L-Arginine Overproduction in *Corynebacterium glutamicum* ATCC 13032

by

Grant de V Theron

Thesis presented for the degree of Doctor of Philosophy

In the Department of Molecular and Cell Biology

University of Cape Town

South Africa

16 February 2009

I hereby:

(a) grant the University free license to reproduce the above thesis in whole or in part, for the purpose of research;

(b) declare that:

(i) the above thesis is my own unaided work, both in conception and execution, and that apart from the normal guidance of my supervisor, I have received no assistance apart from that stated below;

(ii) except as stated below, neither the substance or any part of the thesis has been submitted in the past, or is being, or is to be submitted for a degree at this University or any other University.

(iii) I am now presenting the thesis for examination for the Degree of PhD.

Signed

Acknowledgements

I am indebted to my parents - without their support and sacrifice this work would not have been possible.

Thank you to my supervisor, Assoc. Prof. Sharon Reid, for her constant guidance, encouragement and good humour.

Thank you to Nerson Harry for technical assistance with the pCAF21404-ARGC and pLYSE experiments.

Thank you to Pei-yin Liebrich for her oligonucleotide synthesis expertise.

Thank you to Diane James for her invaluable help with DNA sequencing and the primer extension work.

Thank you to the National Research Foundation of South Africa for funding.

Lastly, thank you to all of my friends and colleagues in the Department of Molecular and Cell Biology at the University of Cape Town.

Contents

L-Arginine Overproduction in *Corynebacterium glutamicum* ATCC 13032

Abbreviations

°C	degree(s) Celsius
A	adenine
ACDA	acetylcitrulline deacetylase
AMP	adenosine 5'-monophosphate
Amp	ampicillin
AODA	acetylornithine deacetylase
AOTC	acetylornithine transcarbamylase
ASL	argininosuccinate lyase
ASS	argininosuccinate synthetase
AST	aspartate transaminase
ATCC	American Type Culture Collection
ATP	adenosine 5'-triphosphate
bp	base pair(s)
BSA	bovine serum albumin
C-	carboxy-(terminal)
C	cytosine
ca.	*circa*
cDNA	complementary DNA
CFU	colony forming units
cm	centimetre(s)

CMN	*Corynebacterium-Mycobacterium-Nocardia*
CoA	coenzyme A
CPS	carbamoylphosphate synthetase
CS	citrate synthase
CSPD	disodium 3-(4-methoxyspiro[1,2-dioxetane-3,2'-(5'-chloro) tricycle{3.3.1.1^{3,7}} decan]-4-yl) phenyl phosphate
Cy5	cyanine 5
D	adenine, guanine or thymine nucleobase
D-	dextrorotatory
d-	distilled
DBD	DNA-binding domain
DCO	double cross-over
dd-	dideoxynucleotide(s)
DEPC	diethylpyrocarbonate
DIG	digoxigenin-11- (2'-deoxyuridine 5'-triphosphate)
DNA	deoxyribonucleic acid
dNTP	adenosine, cytosine, guanosine and thymidine
dsDNA	double-stranded DNA
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid

EMSA	electrophoretic mobility shift assay
et al.	*et alia*
fmol	femtomole(s)
g	acceleration due to gravity
g	gram(s)
G	guanine
G+C	guanine and cytosine
GDH	glutamate dehydrogenase
H	adenine, cytosine or thymine
h	hour(s)
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HR	homologous recombination
HTH	helix-turn-helix
Hyg	hygromycin B
I	inosine
ICD	isocitrate dehydrogenase
IPTG	isopropyl ß-D-1-thiogalactopyranoside
K	guanine or thymine
Kan	kanamycin
kb	kilobase(s)
kcal	kilocalorie(s)

K_d	dissociation constant
kDa	kilodalton(s)
KGDH	α-ketoglutarate dehydrogenase complex
K_m	Michaelis-Menten constant
kV	kilovolt(s)
L-	levorotatory
LTTR	LysR-type transcriptional regulators
M	adenine or cytosine
M	molar
MCS	multiple cloning site
mg	milligram(s)
MIC	minimum inhibitory concentration
min	minute(s)
mJ	millijoule(s)
ml	millilitre(s)
mm	millimetre(s)
mM	millimolar
MM	minimal medium
MOPS	3-(*N*-morpholino)-propanesulfonic acid
ms	millisecond(s)
MSG	monosodium L-glutamate

MS-HPLC	mass spectrometer high performance liquid chromatography
N-	amino-(terminal)
N	normal
NADPH	nicotinamide adenine dinucleotide phosphate
NAGK	*N*-acetylglutamate kinase
NAGS	*N*-acetylglutamate synthase
NCBI	National Centre for Bioinformatics
nm	nanometre(s)
nM	nanomolar
nmol	nanomole(s)
NO	nitric oxide
Nx	nalidixic acid
OAT	ornithine acetyltransferase
OD_x	optical density at *x* nanometres
OTC	ornithine transcarbamylase
P	p-value; indicating probability
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PCx	pyruvate carboxylase
PDHC	pyruvate dehydrogenase complex

PEP	phosphoenolpyruvate
PEPCx	PEP carboxylase
PMF	proton motor force
pmol	picomole(s)
PMSF	phenylmethylsulfonyl fluoride
PPP	pentose phosphate pathway
PTS	phosphotransferase system
PWM	position weight matrix
qRT-PCR	quantitative real-time PCR
RBS	ribosomal binding site
RNA	ribonucleic acid
rpm	revolution(s) per minute
rRNA	ribosomal RNA
RT	reverse transcriptase
RT-PCR	semi-quantitative reverse transcriptase PCR
S	cytosine or guanine
s	second(s)
S	svedberg
SCO	single cross-over
SDS	sodium dodecyl sulfate
SEM	standard error of the mean

spp.	species
T	thymine
T_a	annealing temperature
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
TCA	tricarboxylic acid
TE	Tris-EDTA
TMB	3,3',5,5'-tetramethylbenzidine
Tris	trishydroxymethylaminomethane
U	unit(s)
UTP	uridine-5'-triphosphate
UV	ultraviolet
V	volt(s)
v/v	volume per volume
W	adenine or thymine
w/v	weight per volume
wHTH	winged helix-turn-helix
WT	wildtype; *Corynebacterium glutamicum* ATCC 13032
X-gal	5-bromo-4-chloro-3-indolyl-ß-D-galactopyranoside
x^R	resistant to *x*

x^S	sensitive to *x*
Y	cytosine or thymine
YT	yeast tryptone
Δ	delta; typically indicating a mutant allele
ΔG°	Gibbs free energy
µF	microfarad(s)
µg	microgram(s)
µl	microlitre(s)
µM	micromolar
μ_{max}	maximum growth rate
Ω	ohm(s)

Summary

Corynebacterium glutamicum is widely used for the commercial production of a variety of amino acids, including L-lysine, L-glutamate and L-threonine. With the exception of L-arginine, the biosynthesis and regulation of most of these compounds in this bacterium are relatively well characterised in the literature. The research presented here focuses on improving our understanding of the regulation of L-arginine biosynthesis in *C. glutamicum*. This was performed with the ultimate goal of creating strains capable of producing L-arginine commercially.

A novel gene replacement system was initially used for the directed mutation of the L-arginine biosynthetic gene cluster in *C. glutamicum* ATCC 13032. This was met with limited success, however, and the pK19*mobsacB* vector was thus adopted for further mutagenesis of this region. Using this system, a chromosomal deletion of the gene expected to encode ArgR (predicted to be the main transcriptional regulator of L-arginine metabolism) was introduced into this bacterium. This *argR* mutant strain, together with several others also created as part of this study, was compared to the wildtype in terms of its growth, L-arginine and L-citrulline production capacities and resistance to L-canavanine.

The *argR* integration mutant strains used in this strain all appeared to contain stably integrated recombinant DNA and, in addition to the *argR* deletion mutant, displayed a growth rate comparable to that of the wildtype. In contrast, a strain with an insertion in its *argH* gene (which encodes an enzyme critical for the completion of the L-arginine biosynthetic pathway) proved to be very unstable and the presence of the integrated DNA was found to be associated with a weak growth phenotype. The L-arginine biosynthetic gene cluster was shown by semi-quantitative reverse transcriptase polymerase chain reaction to be transcribed as a single polycistronic *argCJBDFRGH* transcript. Furthermore, upon the deletion of *argR* in *C. glutamicum* ATCC 13032, transcriptional repression of this gene cluster in the presence of L-arginine appeared to have been alleviated. As expected, this translated into increased L-arginine biosynthesis in the mutant strain, where this compound accumulated to approximately 5-fold more than the wildtype strain. Additionally, this mutant strain also accumulated L-citrulline at levels approximately 8-fold more than that observed with the wildtype. The increased quantities of L-arginine produced by the *argR* deletion mutant, in addition to those

produced by the other mutant strains in this study, were lastly shown to be correlated with increased resistance to the toxic L-arginine analogue L-canavanine.

Having confirmed the role of ArgR as an important regulator of L-arginine biosynthesis in *C. glutamicum* ATCC 13032, we next proceeded to perform a preliminary characterisation of the ArgR-DNA interaction using electrophoretic mobility shift assays (EMSAs). As expected, a recombinant, hexahistidine-tagged derivative of ArgR was shown to bind *in vitro* in an L-arginine-dependent manner to the *argC* promoter region preceding the L-arginine biosynthetic gene cluster. Binding was also demonstrated for the promoter region of the *carA* gene, which is expected to be partially responsible for the supply of carbamoyl phosphate into the L-arginine biosynthetic pathway in *C. glutamicum* ATCC 13032. By incorporating competitor oligonucleotides into the EMSAs, we were able to locate putative ArgR binding sites, known as ARG boxes, within the promoter regions of the *argC* and *carA* genes. A pair of putative ARG box motifs, separated by a single base pair, was detected with the *argC* region and, via primer extension analysis, was shown to be located in close proximity to the *argCJBDFRGH* transcriptional start site. In contrast, although ArgR was found to exhibit comparable affinities for both the *argC* and *carA* promoter regions, the *latter* was found to contain only a single putative ARG box sequence. A preliminary *C. glutamicum* ATCC 13032 ARG box consensus sequence was constructed from these results and closely resembled the *Escherichia coli* K12 ARG box consensus sequence. Using this latter sequence as part of an *in silico* approach, we were thus able to detect additional putative ARG box sites located throughout the *C. glutamicum* ATCC 13032 genome. Several of these putative ARG box sites represent putative sites for ArgR-mediated regulation and may be of further interest for the characterisation of L-arginine biosynthesis in this bacterium. Finally, the effect of increased expression of the *lysE* gene (encoding an L-arginine exporter in *C. glutamicum* ATCC 13032) on L-arginine production was assessed and found to elevate the biosynthesis of this compound by about 2-fold relative to the wildtype.

The work described in this thesis thus contributes to our understanding of ArgR-mediated regulation in *C. glutamicum* ATCC 13032, where very little is known of the regulation of L-arginine biosynthesis. It is hoped that the information generated here may also be used for the creation of an industrially significant L-arginine overproducer strain of this bacterium.

Chapter One

L-Arginine Overproduction in *Corynebacterium glutamicum*

General Introduction

Contents

Chapter 1

1.1 A brief history and overview of the amino acid industry

In 1908, monosodium L-glutamate (MSG) was isolated from the Japanese *konbu* kelp by a researcher at the University of Tokyo. This compound quickly became known for its potent, taste-enhancing ability and the use of it as a daily food additive rapidly became commonplace throughout Northeast Asia. In order to extract MSG, plant protein sources, such as wheat gluten, were hydrolysed as part of a hazardous chemical extraction process using concentrated hydrochloric acid (Kinoshita, 2005). In addition to the associated health risks, the MSG yields from such a process are naturally limited due to the defined L-amino acid content of the raw starting material.

With a view to overcome these restrictions, the Kyowa Hakko Chemical Company of Tokyo, Japan, launched a research programme in 1956, aimed at obtaining a microorganism capable of accumulating L-glutamic acid extracellularly (Kinoshita *et al.*, 1957). An isolate from a faecal-contaminated soil sample, initially identified as *Micrococcus glutamicus* (later reclassified to *Corynebacterium glutamicum*), possessed an excellent ability to accumulate L-glutamic acid in significant quantities directly from cheap sources of ammonium and sugar (Kinoshita *et al.*, 1957; Kumagai, 2000). From mutational work subsequently carried out on this organism, it quickly became evident that many other L-amino acids, such as lysine, threonine and valine, could be accumulated in considerable quantities by *C. glutamicum* (Leuchtenberger, 1996). In addition to the exceptional amino acid production capacity of this bacterium, *C. glutamicum* also grows rapidly and is non-pathogenic, making it ideal for industrial food fermentations (Liebl, 2005).

The advent of the amino acid industry is inextricably linked with the discovery of *C. glutamicum* and, as of 2003, the global amino acid market was projected to expand at 10% per annum, with the total annual worldwide consumption of amino acids expected to exceed 2 million tons (Hermann, 2003). The majority of these amino acids are commercially produced via microbial fermentation processes, with only L-alanine, glycine and L-methionine continuing to be profitably produced by chemical synthesis (Ikeda, 2003). In 2003, for example, the coryneform bacteria alone were responsible for producing 1.5 million tons of L-glutamic acid, 600000 tons of L-lysine and over 1500 tons of L-arginine (Hermann, 2003; Ikeda, 2003). Whereas L-glutamic acid is chiefly used as a food flavourant, L-lysine,

along with L-arginine, L-threonine and L-tryptophan, are predominantly used as animal feed additives. The addition of 0.5% L-lysine, for example, increases feed quality by as much as adding approximately 20% more soybean meal (Hermann, 2003). L-Arginine also serves as an important pharmaceutical supplement and has been shown to possess immunostimulant properties (Morris *et al.*, 2004). Other L-amino acids, such as phenylalanine and valine, are produced by specialised *Escherichia coli* strains and are typically applied as cosmetic supplements and food sweeteners (Hermann, 2003).

1.2 Physiology and taxonomy of *C. glutamicum*

The corynebacteria are in possession of genomic DNA with a high molar G+C content of between 51% and 68% and represent an evolutionary line phylogenetically distinct from the low G+C Gram-positive bacteria (Goodfellow and Minnikin, 1981; Liebl, 2005). Grouped within the *Corynebacteriaceae* family and forming part of the *Actinomycetales* order of Gram-positive eubacteria, the *Corynebacterium* genus is composed of over 50 members and is phenotypically diverse (Liebl, 2005). Typical members of this genus are catalase positive, non-sporulating, non-motile and either facultatively anaerobic or aerobic (Liebl, 1991). Corynebacteria cells are frequently arranged in a V-shaped formation and are considered pleomorphic because, depending on their growth phase, they often appear to have either an irregular rod- or club-shaped morphology (Collins and Cummins, 1986). Over half of the known members of the *Corynebacterium* genus have been identified within the last decade from a diversity of environments, including dairy products and human skin samples (Liebl, 2005). This genus includes industrially and medically important microbes, such as *Corynebacterium ammoniagenes*, which is used commercially for nucleotide production, and *Corynebacterium diphtheria*, which is responsible for the respiratory tract disease diptheria (Liebl, 1991).

Early chemotaxonomic studies on the cell wall composition of the *Corynebacterianeae* suborder of the *Actinomycetales* order revealed the *Corynebacterium*, *Mycobacterium*, *Nocardia* and *Rhodococcus* genera to share a unique polysaccharide and lipid cell wall structure due to the inclusion of mycolic acids (Lechevalier and Lechevalier, 1970; Stackebrandt *et al.*, 1997). These bacteria have therefore been grouped into the robust, monophyletic *Corynebacterium-Mycobacterium-Nocardia* taxon (CMN), which has

subsequently been updated to include the *Gordoniaceae*, *Dietziaceae* and *Tsukamurellaceae* families (Barksdale, 1970; Goodfellow, 1992). In addition to serving as an important taxonomic marker, this distinctive cell wall organisation found in the CMN bacteria is significant when considering L-amino acid overproduction in *C. glutamicum* and is discussed later in the context of L-amino acid export (Section 1.6.3) (Eggeling and Sahm, 2003).

1.3 An overview of L-amino acid biosynthesis in *C. glutamicum*

The sequencing of the *C. glutamicum* genome has provided insight into the inherent, excellent L-amino acid producing ability of this organism (Kalinowski *et al.*, 2003; Nakagawa, 2002). Indeed, an *in silico* functional analysis of the predicted *C. glutamicum* proteome revealed it to putatively possess fewer enzymes responsible for L-amino acid catabolism relative to most other known microorganisms (Kalinowski *et al.*, 2003). For example, *C. glutamicum* is unable to degrade L-arginine and L-lysine and has no known homologues to either the tryptophanase gene, *tnaA*, or the glycine cleavage system, *gvc*, both of which are found in *E. coli* (Ikeda and Nakagawa, 2003). Furthermore, *C. glutamicum* is capable of many single-step enzyme reactions at key L-amino acid biosynthetic points that, in other organisms, typically require the action of several enzymes. For example, aspartokinase, which catalyses the first step of the biosynthesis of aspartate-derived amino acids, such as L-lysine and L-threonine, consists of a single isoform in *C. glutamicum*, whereas three isoforms are required in *E. coli* for the biosynthesis of these amino acids (Nakagawa, 2002).

As part of another *in silico* analysis of the *C. glutamicum* genome, Mitsuhashi *et al.* (2002) predicted this bacterium to possess carbonic anhydrase, an enzyme typically found in photosynthetic organisms, where it has been shown to be important for the efficient supply of bicarbonate and carbon dioxide for use in carboxylation reactions. This is of particular biotechnological significance because the efficiency of these reactions is tied directly to L-amino acid yields (Ikeda and Nakagawa, 2003). *C. glutamicum* is also in possession of several specific L-amino acid export systems, which are capable of rapidly depleting intracellular L-amino acid concentrations and can thereby aid the downstream recovery of L-amino acids from the supernatant (Krämer, 1994b). This organism is thus naturally geared

towards producing large amounts of L-amino acids and, indicative of its massive industrial importance, a review showed at least 1500 of the approximately 3000 known *C. glutamicum* genes to be covered by international patent applications (Ikeda and Nakagawa, 2003).

1.3.1 Central metabolism

The central metabolism of *C. glutamicum* is of key significance to L-amino acid biosynthesis as, during cell growth, metabolites are continuously withdrawn from shared metabolic pathways within the cell (Figure 1.1). As L-amino acid overproducing strains are usually mutated in such a way that their biosynthetic pathways are deregulated, a significant metabolic shift within the internal metabolism of the cell may occur. Due to this altered flux, certain metabolites may no longer be available for the biosynthesis of other molecules, which can result in a weakened or growth-impaired strain. By understanding how the central metabolism of *C. glutamicum* operates in relation to L-amino acid biosynthesis, the full effect of any mutation introduced into this microbe can be better understood.

C. glutamicum possesses at least three different phosphotransferase systems which enable the import of glucose, fructose and sucrose saccharides into the cell (Yokota and Lindley, 2005). After uptake, these sugar units are phosphorylated and further metabolised via the canonical glycolysis and the pentose phosphate pathways (PPP). The PPP, which is able to bypass the glycolysis cycle by forming a branch point at glucose-6-phosphate, generates reducing power in the form of NADPH and is also responsible for the non-oxidative interconversion of various sugars (Berg *et al.*, 2001). One such sugar, erythrose-4-phosphate, is required in large quantities by strains optimised for the production of aromatic amino acids, such as L-tryptophan and L-phenylalanine (Yokota and Lindley, 2005). A similar situation holds true for nucleotide- and nucleoside-producing *C. ammoniagenes* strains, which require constant replenishment of the ribose-5-phosphate sugar as a starting material for nucleotide biosynthesis (Yokota and Lindley, 2005). As the biosynthesis of L-amino acids requires a large amount of both pyruvate and NADPH, considerable attention has focused on controlling the flux between the glycolytic and PPP as part of strain improvement programmes.

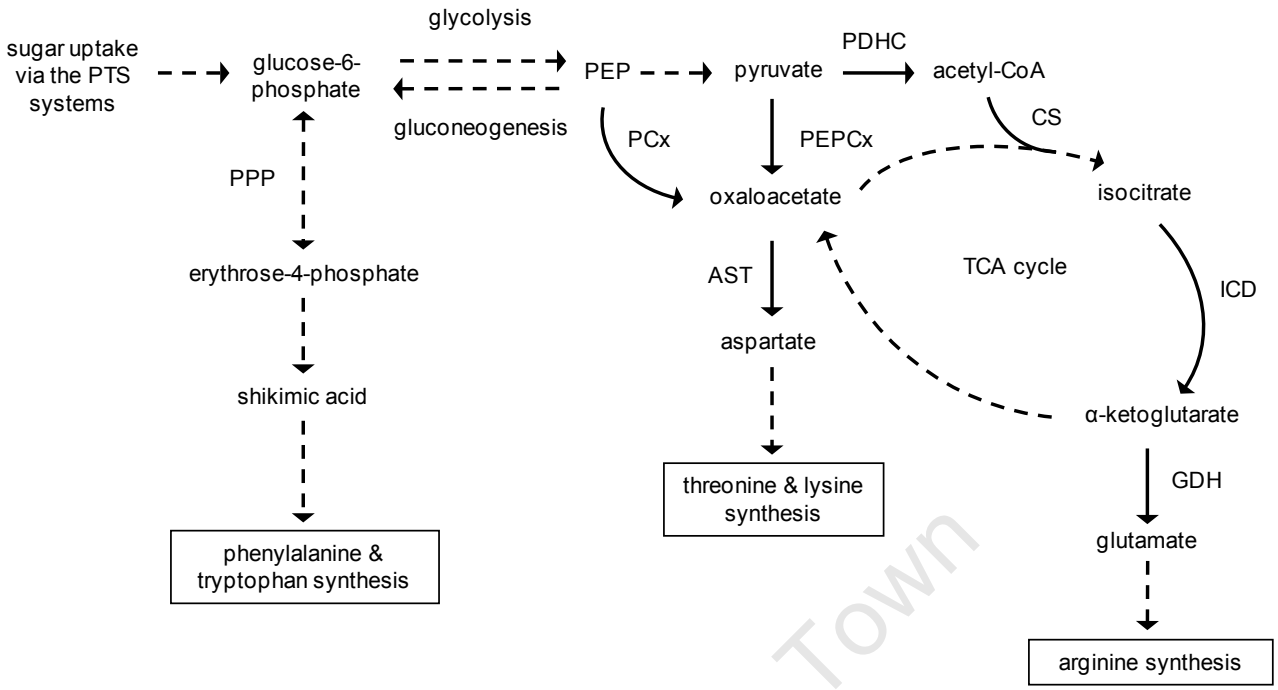


Figure 1.1 The central metabolism of *C. glutamicum* is shown in relation to L-amino acid biosynthesis. Glucose-6-phosphate is converted via the PPP to form erythrose-4-phosphate. This sugar subsequently condenses with PEP to form shikimic acid, after which the biosynthetic pathways for each of the aromatic amino acids (L-phenylalanine, L-tryptophan and L-tyrosine) branch off. L-Aspartate is derived from oxaloacetate and serves as an important branch point for the biosynthesis of the asparagine, isoleucine, lysine, threonine and methionine L-amino acids. The biosynthetic pathways of L-arginine, L-glutamine and L-proline all begin with L-glutamate, which is derived from α-ketoglutarate by GDH. The anaplerotic reactions catalysed by PCx and PEPCx are able to replenish oxaloacetate through the CO_2-mediated carboxylation of either pyruvate or PEP. Oxaloacetate is required later for the biosynthesis of the aspartate- and glutamate-derived L-amino acids. Not all of the *C. glutamicum* L-amino acid biosynthetic pathways are shown and the boxes represent L-amino acid biosynthetic pathways currently commercially exploited in *C. glutamicum*. Solid arrows represent single enzyme reactions and dashed arrows represent steps involving multiple enzymes. Abbreviations: aspartate transaminase (AST), citrate synthase (CS), glutamate dehydrogenase (GDH), isocitrate dehydrogenase (ICD), pyruvate dehydrogenase complex (PDHC), phosphoenol pyruvate (PEP), phosphoenol carboxylase (PEPCx), phosphotransferase system (PTS), pentose phosphate pathway (PPP), pyruvate carboxylase (PCx), tricarboxylic acid (TCA). Adapted from Eikmanns (2005) and Yokota and Lindley (2005).

The tricarboxylic acid (TCA) cycle, which is the final common pathway for the oxidation of fuel molecules and the generation of ATP, is an anabolic cycle which provides precursor molecules for a variety of processes, including L-amino acid biosynthesis and fatty acid metabolism. In relation to L-amino acid biosynthesis in *C. glutamicum*, the two most important intermediates of the TCA cycle are α-ketoglutarate, which is an L-glutamate precursor, and oxaloacetate, which is an aspartate precursor. In-depth reviews of the biosynthesis of the aromatic amino acids in *C. glutamicum*, in addition to the biosynthesis of the aspartate- and glutamate-derived amino acids, are provided respectively by Ikeda (2005), Willis *et al.* (2005) and Kimura (2005). The step-by-step biosynthesis of L- arginine is detailed later in Section 1.4.1.

In order to replenish depleted metabolites in the TCA cycle, *C. glutamicum* is in possession of two anaplerotic enzymes, pyruvate carboxylase (PCx) and PEP carboxylase (PEPCx), which are able to regenerate oxaloacetate from either pyruvate or phosphoenol pyruvate (PEP) with the absorption of bicarbonate (Peters-Wendisch *et al.*, 2001; Shiio *et al.*, 1960). PEPCx overexpression and inactivation has been shown to result in only marginal changes in L-glutamate and L-lysine levels. In contrast, however, elevated expression of PCx results in significantly increased L-glutamate and L-lysine accumulation and the activity of this enzyme is considered to be a major bottleneck for L-amino acid production (Okino *et al.*, 2008; Peters-Wendisch *et al.*, 1993). Furthermore, due to the dependence of this reaction on HCO_3, its activity (and hence the yield of amino acids) is obviously limited by the availability of inorganic carbon (supplied in the form of CO_2) and once L-amino acid production is engineered beyond a certain level, the supply of inorganic carbon can become a constraint for further improving L-amino acid production (Mitsuhashi *et al.*, 2002). The supply of inorganic carbon is also required for the action of carbamoylphosphate synthetase (CPS), which supplies carbamoylphosphate (CP) into the L-arginine biosynthetic pathway (this is detailed later in Section 1.4.1). Maintaining an efficient supply of CO_2 to the bioreactor is therefore an important part of the fermentation process for many industrial *C. glutamicum* amino acid production strains and other bacteria (Arndt and Eikmanns, 2008) For example, without an adequate supply of CO_2 or an active carbonic anhydrase enzyme, *Lactobacillus plantarum* has been shown to be auxotrophic for L-arginine biosynthesis (Nicoloff *et al.*, 2000). Additionally, the amount of dissolved CO_2 present in the readily utilisable bicarbonate form can be increased through elevated expression of the

carbonic anhydrase genes, which has been shown to have the effect of further improving the yield of amino acids such as L-lysine in industrial *C. glutamicum* strains already geared towards overproduction (Mitsuhashi *et al.*, 2003).

In addition to these two carboxylating enzymes, *C. glutamicum* has several decarboxylating anaplerotic enzymes that convert TCA cycle intermediates into either PEP or pyruvate, thereby reducing the net anaplerotic flux directed towards amino acid biosynthesis (Eikmanns, 2005). Previous metabolic flux analysis studies in *C. glutamicum* predict that the elimination of these reverse reactions would result in small increases in L-glutamate and L-lysine production (Riedl *et al.*, 2001).

1.3.2 Major control points in central metabolism and their influence on L-amino acid production

Pyruvate kinase, which catalyses the final irreversible step in glycolysis to yield ATP and pyruvate, is considered one of the most important control points of glycolysis in *C. glutamicum* (Yokota and Lindley, 2005). The activity of this enzyme is inhibited by ATP and positively regulated by AMP. In an attempt to enhance glycolysis in *C. glutamicum*, a gene encoding a defective H^+-ATPase was introduced into the chromosome in order to provoke an energy shortage within the cell (Sekine *et al.*, 2001). As expected, the rate of sugar uptake and cellular respiration increased in conjunction with the rate of glycolysis relative to the parent strain. Although work on improving L-amino acid yields through manipulation of the glycolytic cycle has been minimal, it has been noted that many industrial *C. glutamicum* strains created using traditional random mutagenesis techniques all display increased glycolytic flux and sugar uptake rates in addition to increased L-amino acid yields (Yokota and Lindley, 2005). In the case of the PPP, overexpression of the transketolase responsible for the formation of erythrose-4-phosphate resulted in a concomitant increase in the production of aromatic L-amino acids. Additionally, due to the consequent decrease in PEP availability for aspartate biosynthesis, L-lysine yields were observed to decrease significantly (Ikeda *et al.*, 1999).

Oxaloacetate and α-ketoglutarate are withdrawn from the TCA cycle for the biosynthesis of glutamate- and aspartate-derived L-amino acids and the bioavailability of these precursors therefore influences L-amino acid production. For example, citrate synthase (CS), which

catalyses the condensation of acetyl-CoA and oxaloacetate, is considered to be rate-limiting for the entry of substrates into the TCA cycle and the activity of this enzyme is directly correlated with the formation of the aspartate- and glutamate-derived L-amino acids (Eikmanns, 2005). α-Ketoglutarate also serves as another major flux point within this cycle, as both the α-ketoglutarate dehydrogenase complex (KGDH) and glutamate dehydrogenase (GDH), compete for this shared metabolite. KGDH activity is inversely correlated with L-glutamate formation and the attenuation of KGDH activity is thus an important factor for increasing metabolic flux through the α-ketoglutarate branch of the TCA cycle (Shimizu *et al.*, 2003). Additionally, the activity of the pyruvate dehydrogenase complex (PDHC), which converts pyruvate into acetyl-CoA, strongly influences the biosynthesis of the pyruvate-derived L-amino acids, which include alanine, leucine and valine. For example, a reduction in the intracellular levels of CoA in *C. glutamicum* leads to an overall reduction in carbon flux via the PDHC. This in turn has been shown to result in a dramatic increase in L-valine accumulation in *C. glutamicum* due to increased pyruvate availability (Radmacher *et al.*, 2002). Similarly, Blombach *et al.* (2008) recently noted the inactivation of the gene encoding the PDHC, in addition to the deletion of the genes encoding PCx as well as other enzymes involved in central metabolism, to greatly enhance the biosynthesis of L-valine.

1.4 L-Arginine biosynthesis

L-Arginine is a metabolically versatile, semi-essential basic amino acid that is widely distributed in foodstuffs such as wheat germ and chocolate. It is an essential building block of protein molecules and, in addition to serving as an important precursor for the biosynthesis of polyamines, such as putrescine and spermidine, it is also able to serve as a substrate for a variety of catabolic pathways, thereby serving as a source of carbon, nitrogen and energy (Cunin *et al.*, 1986). Infants are unable to effectively synthesise L-arginine, making it nutritionally essential, whereas adults are able to produce it as part of the urea cycle (Morris *et al.*, 2004). Aside from the aforementioned use of L-arginine as an animal feed additive, it is also commonly used as a pharmaceutical supplement for boosting the adult immune system, where it regulates the activity of the thymus gland and influences the production of white blood cells. L-Arginine is also used by the body to produce nitric oxide (NO), an important signalling molecule responsible for arterial dilation. Due to the NO-stimulating effects of L-arginine, it is commonly used in a variety of therapeutic

regimens to treat hypertension and coronary heart disease (Morris *et al.*, 2004). Since L-arginine is synthesised from L-glutamate through L-ornithine and L-citrulline, microorganisms such as *C. glutamicum*, which possess a strong ability to produce L-glutamate, represent ideal vehicles for L-arginine overproduction.

1.4.1 Enzymatic steps

There are three routes for the biosynthesis of L-arginine from L-glutamate in prokaryotes. The two most common, the so-called linear and recycling pathways, are shown in Figure 1.2. The linear pathway, which is common to the *Enterobacteriaceae* and the archaeobacterial genus *Sulfolobus*, initially involves the consumption of one molecule of acetyl-CoA by *N*-acetylglutamate synthase (NAGS) for the five step conversion of L-glutamate into L-ornithine via a series of acetylated intermediates (Cunin *et al.*, 1986; van de Casteele *et al.*, 1990). This pathway is characterised by the presence of acetylornithine deacetylase (AODA), which hydrolyses *N*-acetylornithine to acetate and L-ornithine. In contrast, most prokaryotes investigated thus far, including the *Corynebacterium* spp., use a more energetically economical process that utilises ornithine acetyltransferase (OAT) as an alternative to AODA (Xu *et al.*, 2007). OAT transfers the *N*-acetyl group from *N*-acetylornithine to L-glutamate in order to generate L-ornithine and *N*-acetylglutamate. *N*-acetylglutamate may then be cycled back into the L-arginine biosynthetic pathway, allowing for OAT to effectively bypass the first and fifth steps of the linear pathway (Bringel *et al.*, 1997). The NAGS enzyme is usually retained in these bacteria, however, as it fulfils an anaplerotic function through the initiation of acetyl cycling by the provision of an *N*-acetylglutamate molecule (Xu *et al.*, 2007). Certain bacteria, however, such as *Bacillus subtilis* and *Neisseria gonorrhoeae*, have a bifunctional OAT which possesses NAGS activity. This allows for OAT-mediated transacetylation of the acetyl-CoA acetyl group to glutamate, thereby effectively rendering NAGS functionally redundant. Consequently, many microorganisms with a bifunctional OAT lack a NAGS enzyme altogether (Xu *et al.*, 2007).

As part of the only published investigation into L-arginine biosynthesis in *C. glutamicum*, Sakanyan *et al.* (1996) detected NAGS activity in crude extracts of this bacterium grown in the absence of L-arginine. When the *argJ* gene of *C. glutamicum* was cloned into an *E. coli argA* mutant, however, no NAGS activity was detected, thus implying that *C. glutamicum*

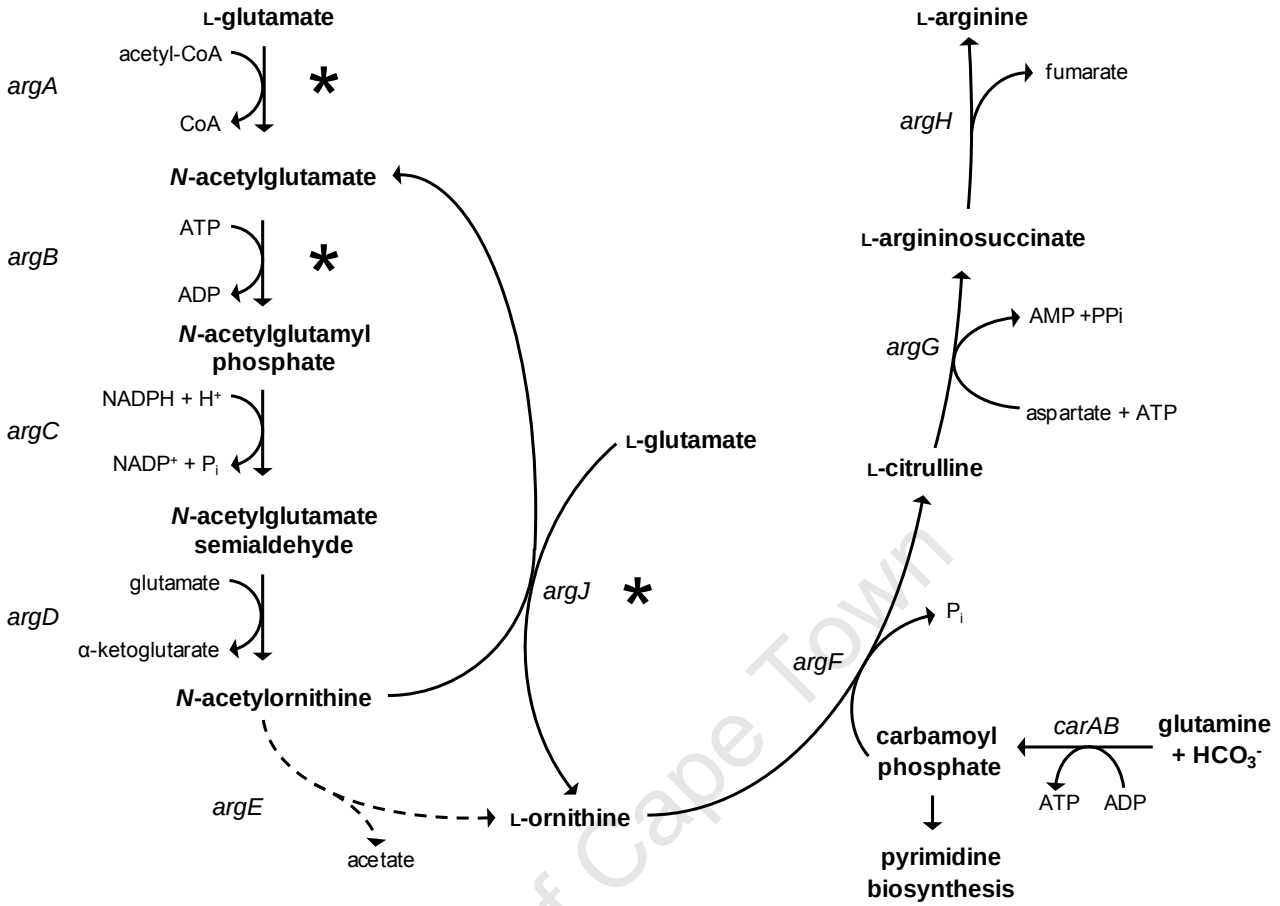


Figure 1.2 Schematic representation of L-arginine biosynthesis via the linear and recycling pathways. L-Arginine biosynthesis may begin with the acetylation of L-glutamate by either NAGS or OAT. *N*-acetylornithine is subsequently deacetylated to L-ornithine by either AODA or OAT, after which L-ornithine is converted into L-arginine as part of the final three biosynthetic steps common to both pathways. As part of the recycling pathway, OAT transfers the *N*-acetyl group of *N*-acetylornithine to L-glutamate, ensuring that NAGS and AODA are not required. Dashed arrows represent steps limited to the linear pathway. Asterisks represent known points of feedback inhibition in *C. glutamicum*. *N*-acetylglutamate synthase and *N*-acetylglutamate kinase are feedback-inhibited by L-arginine, whereas ornithine acetyltransferase is feedback-inhibited by L-ornithine. Each enzyme is encoded by the following genes: *N*-acetylglutamate synthase *(argA)*, *N*-acetylglutamate kinase *(argB)*, *N*-acetylglutamyl phosphate reductase *(argC)*, *N*-acetylornithine transaminase *(argD)*, acetylornithine deacetylase *(argE)*, ornithine transcarbamylase *(argF)*, argininosuccinate synthetase *(argG)*, argininosuccinate lyase *(argH)*, ornithine acetyltransferase *(argJ)*, carbamoylphosphate synthetase *(carAB)*. Adapted from Lu (2006) and Xu *et al.* (2007).

has a monofunctional OAT. Interestingly, in a recent review of prokaryotic L-arginine biosynthesis, Xu *et al.* (2007) found both *C. glutamicum* and *Methanococcus jannaschii* to lack any homologues to known *argA* genes, thereby suggesting that these bacteria have a yet undiscovered mechanism for L-glutamate acetylation and initiation of the acetyl cycle.

The remaining three steps of L-arginine biosynthesis are common to both the linear and recycling pathways (Figure 1.2). Ornithine transcarbamylase (OTC) catalyses the formation of L-citrulline by transferring the carbamoyl moiety of CP to L-ornithine. CP also serves as a precursor for pyrimidine biosynthesis and CPS formation is driven by the availability of bicarbonate, glutamine and ADP. Indeed, derepression of the L-arginine biosynthetic pathway has been shown to cause CP shortages and as result a shortage of pyrimidines and a decreased growth rate (Weerasinghe *et al.*, 2006). Argininosuccinate synthetase (ASS) fulfils the penultimate step of L-arginine biosynthesis, allowing for aspartate, citrulline and ATP to be converted into argininosuccinate. This compound is finally reversibly hydrolysed by argininosuccinate lyase (ASL) to form L-arginine and fumarate.

Lately, a third type of L-arginine biosynthetic pathway has been elucidated in *Xanthomonas compestris* (Shi *et al.*, 2005). This pathway employs the previously undiscovered enzyme acetylornithine transcarbamylase (AOTC) to catalyse the formation of *N*-acetylcitrulline from *N*-acetylornithine. Acetylcitrulline deacetylase (ACDA), which is also active against *N*-acetylornithine, is responsible for the formation of L-citrulline, which is ultimately converted into L-arginine by ASS and ASL (Shi *et al.*, 2005). Interestingly, AOTC and ACDA homologues have subsequently been found in a variety of eukaryotic pathogens, such as *Xylella* spp. and *Bacteroides* spp., where these enzymes are partly responsible for the generation of L-ornithine which, amongst many other uses, is incorporated into pathenogenicity-associated lipids in these bacteria (Xu *et al.*, 2007).

1.4.2 Feedback inhibition

The metabolic flow of acetylated precursors prior to L-ornithine biosynthesis is regulated by end-product feedback inhibition. In bacteria that possess an AODA, yet have no acetyl recycling enzyme, NAGS represents the first committed step of the linear pathway and is inhibited by L-arginine (Figure 1.2) (Xu *et al.*, 2007). In the recycling L-arginine biosynthetic pathway, *N*-acetylglutamate kinase (NAGK) activity can be flow controlling and is inhibited

by L-arginine in many organisms. For example in *C. glutamicum*, NAGK and NAGS activity experience 50% inhibition in the presence of 40 mM and 2 mM L-arginine respectively (Sakanyan *et al.*, 1996). Additionally, bacteria in possession of both a NAGS and an OAT, such as *Pseudomonas aeruginosa,* also experience L-arginine feedback inhibition of NAGK and NAGS (Haas *et al.,* 1972).

In contrast to NAGK and NAGS, OAT activity is not influenced by the concentration of L-arginine and has been shown to instead exhibit strong feedback inhibition in response to elevated L-ornithine concentrations in *C. glutamicum* and *Geobacillus stearothermophilus* (Sakanyan *et al.*, 1993; Udaka, 1966). Although no work to date has examined CPS regulation specifically in *C. glutamicum*, the formation of carbamoylphosphate as part of the L-arginine biosynthetic pathway has generally been found to be tightly controlled, as enzyme activity is activated by L-ornithine and is feedback sensitive to the pyrimidine uridine 5'-monophosphate (Lu, 2006). CPS is also activated by purine molecules, such as inosine monophosphate, thereby ensuring a balance is achieved between purine and pyrimidine biosynthesis (Cunin *et al.*, 1986). Interestingly, the regulation of the activity of the preceding enzyme in the L-arginine biosynthetic pathway, OTC, is considered to be of minor physiological importance due to the highly regulated nature of the subsequent CPS-mediated step. Consequently, there is minimal evidence of OTC feedback inhibition in bacteria (Cunin *et al.*, 1986). Control of NAGK, NAGS, OAT and CPS activity obviously heavily influences L-arginine production and, through the abolishment of their feedback sensitivity, strains capable of overproducing L-arginine strains can be engineered. This is discussed later in Section 1.7 in the context of strain improvement.

1.5 Organisation and regulation of the L-arginine biosynthetic genes

1.5.1 Genetic organisation

The genetic organisation of the prokaryotic L-arginine biosynthetic regulon is diverse and can take the form of multiple operons or individual genes either located in close proximity to each other or scattered throughout the chromosome (Figure 1.3). The *argCJBDFR* operon, for example, is responsible for the biosynthesis of the L-arginine precursor L-citrulline and is found in the majority of Gram-positive bacteria, including *Bacillus* spp. and *Streptomyces coelicolor* (Bringel *et al.*, 1997). Furthermore, in *Streptomyces*

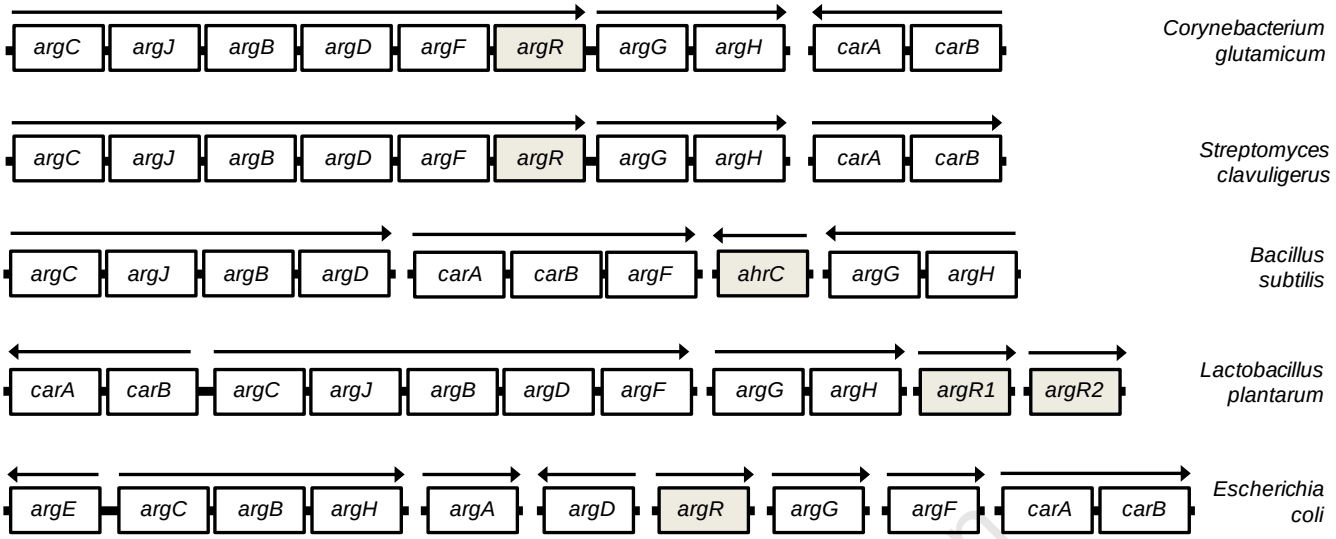


Figure 1.3 The genomic organisation of the L-arginine biosynthetic genes is diverse in prokaryotes. In most Gram-positive species, the L-citrulline biosynthetic genes are typically clustered together and form the *argCJBDF* operon. The remaining genes of the L-arginine biosynthetic pathway either form small operons or are individually scattered throughout the chromosome. With the exception of *C. glutamicum*, each transcriptional unit (indicated by an arrow showing the direction of transcription) of the L-arginine biosynthetic pathway in the above bacteria has been shown to be preceded by a putative ARG box pair (Section 1.5.2). Orthologous genes encoding an ArgR protein are shaded. Genes are represented by boxes. Genes adjacent to each other on the chromosome are joined by a solid black line. The *B. subtilis* arginine repressor is of unusually low homology (~27%) to the classical *E. coli* ArgR protein and is designated AhrC. *L. plantarum* possesses two non-complementary L-arginine repressors, encoded by *argR1* and *argR2*. Adapted from Hänßler *et al.* (2007), Maas (1994), Makarova *et al.* (2001), Nicoloff *et al.* (2004), Rodríguez-García *et al.* (2000) and Sakanyan *et al.* (1996).

clavuligerus, *argR* is transcribed together with the genes that comprise the L-citrulline biosynthetic operon (Rodríguez-García *et al.*, 2000). Most Gram-negative bacteria, however, such as *Pseudomonas* spp. and *Neisseria* spp., lack this operon and instead possess the L-citrulline biosynthetic genes individually distributed throughout the chromosome (Makarova *et al.*, 2001; Sakanyan *et al.*, 1996). In contrast to both, however, certain *Enterobacteriaceae*, such as *E. coli*, have an intermediate grouping of the L-citrulline

biosynthetic genes and form the divergently transcribed bipolar *argECBH* operon (Cunin *et al.*, 1986).

Further evidence of this diverse organisation can be found amongst the remaining genes of the L-arginine biosynthetic pathway. For example, in *B. subtilis, L. plantarum* and *S. clavuligerus, argG* and *argH* are transcribed together as an operon, whereas in most Gram-negative bacteria, these two genes appear to be individually transcribed and, together with the remaining genes that encode the L-arginine biosynthetic pathway, are often scattered throughout the genome (Makarova *et al.*, 2001; Rodríguez-García *et al.*, 2000). This organisation is well documented where, for example, in *E. coli* and *Haemophilus influenzae,* it has been shown to result in a higher degree of regulatory control by allowing for the delicate manipulation of the transcriptional levels of individual genes (Cunin *et al.*, 1986). This empirical form of control can also be seen in the organisation of the *carA* and *carB* genes, which typically form an operon separate from the other L-arginine biosynthetic genes (Makarova *et al.*, 2001). This allows for the fine adjustment of carbamoylphosphate levels that feed into the L-arginine biosynthetic pathway (Cunin *et al.*, 1986).

As part of the earliest study investigating L-arginine biosynthesis in *C. glutamicum*, a cluster of L-arginine biosynthetic genes, comprised of *argJ, argB, argD,* a 3'-portion of *argC* and a 5'-portion of *argF,* were cloned via heterologous complementation into an *E. coli argE* mutant (Sakanyan *et al.*, 1996). Transcription of this *C. glutamicum* DNA fragment in the auxotrophic *E. coli* mutant was shown to proceed from an internal promoter located within the preceding *argJ* gene. Furthermore, *argJ* transcription was shown to be vector dependent, suggesting that *argC* and *argJ* are transcribed together from the putative promoter region preceding *argC* (Figure 1.3). Interestingly, as part of a recent examination of the transcription of the *C. glutamicum* L-arginine biosynthetic genes by Hänßler *et al.* (2007), these genes were shown using RT-PCR to be transcribed as two polycistronic transcripts (an *argCJBDFR* transcript and an *argGH* transcript), thereby suggesting that the internal *argJ* promoter detected by Sakanyan *et al.* (1996) is secondary. Moreover, the genetic organisation of the *argC, argJ, argB, argD* and *argF* genes proposed by Sakanyan *et al.* (1996) was confirmed through the sequencing of the *C. glutamicum* genome by Kalinowski *et al.* (2003), who additionally predicted the *carA* and *carB* genes to be

transcribed as part of a single transcript, similar to what has been found in *E. coli* and the closely related *S. clavuligerus* (Rodríguez-García *et al.*, 2000).

1.5.2 The L-arginine repressor ArgR

Despite significant differences in genetic organisation, the regulation of the L-arginine biosynthetic genes is surprisingly similar across a broad range of archaea, Gram-positive and Gram-negative bacteria. In general, regulation of this pathway is exerted by the binding of the arginine repressor, ArgR, in conjunction with L-arginine, to specific operator sites preceding the target gene, known as ARG boxes (Figure 1.4) (Maas, 1994). Usually found in pairs, each ARG box typically consists of an imperfect 18 bp palindromic sequence overlapping the promoter region of the target gene (Makarova *et al.*, 2001). This pair of flanking ARG box sequences is usually separated by 2-3 bp and this spacing often means that both boxes are aligned on the same side of the DNA double helix (Maas, 1994). When ArgR binds to an ARG box, it is able to sterically hinder the initiation of transcription by RNA polymerase, resulting in repression of the L-arginine biosynthetic genes. The extent of this ArgR-mediated repression is often dependent on a multitude of factors, including the sequence variability of the ARG box pair, their spacing and their location relative to the transcriptional initiation start site (Makarova *et al.*, 2001). Although ArgR cannot serve as a transcriptional activator itself, it is also able to play an architectural role in the initiation of transcription through the promotion of distal contacts between other transcriptional activators, such as NtrC and cAMP-CRP, and RNA polymerase (Kiupakis and Reitzer, 2002). This mechanism is responsible for induction of the L-arginine catabolic operons in *B. subtilis* and *E. coli*, where ArgR binds to a single ARG box upstream of their relevant promoter regions (Lu and Abdelal, 1999).

Although only 27% amino acid sequence identity exists between the *E. coli* and *B. subtilis* L-arginine repressor molecules, recent crystallisation studies on this protein from *B. subtilis*, *E. coli*, *G. stearothermophilus* and *Mycobacterium tuberculosis* have revealed a highly conserved structural organisation and folding pattern, which has provided insight into a common mode of L-arginine repression across a broad range of microorganisms (Cherney *et al.*, 2008; Dennis *et al.*, 2002; Ni *et al.*, 1999; Sunnerhagen *et al.*, 1997). For example, the *B. subtilis* L-arginine repressor is able to functionally substitute for ArgR in *E. coli* (Smith *et al.*, 1986). Such a relationship is not found between any other repressor proteins of the L-

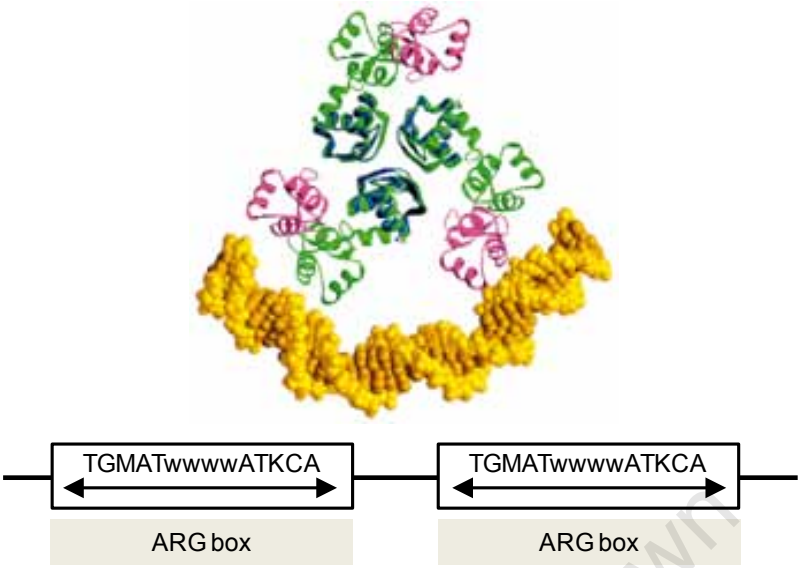


Figure 1.4. Model of the L-arginine-bound ArgR-ARG box interaction in *G. stearothermophilus*. The N-terminal DBD (pink) of each trimer establishes contact with a weakly palindromic ARG box sequence. Subunit oligomerisation occurs at the C-terminal ArgR trimer-trimer interface (blue) and is stabilised by L-arginine binding. ArgR-mediated DNA bending occurs at an angle of 70° and is centred between adjacent ARG boxes. The conserved core ARG box consensus sequence is depicted in each ARG box (where uppercase letters denote strongly conserved nucleotides and lowercase letters denote less conserved nucleotides). Arrows represent inverted repeats that make up ARG box half-sites. Only a single ArgR trimer is depicted for the sake of clarity. Adapted from Makarova *et al.* (2001) and Ni *et al.* (1999).

amino acid biosynthetic pathways common to these bacteria and is thought to be extremely unusual for organisms that are taxonomically distant (Maas, 1994).ArgR is a symmetrical, hexameric protein composed of two trimers stabilised by the fixation of six L-arginine molecules at the trimer-trimer interface (Ni *et al.*, 1999). L-Arginine-binding results in the allosteric activation of ArgR and enhanced ARG box affinity (Holtham *et al.*, 1999). This protein is member of the ArgR/AhrC transcriptional regulator family and each trimer consists of a characteristic winged helix-turn-helix (wHTH) N-terminal DNA-binding domain (DBD) connected through a central hinge region to an acidic C-terminal domain thought to be responsible for trimer oligomerisation and L-arginine binding (Lu, 2006). Following L-arginine binding, the DBD of each trimeric ArgR subunit establishes base-

specific contacts within the major groove of the DNA double helix. This portion of DNA usually consists of two consecutive ARG boxes separated by 2-3 bp and is bent by 70° upon ArgR binding (Figure 1.4) (Tian *et al.*, 1992).

Through the adoption of a comparative genomics approach, Makarova *et al.* (2001) proposed the existence of a putative universal ARG box consensus sequence, which is conserved across all bacterial lineages (TGMATwwwwATKCA; lowercase letters denote weakly conserved nucleotides) (Figure 1.4). ARG box sites were identified according to their similarity with the *E. coli* ARG box consensus sequence and, in agreement with previous experimental work, pairs of these ARG boxes were found to precede known L-arginine biosynthetic transcriptional units in a variety of bacteria (Makarova *et al.*, 2001). Interestingly, conservation of the ARG box sequence was shown to be relatively weak between individual sites within the same genome (Makarova *et al.*, 2001). As a consequence of this ARG box sequence variability, ArgR-binding to a single ARG box site, such as that preceding the autoregulated *E. coli argR* gene, occurs with a relatively low affinity, resulting in only mild repression (Maas, 1994). Due to the common pairwise grouping of ARG boxes, however, DNA-binding affinity of ArgR may be enhanced through the promotion of cooperative interactions between adjacent ArgR binding sites, thereby permitting more stringent repression of the L-arginine biosynthetic genes (Cunin *et al.*, 1986).

Although ArgR formation is typically autoregulated, repressor physiology studies in *E. coli* K12, have revealed ArgR to be maintained at unusually high concentrations (~330 molecules per cell) in the presence of L-arginine, due to the existence of a second, non-autoregulated *argR* promoter (Tian *et al.*, 1994). As ArgR has a relatively low binding affinity for L-arginine (K_d = 10^{-4} M), this mechanism of transcription ensures that a high concentration of ArgR is constantly available for L-arginine binding and regulation (Tian *et al.*, 1994). By maintaining ArgR at high concentrations, the cell is also able to overcome the low binding affinity of ArgR for single ARG boxes sites, whilst simultaneously ensuring that the degree of repression exerted on the L-arginine biosynthetic pathway in determined only by L-arginine availability (Maas, 1994).

From bacterial genome sequence analyses, several low G + C Gram-positive bacteria have been shown to possess multiple ArgR homologues, suggestive of an unusual mechanism of

L-arginine metabolism regulation. *Enterococcus faecalis*, for example, is in possession of two divergently transcribed *argR* homologues that are differentially expressed in response to increased levels of L-arginine and D-glucose (Barcelona-Andres *et al.*, 2002). *Lactococcus lactis* also has two ArgR homologues, encoded by *argR* and *ahrC*, that are interdependent and both essential for the L-arginine-dependent regulation of L-arginine biosynthesis. In this case, only AhrC is involved in the activation of L-arginine catabolism (Larsen *et al.*, 2004). *L. plantarum* has also been shown to possess two *argR* homologues (*argR1* and *argR2)*, which are both required for the repression of L-arginine biosynthesis. Interestingly, this bacterium possesses no L-arginine catabolic genes and the active repressor is proposed to form a heterohexamer, comprised of both ArgR1 and ArgR2, analogous to the classical *E. coli* ArgR (Nicoloff *et al.*, 2004).

In contrast to the canonical ArgR found in *E. coli* and *B. subtilis*, the ArgR of *P. aeruginosa* (ArgRp) is a dimeric protein in possession of a C-terminal helix-turn-helix (HTH) DBD belonging to the AraC/XylS family of transcriptional regulators (Martin and Rosner, 2001). ArgRp recognises specific tandem repeats similar in structure to ARG boxes and is able to serve as both a transcriptional repressor and activator for the L-arginine biosynthetic and catabolic genes respectively (Gallegos *et al.*, 1997). In *P. aeruginosa*, none of the genes preceding L-ornithine formation are subjected to ArgRp-mediated repression (Lu *et al.*, 2004). ArgRp is hence thought to represent a novel class of L-arginine repressor proteins and homologues have been discovered in both the *Burkholderia* and *Azotobacter* genera (Gallegos *et al.*, 1997).

An unexpected function of ArgR has emerged during studies on the replication of the *E. coli* ColE1 plasmid, where ArgR was found to promote the resolution of ColE1 multimers as part of a site-specific recombination system (Stirling *et al.*, 1988; Zakova and Szatmari, 1995). Due to the essential resolving action of ArgR, units of ColE1 are maintained as monomers and are thereby able to be stably inherited by daughter cells. The site of resolution in these plasmids, designated *cer*, has been shown to contain a single ARG box motif for which ArgR-binding was demonstrated *in vivo* (Stirling *et al.*, 1988). The existence of these two seemingly unrelated roles for ArgR, as both a transcriptional regulator and a mediator of homologous recombination (HR), is striking and is thought to be unprecedented amongst prokaryotic repressors (Maas, 1994). Interestingly, various histone-like proteins,

such as integration host factor, promote the HR-mediated integration of phage DNA into the host chromosome via DNA-bending (Robertson and Nash, 1998). As discussed earlier in this section, the ArgR-mediated bending of DNA in possession of ARG box sequence motifs has previously been demonstrated (Tian *et al.*, 1992).

1.5.3 FarR - a new regulator of L-amino acid metabolism

As part of a recent transcriptome-wide analysis of a *C. glutamicum* mutant, a new putative L-amino acid metabolic regulator, designated FarR, was identified (Hänßler *et al.*, 2007). Regulators of this type typically form dimers, each of which is in possession of a characteristic HTH domain (Rigali *et al.*, 2002). Similar to ARG box recognition by ArgR, FarR is thought to bind to adjacent symmetrical DNA sequences (Hänßler *et al.*, 2007; Rigali *et al.*, 2002). A *C. glutamicum farR* mutant strain, created recently by Hänßler *et al.* (2007), displayed a 2- to 3-fold decrease in transcriptional levels of a broad variety of genes involved in carbon and energy metabolism. Notably, this mutant strain also displayed moderately increased transcription levels of the L-arginine biosynthetic genes, suggesting that, like ArgR, FarR is a negative transcriptional regulator of L-arginine biosynthesis in *C. glutamicum*. However, although *in vitro* FarR binding to the intergenic regions preceding the *argC* and *argG* genes was subsequently demonstrated, FarR-deficient cells did not display significantly increased intracellular L-arginine production. The authors subsequently proposed this to be due to L-arginine-mediated feedback inhibition of NAGK (Hänßler *et al.*, 2007). Interestingly, regulators of this type typically required the binding of a specific ligand to stimulate the DNA-binding activity of the protein, however, as part of this study by Hänßler *et al.* (2007), a number of different putative effector molecules, including L-arginine, failed to affect the DNA-binding activity of FarR.

1.6 L-Amino acid export

During growth on complex media, peptide uptake results in increased intracellular L-amino acid concentrations. If these L-amino acids are not catabolised or incorporated into the cell for growth, they need to be removed from the cytosol in order to avoid bacteriostasis (Bellman *et al.*, 2001). Whereas various hydrophobic L-amino acids, such as L-isoleucine and L-phenylalanine, may be released speedily from *C. glutamicum* via passive diffusion (with first order diffusion rate constants of 0.08 min^{-1} and 0.45 min^{-1} respectively), the

cytoplasmic membrane of this bacterium represents an impermeable barrier against the efflux of polar L-amino acids, such as L-lysine (which has a first order diffusion rate constant of 0.002 min^{-1}) (Milner and Wood, 1987; Vrljić *et al.*, 1996; Zittrich and Krämer, 1994). Furthermore, during growth on complex media, translocation of these surplus L-amino acids needs to occur against a concentration gradient (Krämer, 1994a). In order to overcome this, *C. glutamicum* has several specific active carrier proteins within the cytoplasmic membrane that are responsible for the export of L-amino acids from the cytosol into the surrounding media. These carriers control intracellular L-amino acid concentrations and are thus able to influence the state of repression of the L-amino acid biosynthetic pathways (Eggeling, 2005). Increased L-amino acid availability in the supernatant also aids product recovery and can allow for a more cost efficient purification process (Kelle *et al.*, 2005). For these reasons, an understanding of L-amino acid excretion in *C. glutamicum* is important for the creation of strains with improved yields.

1.6.1 The L-lysine / L-arginine exporter LysE

LysE of *C. glutamicum* was the first bacterial L-amino acid exporter to be identified and, due to its unique structure and function, serves as a prototype for the novel LysE superfamily of translocator proteins (Vrljić *et al.*, 1996; Vrljić *et al.*, 1999). LysE homologues are classified into the LysE, CadD and RhtB families and have been identified in a diverse range of bacteria and archaea (Vrljić *et al.*, 1999). Typically consisting of six hydrophobic α-helical segments, each member of the LysE family is involved in the export of small positively charged molecules, including L-amino acids and other structurally related compounds. For example, the *lysE* homologue *attX*, found in *Rhodococcus fasciens*, is responsible for the export of a virulence- inducing L-arginine derivative (Maes *et al.*, 2001). In contrast to the LysE family, members of the CadD family are typically found in *Staphylococcus* spp., where they are responsible for conferring cadmium resistance (Chaouni *et al.*, 1996). The largest group in the LysE superfamily, the RhtB family, has numerous paralogues in a variety of bacteria, including *E. coli* and *B. subtilis*, which are responsible for L-homoserine lactone export and play a role in quorum sensing (Zakataeva *et al.*, 1999). Furthermore, overexpression of the RhtB genes in *E. coli*, resulted in an increased resistance to high intracellular levels of L-threonine, suggesting that these genes are partly responsible for efflux of this L-amino acid (Kruse *et al.*, 2002).

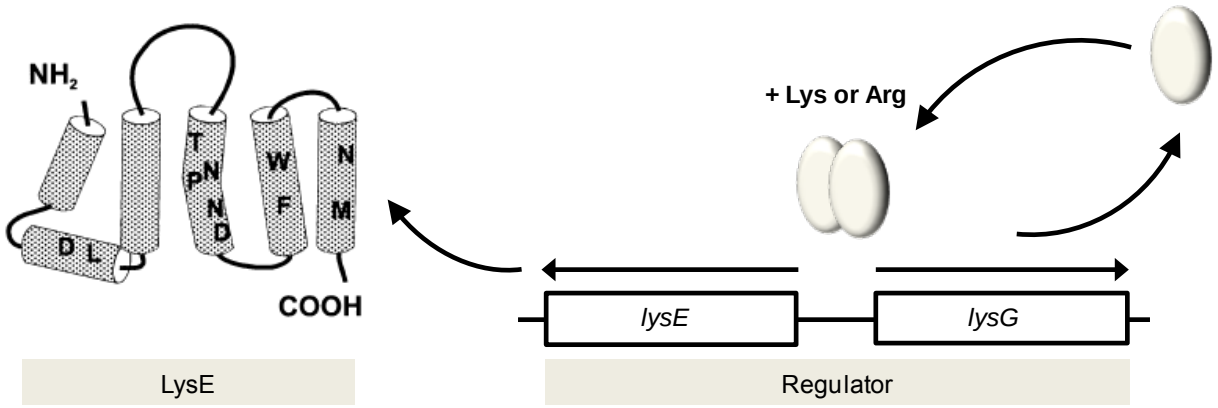


Figure 1.5 The LysEG system regulates L-lysine and L-arginine export. The *lysE* and *lysG* genes are divergently transcribed from overlapping promoters located within the *lysE-lysG* intergenic region. When bound to certain basic amino acids, such as L-lysine or L-arginine, the autogenously controlled transcriptional regulator LysG initiates *lysE* transcription through the recruitment of RNA polymerase. The LysE exporter is characterised by five hydrophobic segments that span the cytoplasmic membrane. Conserved amino acyl residues important to translocation are shown in single letter code. Adapted from Eggeling and Sahm (2001).

LysE is a relatively small membrane protein of 25 kDa consisting of five transmembrane α-helices and an additional hydrophobic segment thought to dip into the cytoplasmic membrane (Figure 1.5) (Vrljić *et al.*, 1996; Vrljić *et al.*, 1999). By analogy with other small transport molecules, such as Emr and AQP1, LysE is proposed to act as a homodimer (Eggeling, 2005; Murata *et al.*, 2000). The *C. glutamicum* LysE actively exports both L-arginine and L-lysine at a maximal rate of 0.75 nmol min^{-1} (mg dry cell mass)$^{-1}$ (Bellman *et al.*, 2001). A proton motor force (PMF) gradient drives translocation and the export of each L-lysine or L-arginine molecule is accompanied by either the symport of two OH^- ions or the antiport of two H^+ ions (Vrljić *et* 1999). As would be expected, the *C. glutamicum* L-lysine uptake system, LysI, exhibits strong substrate affinity for L-lysine (K_m = 10 µM), which allows for the bacterium to cope with low environmental substrate concentrations. In contrast, however, LysE exhibits a relatively weak affinity for L-lysine and L-arginine ($K_m \approx$ 20 mM for both) (Bröer and Krämer, 1991). This, together with the strong LysG-mediated regulation of *lysE* transcription, ensures that under normal physiological conditions, when the intracellular concentration of L-lysine is approximately 5 mM, no significant loss of this

valuable cellular building block occurs from *C. glutamicum* cells (Bellman *et al.*, 2001; Eggeling and Sahm, 2003).

Transcription of *lysE* is tightly controlled by the positive regulator LysG, which belongs to the large group of LysR-type transcriptional regulators (LTTRs) (Figure 1.5) (Schell, 1993). This autogenously controlled regulator enables the 20-fold induction of *lysE* in the presence of elevated intracellular L-lysine and L-arginine concentrations (≥ 35 mM). Additionally, although L-histidine and L-citrulline are not substrates of LysE, they have been shown to also induce *lysE* transcription at similar concentrations to L-arginine and L-lysine (Bellman *et al.*, 2001). Based on known characteristics of the LTTR family, the binding of either L-lysine, L-arginine, L-histidine or L-citrulline to LysG is proposed to result in cooperative intermolecular LysG interactions that give rise to a LysG multimer (Eggeling, 2005; Schell, 1993). This multimer has enhanced affinity for the promoter region between the divergently transcribed *lysE* and *lysG* genes and is able to recruit RNA polymerase, prompting the initiation of *lysE* transcription (Bellman *et al.*, 2001).

The significance of LysE in *C. glutamicum* is apparent upon deletion of *lysE*. In this scenario, growth on complex media is reduced due to elevated intracellular L-amino acid levels and, upon growth in media supplemented with L-lysine and L-arginine dipeptides, these two amino acids accumulate to concentrations of 1100 mM and 900 mM respectively prior to the occurrence of bacteriostasis (Bellman *et al.*, 2001; Vrljić *et al.*, 1996). Not only does this illustrate the impermeable nature of the *C. glutamicum* cytoplasmic membrane, but it also suggests that the rate of L-amino acid excretion may be rate-limiting for L-amino acid production in strains with deregulated biosynthetic pathways.

1.6.2 Other L-amino acid exporters in *C. glutamicum*

Similar to LysE, the L-threonine exporter ThrE of *C. glutamicum* is also representative of a new translocator family found in the bacterial, archaea and fungal kingdoms (Simic *et al.*, 2001). This 52 kDa protein is considerably larger than LysE and consists of ten transmembrane α-helical segments with an unusually long N-terminal domain. ThrE is capable of exporting both L-threonine and L-serine using PMF and, upon overexpression of the *thrE* gene, L-threonine accumulation increases by approximately 40% (Simic *et al.*, 2002). As L-threonine is nonpolar, efflux may, in part, also occur via diffusion. At an

intracellular L-threonine concentration of about 170 mM, 59% of L-threonine efflux is accounted for by ThrE-driven export, 22% is due to passive diffusion and the remaining 19% is due to at least one as yet unidentified L-threonine carrier (Simic *et al.*, 2001).

Although efflux of the branched L-amino acids isoleucine, leucine and valine is partly mediated by diffusion, export may also occur via the PMF-driven two-component BrnFE permease system (Kennerknecht *et al.*, 2002). The BrnF polypeptide is predicted to span the membrane seven times and the smaller BrnE protein four times. Together these two proteins comprise members of the novel LIV-E translocator. Interestingly, neither overexpression of the *brnFE* locus or the deletion of the L-isoleucine uptake system led to increased intracellular L-isoleucine accumulation (Eggeling, 2005; Tauch *et al.*, 1998). Thus, for non-polar L-amino acids, the rate of diffusion across the cytoplasmic membrane appears to be a decisive factor responsible for their intracellular accumulation (Eggeling, 2005).

1.6.3 The influence of the *C. glutamicum* cell wall on L-amino acid efflux

In addition to active L-amino acid carrier proteins localised within the cytoplasmic membrane, other components within the cell wall of the CMN bacteria influence L-amino acid efflux. Members of this taxon, which include *C. glutamicum* and *M. tuberculosis*, possess, in addition to the standard inner cytoplasmic membrane, an outer lipid bilayer through which L-amino acids are required to permeate in order to be successfully exported (Figure 1.6) (Daffé, 2005). This outer membrane is uniquely comprised of mycolic acids that are partially esterified to arabinogalactan polysaccharide units and other lipids within the cell wall, including lipomannan and certain trehalose derivatives. A cross-linked peptidoglycan layer, which contains significant amounts of mannose and glucose, is covalently bound to the arabinogalactan polysaccharide layer and is located above the cytoplasmic membrane (Puech *et al.*, 2000). This porous peptidoglycan, together with the mycolic acids, results in the formation of a structured lipid bilayer close to the cell surface that heavily influences cell wall permeability (Daffé, 2005; Schleifer and Kandler, 1972). Several pore-forming proteins and ion channels, similar to those found in Gram-negative bacteria, span this outer lipid bilayer and are thought to aid solute flux across the cell wall (Puech *et al.*, 2001).

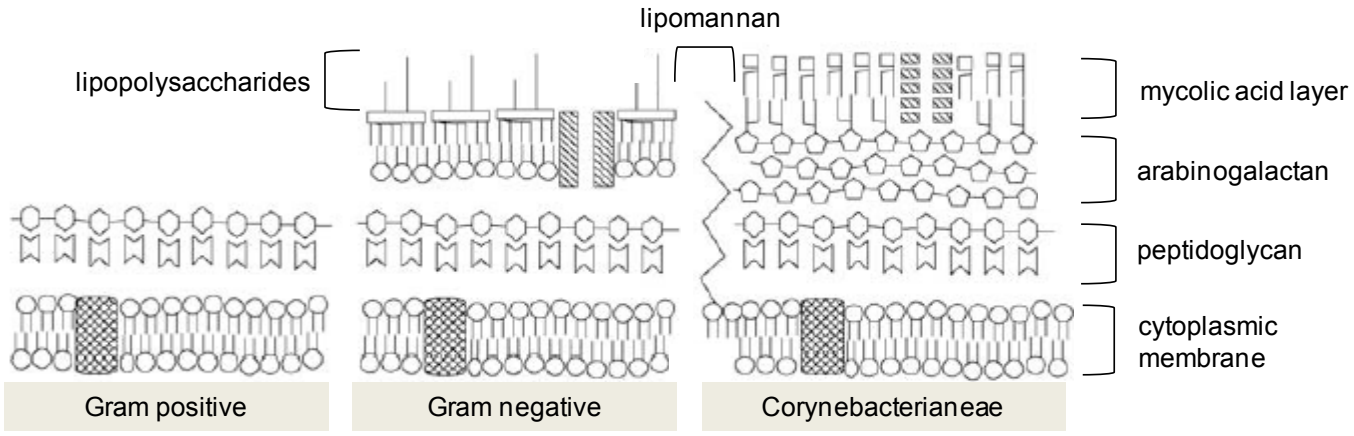


Figure 1.6 The corynebacterial cell envelope is different from the classical cell wall structures of Gram-positive and Gram-negative bacteria. In addition to the cytoplasmic membrane and peptidoglycan-containing cell wall of Gram-positive bacteria, the corynebacterial cell wall contains mycolic acids which, in conjunction with arabinogalactan, form the outer leaflet of an outer permeability barrier, similar in structure to the plasma membrane common to Gram-negative bacteria. Active carriers such as LysE and ThrE are located in the cytoplasmic membrane. Porins and channel proteins in the outer lipid bilayer permit the passage of solutes through the mycolic acid layer via diffusion. Adapted from Eggeling and Sahm (2003).

This atypical cell wall architecture is partly responsible for the highly drug-resistant nature of the human pathogen *M. tuberculosis*, where it presents a highly efficient barrier against antibiotic influx. When the mycoloyltransferase genes of this bacterium are disrupted, a decrease in the mycolic acid content of the cell wall occurs, which results in an improvement in the permeability of the cell wall (Jackson *et al.*, 1999). A similar effect is observed in *C. glutamicum* cells upon disruption of the *cps1* gene, encoding the PS1 mycoloyltransferase, which resulted in elevated L-lysine and L-glutamate excretion (Gebhardt *et al.*, 2007; Puech *et al.*, 2000). Interestingly, the induction of L-glutamate excretion by *C. glutamicum* is dependent on a diverse range of factors, including biotin limitation and treatment with penicillin and certain detergents (Kimura, 2005). Furthermore, strains that are auxotrophic for glycerol or fatty acids, tend to exogenously accumulate L-glutamate in large quantities (Okazaki *et al.*, 1967). Although the exact model of L-glutamate excretion is poorly understood, each of these factors have been shown to influence the composition of the outer lipid bilayer, where they are thought to modulate the permeability of the mycolic acid layer and thereby permit increased diffusion across the

cell wall (Eggeling and Sahm, 2001). The importance of the outer lipid layer in controlling L-amino acid efflux via diffusion is becoming increasingly apparent and, in contrast to *M. tuberculosis*, the absence of certain enzymes involved in the biosynthesis of cell wall synthesis appears to be non-lethal to *C. glutamicum* cells (Kimura, 2005). This raises the possibility of engineering strains impaired in the production of specific cell wall constituents that display enhanced L-amino acid efflux capabilities.

1.7 Strain improvement technology

Due to the exceptional economic value of *C. glutamicum*, this organism is the subject of strain improvement programmes that aim to enhance its industrial productivity. A brief overview of the various technical approaches adopted within the last 60 years to improve metabolite production in this organism is shown in Figure 1.7. During the early stages of *C. glutamicum* strain mutation and development, the breeding of production strains was largely dependent on the iterative random and selection of overproducing strains. This approach allowed for the stepwise assembly of beneficial mutations and many potent L-amino acid producers have since been created using this methodology (Ikeda and Nakagawa, 2003). For example, the most prominent *C. glutamicum* L-arginine producer reported to date was derived from several rounds of alternating X-ray and diethyl sulfate treatment and is capable of accumulating L-arginine up to a concentration of 25.3 g/l (Utagawa, 2004).

The advent of recombinant DNA technology, together with increased knowledge regarding L-arginine metabolism derived from molecular studies, have permitted for the directed, rational development of L-arginine producing strains. In principle, L-arginine yields in *C. glutamicum* may be improved by increased biosynthesis and excretion. Interestingly, although published research on the biosynthesis and overproduction of other L-amino acids in *C. glutamicum* is extensive, only a single study concerned with characterising L-arginine biosynthesis in this bacterium has been published to date (Sakanyan *et al.*; 1996). Furthermore, barring a very recent study by Ikeda *et al.* (2009), there have been no reported attempts to enhance L-arginine biosynthesis in *C. glutamicum*. As part of the study by Ikeda *et al.* (2009), sequencing information of the L-arginine biosynthetic genes in classically mutated *C. glutamicum* L-arginine and L-citrulline producer strain was used to introduce a single point mutation into the *argB* gene of the WT strain, thereby causing it to

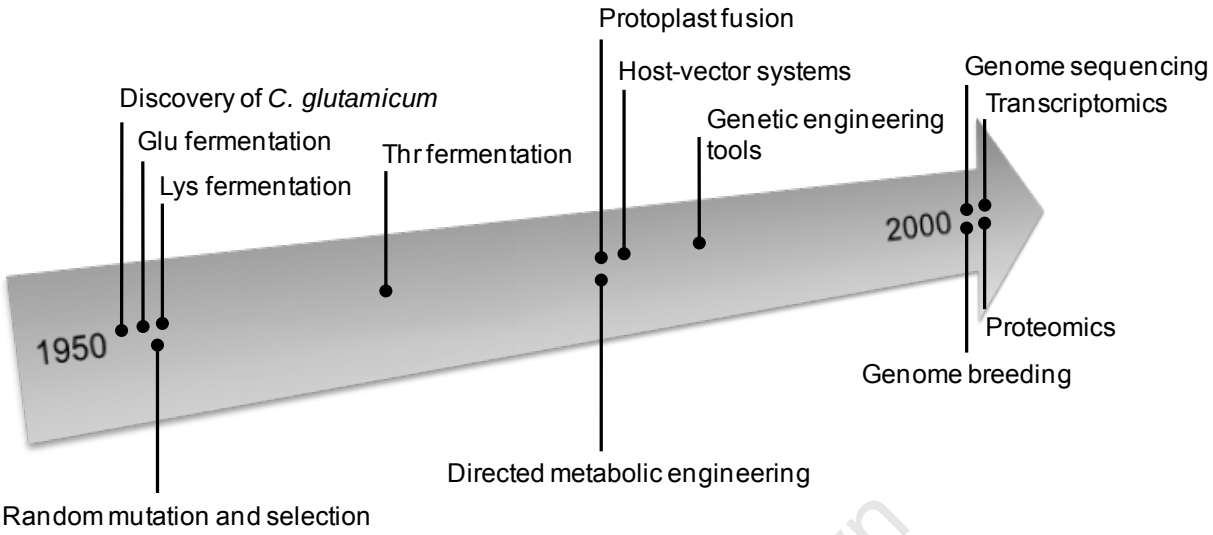


Figure 1.7 History of L-amino acid fermentation and strain development technology in *C. glutamicum*. Strain improvement strategies and their associated techniques are shown along an approximate timeline. Adapted from Ikeda and Nakagawa (2003).

encode a feedback resistant NAGK enzyme (Ikeda *et al.*, 2009). This alone did not significantly increase L-arginine production, however, and, as discussed extensively in Chapter 3, only upon the deletion of the a*rgR* gene in the strain containing the single *argB* point mutation did L-arginine production improve dramatically (where it reached levels of 81 mM). Although the lack of literature regarding the biosynthesis of this amino acid in *C. glutamicum* is in no doubt partly due to the commercially sensitive nature of this research, surprisingly little has also been published regarding the enhancement of L-arginine production yields in model organisms such as *E. coli* (Rajagopal *et al.*, 1998; Tuchman *et al.*, 1997).

As part of one such study, a recombinant *E. coli* strain with a 'derepressed' L-arginine biosynthetic regulon was created (Tuchman *et al.*, 1997). ARG boxes were introduced into a plasmid encoding the *argF* and *carAB* genes, which resulted in ArgR titration away from the chromosomal biosynthetic regulon and, therefore, very high intracellular quantities of the CPS and OTC enzymes. The biosynthetic capacity of this recombinant pathway, however, was only significantly improved upon supplementation with exogenous L-ornithine, suggesting that biosynthesis of this metabolite is rate-limiting for L-arginine

production. To improve the supply of L-ornithine, site-specific mutations were introduced into the *argA* gene (which represents the first committed step of L-arginine biosynthesis), so as to create an L-arginine feedback resistant NAGS enzyme (Rajagopal *et al.*, 1998). Plasmid-based overexpression of this modified *argA* gene as part of the aforementioned system allowed for endogenously formed L-ornithine to completely saturate the L-arginine biosynthetic pathway and increased L-arginine yields approximately 35-fold (to 107 nmol h^{-1} [mg dry cell mass]$^{-1}$).

The sequencing of the *C. glutamicum* genome has given rise to many interesting post-genomic technologies, including DNA arrays and transcriptome analysis, which allow for a comprehensive overview of the total metabolic activity of the organism (for in-depth reviews see Ikeda and Nakagawa [2003] and Vertès *et al.* [2005]). Of particular interest is the use of traditional recombinant methods in conjunction with next-generation 'genome-breeding' technology to further improve highly developed industrial strains (Ikeda and Nakagawa, 2003). As the majority of these producer strains are descendants of strains subjected to random mutagenesis, they often contain many uncharacterised, detrimental mutations. Through the comparative analysis of genome sequences, potentially useful individual mutations may be selected based upon known biochemical information and, analogous to the approach used by Ikeda *et al.* (2009), introduced on a one-on-one basis into the wild-type genome via allelic replacement. If the introduced mutation is shown to exert a positive effect on production of the desired metabolite, it is maintained in the genome and the resulting mutant subsequently used as a parent for the introduction of the next mutation. This iterative approach can allow for the creation of a minimally mutated strain in possession of no deleterious mutations.

1.8 Project aims

An improved understanding of the regulation of L-arginine biosynthesis in *C. glutamicum* may be achieved through the characterisation of *C. glutamicum* strains deficient in specific genes of the L-arginine biosynthetic pathway. This could allow for the creation of L-arginine overproducing *C. glutamicum* mutants that have industrial significance.

The aims of this study were thus to characterise ArgR-mediated regulation of the L-arginine biosynthetic operon in *C. glutamicum* ATCC 13032 with a view to overproduce of L-arginine:

- To employ a novel, recombinant vector system, previously used only in *M. tuberculosis*, to generate double unmarked *C. glutamicum* ATCC 13032 *argH* and *argR* mutants.

- To test the industrial suitability of these strains, together with a *C. glutamicum argR* double cross-over mutant and a previously created *C. glutamicum argR* integration mutant strain, through stability, growth and metabolite production studies.

- To characterise the L-arginine biosynthetic genes with respect to their arrangement and transcriptional regulation using the semi-quantitative reverse transcriptase polymerase chain reaction.

- To characterise the role of the ArgR protein in regulation by DNA-binding studies. This could reveal potential new targets for the manipulation of L-arginine production in this bacterium.

- Finally, the impact of increased *lysE* gene dosage on L-arginine biosynthesis in *C. glutamicum* ATCC 13032 was investigated.

Chapter Two

Construction of *Corynebacterium glutamicum* *argH* and *argR* mutant strains

Contents

Chapter 2

2.1 Summary

A better understanding of L-arginine biosynthesis in the industrially important microbe *Corynebacterium glutamicum* can allow for the creation of overproducing strains. Here, by adopting the p2NIL/pGOAL19 gene replacement system previously used in *Mycobacterium tuberculosis*, we aimed to construct a *C. glutamicum* unmarked *argH* deletion mutant, which is deficient in argininosuccinate lyase activity and dependent on exogenous L-arginine for growth. Additionally, we aimed to create a *C. glutamicum* unmarked *argR* deletion mutant, in possession of a non-functional L-arginine repressor gene, for use in future studies on the regulation of L-arginine biosynthesis. Two suicide delivery vectors, pARGHKO and pARGRKO, each harbouring in-frame deletions of either the *C. glutamicum* ATCC 13032 *argH* or *argR* genes, were constructed and used for the generation of *C. glutamicum argH* and *argR* integration mutant strains (designated SCOH and SCOR respectively). Analysis using the polymerase chain reaction, in conjunction with Southern blotting, revealed that these strains to possess at least a single copy of the appropriate mutant allele, in addition to an intact copy of the wildtype gene. Despite the prolonged incubation of these strains in sucrose-containing media in the absence of antibiotic selection pressure, we were unable to force a second cross-over event that would result in the exclusion of the relevant wildtype allele. The pK19*mobsacB* allelic exchange system, which has previously been used successfully in *C. glutamicum* ATCC 13032, was, after verification of the functionality of the plasmid-encoded *sacB* gene, successfully used for the generation of a new *argR* integration mutant strain (designated SCOR-D3). An unmarked *C. glutamicum argR* deletion strain (designated DCOR-D3) was generated from this strain, and was confirmed to contain a single *argR* mutant allele and no wildtype *argR* allele by both the polymerase chain reaction and Southern blotting.

2.2 Introduction

C. glutamicum is an industrially important microbe responsible for the worldwide production of over 2 million tons of L-amino acids per annum (Hermann, 2003). The majority of this production is constituted by L-lysine, which is commonly used as an animal feed additive (Ikeda, 2003). Due to the massive market potential of this L-amino acid, its biosynthesis in *C. glutamicum* is exceptionally well-understood and is the subject of several advanced strain improvement projects (Section 1.7) (Ikeda and Nakagawa, 2003; Kelle *et al.*, 2005; Kjeldsen and Nielsen, 2008). In contrast, however, the genetic improvement of production in *C. glutamicum* is relatively unexplored and only a single study of the biosynthesis of this L-amino acid in this bacterium has been published to date (Sakanyan *et al.*, 1996). Furthermore, with the exception of a very recent study by Ikeda *et al.* (2009), there appears to have been no other published attempts to create a rationally mutated *C. glutamicum* L-arginine overproducer strain. Consequently, the approximately 1500 tonnes of L-arginine produced annually by *C. glutamicum* are from relatively uncharacterised overproducer strains created through random mutagenesis programmes (Lu, 2006; Utagawa, 2004). Globally, the demand for L-arginine has been found to be increasing by roughly 10% per annum and it is becoming increasingly used as a human immunostimulant (Hermann, 2003; Morris *et al.*, 2004). The aim of this work was to create *C. glutamicum* mutants with altered L-arginine biosynthetic pathways. As part of later studies, these mutant strains could subsequently be characterised in terms of their L-arginine production ability, thereby completing an important step in the construction of a commercially significant L-arginine overproducing strain.

C. glutamicum employs the so-called recycling biosynthetic pathway for L-arginine production (Figure 1.2) (Xu *et al.*, 2007). In contrast to the *Enterobacteriaceae*, which utilise a linear L-arginine biosynthetic pathway, this recycling pathway is energetically more efficient, as it involves the cycling of an acetyl group from *N*-acetylornithine to L-glutamate for the generation of both *N*-acetylglutamate, which is now able to re-enter this pathway, and L-ornithine (Section 1.4.1) (Cunin *et al.*, 1986). Argininosuccinate lyase (ASL), encoded by *argH*, catalyses the final step of the L-arginine biosynthetic recycling pathway, which involves the reversible breakdown of argininosuccinate into L-arginine and fumarate. The L-arginine repressor protein, ArgR, is, by analogy with other better studied organisms,

thought to be a negative transcriptional regulator of the L-arginine biosynthetic genes in *C. glutamicum* (Section 1.5.2) (Xu *et al.*, 2007). Through the introduction of non-functional *argH* and *argR* genes into *C. glutamicum*, we would be able to construct strains that are either auxotrophic for L-arginine biosynthesis, or which display derepressed transcription of the L-arginine biosynthetic regulon.

Previously, the p2NIL/pGOAL19 vector system has been used with great success in *Mycobacterium tuberculosis* for the generation of unmarked double cross-over (DCO) deletion mutants (Parish *et al.*, 1999; Gordhan and Parish, 2001). To our knowledge, this system has not been used in *C. glutamicum* before and, as it possesses a relatively high number of selectable marker genes, it may also prove more reliable than conventional mutagenesis vectors, such as pK19*mobsacB*. The p2NIL/pGOAL19 vector system allows for the rapid assembly of a suicide delivery vector containing a mutant allele of the target gene. Following the integration of this suicide vector into the chromosome, the counter-selectable *sacB* marker gene, which is lethal in the presence of sucrose, permits cells which have lost the integration vector to be screened for sucrose resistance. Ideally, some of these cells should retain the mutant allele and thus possess an unmarked DCO mutation. Importantly, as the internal deletion can be designed to occur in-frame, it is ensured that, in the case of the target gene being located within an operon, any downstream genes are not shifted out of frame. The creation of an auxotrophic *C. glutamicum* mutant is desirable from an industrial point of view, as the plasmid-based expression of the corresponding WT allele can serve as a plasmid selection marker, thereby potentially reducing antibiotic usage (Vertès *et al.*, 2005). Lastly, the establishment of an improved allelic exchange system in this bacterium would allow for the easier introduction of subtle, point mutations into the relevant target gene. This would be extremely useful for the creation of alleles encoding feedback resistant ArgA, ArgB and ArgJ enzymes (Section 1.4.1).

In this study, we initially aimed to use the p2NIL/pGOAL19 vector system to generate double unmarked *argH* and *argR* mutants in the *C. glutamicum* ATCC 13032 type strain. This strain is the parent strain for most of the *C. glutamicum* strains used industrially today (Liebl, 2005). Alleles of these genes containing internal, in-frame deletions were successfully integrated into the *C. glutamicum* ATCC 13032 chromosome by homologous recombination (HR), however, intact copies of each target gene could ultimately not be

excluded. As a DCO *argR* mutant was required for future work, the ability of the pK19*mobsacB* plasmid to generate an *argR* deletion mutant was subsequently assessed, where it led to the successful creation of a *C. glutamicum* strain in possession of an unmarked *argR* deletion.

2.3 Materials and Methods

2.3.1 Bioinformatics

The *C. glutamicum* ATCC 13032 ArgH and ArgR amino acid sequences (Genbank accession numbers YP_225688 and YP_225686) were retrieved from the National Centre for Bioinformatics (NCBI) web server (http://www.ncbi.nlm.nih.gov). Similar sequences were retrieved from the NCBI Entrez protein database (http://www.ncbi.nlm.nih.gov/sites /entrez?db=Protein) using the BLASTP 2.2.17 algorithm (http://www.ncbi.nlm.nih.gov /BLAST) (Altschul *et al.*, 1997; Schäffer *et al.*, 2001). Multiple amino acid sequence alignments were constructed using the AlignX module of the Vector NTI Advance software suite (version 10.1.1; Invitrogen) (Lu and Moriyama, 2004). Orthologous ArgH or ArgR amino acid sequences from each of the following organisms were used for the construction of multiple sequence alignments (ArgH and ArgR Genbank accession numbers are indicated respectively in parentheses for each organism): *M. tuberculosis* CDC1551 (NP_336152; NP_336150), *Streptomyces clavuligerus* (CAA88927; CAB82483), *Escherichia coli* CFT073 (NP_756773; NP_755858), *Bacillus subtilis subsp. subtilis str.* 168 (NP_390822; NP_390305) and *Lactobacillus plantarum* WCFS1 (NP_784524; NP_785197).

2.3.2 Bacterial strains, plasmids and media

The bacterial strains and plasmids used in this study are listed in Appendix A. *E. coli* DH5α cells were grown in Luria-Bertani broth at 37 °C with lateral shaking (Sambrook *et al.*, 2001). *C. glutamicum* cells were routinely grown in 2 × Yeast-Tryptone (2 × YT) broth in the presence of Nx (30 µg ml^{-1}) in baffled flasks (Schott) at 30 °C on an orbital shaker at 120 rpm (Sambrook *et al.*, 2001). *C. glutamicum* 9A was a gift from N. Smazanowiz. *C. glutamicum* CGXII minimal medium (MM) was produced according to the protocol of Eggeling and Reyes (2005), which is described in Appendix B, and was supplemented with 10 mM L-arginine (Sigma-Aldrich) if required. Agar (Merck) was incorporated into solid media at a concentration of 1.5% (w/v) for plating. When necessary, Amp was used at 100 µg ml^{-1}, Hyg at 100 µg ml^{-1}, Kan at 30 µg ml^{-1}, IPTG at 1 mM, X-gal at 50 µg ml^{-1} and sucrose at 10% (w/v). Hyg was obtained from Roche and all other antibiotics were purchased from Sigma-Aldrich. Sucrose was obtained from Merck, whereas IPTG and X-gal were obtained from Bioline.

2.3.3 Preparation and manipulation of DNA

Plasmid DNA was routinely isolated from *E. coli* DH5α using the alkaline lysis method of Ish-Horowicz and Burke (1981). Plasmid DNA for use in cloning or sequencing was isolated from *E. coli* DH5α cells using either a Plasmid Midi Kit (Qiagen) or an E.Z.N.A. Plasmid Miniprep Kit (Peqlab Biotechologie GmbH) as per the manufacturers' protocols. *E. coli* DH5α competent cells were prepared by $CaCl_2$ treatment and transformed using the heat shock protocol of Dagert and Ehrlich (1979). Chromosomal DNA was extracted from cells of each *C. glutamicum* strain using either the DNeasy Blood and Tissue Kit (Qiagen) or the procedure described by Eikmanns *et al.* (1994). DNA fragments were routinely analysed by electrophoresis through either 0.8% (w/v) or 2.0% (w/v) agarose gels in TAE buffer (40 mM Tris-acetate, 0.5 mM EDTA), visualised using ethidium bromide staining and a GelDoc-XR UV System (Bio-Rad) and, when required for cloning, purified using the QIAquick Gel Extraction Kit (Qiagen) (Sambrook *et al.*, 2001). Lambda bacteriophage DNA (Promega) digested with *Pst*I was routinely used as a DNA molecular size marker. DNA quality and concentration were measured with a *ND*-1000 spectrophotometer (NanoDrop Technologies). Restriction digests and ligations were performed according to standard methods (Sambrook *et al.*, 2001). The *Pac*I restriction enzyme was obtained from New England Bioproducts and, unless otherwise stated, all other enzymes from Fermentas. DNA samples were stored at -20 °C in sterile dH_2O.

2.3.4 DNA amplification using the polymerase chain reaction

Polymerase chain reaction (PCR) mixtures (50 µl) typically consisted of (final concentrations indicated in parentheses): 1 µl genomic or plasmid template DNA (2 ng μl^{-1}), 2.5 µl of each primer (0.5 µM), 25 µl of 2 × KAPATaq ReadyMix (10 mM Tris-HCl [pH 8.6], 50 mM KCl, 0.05% [v/v] Tween 20, 0.5 mM DTT, 5% [v/v] glycerol, 1.5 mM $MgCl_2$, 1.25 U KAPATaq and 0.2 mM of each dNTP) (Kapa Biosystems) and 19 µl of sterile dH_2O. The thermal cycling for each PCR was carried out using either GeneAmp PCR System 9700 (Applied Biosystems) or an Advanced Primus 25 Thermal Cycler (Peqlab Biotechologie GmbH). The different primer combinations and conditions used for each PCR in this chapter are described in Appendix C. DNA fragments used for the construction of pARGHKO and pARGRKO were amplified using the High Fidelity PCR System available from Roche. PCR product fragments were purified using either the QIAquick PCR

Purification Kit (Qiagen) or the Biospin PCR Purification Kit (Bioflux). Primer oligonucleotides were synthesised by Pei-Yin Liebrich (Department of Molecular and Cell Biology, University of Cape Town, South Africa) using an Oligo 1000M DNA Synthesiser (Beckman).

2.3.5 DNA sequencing

For the sequencing of most PCR products, each DNA fragment was initially ligated into the TA-tailed vector pTZ57R/T using the InsTAclone PCR Cloning Kit (Fermentas). Each labelling reaction was subsequently performed using either the appropriate supplied primers or the M13 oligonucleotides by Diane James (Department of Molecular and Cell Biology, University of Cape Town, South Africa). The reaction products were thereafter sequenced by Macrogen (South Korea).

2.3.6 Generation of *C. glutamicum* mutants created using the p2NIL/pGOAL19 system

2.3.6.1 Strategy for suicide delivery vector construction

The cloning strategy used to generate suicide delivery vectors and plasmid maps of p2NIL and pGOAL19 is shown in Figure 2.1. Both vectors were gifts from Tanya Parish (Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom). p2NIL was used for manipulation of the gene of interest and contains an *E. coli* origin of replication (*oriE*), a gene conferring Kan^R and a multiple cloning site (MCS) with a single *Pac*I site (Parish and Stoker, 2000). pGOAL19 harbours an *oriE*, a gene conferring Amp^R and a marker gene cassette flanked by two *Pac*I sites (Parish and Stoker, 2000). This marker cassette contains a gene conferring Hyg^R, the *lacZ* gene preceded by the mycobacterial Ag85 promoter and a derivative of the *B. subtilis sacB* gene preceded by the mycobacterial *hsp60* promoter (Parish and Stoker, 2000). The product of the *sacB* gene, levansucrase, has previously been documented to be toxic to *C. glutamicum* ATCC 13032 cells incubated in the presence of sucrose (Jäger *et al.*, 1992). Following the construction of an allele of the target gene harbouring an internal deletion (Section 2.3.6.1; Section 2.3.6.2); this mutant allele was subsequently cloned into the MCS of p2NIL. The *Pac*I marker gene cassette from pGOAL19 was now inserted into *Pac*I-

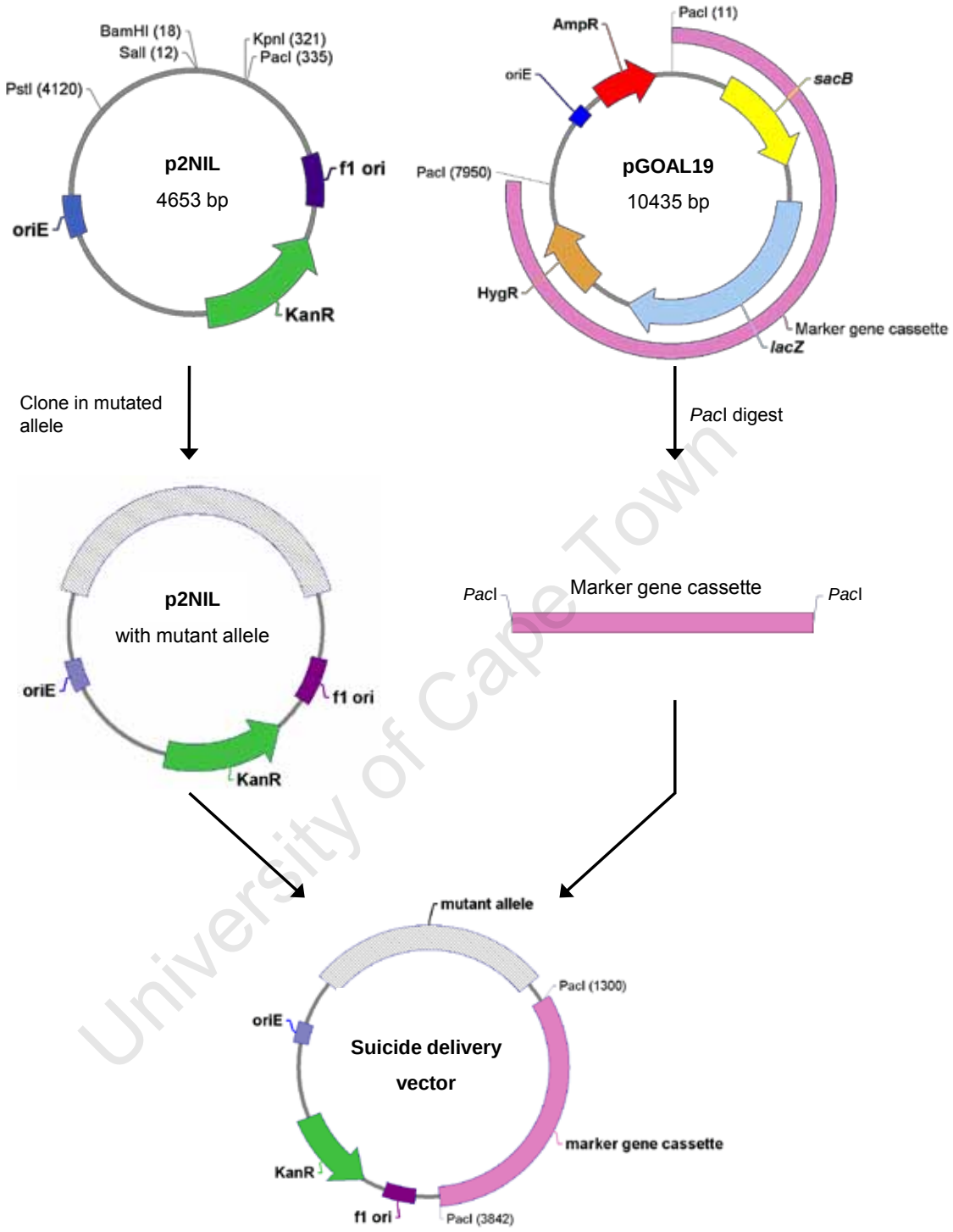


Figure 2.1 Cloning strategy adopted for the construction of suicide delivery vectors. The mutated allele was first cloned into p2NIL. The *Pac*I cassette, containing the Hyg^R, Kan^R, *lacZ* and *sacB* marker genes, was next excised from pGOAL19 and cloned into the *Pac*I site of p2NIL harbouring the mutated allele. The final vector therefore contains the *Pac*I marker gene cassette, an *oriE* and the Kan^R gene.

digested p2NIL containing the mutant allele, thereby resulting in the formation of a mutant delivery vector that harbours several selectable markers and which is incapable of replication in *C. glutamicum* ATCC 13032 cells. This strategy was used for the construction of pARGHKO and pARGRKO, which possesses mutant alleles of the *C. glutamicum* ATCC 13032 *argH* and *argR* genes respectively.

2.3.6.1.1 Construction of pARGHKO

In order to create an allele of the *C. glutamicum* ATCC 13032 *argH* gene containing an internal deletion, two separate fragments, *argH* front and *argH* back, were amplified by PCR from *C. glutamicum* ATCC 13032 genomic DNA (Figure 2.2A). Each of these fragments contains approximately 1 kb of *C. glutamicum* ATCC 13032 genomic DNA flanking either side of the desired deletion. Following amplification, these fragments were purified and individually ligated into pTZ57R/T using the InsTAclone PCR Cloning Kit (Fermentas). The *argH* back fragment cloned into pTZ57R/T was excised with *Bam*HI-*Kpn*I and the resultant 1005 bp fragment cloned into *Bam*HI-*Kpn*I digested p2NIL. The presence of this insert in p2NIL was confirmed by PCR using the p2NIL-F_2 and p2NIL-R primers, which flank the p2NIL *Bam*HI and *Kpn*I restriction sites respectively. The *argH* front fragment cloned into pTZ57R/T was excised with *Pst*I-*Sal*I and the resultant 981 bp fragment cloned into *Pst*I-*Sal*I digested p2NIL harbouring the *argH* back fragment. The *argH* front and *argH* back fragments are now adjacent to each other, thereby forming the 2016 bp *ΔargH* fragment, which contains a mutant allele of the *argH* gene in possession of an internal, in-frame 563 bp deletion. The *argH* junction oligonucleotides, which bind to the 3'-terminal of the *argH* front fragment and the 5'-terminal of the *argH* back fragment, were used to confirm the presence of the mutant allele by PCR and DNA sequencing. The *Pac*I cassette from pGOAL19 was now cloned into the *Pac*I site of this plasmid, giving rise to the pARGHKO suicide vector.

2.3.6.1.2 Construction of pARGRKO

A similar methodology was used to create a mutant allele of the *C. glutamicum* ATCC 13032 *argR* gene (Figure 2.2B). A 2091 bp *argR* fragment, in possession of two primer-introduced restriction enzyme sites (a 5'-*Sal*I site and a 3'-*Bam*HI sites), was PCR-amplified from genomic DNA using the *argR* oligonucleotides and cloned into pTZ57R/T.

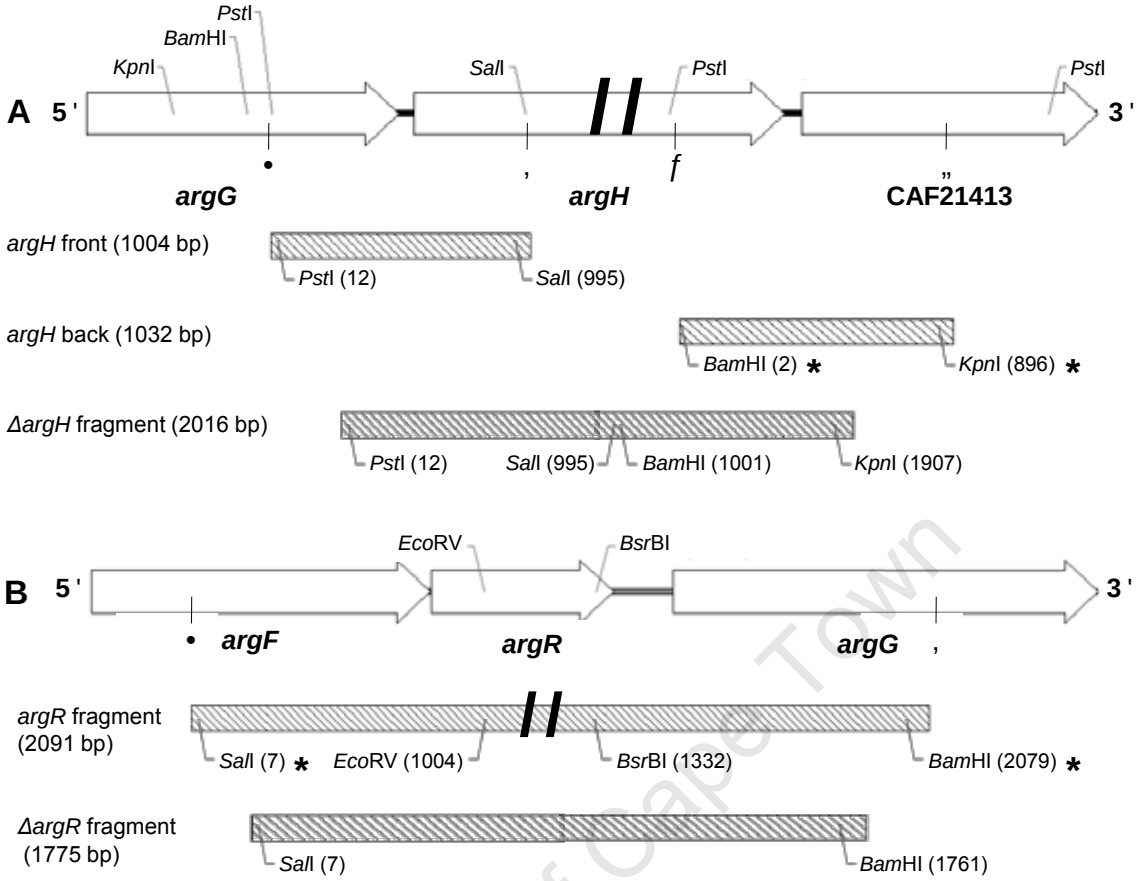


Figure 2.2 Strategy for generating DNA fragments in possession of internal *argH* and *argR* deletions. Asterisks indicate primer-introduced restriction sites. **(A)** In order to construct an internally deleted *argH* gene, the *argH* front and back fragments were amplified by PCR and assembled into p2NIL. The assembly of these two fragments resulted in the formation of the *ΔargH* fragment, which is homologous to a ~2 kb portion of the *C. glutamicum* ATCC 13032 chromosome and contains an in-frame 563 bp *argH* deletion (indicated by the // symbol). The chromosomal position of the *argH* front F and R primers are indicated by • and ‚, whereas the position of the *argH* back F and R primers are indicated by ƒ and „ respectively. **(B)** For the creation of an internally deleted *argR* gene, a ~2 kb portion of the *C. glutamicum* chromosome was amplified using the primers *argR* fragment F and R (the positions of which are indicated by • and ‚ respectively). The central *Eco*RV-*Bsr*BI region was now excised and the terminal fragments of the *argR* fragment rejoined, resulting in the formation of the *ΔargR* fragment, which is homologous to a ~1.8 kb portion of the *C. glutamicum* ATCC 13032 chromosome and contains an in-frame 318 bp *argR* deletion (indicated by the // symbol).

The resulting 2075 bp *Sal*I-*Bam*HI *argR* fragment was excised from this construct and digested with *Eco*RV and *Bsr*BI. Two of the three resulting digested with both *Eco*RV and *Bsr*BI. Two of the three resulting fragments (998 bp and 737 bp) were gel-extracted, blunt-end ligated together and cloned into *Sal*I-*Bam*HI digested pBlueScript SK(+). This resulted in the formation of a 1775 bp *ΔargR* fragment, in possession of in-frame 318 bp deletion within the *argR* gene. The presence of this deletion was confirmed by DNA sequencing and PCR with the *argR* junction primer pair, which flank the *Eco*RV and *Bsr*BI sites. This *Sal*I-*Bam*HI *ΔargR* fragment was next excised from the resultant plasmid and cloned into *Sal*I-*Bam*HI digested p2NIL. The presence of this insert in p2NIL was confirmed by PCR using the p2NIL-F_1 and p2NIL-R primers, which flank the p2NIL *Pst*I and *Kpn*I restriction sites. The *Pac*I cassette was now inserted as outlined previously in order to construct pARGRKO.

2.3.6.2 Generation of single cross-over mutants

In order to improve upon the frequency of appearance of single-cross over (SCO) mutants of *C. glutamicum* ATCC 13032, each suicide delivery vector was UV irradiated in order to introduce pyrimidine dimers into each plasmid molecule, thereby stimulating HR within the host cell after electroporation (Hinds *et al.*, 1999). Approximately 10 µg of either pARGHKO or pARGRKO DNA was resuspended in 10 µl of dH_2O in a sterile Costar microtitre plate well (Corning) and UV irradiated (100 mJ cm^{-2}) using a Hoefer UVC 500 Crosslinker (Amersham Biosciences) (Gordhan and Parish, 2001). An aliquot of the irradiated DNA (1-2 µl) was subsequently electroporated into *C. glutamicum* ATCC 13032 cells using the protocol outlined in Appendix D. Potential SCOs were isolated on 2 × YT solid medium containing Hyg, Kan, IPTG, Nx and X-gal and several individual blue, Hyg^R and Kan^R colonies were selected for further analysis. PCR with the *argH* or *argR* junction primers, in addition to Southern hybridisation, was used to confirm the presence of both WT and mutant alleles in each strain. The p2NIL F_2 and p2NIL R primers, in conjunction with the *argR-argG* F and *CAF21413-CAF21414* R primers, were used to confirm vector integration in cells of the *C. glutamicum* ATCC 13032 *argH* SCO mutant (henceforth designated SCOH). For the PCR-based confirmation of this integration event in genome of the *C. glutamicum* ATCC 13032 *argR* SCO mutant strain (henceforth designated SCOR), the p2NIL F_1 and p2NIL R primers were used in conjunction with the *argD-argF* F and

argG-argR R primers. Any PCR product fragments that matched the expected size were gel-extracted and sequenced to confirm their identity.

2.3.6.3 Southern blot analysis of SCOH and SCOR

Southern blot analysis of the *argH* and *argR* SCO mutants was performed according to the DIG Application Manual for Filter Hybridisation (Roche). pARGHKO and pARGRKO (10 ng each), together with approximately 10 µg of genomic DNA from *C. glutamicum* ATCC 13032 cells and each of the SCO mutants, was digested to completion with either *Bgl*II or *Pvu*II for 16 h at 37 °C. In order to confirm the presence of both *argH* alleles in SCOH, *Bgl*II-digested DNA from pARGHKO, *C. glutamicum* ATCC 13032 genomic DNA and SCOH genomic DNA were separated on a 0.8% (w/v) agarose gel in TAE buffer, contact-blotted onto a Hybond-N^+ nylon membrane (Amersham Biosciences) and hybridised with an internal 1.6 kb DIG-labelled probe for 16 h at 48 °C. An alkaline phosphatase-conjugated anti-DIG antibody (Roche) was used to detect hybridised probe. Chemiluminescent signals were subsequently generated using CSPD (Roche) and captured using a Gene-Gnome Bioimaging System (Syngene). Digital autoradiographs were visualised using the GeneSnap software (version 7.04; Syngene). A similar approach was adopted so as to confirm the presence of both WT and mutant *argR* alleles in the SCOR genome, except *Pvu*II-digested DNA was utilised and an internal 1.5 kb DIG-labelled probe fragment was used for hybridisation at 43 °C.

2.3.6.4 Attempted isolation of double cross-over mutants

Different molecular and phenotypic approaches were adopted in order to screen for *ΔargH* and *ΔargR* DCO mutants. Several SCO colonies of each type was restreaked on 2 × YT agar supplemented with IPTG, Nx and X-gal in order to allow for a second cross-over event to occur. After 48 hours growth (equivalent to eleven generations) at 30 °C, the contents of each plate were scraped off in 1 ml of 2 × YT broth and serially diluted. Each dilution series was plated on 2 × YT agar with or without sucrose and supplemented with IPTG, Nx and X-gal. For the isolation of *ΔargH* mutants, L-arginine was also incorporated into this medium. Colonies that appeared white and sucrose resistant are possible DCO mutants. In order to screen for L-arginine auxotrophy, over 1000 white and sucrose resistant potential *ΔargH* DCO mutants were plated in replicate on CGXII MM agar supplemented with or without L-

arginine. Moreover, for the attempted isolation of DCO mutants, approximately 500 white, sucrose resistant potential colonies from each strain were screened for Kan^S by replica plating on 2 × YT agar supplemented with IPTG, Nx, sucrose, X-gal and Kan. Any colonies that appeared either L-arginine auxotrophic or Kan^S were analysed further by PCR using both the appropriate junction primers and the *sacB* primers, which were designed to anneal to an internal fragment of the *sacB* gene located within the *Pac*I marker gene cassette. Importantly, this screening process was applied to potential DCO colonies derived from several *argH* and *argR* SCO mutants, not just SCOH and SCOR.

2.3.7 Generation of an *argR* double cross-over mutant using the pK19*mobsacB* vector

As we were unable to generate any *C. glutamicum* ATCC 13032 *argR* DCO mutant cells using the p2NIL/pGOAL19 vector system, the pK19*mobsacB* suicide plasmid, which has previously been used extensively for allelic exchange in this bacterium, was used to create a new *argR* SCO mutant (designated SCOR-D3) (Jakoby *et al.*, 2000; Schäfer *et al.*, 1994; Simic *et al.*, 2001; Tauch *et al.*, 2002). An *argR* DCO mutant (designated DCOR-D3) was subsequently successfully isolated from SCOR-D3.

2.3.7.1 Verification of *sacB* integrity

Prior to the construction of a new pK19*mobsacB*-derived *argR* suicide plasmid (designated pARGR-MS), the integrity of the vector-encoded *sacB* gene was verified (Eggeling and Reyes, 2005). Briefly, *E. coli* DH5α cells transformed with pK19*mobsacB* were restreaked on 2 × YT solid medium containing Kan and 10% sucrose and, after approximately 12 h of incubation, any colonies that displayed a clear growth disadvantage relative to WT cells were restreaked on the same medium. If these plasmid-containing cells repeatedly exhibited this negative growth phenotype in the presence of sucrose, they were assumed to contain a fully functional *sacB* gene and were hereafter used for the isolation of pK19*mobsacB* DNA (Eggeling and Reyes, 2005).

2.3.7.2 Construction of pARGR-MS

A 1795 bp DNA fragment, in possession of a primer-introduced 5'-*Hind*III site and a 3'-*Bam*HI site, was PCR-amplified from pARGRKO DNA using the *argR*-MS primer pair. This

argR-MS product fragment, which is in possession of the mutant *argR* allele described in Section 2.3.6.3, was next gel-purified, digested with *Bam*HI and *Hind*III and ligated into *BamH*I-*Hind*III digested pK19*mobsacB* DNA. The presence of insert in the resulting plasmid was confirmed by sequencing using the pK19 primer pair, which anneal on either side of the pK19*mobsacB* MCS. This plasmid was now UV-irradiated and electroporated into *C. glutamicum* ATCC 13032 cells as described respectively in Section 2.3.8 and Appendix D.

2.3.7.3 Generation of single and double cross-over mutants

In order to screen for colonies in possession of only the *ΔargR* allele, ten putative chromosomal integrants were selected on 2 × YT medium containing Kan, individually inoculated into 5 ml of 2 × YT broth and allowed to grow for 16 h. A serial dilution of each culture was plated in quadruplicate on 2 × YT agar with or without either sucrose, Kan or both. Cultures that yielded a relatively low colony count on 2 × YT medium containing Kan and sucrose, yet which also displayed comparatively high numbers on 2 × YT medium containing Kan, are likely to contain a subpopulation of sucrose sensitive SCO cells and hence represent an ideal population for the isolation of DCO cells (Eggeling and Reyes, 2005). If a particular culture displayed these characteristics, it was restreaked in duplicate from 2 × YT agar onto sucrose-containing medium with and without Kan. If any of these colonies displayed both sucrose resistance and Kan^S, they are likely to represent either WT revertants or DCO mutants and were therefore selected for further molecular analysis.

2.3.7.4 Mutant confirmation

A PCR-based approach using the *argR* gene primer pair, which binds externally to the *argR* gene, was used to initially confirm the presence of the appropriate *argR* allele in both SCO and DCO mutant cells. If any of these mutant cells reliably displayed PCR-product fragments indicative of the presence of either only the mutant allele or both the mutant and WT alleles, their genomic DNA was isolated and, in conjunction with WT DNA genomic DNA and pARGR-MS DNA, subsequently analysed by Southern blotting in a manner similar to that outlined in Section 2.3.6.3. This DNA was digested to completion with *Eco*RV and hybridised at a temperature of 45 °C to a 522 bp DIG-labelled DNA probe internal to *argR* produced via PCR performed with the *argR* gene primer pair.

2.4 Results

2.4.1 Sequence analysis of the *C. glutamicum* ATCC 13032 ArgH and ArgR proteins

The predicted *C. glutamicum* ATCC 13032 ArgH and ArgR amino acid sequences were functionally annotated by Kalinowski *et al.* (2003) according to their homology with similar, previously biochemically characterised proteins. In order to confirm that the *C. glutamicum* ATCC 13032 *argH* and *argR* genes do in fact appear to encode proteins with either ASL or L-arginine repressor activity, multiple amino acid sequence alignments were constructed using ArgH and ArgR sequences from a variety of different organisms shown in Figure 2.3 and Figure 2.4. Comparison of the predicted ArgH sequences from these organisms revealed a relatively high degree of similarity between ArgH proteins from *C. glutamicum* ATCC 13032 and the closely related bacteria *M. tuberculosis* CDC1551 and *S. clavuligerus* (possessing 73% and 61% amino acid identity respectively) (Figure 2.3). A significantly lower degree of ArgH amino acid identity was observed between *C. glutamicum* ATCC 13032 and *E. coli* CFT073, *B. subtilis subsp. subtilis str.* 168 and *L. plantarum* WCFS1 (all of which share approximately 40% identity with the *C. glutamicum* ATCC 13032 ArgH protein). A high degree of amino acid conservation was observed in these bacteria between residues shown previously to be responsible for the formation of the ArgH active site in *Homo sapiens* (Turner *et al.*, 1997). As evident from Figure 2.3, the in-frame 563 bp *argH* deletion introduced into *C. glutamicum* ATCC 13032 would result in the formation of an internally deleted ArgH protein lacking the majority of these active sites.

The predicted ArgR sequences share significant amino acid sequence identity with the *C. glutamicum* ATCC 13032 ArgR protein (56% and 49% amino acid identity respectively for *M. tuberculosis* CDC1551 and *S. clavuligerus*) (Figure 2.4). Distantly related organisms, such as *E. coli* CFT073, possess a smaller degree of ArgR amino acid sequence identity relative to *C. glutamicum* ATCC 13032 (28%). Furthermore, the *C. glutamicum* ATCC 13032 ArgR is in possession of several conserved amino acid residues that have previously been shown to be essential for the formation of critical secondary structure elements in the ArgR protein of *Geobacillus stearothermophilus* (Figure 2.4) (Ni *et al.*, 1999). The in-frame 318 bp deletion introduced into the *C. glutamicum* ATCC 13032 *argR* would thus result in the formation of an internally deleted ArgR protein, lacking significant portions of

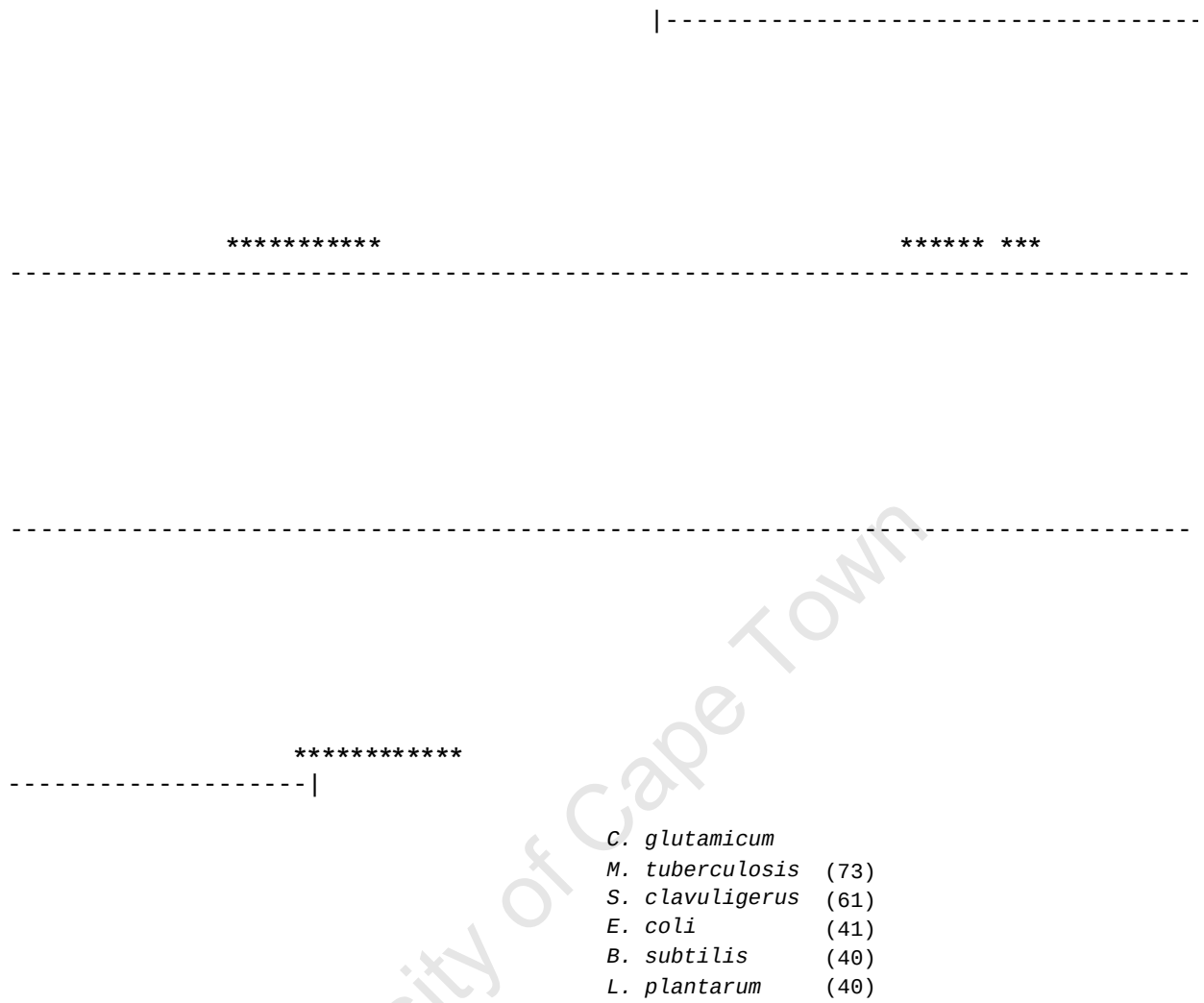


Figure 2.3 Comparison of the predicted ArgH amino acid sequences from various organisms. A fragment of each ArgH containing the putative active site is shown for each organism. Highly conserved residues essential to the formation of the ASL active site are marked with asterisks below the sequence. Identical amino acids are shaded black, conserved residues are shaded dark grey and similar residues are shaded light grey. The dashed line above the sequence represents the region of the *C. glutamicum* ArgH proposed to be deleted in the mutant *argH* allele (amino acids 144 to 367). The percentage of ArgH amino acid sequence identity for each organism relative to the *C. glutamicum* ATCC 13032 ArgH is indicated in parentheses.

the N-terminal DNA binding domain and the C-terminal hexameric oligomerisation domain.

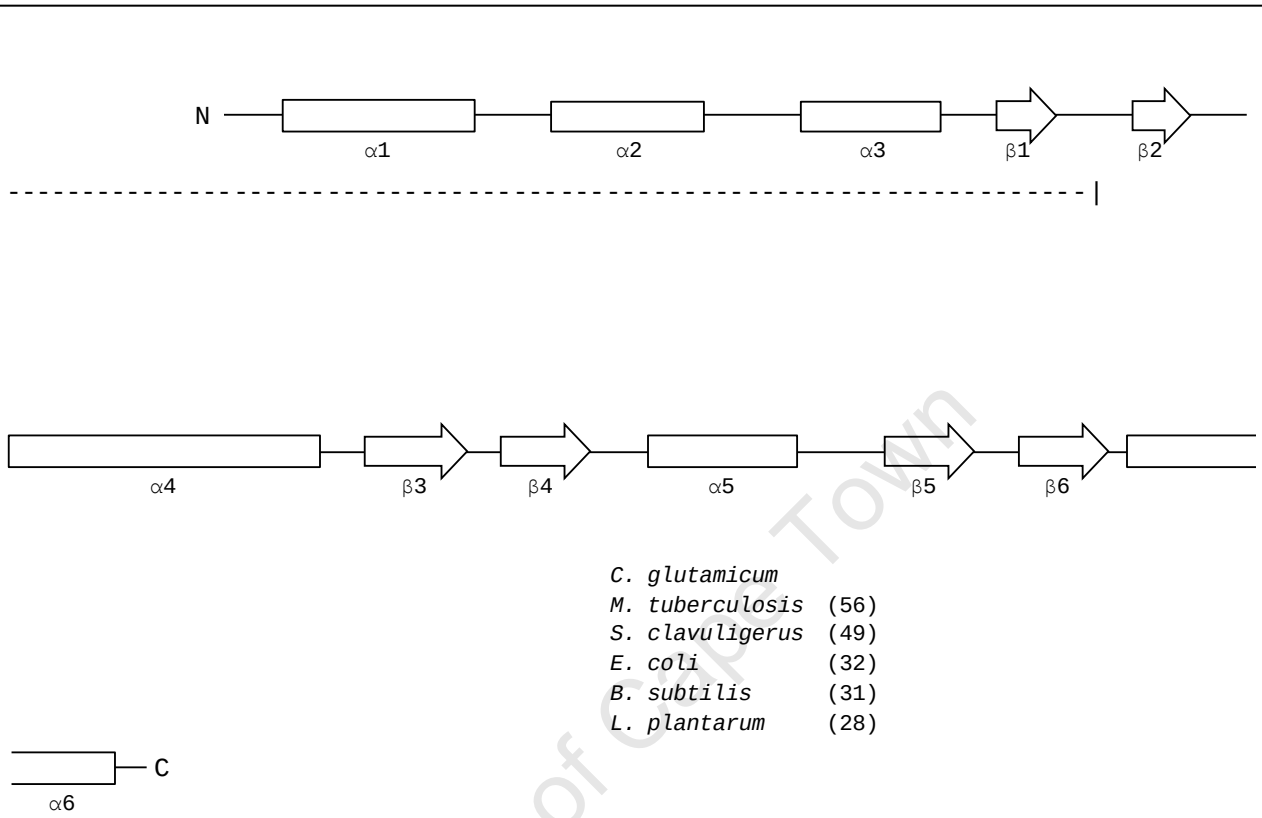


Figure 2.4 Comparison of predicted ArgR amino acid sequences from various organisms. Identical amino acid residues are shaded black, conserved residues are shaded dark grey and similar residues are shaded light grey. The dashed line above the sequence represents the region of the *C. glutamicum* ArgR proposed to be deleted in aberrant ArgR molecules encoded by the mutant *argR* allele (amino acids 43 to 155). α-Helices (rectangles) and ß-strands (arrows) are positioned according to the crystal structure of *G. stearothermophilus* (Ni *et al.*, 1999). The N-terminal DNA binding domain (amino acids 1 to 80) of the *E. coli* ArgR, which terminates after the second ß-strand, and the C-terminal hexameric oligomerisation domain (amino acids 86 to 180) of the same protein, which starts at the beginning of the fourth α-helix, are shown. The percentage of amino acids of each ArgR that are identical to the *C. glutamicum* ATCC 13032 ArgR is indicated in parenthesis for each organism.

The *C. glutamicum argH* and *argR* mutant alleles are hence proposed to encode aberrant proteins that lack key structural features and are thus likely to be non-functional. *C. glutamicum argH* and *argR* integration mutants, either possessing *ΔargH* or *ΔargR* in

addition to the appropriate WT allele, were now generated with the end goal of producing *C. glutamicum* DCO strains that harbour only the appropriate mutant allele.

2.4.2 Construction of *C. glutamicum argH* and *argR* single cross-over mutants using the p2NIL/pGOAL19 system

The p2NIL/pGOAL19 system used here to generate SCO mutants in *C. glutamicum* relies on a recombination event occurring between two homologous regions of DNA (Friedberg *et al.*, 1995). Following the transformation of the suicide vector into *C. glutamicum* ATCC 13032, the entire non-replicating plasmid may integrate into a homologous region on the chromosome via HR, thereby resulting in the generation of an integration mutant in possession of the plasmid marker genes in addition to both forms of the target gene (Figure 2.5; Figure 2.7) (Gordhan and Parish, 2001). In this case, HR would occur either between the chromosomal *argH* or *argR* fragments, which contain the *argH* or *argR* WT alleles in addition to 2 kb of flanking chromosomal DNA, and the plasmid encoded *ΔargH* or *ΔargR* alleles, which are located respectively on pARGHKO or pARGRKO.

2.4.2.1 Confirmation of an *argH* integration mutant

The integration of pARGHKO into the *C. glutamicum* ATCC 13032 *argGH* operon can result in the formation of two different types of SCO mutants that vary according to the site of plasmid integration (Figure 2.5). Following the electroporation of pARGHKO into *C. glutamicum* ATCC 13032, three blue, Hyg^R and Kan^R putative SCO colonies were obtained and selected for further analysis. Each of these colonies grew on 2 × YT agar supplemented with Nx and sucrose, although growth was generally much slower relative to WT *C. glutamicum* ATCC 13032 cells on the same medium. As evident from Figure 2.6A, when these colonies were analysed by PCR using the *argH* junction oligonucleotide primers, each displayed PCR product fragments of the expected sizes of 1279 bp and 716 bp. Each of these, in turn, appeared to be present in variable quantities for each colony. These correspond in size to the expected PCR products indicative of the WT and mutant *argH* alleles, which are shown respectively in lanes one and two of Figure 2.6A. The appearance of a third PCR product of 900 bp was also observed for these SCO colonies, although such a fragment is likely due to non-specific annealing of the *argH* junction primers, as this product was also observed under identical conditions when a *C. glutamicum* ATCC 13032

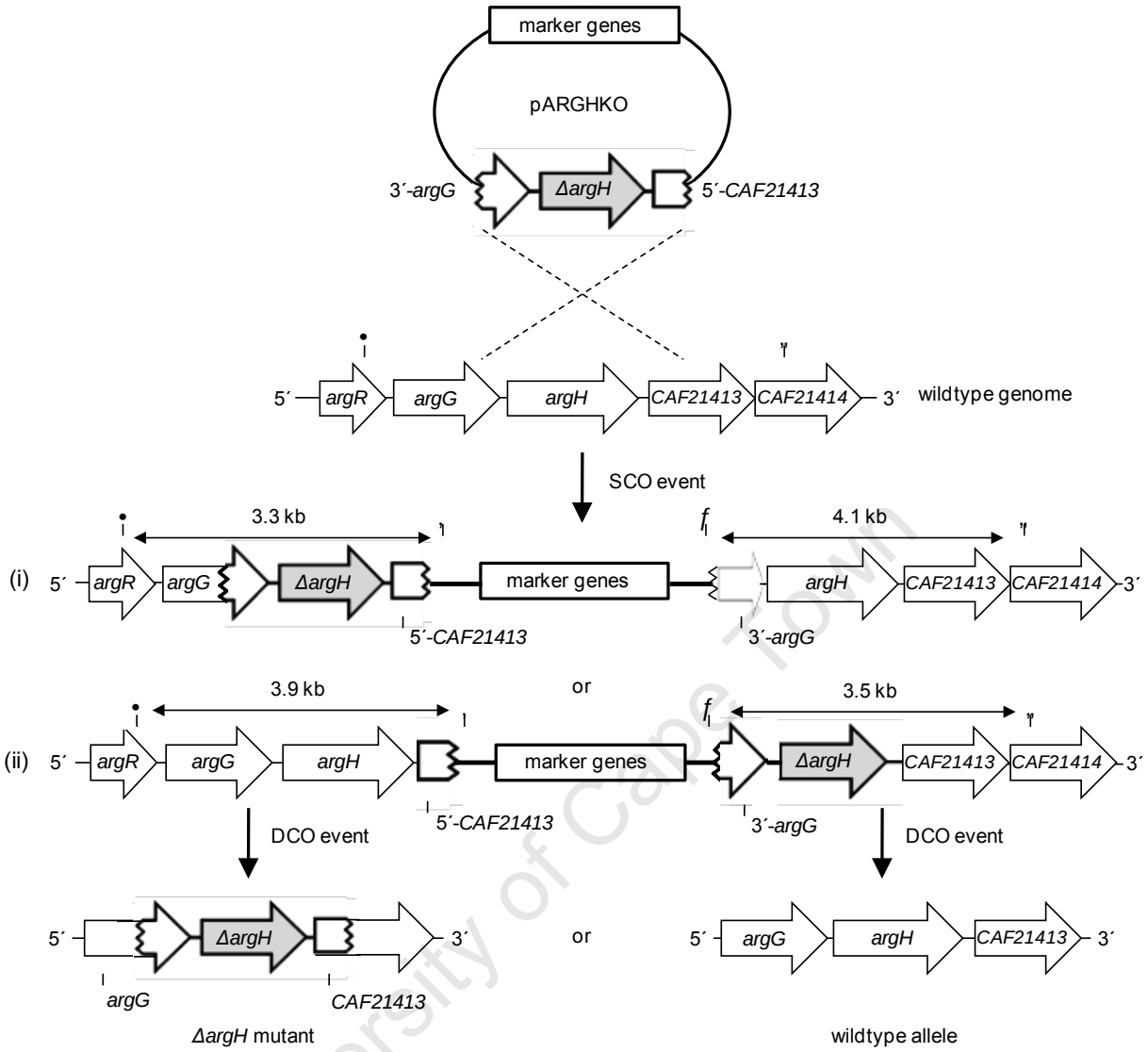


Figure 2.5 Allelic replacement of *argH* by HR. The non-replicating pARGHKO was electroporated into *C. glutamicum* ATCC 13032, where HR may lead to the formation of a SCO mutant that carries both WT and mutant *argH* alleles, in addition to any plasmid marker genes. DCO colonies carry a single copy of either the WT or mutant gene and possess no plasmid marker genes. PCR with primers which anneal to either vector or genomic DNA was used to confirm the location of the pARGHKO integration event. The chromosomal position of the *argR-argG* F and p2NIL R primers are indicated by • and ‚, whereas the position of the p2NIL F_2 and *CAF21413-CAF21414* R primers are indicated by ƒ and „. The thick black line represents vector DNA and the thin black line represents *C. glutamicum* genomic DNA. The black arrows represent the expected approximate PCR product sizes for each primer pair. (i) and (ii) denote the proposed outcome of different integration events. Dashed lines represent possible HR sites. The mutant *argH* allele is shaded blue. Diagram not to scale. Adapted from Gordhan and Parish (2001).

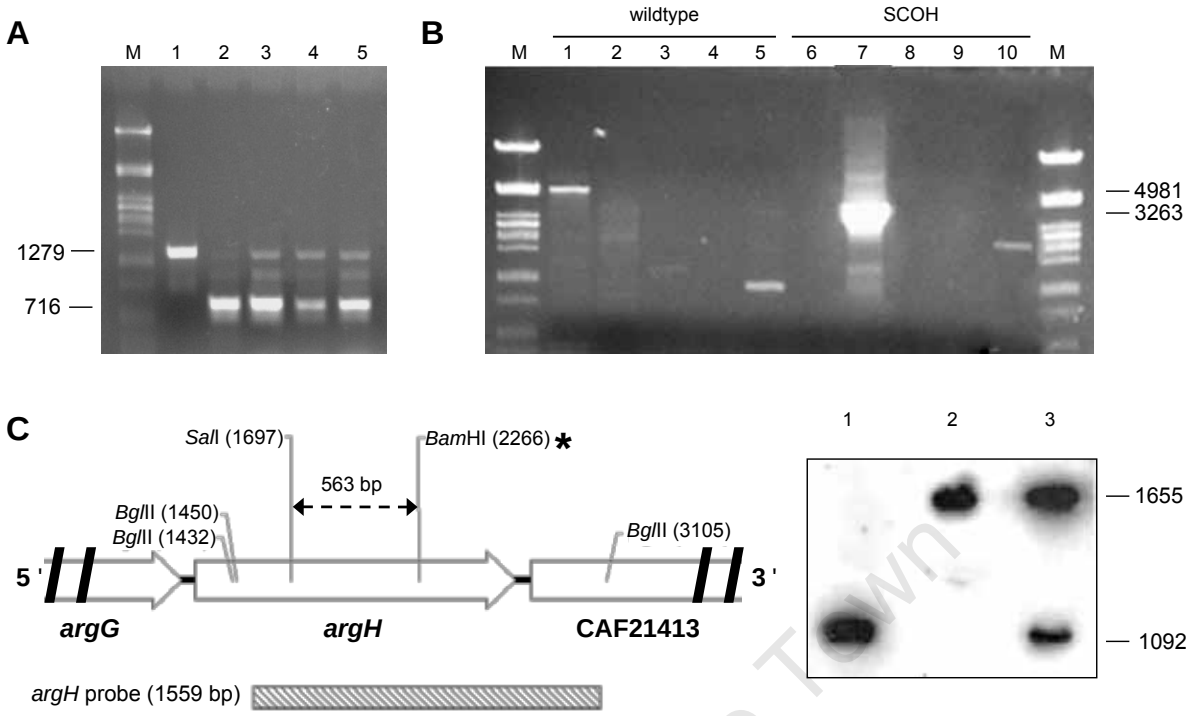


Figure 2.6 Confirmation of pARGHKO integration into *C. glutamicum* SCOH. Fragment sizes are indicated in bp. Lanes M: marker DNA. **(A)** Results of a PCR performed using the *argH* junction primers showing the presence of *ΔargH* in different DNA templates. Lane 1: WT genomic DNA; lane 2: pARGHKO; lanes 3-5: putative SCOH colonies. **(B)** PCR-based confirmation of pARGHKO insertion into the *C. glutamicum* ATCC 13032 *argH* region performed with different primer combinations. Lanes 1-5 contain PCR products amplified from WT genomic DNA whereas lanes 6-10 contain PCR products amplified from SCOH genomic DNA. The PCR primer combinations used for each reaction are as follows: lanes 1 and 6, *argR-argG* F and *CAF21413-CAF21414* R; lanes 2 and 7, *argR-argG* F and p2NIL R; lanes 3 and 8, *argR-argG* F and p2NIL F_2; lanes 4 and 9, *CAF21413-CAF21414* R and p2NIL R; lanes 5 and 10, *CAF21413-CAF21414* R and p2NIL F_2. Their respective binding sites are shown in Figure 2.5. **(C)** Southern blot confirmation of *argH* and *ΔargH* in SCOH. A restriction map of the *C. glutamicum* ATCC 13032 *argH* gene, including flanking chromosomal regions, is shown on the left. The dashed arrow represents the region deleted in *ΔargH*. The hatched rectangle represents the DIG-labelled probe DNA fragment used for hybridisation. The *Bam*HI site indicated by the asterisk is not present in the WT chromosome and was introduced into *ΔargH* using the *argH* back F primer. The autoradiograph shows *Bgl*II-digested DNA hybridised with the DIG-labelled ~1.6 kb *argH* probe. Lane 1: pARGHKO DNA; lane 2: WT genomic DNA; lane 3: SCOH genomic DNA.

WT colony was used as template DNA. Elevating the PCR annealing temperature did not reduce the appearance of this 900 bp fragment. A single putative *argH* SCO colony, designated SCOH, was now selected for further analysis by PCR and Southern blotting.

With a view to confirm the position and orientation of the pARGHKO integration event in SCOH, different combinations of various PCR oligonucleotides, which anneal to either genomic or vector DNA, were used to analyse SCOH (Figure 2.6B). As expected, a 5 kb product fragment was observed using the two *argR-argG* F and *CAF21413-CAF21414* R primers, which are predicted to bind to only *C. glutamicum* ATCC 13032 genomic DNA. This PCR product fragment was not observed when the same primer pair was used with SCOH genomic DNA, likely due to the integration of pARGHKO. As shown in lane seven of Figure 2.6B, the *argR-argG* F and p2NIL R primer pair produced large quantities of a 3.3 kb PCR product fragment that corresponds to the first type of SCO event shown in Figure 2.5. Sequencing revealed this fragment to possess the *argH* mutant allele, in addition to pARGHKO vector DNA, and match the expected region shown in Figure 2.5. In accordance with this position of integration, a 4.1 kb PCR product fragment was expected from SCOH using the *CAF21413-CAF21414* R and p2NIL F_1 primers, although, as shown in lane ten of Figure 2.6B, only a 2.8 kb fragment was obtained. Reducing the PCR annealing temperature did not improve the formation of the expected 4.1 kb fragment. The 2.8 kb PCR product fragment was sequenced and found to be as a result of non-specific primer binding. Additionally, Southern blot analysis of SCOH revealed the *argH* probe fragment to hybridise to two *Bgl*II-DNA fragments (1655 bp and 1092 bp) that correspond to the *argH* WT and mutant alleles (Figure 2.6C).

SCOH contains both *argH* alleles and the integration of pARGHKO into this strain appears to have occurred in the first position outlined in Figure 2.5. Hereafter this strain was considered an authentic *C. glutamicum argH* SCO integration mutation and was subsequently used for the attempted isolation of *argH* DCO mutants.

2.4.2.2 Confirmation of an *argR* integration mutant

The integration of pARGRKO into the *C. glutamicum* ATCC 13032 *argCJBDF* operon was confirmed using an approach similar to that adopted for SCOH (Figure 2.7; Figure 2.8). Following the electroporation of pARGRKO into *C. glutamicum* ATCC 13032, ten blue,

Hyg^R and Kan^R putative SCO colonies were obtained and five selected for further analysis. Similar to SCOH, all of these colonies displayed slower growth relative to *C. glutamicum* ATCC 13032 in the presence of sucrose (result not shown). As evident from Figure 2.8A, each of these putative SCO colonies displayed *argR* junction PCR product fragments that correspond to the *argR* WT (1236 bp) and mutant alleles (918 bp) shown in lanes one and two respectively. PCR-based analysis of one such colony (designated SCOR) with the primer combinations shown in Figure 2.7 produced a large number of non-specific product fragments, whose formation could not be limited through the adjustment of PCR conditions without a significant reduction in the formation of the desired PCR fragments. As expected, primers *argD-argF* F and *argG-argH* R, produced a 4.3 kb PCR product fragment when bound to *C. glutamicum* ATCC 13032 genomic DNA. Indicative of the first type of SCO event shown in Figure 2.7, the *argD-argF* F and p2NIL R primer pair produced large quantities of a 3.8 kb PCR product fragment from SCOR genomic DNA (as shown in lane seven of Figure 2.8B). Sequencing revealed this fragment to possess the *argR* mutant allele, in addition to a fragment of pARGRKO vector DNA, and to be homologous to the expected region shown in Figure 2.7. Similar to the *argR-argG* F and p2NIL R primer combination used for SCOH confirmation, no significant PCR product fragments were observed with the *argG-argH* R and p2NIL F_1 primer pair used on SCOR genomic DNA for confirmation of pARGRKO integration (as shown in lane ten of Figure 2.8B). A Southern blot analysis of SCOR, however, revealed an internal *argR* probe fragment to hybridise to two *Pvu*II-digested DNA fragments (1532 bp and 1217 bp), which correspond to the *argR* WT and mutant alleles (Figure 2.8C).

SCOR contains both WT and mutant alleles of *argR* and, similar to what was observed for pARGHKO integration into SCOH, the integration of pARGRKO into SCOR appears to have occurred in the first position indicated in Figure 2.7. This *C. glutamicum argR* SCO integration mutation was now used for the attempted isolation of *argR* DCO mutants.

2.4.2.3 Screening for *C. glutamicum argH* and *argR* double cross-over mutants

In order to generate unmarked *argH* and *argR* DCO mutants, several SCO colonies, including SCOH and SCOR, were restreaked in the absence of antibiotic selective pressure (Gordhan and Parish, 2001). Due to the absence of any plasmid marker genes after the DCO event, these colonies would be sucrose resistant and appear white on media

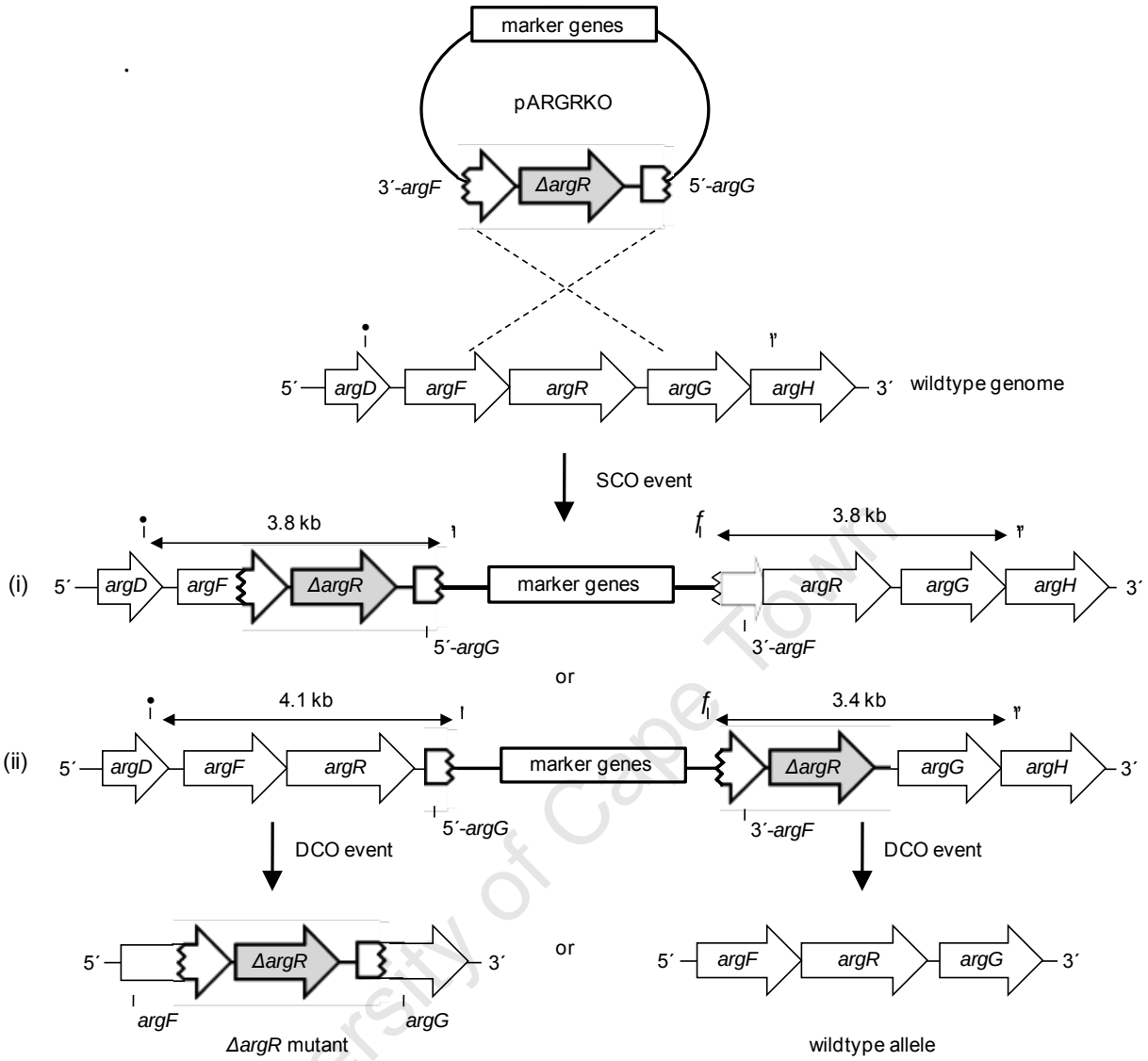


Figure 2.7 Allelic replacement of *argR* by HR. pARGRKO was electroporated into *C. glutamicum* ATCC 13032, where it may integrate into a homologous region of the chromosome and result in the formation of a SCO mutant. DCOs carry a single copy of either the wildtype or mutant gene and possess no plasmid marker genes. PCR with primers which anneal to either vector or genomic DNA may be used to confirm the pARGRKO integration event. The chromosomal position of the *argD-argF* F and p2NIL R primers are indicated by • and , , whereas the position of the p2NIL F_1 and *argG-argH* R primers are indicated by *ƒ* and „ . The black arrows represent the expected approximate PCR product sizes for each primer pair. (i) and (ii) denote the proposed outcome of different integration events. The dashed lines represent possible sites of HR. The thick black line represents vector DNA and the thin black line represents *C. glutamicum* genomic DNA. The mutant *argR* allele is shaded blue. Diagram not to scale. Adapted from Gordhan and Parish (2001).

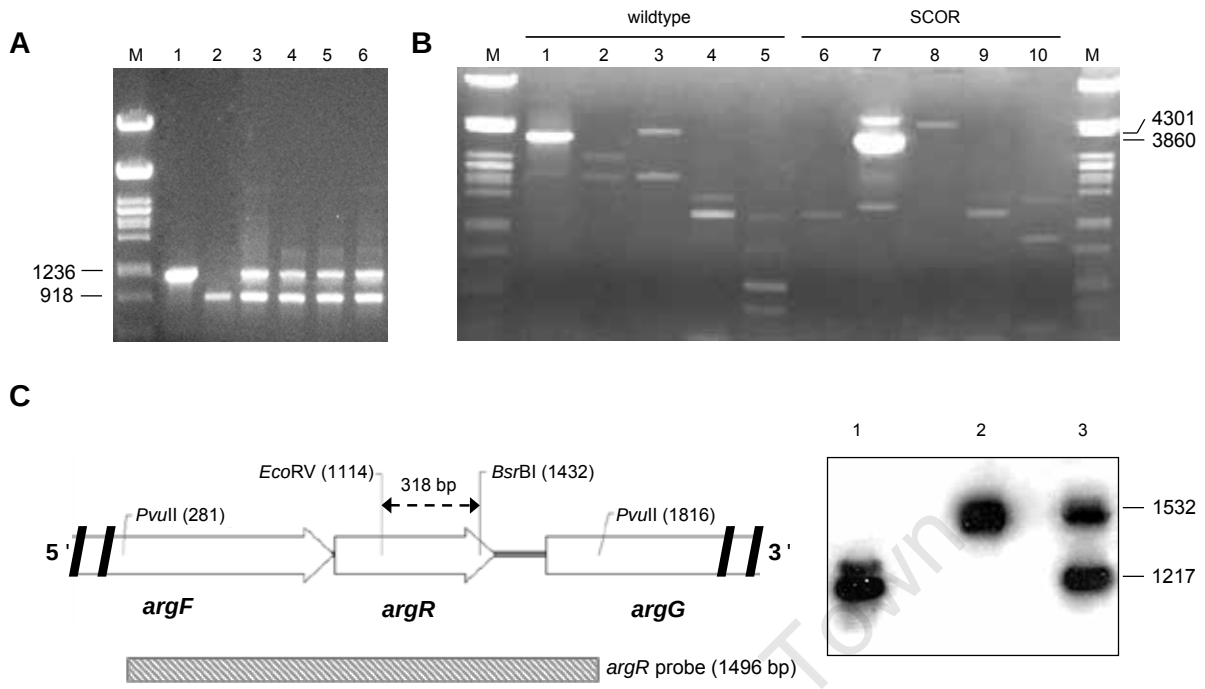


Figure 2.8 Confirmation of pARGRKO integration in *C. glutamicum* ATCC 13032. Fragment sizes are indicated in bp. Lanes M: marker DNA. **(A)** PCR-analysis with the *argR* junction primers to show construction of the *ΔargR* fragment. Each lane contains PCR products amplified from different DNA templates. Lane 1: wildtype genomic DNA; lane 2: pARGRKO; lanes 3-5: putative SCOR colonies. **(B)** PCR-based analysis performed with different primer combinations to confirm the insertion of pARGRKO into the chromosomal regions flanking the *C. glutamicum* ATCC 13032 *argR* gene. Lanes 1-5 contain PCR products amplified from wildtype genomic DNA whereas lanes 6-10 contain PCR products amplified from SCOR genomic DNA. The PCR primer combinations for each lane are as follows: lanes 1 and 6, *argD-argF* F and *argG-argH* R; lanes 2 and 7, *argD-argF* F and p2NIL R; lanes 3 and 8, *argD-argF* F and p2NIL F_1; lanes 4 and 9, *argG-argH* R and p2NIL R; lanes 5 and 10, *argG-argH* R and p2NIL F_1. Their respective binding sites are shown in Figure 2.7. **(C)** Southern blot confirmation of *argR* and *ΔargR* in SCOR. A restriction map of the *C. glutamicum* ATCC 13032 *argR* gene, including flanking chromosomal regions, is shown on the left. The dashed arrow represents the region deleted in *ΔargH*. The hatched box represents the DIG-labelled probe DNA fragment used for hybridisation. The autoradiograph contains *Pvu*II-digested DNA in each lane hybridised with the DIG-labelled ~1.6 kb *argR* probe. Lane 1: pARGRKO DNA; lane 2: wildtype genomic DNA; lane 3: SCOR genomic DNA.

containing IPTG, sucrose and X-gal. Upon plating onto this medium, roughly 10% of the colonies appeared blue and grew readily. Such colonies are spontaneous sucrose resistant mutants that contain a defective *sacB* gene and are still in possession of the marker gene cassette (Gordhan and Parish, 2001). In contrast, when an aliquot of the same cell mixture was plated on a similar medium lacking sucrose, approximately 10-fold more CFU were observed, the majority of which appeared blue. Colonies of the white, sucrose resistant phenotype were initially analysed directly by PCR using the appropriate junction primers. After screening approximately 100 colonies in this manner, however, all produced PCR product fragments indicative of both alleles of the target gene. Therefore, in an attempt to improve upon the isolation of *argH* and *argR* DCO mutant colonies, additional rounds of plating in the absence of antibiotic selective pressure were introduced. Furthermore, so as to enhance the emergence of mutant DCO colonies, a strategy that employed an additional selection step, by screening for either L-arginine auxotrophy or Kan^S, followed by two PCR-based confirmation steps, was now adopted.

Putative *ΔargH* DCO colonies that displayed a weak growth phenotype on CGXII MM solid medium relative to *C. glutamicum* ATCC 13032 cells were analysed by PCR using the *argH* junction primers (Figure 2.9A). Furthermore, potential *ΔargH* DCO colonies, together with potential *ΔargR* DCO colonies, were screened for Kan^S, followed by screening for the presence of the mutant allele using a PCR analysis step with the appropriate junction primers (Figure 2.9A; Figure 2.9C). Regardless of the presence of a PCR product fragment indicative of the WT allele of the target gene, any potential *ΔargH* or *ΔargR* DCO colonies that displayed a PCR product fragment that corresponded to the appropriate mutant allele were analysed further using the *sacB* primers (Figure 2.9B; Figure 2.9D).

Most of the putative *ΔargH* DCO colonies screened for L-arginine auxotrophy generated significant quantities of a PCR product fragment suggestive of the presence of the *argH* mutant allele (716 bp), in addition to smaller quantities of a product fragment indicative of the WT *argH* allele (1279 bp) (Figure 2.9A). In order to confirm that these colonies had lost the marker gene cassette, they were analysed by PCR using the *sacB* primers, for which they all proved to be positive (Figure 2.9B). These colonies appear to still be SCO mutants and, although the allowance of more time for the DCO event in the absence of antibiotic selection did reduce the appearance of SCOH colonies, *ΔargH* DCO mutants were still not

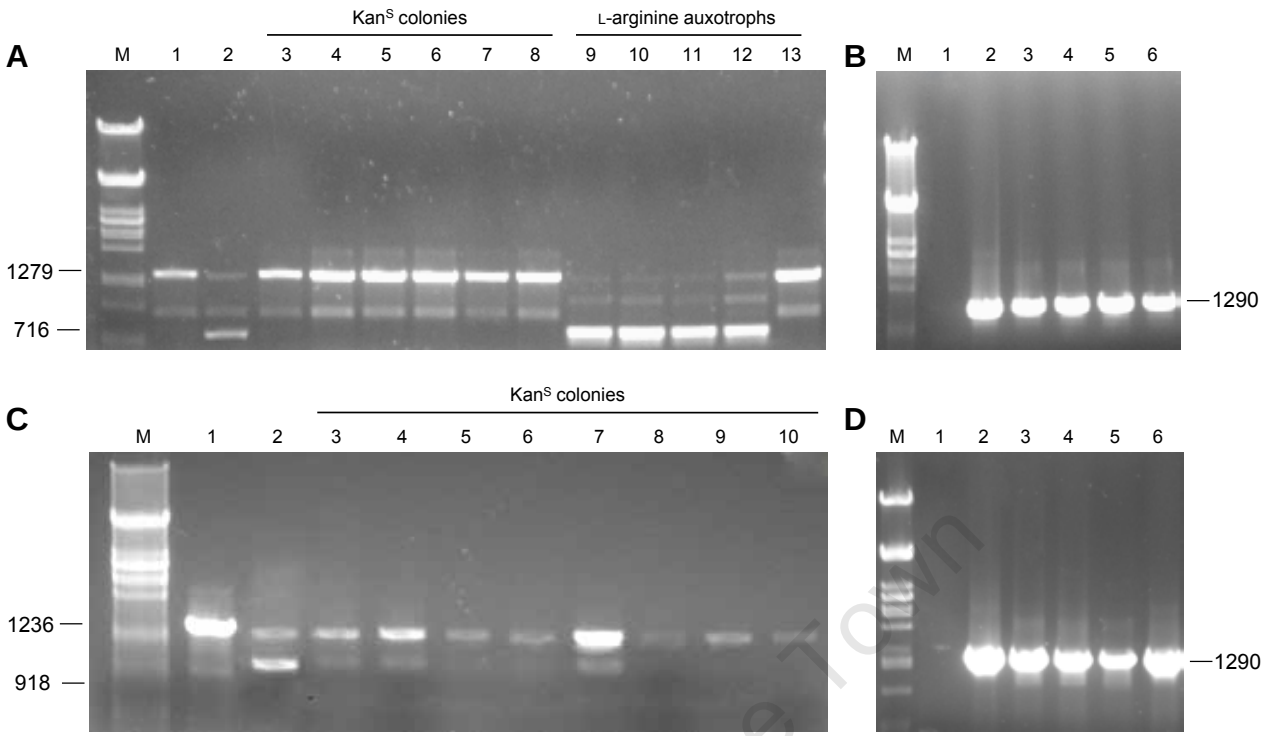


Figure 2.9 PCR-based screening of putative *ΔargH* and *ΔargR* DCO mutants derived from SCOH or SCOR. Several hundred colonies derived from each SCO mutant strain were PCR-screened and representative results are shown. Fragment sizes are indicated in bp. Lanes M: marker DNA. **(A)** Results of a colony PCR performed on putative *ΔargH* DCO mutants using the *argH* junction primers. White, sucrose resistant colonies that appeared either Kan^S (lanes 3-8) or auxotrophic for L-arginine biosynthesis (lanes 9-13), were screened for the presence of the mutant *argH* allele. Control PCR reactions using either wildtype genomic DNA (lane 1) or SCOH genomic DNA (lane 2) are shown. **(B)** PCR-based analysis performed using the *sacB* primers on white, sucrose resistant colonies that displayed the *argH* mutant allele in lanes 9-12 of (A). Control PCR reactions using wildtype genomic DNA (lane 1) or SCOH genomic DNA (lane 2) are shown. PCR products obtained from these colonies using the *sacB* primers are shown in lanes 3-6. **(C)** Results of a colony PCR performed using the *argR* junction primers on putative *ΔargR* DCO mutants. White, sucrose resistant colonies that appeared Kan^S (lanes 3-10) were screened for the presence of the mutant *argR* allele. Control PCR reactions using either wildtype genomic DNA (lane 1) or SCOR genomic DNA (lane 2) are shown. **(D)** PCR-based screening with the *sacB* primers performed on white, sucrose resistant colonies which displayed the *argR* mutant allele (lanes 3-6) in (C). Control PCR reactions using either wildtype genomic DNA (lane 1) or SCOR genomic DNA (lane 2) as template DNA are shown.

detected. Following this increased time allowed for the DCO event, however, an increased proportion of colonies that displayed *argH* junction PCR product fragments indicative of the WT allele were observed (Chapter 3). Furthermore, any putative *ΔargH* DCO colonies screened directly for Kan^S all appeared to be revertants, harbouring only the WT allele of the *argH* gene (Figure 2.9A). Similarly, populations of putative *ΔargR* DCO colonies screened for Kan^S mostly consisted of WT revertants and SCOR mutants that had persisted, despite the removal of antibiotic selective pressure (Figure 2.9C; Figure 2.9D).

The repeated phenotype- and PCR-based screening for putative DCO colonies derived from SCOH and SCOR thus failed to isolate a true unmarked mutant in possession of only the target mutant allele. The use of the pK19*mobsacB* plasmid for DCO mutant generation was therefore subsequently investigated and the SCOH and SCOR strains, which possess disrupted L-arginine biosynthetic operons, were stored for use in further studies (Chapter 3).

2.4.3 Construction of a *C. glutamicum argR* double cross-over mutant using pK19*mobsacB*

In order to isolate DCO mutants created using the pK19*mobsacB* plasmid, the functional stability of the counter-selectable *sacB* marker was first verified. The stability of the sucrose sensitivity phenotype conferred by this gene was assessed through the screening of *E. coli* DH5α pK19*mobsacB* transformants, of which approximately 40% displayed a stable phenotype. Upon the isolation of pK19*mobsacB* DNA from these sucrose sensitive bacteria, however, and, after the incorporation of the *ΔargR* allele present in pARGRKO, the resulting pARGR-MS suicide construct was found to reliably confer a sucrose sensitive phenotype onto both *E. coli* DH5α and *C. glutamicum* ATCC 13032 host cells.

When approximately 10 µg of pARGR-MS were electroporated into *C. glutamicum* ATCC 13032 cells, a similar number of chromosomal integrants were observed compared to those attained for both *argH* and *argR* using the p2NIL/pGOAL19 system. Approximately ten Kan^R transformants from a single electroporation event were screened for sucrose sensitivity and seven were typically found to display drastically reduced colony numbers when plated on sucrose-containing medium (an approximate 100-fold difference). Furthermore, when the sucrose resistant colonies derived from each integrant were

screened for Kan^R, approximately 80% proved to be sensitive, thereby indicating that loss of the plasmid marker genes had occurred. Indeed, when these colonies were PCR-screened using the *argR* gene primers, they were, as expected, found to consist of roughly equal populations of cells either in possession of the WT *argR* gene or only the *ΔargR* allele. As shown in Figure 2.10A, one such colony (designated DCO-D3), reliably displayed a 204 bp PCR product fragment indicative of the presence of the *ΔargR* allele. As is also shown in this Figure, its parent SCO colony (designated SCO-D3) displayed, in addition to this 204 bp fragment, a 522 bp PCR product fragment and together these are indicative of the presence of both *argR* alleles.

Southern blot analysis was used to subsequently confirm the identity of these two mutants and, as expected, the DIG-labelled *argR* gene probe hybridised to two fragments of *Eco*RV-digested SCO-D3 genomic DNA (Figure 2.10B and C). In this case, the smaller approximate 3 kb DNA fragment corresponds to the region between the first two *Eco*RV sites (situated at 505 bp and 3257 bp respectively and 2752 bp long) and was also detected in WT genomic DNA. In contrast, the second, larger fragment detected in SCO-D3 genomic DNA is partly due to the integration of the 7455 bp pARGR-MS plasmid, which does not contain any *Eco*RV sites. Additionally, the final *Eco*RV site shown in Figure 2.10B is located 1544 bp downstream of the previous *Eco*RV site (situated at 3257 bp) and therefore, when combined with the insertion of pARGR-MS, results in the formation of an 8999 bp fragment. This also suggests that the insertion of pARGR-MS in SCOR-D3 has occurred in an orientation similar to that previously outlined in part (ii) of Figure 2.7. Lastly, within the *ΔargR* allele, the *Eco*RV site is not reconstituted following the removal of the internal *Eco*RV-*Bsr*BI fragment and we would thus expect the *argR* gene probe to hybridise to only a single fragment of DCOR-D3 genomic DNA. As expected, an single fragment of approximately 4 kb was observed and, after taking into account the occurrence of the 318 bp *argR* deletion, was found to correspond approximately in size to the region located between the first and last *Eco*RV sites shown in Figure 2.10B (3978 bp).

2.4.4 Confirmation of an *argR* insertional mutation in *C. glutamicum* 9A

C. glutamicum 9A is an *argR* insertion mutant previously created by N. Smazanowiz using pK19*mobsacB* in possession of an internal fragment of the *C. glutamicum* ATCC 13032 *argR* gene (pARGR-INT) (Schäfer *et al.*, 1994). In order to confirm the position of the

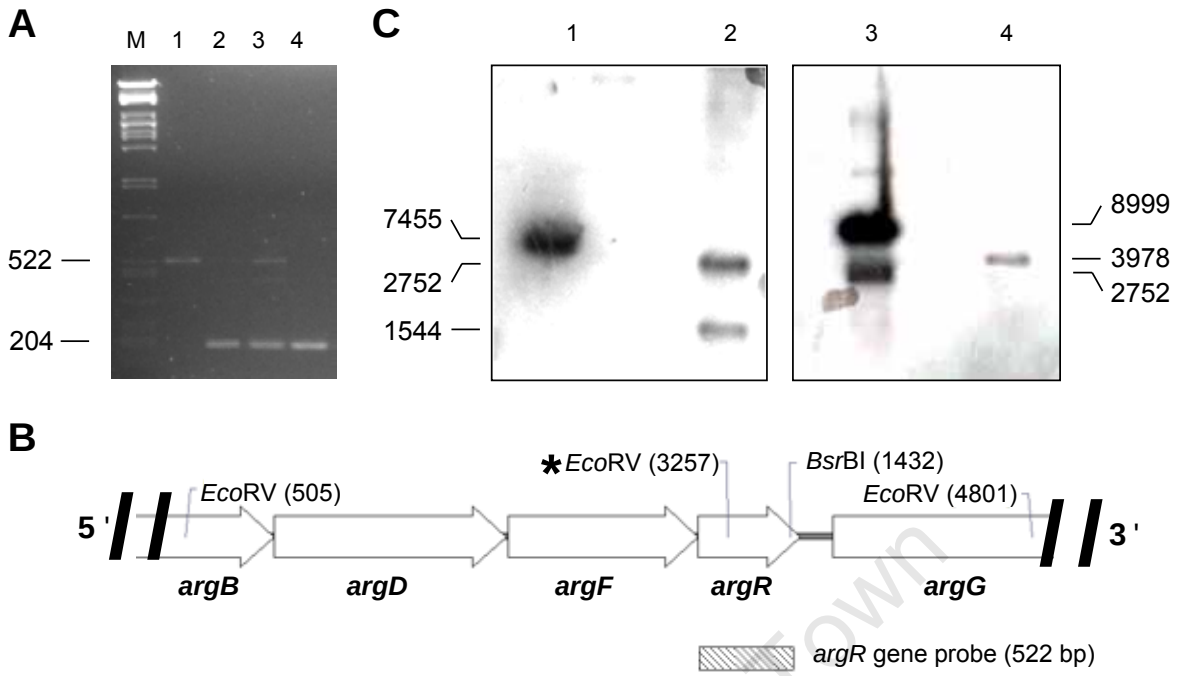


Figure 2.10 Confirmation of pARGR-MS-directed mutagenesis of the *C. glutamicum* ATCC 13032 *argR* gene in SCOR-D3 and DCO-D3. Fragment sizes are indicated in bp. **(A)** Results of a PCR performed with the *argR* junction primers showing the construction of the *ΔargR* fragment. Lane M: marker DNA. Each remaining lane contains PCR products amplified from different DNA templates. Lane 1: wildtype genomic DNA; lane 2: pARGR-MS; lanes 3: SCOR-D3; lane 4: DCOR-D3. The fragment sizes shown here are the expected sizes calculated from the sequence. **(B)** Partial restriction map of the *C. glutamicum* ATCC 13032 *argCJBDR* operon showing the design of the *argR* gene probe for use in Southern blotting. The hatched box corresponds to the position of the 522 bp DIG-labelled probe. The *Eco*RV site indicated by the asterisk is not present in *ΔargR*. Restriction site positions are indicated relative to the start of the *argCJBDR* operon. **(C)** Southern blot analysis of SCOR-D3 and DCO-D3. The autoradiograph shows *Eco*RV-digested DNA in each lane hybridised with the DIG-labelled 522 bp *argR* gene probe. Lane 1: pARGR-MS DNA; lane 2: wildtype genomic DNA; lane 3: SCOR-D3 genomic DNA; lane 4: DCOR-D3 genomic DNA.

pARGR-INT integration site in the *C. glutamicum* 9A chromosome, a similar approach was adopted to that used for SCOH and SCOR (Figure 2.11). By using different combinations of primers, which either bind to genomic DNA that flanks the putative site of integration, or which putatively bind exclusively to integrated vector DNA, the position of the pK19*mobsacB*::*argR* integration event may be confirmed.

PCR-mediated amplification of *C. glutamicum* 9A DNA using the *argF-argR* F and pK19 R primer pair and the *argR-argG* R and pK19 R primer pair produced fragments of 1.2 kb and 1.0 kb respectively (Figure 2.11A). Upon sequencing, these DNA fragments were confirmed to possess homology to both pK19*mobsacB* and the chromosomal regions that flank the *C. glutamicum* ATCC 13032 *argR*, thereby indicating the occurrence of a successful integration event (Figure 2.11B). Furthermore, PCR using the *argR* gene primers, which anneal to regions flanking the *argR* gene, failed to produce a product fragment of 522 bp from *C. glutamicum* 9A genomic DNA, whereas the same product was easily obtained from *C. glutamicum* ATCC 13032 genomic DNA (Figure 2.11C). *C. glutamicum* 9A therefore appears to contain a disrupted *argR* gene and this strain was used for future experimentation in conjunction with the other mutant strains created as part of this work (Chapter 3).

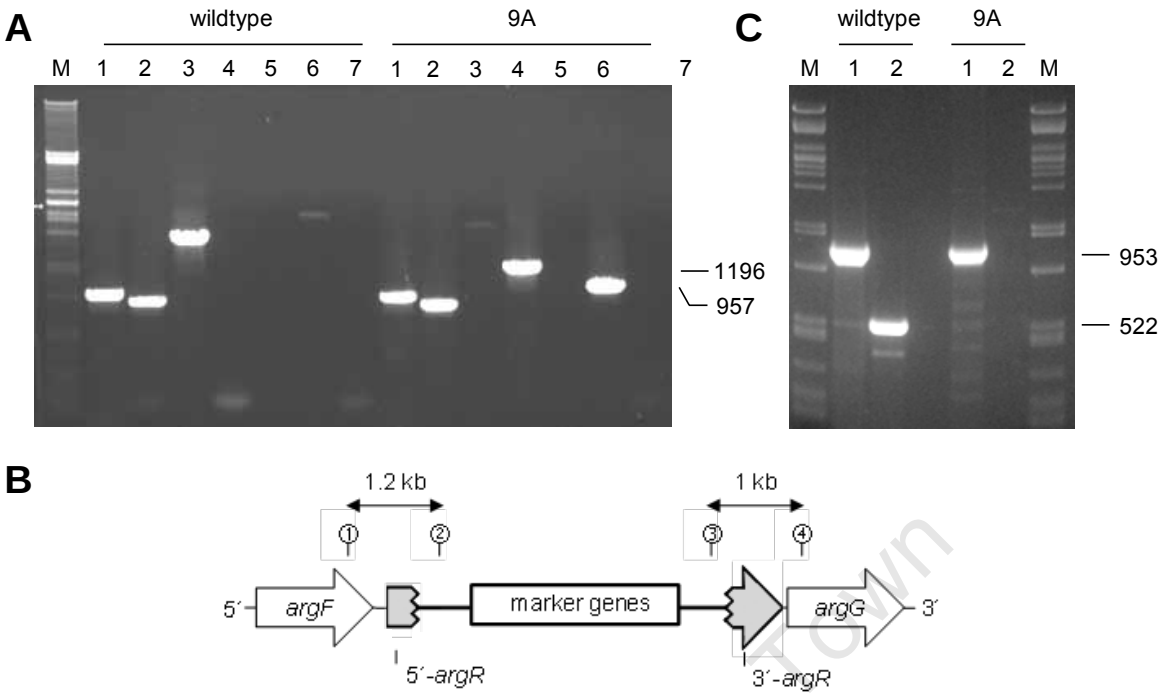


Figure 2.11 Integration of pARGR-INT into the *argR* gene of *C. glutamicum* 9A. Lanes M: marker DNA. Fragment sizes are indicated in bp. **(A)** PCR amplification was performed using different combinations of the pK19 primers, which bind to pARGR-INT vector DNA, and the *argF-argR* F and *argR-argG* R primers, which bind to the *C. glutamicum argF* and *argG* genes respectively. The remaining lanes contain PCR products amplified from either *C. glutamicum* ATCC 13032 genomic DNA or *C. glutamicum* 9A genomic DNA. Primer combinations for each lane: lane 1, *argF-argR* F and R; lane 2, *argR-argG* F and R; lane 3, *argF-argR* F and *argR-argG* R; lane 4, *argF-argR* F and pK19 R; lane 5, *argF-argR* F and pK19 F; lane 6, *argR-argG* R and pK19 F; lane 7, *argR-argG* R and pK19 R. **(B)** To confirm the integration of pARGR-INT into the *C. glutamicum* ATCC 13032 *argR*, PCR with primers that anneal to either vector or genomic DNA were used. The chromosomal position of the *argF-argR* F and pK19 R primers are indicated by • and , , whereas the position of the pK19 F and *argR-argG* R primers are indicated by ƒ and „. The black arrows represent the expected approximate PCR product sizes for each primer pair. The thick black line represents vector DNA and the thin black line represents *C. glutamicum* genomic DNA. The interrupted gene is shaded blue. **(C)** Absence of an intact *argR* gene in *C. glutamicum* 9A. PCR amplification using either the *CAF21404-argC* primers (lanes 1) or the *argR* gene primers (lanes 2) was conducted using either *C. glutamicum* ATCC 13032 genomic DNA or *C. glutamicum* 9A genomic DNA as template.

2.5 Discussion

This chapter reports the generation and confirmation of various *C. glutamicum argH* and *argR* mutant strains created using two different types of suicide plasmids. The p2NIL/pGOAL19 system was first used for the generation of *C. glutamicum* insertional mutants and was initially favoured because, relative to conventional *C. glutamicum* mutant delivery vectors like pK19*mobsacB*, it possesses additional marker genes that should permit for a reduced number of colonies to be screened for the desired mutation. SCO *argH* and *argR* mutant strains were successfully created using this system and, in conjunction with a SCO *argR* mutant created using the pK19*mobsacB* plasmid, the ability of these strains to undergo a second recombination event resulting in unmarked DCO mutations was assessed.

2.5.1 Generation of single cross-over mutants using the p2NIL/pGOAL19 system

The p2NIL/pGOAL19 system employs a two plasmid strategy that allows for the easy, flexible construction of suicide delivery vectors. For example, a variety of different pNIL and pGOAL vectors exist, the latter each harbouring several different selectable markers within the *Pac*I cassette (Parish and Stoker, 2000). This allows for the introduction of different combinations of selectable markers into the *Pac*I insertion site common to the pNIL vectors, each of which is in possession of a different MCS (Gordhan and Parish, 2001). As a two-step strategy is adopted, in which a SCO mutant is first generated and subsequently used to produce DCO colonies, only a single SCO colony is required from the electroporation event in order to produce multiple DCO mutants (Parish and Stoker, 2000). Factors that typically hamper transformation efficiency, such as large vector size and the use of microbes intrinsically difficult to electroporate, are therefore limited (Parish and Stoker, 2000). Importantly, the pGOAL19 *Pac*I cassette is in possession of several counter-selectable marker genes, which permit for the identification of putative DCO colonies that have lost this integration cassette. This is particularly useful when the desired DCO mutant does not have an easily screenable phenotype, such as in the case of *ΔargR*.

The use of this novel system in *C. glutamicum* ATCC 13032 was successful for the generation of SCO mutants. After the electroporation of either pARGHKO or pARGRKO into *C. glutamicum* ATCC 13032 cells, only SCO mutant colonies that displayed both forms

of the appropriate target gene were obtained. Spontaneous antibiotic resistance does not therefore appear to be a concern for this system in *C. glutamicum* ATCC 13032, although in certain Gram-positive bacteria, such as *M. tuberculosis*, the spontaneous occurrence of Kan^R is problematic, the inclusion of the Hyg^R marker in conjunction with *lacZ* on the *Pac*I cassette would likely abolish this problem (Parish *et al.*, 1999). Interestingly, SCO colonies obtained after the integration of each construct only appeared blue after a prolonged period of incubation over four days at 30 °C. *C. glutamicum* has been noted to be relatively impermeable to IPTG and this is likely further exacerbated by the *lacZ* gene lying under the control of a mycobacterial promoter (Parish and Stoker, 2000; Pátek *et al.*, 2003). Obtaining an efficient *C. glutamicum* electroporation procedure proved critical for the appearance of SCO colonies as, although only a single SCO is required for DCO generation, a successful integration event is dependent on both the efficiency of electrotransformation and the frequency of HR between the suicide vector and the chromosome.

The mycolic acid-containing outer lipid bilayer of *C. glutamicum* represents a significant permeability barrier against the passage of DNA molecules into the cytoplasm and, in order to obtain SCO colonies, several improvements over conventional electrotransformation protocols had to be implemented (Section 1.6.3; Appendix D). Electrotransformation efficiency was assessed using the pEKEX2 vector and, following the implementation of the improvements described in Appendix D, was found to range between 1000 and 5000 CFU μg^{-1} vector DNA. This is a 100-fold improvement over what has previously been achieved in our hands using the pEKEX2 vector with *C. glutamicum* ATCC 13032 (Halsey, 2003). Interestingly, this improved value is still approximately 500-fold less than what has previously been reported for electroporation under analogous conditions (Liebl *et al.*, 1989; van der Rest *et al.*, 1999).

The electroporation of 10 µg pARGHKO or pARGRKO into *C. glutamicum* ATCC 13032 cells typically resulted in the generation of four SCOH colonies and ten SCOR colonies. Assuming electroporation has occurred with the aforementioned optimal efficiency of 5000 CFU μg^{-1}, this is suggestive of an average frequency of HR-mediated integration of the suicide vector into the *C. glutamicum* ATCC 13032 chromosome of approximately 1.5×10^{-4} for the *argH* fragment and 5×10^{-3} for the *argR* fragment. As the suicide delivery vectors used in this study are substantially larger than pEKEX2 (approximately 14 kb relative to

8.2 kb), it is likely that, as electrotransformation efficiency is partly limited by vector size, the frequency of HR between portions of the L-arginine biosynthetic operons and the suicide vectors is, in reality, significantly higher than the values reported here. Furthermore, in the closely-related bacterium *M. tuberculosis*, it has been observed that genes that fulfil critical biosynthetic functions often prove refractory to insertional inactivation, thereby indicating that the frequency of allelic inactivation is gene dependent (Gordhan and Parish, 2001). This may explain why pARGRKO integrates more readily into the chromosomal *argR* fragment than pARGHKO into the corresponding *argH* fragment. Furthermore, the frequency of HR using this system in *M. tuberculosis* has been reported to be as low as 10^{-5} and the values reported here for its use in *C. glutamicum* ATCC 13032 fall well within this lower limit (Gordhan and Parish, 2001).

Both PCR and Southern blot analysis revealed SCOH and SCOR to harbour both WT and mutant alleles of their respective target genes and was therefore suggestive of a successful integration event. Indeed, an additional PCR-based analysis step, performed using primers designed to anneal to either vector DNA integrated into the *C. glutamicum* ATCC 13032 chromosome or genomic DNA flanking the putative site of integration, produced a large quantity of non-specific PCR products, in addition to further evidence of a successful integration event (Figure 2.6B; Figure 2.8B). Although the relatively high G+C molar content of the *C. glutamicum* ATCC 13032 genome is likely partly responsible for this non-specific primer binding, a single PCR product fragment, with significant homology to both the mutant target gene and the region downstream of the putative site of integration, was successfully obtained for each strain. Ideally, we would also expect a second PCR product fragment, obtained with primers that bind downstream of the putative site of integration, which would be in possession of significant homology to both the WT target gene and the downstream flanking region (indicated by (i) in Figure 2.5 and Figure 2.7 for SCOH and SCOR respectively). Such a product fragment, however, could not be obtained from either strain, despite significant alterations to the PCR conditions.

It is difficult to surmise as to why this has occurred, although the high G+C content of this bacterium's genome could give rise to various nucleic acid secondary structural elements that would be capable of impeding the progression of DNA polymerase along the template strand, thereby perhaps accounting for the lack of a PCR product obtained from the region

downstream of the putative site of integration in each strain (Kalinowski, 2005). This seems unlikely, however, as, although hairpin-loop terminator structures may exist within each target region, they do not appear to hamper their PCR-based amplification because, as part of later semi-quantitative reverse transcriptase PCR studies, the intergenic DNA fragments located within each upstream region were easily amplified, albeit as part of much smaller product fragments (Section 3.4.3). Alternatively, the occurrence of repeated inverted insertions of the suicide delivery vector into the *C. glutamicum* ATCC 13032 chromosome could render the primer combinations used for amplification of the appropriate upstream target region incompatible, thereby resulting in none of the desired PCR product being formed. However, although the integration of some suicide plasmid systems can result in the insertion of adjacent repeats due to the rolling-circle mechanism of plasmid replication, it would be extremely unusual for such repeats to occur in opposite orientation to each other (Russell and Klaenhammer, 2001). Moreover, it should be noted that this PCR-based strategy used to confirm the position of plasmid integration does not exclude other integration sites and, as noted by the different band intensities for each target allele observed in Figure 2.6A and Figure 2.8A, it is indeed possible for each strain to contain multiple, non-specific suicide plasmid insertions. For example, as evident from Figure 2.8A, *argR* junction PCR product fragments obtained from SCOR genomic DNA representative of each *argR* allele, appear to be present in relatively equal quantities, thereby indicating that only a single pARGRKO integration event has occurred. In contrast, however, as shown in Figure 2.6A, the relative quantities of *argH* junction PCR product fragments derived from SCOH genomic DNA indicate that the number of *argH* mutant alleles present per chromosome exceeds those of the *argH* WT allele. Such a phenomenon would likely be due to multiple pARGHKO insertions and a further Southern blot analysis performed with a probe in possession of homology to the target gene's flanking regions would be able to provide more information regarding the number and location of plasmid insertions. Lastly, a plasmid-rescue based strategy could also be adopted in order to generate sequence information regarding the regions that flank each suicide vector insertion site in SCOH and SCOR (Schäfer *et al.*, 1997).

2.5.2 Attempted double cross-over mutant generation from SCOH and SCOR

The combination of *lacZ* and *sacB* within the *Pac*I marker gene cassette found in pARGHKO and pARGRKO permits fewer colonies derived from SCO mutants to be screened for the occurrence of a second cross-over event. Spontaneous resistance to the lethal action of levansucrase, encoded by *sacB*, in the presence of sucrose is a relatively common occurrence and can become problematic when this gene is used by itself as a counter-selectable marker (Parish *et al.*, 1999). For example, when screening for potential *M. tuberculosis* DCO mutants in the presence of sucrose, between 10 to 90% of colonies that displayed strong growth, thereby appearing to be true DCO colonies, were, in fact, persistent SCO mutants in possession of a defective chromosomal *sacB* gene (Parish and Stoker, 2000). Therefore, in order to allow for a more efficient screening of DCO colonies, the *lacZ* gene was incorporated into the *Pac*I cassette. This permitted spontaneous sucrose resistance colonies, which previously appeared as DCO colonies, to be identified by the persistence of the insoluble blue colour complex arising from LacZ activity in the presence of IPTG and X-gal (Parish and Stoker, 2000). In contrast, the *C. glutamicum* ATCC 13032 suicide vector pK19*mobsacB*, which harbours *sacB* in addition to the Kan^R gene, lacks the counter-selectable *lacZ* and, as a result, false positive DCO colonies are more likely to arise and an additional Kan^S screening step is therefore usually required (Eggeling and Reyes, 2005).

Following the plating of potential DCO cells onto medium supplemented with IPTG, Nx, sucrose and X-gal, a mix of blue and white sucrose resistant colonies was observed. Spontaneous sucrose resistant colonies are blue, since they still carry the *lacZ* gene, whereas DCO colonies are white, having lost the *lacZ* gene. It was difficult to distinguish between colonies that are spontaneous sucrose resistance mutants and persistent SCO mutants as, although the *sacB* gene should be lethal to the latter strain in the presence of sucrose, it proved to result in only a weaker growth phenotype, rather than the total inhibition of growth. Due to this, any colonies screened for levansucrase activity could not be allowed to grow for extended periods of time, as beyond a certain point, any *sacB*-containing colonies appeared morphologically identical to WT colonies. Unfortunately, similar to what was observed with the SCO mutant colonies, the blue colour complex resulting from LacZ activity was only observable after a prolonged incubation period.

Furthermore, *lacZ* expression off the mycobacterial Ag85 promoter appeared highly variable, as some SCO and potential DCO colonies would either be intensely coloured, or appear only pale blue after long periods of incubation. This rendered potential DCO mutant colonies difficult to screen and, for example, if a colony appeared white and sucrose resistant, the appropriate junction PCR frequently revealed the presence of both alleles of the target gene. In such a situation, the *sacB* PCR was subsequently used to confirm both the presence of the marker gene cassette and the occurrence of spontaneous sucrose resistance.

Several measures, including the allowance of increased time for the DCO event to occur (up to approximately fifty generations of growth), in addition to the incorporation of several media-based screening methods, failed to enrich the appearance of DCO mutants. This is highly unusual, as the previous use of this system in *M. tuberculosis* has proved the isolation of DCO mutants to be relatively easy and not to require any additional screening steps (Gordhan *et al.*, 2002*;* Gordhan and Parish, 2001; Parish *et al.*, 1999; Parish and Stoker, 2000). For example, an unmarked *M. tuberculosis argF* deletion mutant, created using the pNIL and pGOAL vectors, was isolated by Gordhan *et al.* (2002) after screening 184 sucrose resistant colonies, whereas Gordhan and Parish (2001) recommended screening only 40 sucrose resistant colonies in order to isolate several DCO mutants. In this study, over 1000 potential *argH* mutants and 500 potential *argR* mutants were screened, none of which, in the case of the *argH* mutant, were auxotrophic for L-arginine, or, in the case of both *argH* and *argR* mutants, lacked a copy of the appropriate WT allele. Additionally, any colonies that initially exhibited a mutant allele were found to possess an intact *sacB* gene, indicative of the persistence of the SCO insertion. In order to ensure HR occurred with an equal frequency on either side of the site of the SCO integration to give both WT and mutant alleles following the DCO event, ~1 kb flanking fragments were incorporated adjacent to the desired deletion on the suicide plasmid (Gordhan and Parish, 2001). This, however, did not enhance the emergence of any mutant DCO colonies and, even after SCOH and SCOR were permitted to grow in the absence of antibiotic selection pressure for prolonged periods, only SCO mutants or WT revertant colonies were detected.

If *argH* were essential for cellular function, it may explain why no DCO colonies in possession of this mutant allele were obtained. Such a scenario appears unlikely, however,

as, in most other microorganisms, ArgH fulfils a relatively routine biosynthetic function. For example, *E. coli* and *Saccharomyces cerevisiae* mutants deficient in ASL activity, which exhibit normal growth in L-arginine supplemented media, have been created successfully (Cunin *et al.*, 1986). Interestingly, however, *argH* SCO mutants, still in possession of an intact *argH* gene, have been created in the archaeon *Methanococcus maripaludis* and, similar to what was observed in this work, a DCO mutant lacking an intact copy of *argH* could not be isolated (Cohen-Kupiec *et al.*, 1999). The authors subsequently concluded ASL to putatively fulfil an essential role in this bacterium, and proposed it to be responsible for the maintenance of intracellular L-glutamate and ammonia content through participation in the urea cycle. As part of this cycle, L-arginine is catabolised into urea and L-ornithine by an enzyme in possession of arginase activity. As most ArgH enzymes are capable of catalysing the reversible breakdown of argininosuccinic acid into L-arginine and fumarate, Cohen-Kupiec *et al.* (1999) proposed this enzyme to critically supply L-arginine into the urea cycle of *M. maripaludis*. Moreover, recent genetic evidence in *C. glutamicum* ATCC 13032 indicates that this organism may indeed encode the enzymes necessary for the detoxification of ammonia through the urea cycle and this thus leaves open the possibility that the *C. glutamicum* ATCC 13032 ASL might fulfil a similar function to that proposed to occur in *M. maripaludis* (Cohen-Kupiec *et al.*, 1999; Kalinowski *et al.*, 2003).

Unlike *argH*, there is little evidence to suggest *argR* fulfils a critical cellular function. For example, *argR* deletion and integration mutants have been successfully created in a variety of bacteria, including *E. coli, Lactococcus lactis* and *Pseudomonas aeruginosa*, with no effect on cell viability (Larsen *et al.*, 2004; Maas, 1994; Park *et al.* 1997). Furthermore, as shown later in this work, *C. glutamicum* 9A, a previously created strain which harbours a disrupted *argR* gene, clearly remains viable (Section 3.4.1). This suggests that the *C. glutamicum* ATCC 13032 *argR* gene is, in all likelihood, not essential for cell survival and, at this stage of the work, it was thus concluded that our inability to isolate an *argR* DCO mutant from SCOR was probably a direct result of the occurrence of a spurious pARGRKO integration event capable of interfering with the second HR step required for the removal of the WT *argR* gene. It is fully possible that this also applies to the integration of pARGHKO in SCOH and it is important to note that, especially as several SCO colonies of each variety were screened unsuccessfully for the ability to generate DCOs, the occurrence of such an event appears to be widespread.

2.5.3 Generation of an *argR* double cross-over mutant using pK19*mobsacB*

SCO mutants of *argR* were created using the pK19*mobsacB* plasmid in a manner similar to that adopted for the creation of SCOH and SCOR with the p2NIL/pGOAL19 vector system. Interestingly, relative to the SCO mutants created using the p2NIL/pGOAL19 system, SCO colonies generated using the pK19*mobsacB*-based pARGR-MS plasmid generally displayed a much stronger sucrose sensitive phenotype (a several hundred fold difference in terms of colony numbers as opposed to merely a slight growth disadvantage). It thus appears that the addition of the sucrose sensitivity screening step of *E. coli* DH5α cells in possession of pK19*mobsacB* greatly aided in the selection of vector DNA that encodes a prominent sucrose sensitive phenotype. Indeed, *E. coli* DH5α transformants carrying pARGR-MS certainly displayed a much stronger sensitivity to sucrose compared to those in possession of pARGHKO or pARGRKO and, as this phenotype was clearly transferred to each of the SCO integrants, it would appear to be critical in the determination of the frequency of appearance of DCO colonies subsequently isolated from each integrant. The occurrence of such a DCO event therefore appears to be exceptionally rare as, without the presence of a strong selection marker, such as sucrose sensitivity, to promote the detection of DCO cells, it is, as detailed in Section 2.5.2, especially difficult to detect cells in which this event has occurred. In retrospect, the presence of a strong and stable sucrose sensitive phenotype should have therefore first been confirmed in cells in possession of the vector DNA used for the subsequent construction of pARGHKO and pARGRKO via the p2NIL/pGOAL19 vector system.

2.6 Concluding remarks

The p2NIL/pGOAL19 vector system has been successfully adapted for the generation of *C. glutamicum* insertion mutants, but should be further optimised before it may be readily used for the generation of unmarked DCO mutants. For example, the placement of the *lacZ* gene under the control of a strong, corynebacterial promoter, such as P-45, would eliminate much of the difficulty associated with weak colour development resulting from variable *lacZ* expression (Pátek *et al.*, 2003). Additionally, the placement of the *sacB* gene under such a promoter would enhance its efficacy as a counter-selectable marker in *C. glutamicum* strains because it should limit the emergence of slow-growing, *sacB*-containing SCO colonies. Most importantly, however, this work has highlighted the importance of ensuring that a strong sucrose sensitive phenotype in present mutagenesis systems that rely on the activity of the counter-selectable *sacB* gene prior to the construction of a suicide vector.

In addition to the SCO integration mutants in possession of mutant *argR* and *argH* genes created using the novel p2NIL/pGOAL19 gene inactivation system, we also successfully constructed another *argR* integration mutant of *C. glutamicum* ATCC 13032 using the pK19*mobsacB* plasmid and an unmarked *argR* mutant was subsequently derived from this strain. The effects of each integration event, as well as that of the *argR* deletion, on the growth kinetics and transcription of the L-arginine biosynthetic genes, will next be investigated in each of these strains, in addition to their ability to produce L-arginine and L-citrulline. These characteristics will also be studied in *C. glutamicum* 9A, which possesses pK19*mobsacb* inserted into the *argR* gene.

Chapter Three

Characterisation of *Corynebacterium glutamicum* mutant strains

Contents

Chapter 3

3.1 Summary

The study of *Corynebacterium glutamicum* strains in possession of altered L-arginine biosynthetic pathways can enhance our understanding of the biosynthesis of this amino acid in this bacterium. In this chapter, through the characterisation of different *argR* mutant strains derived from *C. glutamicum* ATCC 13032, we first show that various mutations introduced into these bacteria do not influence their maximum growth rates and, in the absence of selective pressure, are stably integrated into the host chromosome. In contrast, an insertional mutation introduced into L-arginine biosynthetic gene cluster of strain SCOH, was found to be unstable in the absence of selective pressure and also had the effects of rendering this strain auxotrophic for L-arginine biosynthesis and severely restricting its growth rate. An analysis of both the transcriptional organisation and regulation of the L-arginine *argCJBDFRGH* biosynthetic gene cluster in an *argR* deletion mutant indicated that these genes are transcribed together as a single transcript and, in contrast to the wildtype bacterium, their transcription regulation is derepressed in the presence of L-arginine. Lastly, mutagenesis of the *argR* gene was found to result in a three- to five-fold increase in L-arginine biosynthesis and an eight-fold increase in the biosynthesis of L-citrulline (where levels as high as 23.18 and 34.14 µmol [mg dry cell mass]$^{-1}$ were obtained respectively for each compound). The L-arginine production capacity of each of these strains was also found to be correlated with their ability to grow in the presence of the toxic L-arginine analogue L-canavanine, where the *argR* mutant strains 9A and DCOR-D3 grew strongly in media supplemented with 100 µg ml^{-1} of L-canavanine and wildtype cells failed to grow in media supplemented with as little as 25 µg ml^{-1} of this compound.

Chapter 3

3.2 Introduction

The adoption of targeted, recombinant DNA techniques is a relatively common approach used to manipulate bacteria intended for industrial use. By providing a means whereby these organisms can be engineered to produce abnormal amounts of commercially valuable compounds, these tools often permit for an improved understanding of their biosynthesis. Curiously, although this approach has been used extensively to engineer the production of various fine chemicals, including certain amino acids, the use of these techniques to manipulate L-arginine biosynthesis in otherwise commonly-used industrial bacteria, such *C. glutamicum*, appears to remain relatively unexplored (Hermann, 2003; Ikeda, 2003). Indeed, one of the most competent industrial *C. glutamicum* L-arginine overproducer strains available was created via random mutagenesis and now, as the study of the biosynthesis of other amino acids, such as L-lysine, rapidly advances into a system-wide, post-genomics era, our understanding of L-arginine production in this bacterium remains prohibitively limited (Utagawa, 2004). Until very recently, there has been a consequent lack of rationally mutated *C. glutamicum* L-arginine overproducer strains available for further study and our understanding of the biosynthesis of this amino acid in this bacterium is largely understood by analogy with other, better studied organisms.

As part of the work conducted in Chapter 2, preliminary steps towards the construction of an industrially-suited L-arginine overproducer strain were undertaken. Various *C. glutamicum* strains, in possession of targeted mutations within their L-arginine biosynthetic genes, were constructed using different molecular tools. Now, in order to gain a better understanding of the regulation and biosynthesis of this amino acid in *C. glutamicum* ATCC 13032, the effect of these mutations on the ability of each strain to produce both L-arginine and L-citrulline was examined. Further aspects of the physiology of these strains were also assessed, such as their growth kinetics, the transcriptional regulation of their genes and, where relevant, the stability of recombinant DNA inserted into their host chromosomes.

3.3 Materials and Methods

3.3.1 Bacterial strains and media

The *Escherichia coli* DH5α cells and *C. glutamicum* strains used in this study, which are described in Appendix A, were routinely grown under the conditions described in Section 2.3.2. Antibiotics and L-arginine were also used as described in Section 2.3.2.

3.3.2 Measurement of broth culture growth

For the measurement of *C. glutamicum* broth culture growth, a single colony of cells, grown on 2 × Yeast-Tryptone (2 × YT) solid medium, was used to inoculate 5 ml of either 2 × YT broth or CGXII minimal medium (MM) broth. Each of these seed cultures was supplemented with Nx and, if required, Kan and L-arginine. After culture growth to mid-exponential phase ($OD_{600} \approx 7$), each 5 ml culture was transferred to 45 ml of either 2 × YT broth or CGXII MM broth in a 500 ml fluted flask. Each 45 ml culture was supplemented with Nx and, if required, L-arginine. Broth culture growth was subsequently monitored by OD_{600} measurement using a Novaspec-II spectrophotometer (Pharmacia Biotech). The generation time and maximum growth rate of each *C. glutamicum* strain was calculated according to the method of Willey *et al.* (2007).

3.3.3 DNA manipulation and amplification

DNA fragments were prepared and manipulated as previously described (Section 2.3.3) and amplified using the polymerase chain reaction (PCR) (Section 2.3.4). The thermal cycling conditions and oligonucleotide primers used for each amplification reaction are detailed in Appendix C.

3.3.4 Assessment of suicide vector integration stability

For the measurement of suicide vector stability within the genomes of *C. glutamicum* 9A, SCOH and SCOR cells, each strain was grown in the absence of both Kan and hygromycin B selective pressure. A single colony of cells grown on 2 × YT solid medium was initially used to inoculate 50 ml of 2 × YT broth supplemented with Nx. After growth to saturation ($OD_{600} \approx 12$), 1% of each culture volume was transferred to 50 ml of 2 × YT broth containing Nx and permitted to grow to stationary phase ($OD_{600} \approx 12$). Each passage was repeated for

ca. thirty generations of exponential phase growth and the emergence of Kan^S cells was monitored at different time points by replica plating (Russell and Klaenhammer, 2001). Plating was performed in triplicate for each strain for three independent repeats. The stability of vector insertion in SCOH and SCOR cells in the absence of selective pressure was also assessed on a molecular level, with either the *argH* or *argR* junction oligonucleotides being used as part of a PCR to monitor the presence of the relevant mutant allele in each culture. The *argH* and *argR* junction primer pairs, together with their appropriate PCR conditions, are described in Appendix C. Each PCR mixture contained 1 µl of broth culture from each time point and the remaining components were added as described in Section 2.3.4.

3.3.5 RNA isolation and analysis

The total RNA from *C. glutamicum* cells grown in CGXII MM broth was extracted using a RNeasy Mini Kit (Qiagen). At the appropriate growth phase, a 1.5 ml aliquot of broth culture was immediately added to two volumes of RNAProtect Bacteria Reagent (Qiagen). After 5 min of stabilisation at room-temperature, the treated cells were centrifuged (15 000 × *g*, 10 min), resuspended in 100 µl of TE buffer (10 mM TrisCl, 1 mM EDTA, pH 8.0) in the presence of lysozyme (Sigma-Aldrich) (15 mg ml^{-1}) and proteinase K (Fermentas) (1 mg ml^{-1}) and incubated for 10 min at room temperature with intermittent mixing. An aliquot of Buffer RLT (Qiagen) (350 µl), containing 1% (v/v) ß-mercaptoethanol (Sigma-Aldrich), was added to the cell suspension and the entire mixture was dispensed into a 2 ml screw-top microcentrifuge tube (BioSpec Products) together with 250 µl of sterile zirconia-silica beads (0.1 mm diameter; BioSpec Products). The cells were now mechanically disrupted using a Mini-Beadbeater (BioSpec Products) with three consecutive 45 s cycles of beating at 4 800 rpm. After the collection of insoluble cell debris and glass beads by centrifugation (14 000 × *g*, 10 s, 4 °C), the RNA present in the supernatant was extracted according to the protocol supplied by the manufacturer. In order to completely remove contaminating chromosomal DNA, a 16 h DNAse digestion was performed on each RNA isolate at 37 °C using RNAse-free DNAse I (Roche). The RNeasy Mini Kit was subsequently used to again purify each DNAse-treated RNA sample and, in order to confirm that the complete removal of genomic DNA had occurred, a negative control PCR using the *CAF21404-argC* primer pair and 5 µg of DNAse-treated RNA, was performed (Section 2.3.4; Appendix C). The

quality and integrity of each RNA sample was routinely assessed by spectrophotometry and MOPS-buffered electrophoresis through formaldehyde-agarose gels with ethidium bromide staining (Sambrook *et al.*, 2001). RNA samples were stored in DEPC-treated dH_2O at -80 °C.

3.3.6 Semi-quantitative reverse transcriptase PCR

Reverse transcriptase (RT) reactions were performed using the FirstStrand cDNA Synthesis Kit (Fermentas) on total RNA isolated from mid-exponential phase ($OD_{600} \approx 7$) *C. glutamicum* cells. Briefly, 1 µl of a mixture of random hexameric oligonucleotides (0.2 µg μl^{-1}) was incubated with 1 µg of RNA in DEPC-treated dH_2O at 70 °C for 5 min. The solution was hereafter cooled on ice and brought to a volume of 19 µl by the addition of 4 µl of 5 × RT buffer (250 mM Tris-HCl [pH 8.3], 250 mM KCl, 20 mM $MgCl_2$, 50 mM DTT), 2 µl of RiboLock ribonuclease inhibitor (10 U μl^{-1}) and 2 µl of a 10 mM dNTP mix. After a second incubation period (25 °C, 5 min), 40 U of recombinant murine leukemia virus RT was added to the mixture before incubation at 25 °C for 10 min. The RT reaction was now allowed to proceed at 37 °C for 1 h and was finally stopped by heat inactivation at 70 °C for 10 min. The reaction contents were stored at -20 °C and cDNA fragments were subsequently amplified via the PCR in 50 µl reaction volumes. Each PCR was constituted as described previously in Section 2.3.4 and incorporated 3 µl of the heat-inactivated RT reaction as template DNA. The oligonucleotide primers and thermal cycling conditions used for each RT-PCR are listed in Appendix C.

3.3.7 Sample preparation for L-arginine and L-citrulline quantitation

A modified version of the protocol of Rajagopal *et al.* (1998) was adopted in order to allow for the accumulation of measureable levels of L-arginine and L-citrulline by different *C. glutamicum* cells. Initially, a single colony of cells from each strain grown on 2 × YT solid medium was used to inoculate 50 ml of 2 × YT broth supplemented with Nx and Kan. After growth to mid-exponential phase ($OD_{600} \approx 7$), each culture was diluted 10-fold with 2 × YT broth supplemented with Nx and Kan, grown to an OD_{600} of 1 and pelleted by centrifugation (5 000 × *g*, 10 min, 4 °C). These cell pellets were now washed twice with CGXII MM broth prior to resuspension in a 20 ml volume of CGXII MM broth. The resuspended cells were next transferred to a baffled 500 ml flask and incubated at 30 °C

with orbital shaking (120 rpm). Throughout this incubation period, 0.5 ml and 1.5 ml aliquots of culture were withdrawn from each flask at different time intervals and stored at -20 °C. After thawing, each 0.5 ml aliquot was centrifuged (15 000 × *g*, 2 min) and the harvested cells were dried at 65 °C for 16 h in a desiccator. Each dried pellet was subsequently weighed in order to determine the dry cell mass within each aliquot. The cells present in each 1.5 ml aliquot of culture were mechanically disrupted by 10 min of sonication with a Virsonic Digital 600 sonicator (Virtis) using alternating 1 min pulsing and cooling cycles. After the removal of insoluble cell debris by centrifugation (15 000 × *g*, 5 min, 4 °C), 50% (v/v) trichloroacetic acid (Sigma-Aldrich) was added to the supernatant at a final concentration of 10% (v/v) and the resultant mixture was incubated at 4 °C for 10 min so as to allow for the precipitation of soluble protein (Sambrook *et al.*, 2001). Any precipitated cell material was subsequently removed by centrifugation (15 000 × *g*, 5 min, 4 °C) and the L-arginine and L-citrulline levels in the clarified sample supernatant now measured (Section 3.3.8; Section 3.3.9).

3.3.8 L-Arginine quantitation

3.3.8.1 Colorimetric assay

A method developed by Sakaguchi (1925), incorporating several improvements recommended by Rosenberg *et al.* (1955) and Tomlinson and Viswanatha (1974), was adopted for the quantitation of L-arginine. A developing solution was first prepared by mixing 2 ml of a 25% (w/v) α-naphthol solution, prepared in *n*-propanol, with 250 µl of a 1% (v/v) diacetyl solution, also prepared in *n*-propanol. The final solution was now made up to 10 ml with *n*-propanol. All solutions were prepared fresh upon use. Diacetyl and α-naphthol were obtained from Merck, whereas *n*-propanol was purchased from Kimax. For each reaction, 0.6 ml of dH_2O, 0.1 ml of 3 *N* NaOH, 0.1 ml of sample and 0.2 ml of developing solution were mixed in a 1.5 ml Eppendorf tube. The colour development reaction was subsequently allowed to proceed at room temperature in the dark and, after 30 min of incubation, the intensity of the resultant pink colour complex was quantified by OD_{535} measurement using a Novaspec-II spectrophotometer. CGXII MM broth, containing 10% (v/v) trichloroacetic acid, was used as a reaction blank and L-arginine supplied by Sigma-Aldrich was used for the construction of a standard curve (Appendix E). L-Arginine levels were normalised to the culture dry cell mass for each time point.

3.3.8.2 Mass spectrometer high performance liquid chromatography

The L-arginine content of the samples prepared in Section 2.3.7 was also quantitated using mass spectrometer high performance liquid chromatography (MS-HPLC). This was performed by Marietjie Stander (Department of Biochemistry, University of Stellenbosch, Stellenbosch, South Africa) using a Quattro Micro API (Waters). The standard EZ:Faast method was adopted as described in the user guide. Homoarginine and homophenylalanine were included as internal standards.

3.3.9 L-Citrulline quantitation

A protocol based on the colorimetric method of Knipp and Vašák (2006) was used for the quantitation of L-citrulline. Solution A was prepared first by adding diacetyl monoxime (Sigma-Aldrich) and thiosemicarbazide (Sigma-Aldrich) to final concentrations of 80 mM and 2 mM respectively in a 200 ml aqueous solution. Solution A was stored in the dark at 4 °C and not used if more than a month old. Solution B was prepared as follows: 200 ml of 85% H_3PO_4 (Merck) was added to 450 ml of dH_2O under gentle stirring. To this solution, 330 ml of 98% H_2SO_4 (Merck) was slowly added with gentle mixing, followed by the addition of 750 mg of $NH_4Fe(SO_4)_2 \cdot 12H_2O$ (Sigma-Aldrich). The solution was allowed to cool to room temperature and the final volume adjusted to one litre with dH_2O. A colour developing solution, consisting of one volume of Solution A and three volumes of Solution B, was freshly prepared for each assay and used within 1 h. For each colorimetric reaction, 240 µl of sample was mixed with 800 µl of colour developing solution in a 1.5 ml Eppendorf tube and heated to 95 °C in a heating block with a hot metal plate resting on top of the tubes. After 15 min, each tube was allowed to cool at room temperature for 10 min and the optical density of each sample was measured at 541 nm using a Novaspec-II spectrophotometer. CGXII MM broth supplemented with 10% (v/v) trichloroacetic acid was used as a reaction blank. L-Citrulline, supplied by Sigma-Aldrich, was used for the construction of a standard curve (Appendix E). L-Citrulline levels were normalised to the culture dry cell mass at each time point.

3.3.10 Assessment of growth in the presence of L-canavanine

The ability of *C. glutamicum* strains to grow in the presence of the L-arginine analogue L-canavanine was assessed in CGXII MM supplemented with 40 µg µl^{-1} uracil (Merck). Briefly, 100 µl of saturated culture was used to inoculate 5 ml of CGXII MM broth containing uracil supplemented with L-canavanine before being incubated for 16 h. Different concentrations of L-canavanine were used for each strain and the degree of growth was scored the following day.

3.3.11 Statistical analysis

Where appropriate, data values are represented in the mean ± SEM format. Unless otherwise stated, all experiments were performed on three biological replicates. The unpaired, two-tailed Student's t-test was used to analyse experimental variance. If a *P* value of less or equal to 0.05 was obtained using this test, each of the analysed mean values were reported to be significantly different from the other.

3.4 Results

3.4.1 Maximum growth rate determination

Any mutation introduced into an industrial strain should ideally not deleteriously affect its viability. For this reason, the various *C. glutamicum* strains (SCOH, SCOR, SCOR-D3, DCOR-D3 and 9A) used in this study, were grown in CGXII MM in the absence of selective pressure in order to determine how their respective mutations affect their maximum growth rates (μ_{max}) (Figure 3.1; Table 3.1). The μ_{max} for each *argR* mutant (SCOR, SCOR-D3, 9A and DCOR-D3) and the WT strain was found to occur between approximately 7 h and 14 h of growth and, for these mutant strains, was reduced relative to that of the WT strain regardless of the presence of L-arginine in the growth medium. Upon further analysis, however, these apparent differences were found not to be statistically significant, with the *P* value for each strain exceeding 0.10 irrespective of L-arginine supplementation. SCOH, on the other hand, did not grow in CGXII MM and, when supplemented with L-arginine, attained a significantly lower growth rate ($0.02 \pm 0.00\ h^{-1}$) compared to the other strains within the 7 h to 14 h time period ($P < 0.001$ compared to the WT). The final optical density of the SCOH culture, however, appeared to ultimately reach similar numbers to that for the WT after 23 h of growth, with the μ_{max} ($0.48 \pm 0.05\ h^{-1}$) for the entire incubation period occurring between 17 h and 23 h (Figure 3.1; Table 3.1). This sudden increase in the cell numbers of this culture is suggestive of the emergence of a sub-population of cells lacking the *argH* mutation and, as part of the work in the next section, this aspect was examined further.

3.4.2 Assessment of suicide vector integration stability in 9A, SCOR and SCOH

Also from an industrial point of view, the stability of any introduced mutations is also important. The stability of integration of pARGR-INT, pARGRKO and pARGHKO in the chromosomes of *C. glutamicum* 9A, SCOR and SCOH was thus assessed. This was performed by propagating these strains in the absence of antibiotic selective pressure for approximately 48 generations (Figure 3.2A). Cultures inoculated with either strain 9A or SCOR, which respectively possess either pARGR-INT or pARGRKO inserted into their *argR* genes, each displayed at the end of their incubation periods, a sub-population of Kan^R cells that, as a percentage of the total cell population, did not differ significantly from the

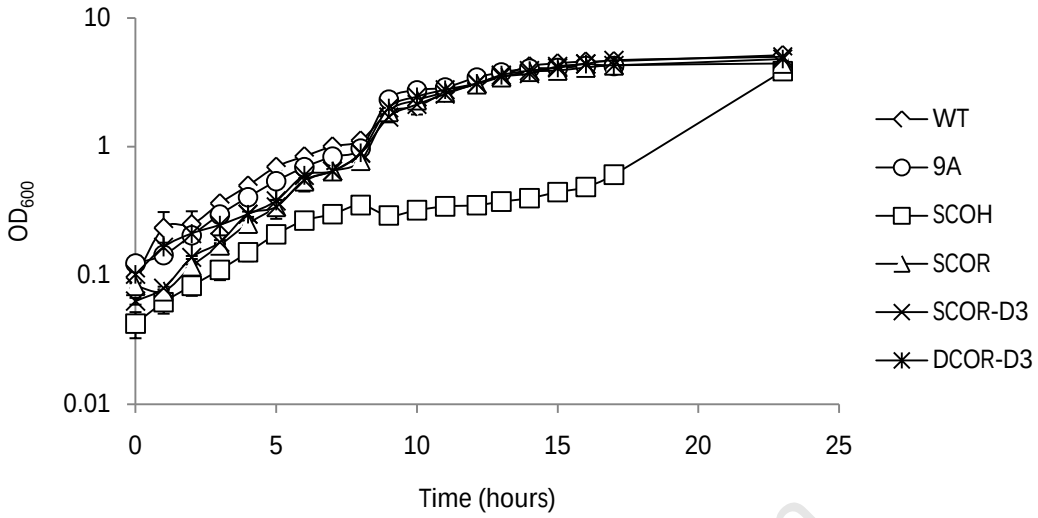


Figure 3.1 A growth curve of *C. glutamicum* ATCC 13032 and various *argH* and *argR* mutants grown in CGXII MM broth in the presence of L-arginine. Except for SCOH, growth in the absence of L-arginine was similar for each strain and, for the sake of clarity, hence omitted from the above graph. SCOH did not grow in the absence of L-arginine.

Table 3.1 Maximum growth rates (μ_{max}) ± SEM of *C. glutamicum* ATCC 13032 and various *C. glutamicum* recombinant strains grown in CGXII MM supplemented with L-arginine.

Strain	μ_{max} (h^{-1}) [a]	
	+ L-arginine	- L-arginine
ATCC 13032	0.45 ± 0.04	0.43 ± 0.07
SCOH	0.48 ± 0.05 (0.02 ± 0.00) [b]	No growth
SCOR	0.39 ± 0.05	0.36 ± 0.04
SCOR-D3	0.37 ± 0.04	0.38 ± 0.06
DCOR-D3	0.42 ± 0.03	0.41 ± 0.06
9A	0.41 ± 0.01	0.41 ± 0.03

[a] typically occurred between 7 h and 14 h of growth for each strain and was calculated by the method of Willey *et al.* (2007).

[b] for SCOH, μ_{max} occurred between 17 h and 23 h of growth; its growth rate between 7 h and 14 h is indicated in parenthesis.

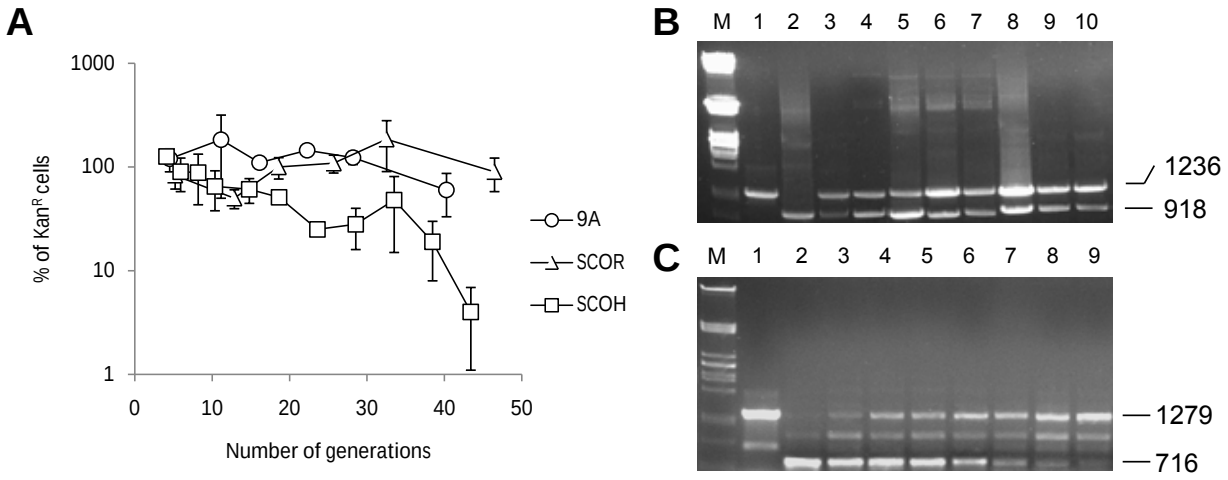


Figure 3.2 Stability of pARGR-INT, pARGRKO and pARGHKO integration into *C. glutamicum* ATCC 13032. **(A)** The percentage of Kan^R cells in broth cultures of *C. glutamicum* 9A, SCOR and SCOH was determined by replica plating on 2 × YT solid medium in the presence of Kan. The generation time for each strain in 2 × YT broth was calculated according to Willey *et al.* (2007). **(B)** Results of a PCR-based analysis showing the persistence of both *argR* alleles in SCOR. Fragment sizes are indicated in bp. Lane M: marker DNA. Lanes 1, 2 and 3 contain PCR products amplified from WT genomic DNA, pARGRKO DNA and SCOR genomic DNA respectively. The remaining lanes contain PCR products derived from cells sampled at different generations (indicated in parenthesis for each lane): lane 4 (0), lane 5 (5), lane 6 (13), lane 7 (18), lane 8 (26), lane 9 (33) and lane 10 (47). **(C)** Results of a PCR-based analysis showing the loss of the *ΔargH* allele in SCOH. Fragment sizes are indicated in bp. Lane M: marker DNA. Lanes 1, 2 and 3 contain PCR products amplified from WT genomic DNA, pARGHKO DNA and SCOH genomic DNA respectively. The remaining lanes contain PCR products amplified from cells sampled at different generations (the number of generations is indicated in parenthesis for each lane): lane 4 (0), lane 5 (8), lane 6 (15), lane 7 (24), lane 8 (34) and lane 9 (43).

percentage of Kan^R cells present at the beginning of their incubation periods. For 9A, for example, after zero generations of growth, all cells appeared to be Kan^R, whereas, after 41 generations of growth, approximately 73 ± 6.36% of the total cell population displayed Kan^R (P = 0.08). Similarly, all SCOR cells at the start of the incubation period expectedly proved to be Kan^R and, after 47 generations of growth, approximately 90 ± 4.41% of the total cell population displayed Kan^R (P = 0.14). In contrast to 9A and SCOR, however, the percentage

of Kan^R cells present within each SCOH culture declined significantly after prolonged growth. For example, only 4.0 ± 1.20% of the total SCOH cell population displayed Kan^R after 43 generations of growth ($P < 0.001$ when compared to the percentage of 9A cells displaying Kan^S after the same period of incubation).

As evident from Figure 3.2B, two varieties of *argR* junction PCR products, a 1236 bp fragment and a 918 bp fragment, respectively representative of the WT and *ΔargR* alleles, are continuously present within the population of SCOR cells. In agreement with the relatively constant proportion of SCOR cells in possession of Kan^R throughout this period (shown in Figure 3.2A), the proportion of *ΔargR* to *argR* amplicons shown in Figure 3.2B did not appear to significantly diminish with time. In contrast, a similar analysis conducted using the *argH* junction primers on SCOH cells revealed the appearance of the 716 bp *ΔargH* PCR fragment to diminish as the duration of incubation increased (Figure 3.2C). Indeed, this *ΔargH* allele was barely detected towards the end of the SCOH incubation period. The appearance of the 1279 bp product fragment representative of the WT *argH* allele, however, appeared to increase in intensity throughout the incubation period. It thus appears that the loss of the integrated *ΔargH*-associated pARGHKO plasmid is occurring throughout this incubation period, and explains the high growth rate seen during 17 h and 23 h in this strain (Table 3.1).

3.4.3 Transcriptional organisation of the L-arginine biosynthetic genes

To gain a better understanding of how mutagenesis of the L-arginine biosynthetic gene cluster may affect the transcription of these genes, their transcriptional organisation should first be characterised in *C. glutamicum* ATCC 13032. As part of a preliminary characterisation, this was accomplished using a RT-PCR-based strategy, which suggested that the L-arginine biosynthetic genes shown in Figure 3.3 are all transcribed together, as part of a single polycistronic *argCJBDFRGH* transcript approximately 8.9 kb length. Oligonucleotide primers, forming amplicons that individually bridge each of the intergenic regions shown in Figure 3.3A, were used as part of a two-step RT-PCR reaction to amplify cDNA derived from mid-exponential phase cells grown in CGXII MM. For each primer pair, genomic DNA was included as a positive control, together with RNA as a negative control, which was used at a concentration in excess to that predicted to occur in aliquots of cDNA used for each experimental reaction. The expected PCR product sizes for each primer

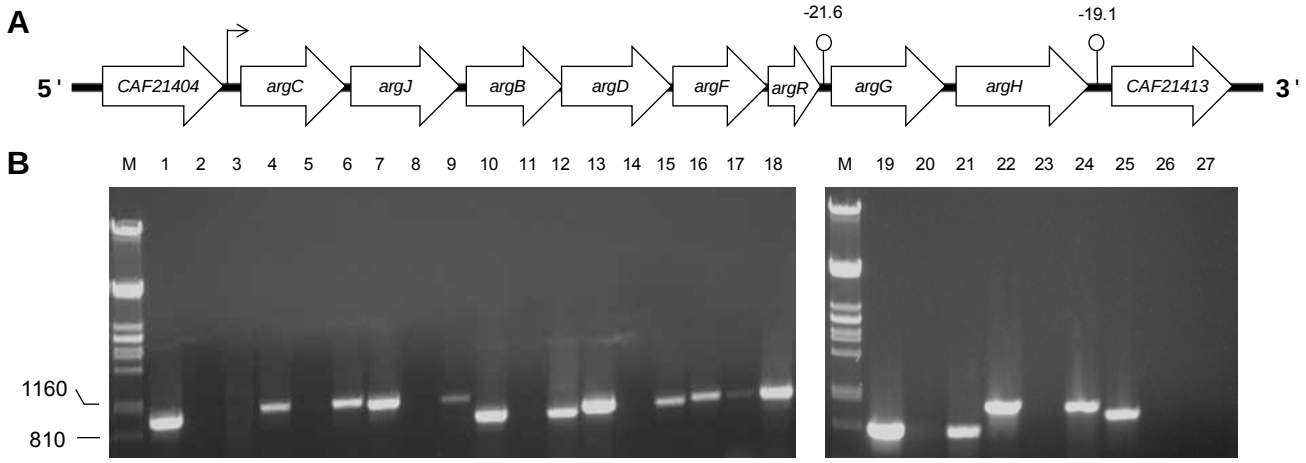


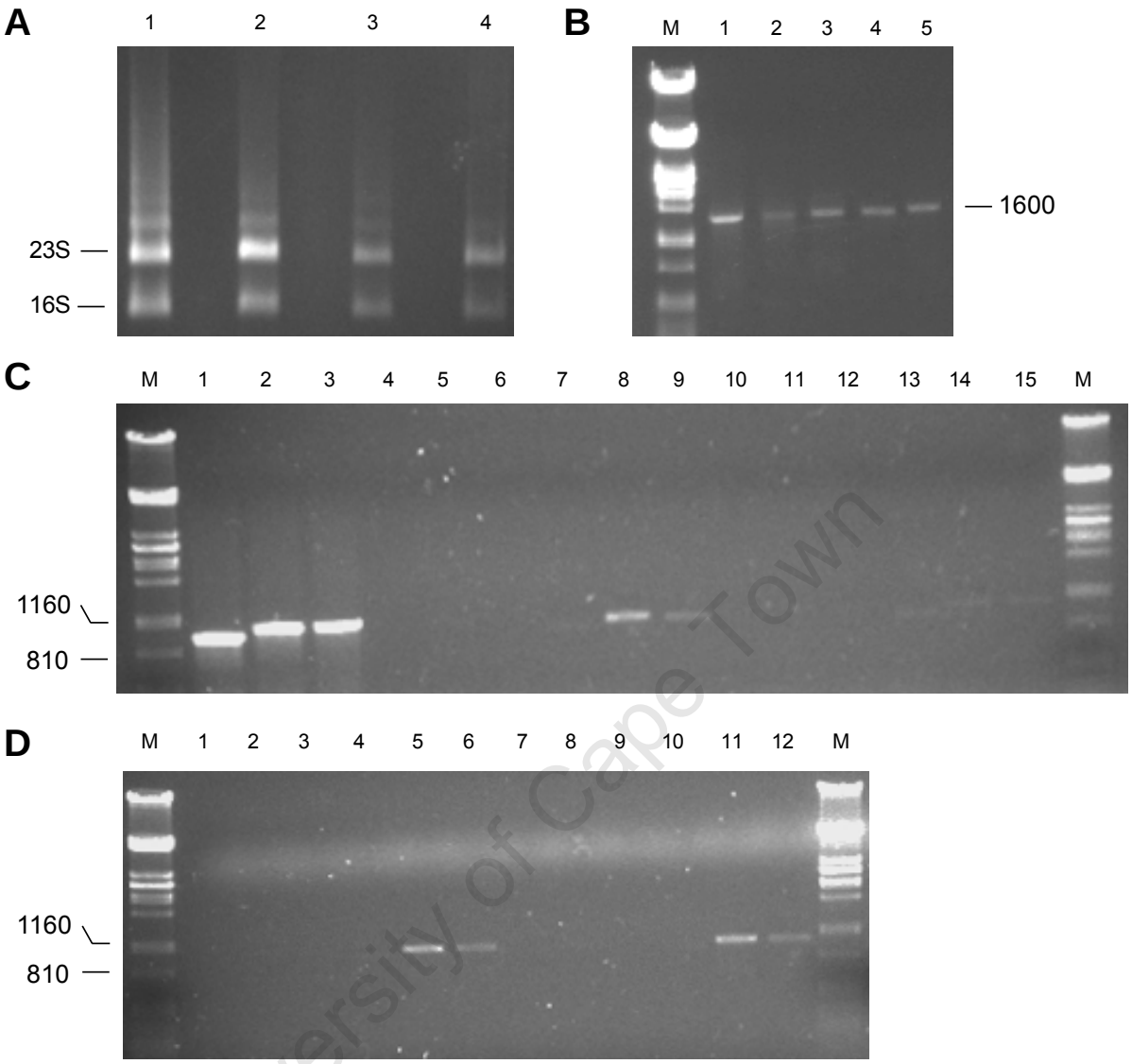
Figure 3.3 RT-PCR analysis of the transcriptional organisation of the L-arginine biosynthetic genes in mid-exponential phase *C. glutamicum* ATCC 13032 grown in CGXII MM. **(A)** The organisation of the L-arginine biosynthetic genes is annotated according to the *C. glutamicum* ATCC 13032 genome sequence (Genbank accession number: NC_006958) available on the National Centre for Bioinformatics web server (http://www.ncbi.nlm.nih.gov). The putative transcriptional start site is indicated by the bent arrow and putative transcription termination points, detected using the Mfold program, are indicated by the stem-loop structures. The calculated ΔG° value at 30 °C for each stem-loop is displayed in kcal per mol above each structure. Diagram not to scale. **(B)** The amplification products of a two-step RT-PCR analysis performed on the L-arginine biosynthetic genes are shown. A positive genomic DNA control and a RNA-only negative control were included for each primer reaction. Fragment sizes are indicated in bp. Lanes M: marker DNA. Lanes 1, 4, 7, 10, 13, 16, 19, 22 and 25 contain the products of positive control reactions performed on *C. glutamicum* ATCC 13032 genomic DNA. Lanes 2, 5, 8, 11, 14, 17, 20, 23 and 26 contain the products of negative control reactions performed on total RNA samples extracted from *C. glutamicum* ATCC 13032 cells. Lanes 3, 6, 9, 12, 15, 18, 21, 24, 27 contain the products of reactions using cDNA synthesised from total RNA samples as template DNA. The amplicon of each primer combination spans a specific intergenic region and the primer combinations for each lane are: lanes 1-3, *CAF21404-argC* F and *CAF21404-argC* R; lanes 4-6, *argC-argJ* F and *argC-argJ* R; lanes 7-9, *argJ-argB* F and *argJ-argB* R; lanes 10-12, *argB-argD* F and *argB-argD* R; lanes 13-15, *argD-argF* F and *argD-argF* R; lanes 16-18, *argF-argR* F and *argF-argR* R; lanes 19-21, *argR-argG* F and *argR-argG* R; lanes 22-24, *argG-argH* F and *argG-argH* R; lanes 25-27; *argH-CAF21413* F and *argH*-CAF21413 R. The expected product sizes for each primer pair are detailed in Appendix C.

combination are listed in Appendix C and, as expected, such products were obtained from each reaction when genomic DNA was used as a template. PCR product fragments of the expected size were also obtained from cDNA as part of each experimental reaction, with the exception of those that use primers which fall outside of the L-arginine biosynthetic gene cluster (namely the *CAF21404-argC* F primer and the *argH-CAF21413* R primers). This suggests that the L-arginine biosynthetic genes form a single transcript and, because each primer pair did not produce any PCR products of the appropriate size when only DNAse I-treated total RNA was used as a template, this result does not appear to be due to the presence of contaminating genomic DNA within each cDNA sample. Indeed, upon the examination of the sequence of this gene cluster using the DNA Mfold program (available from http://mfold.bioinfo.rpi.edu/), a strong putative transcriptional terminator stem-loop structure was detected downstream of the *argH* gene (Zuker, 2003). A similar, second structure was also detected within the *argR-argG* intergenic region, yet in this case, it does not seem to completely abolish the formation of an *argCJBDFRGH* transcript.

3.4.4 Transcription of the L-arginine biosynthetic genes in DCOR-D3

A further RT-PCR-based analysis of the L-arginine gene cluster was performed in order to determine how the deletion of the *argR* gene in DCOR-D3 affected the transcription of the L-arginine biosynthetic genes in this bacterium (Figure 3.4). The *CAF21404-argC*, *argC-argJ*, *argG-argH* and F27-R5 16S primer pairs were therefore used to analyse cDNA derived from *C. glutamicum* ATCC 13032 and DCOR-D3 cells grown with or without L-arginine supplementation.

As expected from Section 3.4.3, the *CAF21404-argC* primer pair failed to produce a PCR product fragment off cDNA derived from either *C. glutamicum* ATCC 13032 cells or DCOR-D3 cells (Figure 3.4). In contrast, when the *argC-argJ* and *argG-argH* primer pairs were used for the amplification of cDNA derived from these cells grown without L-arginine supplementation, the expected 1037 bp and 1036 bp product fragments were observed for each primer pair (Figure 3.4C). Such PCR product fragments could not, however, be attained when cDNA from WT cells grown in the presence of L-arginine was used in each reaction, despite the fact that a 1600 bp product fragment was obtained from the same cDNA samples upon using the F27-R5 16S primer pair (Figure 3.4B). Furthermore, irrespective of whether or not L-arginine was included in the growth medium, cDNA derived



The caption for **Figure 3.4** is shown on the next page.

from DCOR-D3 cells was able to give rise to product fragments of the expected size when amplified using the *argC-argJ* and *argG-argH* primer pairs (Figure 3.4D). The occurrence of these bands does not appear to be due to the presence of contaminating genomic DNA as each primer pair was unable to produce any PCR products of the predicted size from each RNA sample (Figure 3.4C and D; Appendix C). Interestingly, the *argG-argH* primer pair typically produced significantly less of its expected PCR product, regardless of the variety of cDNA used.

Figure 3.4 RT-PCR analysis of the expression of the L-arginine biosynthetic gene cluster in mid-exponential phase *C. glutamicum* ATCC 13032 and DCOR-D3 cells grown in CGXII MM in the presence of L-arginine. Fragment sizes are indicated in bp. Lanes M: marker DNA. **(A**) MOPS-buffered electrophoresis was routinely used to assess the integrity of RNA extracted from either *C. glutamicum* ATCC 13032 cells (lanes 1 and 2) or DCOR-D3 cells (lanes 3 and 4), grown with (lanes 2 and 4) or without (lanes 1 and 3) L-arginine in the medium. **(B)** A control RT-PCR performed using the F27 and R5 16S primer pair on cDNA derived from *C. glutamicum* ATCC 13032 cells (lanes 2 and 3) or DCOR-D3 cells (lanes 4 and 5) grown with (lanes 3 and 5) or without (lanes 2 and 4) L-arginine supplementation. WT genomic DNA was used as template for the reaction products shown in lane 1. **(C)** The transcription levels of the *argCJBDFRGH* gene cluster in *C. glutamicum* ATCC 13032 cells was assessed using the primers *CAF21404-argC* (lanes 1, 4, 7, 10 and 13), *argC-argB* (lanes 2, 5, 8, 11 and 14) and *argG-argH* (lanes 3, 6, 9, 12 and 15). The following templates were used for each reaction: lanes 1-3, WT genomic DNA; lanes 4-6, RNA from WT cells grown in the absence of L-arginine; lanes 7-9, cDNA from WT cells grown in the absence of L-arginine; lanes 10-12, RNA from WT cells grown in the presence of L-arginine; lanes 13-15, cDNA from WT cells grown in the presence of L-arginine. **(D)** The transcriptional levels of the *argCJBDFRGH* gene cluster in DCOR-D3 cells were assessed using the primer pairs *CAF21404-argC* (lanes 1, 4, 7 and 10), *argC-argJ* (lanes 2, 5, 8, 11 and 14) and *argG-argH* (lanes 3, 6, 9 and 12). The following templates were used for each reaction: lanes 1-3, RNA from DCOR-D3 cells grown in the absence of L-arginine; lanes 4-6, cDNA from DCOR-D3 grown in the absence of L-arginine; lanes 7-9, RNA from DCOR-D3 cells grown in the presence of L-arginine; lanes 10-12, cDNA from DCOR-D3 cells grown in the presence of L-arginine.

3.4.5 L-Arginine production

The impact of each mutation on the L-arginine production ability of *C. glutamicum* was next assessed (Figure 3.5A and B). Relative to the WT, SCOR and 9A generally accumulated significantly more L-arginine throughout the 24 h incubation period. For example, after 6 h of incubation, these mutant strains had accumulated a total of 14.86 ± 3.57 ($P < 0.05$) and 21.59 ± 2.56 µmol of L-arginine (mg dry cell mass)$^{-1}$ ($P < 0.01$) respectively, whereas the WT strain had accumulated only 8.13 ± 2.56 µmol L-arginine (mg dry cell mass)$^{-1}$. Similarly, after 24 h of incubation, these amounts had reached 23.18 ± 1.31 and 14.03 ± 1.15 µmol (mg

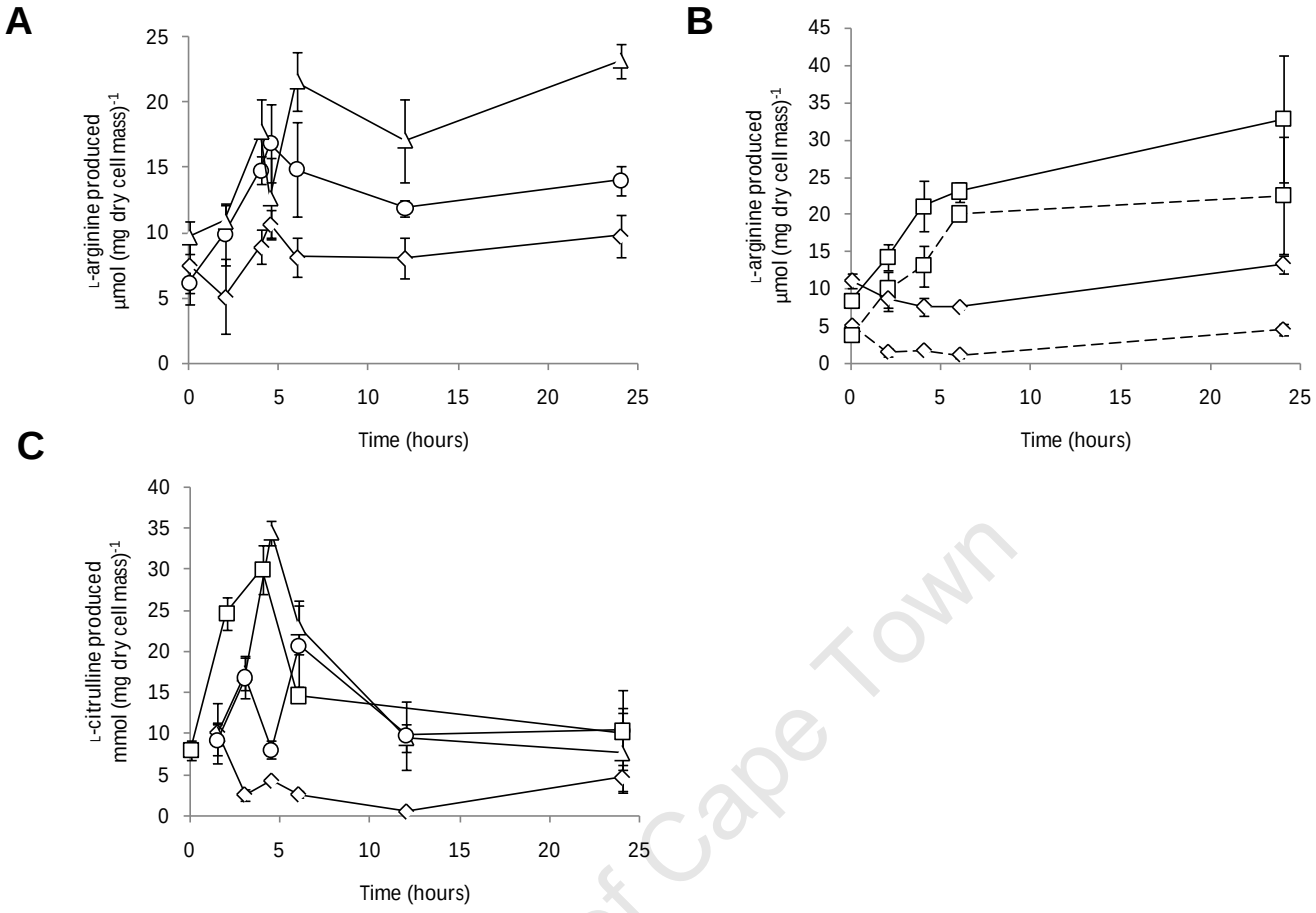


Figure 3.5 L-Arginine **(A and B)** and L-citrulline **(C)** accumulation by various *C. glutamicum argR* mutant strains incubated in CGXII MM broth. Production of the relevant compound by the WT strain (diamonds) is shown in all diagrams. L-Arginine production by 9A (triangles) and SCOR (circles) is shown in **A** and L-arginine production by DCOR-D3 (squares) is shown in **B**. L-Citrulline accumulation by strains 9A (triangles) and DCOR-D3 (squares) is shown in **C**. Solid lines represent the concentration of compounds measured using a colorimetric method, whereas the dash lines represent those measured using MS-HPLC.

dry cell mass)$^{-1}$ for 9A and SCOR respectively, with both differing significantly from the final quantity of L-arginine obtained by the WT strain (9.75 ± 1.6 µmol L-arginine [mg dry cell mass]$^{-1}$) ($P < 0.05$ and $P < 0.01$) as well as from each other ($P < 0.01$). As evident from Figure 3.5B, DCOR-D3 also reached significantly higher levels of L-arginine relative to the WT strain when measured using the colorimetric method. This occurred after 4 h and 6 h, reaching L-arginine levels of 21.04 ± 3.36 and 23.07 ± 1.33 µmol (mg dry cell mass)$^{-1}$ respectively, compared to only 7.58 ± 1.10 and 7.49 ± of L-arginine (mg dry cell mass)$^{-1}$

accumulated by the WT strain ($P < 0.05$ for both time points). MS-HPLC also confirmed these two strains to differ significantly after 4 h and 6 h of incubation and returned values for DCOR-D3 of 13.07± 2.69 and 20.01 ± 0.36 µmol (mg dry cell mass)$^{-1}$ for each time point, whereas the WT strain only achieved L-arginine quantities of 1.71 ± 0.13 and 1.13 ± 0.41 µmol (mg dry cell mass)$^{-1}$ for each of these time points respectively (both P values < 0.05).

3.4.6 L-Citrulline production

Strains in possession of a disrupted L-arginine biosynthetic pathway may experience a distorted metabolic flux. One such metabolite in this pathway, L-citrulline, is present at an important control point and, if a metabolic bottleneck occurs within this biosynthetic pathway under conditions of derepression, this compound may accumulate to atypical levels. To assess this, the levels of L-citrulline in the clarified cell free extracts used in Section 3.4.5 were quantitated.

After 4 h of incubation DCOR-D3 had produced significantly more L-citrulline (30.01 ± 2.98 µmol [mg dry cell mass]$^{-1}$ than the WT strain (4.07 ± 0.29 µmol [mg dry cell mass)$^{-1}$]) ($P <$ 0.001). A similar trend was observed for 9A, which reached levels of 34.41 ± 1.44 µmol of L-citrulline (mg dry cell mass)$^{-1}$ after 4.5 h, whereas the WT had reached only 4.28 ± 0.24 µmol (mg dry cell mass)$^{-1}$ after the same period ($P < 0.001$). In contrast, the level of L-citrulline achieved by SCOR peaked after 6 h of incubation (reaching a level of 20.73 ± 4.92 µmol [mg dry cell mass]$^{-1}$ as opposed to 2.63 ± 0.40 µmol [mg dry cell mass]$^{-1}$ for the WT at this time point [$P < 0.01$]).

3.4.7 L-Canavanine resistance

L-Canavanine is a non-functional structural analogue of L-arginine. Most bacteria are unable to distinguish between these two compounds and, when L-canavanine is mistakenly incorporated into proteins instead of L-arginine, they become structurally aberrant and can result in cell death (Kisumi *et al.,* 1971; Volcani and Snell, 1948). All of the *argR* mutants displayed an increased resistance to L-canavanine in the presence of uracil relative to the WT. For example, an L-canavanine concentration of 25 µg ml^{-1} was sufficient to completely inhibit WT growth. Growth by SCOR, on the other hand, was severely weakened at an L-

canavanine concentration of 65 µg ml^{-1} and this strain ceased to grow at 100 µg ml^{-1}. 9A and DCOR-D3, however, continued to display strong growth at this higher concentration.

3.5 Discussion

In this chapter, the *C. glutamicum* mutant strains created in Chapter 2 were characterised using different molecular and culture dependent techniques. The industrial suitability of each was initially assessed though an investigation of their viabilities and, where appropriate, the stability of any integrated plasmid DNA was also examined. This was followed by a characterisation of the transcriptional organisation of the L-arginine biosynthetic gene cluster and the subsequent determination of the effect of *argR* deletion in DCOR-D3 cells on the transcription of some of these biosynthetic genes. Their L-arginine and L-citrulline production capacities were also examined in addition to their ability to grow in the presence L-canavanine.

3.5.1 Growth characteristics

With the exception of SCOH, the *C. glutamicum* mutant strains used in this study displayed a μ_{max} in CGXII MM broth that did not differ significantly from the μ_{max} of *C. glutamicum* ATCC 13032 cells grown under the same conditions. Additionally, the μ_{max} determined for the latter strain in the absence of L-arginine supplementation (0.45 h^{-1}) was similar to that previously reported where, for example, Lange *et al.* (2003) found *C. glutamicum* ATCC 13032 to possess a μ_{max} of approximately 0.38 h^{-1} when grown in CGXII MM. Similar to what has been observed for other *argR* integration mutant strains, such those derived from *Lactococcus lactis*, each of the *argR* mutant strains in this study was able to grow strongly in MM broth regardless of L-arginine supplementation (Larsen *et al.*, 2004). Interestingly, this is different to what has previously been observed upon the deletion of *argR* in *E. coli*, where it was found to cause a significant decrease in growth rate. This was found to be due to the limited availability of carbamoyl phosphate for pyrimidine biosynthesis due to the increased activity of carbamoyl phosphate synthetase CPS) and increased flux down the L-arginine biosynthetic pathway (Weerasinghe *et al.*, 2006). The reduction in grown rate was subsequently alleviated by the addition of pyrimidines to the growth medium. The fact that a decreased growth was not observed in DCOR-D3 suggests that, even after *argR* deletion, the *C. glutamicum* CPS perhaps remains capable of fulfilling the needs of both the pyrimidine and L-arginine biosynthetic pathways and that further L-arginine overproduction might be constrained by another, more limiting property of the cell. If the availability of bicarbonate, glutamine or ADP (all of which are

required for the CPS activity) became limited, however, it remains likely the growth rate of DCOR-D3 would be significantly reduced as the cell would not be able to fully repress L-arginine biosynthesis and the diminished quantity of CP now available would probably not be sufficient to fill the needs of both pathways and pyrimidine biosynthesis would now therefore be restricted.

SCOH cells, on the other hand, clearly appear to be auxotrophic and, in line with what was discovered during the attempted generation of *argH* double cross-over mutants in Section 2.5.2, it proved to display a very slow growth rate for the majority of its incubation period even when supplemented with L-arginine. The inability of this compound to rescue the growth phenotype of SCOH immediately points to a second role for argininosuccinate lyase important for cell viability. As discussed in Section 2.5.2, however, the evidence for such a role is scant and, although it has been proposed that this enzyme may play a critical role in the detoxification of ammonia in certain bacteria, such a role remains to be clearly elucidated (Cohen-Kupiec *et al.,* 1999). Therefore, while *argR* deletion appears to have no effect on the μ_{max} of *C. glutamicum* cells, the mutation of *argH* is, from an industrial point of view, undesirable as a potential auxotrophic marker.

3.5.2 Insertional mutant stability

By monitoring the presence of Kan^R in cultures of 9A, SCOR and SCOH, the stability of the plasmid DNA integrated into each host's chromosome was assessed. This was also evaluated using primers that anneal to the appropriate plasmid DNA-encoded mutant allele. Whereas 9A and SCOR were shown to contain integrated plasmid DNA that was relatively stable, the plasmid-encoded *ΔargH* allele in SCOH cells proved to be transitory in the absence of selective pressure. It thus appears that the dramatic increase in the number of SCOH cells seen after 17 h of growth shown in Figure 3.1, is due to the emergence of a fast growing sub-population of Kan^S cells that have lost their plasmid associated marker genes and *ΔargH* allele.

3.5.3 Transcriptional organisation of the L-arginine biosynthetic genes

At the time of the work conducted in Section 3.4.4, the transcriptional organisation of the L-arginine biosynthetic gene cluster was unpublished. Using RT-PCR, we subsequently

showed that these genes appear to be transcribed together as a single transcript. This, however, appears to be contradictory to the L-arginine auxotrophy observed amongst the *argR* integration mutants in Section 3.4.1, as this implies that a second transcriptional initiation site that is located downstream of the putative site of plasmid integration and precedes the *argG* coding region. Furthermore, as part of work recently published on FarR, a negative transcriptional regulator of the L-arginine biosynthetic gene cluster in *C. glutamicum* ATCC 13032, these genes were reported from RT-PCR analysis to be transcribed as two separate *argCJBDFR* and *argGH* transcripts (Hänßler *et al.*, 2007). Also taking into account the possible presence of a stem-loop structure in possession of significant free energy within the *argR-argG* intergenic region as part of this work, we repeated the RT-PCR several times using different RNA samples extracted from independent *C. glutamicum* ATCC 13032 cultures. In each instance, however, the results returned verified that all these genes were indeed transcribed together. Importantly, the RNA samples used for cDNA synthesis were also repeatedly shown to be free of contaminating genomic DNA.

The transcription of the *argCJBDFRGH* genes as a single unit is, to the author's knowledge, so far unreported in other microbes. This would indeed be highly unusual, because, as discussed in Section 1.5.1, other closely-related gram positive bacteria, such as *Streptomyces clavuligerus* and *Lactobacillus plantarum*, generally possess a transcriptional organisation of the L-arginine biosynthetic genes similar to that proposed by Hänßler *et al.* (2007). Taking this into account, it is conceivable that the *C. glutamicum* ATCC 13032 L-arginine regulon has a *argCJBDFR* transcriptional terminator that could be regulated via another mechanism and thus give rise to a significant sub-population of more numerous *argCJBDFR* and *argGH* transcripts as well as the larger *argCJBDFRGH* transcript. Notably, in this instance, RT-PCR would detect the largest transcript that spans the *argR-argG* intergenic region and thus could mask the presence of these smaller sub-transcripts. Also of consideration, is the fact that large transcripts such as *argCJBDFRGH*, are generally less amenable to cDNA conversion and can be biased against depending on the RNA extraction protocol (Sambrook *et al.*, 2001). These reasons could account for this transcript not been detected by Hänßler *et al.* (2007).

3.5.4 Transcription of the L-arginine biosynthetic genes in DCOR-D3

An analysis of the L-arginine biosynthetic gene cluster in DCOR-D3 cells indicated that in the presence of L-arginine these genes continue to be transcribed. If, as proposed by Hänßler *et al.* (2007), the L-arginine biosynthetic genes are indeed transcribed as two separate transcripts, the appearance of the *argGH* transcript in DCOR-D3 cells grown in the presence of L-arginine (shown in Figure 3.4) suggests that ArgR normally represses its formation in addition to that of the larger *argCJBDFR* transcript. In contrast to the mutant strain, when *C. glutamicum* ATCC 13032 cells were grown under the same conditions, cDNA fragments indicative of the transcription of these genes could not be detected. Importantly, as the transcriptional levels of the 16S gene in both strains appears to remain relatively constant irrespective of the presence of L-arginine, we can conclude that transcriptional repression of the L-arginine biosynthetic gene cluster in DCOR-D3 cells has been removed by the deletion of the *argR* gene. According to our current understanding of ArgR-mediated regulation, we would thus expect this protein to bind within the *argR-argG* intergenic region, as well as to the region preceding the *argC* gene. At the time of this experiment, this provided the first known experimental evidence of the role of ArgR as a transcriptional repressor of the L-arginine biosynthetic genes in *C. glutamicum* ATCC 13032. At this stage, it is important to emphasise the limitations of this approach and to avoid making any inferences regarding the relative quantities of each transcript in the absence of a functional *argR* gene (this is chiefly due to the absence of proper normalisation and the non-linearity of the RT-PCR performed here). A technique such as quantitative real-time PCR (qPCR) or Northern hybridisation would thus serve extremely useful here in order to accurately calculate the relative fold expression of each gene.

Indeed, very recently this role of ArgR in *C. glutamicum* ATCC 13032 had been confirmed as part of a GeneChip-based study of the transcriptional levels of the arginine biosynthetic gene cluster, where the inactivation of the *argR* gene in a *C. glutamicum* stain already in possession of an *argB* point mutation was found to cause the significant derepression of all of these biosynthetic genes (Ikeda *et al.*, 2009). Interesting, while the level of derepression in this study by Ikeda *et al.* (2009) is similar for the *argCJBDFR* genes (approximately 20-fold), it appears to be considerably less for the *argGH* genes (approximately 5-fold). Moreover, as observed in Section 3.4.4, the RT-PCR products amplified from *C. glutamicum*

ATCC 13032 or DCOR-D3 cDNA using primers that span either the *argC-argB* or *argG-argH* intergenic regions appeared to be present in relatively different quantities. Although this could, for example, be influenced by various experimental factors including the differing efficiencies of primer binding, cDNA conversion or the actual RT-PCR itself; it nonetheless appears to corroborate the work of Ikeda *et al.* (2009). Indeed, as shown by Hänßler *et al.* (2007) and discussed in Section 1.5.3, the *argR-argH* intergenic region contains a putative binding site for the FarR regulator and, although the deletion of *farR* does not affect L-arginine production, it appears that the binding of this or other, as yet uncharacterised, transcriptional regulators may allow for the finer control of *argGH* transcription than that provided by the relatively large ArgR-mediated *argCJBDFRGH* transcript.

3.5.5 Production of L-arginine

Both a colorimetric assay and a MS-HPLC-based method were used for the quantitation of L-arginine and both methods indicated maximal levels of this compound to be produced by the DCOR-D3 culture after 24 h of incubation where, according to the colorimetric assay, this strain accumulated approximately treble the amount of L-arginine compared to *C. glutamicum* ATCC 13032 cells. In contrast, MS-HPLC reported this difference to be approximately 5-fold and, although the L-arginine levels attained at this time point for DCOR-D3 are comparatively similar according to either method, this discrepancy is largely due to a significant reduction in the amount of WT-accumulated L-arginine detected by the MS-HPLC technique as opposed to the colorimetric assay. This latter method does, to a certain extent, possess a small degree of non-specific cross-reactivity with the guanidinium groups of other molecules and, from a technical standpoint, the MS-HPLC-based method appears to less affected by background levels of cell material present in the clarified cell extracts assayed (Tomlinson and Viswanatha, 1974). This MS-HPLC-based technique may thus be considered to be more accurate and more sensitive than the colorimetric assay used in this study.

Interestingly, although there is a clear relative fold difference observed between the experimental and control samples, the L-arginine quantities reported here are significantly lower than what has been previously observed for *C. glutamicum* ATCC 13032. For example, whereas the normalised total L-arginine values obtained here for the DCO mutant

reach amounts of least 20 µmol (mg dry cell mass)$^{-1}$ (equivalent to an L-arginine sample concentration of approximately 0.5 mM), Bellman *et al.* (2001) and Hänßler *et al.* (2007) have noted WT cells to accumulate L-arginine at concentrations as high as 5.0 mM upon incubation in MM broth. The extracellular L-arginine produced by these cells has also been shown to reach levels of approximately 2.0 mM (Bellman *et al.*, 2001). Importantly, as the concentrations presented in these studies appear not to be normalised and were detected using varying methods, it is probably not ideal to compare these absolute values across different experiments. Encouragingly, however, an L-arginine deregulated *L. plantarum* mutant strain created by Nicoloff *et al.* (2004) was also shown to accumulate L-arginine to a concentration of approximately 0.5 mM. Additionally, the inclusion of internal controls in our work should permit for any improvement in L-arginine production to be compared accurately in terms of relative fold increase. For example, although a 5-fold increase in L-arginine accumulated is reported here upon the deletion of the *C. glutamicum* ATCC 13032 *argR* gene, a study of an *E. coli* strain in possession of an inactivated *argR* gene found L-arginine production to increase roughly 13-fold relative to the WT strain (Momose *et al.*, 1984). This *E. coli* strain was already an L-arginine overproducer and had been created through random mutagenesis of the host genome.

Similarly, as part of a recent study published shortly after the completion of the work reported in this chapter, a *C. glutamicum* strain in possession of mutant *argB* gene encoding a feedback resistant *N*-acetylglutamate kinase (NAGK) enzyme was examined in terms of its L-arginine production capability (Ikeda *et al.*, 2009). This alone, however, was insufficient to significantly increase L-arginine biosynthesis and an *argR* DCO mutation was subsequently introduced into this strain in a manner similar to that adopted in Chapter 2 (Ikeda *et al.*, 2009). Here, although the absolute concentrations of L-arginine detected differ from those observed in our study, this latter mutation resulted in an approximate 28-fold increase in L-arginine production relative to the WT strain (where a final L-arginine concentration of 85 mM was attained for the mutant). This discrepancy in fold-increase in L-arginine production between our study and the work by Ikeda *et al.* (2009) suggests that further directed mutagenesis of the *C. glutamicum* ATCC 13032 genome that could, for example, focus on engineering feedback resistant NAGK and *N*-acetylglutamate synthase enzymes, further amplify the positive effect of *argR* deletion on L-arginine production. Indeed, the *argB* point mutations in the study by Ikeda *et al.* (2009) were

introduced into the WT strain based on nucleotide changes detected in the L-arginine biosynthetic genes of a classically mutated *C. glutamicum* L-arginine and L-citrulline producer strain. This mutant strain was also found to contain an *argR* point mutation, expected to cause the gene to encode a non-functional ArgR protein. However, upon the introduction of all mutations detected in the L-arginine biosynthetic gene cluster of this classically mutated strain into the WT genome, L-arginine production nonetheless failed to reach levels previously observed in the classical producer, thereby suggesting that further metabolic bottlenecks that limit L-arginine production likely continue exist upon the deletion of the *argR* gene. In summary, the enhanced L-arginine production levels displayed by DCOR-D3 clearly further boosts evidence for the role of the *C. glutamicum* ATCC 13032 ArgR as a repressor of L-arginine biosynthesis. Finally, although it was not investigated directly as part of our work, it appears possible that the tendency of SCOR to produce slightly reduced levels of L-arginine relative to DCOR-D3 could be due its possession of an intact copy of the *argR* gene.

3.5.6 Production of L-citrulline

This work has shown that transcription of the *C. glutamicum* ATCC 13032 *argCJBDFR* and *argGH* gene clusters are negatively regulated by ArgR in the presence of L-arginine. If, as suggested by Hänßler *et al.* (2007), these gene clusters are transcribed separately, it stands to reason that their transcription may be differentially affected by the deletion of the *argR* gene. Although the relative quantities of these transcripts has not been studied here, if an increased amount of L-citrulline is detected within DCOR-D3 cells, it may indicate that the heightened expression of the *argCJBDF* genes responsible for the synthesis of this compound are causing it to be produced faster than it may be utilised by the downstream argininosuccinate synthetase and argininosuccinate lyase enzymes, which are encoded by the *argGH* gene cluster. Alternatively, deletion of the *argR* gene may cause the *argGH* genes to be expressed less efficiently due to the destabilisation of the operon transcript downstream of the site of deletion. All the *argR* mutant strains in this study accumulated dramatically more L-citrulline than the WT (in the DCOR-D3 mutant L-citrulline production increased more than 8-fold), thereby suggesting that a potential metabolic bottleneck exists here and, as part of future strain improvements, improving the expression or the activity of the ArgG and ArgH enzymes would likely further enhance L-arginine production in ArgR-

deregulated *C. glutamicum* strains. This finding has recently been confirmed by Ikeda *et al.* (2009), who noted L-citrulline production to increase approximately 10-fold relative to *C. glutamicum* ATCC 13032 upon the deletion of the *argR* gene in a mutant *C. glutamicum* strain already encoding a feedback resistant NAGK enzyme.

3.5.7 Resistance to L-canavanine

As L-canavanine is competitively included in host proteins in the presence of L-arginine, we can expect cells that overproduce L-arginine to display an increased tolerance to L-canavanine. As part of this work, strains that produced the highest levels of L-arginine consequently displayed higher levels of L-canavanine resistance. Similarly, strains that produced only slightly improved levels of L-arginine, such as SCOR, grew relatively weakly compared to the DCO mutants yet still displayed increased L-canavanine tolerance relative to the WT. Indeed, similar to that previously observed for *L. plantarum*, growth by *C. glutamicum* ATCC 13032 cells was completely inhibited by 25 µg ml^{-1} of L-canavanine whereas, upon the inactivation of each strain's respective *argR* genes, the approximate minimum inhibitory concentration of L-canavanine for each increased to at least 100 µg ml^{-1} (Nicoloff *et al.*, 2004).

3.6 Concluding remarks

In this work, we used a range of techniques to assess the physiological effects of the different *argR* and *argH* mutations introduced into the L-arginine biosynthetic gene cluster of different *C. glutamicum* mutant strains in Chapter 2. For example, by using RT-PCR, the L-arginine biosynthetic genes were demonstrated to be transcribed as a single *argCJBDFRGH* transcript although this was in contrast to the results reported by Hänßler *et al.* (2007). In light of this, there is a clear need for the true transcriptional organisation of these genes to be further elucidated using a technique such as Northern hybridisation. Also via RT-PCR, we demonstrated that the transcription of the L-arginine biosynthetic genes is deregulated in the absence of an intact copy of the *argR* gene. This, in conjunction with the increased levels of L-arginine displayed by such a strain, indicates that, as observed in similar bacteria, ArgR is a repressor of L-arginine biosynthesis in *C. glutamicum* ATCC 13032. Importantly, in contrast to the mutation of the *argH* gene, this deletion of the *argR* gene in *C. glutamicum* ATCC 13032 did not negatively impact upon its growth characteristics. It should be noted here that the approach used to analyse the transcription of these L-arginine biosynthetic genes is limited and, in order to properly quantitate the change in expression experienced by the deregulated mutant, a technique such as qPCR could be adopted in future. Interestingly, we also found L-citrulline to accumulate to significant levels in these *argR* mutants and thus identified a potential metabolic bottleneck, possibly limiting the production of more L-arginine in these mutants. The adoption of a system-wide metabolomics approach could further identify other regulatory mechanisms or rate limiting steps within these deregulated strains that may be targeted as part of future strain improvement programmes. In the next chapter, further mechanisms, by which L-arginine production in this bacterium can be enhanced, such as through the manipulation of amino acid export, shall be investigated. This will be performed in conjunction with an analysis of DNA-binding specificity of the ArgR protein in *C. glutamicum* ATCC 13032 cells.

Chapter Four

Characterisation of ArgR DNA-binding activity and *lysE* overexpression in *Corynebacterium glutamicum* ATCC 13032

Contents

Chapter 4

4.1 Summary

In the previous chapter, we demonstrated transcriptional repression of the L-arginine biosynthetic genes to be alleviated in DCOR-D3 cells grown in the presence of L-arginine. This suggests that ArgR is the main transcriptional regulator of these genes. In this chapter, a recombinant, hexahistidine-tagged form of the *Corynebacterium glutamicum* ATCC 13032 ArgR protein, designated $ArgR^{His}$, was shown to bind *in vitro* to the promoter regions of the *argC* and *carA* genes with similar dissociation constants ranging between 200-250 nM. In contrast, $ArgR^{His}$ was found to not bind to the *argG* promoter region. Binding to this latter promoter region was shown to be dependent on the inclusion of 10 mM of either L-arginine or L-canavanine in the polyacrylamide gel matrix. A series of double stranded competitor oligonucleotides were also tested for their ability to prevent $ArgR^{His}$ binding to labelled DNA. For the *argC* promoter region, competitor oligonucleotides in possession of two similar motifs each resembling the *Escherichia coli* K12 ARG box consensus sequence and separated by a single base pair resulted in a large quantity of free probe to remain unshifted by $ArgR^{His}$. A similar effect was observed for the *carA* promoter region upon the inclusion of an oligonucleotide in possession of a single motif resembling the *Escherichia coli* K12 ARG box consensus sequence in the binding reaction, where it was shown to result in the partial release of free probe by $ArgR^{His}$. The consensus sequence for these three *C. glutamicum* ATCC 13032 ARG box motifs is 5'-HMT GMA TSW ADW WTW TDY-3' and the core sequence (underlined here) was found to be particularly well conserved throughout the *C. glutamicum* ATCC 13032 genome, where it was detected to precede many diverse targets of putative ArgR-mediated regulation in this bacterium. Finally, increased dosage of either the *CAF21404-argC* intergenic region or the *lysE* region was shown to result in increased L-arginine production in *C. glutamicum* ATCC 13032 cells by approximately 2-fold. Whereas the latter result is indicative of this region containing ARG box motifs, increased intracellular *lysE* expression clearly represents another strategy for further stimulating L-arginine biosynthesis by *C. glutamicum* ATCC 13032 cells.

4.2 Introduction

In prokaryotes, the genetic regulation of L-arginine metabolism differs from that typically found for the other amino acids in that, instead of relying on transcriptional attenuation, regulation is instead usually exerted via the DNA-binding action of an ArgR-type transcriptional regulator (Merino and Yanofsky, 2005). As detailed in Section 1.5.2, this interaction is normally L-arginine-dependent and occurs where defined weakly palindromic 18 bp sequence motifs, known as ARG box motifs, are situated (Maas, 1994). Indeed, although there are some subtle differences, this mechanism is thought to be unusually well conserved in spite of the diverse organisation of the L-arginine biosynthetic genes and has, for example, been demonstrated across a wide variety of bacteria, including *E. coli*, *Geobacillus stearothermophilus* and *Thermotoga maritima* (Maas, 1994; Morin *et al.*, 2003; Ni *et al.*, 1999).

ARG boxes have previously been demonstrated to precede genes usually involved in the L-arginine biosynthetic and catabolic pathways and recently some evidence of a broader regulatory role for ArgR has started to emerge in a variety of bacteria (Cunin *et al.*, 1986; Maas, 1994). For example, the *L. lactis* and *E. coli* ArgRs have been demonstrated to bind *in vitro* to the promoters of genes that encode amino acid export systems in each of these bacteria (Caldara *et al.*, 2007; Larsen *et al.*, 2004). The existence of putative ARG box motifs not directly involved in L-arginine metabolism, however, remains relatively unexplored.

The work conducted in Chapter 3, together with a recent study by Ikeda *et al.* (2009), are indirectly suggestive of the *C. glutamicum* ATCC 13032 ArgR binding to the L-arginine biosynthetic genes. This has not, however, been demonstrated directly in this bacterium and now, using electrophoretic mobility shift assays (EMSAs) in conjunction with various competitor oligonucleotides and putative effector molecules, we attempted to characterise the role of ArgR in *C. glutamicum* ATCC 13032. Additionally, we investigated the feasibility of further improving L-arginine biosynthesis through the heightened expression of LysE exporter system, in addition to an *in silico* analysis of putative ArgR binding sites in this bacterium.

4.3 Materials and Methods

4.3.1 Bacterial strains, media and growth kinetics

The *C. glutamicum* ATCC 13032 cells and *E. coli* strains used in this work were grown as described in Section 2.3.2. CGXII minimal medium (MM), Luria-Bertani (LB) medium and 2 × Yeast-Tryptone (2 × YT) medium were prepared as discussed elsewhere (Appendix B) Sambrook *et al.*, 2001). Antibiotics were used for plasmid selection as described in Section 2.3.2 and broth culture growth was monitored as outlined in Section 3.3.2.

4.3.2 DNA manipulation and amplification

DNA was prepared and manipulated as described in Section 2.3.3. DNA fragments were amplified using the polymerase chain reaction (PCR) as discussed earlier in Section 2.3.4. The thermal cycling conditions and oligonucleotide primers used for each amplification reaction are listed in Appendix C.

4.3.3 Construction of pLYSE and pCAF21404-ARGC, DNA sequencing and electroporation

In order to construct pLYSE, the *C. glutamicum* ATCC 13032 *lysE* gene (Genbank accession number NC_006958), was amplified from *C. glutamicum* ATCC 13032 genomic DNA by a PCR performed with the *lysE* gene primers, purified, ligated into pTZ57R/T and sequenced. An approximately 800 bp *Bam*HI-*Xba*I fragment, containing an intact copy of the *lysE* gene, was subsequently excised from this vector and subcloned into *Bam*HI-*Xba*I digested pUC19. A DNA fragment, containing *lysE* was next excised from this plasmid using *Bam*HI and *Pst*I and inserted into *Bam*HI-*Pst*I digested pEKEX2, which harbours the IPTG-inducible *tac* promoter. The resulting plasmid was designated pLYSE. For the construction of pCAF21404-ARGC, a 115 bp intergenic region preceding the *C. glutamicum* ATCC 13032 *argC* gene was amplified via PCR using the *CAF21404-argC* fragment primer pair as part of a larger 491 bp fragment which, following purification, was ligated into pTZ57R/T (Fermentas) and sequenced (Section 2.3.5). This 491 bp fragment, which includes approximately 300 bp of the 3'-region of *CAF21404* and 80 bp of the 5'-region of *argC*, in addition to the *CAF21404-argC* intergenic region, was next excised from pTZ57R/T with *Bam*HI-*Eco*RI and cloned into *Bam*HI-*Eco*RI digested pEKEX2, thus forming pCAF21404-

ARGC. Restriction analysis was used to confirm the identity of pLYSE and pCAF21404-ARGC. These plasmids, together with pEKEX2, were now individually electroporated into competent *C. glutamicum* ATCC 13032 cells using the protocol described in Appendix D.

4.3.4 L-Arginine assays

C. glutamicum ATCC 13032 cells harbouring pCAF21404-ARGC, pEKEX2 or pLYSE were grown in CGXII MM broth as described previously (Section 3.3.7) in order to allow for the accumulation of measurable quantities of L-arginine. In the case of cells containing pLYSE or the control plasmid, pEKEX2, IPTG at a final concentration of 1 mM was added to each culture at the beginning of the incubation period so as to induce transcription of the plasmid-borne *lysE* genes off the *tac* promoter. After incubation, cell-free extracts of each strain were prepared as outlined in Section 3.3.7. In order to measure extracellular L-arginine accumulation by *C. glutamicum* ATCC 13032 cells harbouring either pEKEX2 or pLYSE, the 1.5 ml samples collected at each time point were immediately centrifuged (15 000 × *g*, 5 min, 4 °C) and the supernatant collected. After the removal of soluble proteins within both these extracellular samples and the total L-arginine samples via trichloroacetic acid treatment, the amount of L-arginine in each sample was quantitated colorimetrically as described in Section 3.3.8.1.

4.3.5 Expression, purification and detection of recombinant ArgR^{His}

4.3.5.1 Construction of pARGR-P

With the view of creating a construct for the inducible overproduction of ArgR, the *argR* gene (Genbank accession number YP_225686) was amplified from *C. glutamicum* ATCC 13032 genomic DNA by a PCR performed with the *argR*-P primer pair (Appendix C). As confirmed by DNA sequencing, the resultant 346 bp product fragment contained several oligonucleotide-introduced features. These included a 5'-*Bam*HI restriction site upstream of the *argR* start codon, a 3'-*Eco*RI restriction site downstream of the *argR* coding region and a point mutation within the stop codon causing it to now encode an L-histidine residue. This fragment was subsequently digested with *Bam*HI and *Eco*RI, ligated into pET22b(+) (Novagen) digested with the same enzymes and finally transformed into *E. coli* BL21 (DE3) cells. The resulting construct, designated pARGR-P, encodes ArgR in possession of both a

C-terminal hexahistidine fusion and an N-terminal PelB leader sequence. The hexahistidine tag permits for the affinity chromatography-based purification of this protein and the PelB sequence can allow for the potential periplasmic localisation of recombinant protein in *E. coli* BL21 (DE3) cells (Georgiou and Segatorei, 2005; Wolfe *et al.*, 1983).

4.3.5.2 Overexpression of ArgR^{His}

The overproduction of ArgR^{His} was induced as outlined in the pET System Manual (Novagen) with various modifications included so as to limit the formation of insoluble inclusion bodies. Firstly, a 5 ml seed culture of *E. coli* BL21 (DE3) cells containing pARGR-P was grown for 16 h in LB broth and was subsequently used to subinoculate 100 ml of LB broth. This culture was grown to an OD_{600} of 0.4-0.6 in a baffled flask before IPTG was added at a final concentration of 0.5 mM in order to induce ArgR^{His} expression. It was next incubated with vigorous orbital shaking (150 rpm) at 30 °C for 2 h to allow for protein expression. To permit for the potential partial resolubilsation of any insoluble protein aggregates, these cells were pelleted (10 000 × *g*, 5 min, 4 °C), washed twice with LB broth in and resuspended in LB broth without IPTG, after which they were incubated at 30 °C for 4 h with orbital shaking and subsequently harvested for protein extraction by centrifugation (10 000 × *g*, 5 min, 4 °C) (Carrió and Villaverde, 2001).

4.3.5.3 Extraction and purification of ArgR^{His}

Pelleted *E. coli* BL21 (DE3) cells were resuspended in 1.5 ml of 50 mM Tris-HCl (pH 7.5) containing 1 mM PMSF (Merck) and were mechanically disrupted by 2.5 min of sonication using a Virsonic Digital 600 sonicator (Virtis) and three alternating pulsing (30s) and cooling (20s) cycles. The crude lysates were next centrifuged (15 000 × *g*, 2 min, 4 °C) and the soluble fraction collected and eluted through a HIS-Select nickel affinity agarose column (Sigma-Aldrich). Bound cell material on the column was rinsed twice with wash buffer (20 mM imidazole, 0.4 mM NaCl [pH 8], 50 mM $NaPO_4$, 1 mM PMSF, 0.1% [w/v] Triton X-100) and any remaining compounds were eluted using buffer containing 250 mM imidazole, 0.4 M NaCl, 100 mM $NaPO_4$ (pH 8.0) and 25% (w/v) glycerol. The imidazole was removed from the purified protein mix by exchanging the elutant against buffer containing 0.2 M NaCl, 30 mM Tris-HCl (pH 8.0), 25% (w/v) glycerol, 2 mM DTT and 1 mM PMSF using an Amicon Ultra-15 filter unit (Millipore) with a molecular weight cut off of 10 kDa.

The volume of the final elutant was approximately 300 µl and purified protein samples of $ArgR^{His}$ were subsequently stored in single-use aliquots at -80 °C. Where relevant, molar amounts of $ArgR^{His}$ are given in monomeric quantities.

4.3.5.4 Protein quantitation, separation and detection

The total protein content of crude cell lysates and purified $ArgP^{His}$ aliquots was quantitated according to the method of Bradford (1976) using Coomassie Brilliant Blue G-250 dye (Bio-rad) and BSA (Fermentas) for the construction of a calibration curve. Prior to separation, protein samples were denatured in the presence of an equal volume of loading buffer (125 mM Tris-HCl [pH 6.8], 2.5% [w/v] SDS, 20% [v/v] glycerol, 0.2% [w/v] bromophenol blue) by heating at 100 °C for 5 min. Proteins were then separated according to molecular weight by SDS-polyacrylamide (37:1) gel electrophoresis (PAGE) using a Mini-PROTEAN Electrophoresis System (Bio-rad) with a 4% stacking gel and a 15% resolving gel as described by Laemmli (1970). The prestained PageRuler protein molecular weight marker, available from Fermentas, was used to determine the apparent molecular weights of the separated proteins. Hexahistidine-tagged proteins were detected according to the standard Western blot method using an anti-6-His rabbit antibody (1:5000 dilution) and a horse radish peroxidise-conjugated anti-rabbit goat secondary antibody (1:5000 dilution) in conjunction with the chromogenic TMB membrane peroxidase substrate (all manufactured by GeneTex) (Sambrook *et al.*, 2001).

4.3.6 Electrophoretic mobility shift assays

4.3.6.1 Oligonucleotide labelling

Before DNA-binding by $ArgR^{His}$ could be visualised, each target fragment was terminally labelled with a 3'-DIG molecule. The *argC* promoter, *argG* promoter, *carA* promoter and *narKGHJI* primer pairs were each used for the PCR-based amplification of their relevant intergenic target regions from *C. glutamicum* ATCC 13032 genomic DNA as described in Appendix C. Following the gel purification of each, 10 µl of each fragment was resuspended in TEN buffer (10 mM Tris-HCl, 1 mM EDTA, 0.1 M NaCl, pH 8) (typically containing 2.0-5.0 pmol of each fragment), heat-treated at 95 °C for 10 min and cooled to room temperature. This solution was subsequently added to the labelling reaction mix at 4 °C

and containing 4 µl of 5 × labelling buffer (1 M potassium cacodylate, 0.125 M Tris-HCl, 1.25 mg ml^{-1} BSA, pH 6.6), 4 µl of 25 mM $CoCl_2$, 1 µl of 1.0 mM DIG-ddUTP and 20 U of terminal transferase (all reagents supplied as part of the DIG 2nd Generation Gel Shift Kit). The labelling reaction was allowed to proceed at 37 °C for 10 min before 2 µl of 0.2 M EDTA (pH 8.0) was added to stop the reaction. For each reaction, the efficiency of labelling was assessed using the standard DIG-CSPD chemiluminescence protocol outlined in the DIG Application Manual for Filter Hybridisation (Roche) and visualised as in Section 2.3.6.2. As little as 1 fmol of labelled dsDNA was detectable using this system and each labelled fragment was subsequently diluted to a concentration of 20 fmol $µl^{-1}$ in TEN buffer prior to incubation in the presence of $ArgR^{His}$.

4.3.6.2 Binding reactions

Protein DNA-binding reactions were prepared on ice and included 2 µl of 5 × labelling buffer (100 mM HEPES [pH 7.6], 5 mM EDTA, 50 mM $(NH_4)_2SO_4$, 5 mM DTT, 1% [w/v] Tween 20, 150 mM KCl). In order to reduce non-specific DNA-protein interactions, 0.5 µl of a poly-(d[I-C]) solution (1 µg $µl^{-1}$) was added to the reaction mix together with 0.5 µl of a poly-L-lysine solution (1 µg $µl^{-1}$), which has previously been shown to enhance the stability of various specific protein-DNA interactions (Kozmík *et al.*, 1990). Unless otherwise stated 10 mM of L-arginine was included in the binding reaction as an effector molecule. Typically 20 fmol of the DIG-labelled target DNA was included in the reaction together with 5 pmol of purified $ArgR^{His}$ protein. The reaction mix was lastly brought to a volume of 10 µl using dH_2O and incubated at room temperature for 30 min. Prior to the separation of unbound DNA and protein-DNA complexes by PAGE, 2.5 µl of loading buffer (0.25 × TBE buffer [22.25 mM Tris-borate, 0.5 mM EDTA], 60% [w/v] glycerol, 0.2% [w/v] bromophenol blue) was added to each reaction. Each of these reagents was supplied as part of the DIG 2nd Generation Gel Shift Kit. In order to calculate the apparent dissociation constant (K_d; the concentration of protein at which 50% of the free probe is shifted) of $ArgR^{His}$ binding to different promoter regions, the method of Sekino *et al.* (2000) was adopted. Band intensities were quantified using the open source ImageJ software (version 1.40g; United States National Institute of Health). For binding reactions involving competitor oligonucleotides, unlabelled dsDNA fragments were included at a 1500-fold molar excess relative to the DIG-labelled target fragment. These oligonucleotides were synthesised by Inqaba Biotechnical

Industries and are listed in Appendix F. Prior to their inclusion in each reaction, complementary oligonucleotide pairs were mixed in TEN buffer and heat denatured at 95 °C for 5 min, after which they were cooled to room temperature at rate of 1 °C per minute using a thermal cycler.

4.3.6.3 Polyacrylamide gel electrophoresis

Protein DNA-binding reactions were analysed by electrophoresis through a native 8% (w/v) polyacrylamide gel (29:1; Sigma-Aldrich) in 0.25 × TBE buffer at 80 V using a Mini-PROTEAN Electrophoresis System at 4 °C. Prior to loading, gels were pre-electrophoresed for 15 min. Unless otherwise stated, 10 mM of L-arginine was included in the polyacrylamide solution. Gels were typically electrophoresed for approximately 3 h until the loading dye had run off and were thereafter transferred to a Hybond-N^+ nylon membrane (Amersham Biosciences) via electroblotting as described in the DIG 2nd Generation Gel Shift Kit manual. Transferred labelled nucleic acid was finally visualised using the chemiluminescent CSPD substrate (Roche) as outlined in Section 2.3.6.2 and the DIG Application Manual for Filter Hybridisation.

4.3.7 Primer extension

A primer extension approach was adopted in order to determine the putative transcriptional start site of the *argC* gene. An unlabelled 20 bp oligonucleotide primer (5'-GGC TCC TGC GAT TGC AAC CT-3'), designed to anneal 40 bp downstream of the putative *argC* start codon, was initially tested for non-specific cross-hybridisation via PCR using *C. glutamicum* ATCC 13032 L-arginine biosynthetic gene cluster as template DNA. The reaction products were subsequently sequenced and the results found to indicate that here this primer had annealed specifically. This oligonucleotide was resynthesised with a 5'-Cy5 molecule by Pei-Yin Liebrich (Department of Molecular and Cell Biology, University of Cape Town, South Africa). Approximately 40 µg total RNA, extracted from *C. glutamicum* ATCC 13032 cells grown in CGXII MM as outlined in Section 3.3.5, was now denatured in a 10 µl volume of dH_2O at 95 °C for 10 min and rapidly cooled on ice. Fluorescently-labelled primer (approximately 5 pmol) was now added to this RNA solution together with 1 µl of RiboLock ribonuclease inhibitor (Fermentas) (40 U μl^{-1}) and the resultant mixture was incubated at 45 °C for 16 h in the dark in order to allow for the

primer to anneal. After this incubation period, the following reagents supplied by Fermentas were added: 4 µl of 5 × reverse transcription (RT) buffer (250 mM Tris-HCl [pH 8.3], 250 mM KCl, 20 mM $MgCl_2$, 50 mM DTT), 1.5 µl of a 10 mM dNTP mix, 2 µl of RiboLock ribonuclease inhibitor (40 U μl^{-1}) and 2 µl of avian myeloblastosis virus RT (20 U μl^{-1}). When required, 1.5 M betaine (Sigma-Aldrich) was included in the reaction in order to limit secondary structure formation (Henke *et al.*, 1997). RT was allowed to proceed for 16 h at 50 °C before the termination of the reaction through the addition of 10 µl of 1 M NaOH and 10 µl of 0.5 M EDTA and a 15 min period of incubation at 65 °C. A 10 µl volume of 1 M HCl was finally added to the reaction mix and the RNeasy MinElute Cleanup Kit (Qiagen) was used to purify cDNA fragments according to the manufacturer's instructions. Solutions of cDNA were hereafter concentrated down to approximately 8 µl using a Speed Vac Concentrator System (Savant) and stored in the dark at -80 °C. The fluorescently-labelled reaction products were now analysed using an ALFexpress Automated DNA Sequencer (Pharmacia Biotech) together with reaction products obtained from a sequencing reaction using the same primer. The primer extension data was processed using the ALFwin Sequence Analyser software (version 2.10; Pharmacia Biotech).

4.3.8 Bioinformatics

C. glutamicum ATCC 13032 nucleotide sequences were retrieved from the genome sequence of this organism (Genbank accession number NC_006958) stored on the National Centre for Bioinformatics web server (http://www.ncbi.nlm.nih.gov). Sequences similar to the *E. coli* K12 ARG box consensus sequence were detected using a position weight matrix (PWM) of this sequence retrieved from the Prodoric database (http://prodoric.tu-bs.de/) and the Virtual Footprint software (version 3.0; Technical University Braunschweig) (available online from http://prodoric.tu-bs.de/vfp/index2.php) (Münch *et al.*, 2003; Münch *et al.*, 2005). Nucleotide alignments were constructed using the AlignX module of the Vector NTI Advance software suite (version 10.1.1; Invitrogen) (Lu and Moriyama, 2004). Sequence logo diagrams were created using the online WebLogo tool (version 3.0; University of California) available from http://weblogo.berkeley.edu/ (Crookes *et al.*, 2004).

4.3.9 Statistical analysis

The statistical methods used to analyse and present data are described in Section 3.3.11.

4.4 Results

4.4.1 Effect of *lysE* dosage on L-arginine production

If the rate of L-arginine export from *C. glutamicum* ATCC 13032 cells may be enhanced by, for example, the overexpression of *lysE*, we can expect to see improved L-arginine biosynthesis in this bacterium as this would result in reduced intracellular levels of L-arginine and thereby cause less ArgR to be present in the activated, L-arginine-bound form. As evident from Figure 4.1B, *C. glutamicum* ATCC 13032 cells containing increased copies of *lysE* produce more L-arginine relative to *C. glutamicum* ATCC 13032 (pEKEX2) cells. For example, as shown in Section 4.4.1, the total L-arginine levels produced by the latter strain remained relatively low throughout the incubation period (peaking at 3.40 ± 0.28 µmol [mg dry cell mass]$^{-1}$), whereas *C. glutamicum* ATCC 13032 (pLYSE) cells displayed significantly higher L-arginine concentrations of 3.88 ± 0.32 ($P < 0.01$), 5.4 ± 0.09 ($P < 0.005$) and 5.2 ± 0.40 µmol (mg dry cell mass)$^{-1}$ ($P < 0.05$) after 3, 6 and 9 h of incubation respectively. Furthermore, as shown in Figure 4.1C, increased dosage of the *lysE* gene in *C. glutamicum* ATCC 13032 cells significantly improved the recovery of L-arginine from the supernatant, indicating that the increased quantities of L-arginine produced by cells harbouring pLYSE are actively exported by the overexpressed LysE protein. For instance, significantly higher levels of extracellular L-arginine were detected in the supernatant following the incubation of *C. glutamicum* ATCC 13032 cells that harbour pLYSE relative to cells in possession of pEKEX2 after 9 h (6.02 ± 0.85 compared to 2.85 ± 0.07 µmol [mg dry cell mass]$^{-1}$ [$P < 0.05$]) and 24 h (5.29 ± 0.36 compared to 3.51 ± 0.03 µmol [mg dry cell mass]$^{-1}$ [$P < 0.05$]) of incubation. Indeed, the total L-arginine content extracted from cells in possession of pLYSE reached comparatively similar values to that observed in the cell-free supernatant isolated from this same strain after 9 h and 24 h of incubation (with both reaching L-arginine levels of between 5.22-6.22 µmol [mg dry cell mass]$^{-1}$). Most of the L-arginine produced by *C. glutamicum* ATCC 13032 (pLYSE) cells thus appears to be located in the supernatant at these time points, suggesting that here the intracellular L-arginine content within these cells is either negligible or below the detection limit of this colorimetric assay.

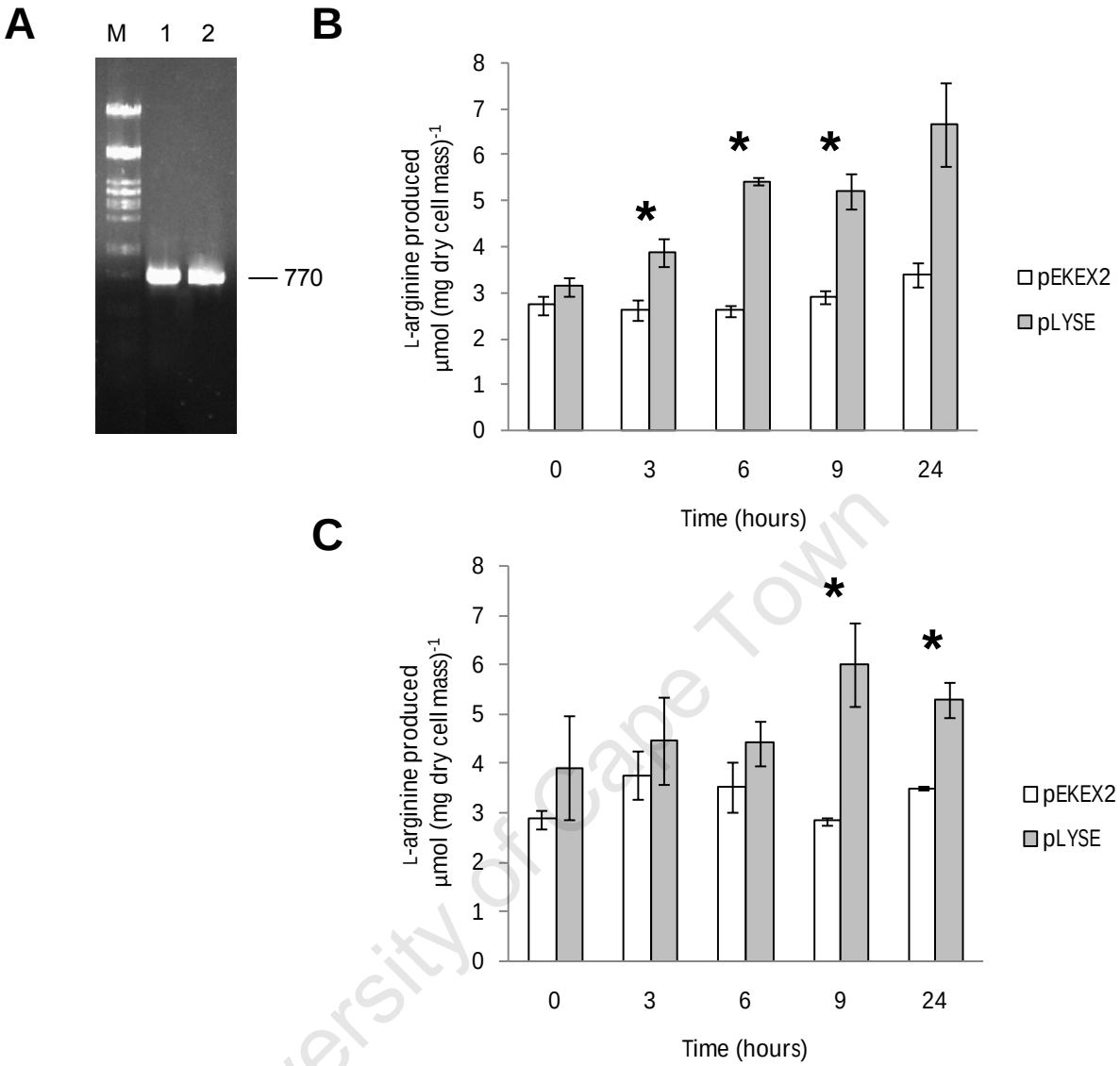


Figure 4.1 The effect of increased dosage of the *lysE* gene on L-arginine accumulation by *C. glutamicum* ATCC 13032 cells. **(A)** PCR amplification of the 770 bp *lysE* gene using either *C. glutamicum* ATCC 13032 genomic DNA (lane 1) or pLYSE (lane 2) as template DNA. Lane M: marker DNA. Fragment sizes are indicated in bp. Total **(B)** or extracellular **(C)** L-arginine produced by *C. glutamicum* ATCC 13032 cells in possession of pLYSE or pEKEX2 is shown. Asterisks represent significantly different data points.

4.4.2 Effect of pCAF21404-ARGC on L-arginine production

In several bacteria, including, *L. lactis* and *Streptomyces clavuligerus*, the ArgR repressor protein has been shown to bind to intergenic DNA motifs located directly upstream of

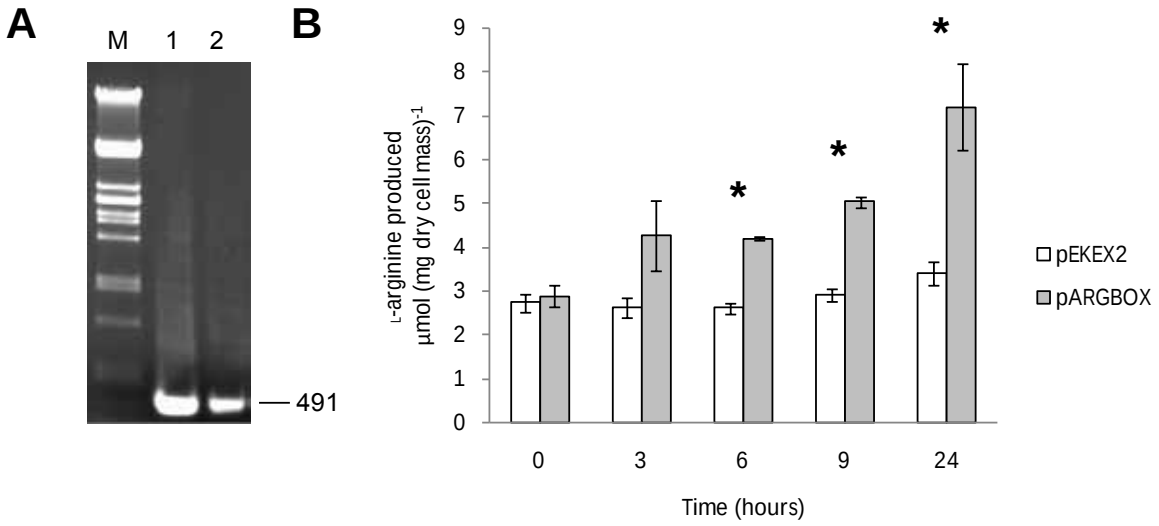


Figure 4.2 The effect of increased copy number of the *CAF21404-argC* fragment on L-arginine production in *C. glutamicum* ATCC 13032 cells. **(A)** PCR amplification of the 491 bp *CAF21404-argC* intergenic fragment from *C. glutamicum* ATCC 13032 genomic DNA (lane 1) and pARGBOX (lane 2). Lane M: marker DNA. Fragment sizes are indicated in bp. **(B)** *C. glutamicum* ATCC 13032 cells harbouring either pEKEX2 or pARGBOX were incubated in CGXII MM for 24 h and the total L-arginine levels assayed throughout this period. Asterisks represent significantly different data points.

the *argC* gene and down-regulate the operon (Larsen *et al.*, 2005; Rodríguez-García *et al.*, 1997). By analogy with these other Gram-positive bacteria, we can expect a similar phenomenon to occur within *C. glutamicum* ATCC 13032 cells. If this is the case, the introduction of additional copies of the *CAF21404-argC* intergenic region into this bacterium should lead to the alleviation of transcriptional repression of the L-arginine biosynthetic genes through the titration of ArgR molecules away from the L-arginine biosynthetic regulon. In order to gather evidence of sequence motifs within this region that the *C. glutamicum* ATCC 13032 ArgR is capable of binding, pCAF21404-ARGC, which harbours the *CAF21404-argC* intergenic region on a DNA fragment inserted into pEKEX2, was introduced into these cells and the formation of L-arginine throughout their incubation in CGXII MM broth was monitored. As expected, *C. glutamicum* ATCC 13032 (pCAF21404-ARGC) cells displayed significantly elevated levels of total L-arginine relative to *C. glutamicum* ATCC 13032 (pEKEX2) cells (Figure 4.2B). For example, the amount of L-arginine produced by the latter variety of cells increased significantly throughout the incubation period, reaching levels of approximately 4.20 ± 0.03 ($P < 0.005$), 5.03 ± 0.13 ($P <$

0.0005) and 7.21 ± 0.99 µmol (mg dry cell mass)$^{-1}$ ($P < 0.05$) after 6, 9 and 24 h of incubation respectively, whereas *C. glutamicum* ATCC 13032 (pEKEX2) cells maintained relatively unchanged total L-arginine levels throughout the 24 h incubation period (ranging between 2.61 ± 0.13 and 3.40 ± 0.28 µmol [mg dry cell mass]$^{-1}$). These data clearly suggest that the *CAF21404-argC* fragment harboured in pCAF21404-ARGC is capable of alleviating repression of the L-arginine biosynthetic genes and, in order establish whether or not ArgR binds specifically to the *CAF21404-argC* intergenic region, EMSAs were hereafter conducted using recombinant ArgR^{His}.

4.4.3 Production and purification of ArgR^{His}

In order to study the DNA-binding activity of the *C. glutamicum* ATCC 13032 ArgR protein, a C-terminally hexahistidine-tagged derivative of it, designated ArgR^{His}, was overexpressed in *E. coli* BL21 (DE3) cells. Generally, very large amounts of the expressed recombinant protein appeared to be present in the insoluble fraction and thus unsuited for functional studies. By altering the conditions under which this protein was expressed, however, we were able to shift large quantities of ArgR^{His} into the soluble fraction, from which sufficient amounts could subsequently be isolated by affinity chromatography (Figure 4.3B). A large amount of protein with a molecular weight of approximately 21 kDa was found to be expressed by induced cells in possession of pARGR-P. Interestingly, upon the purification of hexahistidine-tagged protein to apparent homogeneity via affinity chromatography, a slightly larger form of protein (approximately 24 kDa) appeared to have been co-purified in approximately equal quantities with the 21 kDa protein (Figure 4.3C). The DNA-binding activity of these hexahistidine-tagged proteins for different *C. glutamicum* ATCC 13032 promoter regions was subsequently assessed using EMSAs.

4.4.4 Electrophoretic mobility shift assays

4.4.4.1 L-Arginine-dependent *in vitro* DNA-binding activity of ArgR^{His}

Prior to the purification of ArgR^{His}, total soluble protein isolated from IPTG-induced *E. coli* BL21 (DE3) (pARGR-P) cells was used as part of a preliminary EMSA analysis, performed in conjunction with DIG-labelled *argC* promoter DNA, in order to confirm the existence of *in vitro* DNA-binding activity within this extracted protein (Figure 4.4A). These protein

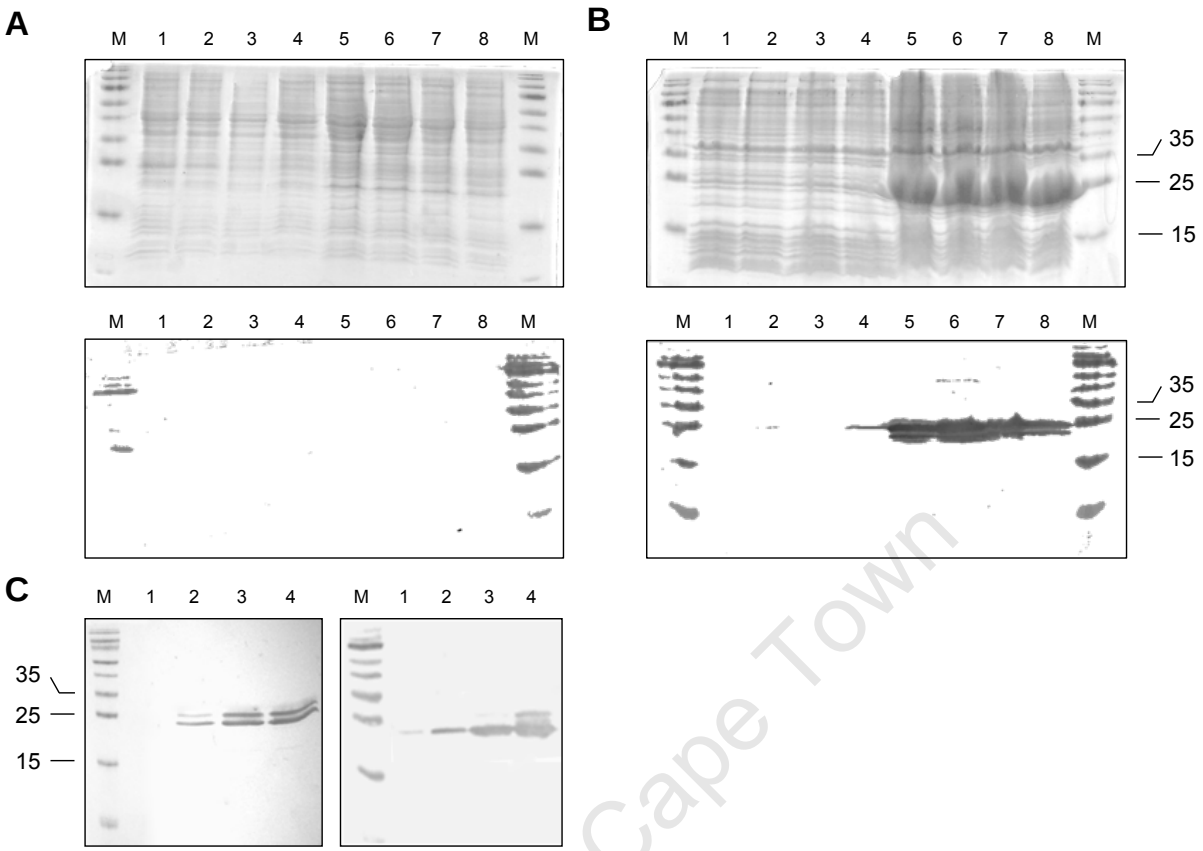


Figure 4.3 Production and purification of ArgR^{His}. Lanes M: PageRuler protein molecular weight marker. Approximate protein molecular weights are indicated in kDa. **(A and B)** Total soluble protein extracted from *E. coli* BL21 (DE3) cells containing pET22b(+) **(A)** or pARGR-P **(B)**. A Western blot analysis using an anti-6-His rabbit antibody was performed on each sample as is shown below each Coomassie-stained SDS-polyacrylamide gel. In each diagram, lanes 1-4 contain protein extracted from uninduced cells whereas lanes 5-8 contain protein extracted from induced cells. Lanes 3, 4, 7 and 8 contain samples from a replicate culture grown under identical conditions and are thus similar to lanes 1, 2, 5 and 6. Lanes 1, 3, 5 and 7 contain protein extracted from cells 2 h after the culture had reached an OD_{600} of 0.4-0.6 (i.e. in the case of the induced samples, 2 h after the addition of IPTG) and lanes 2, 4, 6 and 8 contain protein extracted from cells at the end of the second incubation period of 4 h. **(C)** Aliquots of protein extracted from *E. coli* BL21 (DE3) cells expressing pARGR-P were bound and eluted from a nickel-agarose column and separated through an SDS-polyacrylamide gel. The cognate anti-6-His Western blot is shown alongside this gel. Lanes 1-4 each contain the following quantities of protein (mg): 0.3, 1.5, 3.0 and 6.0.

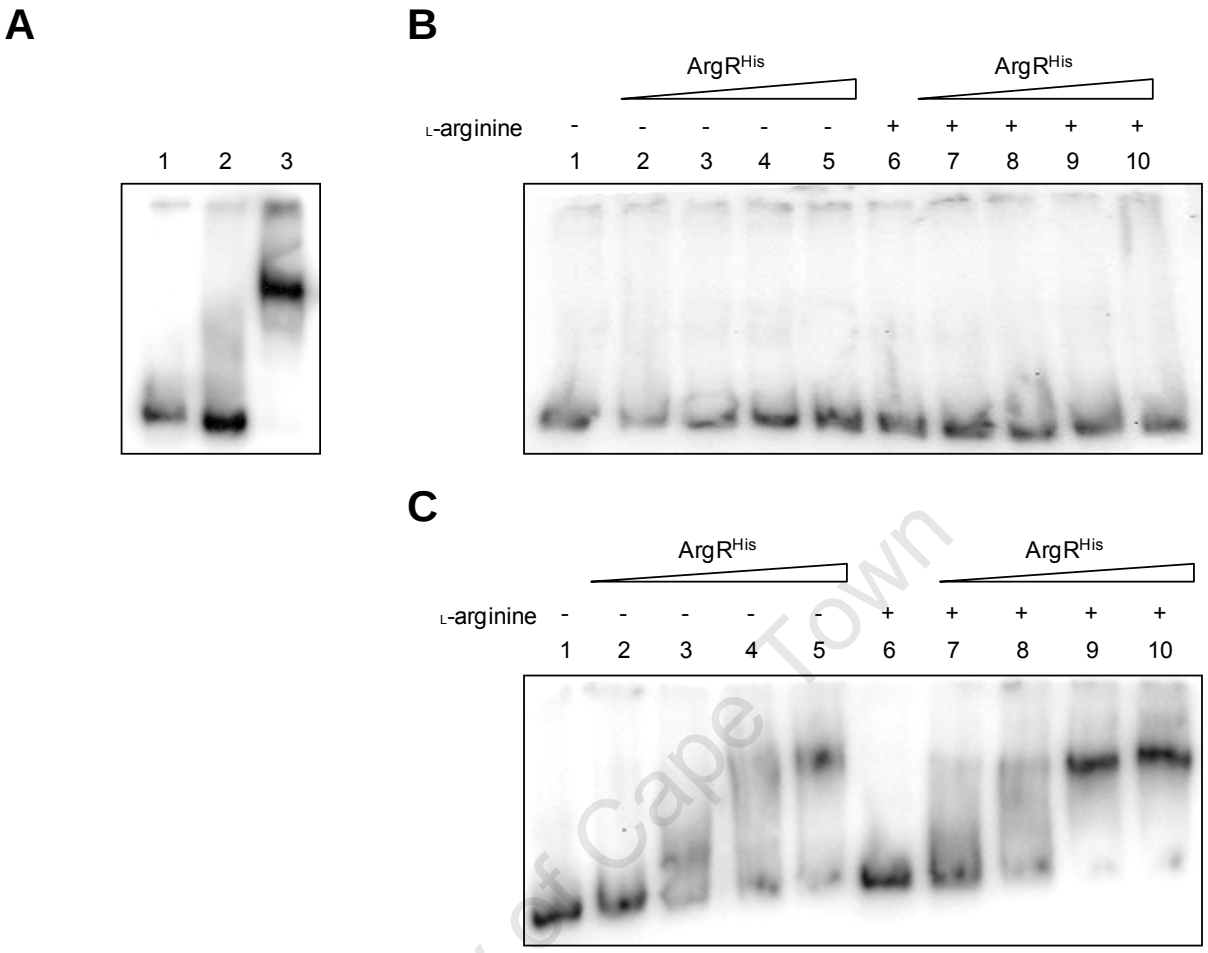


Figure 4.4 EMSA-based analysis of the *argC* promoter-binding activity of protein extracted from *E. coli* BL21 (DE3) (pARGR-P) cells and the L-arginine-dependent interaction between purified ArgR^{His} and the *argC* promoter region. **(A)** *argC* promoter-binding activity of total soluble protein extracted from induced *E. coli* BL21 (DE3) cells in possession of either pET22b(+) or pARGR-P. Lane 1 contains the products of a negative control binding reaction performed with 20 fmol of DIG-labelled *argC* promoter fragments and no protein, whereas lanes 2 and 3 contain the products of binding reactions performed using 2.0 mg of total protein extracted from cells in possession of either plasmid. L-Arginine (10 mM) was included here in the binding reactions and the polyacrylamide gel matrix. **(B and C)** The DIG-labelled *argC* promoter fragments were incubated with varying amounts of ArgR^{His} in the presence or absence of L-arginine (indicated by the plus or minus sign) for each binding reaction. The products of each were subsequently electrophoresed through polyacrylamide gels supplemented with **(C)** or without **(B)** 10 mM L-arginine. ArgR^{His} was used in the following amounts (pmol): lanes 1 and 6, 0; lanes 2 and 7, 1.09; lanes 3 and 8, 2.19; lanes 4 and 9, 4.1; lanes 5 and 10, 5.47.

extracts reliably caused a clear, distinct reduction in the electrophoretic mobility of the labelled-*argC* promoter fragment, thereby indicating the occurrence of a stable DNA-protein interaction. Moreover, this interaction appears to be due to the specific interaction of ArgR^{His} as this shift was not observed when protein extracted from cells lacking the *C. glutamicum* ATCC 13032 *argR* gene was used as part of the binding reaction. Based on the previously observed *in vitro* L-arginine dependency of this interaction in other bacteria, L-arginine was included in both the binding reactions and the polyacrylamide gels here at a concentration of 10 mM as part of this preliminary analysis, although the L-arginine-dependent nature of this interaction was investigated as part of the next experiment (Caldara *et al.*, 2007; Larsen *et al.*, 2005).

As shown in Figure 4.4B and C, the inclusion of 10 mM L-arginine within the polyacrylamide gel matrix appears to be required for the stable *in vitro* formation of the *argC* promoter-ArgR^{His} complex. Indeed, in the absence of this compound within the gel matrix, a stable complex resulting in an electrophoretic shift of the *argC* promoter DNA fragment could not be observed, even upon the inclusion of L-arginine within each binding reaction (Figure 4.4B). Additionally, as evident from Figure 4.4C, the inclusion of this compound within each binding reaction was not critical for the stable formation of this DNA-ArgR^{His} complex if L-arginine was included within the gel matrix. Interestingly, however, a greater portion of labelled fragment consistently appeared to be present in the bound form when L-arginine was included in both the reaction and the gel (this may be observed, for example, by comparing lanes 4 and 9 in Figure 4.4C). The inclusion of L-arginine at 10 mM within both the reaction and the gel therefore appeared to improve either the stability of the *argC* promoter-ArgR^{His} complex or the *argC* promoter-binding affinity of ArgR^{His} and thus this compound was included at this concentration in both the reaction mix and the polyacrylamide gel for future EMSA-based experiments.

The binding activity of ArgR^{His} to various *C. glutamicum* ATCC 13032 promoter regions expected to be involved in L-arginine biosynthesis was next assessed (Figure 4.5). The *narKGHJI* promoter region, which precedes various genes involved in nitrate metabolism and is not expected to play a role in L-arginine biosynthesis, was included as a negative control in these experiments (Nishimura *et al.*, 2008). In agreement with our earlier findings, ArgR^{His} bound strongly to the *argC* promoter region, with approximately 4.1 pmol

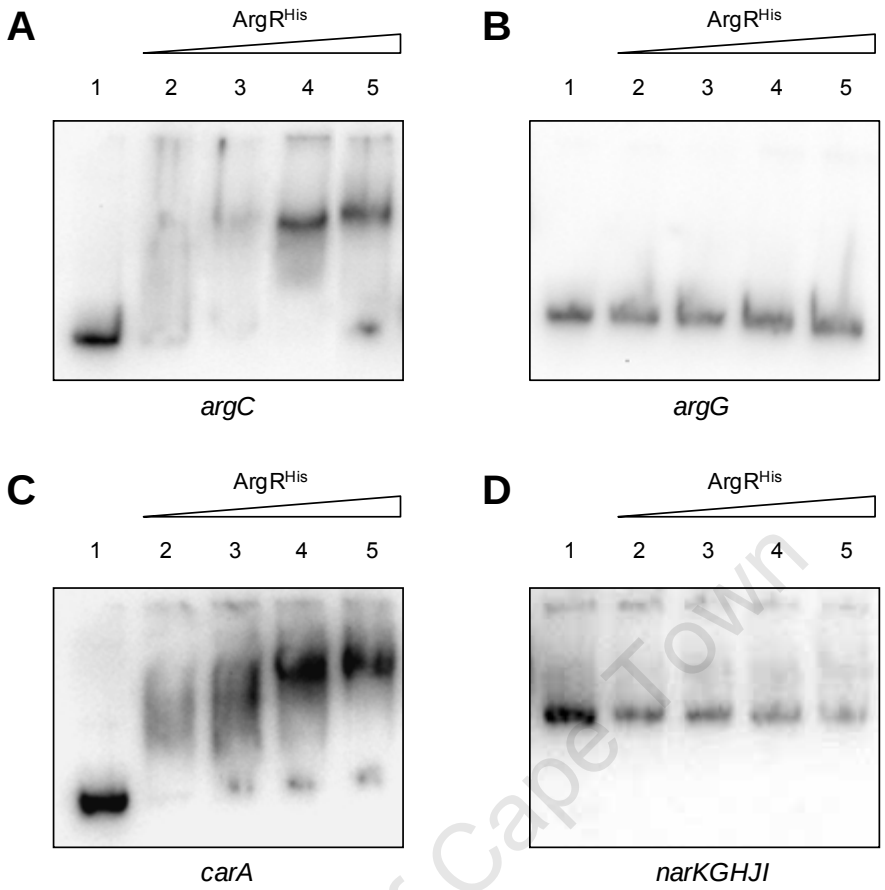


Figure 4.5 Analysis of the interaction between ArgR^{His} and various *C. glutamicum* ATCC 13032 promoter regions. Each EMSA shows the contents of a binding reaction involving 20 fmol of the relevant promoter region (indicated below each blot) and the following amounts of ArgR^{His} (pmol): lane 1, 0; lane 2, 1.09; lane 3, 2.19; lane 4, 4.1; lane 5, 5.47. Each binding reaction and polyacrylamide gel was supplemented with 10 mM L-arginine.

of ArgR^{His} appearing sufficient to fully shift 20 fmol of labelled fragment (Figure 4.5A). This protein also did not display any apparent binding activity at this concentration for the *narKGHJI* promoter region (Figure 4.5 D). Interestingly, labelled DNA fragments encoding the *argG* promoter region were not shifted by ArgR^{His}, even when relatively high amounts of protein were used (Figure 4.5B). In contrast, however, the labelled *carA* promoter DNA fragments also appeared to be strongly bound by the ArgR^{His} protein (Figure 4.5C).

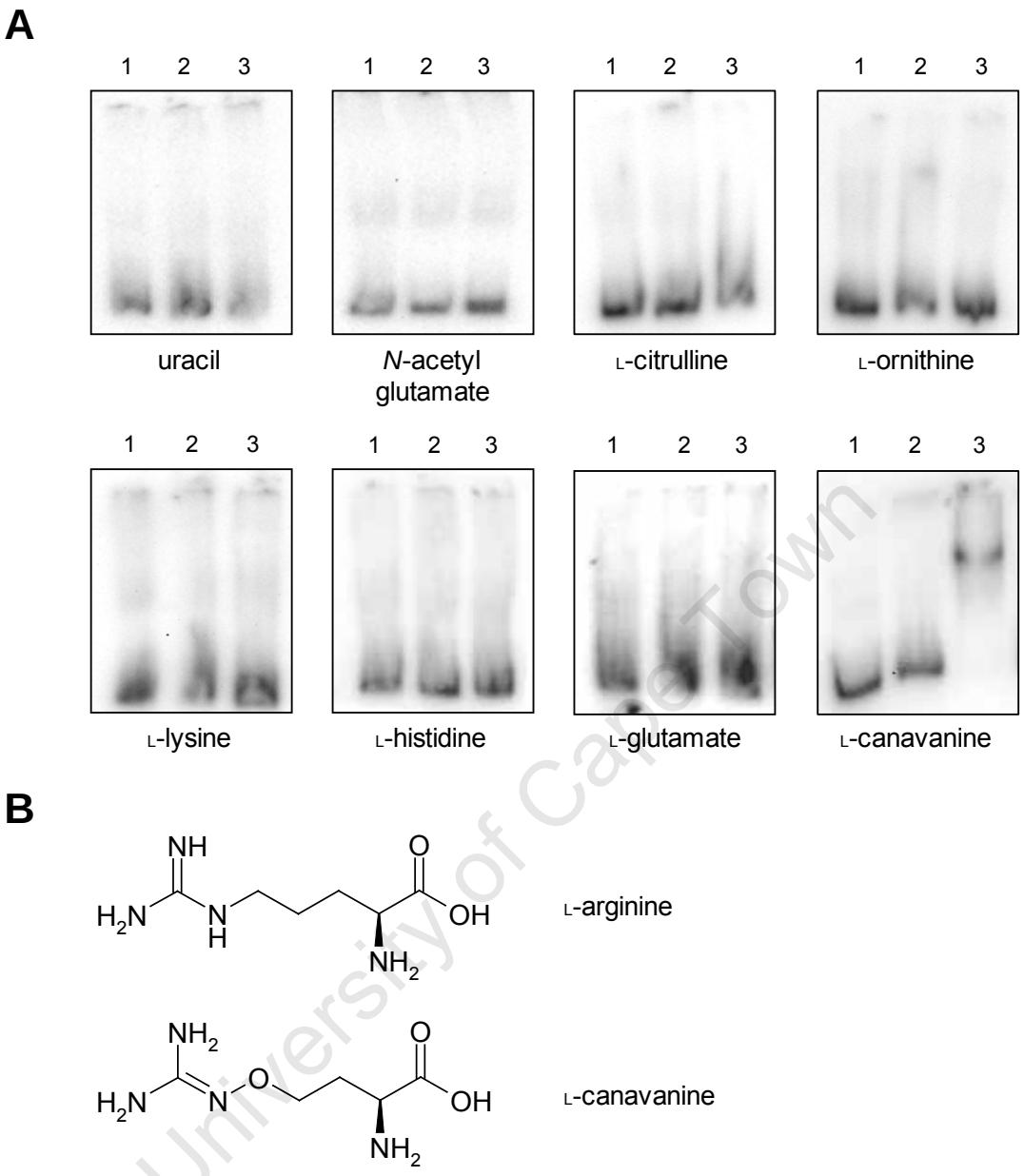


Figure 4.6 Analysis of the interaction between ArgR^{His} and the *argC* promoter region in the presence of different effector molecules. **(A)** Different types of effector molecules were included in each binding reaction and each polyacrylamide gel at a concentration of 10 mM. Each binding reaction also included 20 fmol of labelled *argC* promoter DNA and 5.47 pmol of ArgR^{His}. The lane configuration for each blot follows: lane 1, negative control binding reaction performed using no protein and no effector; lane 2, reaction performed with ArgR^{His} and no effector; lane 3, reaction performed with ArgR^{His} and the relevant effector molecule. **(B)** The molecular structures of L-arginine and L-canavanine are shown.

4.4.4.2 ArgR^{His}-*argC* promoter binding activity in the presence of various effectors

A variety of different amino acids and other metabolites that form part of the L-arginine biosynthetic pathway were individually included as effector molecules in both the binding reactions and polyacrylamide gel matrices at a concentration of 10 mM in order to examine if they are capable of activating *in vitro* ArgR^{His}-*argC* promoter binding (Figure 4.6). At this concentration, none of the compounds shown, with the exception of the L-arginine analogue L-canavanine, resulted in the stable *in vitro* formation of an ArgR^{His}-*argC* promoter complex. As indicated in Figure 4.6A, the inclusion of this compound in each binding reaction consistently resulted in the formation of a stable protein-DNA complex with decreased electrophoretic mobility relative to the unbound labelled DNA.

4.4.4.3 Apparent dissociation coefficient determination for the ArgR^{His}-*argC* and -*carA* promoter interaction

The observed change in the electrophoretic mobility of the labelled *argC* and *carA* promoter fragments in response to increasing concentration of ArgR^{His}, shown for example in Figures 4.4 and 4.5, allows for the apparent K_d value for ArgR^{His} binding to these fragments to be calculated for the conditions used in this study. Several EMSAs, using a variety of ArgR^{His} concentration gradients in the presence of 10 mM L-arginine, were thus performed on 20 fmol quantities of each of these labelled DNA fragments. For the *argC* promoter-ArgR^{His} interaction, an apparent K_d value of 247.90 ± 46.80 nM (for monomers of ArgR^{His}) was calculated from six independent EMSAs, whereas, for the *carA* promoter region, a value of 205.26 ± 114.29 nM was calculated from three independent EMSAs. Clearly, these K_d values are not significantly different ($P > 0.1$) and, as reflected by the large SEM values, a prohibitive amount of variation was encountered between each EMSA.

4.4.4.4 Elucidation of the ArgR^{His}-*argC* and -*carA* promoter binding sites

Next, in order to locate any sequence motifs within these promoter regions responsible for the ArgR^{His}-DNA interaction, competitive EMSAs were conducted using unlabelled oligonucleotides complementary to specific portions of the *argC* and *carA* promoter regions. These double stranded competitor fragments were included in each binding reaction at a 1500-fold excess relative to the labelled target fragments and, if they are in possession of

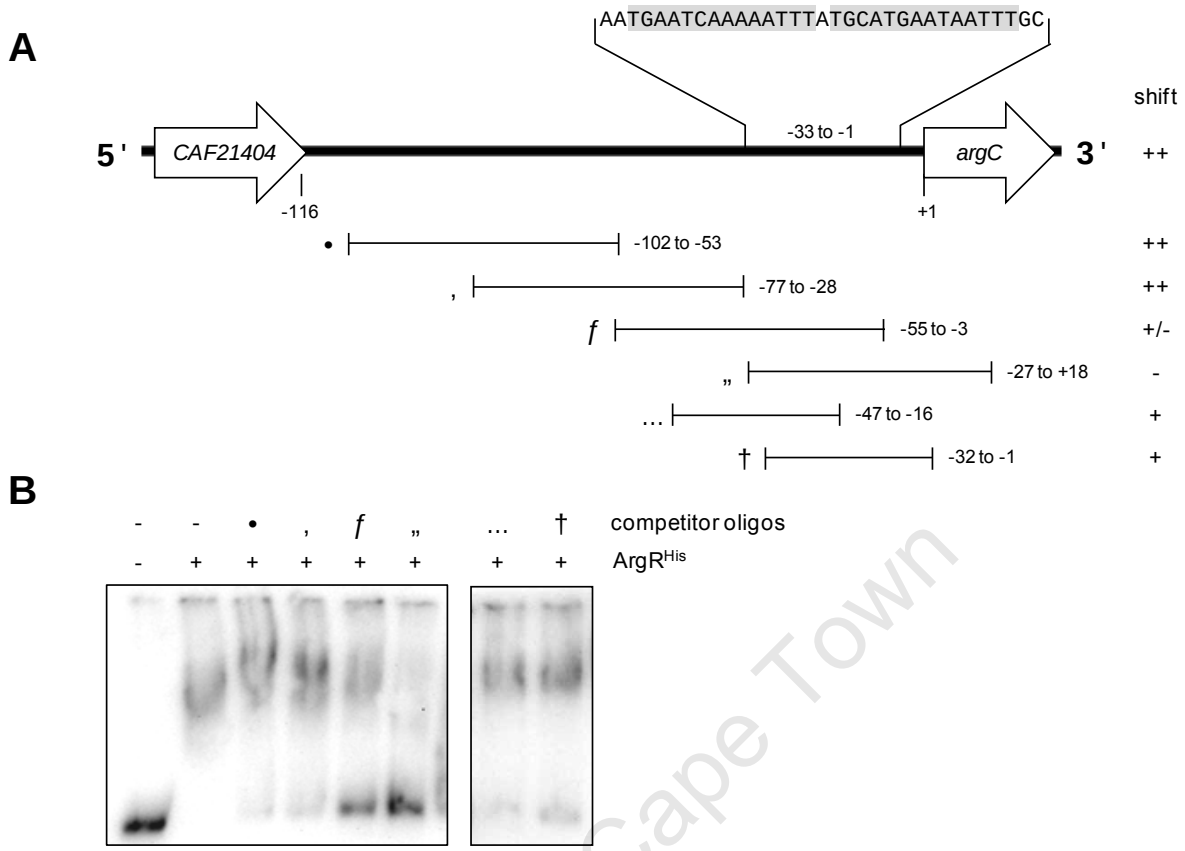


Figure 4.7 Identification of ArgR^{His} binding motifs within the *argC* promoter region by competitive EMSA studies in the presence of L-arginine. **(A)** Detailed view of the *CAF21404-argC* intergenic region showing the position of double stranded competitor oligonucleotides (labelled by the circled numbers). The coordinates for each fragment are supplied relative to the *argC* start codon (+1). The occurrence of an observed ArgR^{His}-meditated mobility shift in the presence of each competitor is indicated. The bracketed sequence contains ArgR^{His}-binding motifs and the highlighted portions possess significant homology to the *E. coli* K12 ARG box core consensus sequence. Diagram not to scale. **(B)** EMSA performed using 20 fmol of DIG-labelled *argC* promoter DNA and 5.47 pmol of ArgR^{His} in each reaction. The presence of each of these reagents, in addition to the relevant competitor fragment (30 pmol), is indicated above each lane.

any ARG box protein binding motifs, they should titrate ArgR^{His} away from the labelled DNA and thus cause the majority of probe to remain in the unbound form. The sequences and lengths of the competitor oligonucleotides are described in Appendix F. As is evident from Figure 4.7, when oligonucleotides corresponding to different portions of the *argC* promoter region were included in various binding reactions, the largest amount of unshifted

probe was observed upon the inclusion of competitor oligonucleotide four, which is complementary to the -27 to +18 region shown in this diagram. An analysis of this region using the Virtual Footprint software subsequently detected a pair of 14 bp ARG box sequence motifs (5'-TGA ATC AAA AAT TT-3' and 5'-TGC ATG AAT AAT TT-3'; designated *argC*$_1$ and *argC*$_2$ respectively) separated by 1 bp and similar to the *E. coli* K12 ARG box core consensus sequence (Figure 4.7A) (Maas, 1994). Interestingly, the inclusion of competitor oligonucleotide six, which contains this putative ARG box pair, in the binding reaction did not appear to result in as much probe remaining unshifted as observed for oligonucleotides three and four. An analogous approach was adopted to identify ARG box sequence motifs within the *carA* promoter region, except here the Virtual Footprint tool was first used to identify putative ARG box motifs within the entire region, before the use of competitive EMSAs to detect ArgR^{His} DNA-binding (Figure 4.8). A relatively small quantity of free *carA* probe was observed upon the inclusion of competitor oligonucleotides two and three in the binding reaction (Figure 4.8B). When an excess of oligonucleotide fragment two, which spans a putative 14 bp ARG box core sequence motif (-61 to -88) (5'-TGA ATG TAG TTT AT-3'; designated *carA*$_1$) detected by the Virtual Footprint software, was included in the binding reaction, a larger amount of free probe remained unbound compared to that obtained in the presence of the other competitor oligonucleotides. Similar to what was seen for the competitive *argC* experiments, the inclusion of the shorter oligonucleotide one in the ArgR^{His} DNA-binding reaction caused less probe to remain unbound relative to the longer oligonucleotide two upon its inclusion, even though it also contains the putative *carA*$_1$ ARG box sequence. A detailed sequence analysis of both these *argC* and *carA* promoter ArgR^{His} binding motifs is presented in Section 4.4.5.

4.4.5 ARG box sequence analysis

An analysis of the these ARG box motifs detected within the *argC* and *carA* promoter regions using competitive EMSAs and the Virtual Footprint program is shown in Figure 4.9. Using these motifs, for which *in vitro* ArgR^{His} binding activity has now been demonstrated, a 14 bp proposed ARG box consensus sequence for *C. glutamicum* ATCC 13032 was constructed (5'-HMT GMA TSW ADW WTW TDY-3') (Figure 4.9A). As evident from the sequence logo diagram in Figure 4.9B, this consensus sequence consists of a core region (positions 3-16) that contains several highly conserved residues in positions 3, 4, 7,

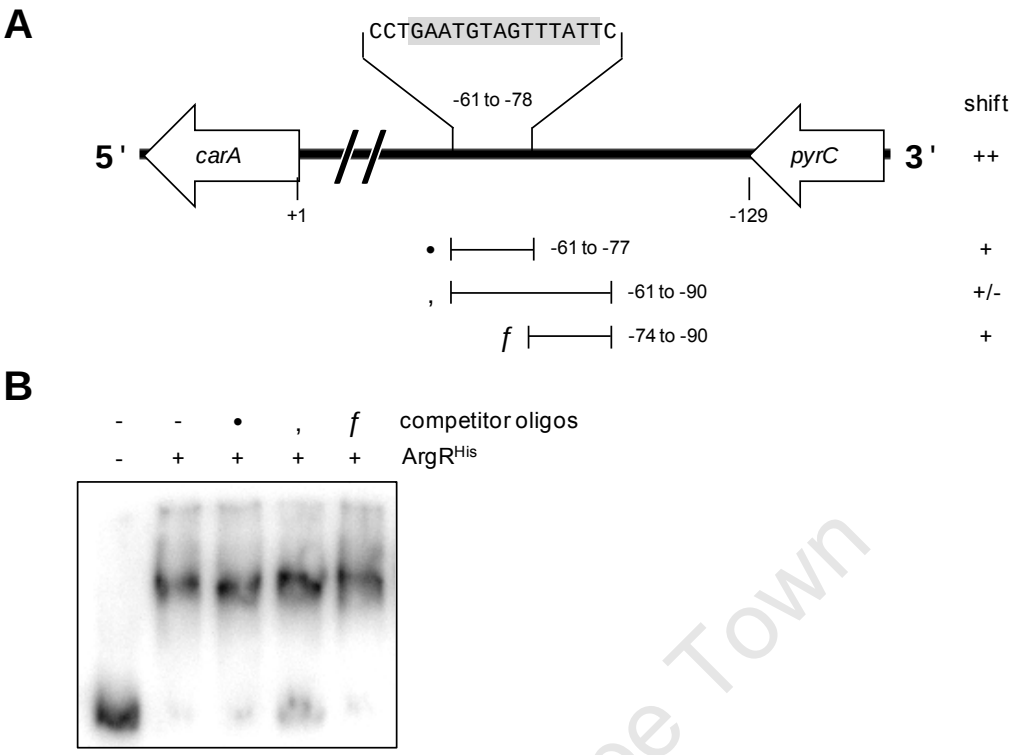


Figure 4.8 Identification of a ArgR^{His} binding motif within the *carA* promoter region by competitive EMSA studies in the presence of L-arginine. **(A)** Detailed view of the *pyrC-carA* intergenic region showing the position of double stranded competitor oligonucleotides (labelled by the circled numbers). The coordinates for each fragment are supplied relative to the *carA* start codon (+1). The occurrence of an observed ArgR^{His}-meditated mobility shift in the presence of each competitor is indicated. The bracketed sequence contains a putative ArgR^{His}-binding motif and the highlighted portion possesses significant homology to the *E. coli* K12 ARG box core consensus sequence. Diagram not to scale. **(B)** EMSA performed using 20 fmol of DIG-labelled *carA* promoter DNA and 5.47 pmol of ArgR^{His} in each reaction. The presence of each of these reagents, in addition to the relevant competitor fragment (30 pmol), is indicated above each lane.

14 and 16 that are similar to the universal core consensus sequence proposed by Makarova *et al*. (2001). Ideally, we would want to use this proposed *C. glutamicum* ATCC 13032 ARG box consensus sequence to screen the remainder of the genome for similar motifs, however, it is difficult to construct an accurate PWM score of such a motif from only three experimentally determined sequences and, for this reason, a PWM of the *E. coli* K12 ARG box core consensus sequence (5'-TGN ATW WW WAT NCA-3') was used (Maas, 1994). Using the Virtual Footprint software, 417 similar motifs located within intergenic regions

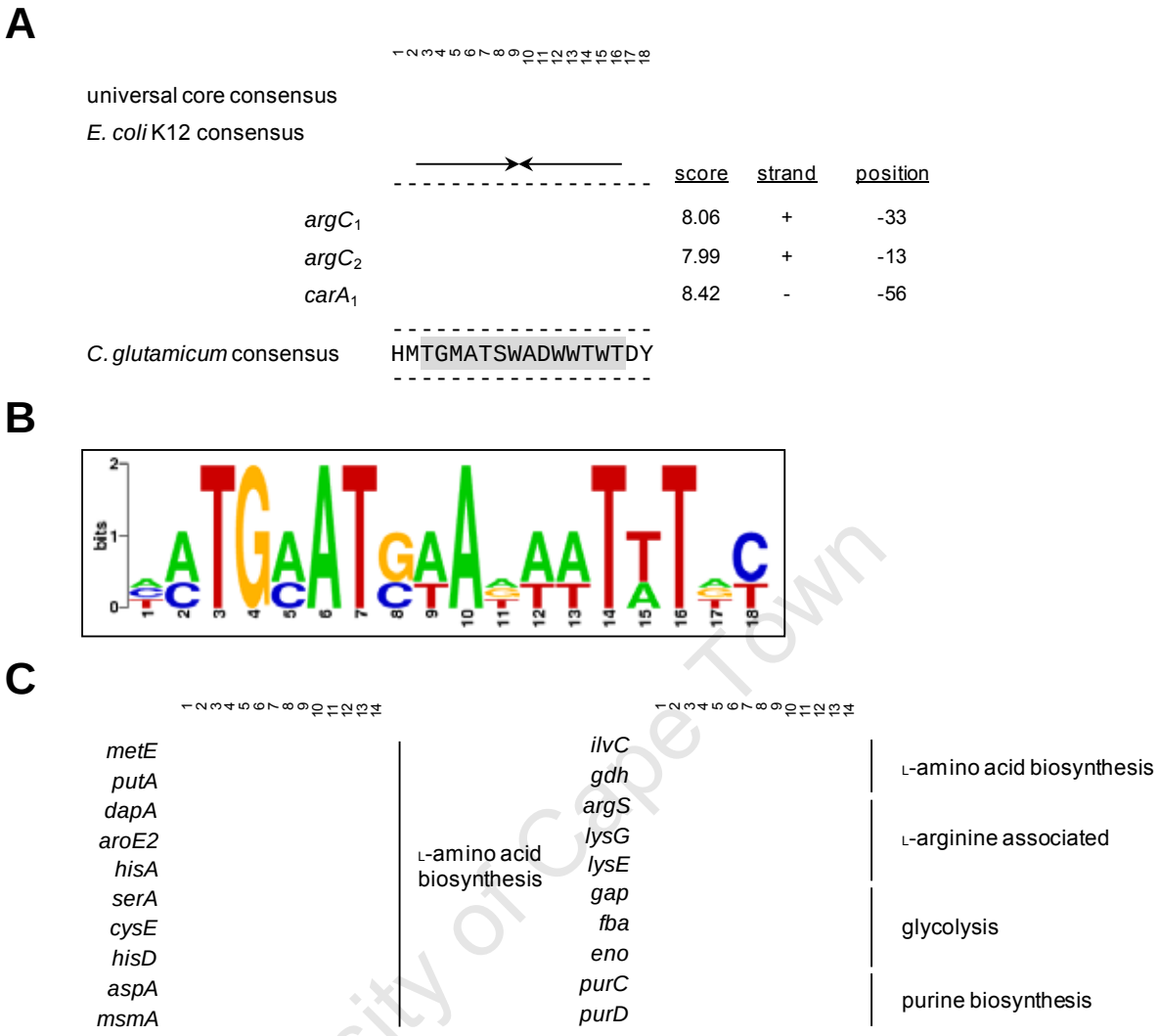


Figure 4.9 Overview of ArgR^{His} binding sites in *C. glutamicum* ATCC 13032. **(A)** Multiple sequence alignment of the putative ARG box universal core consequence sequence of Makarova *et al*. (2001), the extended ARG box *E. coli* K12 consensus sequence of Maas (1994) and the *C. glutamicum* ATCC 13032 extended ArgR^{His}-binding sites within the *argC* and *carA* promoter regions as identified via competitor EMSAs and the Virtual Footprint tool. Where relevant, the PWM score of each core sequence (positions 3-16) relative to the *E. coli* K12 core consensus sequence is indicated. The position of the first nucleotide of each motif is indicated in bp relative to the first nucleotide of the start codon of the relevant gene. The convergent arrows indicate hyphenated dyad symmetry. **(B)** The proposed *C. glutamicum* ATCC 13032 extended ARG box consensus sequence is shown in a sequence logo diagram constructed using WebLogo. **(C)** *In silico* analysis of intergenic *C. glutamicum* ATCC 13032 genomic motifs similar to the *E. coli* K12 core consensus sequence. A selection of motifs that precede genes with predicted ontologies are shown.

and at least 250 bp upstream of the relevant ATG start codon in the *C. glutamicum* ATCC 13032 genome were detected. Over 140 of the 417 genes that these motifs precede are classified ontologically within the Genbank database according to their sequence similarity with other experimentally characterised genes. A selection of these, together with their proposed ontologies as retrieved from Genbank, is shown in Figure 4.9C. Similar motifs were detected preceding the gene clusters encoding the final biosynthetic steps for eleven amino acids, including L-methionine, L-cysteine and L-isoleucine. These motifs were also detected preceding genes such as *aroE2* and *gdh*, which encode important intermediate steps in the biosynthetic pathways of the aromatic and glutamate-derived L-amino acids respectively (Eikmanns, 2005). Putative ARG box motifs were also detected preceding both *lysE* and *lysG*, which respectively encode an important exporter of L-arginine and L-lysine as well its main transcriptional regulator in *C. glutamicum* ATCC 13032 (Section 1.6.1) (Schnell, 1993; Vrljić *et al.*, 1996). In addition to being situated upstream of various genes with putative roles in the glycolytic cycle and purine biosynthesis, these motifs were also detected preceding a variety of other genes, including several proposed to be involved in fatty acid metabolism, the phosphotransferase system and the tricarboxylic acid cycle (Kalinowski *et al.*, 2003). These motifs detected *in silico* closely resemble the preliminary *C. glutamicum* ATCC 13032 ARG box consensus sequence (of which the core sequence is located at positions 3-16 in Figure 4.9A). Notably, all of these motifs contain T, G, T and T residues that each appear to be completely conserved at base positions 1, 2, 5 and 11 of the core consensus sequence and, similar to what has been documented previously in closely related organisms such as *Bacillus subtilis*, were observed to possess a high A+T content between base positions 3 and 11 of their core sequences (Figure 4.9C) (Garnett *et al.*, 2008).

4.4.6 Analysis of the *argC* promoter region

Primer extension analysis of the mRNA encoding the *argC* gene consistently revealed the presence of a putative transcriptional start site located 109 bp upstream of the first base of the *argC* start codon (Figure 4.10). Preceding this initiation site, a sequence (5'-TGG CTA CTA AAA A-3') (core hexamer underlined) within the extended -10 region (positions -17 to -5) and bearing close resemblance to the extended *C. glutamicum* ATCC 13032 -10 consensus sequences (5'-TgtGsTAtAATGG-3') was detected (core hexamer underlined; lowercase letters denote weakly conserved nucleotides) (Pátek, 2005). A -35 hexameric

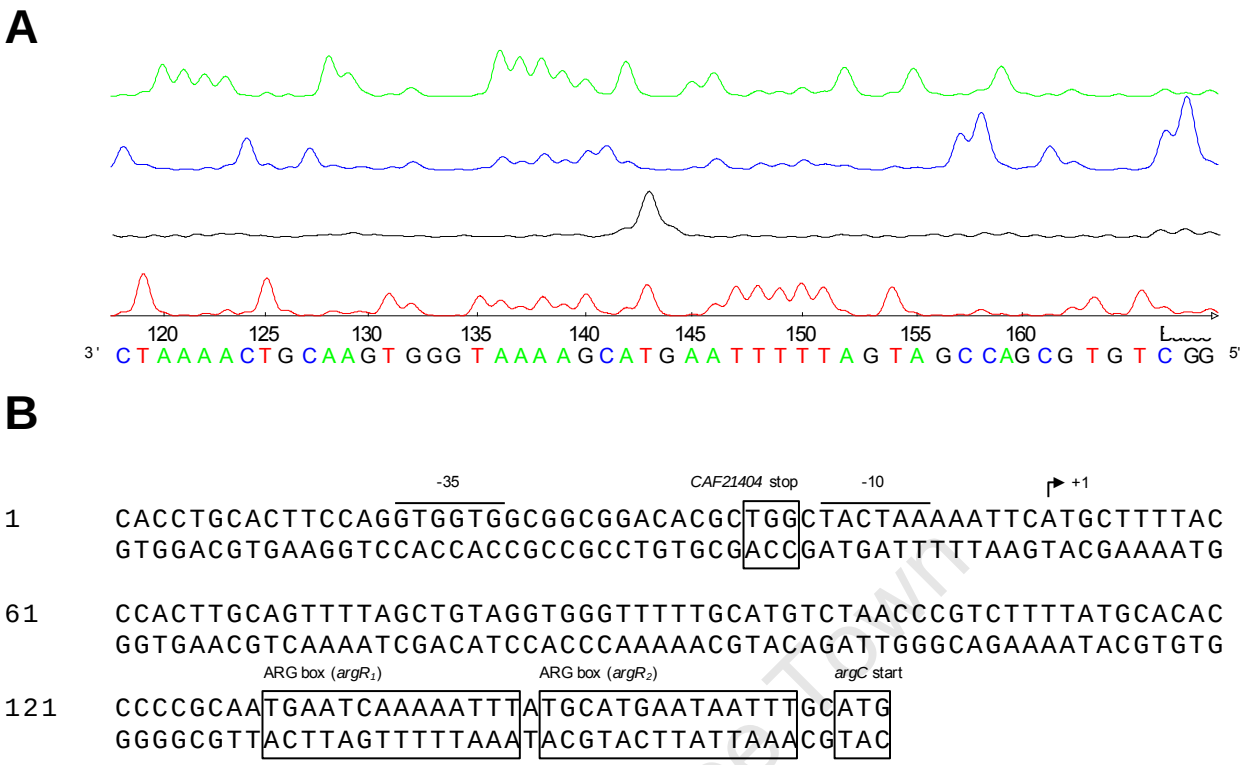


Figure 4.10 Determination of the *argC* promoter putative transcriptional start site. **(A)** A chromatogram of the products of different sequencing reactions (green, blue and red lanes; representing the A, C and T nucleotides respectively) and a primer extension reaction (black lane) is shown. All reactions were conducted using an oligonucleotide primer designed to anneal approximately 40 bp downstream of the *argC* start codon and the products analysed using an ALFexpress Automated DNA Sequencer. The primer extension reaction was performed on RNA extracted from *C. glutamicum* ATCC 13032 cells grown in CGXII MM and the fluorescently labelled RT reaction products were electrophoresed in parallel with the products of sequencing reactions that used plasmid DNA as template. A black peak corresponding in position to the highlighted T nucleotide was observed. The numbers along the x-axis represent the run duration in minutes. **(B)** Proposed structure of the *argC* promoter region. The ARG boxes *argR*$_1$ and *argR*$_2$, the putative *argC* transcriptional start site (indicated by an arrow) and the putative -10 and -35 hexameric sequences are all shown.

sequence (5'-GTG GTG-3'), resembling several known *C. glutamicum* ATCC 13032 -35 motifs that have been shown to be relatively poorly conserved, was identified (Pátek, 2005). Nonetheless, the hexamer identified here partially conforms to the 5'-ttGcca-3' (lowercase letters denote weakly conserved nucleotides) motif identified by Pátek *et al.* (1996).

4.5 Discussion

This chapter reports the use of plasmids harbouring either the *lysE* gene or the *CAF21404-argC* intergenic region to increase L-arginine production in *C. glutamicum* ATCC 13032 cells in addition to the characterisation of the *in vitro* DNA-binding of hexahistidine-tagged ArgR protein to various regions of *C. glutamicum* ATCC 13032 promoter DNA. An analysis of the *argC* promoter and *carA* promoter ARG box sequences is also presented.

4.5.1 Increased *lysE* dosage enhances L-arginine biosynthesis

The function of LysE in *C. glutamicum* ATCC 13032 cells is to regulate the intracellular concentration of L-arginine and L-lysine (Bellman *et al.*, 2001). Here, through the introduction of extra copies of *lysE* into *C. glutamicum* ATCC 13032 cells, both L-arginine production and export were significantly improved. Not only does this confirm the ability of LysE export L-arginine, but it also illustrates that, by overexpressing the LysE protein, the limiting effects of *lysE* transcriptional repression on L-arginine export may be partly alleviated (Section 1.6.1) (Bellman *et al.*, 2001). This thus represents a mechanism for reducing both L-arginine-mediated repression of the L-arginine biosynthetic pathway and the inhibition of L-arginine feedback sensitive enzymes, such as *N*-acetylglutamate kinase (NAGK) and *N*-acetylglutamate synthase (NAGS) (Sakanyan *et al.*, 1996). Furthermore, the overexpression of *lysE* resulted in increased rates of L-arginine export relative to WT cells and it appears, that in the event of a cell accumulating high levels of intracellular L-arginine, increased *lysE* expression may be used to enhance production further. This could also have the effect of alleviating any bacteriostasis that may arise from increasingly unregulated L-arginine accumulation. L-Arginine biosynthesis thus appears to be partially limited by the rate of LysE-mediated L-arginine export and, importantly from an industrial viewpoint, improved export may also allow for the easier purification of this amino acid from bacterial fermentation cultures (Kelle *et al.*, 2005).

4.5.2 pCAF21404-ARGC enhances L-arginine biosynthesis

Elevated dosage of the putative ARG box-containing *CAF21404-argC* intergenic region in *C. glutamicum* ATCC 13032 cells resulted in an approximate 2- to 2.5-fold increase in L-arginine production relative to the pEKEX2 control. This suggests that the derepression of

L-arginine biosynthesis has occurred through the titration of ArgR molecules away from any genome-encoded ARG box sites which, as documented in a variety of other prokaryotes, are thought to precede the L-arginine biosynthetic genes. Interestingly, a similar fold increase in the activity of the ornithine acetyltransferase and ornithine transcarbamylase enzymes, which both fulfil critical steps in the recycling L-arginine biosynthetic pathway, has been reported in *Streptomyces coelicolor* and *Streptomyces lividans* cells that harbour plasmid-borne fragments of the regions located immediately upstream of their respective *argC* genes (Rodríguez-García *et al.*, 1997). Each of these regions was later shown to contain ARG box sequence motifs capable of ArgR-binding through EMSA analysis (Rodríguez-García *et al.*, 1997; Soutar and Baumberg, 1996). Furthermore, *E. coli* K12 cells are known to maintain approximately 300 ArgR molecules per cell during growth in complete medium and, assuming a similar number of ArgR molecules occur within *C. glutamicum* ATCC 13032 cells under the conditions in this study, partial ArgR titration by pCAF21404-ARGC certainly appears feasible. This is because this plasmid, which possesses an origin of replication derived from the cryptic corynebacterial pBL1 plasmid, is estimated to be maintained at roughly 10-30 copies per cell (Miwa *et al.*, 1984; Tauch, 2005; Tian *et al.*, 1994). This experiment clearly demonstrates that the *CAF21404-argC* fragment, which appears to not contain any protein-encoding DNA, contains binding sites for a transcriptional regulator, mostly probably that of ArgR. In order to confirm this and further study the L-arginine biosynthetic gene cluster via EMSA studies, recombinant ArgR protein was now produced.

4.5.3 Preparation of purified ArgR^{His}

The addition of the PelB leader sequence to the ArgR^{His} N-terminal may aid in the recovery of soluble protein extracted from *E. coli* BL21 (DE3) cells because this signal peptide marks it for export from the cytoplasm into the periplasmic space, where conditions are more conducive to promoting proper protein folding (Georgiou and Segatorei, 2005; Wolfe *et al.*, 1983). This thus helps restrict the formation of inclusion bodies and, upon the export of this fusion protein by the Sec system, the 2.45 kDa N-terminal PelB fusion sequence is cleaved by *E. coli* Signal Peptidase I (Wickner *et al.*, 1991). The approximate 24 kDa and 21 kDa protein bands observed in Section 4.4.3 therefore appear to correspond to cleaved and uncleaved forms of ArgR^{His} (where the *argR* coding region encodes an 18.47 kDa protein

and the C-terminal hexahistidine tag encodes a 2.45 kDa fusion). The efficiency at which a PelB-tagged protein is exported into the periplasmic space is closely associated with the intrinsic properties of the target protein and the larger form of hexahistidine-tagged protein detected here in the purified protein extract clearly appears to not have been exported (Georgiou and Segatorei, 2005). In an attempt to isolate only cleaved ArgR^{His}, a periplasmic-only extraction of soluble protein from *E. coli* BL21 (DE3) cells expressing pARGR-P was attempted several times as per the pET System Manual, however, here the uncleaved form of ArgR^{His} still persisted at significant levels within each extract. Ideally, we would want to use a completely homogenous mixture of ArgR^{His} protein for the EMSA studies, however, this N-terminal PelB leader sequence is thought to not generally interfere with the functionality of the fused target protein (Georgiou and Segatorei, 2005). Indeed, several studies have successfully used unprocessed PelB-tagged protein as part of the EMSA-based study of transcriptional regulators which, similarly to ArgR^{His}, are in possession of an N-terminal DNA-binding domain (Horn *et al.*, 2001; Mura-Pastor *et al.*, 2004). It was therefore decided to proceed with using this mixture of ArgR^{His} protein in the EMSA studies and this should be noted especially in light of the calculation of the molar excesses of protein used (where a 50:50 ratio of each variety was assumed) and consequently the determination of the K_d values for *in vitro* ArgR^{His} binding to the *argC* and *carA* promoter regions.

4.5.4 ArgR^{His} DNA-binding

In this study, we demonstrated ArgR^{His} to bind to a pair of ARG box motifs located within the *C. glutamicum* ATCC 13032 *argC* promoter region. The L-arginine dependent manner of this interaction is consistent with that found in *E. coli* K12, where the binding of six L-arginine molecules to the ArgR hexamer was shown to be responsible for activating ARG box binding affinity (Maas, 1994). Such an interaction has also been documented in other bacteria, such as *L. plantarum* and *Streptomyces clavuligerus*, although it differs from that found, for example, in *Lactococcus lactis* and *T. maritima*, where ArgR has been shown to interact in an L-arginine independent manner with similar promoter regions (Larsen *et al.*, 2005; Morin *et al.*, 2003; Nicoloff *et al.*, 2004; Rodríguez-García *et al.*, 1997). Importantly, the ability of homohexameric recombinant ArgR protein to bind DNA distinguishes it from,

for example, the ArgR of *L. plantarum*, which is expected to form a heterohexamer, comprised of the products of two separate genes (Section 1.5.2) (Nicoloff *et al.*, 2004).

Interestingly, the *C. glutamicum* ATCC 13032 $ArgR^{His}$-*argC* promoter interaction was also found to occur in the presence of L-canavanine. As shown in Figure 2.6B, this molecule differs from L-arginine by the shifting of a double bond from the primary to the secondary amine, in addition to the substitution of a carbon atom with an oxygen atom. In *G. stearothermophilus*, Ni *et al.* (1999) have proposed the activation of ArgR DNA-binding activity to rely on the formation of a salt bridge between the guanidino group of each L-arginine molecule and conserved L-aspartic acid residues located on both sides of the ArgR trimer-trimer interface. The guanidino group remains intact in L-canavanine and, in spite of the differences between this molecule and L-arginine, the salt bridge-mediated stabilisation and activation of the ArgR trimers still appears to occur in the presence of this molecule. In contrast, other molecules that are in possession of a guanidino group and that were tested as possible co-effectors in this study, such as L-glutamate, failed to promote *in vitro* $ArgR^{His}$ binding, although this is probably because they individually differ more substantially from L-arginine than L-canavanine. Although they were ultimately not able to effect ArgR-*argC* promoter binding here, these and other metabolites related to amino acid biosynthesis, such as L-citrulline and L-histidine, were included in this study as they have been found to activate the DNA-binding of other related regulators in *C. glutamicum* ATCC 13032, such as LysG (Section 1.6.1) (Bellman *et al.*, 2001). Lastly, the failure of L-ornithine and L-citrulline to stimulate $ArgR^{His}$ binding is in agreement with a study previously performed by Zidwick *et al.* (1984), who found these two compounds not to activate the *in vivo* binding of the *E. coli* ArgR to the promoter region of the *argECBH* operon.

In the work performed here, $ArgR^{His}$ also failed to bind *in vitro* to the putative promoter region preceding the *argGH* genes. Indeed, this is in agreement with the semi-quantitative RT-PCR results discussed in Sections 3.5.3 and 3.5.4, which suggested that the entire L-arginine biosynthetic gene cluster is transcribed as a single transcript. The transcription of the *argGH* genes, which was shown to be depressed in DCOR-D3, therefore appears to be influenced by ArgR binding to the *argC* promoter region. As discussed in Section 3.5.4, however, it remains possible that the FarR regulator, for which L-arginine-independent

binding to the *argGH* promoter region has been demonstrated *in vitro*, may also exert some form of transcriptional repression over these two genes (Hänßler *et al.*, 2007).

Current models for L-arginine biosynthesis in bacteria such as *L. plantarum*, indicate that the ArgR hexamer (here consisting of the products of the *argR1* and *argR2* genes) is expected to bind to the *argGH* promoter region (Nicoloff *et al.*, 2004). Similarly, ArgR is also proposed to bind to this region in *L. lactis* (Larsen *et al.*, 2004). In *C. glutamicum* ATCC 13032, however, no putative ARG box motifs were detected preceding this region and, as the ArgR-mediated regulation of the *argGH* genes thus seems to solely occur via binding to the *argC* promoter region, it appears that an unusual form of transcriptional control of the L-arginine biosynthetic genes may be in place here. In contrast, ArgR binding to the *carA* promoter region in *C. glutamicum* ATCC 13032 is in agreement with these *L. plantarum* and *L. lactis* L-arginine biosynthesis regulatory models and this type of interaction is indeed expected to be present in the majority of prokaryotes (Larsen *et al.*, 2005; Maas, 1994; Nicoloff *et al.*, 2004). Interestingly, this interaction, which presumably would result in the transcriptional repression of the *carAB* genes, would strongly limit the supply of the carbamoyl phosphate (the compound produced by the products of these genes) and hence also exert a strong negative effect on pyrimidine biosynthesis (Section 1.4.1).

Under the *in vitro* conditions used in this study, $ArgR^{His}$ was found to bind to the DIG-labelled *argC* promoter region in the presence of L-arginine with a K_d of 247.90 ± 46.80 nM; higher than that obtained for this protein for the binding to the *carA* promoter region (205.26 ± 114.29 nM) (both values provided for monomeric units of protein). It is tempting to speculate that this enhanced affinity for the *argC* promoter region is due to the presence of an additional ARG box sequence relative to the *carA* promoter region, however, it is unwise to make this assumption given the high amount of experimental variance experienced. Nevertheless, these K_d values are similar to those previously published for the DNA-binding activity of monomeric units of ArgR in other organisms, such as *E. coli*, *L. lactis* and *T. maritima*, which were all determined using a similar EMSA-based method (Caldara *et al.*, 2007; Larsen *et al.*, 2005; Maas, 1994; Morin *et al.*, 2003). For example, Maas *et al.* (1994) reports the L-arginine-dependant ArgR-DNA interaction in *E. coli* cells to typically possess a K_d value ranging between 6-600 nM depending on type of target region. Furthermore, Morin *et al.* (2003) found the *T. maritima* ArgR protein to bind *in vitro* to an

argG promoter region in possession of two ARG box motifs with a collective K_d value of approximately 100 nM. As part of study on ArgR regulation of the L-arginine and L-histidine transport genes in *E. coli* K12, Caldara *et al.* (2007) reported ArgR to bind to the promoter regions of these genes, each in possession of a pair of ARG boxes, with K_d values ranging between 400-600 nM. Similar to what was found here for the $ArgR^{His}$-*argC* promoter interaction, a SEM of roughly 25% was obtained for each of these values. Clearly further work needs to be performed on the ArgR-DNA interaction in *C. glutamicum* ATCC 13032 in order to accurately elucidate the binding affinity of ArgR for promoter regions in possession of variable ARG box sequence motifs. In this context, experimental variance could be reduced by, for example, ensuring each binding reaction is conducted at a constant temperature. Alternative DNA-labelling methods, such as the use of SYBR-Green or radiolabelled DNA, which seemingly allow for more accurate band quantitation, may be used instead of DIG-based methods.

4.5.5 ARG box sequence motifs

The ArgR binding sites identified via competitive EMSAs are highly similar to both the 18 bp ARG box motifs of *E. coli* K12 and the 14 bp ARG box universal core consensus sequence (Section 1.5.2) (Maas, 1994; Makarova *et al.*, 2001). Although ARG boxes are typically spaced by 2-3 bp in organisms such as *E. coli, G. stearothermophilus* and *Thermotoga neopolitana*, the ARG box pair in this study located within the *C. glutamicum* ATCC 13032 *argC* promoter region is separated by only a single base pair (Maas, 1994; Song *et al.*, 2002). By combining the two ARG box sequence motifs within this region with the single ARG box motif detected within the *carA* promoter region, we were able to construct a preliminary *C. glutamicum* ATCC 13032 ARG box consensus sequence (5'-HMT GTA TSW ADW WTW TDY-3' (Figure 4.9A).

Of note within this context is that when relatively short oligonucleotides containing either the $argR_1$ and $argR_2$ ARG box motifs (oligonucleotide six [31 bp] in Figure 4.7) or the $carA_1$ ARG box motif (oligonucleotide one [16 bp] in Figure 4.8) were included in each binding reaction, significantly less probe remained unshifted relative to reactions that instead included longer oligonucleotides also in possession of these motifs (i.e. oligonucleotide four [45 bp] for the *argC* promoter shift and oligonucleotide two [29 bp] for the *carA* promoter shift). $ArgR^{His}$ thus clearly appears to be more effectively titrated away from the DIG-

labelled DNA by these larger fragments. In *E. coli* K12, for example, ArgR DNA-binding has been shown to occur on a single face of the DNA molecule where, in addition to interacting directly with the base pairs within the ARG box motif, it also comes into direct contact with the minor and major groove elements that span four helical turns (corresponding to about 40 bp) (Wang *et al.,* 1998). Thus, assuming a similar situation holds true for $ArgR^{His}$, it appears that the shorter oligonucleotides used here may not be able to adopt the appropriate topological structure required to bind $ArgR^{His}$ with as high as affinity as observed for their longer counterparts. Indeed, this reason would also explain the apparent discrepancy between the amount of free probe observed for competitor oligonucleotide two and the *carA* promoter, which resulted in only a relatively small amount of probe to remain unbound, and that seen for oligonucleotide four and the *argC* promoter, which resulted in significantly more probe remaining unbound upon its inclusion in the bind reaction.

The screening of the *C. glutamicum* ATCC 13032 genome using the *E. coli* K12 core consensus sequence revealed a variety of similar sequences that precede genes that are diverse in terms of their predicted functions (a selection of which is presented in Figure 4.9C). This could be indicative of a possibly wider-regulatory role for ArgR in this bacterium. Indeed, the study of *in vitro* ArgR binding to promoter regions preceding genes thought not to be directly involved in L-arginine metabolism appears to be relatively unexplored in the literature. Notable exceptions to this, however, include the work of Larsen *et al.* (2004) and Caldara *et al.* (2007), who demonstrated ArgR to bind to the promoter region preceding genes responsible for L-arginine export in *L. lactis* and *E. coli* respectively. Similarly, one of the putative ArgR target motifs identified here as part of our *in silico* screening of the *C. glutamicum* ATCC 13032 genome precedes the *lysE* gene, which encodes an important exporter of both L-arginine and L-lysine (Section 1.6.1) (Vrljić *et al.,* 1996). This immediately implies that the *C. glutamicum* ATCC 13032 ArgR probably plays an additional regulatory role in L-lysine biosynthesis and the inactivation of *argR* in this bacterium could thus prove useful for the creation of an L-lysine overproducing strain. Moving towards a comprehensive characterisation of the genome-wide binding capabilities of the *C. glutamicum* ATCC 1032 ArgR, a logical next step here would be to perform a transcriptome-based analysis of the DCOR-D3 strain. For example, such a strategy was recently successfully adopted by Buchinger *et al.* (2008) to reveal several novel targets of

the AmtR master nitrogen regulator in *C. glutamicum* ATCC 13032. Once transcriptional evidence of ArgR-mediated regulation of these genes has been gathered, ArgR binding may be verified *in vitro* and, in order to assess the contribution of each gene towards a particular phenotype, such as L-arginine biosynthesis, a knock out mutant could be constructed and the appropriate assay or metabolomic analysis thereafter conducted.

Finally, the putative core ARG box sequences detected both either *in silico* or *in vitro* via competitive EMSAs are all in possession of a unanimously conserved T nucleotide at position 11 (Figure 4.9C). In contrast, the nucleotide at this position in the ARG boxes of other bacteria, such as *E. coli* K12 and various *Streptomyces* spp., is less conserved (Maas, 1994; Rodríguez-García *et al.*, 1997). This may indicate that the *C. glutamicum* ATCC 13032 ARG box sequence motifs are more conserved on an intra-genome level relative to that found in other bacteria and, in future, the contribution of individual nucleotides within these motifs towards the overall binding affinity of ArgR may be assessed through the introduction of single point mutations into oligonucleotides in possession of ARG boxes (Makarova *et al.*, 2001). By analogy with a recent study by Garnett *et al.* (2008), however, we can expect some of the base pairs detected here within the core ARG box sequences, such as the conserved G residue and the semi-conserved A residue (located respectively at base positions 2 and 10 of each sequence shown in Figure 4.9C), to be especially important for ArgR DNA-binding in *C. glutamicum* ATCC 13032 as, in *B. subtilis*, they have been shown to play an important role in establishing specific contacts with ArgR amino acids that are critical for formation of the protein-DNA complex. Lastly, in order to map the exact site of the ArgR-ARG box interaction within each promoter region, a DNA footprinting approach, similar to that used by Caldara *et al.* (2007), may be adopted.

4.5.6 Structure of the *argC* promoter region

As part of this work, the putative start site of a RNA transcript containing the *argC* gene was located within the *argC* promoter region. Both a -10 and a -35 sequence resembling the conserved motifs of Pátek (2005) were detected in close proximity to this transcriptional start site and it appears feasible that the binding of ArgR to the motifs identified in Section 4.4.4.4 would obstruct RNA polymerase from moving along the DNA strand. It should be noted, however, that during the initial primer extension analysis of the *argC* promoter, multiple signals indicative of several transcriptional start points were detected. Upon

further analysis of this sequence using the RNA Mfold program (available from http://mfold.bioinfo.rpi.edu/), the majority of these signals were found to be located near or in close proximity to strong secondary structure elements, such as stem-loop structures (Zuker, 2003). Consequently, after the inclusion of betaine in the reaction volume, in addition to an elevation of the extension temperature of the reaction, many of these signals were no longer detected (Henke *et al.*, 1997). This number was reduced further through the use of avian myeloblastosis virus RT which, according to the manufacturer, is more tolerant of secondary structure elements than the avian myeloblastosis virus RT used initially for primer extension. After these adaptations were made, a single signal was reliably detected within 300 bp of the *argC* start codon. It should be noted here that the A residue detected in the sixth position of the 5'-TACTAA-3' *argC* -10 motif is unusual as in most previously characterised *C. glutamicum* ATCC 13032 -10 regions a T residue is present at this position (Pátek, 2005). It is not, however, unprecedented and a similar motif with an A in this position has been detected by Mateos *et al.* (1994) preceding the *thrB* gene. Given the rarity of the motif detected here, however, this result should be confirmed by a second primer extension analysis with a different primer.

A similar approach was adopted in order to locate any potential transcriptional start sites preceding the *argG* gene. Using a primer whose specificity was first verified as part of a positive control experiment, multiple signals possibly indicative of putative transcriptional start sites were initially identified. Like that found for the *argC* promoter, many of these signals were found to coincide with stem-loop structures or were located unfeasibly far from the *argG* start codon. Upon the adoption of different combinations of the modifications described above with a view to make the RT process more stringent, we were not, however, able to reliably detect any strong signals. For this reason, we were unable to accurately determine the existence and location of an *argG* transcriptional start site with confidence. The possible causes for this are diverse. For example, although a large range of different RNA quantities were used as part of the RT reaction, any population of putative *argGH* transcripts, if present, may remain too small and could be undetectable via this method. Also, the equipment used here is not able to accurately detect labelled RT products less than 50 bp in length and, in future, a method such as the rapid amplification of 5'-cDNA ends may be better suited for the accurate determination of multiple transcriptional start sites.

4.6 Concluding remarks

In this chapter we provide the first characterisation of the ArgR-DNA interaction found in *C. glutamicum* ATCC 13032 cells. The detection of ArgR^{His} binding to the *argC* and *carA* promoter regions, together with the elucidation of the sequences of the ARG box motifs within these regions, further extends our knowledge of the regulation of L-arginine biosynthesis in *C. glutamicum* ATCC 13032 which, in contrast to bacteria such as *L. lactis*, appears to employ only a single transcriptional regulator encoded by the *argR* gene. Although the L-arginine-dependent nature of the ArgR^{His}-*argC* promoter interaction is in good agreement with canonical models of regulation found for example in *E. coli* K12, it is unusual in that this protein does not appear to interact with the *argG* promoter region. Additionally, the ability of L-canavanine to activate ArgR^{His} binding to the *argC* promoter region is interesting from a molecular point of view and this might hold relevance for future structural studies performed on ArgR and the examination of the mechanism of activation of this protein. Importantly, the ARG box motifs with the *argC* and *argG* promoter regions detected via competitive EMSAs were found to be similar to the *E. coli* K12 ARG box consensus sequence and this latter sequence was thus successfully used for the *in silico* detection of other putative sites of ArgR interaction within the WT genome. This revealed a possibly broad regulatory role for the *C. glutamicum* ATCC 13032 ArgR and, once binding to these sites has been studied *in vitro* in a similar manner to that used in this study, the preliminary ARG box consensus sequences constructed here from experimentally-verified motifs can be further enhanced and subsequently used for the more accurate detection of putative ARG box motifs located throughout the genome of this bacterium. This study has thus laid the preliminary groundwork for the further characterisation of the role of this protein in this bacterium and this body of work would, in future, be greatly enhanced by a transcriptomal analysis of gene expression in strains lacking a functional *argR* gene. This could reveal potentially novel, previously unknown targets for mutagenesis that, similar to *lysE*, improve the biosynthesis of L-arginine or other amino acids upon the manipulation of their expression levels.

Chapter Five

General conclusions and future research

Prior to this study, the physiology and ArgR-mediated regulation of L-arginine biosynthesis in *Corynebacterium glutamicum* ATCC 13032 had not been investigated to any real extent. As outlined in Chapter 1, this organism is naturally geared towards the efficient biosynthesis of various L-amino acids. From an industrial viewpoint, however, very few L-arginine overproducer strains of this bacterium have been described to date in the literature. Furthermore, barring a very recent study by Ikeda *et al.* (2009), all of these strains have been created via a classical mutagenesis approach (Utagawa, 2004). This prompted us to investigate how L-arginine biosynthesis in *C. glutamicum* ATCC 13032 may be further enhanced by directed mutagenesis of the host genome. The *argR* gene, previously shown to be an autoregulated transcriptional repressor of the L-arginine biosynthetic genes in several organisms, was initially targeted. Additionally, in order to provide a preliminary understanding of the nature of the ArgR-mediated regulation in this bacterium, we also sought to characterise the DNA-binding activity of ArgR. In this chapter, a broad summary of these results is presented, together with recommended guidelines regarding the future continuation of this work.

The p2NIL/pGOAL19 vector system was initially used for directed mutagenesis of the *C. glutamicum* ATCC 13032 genome. These vectors, previously used only in *Mycobacterium tuberculosis*, was chosen over the conventional pK19*mobsacB* system as they encode several additional markers that aid in the selection and isolation of double cross-over (DCO) mutants (Eggeling and Reyes, 2005; Parish and Stoker, 2000). Single cross-over (SCO) *argR* and *argH* mutants were successfully created using this method, although we were ultimately unable to derive DCO mutants from these cells. This appears to be because insufficient screening of the relatively unstable plasmid-encoded sucrose sensitivity phenotype was performed prior to the construction of the *argR* and *argH* gene knock-out vectors. As the DCO event is especially rare and hence difficult to isolate in the absence of an effective negative selection marker, the downstream isolation of DCO mutants in possession of a mutant allele was thereby rendered problematic. The pK19*mobsacB* vector was thus used for the creation of an *argR* chromosomal deletion mutant and, as

recommended by Eggeling and Reyes (2005), an initial sucrose sensitivity screening step of the cells was performed. Our recommendation is thus that such a step should always be implemented prior to the use of the p2NIL/pGOAL19 system. This would ultimately permit for the molecular-based screening of fewer potential DCO mutants than what would typically be required of those created using a pK19*mobsacB*-derived suicide vector.

The *argR* mutants created in this study all displayed physiological characteristics favourable from an industrial standpoint. For example, their maximum growth rates were unaffected by the introduced mutation and their production of L-arginine and L-citrulline was elevated significantly (with the *argR* DCO mutant accumulating each of these compounds at approximately 5- and 8-fold relative to the WT). Semi-quantitative reverse transcriptase polymerase chain reaction (PCR) experiments suggested that the L-arginine biosynthetic genes are transcribed together as a single *argCJBDFRGH* transcript (the proposed transcriptional organisation, together with a proposed model of L-arginine biosynthesis in *C. glutamicum* ATCC 13032 is shown in Figure 5.1). This notion is reinforced by the fact that primer pairs targeted within both the *argC-argJ* and *argG-argH* regions showed evidence of the transcriptional deregulation of these genes upon *argR* deletion. As ArgR appears to bind to the *argC* promoter region (and not to the *argG* promoter region) in *C. glutamicum* ATCC 13032, it appears that the ArgR-*argC* promoter interaction is able to exert control over the formation of transcripts encoding the *argG* and *argH* genes. It should be noted, however, that, although this latter form of control appears to seem more likely, ArgR-mediated regulation of the transcription of *argG* and *argH* may not necessarily occur directly via this mechanism. It is conceivable that the ArgR may regulate the formation of another regulator, such as FarR, which is capable of binding to the *argG* promoter region as a negative transcriptional regulator (Hänßler *et al.*, 2007). Indeed, further evidence of a mixed population of *argCJBDFRGH*, *argCJBDFR* and *argGH* transcripts is provided by the *argR* integration mutants created as part of our study that lacks an intact *argR* gene and is capable of growth on minimal medium. The integration event here is predicted to throw the *argCJBDFRGH* transcript out of frame and this is therefore suggestive of the existence of *argGH* transcriptional start site downstream of the site of integration. As discussed extensively in Chapter 3, further work in the form of a Northern blot analysis or quantitative real-time PCR is clearly needed here.

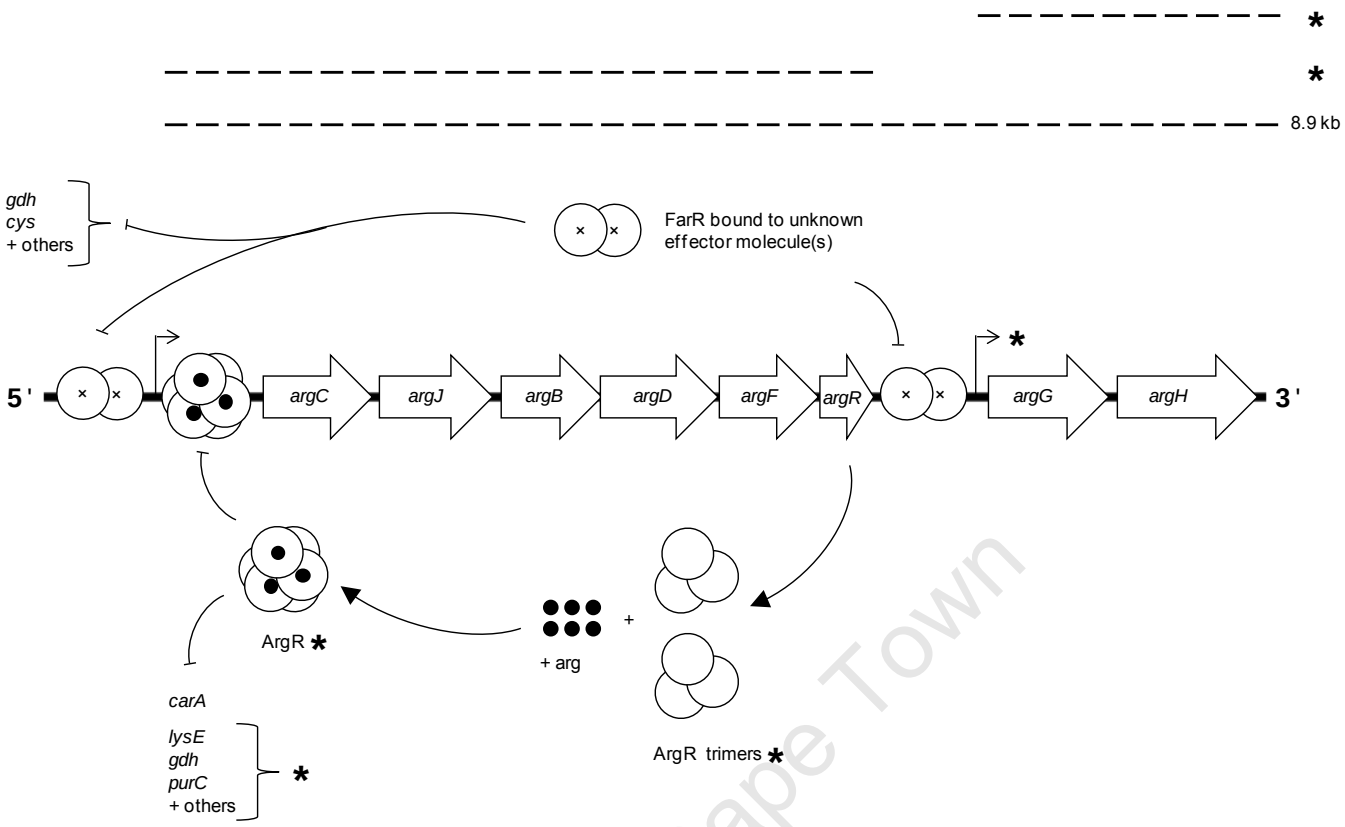


Figure 5.1 Proposed model of regulation of the L-arginine biosynthetic gene cluster in *C. glutamicum* ATCC 13032. ArgR trimer oligomerisation is expected to be activated by the binding of six L-arginine molecules. Hexameric ArgR is expected to bind to ARG boxes within the *argC* and *carA* promoter regions, in addition to several other putative targets identified via an *in silico* analysis of the *C. glutamicum* ATCC 13032 genome. The form of regulation exerted on these other targets, including *carA*, by ArgR remains to be experimentally elucidated. The interaction of dimeric FarR with the *argC* and *argG* promoters has been demonstrated by Hänßler *et al.* (2007), although the exact location of DNA-binding motif(s) and the type of effector molecule required is to date undetermined. Hänßler *et al.* (2007) also identified several other sites putatively regulated by FarR, such as the *gdh* and *cys* genes, by analysing the transcriptome of a *farR* deletion mutant. Dashed lines indicate the proposed transcripts that encode the L-arginine biosynthetic genes. Putative transcription start site are indicated by bent arrows. Asterisks represent proposed features that are not yet experimentally verified. Diagram not to scale.

As shown in Figure 5.1 and discussed in detail in Section 1.5.3, Hänßler *et al.* (2007) have recently demonstrated the *C. glutamicum* ATCC 13032 FarR to bind *in vitro* to the *argC* and *argG* promoter regions. Although upon the deletion of *farR* the population of

transcripts encoding the L-arginine biosynthetic genes was elevated 3- to 5-fold, the amount of L-arginine produced by the mutant was found to not increase significantly (Hänßler *et al.*, 2007). Interestingly, as part of our work and similar to that seen by Ikeda *et al.* (2009), the deletion of *argR* resulted in the accumulation of large amounts of L-citrulline by mutant *C. glutamicum* cells. As the enzymes required for the biosynthesis of this compound from L-glutamate are encoded by the FarR- and ArgR- transcriptionally regulated *argCJBDF* genes, it thus appears that FarR-mediated regulation of *argGH* may be important for the fine control of L-citrulline utilisation by the final two steps of the L-arginine biosynthetic pathway. It therefore stands to reason that, through the deletion of *farR* in the *argR* deletion mutant, we can expect to see the levels of all the transcripts shown in Figure 5.1 to increase and L-arginine production to be further elevated. Consequently, we propose that the unchanged L-arginine levels observed upon *farR* deletion by Hänßler *et al.* (2007) could be due to a form of ArgR-mediated transcriptional super-repression of the L-arginine biosynthetic gene cluster.

As detailed in Chapter 4, several putative targets for ArgR binding were detected preceding a variety of genes via an *in silico* analysis of the *C. glutamicum* ATCC 13032 genome. The alteration of the expression levels of some of these genes would appear to represent a viable strategy for further enhancing L-arginine production in this bacterium. For example, we showed in this study that a 2-fold improvement in L-arginine biosynthesis may be accomplished by increasing the copy number of the *lysE* gene. Although transcription of many of the genes identified here would already be elevated by the deletion of the *argR* gene, we would expect both the enhancement of their promoter activity and their increased dosage within the host chromosome to further improve L-arginine production. Such an approach, for example, has previously been used successfully to increase L-lysine biosynthesis in *C. glutamicum* ATCC 13032 through the chromosomal introduction of a *dapA* gene in possession of a mutant promoter (Pfefferle *et al.*, 2003; Vašicová *et al.*, 1998). Of interest, several of these *in silico* identified putative ArgR targets, such as the *gdh* and *cys* genes, were found to also be transcriptionally regulated by FarR (Hänßler *et al.*, 2007). This is suggestive of a complex as yet unreported interplay between these two transcription factors and the exact nature of such a mechanism remains to be elucidated.

Notably, the L-arginine levels attained here by the DCO mutant are several-fold below those documented in classically-mutated producer strains (Utagawa, 2004). A similar affect has been observed by Ikeda *et al.* (2009) and this is indicative of the ability of further directed mutagenesis of the *C. glutamicum* ATCC 13032 genome to improve L-arginine production. Taking the increasing efficiency of sequencing methods into account, so-called 'genome-breeding' technologies, which rely on the sequencing of the genomes of these classical producers in order to identify potentially useful mutations and their subsequent introduction on a one-on-one basis into model strains, should prove increasingly viable for the future creation of recombinant strains capable of producing L-arginine at levels exceeding those produced by the classical producers (Section 1.7).

Once the DNA-binding activity of ArgR for some of these *in silico*-identified motifs has been demonstrated *in vitro* using a similar approach to that adopted in this study, it would be possible to improve upon the preliminary *C. glutamicum* ATCC 13032 ARG box consensus sequence constructed here (5'-HMT GMA TSW ADW WTW TDY-3'). This would locate other putative ArgR-binding sequences within the genome to be detected with enhanced sensitivity. Furthermore, as discussed in detail in Section 4.5.5, the genome-wide characterisation of ArgR DNA-binding activity would be greatly enhanced via a variety of system-wide approaches of the *argR* deletion mutant, such as transcriptomal- and metabolomic-based analyses.

This study has not addressed the purported secondary role of ArgR – that of a resolvase important for the resolution of plasmid multimers (Section 1.5.2). This had to date been documented chiefly in *E. coli* and, although it lies outside of the scope of L-arginine biosynthesis, it may nonetheless be of interest to investigate plasmid resolution within the *argR* DCO mutant in a manner similar to that of Stirling *et al.* (1988).

Our work has mainly focussed on the genetic means whereby L-arginine biosynthesis may be improved in *C. glutamicum* ATCC 13032. Within this context, there are, of course, additional steps that can be adopted to improve production, such as the introduction of mutations rendering the *N*-acetylglutamate synthase and *N*-acetylglutamate kinase enzymes feedback resistant. Other options include, for example, the supplementation of growth media with intermediates of the L-arginine biosynthetic pathway, such as L-ornithine, and the elevation of the fermentation temperature, which Tuchman *et al.* (1997)

and Ikeda *et al.* (2009) have respectively reported to improve L-arginine biosynthesis significantly. Importantly, as discussed in Section 4.5.4, the apparent L-arginine-dependent nature of the interaction between homohexameric ArgR and the ARG box motifs elucidated here distinguishes this mechanism of L-arginine regulation from that found in other bacteria, such as *Lactobacillus plantarum* and *Lactococcus lactis* (Larsen *et al.*, 2005; Nicoloff *et al.*, 2004).

In conclusion, this thesis has provided the first study of the ArgR protein in *C. glutamicum* ATCC 13032. This work is of applied as well as academic interest, as, not only has it generated a body of information that may be used for the engineering of an industrial L-arginine overproducer strain, but it should serve as a basis for the future elucidation of global ArgR-mediated regulation in *C. glutamicum* ATCC 13032.

Literature Cited

Abe, S., Takayama, K. I., and Kinoshita, S. (1967). Taxonomic studies on glutamic acid-producing bacteria. J. Gen. Appl. Microbiol. 13: 279-304.

Alm, E. W., Oerther, D.B., Laresen, N., Stahl, D. A., and Raskin, L. (1996). The oligonucleotide probe database. Appl. Environ. Microbiol. 62: 3557-3559.

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

Arndt, A., and Eikmanns, B.J. (2008). Regulation of carbon metabolism in *Corynebacterium glutamicum*. *In*: Burkovski, A., (Ed.), Coyrnebacteria: Genomics and Molecular Biology, Caister Academic Press, Erlangen, Germany, pp. 121-148.

Barcelona-Andres, B., Marina, A., and Rubio, V. (2002). Gene structure, organization, expression and potential regulatory mechanisms of arginine catabolism in *Enterococcus faecalis*. J. Bacteriol. 184: 6289 – 6300.

Barksdale, L. (1970). *Corynebacterium diphtheriae* and its relatives. Bacteriol. Rev. 34: 378-422.

Bellman, A., Vrljić, M., Pátek, M., Sahm, H., Krämer, R., and Eggeling, L. (2001). Expression control and specificity of the basic amino acid exporter LysE of *Corynebacterium glutamicum*. Microbiology 147: 1765-1774.

Berg, J. M., Tymoczko, J. L., and Stryer, L. (Eds.) (2001). The Calvin cycle and the pentose phosphate pathway. *In*: Biochemistry, 5th Edition. W. H. Freeman and Company, New York, USA, pp. 551-576.

Blombach, B., Schreiner, M. E., Bartek, T., Oldiges, M., and Eikmanns, B. J. (2008). *Corynebacterium glutamicum* tailored for high-yield L-valine production. Appl. Microbiol. Biotechnol. 79: 471-479.

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

Bringel, F., Frey, L., Boivin, S., and Hubert, J. C. (1997). Arginine biosynthesis and regulation in *Lactobacillus plantarum*: the *carA* gene and the *argCJBDF* cluster are divergently transcribed. J. Bacteriol. 179: 2697-2706.

Bröer, S., and Krämer, R. (1991). Lysine excretion by *Corynebacterium glutamicum*. 1. Identification of a specific secretion carrier system. Eur. J. Biochem. 202: 131-135.

Buchinger, S., Strösser, J., Rehm, N., Hänßler, E., Hans, S., Bathe, B., Schomburg, D., Krämer, R., and Burkovski, A. (2008). A combination of metabolome and transcriptome analyses reveals new targets of the *Corynebacterium glutamicum* nitrogen regulator AmtR. J. Biotechnol. [Article in Press as of 10 February 2009].

Caldara, M., Le Minh, P. N., Bostoen, S., Massant, J., and Chartlier, D. (2007). ArgR-dependant repression of arginine and histidine transport genes in *Escherichia coli* K-12. J. Mol. Biol.373: 251-267.

Carrió, M. M., and Villaverde, A. (2001). Protein aggregation as bacterial inclusion bodies is reversible. FEBS Letters 489: 29-33.

Chaouni, L. B., Etienne, J., Greenland, T., and Vandenesch, F. (1996). Nucleic acid sequence and affiliation of pLUG10, a novel cadmium resistance plasmid from *Staphylococcus lugdunensis*. Plasmid 36: 1-8.

Cherney, L. T., Cherney, M. M., Garen, C. R., Lu, G. J., and James, M. N. (2008). Crystal structure of the arginine repressor protein in complex with the DNA operator from *Mycobacterium tuberculosis*. J. Mol. Biol. 384: 1330-1340.

Cohen-Kupiec, R., Kupiec, M., Sandbeck, K., and Leigh, J. A. (1999). Functional conservation between the argininosuccinate lyase of the archaeon *Methanococcus maripaludis* and the corresponding bacterial and eukaryal genes. FEMS Microbiol. Lett. 173: 231-238.

Collins, M. D., and Cummins, C. S. (1986). Genus *Corynebacterium* Lehman and Neumann, 1896, pp 1266-1276. In: Sneaths, P. H. A., Mair, N. S., Sharpe, M. E., and Holt, J. G. (Eds.), Bergey's Manual of Systematic Bacteriology, vol. 2. Williams and Wilkins, Baltimore, USA.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: A sequence logo generator. Genome Res. 14: 1188-1190.

Cunin, R., Glansdorff, N., Pierard, A., and Stalon, V. (1986). Biosynthesis and metabolism of arginine in bacteria. Microbiol. Rev. 50: 314–352.

Daffé, M. (2005). The cell envelope of Corynebacteria. *In*: Eggeling, L., Bott, M. (Eds.), Handbook of *Corynebacterium glutamicum*, CRC Press, Boca Raton, USA, pp. 121-148.

Dagert, M., and Ehrlich, S.D. (1979). Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. Gene 6: 23–28.

Dennis, C. C., Glykos, N. M., Parsons, M. R., and Philips, S. E. (2002). The structure of AhrC, the arginine repressor/activator protein from *Bacillus subtilis*. Acta. Crystallogr. 58: 421-430.

Edwards, R. A., Helm, R. A., and Maloy, S. R. (1999). Increasing DNA transfer efficiency by temporary inactivation of host restriction. Biotechniques 26: 892-898.

Eggeling, L. (2005). Export of amino acids and other solutes. *In*: Eggeling, L., Bott, M. (Eds.), Handbook of *Corynebacterium glutamicum*. CRC Press, Boca Raton, USA, pp. 187-208.

Eggeling, L., and Reyes, O. (2005). Experiments. *In*: Eggeling, L., Bott, M. (Eds.), Handbook of *Corynebacterium glutamicum*. CRC Press, Boca Raton, USA, pp. 535-556.

Eggeling, L., and Sahm, H. (2001). The cell wall barrier of *Corynebacterium glutamicum* and amino acid efflux. J. Biosci. Bioeng. 92: 201-213.

Eggeling, L., and Sahm, H. (2003). New ubiquitous translocators: amino acid export by *Corynebacterium glutamicum* and *Escherichia coli*. Arch. Microbiol. 180: 155-160.

Eikmanns, B. J. (2005). Central metabolism: tricarboxylic acid cycle and anaplerotic reactions. *In*: Eggeling, L., Bott, M. (Eds.), Handbook of *Corynebacterium glutamicum*. CRC Press, Boca Raton, USA, pp. 241-276.

Eikmanns, B. J., Kleinertz, E., Liebl, W., and Sahm, H. (1991). A family of *Corynebacterium glutamicum*/*Escherichia coli* shuttle vectors for cloning, controlled gene expression, and promoter probing. Gene 102: 93-98

Eikmanns, B. J., Thum-Schmitz, N., Eggeling, L., Luedtke, K., and Sahm, H. (1994). Nucleotide sequence, expression and transcriptional analysis of the *Corynebacterium glutamicum glnA* gene encoding citrate synthase. Microbiology 140: 1817-1828.

Friedberg, E. C., Walker, G. C., and Siede, W. (1995). DNA repair and mutagenesis. American Society for Microbiology Press, Washington DC, USA.

Gallegos, M. T., Schleif, R., Bairoch, A., Hoffmann, K., and Ramos, J. L. (1997). AraC/XylS family of transcriptional regulators. Microbiol. Mol. Biol. Rev. 61: 393-410.

Garnett, J. A., Marincs, F., Baumberg, S., Stockley, P. G., Phillips, S. E. (2008). Structure and function of the arginine repressor-operator complex from *Bacillus subtilis.* J. Mol. Biol. 379: 284-298.

Gebhardt, H., Meniche, X., Tropis, M., Krämer, R., Daffé, M., and Morbach, S. (2007). The key role of the mycolic acid content in the functionality of the cell wall permeability barrier in *Corynebacteriaceae*. Microbiology 153: 1424-1434.

Georgiou, G., and Segatorei, L. (2005). Preparative expression of secreted proteins in bacteria: status report and future prospects. Curr. Opin. Biotechnol. 16: 538-545.

Goodfellow, M. (1992). The family *Nocardiaceae*. *In*: Balows, A., Trüper, H. G., Dworkin, M., Harder, W., and Schleifer, K. H. (Eds.), The Prokaryotes, 2nd Edition. Springer-Verlag, New York, USA, pp. 1188-1213.

Goodfellow, M., and Minnikin, D. E. (1981). Introduction to the coryneform bacteria. *In*: Starr, M. P., Stolp, H., Trueper, H. G., Balows, A., and Schlegel, H. G. (Eds.), The Prokaryotes, vol. 2. Springer-Verlag, Berlin, Germany, pp. 1811-1826.

Gordhan, B. G., and Parish, T. (2001). Gene replacement using pretreated DNA. *In*: Parish, T., and Stoker, N. G. (Eds.), *Mycobacterium tuberculosis* Protocols, vol. 54. Humana Press Inc., Totowa, USA, pp. 77-92.

Gordhan, B. G., Smith, D. A., Alderton, H., McAdam, R. A., Bancroft, G. J., and Mizrahi, V. (2002). Construction and phenotypic characterization of an auxotrophic mutation of *Mycobacterium tuberculosis* defective in L-arginine biosynthesis. Infect. Immun. 70: 3080-3084.

Haas, D., Kurer, V., and Leisinger, T. (1972). *N*-Acetylglutamate synthetase of *Pseudomonas aeruginosa*: an assay in vitro and feedback inhibition by arginine. Eur. J. Biochem. 31: 290-295.

Halsey, R. J. (2003). Investigation of a putative glycogen debranching enzyme in *Corynebacterium glutamicum* encoded by *glgE*. Bachelor of Science (Honours) Thesis. University of Cape Town, Cape Town, South Africa.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol., 136: 557-580.

Hänßler, E., Müller, T., Oberger, N., Mzke, A., Kalinowski, J., Krämer, R., and Burkovski, A. (2007). FarR, a putative regulator of amino acid metabolism in *Corynebacterium glutamicum*. Appl. Microbiol. Biotechnol. 76: 625-632.

Haynes, J. A., and Britz, M. L. (1990). The effect of growth conditions of *Corynebacterium glutamicum* on the transformation frequency obtained by electroporation. J. Gen. Microbiol. 136: 255–263

Henke, W., Herdel, K. H., Schmorr, D., and Loening, S.A. (1997). Betaine improves the PCR amplification of GC-rich DNA sequences. Nucleic Acids Res. 25: 3957-3958.

Hermann, T. (2003). Industrial production of amino acids by coryneform bacteria. J. Biotechnol. 104: 155-172.

Hinds, J., Mahenthiralingam, E., Kempsell, K. E., Duncan, K., Stokes, R. W., Parish, T., and Stoker, N. G. (1999). Enhanced gene replacement in mycobacteria. Microbiology 145: 519-527.

Holtham, C. A., Jumel, K., Miller, C. M., Harding, S. E., Baumberg, S., and Stockley, P. G. (1999). Probing activation of the prokaryotic arginine transcriptional regulator using chimeric proteins. J. Mol. Biol. 289: 707-727.

Horn, J., Dietz-Schmidt, A., Zündorf, I., Garin, J., Dingermann, T., and Winckler, T. (2001). A *Dictyostelium* protein binds to distinct oligo(dA)·oligo(dT) DNA sequences in the C-module of the retrotransposable element DRE. Eu. J. Biochem. 265: 441-448.

Ikeda, M. (2003). Amino acid production processes. Adv. Biochem. Eng. Biotechnol. 79: 1-35.

Ikeda, M. (2005). L-Tryptophan production, *In*: Eggeling, L., Bott, M. (Eds.), Handbook of *Corynebacterium glutamicum*, CRC Press, Boca Raton, USA, pp. 489-509.

Ikeda, M., and Nakagawa, S. (2003). The *Corynebacterium genome*: features and impacts on biotechnological processes. Appl. Microbiol. Biotechnol. 62: 99-109.

Ikeda, M., Mitsuhashi, S., Tanaka, K., and Hayashi, M. (2009) Re-engineering of an L-arginine and L-citrulline producer of *Corynebacterium glutamicum*. Appl. Environ. Microbiol. [Published electronically ahead of print on 9 January 2009].

Ikeda, M., Okamoto, K., and Katsumata, R. (1999). Cloning of the transketolase gene and the effect of its dosage on aromatic amino acid production in *Corynebacterium glutamicum*. Appl. Microbiol. Biotechnol. 50: 375-378.

Ish-Horowicz, D., and Burke, J. F. (1981). Rapid and efficient cosmid cloning. Nucleic Acids Res. 9: 2989-2998.

Jackson, M., Raynaud, C., Lanéelle M-A., Guilhot, C., Laurent-Winter, C., Ensergueix, D., Gicquel, B., and Daffé, M. (1999). Inactivation of the antigen 85C gene profoundly affects the mycolate content and alters the permeability of the *Mycobacterium tuberculosis* cell envelope. Mol. Microbiol. 31: 1573-1587.

Jäger, W., Schäfer, A., Pühler, A., Labes, G., and Wohlleben, W. (1992). Expression of the *Bacillus subtilis sacB* gene leads to sucrose sensitivity in the gram-positive bacterium *Corynebacterium glutamicum* but not in *Streptomyces lividans*. J. Bacteriol. 174: 5462-5465.

Jakoby, M., Nolden, L., Meier-Wagner, J., Krämer, R., and Burkovski, A. (2000). AmtR, a global repressor in the nitrogen regulation system of *Corynebacterium glutamicum*. Mol. Microbiol. 37: 964-977.

Kalinowski, J. (2005). The genomes of amino-acid producing Corynebacteria. *In*: Eggeling, L., Bott, M. (Eds.), Handbook of *Corynebacterium glutamicum*, CRC Press, Boca Raton, USA, pp. 37-56.

Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., Dusch, N., Eggeling, L., Eikmanns, B. J., and other authors (2003). The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate derived amino acids and vitamins. J. Biotechnol. 104: 5-25.

Keilhauer, C., Eggeling, L., and Sahm, H. (1993). Isoleucine synthesis in *Corynebacterium glutamicum:* molecular analysis of the *ilvB-ilvN-ilvC* operon. J. Bacteriol. 175: 5595-5603.

Kelle, R., Hermann, T., and Bathe, B. (2005). L-Lysine production. *In*: Eggeling, L., Bott, M. (Eds.), Handbook of *Corynebacterium glutamicum*, CRC Press, Boca Raton, USA, pp. 465-488.

Kennerknecht, N., Sahm, H., Yen, M. R., Pátek, M., Saier, M. H. Jr, and Eggeling, L. (2002) Export of L-isoleucine from *Corynebacterium glutamicum:* a two-gene-encoded member of a new translocator family. J. Bacteriol. 184: 3947-3956.

Kimura, E. (2005). L-Glutamate production. *In*: Eggeling, L., Bott, M. (Eds.), Handbook of *Corynebacterium glutamicum*, CRC Press, Boca Raton, USA, pp. 439-463.

Kinoshita, S. (2005). A short history of the birth of the amino acid industry in Japan. *In*: Eggeling, L., Bott, M. (Eds.), Handbook of *Corynebacterium glutamicum*, CRC Press, Boca Raton, USA, pp. 3-5.

Kinoshita, S., Udaka, S., and Shimono, M. (1957). Studies on the amino acid fermentation. Part I. Production of L-glutamic acid by various microorganisms. J. Gen. Appl. Microbiol. 3: 193–205.

Kisumi, M., Kato, J., Suguira, M., and Chibata, I. (1971). Production of L-arginine by arginine hydroxamate-resistant mutants of *Bacillus subtilis*. Appl. Microbiol. 22: 987-991.

Kiupakis, A. K., and Reitzer, L. (2002). ArgR-independent induction and ArgR-dependent superinduction of the *astCADBE* operon in *Escherichia coli*. J. Bacteriol. 184: 2940-2950.

Kjeldsen, K.R., and Nielsen, J. (2008). *In silico* genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network. Biotechnol. Bioeng. doi:10.1002/bit.22067.

Knipp, M., and Vašák, M. (2006). A colorimetric 96-well microtitre plate assay for the determination of enzymatically formed citrulline. Anal. Biochem. 286: 257-264.

Kozmík, Z., Urbánek, P., and PaCes, V. (1990). Albumin improves formation and detection of some specific protein-DNA complexes in the mobility shift assay. Nucleic Acids Res. 18: 2198.

Krämer, R. (1994a). Secretion of amino acids by bacteria: physiology and mechanism. FEMS Microbiol. Rev. 13: 75-94.

Krämer, R. (1994b). Systems and mechanisms of amino acid uptake and excretion in prokaryotes. Arch. Microbiol. 162: 1-13.

Kruse, D., Krämer, R., Eggeling, L., Rieping, M., Pfefferle, W., Tchieu, J. H., Chung, Y. J., Saier, M. H. Jr., and Burkovski, A. (2002). Influence of threonine exporters on threonine production in *Escherichia coli*. Appl. Microbiol. Biotechnol. 59: 205-210.

Kumagai, H. (2000). Microbial production of amino acids in Japan. *In*: Scheper, T. (Ed.), Advances in Biochemical Engineering, vol. 69. Springer-Verlag, Berlin, Germany, pp. 71–85.

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

Lane, D. J., Pacew, B., Olsen, G. J., Stahl, D. A., Sogin, M. J., and Pace, N. R. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analysis. Proc. Natl. Acad. Sci. USA. 82: 6955-6959.

Lange, C., Rittman, D., Wendisch, V.F., Bott, M., and Sahm, H. (2003). Global expression profiling and physiological characterisation of *Corynebacterium glutamicum* grown in the presence of L-valine. Appl. Environ. Microbiol. 69: 2521-2532.

Larsen, R., Buist, G., Kuipers, O. P., and Kok, J. (2004). ArgR and AhrC are both required for regulation of arginine metabolism in *Lactococcus lactis*. J. Bacteriol. 186: 1147-1157.

Larsen, R., Kok, J., and Kuipers, O. P. (2005). Interaction between ArgR and AhrC controls regulation of arginine metabolism in *Lactococcus lactis*. J. Biol. Chem. 280: 19319-19330.

Lechevalier, M. P., and Lechevalier, H. A. (1970). Chemical composition as a criterion in the classification of aerobic actinomycetes. Int. J. Syst. Bacteriol. 20: 435-443.

Leuchtenberger, W. (1996). Amino acids – technical production and use. *In*: Rehm, H. J., and Reed, G. (Eds.), Biotechnology: Products of Primary Metabolism, vol. 6. VCH-Verlag, Weinheim, Germany, pp. 465-502.

Liebl, W. (1991). *Corynebacterium* nonmedical. *In*: Balows, A., Trüper, H. G., Dworkin, M., Harder., W., Schleifer, K. H. (Eds.), The Prokaryotes, 3rd Edition, Springer-Verlag, New York, USA, pp. 1157-1171.

Liebl, W. (2005). *Corynebacterium* taxonomy. *In*: Eggeling, L., Bott, M. (Eds.), Handbook of *Corynebacterium glutamicum*, CRC Press, Boca Raton, USA, pp. 9-34.

Liebl, W., Bayer, A., Schein, B., Stillner, U., and Schleifer, K. H. (1989). High efficiency electroporation of intact *Corynebacterium glutamicum* cells. FEMS Microbiol. Lett. 65: 299-304.

Lu, C. D. (2006). Pathways and regulation of bacterial arginine metabolism and perspectives for obtaining arginine overproducing strains. Appl. Microbiol. Biotechnol. 70: 261-272.

Lu, C. D., and Abdelal, A. T. (1999). Role of ArgR in activation of the *ast* operon, encoding enzymes of the arginine succinyltransferase pathway in *Salmonella typhimurium*. J. Bacteriol. 181: 1934-1938.

Lu, C. D., Yang, Z., and Li, W. (2004). Transcriptome analysis of the ArgR regulon in *Pseudomonas aeruginosa*. J. Bacteriol. 186: 3855–3861.

Lu, G., and Moriyama, E. N. (2004). Vector NTI, a balanced all-in-one sequence analysis suite. Brief. Bioinform. 5: 378-388.

Maas, W. K. (1994). The arginine repressor of *Escherichia coli*. Microbiol. Rev. 58: 631-640.

Maes, T., Vereecke, D., Ritsema, T., Corenlis, K., Thu, H. N., Van Montagu, M., Holsters, M., and Goethals, K. (2001). The *att* locus of *Rhodococcus fasciens* strain D188 is essential for full virulence on tobacco through the production of an autoregulatory compound. Mol. Microbiol. 42: 13-28.

Makarova, K. S., Mironov, A. A., and Gelfand, M. S. (2001). Conservation of the binding site for the arginine repressor in all bacterial lineages. Genome Biol. 2: research0013.1–research0013.8.

Martin, R. G., and Rosner, J. L. (2001). The AraC transcriptional activators. Curr. Opin. Microbiol. 4: 132–137.

Mateos, L. M., Pisabarro, A., Pátek, M., Malumbres, M., Guerrero, C., Eikmanss, B.J., Sahm, H., and Martin, J.F. (1994). Transcriptional analysis and regulatory signals of the *hom-thrB* cluster of *Brevibacterium lactofermentum*. J. Bacteriol. 176: 7362-7371.

Merino, E., and Yanofsky, C. (2005). Transcription attenuation: a highly conserved regulatory strategy used by bacteria. Trends Genet. 21: 260-264.

Milner, J. L., and Wood, J. M. (1987). Transmembrane amino acid fluxes in bacterial cells. CRC Crit. Rev. Biotechnol. 5: 1-47.

Mitsuhashi, S., Ohnishi, J., Hayashi, M., Ikeda, M. (2002). Physiological role of carbonic anhydrase in *Corynebacterium glutamicum*. *In*: Proc. Annu. Meeting Agric. Chem. Soc. Jpn., Sendai, Japan, 25-27 March, p 289.

Mitsuhashi, S., Ohnishi, J., Hayashi, M., Ikeda, M. (2003). A gene homologous to ß-type carbonic anhydrase is essential for the growth of *Corynebacterium glutamicum* under atmospheric conditions. Appl. Microbiol. Biotechnol. 63: 592-601.

Miwa, K., Matsui, H., Terabe, M., Nakamori, S., Sano, K., and Momose, H. (1984). Cryptic plasmids in glutamic acid-producing bacteria. Agric. Biol. Chem. 48: 2901-2903.

Morin, A., Huysveld, N., Braun, F., Dimova, D., Sakanyan, V., and Charlier, D. (2003). Hyperthermophilic *Thermotoga* arginine repressor binding to full-length cognate and heterologous arginine operators and to half-site targets. J. Mol. Biol. 332: 537-553.

Morris, S. M. J., Loscalzo, J., Bier, D., and Souba, W. W. (Eds.) (2004). Arginine metabolism: enzymology, nutrition and clinical significance. The Journal of Nutrition, Bermuda, UK , pp. 1-66.

Münch, R., Hiller, K., Barg, H., Heldt, D., Linz, S., Wingender, E., and Jahn, D. (2003). PRODORIC: prokaryotic database of gene regulation. Nucleic Acids Res. 31: 266-269.

Münch, R., Hiller, K., Grote, A., Scheer, M., Klein, J., Schobert, M., and Jahn, D. (2005) Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes. Bioinformatics 21: 4187-4189.

Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J. B., Engel, A., and Fujiyoshi, Y. (2000). Structural determinants of water permeation through aquaporin-1. Nature 407: 5999-6005.

Muro-Pastor, M. I., Strauss, J., Ramón, A., and Scazzocchio, C. (2004). A Paradoxical Mutant GATA Factor. Eukaryot. Cell 3: 393-405.

Nakagawa, S. (2002). The complete genome sequencing of *Corynebacterium glutamicum* ATCC 13032. In: Proceedings of the 9th International Symposium on Genetics of Industrial Microorganisms, Gyeongju, Korea, 1-5 July 2002, p 21.

Nakayama, K., and Yoshida, H. (16 Jan. 1973). Process for producing L-arginine by fermentation. United States Patent 3 849 250.

Ni, J., Sakanyan, V., Charlier, D., Glansdorff, N., and Van Duyne, G. D. (1999). Structure of the arginine repressor from *Bacillus stearothermophilus*. Nat. Struct. Biol. 6: 427-434.

Nicoloff, H., Arsène-Ploetze, F., Malandain, C., Kleerebezem, M., and Bringel, F. (2004). Two arginine repressors regulate arginine biosynthesis in *Lactobacillus plantarum*. J. Bacteriol. 186: 6059-6069.

Nicoloff, H., Hubert, J.C., and Bringel, F. (2000). In *Lactobacillus plantarum*, carbamoyl phosphate is synthesized by two carbamoyl-phosphate synthetases (CPS): carbon dioxide differentiates the arginine-repressed from the pyrimidine-regulated CPS. J. Bacteriol. 182: 3416-3422.

Nishimura, T., Teramoto, H., Vertès A.A., Inui, M., and Yukawa, H. (2008). ArnR, a novel transcriptional regulator, represses expression of the *narKGHJI* operon in *Corynebacterium glutamicum*. J. Bacteriol. 190: 3264-3273.

Okazaki, H., Kanzaki, T., Doi, M., Sumino, Y., and Fukuda, H. (1967). L-glutamate acid fermentation. Part II. The production of L-glutamate acid by an oleic acid-requiring mutant. Agric. Biol. Chem. 31: 1314-1317.

Okino, S., Noburyu, R., Suda, M., Jojima, T., Inui, M., and Yukawa, H. (2008). An efficient succinic acid production process in a metabolically engineered C*orynebacterium glutamicum* strain. Appl. Microbiol. Biotechnol. 81: 459-464.

Parish, T., and Stoker, N. G. (2000). Use of a flexible cassette method to generate a double unmarked *Mycobacterium tuberculosis tlyA plcABC* mutant by gene replacement. Microbiology. 146: 1969-1975.

Parish, T., Gordhan, B. G., McAdam, R. A., Duncan, K., Mizrahi, V., and Stoker, N. G. (1999). Production of mutants in amino acid biosynthesis genes of *Mycobacterium tuberculosis* by homologous recombination. Microbiology. 145: 3497-3503.

Park, S., Lu, C., and Abdelal, A. T. (1997). Cloning and characterization of *argR*, a gene that participates in regulation of arginine biosynthesis and catabolism in *Pseudomonas aeruginosa* PAO1. J. Bacteriol. 179: 5300-5308.

Pátek, M. (2005). Regulation of gene expression. *In*: Eggeling, L., Bott, M. (Eds.), Handbook of *Corynebacterium glutamicum*, CRC Press, Boca Raton, USA, pp. 81-98.

Pátek, M., Eikmanns, B. J., Pátek J., Sahm, H. (1996). Promoters from *Corynebacterium glutamicum*: cloning, molecular analysis and search for a consensus motif. Microbiology 142: 1297-1309.

Pátek, M., Nešvera, J., Guyonvarch, A., Reyes, O., and Leblon, G. (2003). Promoters of *Corynebacterium glutamicum*. J. Biotechnol. 104: 311-323.

Peters-Wendisch, P. G., Eikmanns, B. J., Thierbach, G., Bachmann, B., and Sahm, H. (1993). Phosphoenolpyruvate carboxylase in *Corynebacterium glutamicum* is dispensable for growth and lysine production. FEMS Microbiol. Lett. 112: 269-274.

Peters-Wendisch, P. G., Schiel, B., Wendisch, V. F., Katsoulidis, E., Möckel, B., Sahm, H., and Eikmanns, B. J. (2001). Pyruvate carboxylase as a major bottleneck for glutamate and lysine production in *Corynebacterium glutamicum*. J. Mol. Microbiol. Biotechnol. 3: 295-300.

Pfefferle, W., Möckel, B., Bathe, B., and Marx, A. (2003). Biotechnological manufacture of lysine. Adv. Biochem. Engn. Biotechnol. 79: 59-112.

Puech, V., Bayan, N., Salim, K., Leblon, G., and Daffé, M. (2000). Characterization of the *in vivo* acceptors of the mycoloyl residues transferred by the corynebacterial PS1 and related mycobacterial antigens 85. Mol. Microbiol. 35: 1026-1041.

Puech, V., Chami, M., Lemassu, A., Lanéelle., M. A., Schiffler, B., Gounon., P., Bayan, N., Benz, R., and Daffé, M. (2001). Structure of the cell envelope of corynebacteria: importance

of the non-covalently bound lipids in the formation of the cell wall permeability barrier and fracture plane. Microbiology 147: 1365-1382.

Radmacher, E., Vaitsikova, A., Burger, U., Krumbach, K., Sahm, H., and Eggeling, L. (2002). Linking central metabolism with increased pathway flux: L-valine accumulation by *Corynebacterium glutamicum*. Appl. Environ. Microbiol. 68: 2246-2250.

Rajagopal, B. S., DePonte III, J., Tuchman, M., and Malamy, M. H. (1998). Use of inducible feedback resistant *N*-acetylglutamate synthetase (*argA*) genes for enhanced arginine biosynthesis by genetically engineered *Escherichia coli* K-12 strains. Appl. Environ. Microbiol. 64: 1805-1811.

Riedl, C., Ritmann, D., Dangel, P, Möckel, B., Sahm, H., and Eikmanns B. J. (2001). Characterization, expression and inactivation of the phosphoenolpyruvate carboxykinase gene from *Corynebacterium glutamicum* and significance of the enzyme for growth and amino acid production. J. Mol. Microbiol. Biotechnol. 3: 573-583.

Rigali, S., Derouaux, A., Giannotta, F., and Dusart, J. (2002). Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR and YtrA subfamilies. J. Biol. Chem. 227: 12507-12515.

Robertson, C. A., and Nash, H. A. (1998). Bending of the bacteriophage lambda attachment site by *Escherichia coli* integration host factor. J. Biol. Chem. 263: 3554-3557.

Rodríguez-García, A., de la Fuente, A., Pérez-Redondo, R., Martín, J. F., and Liras, P. (2000). Characterization and expression of the arginine biosynthesis gene cluster of *Streptomyces clavuligerus*. J. Mol. Microbiol. Biotechnol. 2: 543-550.

Rodríguez-García, A., Ludovice, M., Martín, J. F., and Liras, P. (1997). Arginine boxes and the *argR* gene in *Streptomyces clavuligerus*: evidence for a clear regulation of the arginine pathway. Mol. Microbiol. 25: 219-228.

Rosenberg, H., Ennor, A. H., and Morrison, J. F. (1955). The estimation of arginine. Biochem J. 63: 153-9.

Russell, W. M., and Klaenhammer, T. R. (2001). Efficient system for directed integration into the *Lactobacillus acidophilus* and *Lactobacillus gasseri* chromosomes via homologous recombination. Appl. Environ. Microbiol. 67: 4361-4364.

Sakaguchi, S. (1925). On a new colour reaction of protein and arginine. J. Biochem. 5: 25.

Sakanyan, V., Legrain, C., Charlier, D., and Kochikyan, A. (1993). *N*-acetylglutamate-5-phosphotransferase of the thermophilic bacterium *Bacillus stearothermophilus:* nucleotide sequence of the gene and enzyme characterization. Genetika. 29: 556-570.

Sakanyan, V., Petrosyan, P., Lecocq, M., Boyen, A., Legrain., C., Demarez, M., Hallet, J., and Glansdorff, N. (1996). Genes and enzymes of the acetyl cycle of arginine biosynthesis in *Corynebacterium glutamicum*: enzyme evolution in the early step of the arginine pathway. Microbiology 142: 99-108.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (2001). Molecular cloning: a laboratory manual. 3rd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, USA.

Schäfer, A., Tauch, A., Droste, N., Pühler, A., and Kalinowski, J. (1997). The *Corynebacterium glutamicum cglIM* gene encoding a 5-cytosine methyltransferase enzyme confers a specific DNA methylation pattern in a McrBC-deficient *Escherichia coli* strain. Gene 203: 95-101.

Schäfer, A., Tauch, A., Jäger, W., Kalinowski, J., Thierbach, G., and Pühler, A. (1994). Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. Gene 145: 69-73.

Schäffer, A. A., Aravind, L., Madden, T. L., Shavirin. S., Spouge, J. L., Wolf, Y. I., Koonin, E. V., and Altschul, S. F. (2001). Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Res. 29: 2994-3005.

Schell, M. A. (1993). Molecular biology of the LysR family of transcriptional regulators. Annu. Rev. Microbiol. 47: 597-626.

Schleifer, K. H., and Kandler, O. (1972). Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriol. Rev. 35: 407-477.

Sekine, H., Shimada, T., Hayashi, C., Ishiguro, A., Tomita, F., and Yokota, A. (2001). H^+-ATPase defect in *Corynebacterium glutamicum* abolishes glutamic acid production with enhancement of glucose consumption rate. Appl. Microbiol. Biotechnol. 57: 534-540.

Sekino, Y., Bruner, S. D., and Verdiene, G. L. (2000). Selective inhibition of Herpes Simplex Virus type-1 uracil-DNA glycosylase by designed substrate analogs. J. Biol. Chem. 275: 36506-36508.

Shi, D., Morizono, H., Yu, X., Roth, L., Caldovic, L., Allewell, N. M., Malamy, M. H., and Tuchmann, M. (2005). Crystal structure of *N*-acetylornithine transcarbamylase from *Xanthomonas campestris*: a novel enzyme in a new arginine biosynthetic pathway found in several eubacteria. J. Biol. Chem. 280: 14366-14369.

Shiio, I., Otsuka, S. I., and Tsunoda, T. (1960). Glutamic acid formation on glucose by bacteria. J. Biochem. 47: 414-421.

Shimizu, H., Tanaka, H., Nakato, A., Nagahisa, K., Kimura, E., and Shioya, S. (2003). Effects of the changes in enzyme activities on metabolic flux redistribution around the 2-oxoglutarate branch in glutamate production by *Corynebacterium glutamicum*. Bioprocess Biosys. Eng. 25: 291-298.

Simic, P., Sahm, H., Eggeling, L. (2001). L-Threonine export: use of peptides to identify a new translocator from *Corynebacterium glutamicum*. J. Bacteriol. 183: 5317-5324.

Simic, P., Willuhn, J., Sahm, H., and Eggeling, L. (2002). Identification of *glyA* (encoding serine hydroxymethyltransferase) and its use together with the exporter ThrE to increase L-threonine accumulation by *Corynebacterium glutamicum*. Appl. Environ. Microbiol. 68: 3321-3327.

Smith, M. C., Mountain, A., and Baumberg, S. (1986). Cloning in *Escherichia coli* of a *Bacillus subtilis* arginine repressor gene through its ability to confer structural stability on a fragment carrying genes of arginine biosynthesis. Mol. Gen. Genet. 205: 176-182.

Song, H., Wang, H., Gigot, D., Dimova, D., Sakanayn, V., Glansdorff, N., and Charlier, D. (2002). Transcription regulation in thermophilic bacteria: high resolution contact probing of *Bacillus stearothermophilus* and *Thermotoga neapolitana* arginine repressor-operator interactions. J. Mol. Biol. 315: 255-274.

Souter, A., and Baumberg, S. (1996). Implication of repression system, homologous to those of other bacteria, in the control of the arginine biosynthesis genes in *Streptomyces coelicolor*. Mol. Gen. Genet. 251: 245-251.

Stackebrandt, E., Rainey, F. A., and Ward-Rainey, N. L. (1997). Proposal of a new hierarchic classification system, Actinobacteria classis nov. Int. J. Syst. Bacteriol. 47: 479-491.

Stirling, C. J., Szatmari, G., Stewart, G., Smith, M. C., and Sherratt, D. J. (1988). The arginine repressor in essential for plasmid-stabilizing site-specific recombination at the ColE1 *cer* locus. EMBO J. 7: 4389-4395.

Studier, F. W., and Moffat, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. 189: 113-130.

Sunnerhagen, M., Nilges, M., Otting, G., and Carey, J. (1997). Solution structure of the DNA-binding domain and model for the complex of multifunctional hexameric arginine repressor with DNA. Nat. Struct. Biol. 4: 819-826.

Tauch, A. (2005). Native plasmids of amino-acid producing corynebacteria, *In*: Eggeling, L., Bott, M. (Eds.), Handbook of *Corynebacterium glutamicum*. CRC Press, Boca Raton, USA, pp. 57-80.

Tauch, A., Götker, S., Pühler, A., Kalinowski, J., and Thierbach, G. (2002). The alanine racemase gene *alr* is an alternative to antibiotic resistance genes in cloning systems for industrial *C. glutamicum* strains. J. Biotechnol. 99: 79-91.

Tauch, A., Hermann, T., Burkovski, A., Krämer, R., Pühler, A., and Kalinowski, J. (1998). Isoleucine uptake in *Corynebacterium glutamicum* ATCC 13032 is directed by the *brnQ* gene product. Arch. Microbiol. 169: 303-312.

Tian, G., Lim, D., Carey, J., and Maas, W. K. (1992). Binding of the arginine repressor of *Escherichia coli* K12 to its operator sites. J. Mol. Biol. 226: 387-397.

Tian, G., Lim, D., Oppenheim, J. D., and Maas, W. K. (1994). Explanation for different types of regulation of arginine biosynthesis in *Escherichia coli* B and *Escherichia coli* K12 caused by a difference between their arginine repressors. J. Mol. Biol. 235: 221-230.

Tomlinson, G., and Viswanatha, T. (1974). Determination of the arginine content of proteins by the Sakaguchi procedure. Anal. Biochem. 60: 15-24.

Tuchman, M., Rajagopal, B. S., McCann, M. T., and Malamy, M. H. (1997). Enhanced production of arginine and urea by genetically engineered *Escherichia coli* K-12 strains. Appl. Environ. Microbiol. 63: 33-38.

Turner, M. A., Simpson, A., Mcinnes, R. R., and Howell, P. L. (1997) Human argininosuccinate lyase: a structural basis for intragenic complementation. Proc. Natl. Acad. Sci. 94: 9063–9068

Udaka, S. (1966). Pathway-specific pattern of control of arginine biosynthesis in bacteria. J. Bacteriol. 91: 617-621.

Utagawa, T. (2004). Production of arginine by fermentation. J. Nutr. 134: 2854S-2857S.

van de Casteele, M., Demarez, M., Legrain, C., Glansdorff, N., and Piérard, A. (1990). Pathways of arginine biosynthesis in extreme thermophilic archaeo- and eubacteria. J. Gen. Microbiol. 136: 1177-1183.

van der Rest, M. E., Lange, C., and Molenaar, D. (1999). A heat shock following electroporation induces highly efficient transformation of *Corynebacterium glutamicum* with xenogenic plasmid DNA. Appl. Microbiol. Biotechnol. 52: 541-545.

Vašicová, P., Abrhámová, Z., Nešvera, J., Pátek, M., Sahm, H., and Eikmanns, B. (1998). Integrating and autonomously replicating vectors for analysis of promoters in *Corynebacterium glutamicum*. Biotechnol. Techniques 12: 743-746.

Vertès, A. A., Inui, M., and Yukawa, H. (2005). Manipulating Corynebacteria, from individual genes to chromosomes. Appl. Environ. Microbiol. 71: 7633-7642.

Volcani, B. E., and Snell, E. E. (1948). The effect of canavanine, arginine and related compounds on the growth of bacteria. J. Biol. Chem. 174: 893-902.

Vrljić, M., Eggeling, L., and Sahm, H. (1996). A new type of transporter with a new type of cellular function: L-lysine export from *Corynebacterium glutamicum*. Mol. Microbiol. 22: 815-826.

Vrljić, M., Garg, J., Bellman, A., Wachu, S. Freudl, R. Malecki, M. J., Sahm, M., Kozina, V. J., Eggeling, L., and Saier, M. H. Jr. (1999). The LysE superfamily: topology of the lysine exporter LysE of *Corynebacterium glutamicum*, a paradyme for a novel superfamily of transmembrane solute translocators. J. Mol. Microbiol. Biotechnol. 1: 327-336.

Wang, H., Glansdorff, N., and Charlier, N. (1998). The arginine repressor of *Escherichia coli* K-12 makes direct contacts to minor and major groove determinants of the operators. J. Mol. Biol. 277: 805-824.

Weerasinghe, J. P., Dong, T., Schertzberg, M. R., Kirchhif, M. G., Sun, Y., and Schellhorn, H. E. (2006). Stationary phase expression of the arginine biosynthetic operon *argCBH* in *Escherichia coli*. BMC Microbiol. 6: 14.

Wickner, W., Driessen, A. J., and Hartl, F. U. (1991). The enzymology of protein translocation across the *Escherichia coli* plasma membrane. Annu. Rev. Biochem. 60: 101-124.

Willey, J. M., Sherwood, L. M., and Woolverton, C. J. (2007). Microbial growth. *In*: Prescott, Harley and Klein's Microbiology, 7th Edition. McGraw-Hill Higher Education, New York, USA, pp. 119-148.

Willis, L. B., Lessard, P. A., and Sinskey, A. J. (2005). Synthesis of L-threonine and branched-chain amino acids. *In*: Eggeling, L., Bott, M. (Eds.), Handbook of *Corynebacterium glutamicum*, CRC Press, Boca Raton, USA, pp. 511-531.

Wolfe, P. B., Wickner, W., and Goodman, J. M. (1983). Sequence of the leader peptidase gene of *Escherichia coli* and the orientation of leader peptidase in the bacterial envelope. J. Biol. Chem. 258: 12073-12080.

Xu, Y., Labedan, B., and Glansdorff, N. (2007). Surprising arginine biosynthesis: a reappraisal of the enzymology and evolution of the pathway in microorganisms. Microbiol. Mol. Rev. 71: 36-47.

Xue, G., Johnson, J. S., and Dalrymple, B. P. (1999). High osmolarity improves the electro-transformation efficiency of the gram-positive bacteria *Bacillus subtilis* and *Bacillus licheniformis*. J. Microbiol. Methods. 34: 183-191.

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.

Yokota, A., and Lindley, N. D. (2005). Central metabolism: sugar uptake and conversion. *In*: Eggeling, L., Bott, M. (Eds.), Handbook of *Corynebacterium glutamicum*. CRC Press, Boca Raton, USA, pp. 215-240.

Zakataeva, N. P., Aleshin, V. V., Tokmakova, L. L., Troshin, P. V., and Livshits, V. A. (1999). The novel transmembrane *Escherichia coli* proteins involved in the amino acid efflux. FEBS Let. 452: 228-232.

Zakova, N., and Szatmari, G. B. (1995). Site-specific recombination between ColE1 cer and NTP16 nmr sites in vivo. Mol. Gen. Genet. 247: 509-514.

Zidwick, N. I., Keller, J. G., and Rogers, P. (1984) Regulation and coupling of *argECBH* mRNA and enzyme synthesis in cell extracts of *Escherichia coli*. J. Bacteriol. 159: 640-646.

Zittrich, S., and Krämer, R. (1994). Quantitative discrimination of carrier-mediated excretion of isoleucine from uptake and diffusion in *Corynebacterium glutamicum*. J. Bacteriol. 176: 6892-6899.

Zucker, M. (2003). Mfold web server for nucleic acid folding and hybridisation prediction. Nucleic Acids Res. 31: 3406-3415.

Appendix A

Bacterial strains and plasmids used

Strain or plasmid	Relevant characteristics	Reference or source
Strains		
E. coli		
DH5α	F$^-$ Φ80d*lacZ* ΔM15 Δ(*lacZYA-argF*)U169 *deoR recA1 endA1 hsdR17* (rK$^-$, mK$^+$) *phoA supE44* λ$^-$ *thi-1 gyrA96 relA1*	Hanahan (1983)
BL21 (DE3)	*ompT hsdS*$_B$(r$_B^-$, m$_B^-$) *gal dcm* (DE3)	Studier and Moffat (1986)
C. glutamicum		
ATCC 13032	Type strain; Nx^R	Abe *et al.* (1967)
9A	ATCC 13032 pARGR-INT integration mutant; Kan^R Nx^R	Nina Smazanowiz
SCOH	ATCC 13032 *argH* pARGHKO integration mutant; Hyg^R Kan^R Nx^R	This study
SCOR	ATCC 13032 *argR* pARGRKO integration mutant; Hyg^R Kan^R Nx^R	This study
SCOR-D3	ATCC 13032 *argR* pARGR-MS integration mutant; Kan^R Nx^R	This study
DCOR-D3	ATCC 13032 *ΔargR* DCO mutant; Nx^R	This study
ATCC 21831	L-arginine overproducer; Nx^R	Nakayama *et al.* (1973)

Appendix A

Plasmids

pTZ57R/T	Cloning vector with 3'-ddT overhangs (*lacZα* $ori_{E.\ coli}$); Amp^R	Fermentas
pBlueScript SK(+)	Cloning vector (*lacZα* $ori_{E.\ coli}$); Amp^R	Stratagene
pUC19	Subcloning vector used for pLYSE construction (*lacZα* $ori_{E.\ coli}$); Amp^R	Yanisch-perron *et al.* (1985)
p2NIL	Gene manipulation vector ($ori_{E.\ coli}$); Kan^R	Parish and Stoker (2000)
pGOAL19	Vector harbouring a *Pac*I DNA fragment in possession of a marker gene cassette (P_{Ag85}-*lacZ* P_{Hsp60}-*sacB* $ori_{E.\ coli}$); Hyg^R Kan^R	Parish and Stoker (2000)
pARGHKO	*C. glutamicum ΔargH* suicide delivery vector harbouring the pGOAL19 *Pac*I marker gene cassette (P_{Ag85}-*lacZ* P_{Hsp60}-*sacB* $ori_{E.\ coli}$); Hyg^R Kan^R	This study
pARGRKO	*C. glutamicum ΔargR* suicide delivery vector harbouring the pGOAL19 *Pac*I marker gene cassette (P_{Ag85}-*lacZ* P_{Hsp60}-*sacB* $ori_{E.\ coli}$); Hyg^R Kan^R	This study
pK19*mobsacB*	Mobilisable *E. coli-C. glutamicum* pK19-derived vector used for the construction of *C. glutamicum* chromosomal insertion and deletion mutants (*sacB lacZα* $ori_{E.\ coli}$ *oriT*); Kan^R	Schäfer *et al.* (1994)
pARGR-MS	*C. glutamicum ΔargR* suicide delivery vector derived from pK19*mobsacB* (*sacB*	This study

	lacZα ori$_{E.\ coli}$ *oriT*); Kan^R	
pEKEX2	*E. coli-C. glutamicum* shuttle vector containing a pK18-derived MCS preceded by the IPTG-inducible *tac* promoter for controlled gene expression (P$_{tac}$ *lacI*^{q} *ori*$_{E.\ coli}$ *ori*$_{C.\ glutamiucm}$); Kan^R	Eikmanns *et al.* (1991)
pCAF21404-argC	pEKEX2 containing a DNA fragment in possession of the *C. glutamicum* ATCC 13032 *CAF21404-argC* intergenic region, a 3'-region of the *CAF21404* gene and a 5'-region of the *argC* gene (P$_{tac}$ *lacI*^{q} *ori*$_{E.\ coli}$ *ori*$_{C.\ glutamiucm}$); Kan^R	This study
pUC19	Subcloning vector used for pLYSE construction (*lacZα ori*$_{E.\ coli}$); Amp^R	Yanisch-perron *et al.* (1985)
pLYSE	pEKEX2 harbouring the *C. glutamicum* ATCC 13032 *lysE* gene (P$_{tac}$ *lacI*^{q} *ori*$_{E.\ coli}$ *ori*$_{C.\ glutamiucm}$); Kan^R	This study
pET22b(+)	IPTG-inducible *E. coli* protein expression vector; contains a N-terminal *pelB* leader sequence and a C-terminal region encoding a hexahistadyl tag (*P*$_{T7}$ *lacI*^{q} *ori*$_{E.\ coli}$); Amp^R	Novagen
pARGRP	pET22b(+) containing the *C. glutamicum* ATCC 13032 *argR* gene fused in-frame to sequences encoding a *pelB* leader sequence and a hexahistadyl tag (*P*$_{T7}$ *lacI*^{q} *ori*$_{E.\ coli}$); Amp^R	This study

Appendix B

CGXII minimal medium

MOPS-buffered minimal medium CGXII, based on the protocol of Keilhauer *et al.* (1993), was made according to the method of Eggeling and Reyes (2005).

Biotin solution:

Biotin (Sigma-Aldrich) was added to 10 ml of dH_2O at a final concentration of 200 µg/ml and dissolved by heating. The solution was filter sterilised and stored at 4 °C.

$CaCl_2$ solution:

$CaCl_2$ was added to 100 ml of dH_2O to a final concentration of 10 mg/ml.

Glucose solution:

50 g of glucose was added to 60 ml of dH_2O and dissolved by heating under stirring. After solubilisation, dH_2O was added to a final volume of 100 ml and the solution autoclaved.

Protocatechuic acid:

300 mg of protocatechuic acid (Sigma-Aldrich) was added to 8 ml of dH_2O. The mixture was dissolved by the addition of 1 ml of 10 *N* NaOH and filter sterilised. This solution was stored at 4 °C and made fresh monthly.

Trace elements solution:

1 g	$FeSO_4 \cdot 7H_2O$
1 g	$MnSO_4 \cdot H_2O$
0.1 g	$ZnSO_4 \cdot 7H_2O$
0.02 g	$CuSO_4 \cdot 5H_2O$
0.002 g	$NiCl_2 \cdot 6H_2O$

The trace elements solution was made up to 100 ml with dH_2O. A few drops of concentrated HCl were added to the solution under gentle stirring until complete solubilisation had occurred. The final mixture was filter sterilised.

Appendix B

CGXII salts solution (per litre):

0.25 g	$MgSO_4 \cdot 7H_2O$
1 g	KH_2PO_4
1 g	K_2HPO_4
5 g	urea
20 g	$(NH_4)_2SO_4$
42 g	MOPS
15 g	minimal medium agar (Oxoid) (for solid media)

Add 800 ml of dH_2O

Add 1 ml $CaCl_2$ solution

Adjust to pH 7 with 10 *N* NaOH

Adjust the final volume to 920 ml with dH_2O

Autoclave

CGXII minimal medium (per litre):

920 ml	CGXII salts solution
80 ml	50% glucose
1 ml	biotin solution
1 ml	protocatechuic acid
1 ml	trace elements solution

Appendix C

Oligonucleotide primers and PCR conditions

Primer Name		Sequence [a] (5' to 3')	Product size (bp)	T_a [b] (°C)	Function / Relevant Characteristics
General use					
M13	F	CGC CAG GGT TTT CCC AGT CAC GAC	Variable	50	Used for confirmation and sequencing of insert DNA in pBlueScript SK(+) and pTZ57R/T
	R	GAG CGG ATA ACA ATT TCA CAC AGG			
T7	F	GTA ATA CGA CTC ACT ATA GGG C	Variable	55	Used for confirmation and sequencing of insert DNA in pET-22b(+)
	R	GCT AGT TAT TGC TCA GCG G			
Suicide plasmid construction and mutant confirmation					
argH front	F	TCC GTA CTG CAG GCT ATT GAA G	1004	51	Amplification of the *argH* front fragment with a 5'-*Pst*I site and a 3'-*Sal*I site
	R	AGG GCG TCG ACA AGC TCG			
argH back	F	CCT GCA GGA AGA TAA GGA TCC AA	1032	51	Amplification of the *argH* back fragment with a 5'-*Bam*HI site and 3'-*Kpn*I site
	R	CGG GTA CCG GCT TGG GTT			
argR fragment	F	GGA GCG TCG ACT CGA GG	wt: 2091 *ΔargR*:	50	Amplification of the *argR* fragment with the introduction of a 5'-*Sal*I
	R	AAT AGC GCG GAT CCA			

		GAC G	1773		site and a 3'-*Bam*HI site
Mutant confirmation					
argH junction	F	TTG GCC TCC AAG GCA CAC G	wt: 1279	52	Anneal to the regions flanking the deletion portion of the *ΔargH* fragment; used to distinguish between *argH* wildtype and mutant alleles
	R	CCG CGT GAG CGG TAC TTG	*ΔargH*: 716		
argH probe	F	CTG ATG AGG ATG TGC ACG G	wt: 1559	51	Amplification of the *argH* *Bgl*II-internal region for use in DIG-labelling and Southern hybridisation of SCOH
	R	GGA TGG GAA CGC GAA ATC G	*ΔargH*: 996		
argR junction	F	CGT GCG GAA TTC GTG GAG C	wt: 1236	53	Anneal to regions flanking the deletion in the *ΔargR* fragment; used to distinguish between wildtype and mutant alleles of *argR*
	R	CGA TGG ACT CAG CTG CAC C	*ΔargR*: 918		
argR probe	F	GTA TTG TCC CGC TAC GTG G	wt: 1496	51	Amplification of the *argR* *Pvu*II-internal region for use in DIG-labelling and Southern hybridisation of SCOR
	R	CAT GTT CTC TCC ACC CTG G	*ΔargR*: 1178		
p2NIL	F_1	CCA GTG CTG CAA TGA TAC C	Variable	48	Confirmation of insert in p2NIL: p2NIL-F_1 binds upstream of the *Pst*I site,
	F_2	TAC CGC TGT TGA GAT			

		CCA G			
	R	ATC ACG ACT GTG CTG GTC			p2NIL-F_2 binds upstream of the *Bam*HI site and p2NIL-R binds downstream of the *Kpn*I site; also used for confirmation of the pARGHKO and pARGRKO integration events
sacB	F	CTT TAC TAC CGC ACT GCT GG	1290	50	Amplification of a *sacB* internal fragment; used for confirmation of the *Pac*I marker gene cassette present in potential SCO mutants
	R	AGG AAG CTT GGC GCA AAC G			
pK19	F	TAA GCC CAC TGC AAG CTA CC	Variable	50	Used for the confirmation of insert in the pK19*mobsacB* MCS; also used for the confirmation of pARGR-INT integration in *C. glutamicum* 9A together with the *argF-argR* F and *argR-argG* R primers
	R	AGC AAC GCG GCC TTT TTA CG			
argR-MS	F	TAA GCT TTC AAC TCG TAC TCG C	wt: 2113	49	Amplification of the *argR*-MS fragment with the introduction of a 5'-*Hind*III site and a 3'-*Bam*HI site
	R	GAT CCA GAC GTT CTG GTCG	*ΔargR*: 1795		

argR gene	F	ATG TCC CTT GGC TCA ACC	wt: 522	47	Used for detection of the *argR* wildtype and mutant alleles; DIG-labelled product used for SCOR-D3 and DCOR-D3 mutant confirmation
	R	GGC GCT TTA AGT GGT GC	*ΔargR*: 204		

Semi-quantitative reverse transcriptase PCR

CAF21404-argC	F	AGG ACA CCA CCG AAG AAA CC	953	50	Amplification of the *CAF21404-argC* intergenic region
	R	AGA AGA GCC AAG GTT GCA CC			
argC-argJ	F	GGT TCA AGC GGG ACT TAT CG	1037	50	Amplification of the *argC-argJ* intergenic region
	R	GCG GTC AGC TGA TCA ATA CC			
argJ-argB	F	GCT TGA CCA CTG ATG CAT CC	999	50	Amplification of the *argJ-argB* intergenic region
	R	CAA CTA AAT CGC GAC CGA CC			
argB-argD	F	TGT GGA AGG TCT GTA CAC CG	777	50	Amplification of the *argB-argD* intergenic region
	R	AAC GCT TCA CGC TTG TCT GG			
argD-argF	F	GGA TGT GGC TGC TAT CTT CC	829	50	Amplification of the *argD-argF* intergenic region; also used for confirmation of the pARGRKO integration event in SCOR
	R	TCG AAG GAG AAG CGA GTA CG			

Appendix C

argF-argR	F	TTC CGG TAG CTC ACA GAT GG	927	50	Amplification of the *argF-argR* intergenic region
	R	AGA CAG TTG TAC CTG GCT GG			
argR-argG	F	GAT CTC GAT GAA CTC GGT GC	783	50	Amplification of the *argR-argG* intergenic region; also used for confirmation of the pARGHKO integration event in SCOH
	R	CAG AAA CCA GTG GGT ACT GC			
argG-argH	F	CAT TGA TTC CAC CCA GGA GC	1036	50	Amplification of the *argG-argH* intergenic region; also used for confirmation of the pARGRKO integration event in SCOR
	R	GCA CGA AGG CGG TTT CAG AT			
argH-CAF21413	F	GCA GAA GAA GAA CCC TGA CG	949	50	Amplification of the *argH-CAF21413* intergenic region
	R	TGG GAA CGC GAA ATC GTA CG			
CAF21413-CAF21414	F	CCA TCA TCC CAG AGC TTA GC	1016	50	Amplification of the *CAF21413-CAF21414* intergenic region; also used for confirmation of the pARGHKO integration event in SCOH
	R	GGT ATC AGC ATA TGC CGT GG			
	R	GCC ATC ATG GCT CGT GAC TA			
16S	F27	AGA GTT TGA TCC TGG CTC AG	~1600	50	Amplification of an internal fragment of the 16S gene (Alm *et al.*, 1996; Lane *et al.*, 1985)
	R5	GGG GTT GCG CTC GTT G			

Appendix C

Construction of pCAF21404-ARGC, pLYSE and pARGRP

CAF21404-argC fragment	F	CAC TGA TTC GTG TGT CTG C	491	50	Amplification of the intergenic region preceding the *argC* gene in addition to a 3'-region of the *CAF21404* gene and a 5'-region of the *argC* gene
	R	CGC GAC TTA ATC AGC TGG			
lysE gene	F	TAC TTC CAT AGG TCA CGA TGG	772	49	Amplification of the *lysE* gene
	R	GGA AGC GGG TTT ACA ATA C			
argR-P	F	CCG GAT CCA GAC ATG TCC CTT GG	346	56	Amplification of the *argR* gene with a 5'-*Bam*HI site, a 3'-*Eco*RI site and a non-functional stop codon
	R	GTG GAA TTC GGG GCG CTT GAA GTG			

Preparation of potential ArgR^{His}-DNA binding regions

argC promoter	F	GCT GGC TAC TAA AAA TTC ATG C	147	48	Amplification of the *CAF21404-argC* intergenic region
	R	TAC ACG TTA TGC ATG ATC ATG C			
argG promoter	F	AAG CGC CCC TAG TTC AAG G	185	50	Amplification of the *argR-argG* intergenic region
	R	CGA TGC GGT TAG TCA TGA GG			
carA promoter	F	ACT CAC GGT GAC TGT TGT CG	141	50	Amplification of the *pyrR-*

	R	GCT TAA CGG GTG GGT GCA TT		*carA* intergenic region
narKGHJI promoter	F	CAT CAA TTT CTC CTC AAG TGG	281	47	Amplification of a portion of the *narKGHJI* promoter region; used as a negative control for binding assays (Nishimura *et al.*, 2008)
	R	TTT GGT GTT GAG TTG TGT CAC		

[a] Underlined nucleotides indicate a restriction site

[b] Annealing temperature. The PCR conditions used for each primer pair are similar, with only the annealing temperature and elongation time differing for most primer combinations. After an initial denaturation step (96 °C, 5 min), each sample was subjected to 30 amplification cycles, consisting of 30 s of denaturation at 95 °C, 30 s of annealing at T_a and either 30 s (for products < 1 kb) or 60 s (for products ≥ 1 kb) of elongation at 72 °C. This was followed by a final elongation step at 72 °C for 7 min.

Appendix D

Electroporation of *C. glutamicum* ATCC 13032

Electrocompetent *C. glutamicum* ATCC 13032 cells were prepared using a method adapted from the protocol of Haynes and Britz (1990). A single colony of *C. glutamicum* ATCC 13032 cells grown on 2 × Yeast-Tryptone (2 × YT) solid medium was used to inoculate 10 ml of 2 × YT broth supplemented with 30 µg ml^{-1} Nx. This seed culture was grown for 16 h at 30 °C and decanted into 500 ml of 2 × YT supplemented with Nx, 2.5 % glycine (Merck), and 4 mg ml^{-1} isonicotinic acid hydrazide (Sigma-Aldrich). Once the culture had reached an OD_{600} of between 0.15 and 0.25, the cells were harvested by centrifugation (6000 × *g*, 5 min, 4 °C), washed twice with a single culture volume of ice-cold 15 % glycerol (Merck), and lastly resuspended in 1 ml of ice-cold 15 % glycerol supplemented with 0.5 M mannitol (Merck). Aliquots of chilled cells (100 µl) were either used immediately for electroporation or stored at -80 °C. Frozen cells were thawed slowly on ice, washed twice with ice-cold 15 % glycerol at 4 °C and resuspended in 100 µl of ice-cold mannitol-containing 15 % glycerol prior to the addition of 1-2 µl of butanol-precipitated DNA (≥1 µg $µl^{-1}$) (Sambrook *et al.*, 2001; Xue *et al.*, 1999). The cell-DNA mixture was now dispensed into an ice-cold electroporation cuvette (1 mm electrode gap; Molecular BioProducts) and the competent cells electroporated with a GenePulser-Pulse Controller Apparatus (Bio-Rad) using a capacitance of 25 µF, a parallel resistance of 200 Ω and a voltage of 2.5 kV (Liebl *et al.*, 1989). This typically resulted in time constants in the range of 2-4 ms. Following electroporation, 1.5 ml of 2 × YT broth was added to the cuvette and used to rinse out the cell suspension, which was immediately subjected to a heat treatment step (46 °C, 6 min) in a 2.0 ml Eppendorf tube in order to limit restriction enzyme activity (Edwards *et al.*, 1999). Cells were subsequently incubated at 30 °C with lateral shaking for 2 h and finally plated with antibiotic selection on 2 × YT agar. Electrotransformation efficiencies were routinely assessed using the pEKEX2 vector, which harbours a Kan^R marker and is capable of replication in *C. glutamicum* ATCC 13032 (Eikmanns *et al.*, 1994).

Appendix E

Standard curves for the colorimetric quantitation of L-arginine and L-citrulline

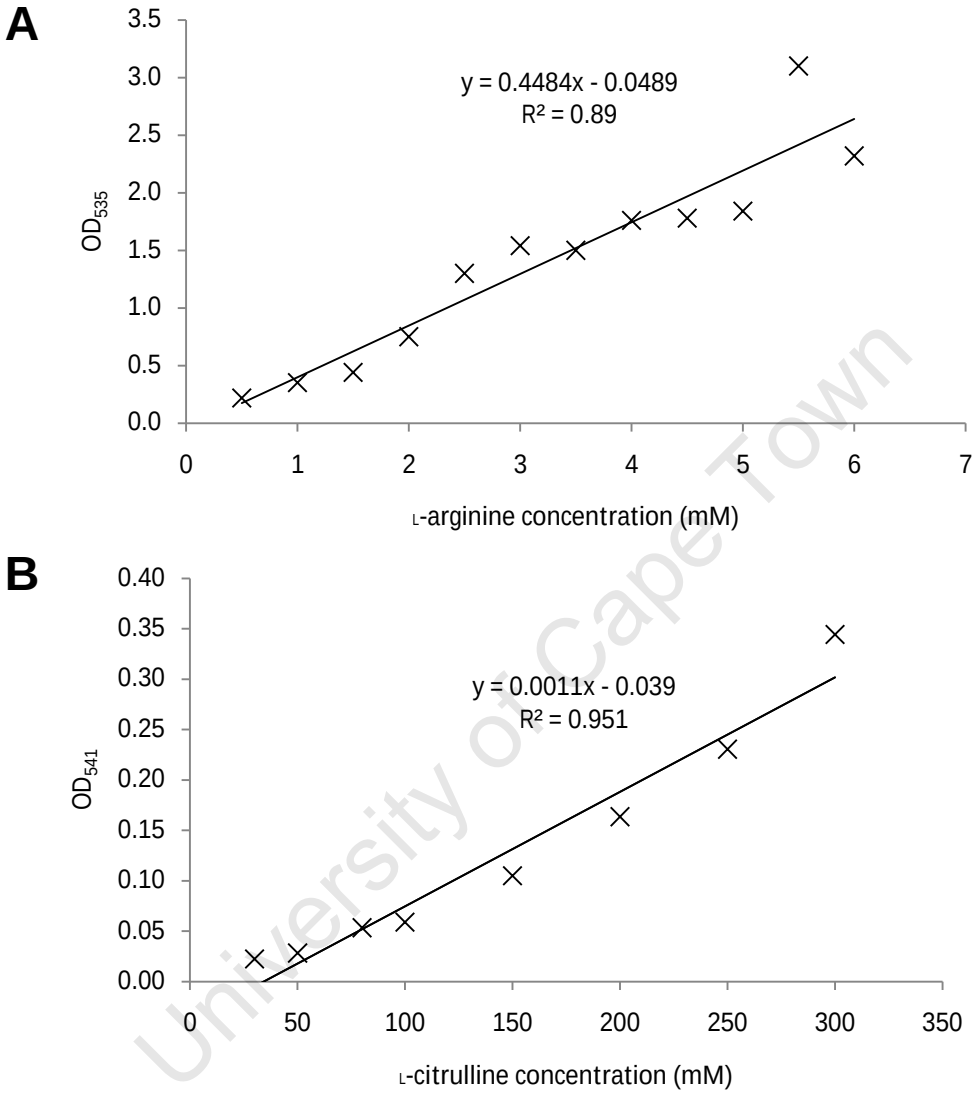


Figure A.1 Calibration curves for the colorimetric quantitation of L-arginine and L-citrulline. All measurements were performed in triplicate for three independent experiments. **(A)** Relationship between L-arginine concentration and colour intensity. L-Arginine was added as indicated to CGXII MM broth supplemented with 10 % trichloroacetic acid, after which the OD_{535} was quantitated as outlined by Tomlinson and Viswanatha (1974). **(B)** Relationship between L-citrulline concentration and colour intensity. L-Citrulline was added as indicated to CGXII MM broth containing 10 % trichloroacetic acid, after which the OD_{541} was quantitated according to the method of Knipp and Vašák (2006).

Appendix F

Competitor oligonucleotides used in the electrophoretic mobility shift assays

Oligonucleotide name		Sequence (5' to 3')	Length (bp)
***CAF21404-argC* intergenic region**			
1	sense	TAC CCA CTT GCA GTT TTA GCT GTA GGT GGG TTT TTG CAT GTC TAA CCC G	50
	antisense	ACG GGT TAG ACA TGC AAA AAC CCA CCT ACA GCT AAA ACT GCA AGT GGG TA	50
2	sense	GTG GGT TTT TGC ATG TCT AAC CCG TCT TTT ATG CAC ACC CCC GCA ATG AA	50
	antisense	TTC ATT GCG GGG GTG TGC ATA AAA GAC GGG TTA GAC ATG CAA AAA CCC AC	50
3	sense	CTT TTA TGC ACA CCC CCG CAA TGA ATC AAA AAT TTA TGC ATG AAT AAT TT	50
	antisense	AAA TTA TTC ATG CAT AAA TTT TTG ATT CAT TGC GGG GGT GTG CAT AAA AG	50
4	sense	TCA AAA ATT TAT GCA TGA ATA ATT TGC ATG ATC ATG CAT AAC GTG	45
	antisense	CAC GTT ATG CAT GAT CAT GCA AAT TAT TCA TGC ATA AAT TTT TGA	45
5	sense	ATG CAC ACC CCC GCA ATG AAT CAA AAA TTT A	31
	antisense	TAA ATT TTT GAT TCA TTG CGG GGG TGT GCA T	31
6	sense	ATG AAT CAA AAA TTT ATG CAT GAA TAA TTT G	31

	antisense	CAA ATT ATT CAT GCA TAA ATT TTT GAT TCA T	31

***pyrR-carA* intergenic region**

1	sense	CTG AAT GTA GTT TAT T	16
	antisense	AAT AAA CTA CAT TCA G	16
2	sense	CTG AAT GTA GTT TAT TCA ATT CAT TGC AT	29
	antisense	ATG CAA TGA ATT GAA TAA ACT ACA TTC AG	29
3	sense	ATT CAA TTC ATT GCA T	16
	antisense	ATG CAA TGA ATT GAA T	16