and Zanetti, G. (1993). Glutamate synthase genes of the diazotroph *Azospirillum brasilense.* Cloning, sequencing, and analysis of functional domains. *J.Biol.Chem.* **268**, 3099-3106.

Platko, J. V. and Calvo, J. M. (1993). Mutations affecting the ability of *Escherichia coli* Lrp to bind DNA, activate transcription, or respond to leucine. *J.Bacteriol* **175**, 1110-1117.

Prescott, S. C. and Dunn, C. G. (1959). 'Industrial Microbiology.' 3 Edn. McGraw-Hill: New York.

Ratti, S., Curti, B., Zanetti, G., and Galli, E. (1985). Purification and characterization of glutamate synthase from *Azospirillum brasilense. J.Bacteriol* **163**, 724-729.

Rawlings, D. E., Jones, W. A., O'Neill, E. G., and Woods, D. R. (1987). Nucleotide sequence of the glutamine synthetase gene and its controlling region from the acidophilic autotroph *Thiobacillus ferrooxidans. Gene* **53**, 211-217.

Reitzer, L. J. (1996). Ammonia assimilation and the biosynthesis of glutamine, glutamate, aspartate, asparagine, L-alanine and D-alanine. In *Escherichia coli and Salmonella typhimurium: cellular and molecular biology, 2nd Edition.* pp. 391-407. ASM Press: Washington DC.

Reitzer, L. J. and Magasanik, B. (1987). Ammonia assimilation and the biosynthesis of glutamine, glutamate, asparate, asparagine, L-alanine and D-alanine. In *Escherichia coli and Salmonella typhimurium: cellular and molecular biology.* pp. 302-20. ASM Press: Washington DC.

Rendina, A. R. and Orme-Johnson, W. H. (1978). Glutamate synthase: on the kinetic mechanism of the enzyme from *Escherichia coli* W. *Biochemistry* **17**, 5388-5393.

Reysset, G. and Sebald, M. (1993). Transformation/electrotransformation of clostridia. In *The Clostridia and Biotechnology.* pp. 151-6. Butterworth-Heinemann: Stoneham, MA.

Rogers, P. and Gottschalk, G. (1993). Biochemistry and regulation of acid and solvent productions in clostridia. In *The Clostridia and Biotechnology.* pp. 201-25. Butterworth-Heinemann: Stoneham, MA.

Roos, J. W., McLaughlin, J. K., and Papoutsakis, E. T. (1985). The effect of pH on nitrogen supply, cell lysis and solvent production in fermentations of *Clostridium acetobutylicum. Biotechnol.Bioeng.* **27**, S401-S403.

Rose, R. E. (1988). The nucleotide sequence of pACYC184. *Nucleic Acids Res.* **16**, 355.

Sakakibara, H., Watanabe, M., Hase, T., and Sugiyama, T. (1991). Molecular cloning and characterization of complementary DNA encoding for ferredoxin-dependent glutamate synthase in maize leaf. *J.Biol.Chem.* **266**, 2028-2035.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). 'Mollecular cloning: a laboratory manual.' 2nd Edn. Cold Spring Harbour Laboratory: New York.

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

Santangelo, J. D., Jones, D. T., and Woods, D. R. (1991). Metronidazole activation and isolation of *Clostridium acetobutylicum* electron transport genes. *J.Bacteriol.* **173**, 1088-1095.

Santangelo, J. D., Kuhn, A., Treuner-Lange, A., and Dürre, P. (1998). Sporulation and time course expression of sigma-factor homologous genes in *Clostridium acetobutylicum. FEMS Microbiol.Lett.* **161**, 157-164.

Saroja, G. N. and Gowrishankar, J. (1996). Roles of SpoT and FNR in NH4+ assimilation and osmoregulation in GOGAT (glutamate synthase)-deficient mutants of *Escherichia coli. J.Bacteriol* **178**, 4105-4114.

Sauer, U., Treuner, A., Buchholz, M., Santangelo, J. D., and Dürre, P. (1994). Sporulation and primary sigma factor homologous genes in *Clostridium acetobutylicum. J.Bacteriol.* **176**, 6572-6582.

Schell, M. A. (1993). Molecular biology of the LysR family of transcriptional regulators. *Annu.Rev.Microbiol.* 47, 597-626.

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 genetics.* pp. 281-98. American Society for Microbiology: Washington D.C.

Schreier, H. J. and Bernlohr, R. W. (1984). Purification and properties of glutamate synthase from *Bacillus licheniformis*. *J.Bacteriol.* **160**, 591-599.

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

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

Schreier, H. J., Smith, T. M., and Bernlohr, R. W. (1982). Regulation of nitrogen catabolic enzymes in *Bacillus* spp. *J.Bacteriol* **151**, 971-975.

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

Schwacha, A. and Bender, R. A. (1993a). The *nac* (nitrogen assimilation control) gene from *Klebsiella aerogenes*. *J.Bacteriol* **175**, 2107-2115.

Schwacha, A. and Bender, R. A. (1993b). The product of the *Klebsiella aerogenes nac* (nitrogen assimilation control) gene is sufficient for activation of the *hut* operons and repression of the *gdh* operon. *J.Bacteriol* **175**, 2116-2124.

Scrutton, N. S., Berry, A., and Perham, R. N. (1990). Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. *Nature* **343**, 38-43.

Senior, P. J. (1975). Regulation of nitrogen metabolism in *Escherichia coli* and *Klebsiella aerogenes*: studies with the continuous-culture technique. *J.Bacteriol* **123**, 407-418.

Shaheen, R., Shirley, M., and Jones, D. T. (2000). Comparative fermentation studies of industrial strains belonging to four species of solvent-producing clostridia. *J.Mol.Microbiol.Biotechnol* **2**, 115-124.

Shapiro, B. M. and Stadtman, E. R. (1968). Glutamine synthetase deadenylylating enzyme. *Biochem.Biophys.Res.Commun.* **30**, 32-37.

Shapiro, S. and Vining, L. C. (1983). Nitrogen metabolism and chloramphenicol production in *Streptomyces venezuelae*. *Can.J.Microbiol.* **29**, 1706-1714.

Singhal, R. K., Krishnan, I. S., and Dua, R. D. (1989). Stabilization, purification, and characterization of glutamate synthase from *Clostridium pasteurianum*. *Biochemistry* **28**, 7928-7935.

Somerville, C. R. and Ogren, W. L. (1982). Mutants of the cruciferous plant *Arabidopsis thaliana* lacking glycine decarboxylase activity. *Biochem.J.* **202**, 373-380.

Somerville, J. E., Shatters, R. G., and Kahn, M. L. (1989). Isolation, characterization, and complementation of *Rhizobium meliloti* 104A14 mutants that lack glutamine synthetase II activity. *J.Bacteriol* **171**, 5079-5086.

Southern, J. A., Parker, J. R., and Woods, D. R. (1989). Novel structure, properties and inactivation of glutamine synthetase cloned from *Bacteroides fragilis* . *J.Gen.Microbiol.* **133**, 2437-2446.

Stabile, H., Curti, B., and Vanoni, M. A. (2000). Functional properties of recombinant *Azospirillum brasilense* glutamate synthase, a complex iron-sulfur flavoprotein. *Eur.J.Biochem.* **267**, 2720-2730.

Stewart, V. (1993). Nitrate regulation of anaerobic respiratory gene expression in *Escherichia coli. Mol.Microbiol.* **9**, 425-434.

Stock, J. B., Ninfa, A. J., and Stock, A. M. (1989). Protein phosphorylation and regulation of adaptive responses in bacteria. *Microbiol.Rev.* **53**, 450-490.

Strauch, M. A., Aronson, A. I., Brown, S. W., Schreier, H. J., and Sonenhein, A. L. (1988). Sequence of the *Bacillus subtilis* glutamine synthetase gene region. *Gene* **71**, 257-265.

Suzuki, A. and Rothstein, S. (1997). Structure and regulation of ferredoxin-dependent glutamase synthase from *Arabidopsis thaliana.* Cloning of cDNA expression in different tissues of wild-type and *gltS* mutant strains, and light induction. *Eur.J.Biochem.* **243**, 708-718.

Switzer, R. L., Turner, R. J., and Lu, Y. (1999). Regulation of the *Bacillus subtilis* pyrimidine biosynthetic operon by transcriptional attenuation: control of gene expression by an mRNA- binding protein. *Prog.Nucleic Acid Res.Mol.Biol.* **62**, 329-367.

Tempest, D. W., Meers, J. L., and Brown, C. M. (1970). Synthesis of glutamate in *Aerobacter aerogenes* by a hitherto unknown route. *Biochem.J.* **117**, 405-407.

Temple, S. J., Bagga, S., and Sengupta-Gopalan, C. (1998). Down-regulation of specific members of the glutamine synthetase gene family in alfalfa by antisense RNA technology. *Plant Mol.Biol.* **37**, 535-547.

Tiboni, O., Cammarano, P., and Sanangelantoni, A. M. (1993). Cloning and sequencing of the gene encoding glutamine synthetase I from the archaeum *Pyrococcus woesei*: anomalous phylogenies inferred from analysis of archaeal and bacterial glutamine synthetase I sequences. *J.Bacteriol* **175**, 2961-2969.

Trotta, P. P., Platzer, K. E., Haschemeyer, R. H., and Meister, A. (1974). Glutamine-binding subunit of glutamate synthase and partial reactions catalyzed by this glutamine amidotransferase. *Proc.Natl.Acad.Sci.U.S.A* **71**, 4607-4611.

Tyler, B. (1978). Regulation of the assimilation of nitrogen compounds. *Annu.Rev.Biochem.* **47**, 1127-1162.

Usden, K. P., Zappe, H., Jones, D. T., and Woods, D. R. (1986). Cloning, expression and purification of glutamine synthetase from *Clostridium acetobutylicum. Appl.Environ.Microbiol.* **52**, 413-419.

Valentin, K., Kostrzewa, M., and Zetsche, K. (1993). Glutamate synthase is plastid-encoded in a red alga: implications for the evolution of glutamate synthases. *Plant Mol.Biol.* **23**, 77-85.

Valenzuela, L., Ballario, P., Aranda, C., Filetici, P., and Gonzalez, A. (1998). Regulation of expression of *GLT1*, the gene encoding glutamate synthase in *Saccharomyces cerevisae. J.Bacteriol.* **180**, 3533-3540.

Valenzuela, L., Guzman-Leon, S., Coria, R., Ramirez, J., Aranda, C., and Gonzalez, A. (1995). A NADP-glutamate dehydrogenase mutant of the petit-negative yeast *Kluyveromyces lactis* uses the glutamine synthetase-glutamate synthase pathway for glutamate biosynthesis. *Microbiology* **141 (Pt 10)**, 2443-2447.

van Heeswijk, W. C., Hoving, S., Molenaar, D., Stegeman, B., Kahn, D., and Westerhoff, H. V. (1996). An alternative PII protein in the regulation of glutamine synthetase in *Escherichia coli*. *Mol.Microbiol.* **21**, 133-146.

van Heeswijk, W. C., Rabenberg, M., Westerhoff, H. V., and Kahn, D. (1993). The genes of the glutamine synthetase adenylylation cascade are not regulated by nitrogen in *Escherichia coli*. *Mol.Microbiol.* **9**, 443-457.

van Heeswijk, W. C., Stegeman, B., Hoving, S., Molenaar, D., Kahn, D., and Westerhoff, H. V. (1995). An additional PII in *Escherichia coli*: a new regulatory protein in the glutamine synthetase cascade. *FEMS Microbiol.Lett* **132**, 153-157.

Vanoni, M. A., Accornero, P., Carrera, G., and Curti, B. (1994a). The pH-dependent behavior of catalytic activities of *Azospirillum brasilense* glutamate synthase and iodoacetamide modification of the enzyme provide evidence for a catalytic Cys-His ion pair. *Arch.Biochem.Biophys.* **309**, 222-230.

Vanoni, M. A. and Curti, B. (1999). Glutamate synthase: a complex iron-sulfur flavoprotein. *Cell Mol.Life Sci.* **55**, 617-638.

Vanoni, M. A., Edmondson, D. E., Rescigno, M., Zanetti, G., and Curti, B. (1991). Mechanistic studies on *Azospirillum brasilense* glutamate synthase. *Biochemistry* **30**, 11478-11484.

Vanoni, M. A., Edmondson, D. E., Zanetti, G., and Curti, B. (1992). Characterization of the flavins and the iron-sulfur centers of glutamate synthase from *Azospirillum brasilense* by absorption, circular dichroism, and electron paramagnetic resonance spectroscopies. *Biochemistry* **31**, 4613-4623.

Vanoni, M. A., Fischer, F., Ravasio, S., Verzotti, E., Edmondson, D. E., Hagen, W. R., Zanetti, G., and Curti, B. (1998). The recombinant alpha subunit of glutamate synthase: spectroscopic and catalytic properties. *Biochemistry* **37**, 1828-1838.

Vanoni, M. A., Mazzoni, A., Fumagalli, P., Negri, A., Zanetti, G., and Curti, B. (1994b). Interdomain loops and conformational changes of glutamate synthase as detected by limited proteolysis. *Eur.J.Biochem.* **226**, 505-515.

Vanoni, M. A., Verzotti, E., Zanetti, G., and Curti, B. (1996). Properties of the recombinant beta subunit of glutamate synthase. *Eur.J.Biochem.* **236**, 937-946.

Wang, Q. and Calvo, J. M. (1993). Lrp, a global regulatory protein of *Escherichia coli*, binds co- operatively to multiple sites and activates transcription of *ilvIH*. *J.Mol.Biol.* **229**, 306-318.

Wang, Q., Sacco, M., Ricca, E., Lago, C. T., De Felice, M., and Calvo, J. M. (1993). Organization of Lrp-binding sites upstream of *ilvIH* in *Salmonella typhimurium*. *Mol.Microbiol.* **7**, 883-891.

Weisberg, R. A. and Gottesman, M. E. (1999). Processive antitermination. *J.Bacteriol* **181**, 359-367.

Wierenga, R. K., De Maeyer, M. C. H., and Hol, W. G. J. (1985). Interaction of the pyrophosphate moieties with α–helixes in dinucleotide binding proteins. *Biochemistry* **24**, 1346-1357.

Wiese, D. E., Ernsting, B. R., Blumenthal, R. M., and Matthews, R. G. (1997). A nucleoprotein activation complex between the leucine-responsive regulatory protein and DNA upstream of the *gltBDF* operon in *Escherichia coli*. *J.Mol.Biol.* **270**, 152-168.

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

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

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

Williams, D. R., Young, D. I., and Young, M. (1990). Conjugative plasmid transfer from *Escherichia coli* to *Clostridium acetobutylicum*. *J.Gen.Microbiol.* **136 (Pt 5)**, 819-826.

Willins, D. A., Ryan, C. W., Platko, J. V., and Calvo, J. M. (1991). Characterization of Lrp, an *Escherichia coli* regulatory protein that mediates a global response to leucine. *J.Biol.Chem.* **266**, 10768-10774.

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

Wong, J., Sass, C., and Bennett, G. N. (1995). Sequence and arrangement of genes encoding sigma factors in *Clostridium acetobutylicum* ATCC 824. *Gene* **153**, 89-92.

Wood, W. I., Gitschier, J., Lasky, L. A., and Lawn, R. M. (1985). Base composition-independent hybridization in tetramethylammonium chloride: a method for oligonucleotide screening of highly complex gene libraries. *Proc.Natl.Acad.Sci.U.S.A* **82**, 1585-1588.

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

Woods, D. R. and Reid, S. J. (1993). Recent developments on the regulation and structure of glutamine synthetase enzymes from selected bacterial groups. *FEMS Microbiol.Rev.* **11**, 273-283.

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

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

Wray, L. V., Jr., Atkinson, M. R., and Fisher, S. H. (1991). Identification and cloning of the *glnR* locus, which is required for transcription of the *glnA* gene in *Streptomyces coelicolor* A3(2). *J.Bacteriol* **173**, 7351-7360.

Wray, L. V., Jr., Atkinson, M. R., and Fisher, S. H. (1994). The nitrogen-regulated *Bacillus subtilis nrgAB* operon encodes a membrane protein and a protein highly similar to the *Escherichia coli glnB*-encoded PII protein. *J.Bacteriol* **176**, 108-114.

Wray, L. V., Jr., Ferson, A. E., and Fisher, S. H. (1997). Expression of the *Bacillus subtilis ureABC* operon is controlled by multiple regulatory factors including CodY, GlnR, TnrA, and Spo0H. *J.Bacteriol* **179,** 5494-5501.

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

Wray, L. V., Jr. and Fisher, S. H. (1993). The *Streptomyces coelicolor glnR* gene encodes a protein similar to other bacterial response regulators. *Gene* **130,** 145-150.

Yamamoto, I., Abe, A., Saito, H., and Ishimoto, M. (1984). The pathway of ammonia assimilation in *Bacteroides fragilis. J.Gen.Microbiol.* **30,** 499-508.

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

Yeung, T. C., Beaulieu, B. B., Jr., McLafferty, M. A., and Goldman, P. (1984). Interaction of metronidazole with DNA repair mutants of *Escherichia coli. Antimicrob.Agents Chemother.* **25,** 65-70.

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

Youngleson, J. S., Santangelo, J. D., Jones, D. T., and Woods, D. R. (1988). Cloning and expression of a *Clostridium acetobutylicum* alcohol dehydrogenase gene in *Escherichia coli. Appl.Environ.Microbiol.* **54,** 676-682.

Zabeau, M. and Stanley, K. K. (1982). Enhanced expression of cro-β-galactosidase fusion proteins under the control of the P_R promoter of bacteriophage λ. *EMBO J.* **1,** 1217-1224.

Zappe, H., Jones, D. T., and Woods, D. R. (1986). Cloning and expression of *Clostridium acetobutylicum* endoglucanase, cellobiase and amino acid biosynthesis genes in *Escherichia coli. J.Gen.Microbiol.* **132 (Pt 5),** 1367-1372.

Zappe, H., Jones, W. A., Jones, D. T., and Woods, D. R. (1988). Structure of an endo-beta-1,4-glucanase gene from *Clostridium acetobutylicum* P262 showing homology with endoglucanase genes from *Bacillus* spp. *Appl.Environ.Microbiol.* **54,** 1289-1292.