SubtilNet2, a functional interaction model for the validation of potential therapeutic targets

Christopher Stephen Hawes

Thesis submitted in partial fulfilment of the requirements of the regulations for the degree of Doctor of Philosophy

Newcastle University

Faculty of Medical Sciences

Institute of Cell and Molecular Biosciences

May 2010

Abstract

Systems Biology- the study of interactions between components of biological systems, and how these can produce new functions and behaviours, is beginning to produce a more comprehensive understanding of biology.

Its development is enabling many new opportunities, including the discovery and development of more effective and targeted therapeutics for a range of different conditions. It was in this context that this investigation began, with focus placed upon identifying therapeutic targets in *Bacillus subtilis* that could be used to limit the development and spread of infection, so called anti-infective targets.

Using an *in silico* data driven Systems Biology approach, our industrial collaborators, e-Therapeutics predicted pairs of genes from *B. subtilis* that could act as anti-infective targets when targeted together. This investigation was tasked with the development and testing of experimental models and approaches that could be used to validate these potential targets.

In a separate collaboration with the Integrative Bioinformatics Group at Newcastle University, a functional interaction network model for *B. subtilis*- SubtilNet2, was generated and tested. Compiled from a range of experimental, bioinformatical and literature based sources, it represented all known functional interactions known to occur within *B. subtilis*. This network was applied to investigate the selection of the predicted targets, and determine any biological basis for the experimental results seen. A single predicted target acting by itself was confirmed to be successful.

As a second component to this investigation, Systems Biology was used to complement traditional hypothesis driven research, specifically the possibility of directed targeting and channelling of substrates between two biosynthetic pathways. This was explored by studying the synthesis of carbamoyl phosphate (CP), an intermediate in both the arginine and uracil biosynthetic pathways. Typically, prokaryotes encode a single heterodimeric carbamoyl phosphate synthetase (CPS) that is used by both the arginine and pyrimidine biosynthetic pathways. *B. subtilis* and its close relatives are unique in encoding arginine- and uracil-specific copies of this enzyme. Moreover, the genes encoding the respective arginine (*carA* and *carB*) and uracil (*pyrAA* and *pyrAB*) specific CPSs are clustered with the other genes in their respective pathways (e.g. *argC,J,B,D-carA,B-argF* and *pyrB,C,AA,AB,K,D,F,E*) This degree of clustering is not found in bacteria with single CPSs.

Experimental and SubtilNet2 analysis approaches were developed to express and individually test for the presence of any interaction between the subunits of each systems CPS's, as well as to other components within associated gene clusters. The presence or absence of interaction would be used to determine if CP produced by one system could be shared with the opposite system. If it couldn't, could the unusual cluster of genes seen to surround each CPS be used to encode a macromolecular complex structure with a single point of entry and exit to channel CP and other substrates within a biosynthetic system?

A failure despite repeated attempts and strategies to produce soluble CPS subunits and other biosynthesis proteins, when expressed independently of one another, suggested a need for the presence of other members of each pathway. SubtilNet2 testing of these components and their functional associations didn't identify any distinct groups or systems being supplied with system specific CP, however this is more likely to result from limitations of the associated approaches, rather than genuine a biological property.

SubtilNet2, a functional interaction model for the validation of potential therapeutic targets

Christopher Stephen Hawes

Thesis submitted in partial fulfilment of the requirements of the regulations for the degree of Doctor of Philosophy

Newcastle University

Faculty of Medical Sciences

Institute of Cell and Molecular Biosciences

May 2010

Abstract

Systems Biology- the study of interactions between components of biological systems, and how these can produce new functions and behaviours, is beginning to produce a more comprehensive understanding of biology.

Its development is enabling many new opportunities, including the discovery and development of more effective and targeted therapeutics for a range of different conditions. It was in this context that this investigation began, with focus placed upon identifying therapeutic targets in *Bacillus subtilis* that could be used to limit the development and spread of infection, so called anti-infective targets.

Using an *in silico* data driven Systems Biology approach, our industrial collaborators, e-Therapeutics predicted pairs of genes from *B. subtilis* that could act as anti-infective targets when targeted together. This investigation was tasked with the development and testing of experimental models and approaches that could be used to validate these potential targets.

In a separate collaboration with the Integrative Bioinformatics Group at Newcastle University, a functional interaction network model for *B. subtilis*- SubtilNet2, was generated and tested. Compiled from a range of experimental, bioinformatical and literature based sources, it represented all known functional interactions known to occur within *B. subtilis*. This network was applied to investigate the selection of the predicted targets, and determine any biological basis for the experimental results seen. A single predicted target acting by itself was confirmed to be successful.

As a second component to this investigation, Systems Biology was used to complement traditional hypothesis driven research, specifically the possibility of directed targeting and channelling of substrates between two biosynthetic pathways. This was explored by studying the synthesis of carbamoyl phosphate (CP), an intermediate in both the arginine and uracil biosynthetic pathways. Typically, prokaryotes encode a single heterodimeric carbamoyl phosphate synthetase (CPS) that is used by both the arginine and pyrimidine biosynthetic pathways. *B. subtilis* and its close relatives are unique in encoding arginine- and uracil-specific copies of this enzyme. Moreover, the genes encoding the respective arginine (*carA* and *carB*) and uracil (*pyrAA* and *pyrAB*) specific CPSs are clustered with the other genes in their respective pathways (e.g. *argC,J,B,D-carA,B-argF* and *pyrB,C,AA,AB,K,D,F,E*) This degree of clustering is not found in bacteria with single CPSs.

Experimental and SubtilNet2 analysis approaches were developed to express and individually test for the presence of any interaction between the subunits of each systems CPS's, as well as to other components within associated gene clusters. The presence or absence of interaction would be used to determine if CP produced by one system could be shared with the opposite system. If it couldn't, could the unusual cluster of genes seen to surround each CPS be used to encode a macromolecular complex structure with a single point of entry and exit to channel CP and other substrates within a biosynthetic system?

A failure despite repeated attempts and strategies to produce soluble CPS subunits and other biosynthesis proteins, when expressed independently of one another, suggested a need for the presence of other members of each pathway. SubtilNet2 testing of these components and their functional associations didn't identify any distinct groups or systems being supplied with system specific CP, however this is more likely to result from limitations of the associated approaches, rather than genuine a biological property.

Acknowledgements

I would like to thank the following people for their help, support and guidance in the completion of this investigation. My family, Professor C.R. Harwood, Professor A. Wipat, Dr J. Hallinan and members of the Harwood lab. I would also like to thank the support staff of the Institute of Cell and Molecular Biology and the Integrative Bioinformatics Group at the University of Newcastle upon Tyne.

Christopher S. Hawes (May 2010)

"Aequam memento rebus in arduis servare mentem"

Horace

Table of Contents

Table of Figures

Table of Tables

Abbreviations

Å	Armstrong
ABC	ATP binding cassette
AdoMet	Adenosylmethionine
Amp^R	Ampicillin resistance
AI2	Auto-inducer 2
APS	Ammonium persulphate
ATP	Adenosine triphosphate
BSA	Bovine serum albumin
Cm^R	Chloramphenicol resistance
CP	Carbamoyl phosphate
CPS-A	Carbamoyl phosphate synthetase- arginine biosynthetic system
CPS	Carbamoyl phosphate synthetase
CPS-P	Carbamoyl phosphate synthetase-pyrimidine biosynthetic system
Cytoscape	A Java based network modelling and interrogation tool
DAPI stain	6-Diamidino-2-phenylindole stain
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
(ds) DNA	Double stranded DNA
E.C	Enzyme commission number
EDTA	Ethylenediaminetetraacetic acid
Em^R	Erythromycin resitance
g	Grams
g	Gravitational force
GFP	Green fluorescent protein
gfpmut1	Gene encoding green fluorescent protein
GST	Glutathionine S-transferase
GO term	Gene ontology term
HCl	Hydrochloric acid
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
Kbp	Kilobase pair
kDa	KiloDalton
KMBA	Keto methylthiobutyric acid
KO	Knockout
lacI	Lac repressor
lacZ	Beta-galactosidase
M	Molar
MBP	Maltose binding protein
Mbp	Megabase pair

MCODE	Molecular complex detection algorithm – a Cytoscape plug-in to detect and analyse clusters
MCP	Methyl-accepting chemotaxis protein
MCS	Multiple cloning site
MgATP	Magnesium adenosine triphosphate
mg	Milligram
ml	Millilitre
mM	Millimolar
(m)RNA	Messenger RNA
ms	Millisecond
MTA	Methylthioadenosine
MTR	Methylthioribose
NaOH	Sodium Hydroxide
OD	Optical Density-(measured at 600 nm unless otherwise stated)
OAS	O-acetyl serine
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
Pmut	Primer sequence for Pspac promoter sequence, to confirm gene orientation
P_{spac}	IPTG inducible pMUTIN4 promoter
P_{tac}	*Tac* promoter
PTS	Phosphotransferase system
P_{xyl}	Xylose inducible pSG1164 promoter
RNA	Ribonucleic acid
rpm	Revolutions per minute
SDS	Sodium dodecyl sulphate
SPR	Surface plasmon resonance
SRH	*S*-ribosylhomocysteine
TEMED	Tetramethylethylenediamine
UMP	Uridine monophosphate
µl	Microlitre
µg	Microgram
V	Volts

Chapter 1

Introduction

1. Introduction

In this chapter will be described both broad and specific aspects that relate and underpin the investigation that has been conducted, these include:

- The choice of organism as well as specific biological properties and systems contained within it will be discussed, both in a broad sense as well as key aspects that relate to the ensuing investigation.
- The concepts behind using mathematics to represent and analyse complex systems such as those found within biological systems, in attempts to gain unparalleled comprehensive understandings of the studied systems will be discussed.
- How the use of integrated mathematical/biological approaches in research could allow the development of a more comprehensive understanding of biological systems and organisms, with the range of benefits that this would produce.
- This approach will be applied to two separate fields- the identification of new drug targets, together with compound design and development, and the investigation of the arginine biosynthetic pathway, its function and unique organisation.
- The specific aims and objectives of the investigation.

1.1 *Bacillus subtilis*

1.1.1 Model Gram-positive bacterium and industrial workhorse

Bacillus subtilis is an aerobic, endospore-forming rod shaped non-pathogenic bacterium, typically found in the upper layers of soil and in associated watercourses (Priest, 1989). It is capable of oxidizing a wide range of organic compounds and has relatively simple growth requirements. *B. subtilis* belongs to the genus *Bacillus*, which contains more than 50 validly described species (Claus, 1989) including both pathogenic and non-pathogenic strains. *B. subtilis*, is the most extensively studied Gram-positive bacterium and used as a laboratory model organism for the study of pathogenic members of this group as well as other Gram-positive bacteria (Zweers *et al.*, 2008).

The members of the genus *Bacillus* have a long history of exploitation by man, from their use in food preservation, in producing the fermented soybean product natto (Zweers *et al.*, 2008) and Chungnkook-Jang (Kim *et al.*, 2004), to their applications in industry, to produce high value and high volume (>20g/L^{-1}) (Harwood & Cranenburgh, 2008) biochemicals such as antibiotics, vitamins, pesticides and hydrolytic enzymes (Harwood, 1992).

1.1.2 Genetic amenability and molecular tools

B. subtilis was the first Gram-positive bacterium to be sequenced (Kunst *et al.*1997 (Kunst *et al.*, 1997)), identifying 4,100 genes encoded within a 4.2 Mbp chromosome. Complimenting this sequencing data is more than 50 years of extensive research into many aspects of the biochemistry, physiology and genetics of the organism (Zweers *et al.*, 2008). This has defined *B. subtilis* as the archetypal model for all Gram-positive bacteria, with knowledge second only to that of *Escherichia coli* (*E. coli*) (Harwood & Wipat, 1996). This combined wealth of information has strengthened both commercial and research interest using this organism, particularly in Japan and Europe that are generating storing and analysing "omic-size" datasets. This existing comprehensive and continually expanding body of knowledge means that *B. subtilis* represents an ideal organism to study using Systems Biology approaches.

The two driving forces for early extensive studies on *B. subtilis* were its ability to sporulate, providing insights into cellular differentiation and its intricate controls, and its amenability to genetic modification (Harwood & Wipat, 1996). *B. subtilis* grows rapidly on a simple salts medium and has the innate ability to be transformed by large pieces (up to 8.5Kb) of extracellular double stranded (ds)DNA (Dubnau, 1991). Under conditions of nutrient starvation, competence genes encoding an efficient DNA uptake system become expressed (Dubnau, 1991). Double-stranded (ds) DNA from the surrounding environment becomes internalised and, providing it has homology with the chromosome, is able to integrate into the *B. subtilis* chromosome (Provvedi & Dubnau, 1999). Extracellular (ds)DNA uptake and integration occurs in four stages (Dubnau, 1991):

1) Expression of adhesion proteins that facilitate attachment of extracellular (ds)DNA to the cell surface.
2) Fragmentation and partial digestion of the (ds)DNA to generate single stranded (ss)DNA.
3) Internalisation of the (ss)DNA.
4) Recombination between homologous regions of the internalised (ss)DNA and the chromosome.

1.1.3 Gene regulation and metabolism

1.1.3.1 Introduction

Bacteria have the ability to adapt rapidly and reversibly to a variety of changes in their internal and external environments (Nicholl, 2002). The ability to selectively express a large repertoire of gene products allows them to survive and function in the face of major changes in their environment (Lopez-Maury *et al.*, 2008). Controlling and coordinating the expression of this repertoire of genes is essential for both survival, and to limit the amount of energy expended in expressing required genes. Because of the importance of coordinating gene expression, the bacterium has evolved a variety of complex regulation systems.

One of the most important levels at which regulatory systems function is that of transcription. A model proposing the control of gene transcription was developed by François Jacob and Jacques Monod (Jacob & Monod, 1961) termed the operon model, following work on lactose metabolism in *E. coli*. The model was later found to be widely applicable to many prokaryotes, and describes the ability to control the expression of collections of genes by a single stimulus.

1.1.3.2 Operon structure

The term operon collectively defines a fundamental transcriptional unit (Bergman *et al.*, 2007) within a bacterial genome. It contains a cluster of co-expressed “structural” genes and their regulatory sequences, often functionally related and required by the cell for a particular process or pathway (Bergman *et al.*, 2007) (Figure 1.1). It has been estimated that ~50% of all bacterial genes are located within an operon (Brouwer *et al.*, 2008). Operons are identified by an upstream promoter or promoters and a downstream

transcription terminator (Figure 1.1). When transcribed into messenger (m)RNA, individual genes, encoding enzymatic or structural components of the cell, are defined by an upstream ribosome binding site and associated start codon, and a downstream stop codon (Hartl & Jones, 2008).

Depending on the environment, individual operons may be constitutively expressed or selectively expressed or repressed. This is facilitated by activator and/or repressor molecules produced in response to a specific stimulus (Klug *et al.*, 2005). During repression, a repressor molecule will bind to an operator-a sequence downstream of the promoter, to either impede the binding or movement of RNA polymerase. During activation of an operon, an activator molecule will bind to an activator binding site, facilitating the binding of RNA polymerase to the promoter.

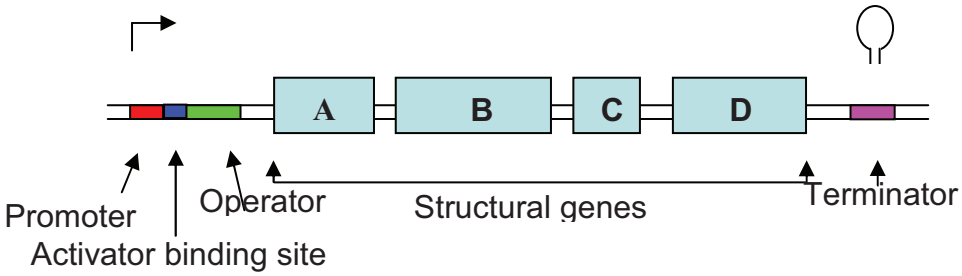


Figure 1.1: The operon structure. A graphical representation of operon structure. Several genes (A-D) are co-transcribed from the promoter region (red) to the terminator region (pink) in response to a specific signal. Repressors and activators of transcription bind to the operator (green) and activator (blue) respectively.

1.1.3.3 Operon function

Several theories exist as to the purpose of operons, with it generally being accepted that their main role is to facilitate the co-regulation of its structural genes (Price *et al.*, 2006). Operons can contain a few or several genes, and can be under the control of the same regulatory mechanism (Klug *et al.*, 2005). They may be constitutively or conditionally activated or repressed by the binding or removal of their cognate activator or repressor from the operator region upstream of the first structural gene. This ultimately affects the manner in which RNA polymerase binds to the promoter and initiates transcription.

A single polycistronic (m)RNA strand is transcribed for each operon in prokaryotes (Lengeler, 1999), and to which are bound numerous ribosomes along its length (Klug *et al.*, 2005) (Figure 1.2). Translation is initiated even before transcription is completed and consequently, these two processes are intimately coupled in bacterial systems. The simultaneous processing of several structural genes allows their translated products to come into close proximity with each other (Dandekar *et al.*, 1998), providing an efficient and targeted mechanism for protein complexes to interact This has led to the proposal that the maintenance of operon structures throughout evolution may be as much to do with the formation of such complexes as with the conventional view that they are required for co-ordinated gene expression.

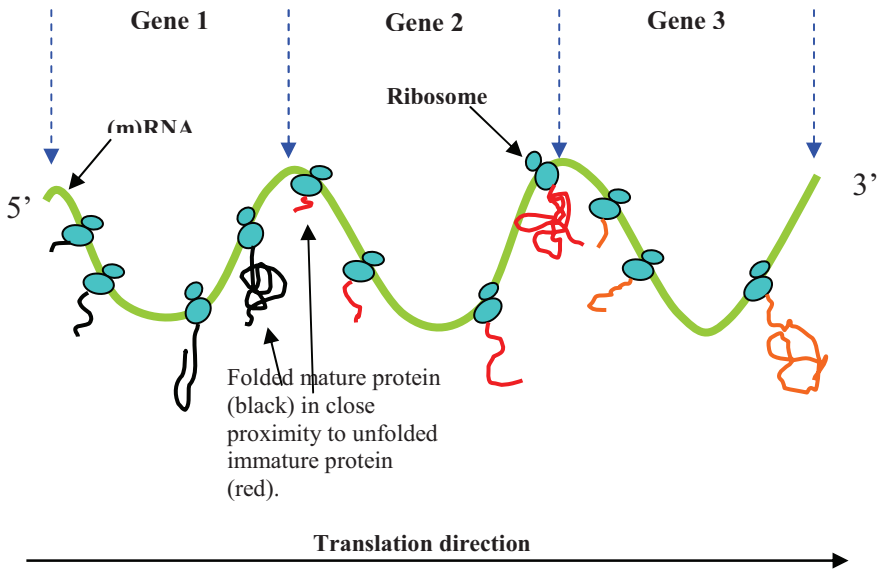


Figure 1.2: The localisation of (m)RNA and translated protein from the ribosome. A graphical representation of translated (m)RNA and close proximity of resultant mature protein, to immature protein.

1.1.3.4 Co-ordinated operon function

Operons distributed around the chromosome can be co-transcribed with one another under particular conditions. Operons that become transcribed with one another in response to a stimulatory signal are referred to as stimulons (Lengeler, 1999; Cases & de Lorenzo, 2005). Operons under the regulatory control of the same protein are known as regulons (Lengeler 1999; Cases & de Lorenzo, 2005). In this manner, the multiple

control of genes required for a biological function, can be achieved efficiently and rapidly, even when separated from one another within the chromosome.

1.1.4 Arginine biosynthesis introduction

L-arginine is a metabolically versatile amino acid (Lu, 2006), that is used as a building block in proteins, as well as a source of carbon, nitrogen and energy through its conversion in a range of catabolic processes (Lu, 2006). Due to the essential nature of this amino acid and its multifunctional role, many bacteria have developed uptake mechanisms allowing it to be assimilated from the environment (Celis *et al.*, 1973; Poolman *et al.*, 1987; Wissenbach *et al.*, 1995). Bacteria have also developed an ability to biosynthesise arginine from L-glutamate, the most abundant amino acid found within bacterial cells (Commichau *et al.*, 2008). The complex way in which arginine biosynthesis is controlled and regulated is considered to be a model of gene expression and regulation with investigations upon it dating back to the 1950's (Cunin *et al.*, 1986; Maas, 1991).

1.1.4.1 The arginine biosynthetic pathway

Arginine is synthesised from L-glutamate in eight enzymatic steps that are under complex regulation, ensuring that pathway genes are only ever induced in the absence of exogenous arginine (Figure 1.3) (Cunin *et al.*, 1986). The first four stages involve the acetylation of L-glutamate and its derivative products (Vogel, 1953) using the enzymes ArgJ/A, ArgB ArgC and ArgD the purpose of which is thought to prevent glutamate from being used in proline biosynthesis (Caldovic & Tuchman, 2003). The end of this stage yields N-acetylornithine. Prokaryotes can then follow one of two routes, a linear or cyclical pathway to produce arginine, differing in the deacetylation method of N-acetylornithine (Sakanyan *et al.*, 1996). This represents step 5 of the pathway.

In the linear route (used typically by members of the *Enterobacteriaceae* and the archaean genus *Sulfolobus* (Casteele *et al.*, 1990)), the deacetylation of N-acetylornithine is performed by acetylornithine deacetylase (ArgE), to produce ornithine and acetate. The acetyl group required to acetylate glutamate for the beginning 4 steps of arginine biosynthesis is supplied by acetyl-CoA (Sakanyan *et al.*, 1996).

In the cyclical route (the most efficient and the pathway typically used by most prokaryotes (Caldovic & Tuchman, 2003) including members of the genus *Bacillus*

(Sakanyan, 1992)) the acetyl group is recycled by its transfer from acetylornithine to glutamate, *via* the enzyme ornithine acetyltransferase (ArgJ) (Sakanyan *et al.*, 1993).

The final 3 steps in the arginine biosynthesis pathway is common to both the linear and cyclical pathways. A carbamoyl phosphate moiety is transferred to the five amino group of ornithine, to generate citrulline via ornithine transcarbamylase (ArgF) (Bringel *et al.*, 1997). Citrulline is converted to argino-succinate by argininosuccinate synthase (ArgG) before being converted to arginine by argininosuccinate lyase (ArgH).

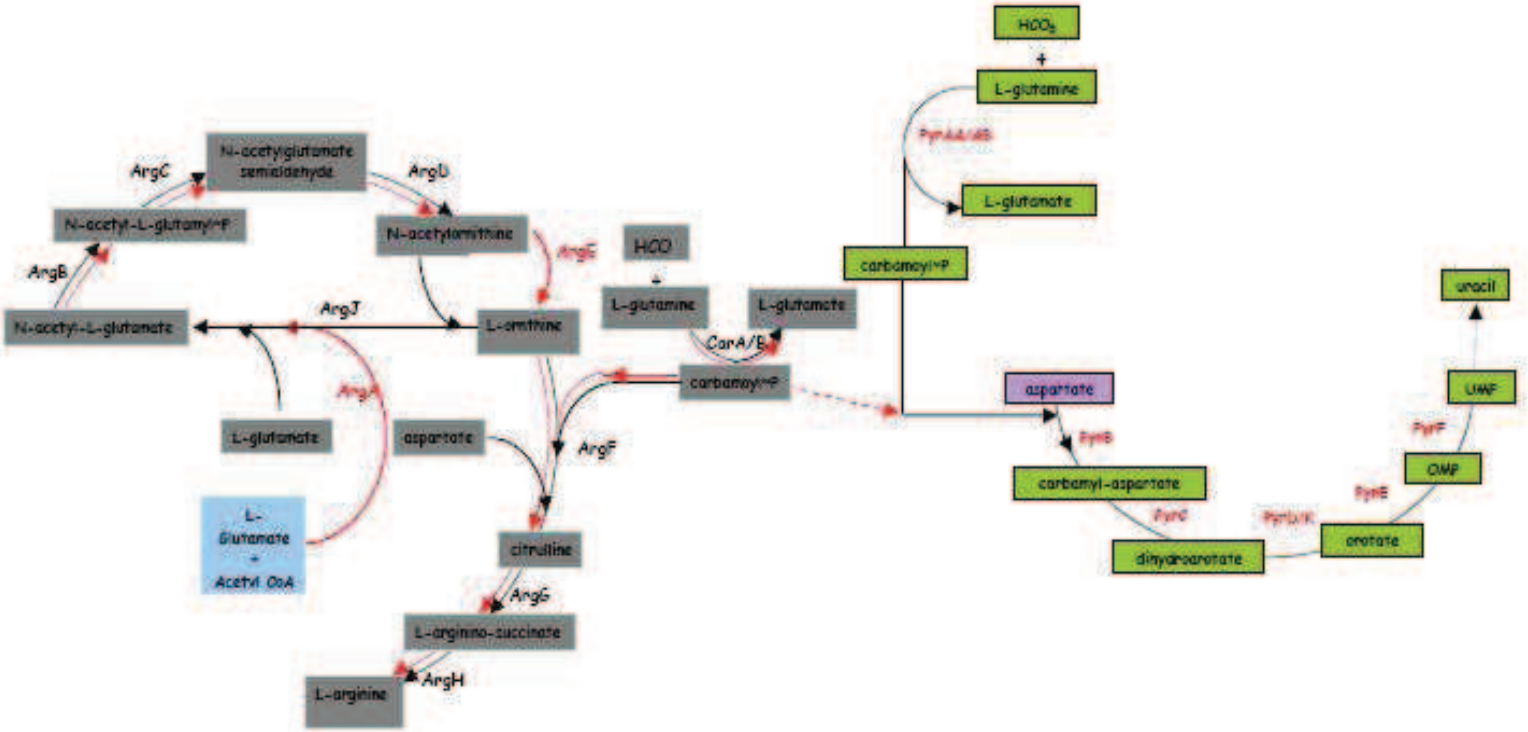


Figure 1.3: The arginine and pyrimidine (uracil) biosynthetic pathways in prokaryotes. A graphical representation of both the linear (red) and cyclical route (black) arginine biosynthetic pathways. Enzymes are identical for both linear and cyclical unless highlighted in the appropriate colour. ArgB, N-acetylglutamate 5-phosphotransferase. ArgC, N-acetylglutamate phosphate reductase. ArgD; N-acetylornithine transaminase. ArgE, Acetylornithine deacetylase. ArgJ, Ornithine acetyltransferase. ArgA, N-acetylglutamate synthase. ArgF, Ornithine transcarbamylase. ArgG, Argininosuccinate synthase. ArgH, Argininosuccinate lyase. CarA, Carbamoyl phosphate synthetase subunit A (arginine biosynthetic system). CarB, Carbamoyl phosphate synthetase subunit B (arginine biosynthetic system). PyrAA, Carbamoyl phosphate synthetase subunit A (pyrimidine biosynthetic system). PyrAB, Carbamoyl phosphate synthetase subunit B (pyrimidine biosynthetic system). PyrB, Aspartate carbamoyltransferase. PyrC, Dihydroorotase. PyrD/K, Dihydroorotate dehydrogenase. PyrE, Orotate phosphoribosyltransferase. PyrF, Orotidine 5'-phosphate decarboxylase.

1.1.4.2 Carbamoyl phosphate synthetase

Carbamoyl phosphate (CP) is a small energy rich compound (Cunin *et al.*, 1986), produced using the enzyme carbamoyl phosphate synthetase (CPS) and the substrates L-glutamine, bicarbonate and MgATP (Thoden *et al.*, 1997). Carbamoyl phosphate is central to the synthesis of arginine and pyrimidines in both prokaryotic and eukaryotic organisms.

Most prokaryotes employ a single CPS for all carbamoyl phosphate requirements. *Bacillus* and other closely related genera employ two distinct CPS designated CPS-A and CPS-P (Nicoloff *et al.*, 2000). Utilising identical substrates, they differ only with respect to the pathways to which they supply carbamoyl phosphate to. CPS-A provides CP to the arginine biosynthetic pathway, and CPS-P to the pyrimidine biosynthetic pathway.

CPS has been the subject of intense study for more than 40 years because of its interesting catalytic properties, large size and important metabolic role (Holden *et al.*, 1999). X-ray crystallographic studies of the *E. coli* CPS (Thoden *et al.*, 1997) has identified it as having an α,ß-heterodimeric structure, comprising of a large and small subunit with the ability to interconvert to an $(\alpha,\beta)_4$-heterooctamer, depending on the presence of specific effector molecules (Kim & Raushel, 2001; Powers *et al.*, 1980). The small subunit of the structure of CPS contains an aminotransferase domain that generates an ammonium ion from glutamine (Matthews & Anderson, 1972). The two domains of the large subunit catalyse the production of carbamoyl phosphate from delivered ammonium, bicarbonate and MgATP.
Analysis of CPS from several other organisms has identified similar sizes of 160 kDa and shown to catalyse the same set of intermediate reactions (Meister, 1989).

1.1.4.3 Metabolic channelling

The production of CP occurs in four separate reactions (Figure 1.4). The presence of bicarbonate on an active site within the large CPS subunit causes the enzyme to be activated through the hydrolysis of a single MgATP to produce carboxyphosphate. The aminotransferase domain of the small subunit catalyses the hydrolysis of a single molecule of L-glutamine. This produces an ammonium ion that launches a nucleophilic attack on carboxyphosphate to produces carbamate. This conducts a nucleophilic attack

on another molecule of MgATP via a second active site within the large CPS subunit to produce CP.

The intermediate products produced in the production of CP are highly reactive and unstable. Carboxyphosphate, with a half life of 70 ms (Sauers, 1975), is rapidly hydrolysed to carbon dioxide and inorganic phosphate (Purcarea *et al.*, 2001), while carbamate, with a half life of just 28 ms (Wang *et al.*, 1972), is rapidly decomposed to ammonia and carbon dioxide. To prevent decomposition and loss of these energetically costly intermediates, CPS has evolved a molecular transport channel to direct these intermediate products to the various active sites within the enzyme (Miles *et al.*, 1999; Thoden *et al.*, 1997). This mechanism of "metabolic channelling", prevents contact of the unstable intermediates with other cellular environments and competing systems, while reducing the transit time is not limited to CPS, and has been observed in other enzymes including tryptophan synthase (Hyde *et al.*, 1988) and glutamine phosphoribosyl pyrophosphate amidotransferase (GPATase)(Krahn *et al.*, 1997).

The X-ray crystallographic studies of *E. coli* CPS has identified a 96Å intermolecular channel connecting the three domains of the enzyme (Thoden *et al.*, 1997). The first section of the channel connects the active site of the small subunit with the carboxyphosphate producing active site of the large subunit and is called the ammonia channel (Kim *et al.*, 2002). The connection of the carboxyphosphate active site with the carbamate synthesising active site occurs by a carbamate channel (Kim *et al.*, 2002).

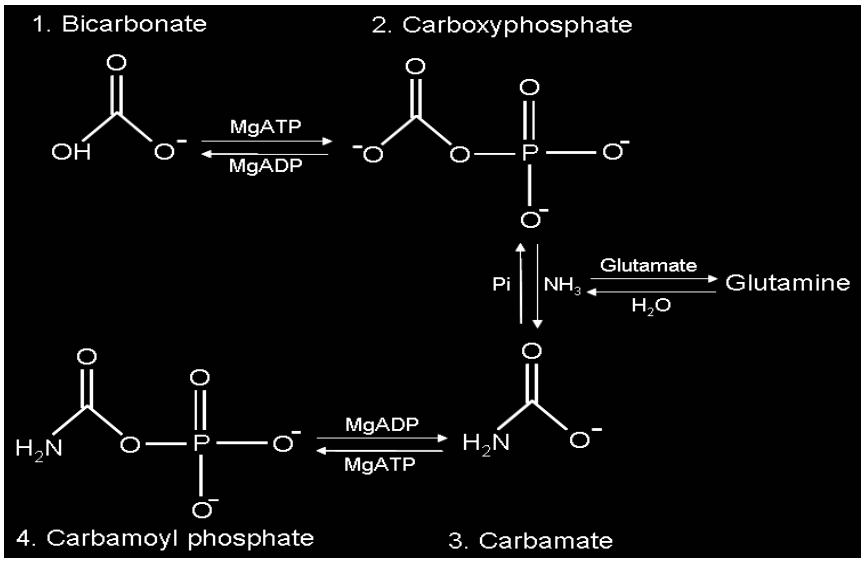


Figure 1.4: The four reactions of carbamoyl phosphate synthetase. The four separate reactions that occur in carbamoyl phosphate synthetase to produce carbamoyl phosphate.

1.1.4.4 Metabolic channelling in a wider context

The analysis of the distribution of arginine and pyrimidine biosynthetic genes in different organisms has revealed two distinct conformations. Organisms utilising a single CPS such as *E. coli* show some clustering in these genes (Mountain *et al.*, 1986), which remain separate from the single CPS (Figure 1.5). Organisms utilising two separate CPS have a greater degree of gene clustering, with distinct clusters that associate with individual biosynthetic systems. In *B. subtilis* the majority of biosynthetic genes for both the arginine and pyrimidine biosynthetic pathways are clustered around CPS-A and CPS-P respectively (Mountain *et al.*, 1986) (Figure 1.5).

The presence of two CP producing and utilising systems within a single organism could suggest that the produced CP may be being targeted to specific systems, especially as the duplication of CPS is potentially costly both in terms of efficiency and energy expenditure to the bacteria. The analysis of the closely related *Lactobacillus plantarum* possessing two separate CPS with the degree of biosynthetic gene clustering surrounding each reduced (Bringel *et al.*, 1997; Nicoloff *et al.*, 2004), identified an inability for CP to be shared when produced by the arginine biosynthetic pathway but

an ability for it to be shared when produced by the pyrimidine biosynthetic pathway (Nicoloff *et al.*, 2000).

The presence of separate CPS and arrangement of the biosynthetic gene clusters surrounding individual CPS could imply the channelling of CP from distinct CPS through a macromolecular complex that is formed by the enzymes involved within the biosynthetic system and having a single entry and exit point. This would prevent substrates and intermediates from entering a central pool, (a view that has recently been challenged by evidence suggesting bacterial cells have cytoskeletons capable of higher order structures (Carballido-Lopez & Errington, 2003)) where they can be used by the next enzyme of a pathway or be lost to an unassociated pathway. The formation of macromolecular complexes would also allow the efficient and timely delivery of substrates to each componant of the biosynthetic system.

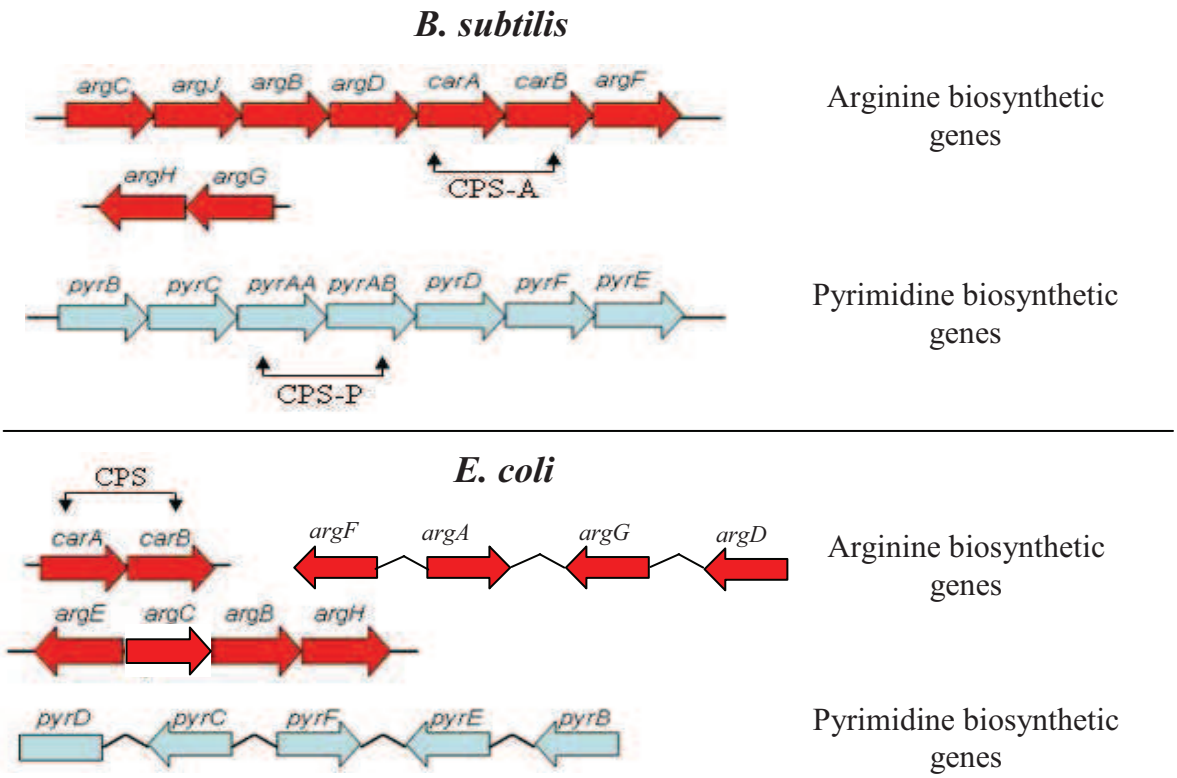


Figure 1.5: Arginine and pyrimidine biosynthetic gene organisation in *E.coli* and *B. subtilis*. CPS-A represents the carbamoyl phosphate synthetase genes of the arginine biosynthetic system, surrounded by other components of the arginine biosynthetic system in *B. subtilis*. CPS-P represents carbamoyl phosphate synthetase genes of the pyrimidine system, surrounded by other components of the pyrimidine biosynthetic system in *B. subtilis*. CPS represents the carbamoyl phosphate synthetase genes of the arginine/pyrimidine system in *E.coli*.

1.2 Systems Biology introduction

Biology is a complex field. To fully understand the properties, behaviours and mechanisms of biological systems, analysis must focus upon investigating complete systems and how their properties and behaviours originate as a property of their constituent components.

Systems Biology is an approach that can be used to investigate this field (Kitano, 2002a; Kitano, 2002b), through the integration of experimental and computational analysis, in a process of iterative refinement.

1.2.1 Complex systems

A complex system is one that is made up of many interacting parts, that can be conveniently represented as a network (Proulx *et al.*, 2005). The components and their interactions with one another lead to large scale behaviours and properties that can't easily, if at all, predicted by investigating individual components (Mitchell, 2002). These properties are said to be "emergent" (Kitano, 2002b).

The investigation and analysis of such systems and how they produce emergent properties is of great interest, due to their abundance and influences on the world around us. Incorporating technical knowledge from many different fields, including cybernetics, general systems theory, chaos theory, non-linear dynamics, mathematics, and physics (Ahn *et al.*, 2006). Its investigation has the potential to allow the understanding and prediction of the behaviour and properties of such systems.

Examples of complex systems are common and widespread, both man-made and natural, exceptionally large and intricately small. The World Wide Web, stock markets and society are examples of the former (Barabasi, 2009), while ecosystems and the climate are examples of the latter (Mitchell, 2002).

The types of emergent behaviour and properties that distinguish complex systems from simply "complicated systems" include:

- An ability to self-organise (Coffey, 1998), (Finnigan, 2005): Patterns form within the system through interactions internal to the system without intervention from external directing influences (Camazine *et al.*, 2001).

- A non-linear organisation (Coffey, 1998): In contrast to a linear organisation where a single perturbation affect only neighbours of the target, a single perturbation in a complex system may be propagated so that it affects multiple components.
- An order/chaos dynamic (Janecka, 2007): As a complex system increases in size its behaviour becomes less predictable.

1.2.2 Modelling complex systems

With the realisation that much of the behaviour of complex systems is attributable to the emergent properties, rather than individual isolated parts, research has come to focus on a systems level approach to understanding complex systems (Kitano, 2002a). Systems level analysis and understanding of multiple interacting systems, containing potentially many thousands of components and interactions, requires a robust, accurate and rapid approach to analysis, beyond that of normal human capabilities. Recent advances in computer technology now make this achievable (Kitano, 2002a).

To perform systems level analysis, components and their interactions must be represented in a form that is conducive to computational analysis, and that will allow the production of a two dimensional abstract network model, often termed as graphs (Figure 1.6). This approach is dependent on the availability of sufficient quantities and categories of quantitative data (Kitano, 2002b) about the components and interactions – that lend themselves very well to mathematical analysis, ultimately producing a graphical representation of the system under study.

The construction of graphs uses nodes to represent components of complex systems, and lines (or edges) to represent interactions between them (Ferrell, 2009). In many fields, including biology, multiple data sets and collection techniques are used. The weight and the direction of connection between interactions can be annotated to edges to fulfil specific requirements of analysis.

The most common systems level models of complex systems are static “snap shots” of what is occurring within the network at the time from which measurements were taken. Network models that display dynamic properties exist, but because of the complexities in constructing and simulating them they are generally considerably smaller than their static counterparts (Albert, 2007; van Riel, 2006).

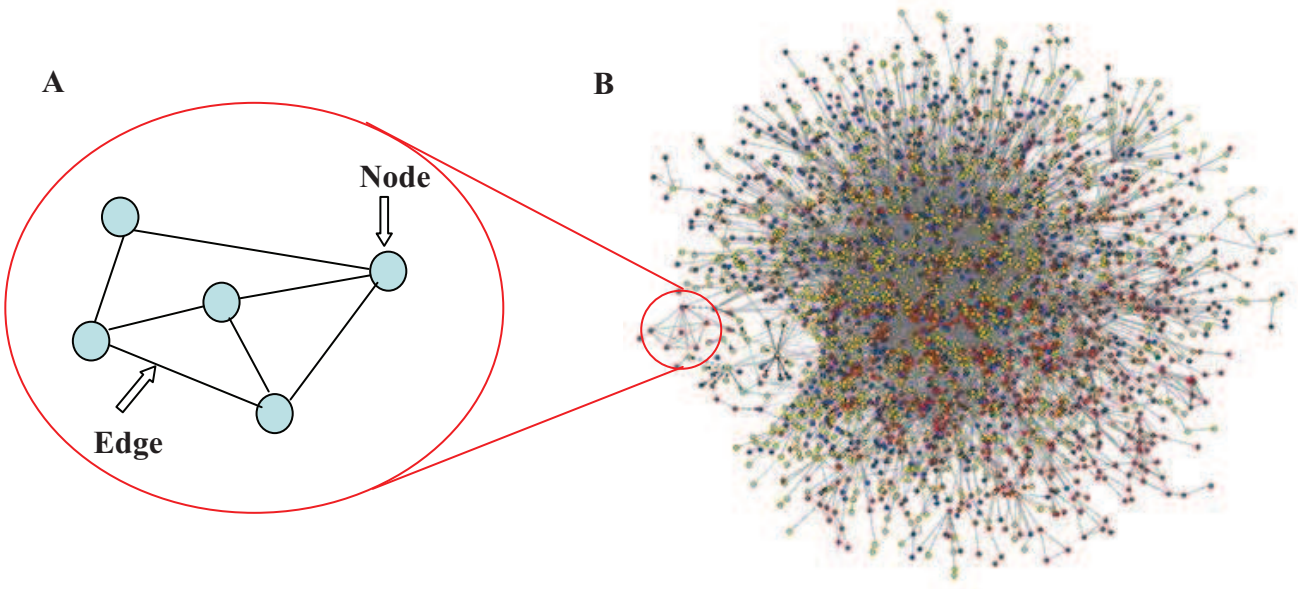


Figure 1.6: A graph model and its components. The nodes representing gene/gene products and edges representing the functional association between the nodes. B, The compilation of nodes and edges as would be seen in a typical complex system. This figure represents one of the first attempts of a protein interaction map for *Caenorhabditis elegans* (*C. elegans*) (Li *et al.*, 2004).

These networks can take on specific spatial arrangements (topology) (Figure 1.7), dependent upon the subject being modelled and can include:

- Ordered networks: Chains, grids, and lattice topologies. Nodes are associated with one another in an ordered and regular structure (Figure 1.7A). This topology facilitates the study of the behaviours of small groups of nodes, without the complexities of large integrated networks (Strogatz, 2001).
- Random networks: Node association is random, as is the generated structure (Figure 1.7B)(Strogatz, 2001).
- Small world networks: A middle ground between ordered graphs and random graphs, representing many real world networks (Strogatz, 2001). These graphs have regular network topologies, but some of the connections between nodes are replaced with random links (Figure 1.7C)(Watts & Strogatz, 1998).

- Scale free networks: A version of a small world network topology, possessing many of its features in addition to some increased attachment to individual nodes- that are collectively known as hubs (Figure 1.7D)(Strogatz, 2001). In this topological configuration, there is an increased probability of picking a random node that is connected to a hub, than one that is not. This topology is commonly seen in biological systems (Albert & Barabasi, 2002).

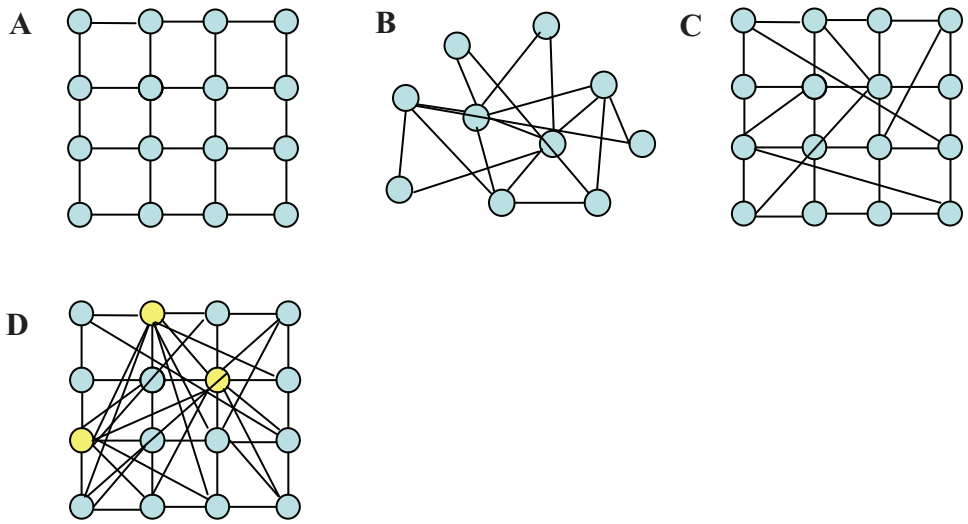


Figure 1.7: The different network topologies. A graphical view of different network topologies. A, An ordered regular lattice network topology. B, A random network topology with random edges between nodes. C, A small world network topology with some random edges. D, A scale free network topology with some preferential attachment to "hub" (yellow) nodes.

The scale free topology (Figure 1.7D) lends itself well to resisting adverse changes, a property known as robustness (Callaway *et al.*, 2000; Csete & Doyle, 2002). The network can resists the random removal of nodes with little detrimental effect to the underlying functioning of the system. Instead data is transmitted through hubs and to other nodes, which does not occur in random networks. Because of this property, scale free networks are found in situations where when random errors occur, it is imperative for the system to continue functioning, (e.g. biological systems and the World Wide Web (Barabasi & Oltvai, 2004)). This network topology however is very susceptible to

disruption when highly connected nodes are targeted (Kitano, 2002a), causing destruction of the network, a weakness not shared by a random network topology.

The degree distribution (P(*k*)), a measure of the probability of selecting a particular node with a particular number of edges (*k*) (Barabasi & Oltvai, 2004; Petermann & Rios, 2004) differs between network topologies. When plotted as histograms, random networks display a bell shaped distribution profile (Figure 1.8A)(Albert, 2005). Nodes picked at random have an average number of edges, with the exception of a few that have more and less and are represented by the sides of the curve. Scale-free networks display a power law distribution (Albert & Barabasi, 2002; Joyce & Palsson, 2006), where most nodes have only a few links. A few nodes have a very large number of links, which are called hubs, and connect the other nodes together (Barabasi & Oltvai, 2004; Hu *et al.*, 2005). The Probability of selecting a node with few edges is high, and nodes with greater numbers of edges decreases even as the network expands (Figure 1.8B). This can occur as a result of preferential attachment of nodes to hubs as well as by the standard growth of the network (Barabasi & Albert, 1999).

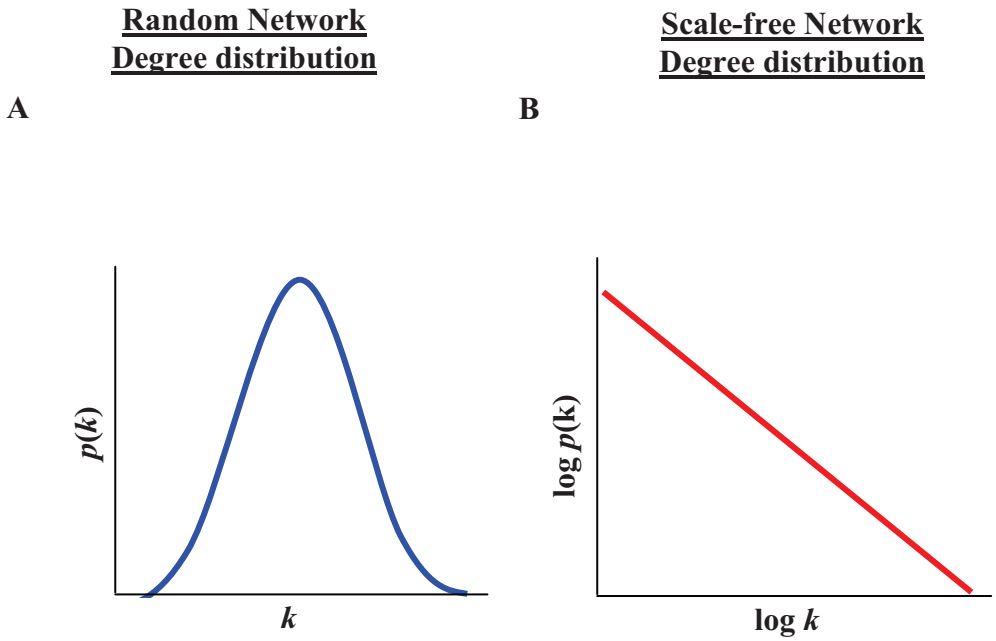


Figure 1.8: The degree distribution of random and scale free graph topologies. A, The degree distribution profile of a random network topology. B, The degree distribution profile of a scale-free network topology. *k* represents the number of edges and P(*k*) represents the probability of selecting a node with an expected number of edges. The degree distribution follows a bell shaped curve for a random networks, and a power law distribution for scale networks.

1.2.3 Complex network analysis

The generation of network models can be studied using a variety of statistical and mathematical tools, which look into the connections of nodes within the network. From these measurements, predictions can be made as to functions and characteristics that may possibly be produced.

There are many different analysis techniques available, with interest in this research field popular and varied, leading to constant developments. For the purpose of this investigation attention will be focused upon those adopted for use in this investigation. For a review of other potential methods, the reader is directed to the following publications (Barabasi & Oltvai, 2004; Mason & Verwoerd, 2007; Strogatz, 2001).

- **Degree:** A measurement of the number of "edge" interactions a node has within a network (Diestel, 2005). A node that has a high degree, could suggest multiple interactions and possible participation in many different reactions.
- **First neighbour network:** A network generated from all nodes sharing an interaction with the target node. The nodes that share this interaction could potentially be functionally related, and could give indications as to the function of unknown nodes by their associations with well characterised nodes and their interactions.
- **Clustering/module analysis:** Clustering also known as modularity, refers to a set of nodes that appear to be more highly connected to one another than to the rest of the network (Mason & Verwoerd, 2007). The identification of clusters within the network and their interactions could indicate the presence of underlying systems.
- **Local cohesive co-efficient;** The local cohesive co-efficient, describes the local cohesiveness of a node or set of nodes to other local nodes within the network (Albert, 2005; Mason & Verwoerd, 2007), potentially indicating the presence of a cluster. These values are given as a ratio. Values close to and including one, indicates that the node and close surrounding area of that node is highly connected. Values approaching and including zero, indicate less of a connection with surrounding nodes.
- **Path:** The path describes the set of edges that are traversed when connecting any two nodes together (Diestel, 2005).

- **Path length**; The path length describes the number of edges that are traversed between target nodes (Narsingh, 2004). If the path length is less than the average for the network, it could indicate a potential pathway between nodes.
- **Shortest path;** Describes the path of the least number of edges linking two nodes together. Multiple shortest paths may exist, and they could indicate the default pathway associations between nodes.

1.2.4 Systems Biology, the study of complex systems in a biological context

The adoption of a systems level approach to improve our understanding of biological systems began with the recognition of the limitations of traditional classical approaches. Molecular biology has traditionally adopted a reductionist approach, where the individual components of a system are considered outside the context of the entire cell. To develop an understanding of the entire system, research would first focus on individual components, trying to reconstruct the biological system to which it belonged in what has become known as a "bottom up" approach (Bruggeman & Westerhoff, 2007; Kitano, 2000). Although such an approach has been successful in explaining the chemical basis of numerous living processes, its limitations for analysing the enormous complexity of biological systems are being increasingly recognised to provide few insights into observed emergent properties (Van Regenmortel, 2004). Bottom up approaches are beginning to be complemented by systems level approaches that attribute specific behaviours and properties to emergent properties of the whole system (Little *et al.*, 1999), replacing the traditional approach, with that of a "top down" holistic approach (Kitano, 2000). Currently both of these complementary approaches are being applied as the technology and methods for the latter are developed and tested.

The first attempts to apply complex systems theory to biology was taken in the 1960's (Short, 2009). However the experimental techniques and approaches available at the time produced insufficient quantitative data to make this approach viable, and interest was lost. With the emergence of various technological and methodological advances, particularly in computational biology and high-throughput technologies that have focused (Galperin & Ellison, 2006), (Alm & Arkin, 2003) on genomics and proteomics (Kitano, 2002a), interest has been revived (Ahn *et al.*, 2006; Short, 2009). These advances have provided researchers with "omic" scale quantitative data sets and the means by which to process, model and analyse them.

The application of Systems Biology, to biological systems analysis will not only increase our current knowledge of biological processes and systems, but will continue to alter the ways in which research is conducted, shifting from traditional hypothesis to data driven experimentation approaches.

1.2.4.1 Functional interaction network models, an integrated approach

The relationships between genes and the proteins they encode within biological systems can be mapped as "traditional" graphs, using physical, functional or genetic data sets, with the coverage and accuracy of the produced graph reliant on the collected data (Hallinan & Wipat, 2007).

Functional interaction networks, have been developed to complement these "traditional" graphs (Lee *et al.*, 2004). Using a variety of integration techniques, data from multiple sources can be combined, to make the best predictions from multiple data sources overcoming the discussed limitations (Hallinan & Wipat, 2007) and giving insights that may previously not have been seen (Halinnan *et al.*, 2009). Such graphs represent a more accurate and better represention of the system that they model.

1.2.4.2 Probabilistic functional networks

Through the course of typical research, researchers acquire a mass of quantitative data, with often only a small proportion being used, the rest simply residing in archives, and databases as a multitude of different data formats. These large dataset collections have often been produced using diverse collection techniques, which have analysed multiple different aspects of organisms, making them ideal to generate comprehensive functional interaction networks. This task is being aided by new technologies and methods that are allowing the integration of the collections of data that are in different formats.

As an example of integrating data sets together, functional interaction networks that are themselves probabilistic (PFIN) – assigning probability of the reliability as to a given functional association can be produced. These networks allow the prediction of functional interactions between node and node products within the network, despite absences in known functional properties for node and node products. PFINs have been compiled for several *Bacillus* species (Craddock, 2008).

Craddock, investigated the secreted proteins (secretome) of 11 *Bacillus* species, both pathogenic and non pathogenic. A number of protein families were found to be either

common to all tested species or specific. A number of protein families with unknown functions were identified, from which a more detailed investigation was conducted.

To aid in determining the function of these unknown secreted protein families, data integration, modelling and analysis frameworks collectively termed SubtilNet were developed and applied to all the species studied. SubtilNet integrated 11 data sources, spanning genetic, biochemical and computational fields, using the approach adopted by Lee (Lee *et al.*, 2004). This data was then scored and weighted using a probabilistic Bayesian approach (Needham *et al.*, 2006), producing a PFIN. This PFIN was used to analyse the uncharacterised secreted protein families, determining their association with known genes and their products and then by using a "guilt by association" approach (Oliver, 2000) determining their function.

1.2.4.3 Examples of Systems Biology generated models

Systems Biology approaches have been used to simulate small and large scale systems, with examples including bacterial chemotaxis (Hansen *et al.*, 2008) feedback circuits (Oda *et al.*, 2005), signal transduction pathways (Bhalla & Iyengar, 1999; Schoeberl *et al.*, 2002) and simplified models of cell cycles (Chen *et al.*, 2000).

Attempts to create larger scale systems include, *E. coli* (Butland *et al.*, 2005) and *Helicobacter pylori* (Rain *et al.*, 2001) to virtual organs that represent essential features *in silico* (Bassingthwaighte *et al.*, 2009; Bassingthwaighte, 2000). The Physiome project, led by Denis Noble has produced one such example. Integrating multiple generated models produced from genetics and physiology data to produce a virtual human heart model (Noble, 2002). This has allowed for the prediction of drug side effects (Noble, 2005) and aid in the design of Ranolazine (Noble, 2008), an FDA approved anti-angina medication.

In another such example the drugs company Pfizer is applying Systems Biology in an attempt to model diabetes, so that it can develop treatments to it, as well as to other insulin resistant diseases.

1.3 Therapeutic compound discovery

1.3.1 Introduction

Illness is often attributed to the malfunction of a limited number of components within a biochemical pathway, and the interconnectedness of this pathway to others, results in wide spread knock on effects and the symptoms associated with illness.

Over the last century, numerous approaches have been used to discover and develop therapeutic compounds. These include the screening of vast repositories of artificially synthesised compounds against multiple targets in a trial and error approach, as occurs in many pharmaceutical establishments, the identification and isolation of active compounds from traditional effective herbal preparations, or the development of existing licensed compounds for new applications (Butcher *et al.*, 2004; Davidov *et al.*, 2003), following the discovery that side effects produced were desirable in other conditions and applications.

Successful candidates although producing desired effects, also often produce a plethora of undesirable effects, such as cellular toxicity, sub-optimal pharmacokinetics or cross reactions with other medications/biological systems. These side effects are a result of the non specific nature of the identified therapeutic compound, binding to and interfering with other biochemical pathways and systems (Hood & Perlmutter, 2004; Tatonetti *et al.*, 2009). As a consequence of this, manufacturers develop and modify these compounds in such a way as to produce a compromise between desired effects, while trying to limit and where possible remove observed side effects while maintaining effectiveness. It is because of this approach that current therapeutic design is subject to multiple stages of testing prior to certification, and use by the mainstream user.

1.3.2 Compound evaluation, certification and marketing

The identification, development and marketing, of successful therapeutic compounds involve multiple stages (Figure 1.9), beginning with the identification of potential therapeutic candidates. These candidates enter an optimisation stage where they are chemically altered, to improve their drug-like qualities. Successfully modified candidates become known as lead compounds, and proceed to the next stage of development, pre-clinical testing. In this stage lead compounds undergo a comprehensive set of analyses with respect to pharmacokinetics and dynamics, toxicity, teratogenicity, and carcinogenicity. This is conducted in a range of models systems that

increases in complexity giving the developer indications as to their likely biological activity and behaviour in clinical contexts. The time taken to proceed through these three stages of testing can range in time up to 10 years (Sindelar, 2002).

The lead compounds successfully passing pre-clinical testing stages enter a first phase of clinical trials, typically lasting up to six years (Sindelar, 2002). Here the compound is tested in 20-80 healthy volunteers to establish safety, tolerability and pharmacokinetics. This analysis is then repeated in 20-80 mildly symptomatic volunteers, while at the same time initial studies into the efficacy of the compound are carried out.

Compounds that pass the first clinical stage of testing enter a second phase of clinical trials, lasting approximately two years. This stage determines the efficacy and optimal dose of the lead compound after it's tested on 100-300 symptomatic volunteers. It is during this stage that existing drugs that are applying for certification for new applications enter testing.

Compound passing the second stage of clinical trials enter into the third and final clinical stage, lasting up to 4 years (Sindelar, 2002). During this stage evidence of the long term efficacy of the compound on 1000-3000 volunteers is determined and provided to licensing bodies.

The final stage of therapeutic compound development involves post market research which is aimed at monitoring the use of the drug for additional benefits/side effects and its suitability for use for other conditions.

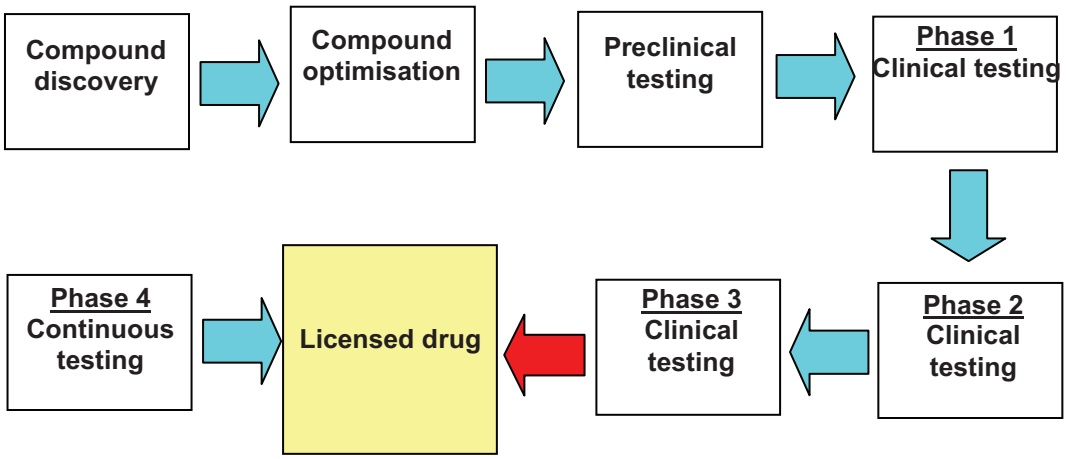


Figure 1.9: The steps involved in drug development. The developmental stages involved in the identification, and development of new therapeutic drugs.

1.3.3 Systems Biology and its application to therapeutic target discovery

The development of a new drug requires significant investments of time and capital. From initial compound testing to market takes on average 12 years, at a cost around 800 million dollars (DiMasi *et al.*, 2003), and with no guarantees of success. Due to this and the failure rates within the developmental process, pharmaceutical companies have always tried to predict and restrict the number of compounds entering the later, more costly developmental stages of drug development.

Typically 5000 candidate compounds will enter initial testing stages, with only one successfully licensed as a therapeutic compound (Carroll, 2007). This licensed compound must then be competitive enough to generate the developing company a profit of at least $500 million a year (Rawlins, 2004) to cover developmental costs. If successful, a drug can earn many times the capital spent on its development, protected for a minimum of 20 years by trade-related intellectual property rights (TRIPS). Lipitor® (Atorvastatin), a cholesterol lowering compound, earned its developers, Pfizer, $12.4 billion US (2008) and Avastin® (Bevacizumab) a monoclonal antibody preparation used to treat a range of cancers, earned its makers, Genentech, $2.6 billion US (2008).

With these factors in mind, developers are reluctant to focus on the development of new drugs for unproven targets and/or for unprofitable medical conditions tending to focus on compounds and conditions that are clinically well understood and with a high probability of being financially successful (Rawlins, 2004).

Realising the causes of illness results from the malfunction of specific components within a highly connected system, system biology is attempting to identify these individual targets, and allowing the development of compounds specifically to them, leaving surrounding targets and systems unaffected by them. This task has been aided by the discovery that many biological networks have a scale-free topology, and are composed of a few highly connected hub nodes which could potentially:

- Identify targets for future drug design
- Identify mechanisms that contribute to system
- Provide insights into how drug resistance develops

This approach will ultimately reduces the development time, produces more efficient, targeted and successful compounds with fewer side effects, reduced development costs and produce highly detailed evidential analysis, prior to any expensive laboratory testing.

The successful adoption of a complete system level analysis to drug design is still many years away. In the short term, systems level analysis will facilitate a better understanding of current drug action, particularly for the surprising large number drugs where their mode of action is poorly understood. This improved knowledge will also provide insights into how their activity can be improved, but ultimately, the wet lab approach to therapeutic design could be replaced with an entirely *in silico* approach. This has the potential to considerably reduce of the time taken to produce and licence a drug as well as holding out the prospect of developing new compounds/targets for currently un-studied/non profitable conditions as well as tailored genetic medications.

1.3.4 Systems Biology's benefits and limitations to future therapeutic target and compound discovery

As technology develops, so too does our understanding of key areas of biology. The technology and knowledge already acquired is slowly being adopted by researchers and pharmaceutical companies for use in system level approaches that compliment existing research methods. A paradigm shift will only be made when there is a more comprehensive adoption of systems level approaches for tackling the more recalcitrant challenges in biology that currently defies resolution by conventional approaches.

1.4 Aims and objectives

The following are the aims and objectives of this investigation:

- To develop and implement a testing method that can be used to validate gene candiates, identified by e-Therapeutics using a data driven approach, that could act as potential therapeutic targets.
- To develop and validate a functional interaction network, that can then be used to identify and analyse possible linkages between the identified candidates.
- To use the newly generated network to develop possible explanation for the experimental observations seen.
- To use a hypothesis driven approach to question the chromosomal organisation of genes involved in arginine and pyrimidine (uracil) biosynthesis in *B. subtilis*, and wether this leads to the targetting of individual system produced CP.
- To use the functional interaction network to investigate the arginine and pyrimidine biosynthetic pathways with respect to the questions raised by experimental analysis.

Chapter 2
Experimental Materials and Methods

2. Experimental Materials and Methods

2.1 Media and buffers

2.1.1 Luria-Bertani (LB) media (Sambrook *et al.*, 1989)

Per litre: 5g yeast extract, 10g sodium chloride, 5g tryptone was added to 900 ml of deionised water. The pH was adjusted to 7.0 with either sodium hydroxide (NaOH), or hydrochloric acid (HCl) and the volume made up to 1 litre with deionised water. The medium was autoclaved at 15 psi.

2.1.2 LB agar

As 2.1.1 but with 10g bacteriological agar added to the final volume of LB, before being autoclaved at 15psi.

2.1.3 LB for salt shock experiments

All LB was made to 1 litre as described in 2.1.1 with the quantity of NaCl replaced with those shown in table 2.1 to produce the final concentrations of 5,9,13,17,21 and 25 % (w/v).

Table 2.1: NaCl quantities to produce specific % NaCl (w/v) LB

NaCl (g/litre)	% (w/v)
50	5
90	9
130	13
170	17
210	21
250	25

2.1.4 Tris-Borate-EDTA (TBE) buffer (5 x stock)

Per litre: 53.9g Tris-base, 27.5g boric acid and 3.75g ethylenediaminetetraacetic acid (EDTA) was added to 800 ml of deionised water. The pH was adjusted to 8.0 using HCl, and the volume made up to 1 litre with deionised water.

2.1.5 Glycerol solution 80% (v/v)

Per 200 ml: 40 ml of deionised water was added to 160 ml of 100% glycerol (v/v). The mixture was autoclaved at 15 psi.

2.1.6 Minimal salts solution (5 x stock)(Bron, 1990)

Per 200 ml: 2 g $(NH_4)_2SO_4$, 14.8g of K_2HPO_4, 5.4g of KH_2PO_4, 1.9g of $Na_3C_6H_5O_7$ and 0.2g of $MgSO_4.7H_20$ was added to 180 ml of deionised water. The pH was adjusted to 7.0 and the volume made up to 200 ml with deionised water. The mixture was autoclaved at 15 psi.

2.1.7 Minimal growth medium (Bron, 1990)

Per 200 ml: 40 ml 5 x minimal salts solution, 2.5 ml glucose (40 % w/v), 400 µl tryptophan (10 mg/ml), 200 µl ammonium iron citrate (2.2 mg/ml) and the volume made up to 200 ml with deionised water, before being filter sterilised.

2.1.8 Starvation media (Bron, 1990)

Per 200 ml: 40 ml 5 x minimal salt solution, 2.5ml glucose (40 % w/v), and the volume made up to 200 ml with deionised water, before being filter sterilised.

2.1.9 Glucose solution 40% (w/v)

Per 10 ml: 4g glucose added to deionised water and the volume made up to 10 ml. The solution was filter sterilised and stored at 4^oC.

2.1.10 Phosphate buffered saline (PBS) (10 x stock)

Per litre: 1.44g of Na_2HPO_4, 8g of NaCl, 0.2g of KCl and 0.24g of KH_2PO_4 added to 800 ml deionised water. The pH was adjusted to pH 7.4 with HCl, and the volume made up to 1 litre with deionised water. The mixture was autoclaved at 15 psi.

2.1.11 Casamino acid solution 20% (w/v)

Per 10 ml: 2g Casamino acid, added to deionised water and the volume made up to 10 ml. The mixture was filter sterilised and stored at 4^{o}C.

2.1.12 Calcium chloride solution (0.1M)

Per litre: 11.1g $CaCl_2$ added to deionised water and made up to a final volume of 1 litre. The mixture was autoclaved at 15 psi.

2.1.13 Magnesium chloride solution (0.1M)

Per litre: 12.6g $MgCl_2$ added to deionised water and made to a final volume of 1 litre. The mixture was autoclaved at 15 psi.

2.1.14 Magnesium sulphate solution (1M)

Per 10 ml: 1.2g $MgSO_4$ added to deionised water and made to a final volume of 10 ml. The mixture was autoclaved at 15 psi.

2.1.15 Lysosyme solution (20 mg/ml)

Per 1 ml: 20 mg of lysosyme was added to 1x TE buffer with the volume being made up to 1 ml. It was then filter sterilised. This solution was prepared fresh.

2.1.16 Tryptophan solution (10 mg/ml)

Per 10 ml: 100 mg tryptophan added to deionised water and made to a final volume of 10 ml. The mixture was filter sterilised and stored at 4^{o}C.

2.1.17 TE buffer (10 x stock)

Per litre: 12.2g Tris-base and 3.0g EDTA was added to 800 ml of deionised water and the pH adjusted to 7.4 using HCl. The solution was made up to 1 litre and autoclaved at 15 psi.

2.1.18 Xylose solution 50% (w/v)

Per 20 ml: 10g xylose added to deionised water and made up to a final volume of 20 ml. The mixture was filter sterilised and stored at 4^{o}C.

2.1.19 IPTG solution (0.1M)

Per 10 ml: 238mg IPTG added to deionised water and made up to a final volume of 10 ml. The mixture was filter sterilised and stored at 4^{o}C.

2.1.20 Ammonium iron citrate solution (2.2mg/ml)

Per 10 ml: 22 mg ammonium iron citrate was added to deionised water and made up to a final volume of 10 ml. The mixture was filter sterilised, wrapped in foil to protect from light and stored at 4^{o}C.

2.1.21 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) running buffer (10 x stock)

Per litre: 30g Tris-base, 144g glycine and 10 g SDS was added to deionised water and made up to a final volume of 1 litre.

2.1.22 SDS resolving gel buffer (4 x stock)

Per litre: 180g Tris-HCl and 4g SDS was added to deionised water and made up to a final volume of 1 litre.

2.1.23 SDS stacking gel buffer (4 x stock)

Per litre: 60g Tris-HCl and 4 g SDS was added to deionised water and made up to a final volume of 1 litre.

2.1.24 Comassie stain

Per litre: 250 ml CH_3OH (HPLC) and 70 ml glacial acetic acid was added to deionised water and made up to a final volume of 1 litre, to which a 0.1% serva blue tablet was added and the solution filtered through Whatman #1 filter paper.

2.1.25 SDS-PAGE de-stain

Per litre: 250 ml CH_3OH (HPLC) and 70 ml glacial acetic acid was added to deionised water and made up to a final volume of 1 litre.

2.1.26 Antibiotics

All antibiotic used in this investigation and their concentration are described in Table 2.2. Antibiotics were filter sterilised and stored at -20^{o}C.

Table 2.2: Antibiotics and working concentrations used in this investigation

Antibiotic	Stock Concentration	Working Concentration (μg/ml)
Ampicillin (Amp)	100 mg/ml in sterile deionised water	50
Lincomycin (Lm)	1 mg/ml (100% ethanol)	25
Erythromycin (Em)	25 mg/ml in ethanol (50% v/v)	0.3
Nalidixic acid (Na)	10 mg/ml in sterile deionised water	300
Vancomycin (Vm)	10 mg/ml in sterile deionised water	2.0
Streptomycin (Str)	10 mg/ml sterile deionised water	80
Kanamycin (Km)	2 mg/ml sterile deionised water	10
Tetracyclin (Tet)	10 mg/ml in methanol 50% (v/v)	2.5
Rifampicin (Rif)	4 mg/ml in methanol 50% (v/v)	0.06
Chloramphenicol (Cm)	10 mg/ml in ethanol 70% (v/v)	7

2.1.27 Bacterial Strains

Table 2.3: Bacterial strains

Strain Name	Genotype	Comments	Source
B. subtilis	*trpC2*	Tryptophan-dependent, wild type strain	(Anagnostopoulos, 1961)
E.coli DH5α	Fϕ80*lac*ZΔM15Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(r_k^-, m_k^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1	General purpose cloning strain	New England Biolabs
E.coli BL21	B F- *dcm ompT hsdS*(r_B^- m_B^-) *gal*	General purpose expression strain, deficient in lon and ompT proteases	New England Biolabs
E. coli TB1	F^- *ara Δ(lac-proAB) [Φ80dlac Δ(lacZ)M15] rpsL*(Str^R) *thi hsdR*	General purpose cloning and expression strain	New England Biolabs
cheB KO	*trpC2 cheB*::ΔpSG1164	*B. subtilis* 168 with ΔpSG1164 disruption of *cheB*. Chloramphenicol resistant	This study
ybfS KO	*trpC2 ybfS*::ΔpSG1164	*B. subtilis* 168 with ΔpSG1164 disruption of *ybfS*. Chloramphenicol resistant	This study
luxS KO	*trpC2 luxS*::ΔpSG1164	*B. subtilis* 168 with ΔpSG1164 disruption of *luxS*. Chloramphenicol resistant	This study
ycdH KO	*trpC2 ycdH*::ΔpSG1164	*B. subtilis* 168 with ΔpSG1164 disruption of *ycdH*. Chloramphenicol resistant	This study

Table 2.3 (continued): Bacterial Strains

Strain Name	Genotype	Comments	Source
yjcH KO	*trpC2* *yjcH*::ΔpSG1164	*B. subtilis* 168 with ΔpSG1164 disruption of *yjcH*. Chloramphenicol resistant	This study
fbaB KO	*trpC2* *fbaB*::ΔpSG1164	*B. subtilis* 168 with ΔpSG1164 disruption of *fbaB*. Chloramphenicol resistant	This study
yacL KO	*trpC2 yacL*::pMUTIN	*B. subtilis* 168 with pMUTIN disruption of *yacL*. Erythromycin/lincomycin resistant	This study
yndH KO	*trpC2* *yndH*::pMUTIN	*B. subtilis* 168 with pMUTIN disruption of *yndH*. Erythromycin/lincomycin resistant	This study
licT KO	*trpC2 licT*::pMUTIN	*B. subtilis* 168 with pMUTIN disruption of *licT*. Erythromycin/lincomycin resistant	This study
ywdH KO	*trpC2* *ywdH*::pMUTIN	*B. subtilis* 168 with pMUTIN disruption of *ywdH*. Erythromycin/lincomycin resistant	This study
abnA KO	*trpC2* *abnA*::pMUTIN	*B. subtilis* 168 with pMUTIN disruption of *abnA*. Erythromycin/lincomycin resistant	This study
yvgQ KO	*trpC2 yvgQ*::pMUTIN	*B. subtilis* 168 with pMUTIN disruption of *yvgQ*. Erythromycin/lincomycin resistant	This study

2.1.28: plasmids

Table 2.4: Plasmids

Plasmid	Phenotype	Comments	Source
pSG1164	Amp^R (*E. coli*) Cm^R (*B. subtilis*)	*B. subtilis* integration vector with a xylose inducible (P_{xyl}) promoter.	(Daniel *et al.*, 1998)
ΔpSG1164	Amp^R (*E. coli*) Cm^R (*B. subtilis*)	*B. subtilis* pSG1164 integration vector deficient in *gfpmut1* gene	This study
pMUTIN	Amp^R (*E. coli*) Em^R Lm^R (*B. subtilis*)	*B. subtilis* integration vector with IPTG inducible (P_{spac}) promoter.	(Vagner *et al.*, 1998)
pMAL-p2X	Amp^R (*E.coli*)	Expression plasmid, used to produce a transcriptional fusion of an inserted gene sequence to a maltose binding protein (MBP)	New England Biolabs (NEB)
pMAL-p2X-*argF*	Amp^R (*E.coli*)	Expression plasmid with the *argF* gene (amplified with *argF* forward and *argF* reverse primers) within EcorI and SalI sites	This study
pMAL-p2X-*carA*	Amp^R (*E.coli*)	Expression plasmid with the *carA* gene (amplified with *carA* forward and *carA* reverse primers) within EcorI and SalI sites	This study
pMAL-p2X-*carB*	Amp^R (*E.coli*)	Expression plasmid with the *carB* gene (amplified with *carB* forward and *carB* reverse primers) within EcorI and SalI sites	This study
pMAL-p2X- *pyrAA*	Amp^R (*E.coli*)	Expression plasmid with the *pyrAA* gene (amplified with *pyrAA* forward and *pyrAA* reverse primers) within EcorI and SalI sites	This study
pMAL-p2X-*pyrAB*	Amp^R (*E.coli*)	Expression plasmid with the *argF* gene (amplified with *argF* forward and *argF* reverse primers) within Xba1 and SalI sites	This study
pGEX-6p-1	Amp^R (*E. coli*)	Expression plasmid, used to produce a transcriptional fusion of inserted gene sequence to a glutathionine S-transferase (GST) tag	GE Healthcare

Table 2.4 (continued): Plasmids

Plasmid	Genotype	Comments	Reference
ΔpSG1164-*cheB* KO	Amp^R (*E. coli*) Cm^R (*B. subtilis*)	Integration vector containing an internal section of the *cheB* gene, (produced from *cheB* KO forward and *cheB* KO reverse primers), between two XbaI sites	This study
ΔpSG1164-*ybfS* KO	Amp^R (*E. coli*) Cm^R (*B. subtilis*)	Integration vector containing an internal section of the *ybfS* gene (produced from *ybfS* KO forward and *ybfS* KO reverse primers), between two XbaI sites	This study
ΔpSG1164-*hxS* KO	Amp^R (*E. coli*) Cm^R (*B. subtilis*)	Integration vector containing an internal section of the *hxS* gene (produced from *hxS* KO forward and *hxS* KO reverse primers), between two XbaI sites	This study
ΔpSG1164-*fbaB* KO	Amp^R (*E. coli*) Cm^R (*B. subtilis*)	Integration vector containing an internal section of the *fbaB* gene (produced from *fbaB* KO forward and *fbaB* KO reverse primers), between two XbaI sites	This study
ΔpSG1164- *yjcH* KO	Amp^R (*E. coli*) Cm^R (*B. subtilis*)	Integration vector containing an internal section of the *yjcH* gene (produced from *yjcH* KO forward and *yjcH* KO reverse primers), between two XbaI sites	This study
ΔpSG1164- *ycdH* KO	Amp^R (*E. coli*) Cm^R (*B. subtilis*)	Integration vector containing an internal section of the *ycdH* gene (produced from *ycdH* KO forward and *ycdH* KO reverse primers), between two XbaI sites	This study
pMUTIN-*yacL* KO	Amp^R (*E. coli*) Em^R Lm^R (*B. subtilis*)	Integration vector containing an internal section of the *yacL* gene (produced from *yacL* KO forward and *yacL* KO reverse primers), between HindIII and BamHI sites	This study
pMUTIN-*yndH* KO	Amp^R (*E. coli*) Em^R Lm^R (*B. subtilis*)	Integration vector containing an internal section of the *yndH* gene (produced from *yndH* KO forward and *yndH* KO reverse primers), between HindIII and sacII sites	This study

Table 2.4 (continued): Plasmids

Plasmid	Genotype	Comments	Reference
pMUTIN-*licT* KO	Amp^R (*E. coli*) Em^R Lm^R (*B. subtilis*)	Integration vector containing an internal section of the *licT* gene (produced from *licT* KO forward and *licT* KO reverse primers), between HindIII and sacII sites	This study
pMUTIN-*ywdH* KO	Amp^R (*E. coli*) Em^R Lm^R (*B. subtilis*)	Integration vector containing an internal section of the *ywdH* gene (produced from *ywdH* KO forward and *ywdH* KO reverse primers), between HindIII and BamHI sites	This study
pMUTIN-*abnA* KO	Amp^R (*E. coli*) Em^R Lm^R (*B. subtilis*)	Integration vector containing an internal section of the *abnA* gene (produced from *abnA* KO forward and *abnA* KO reverse primers), between HindIII and sacII sites	This study
pMUTIN-*yvgQ* KO	Amp^R (*E. coli*) Em^R Lm^R (*B. subtilis*)	Integration vector containing an internal section of the *yvgQ* gene (produced from *yvgQ* KO forward and *yvgQ* KO reverse primers), between NotI and BamHI sites	This study

2.2 Maintenance

2.2.1 Strain maintenance

All strains were streaked onto fresh LB agar plates with the appropriate antibiotic and inducer if applicable (1mM IPTG/1% (w/v) xylose), and incubated overnight at 37°C. The following day a single colony was inoculated using a wire loop into 5 ml LB media containing appropriate antibiotics and grown overnight at 37°C with shaking at 200 rpm. The following day 750 µl was removed and added to 750 µl 80% (v/v) glycerol and frozen at -80°C.

2.2.2 Maintenance of genomic, plasmid and primer DNA

B. subtilis genomic and plasmid DNA was stored in sterile 1 x TE buffer at 4°C. Primers were supplied lysophilised and were re-hydrated to a final concentration of 100 mM with sterile double deionised water and stored at -20°C. Working stocks were also produced at a concentration of 20mM and maintained at 4°C.

2.3 Molecular Methods

2.3.1 Isolation of chromosomal DNA

The extraction of chromosomal DNA was performed using the Qiagen DNeasy blood and tissue kit. A single colony of *B .subtilis* was inoculated into 10 ml of LB with appropriate antibiotics in a 250 ml conical flask and grown overnight at 37°C with shaking at 200 rpm. The following day, 3 ml of culture was removed and distributed to two micro-centrifuge tubes. Tubes were centrifuged at 11,000 *g* for 10 minutes and the supernatent removed.

Lysosyme (20mg) was dissolved in 1 ml of 1 x TE buffer and filter sterilised. 180 µl of this solution was used to resuspend a bacterial pellet and then this mixture used to resuspend the remaining pellet. The sample was incubated for 30 minutes at 37°C, after which 25 µl of proteinase K was added, together with 200 µl of buffer AL. The mixture was incubated for 30 minutes at 56°C, after which time 200 µl ethanol was added and mixed by vortexing. The mixture was then transferred to a separate DNA binding spin column and centrifuged at 6,000 *g* for 1 minute, discarding the flow through. 500 µl of AW1 buffer was added to the spin column and centrifuged for 1 minute at 6,000 *g* discarding the flow through. 500 µl AW2 buffer was added to the spin column and centrifuged at 11,000 *g* for 3 minute discarding the flow through. Finally, the spin

column was transferred to a fresh micro-centrifuge tube, 100 µl of AE buffer added and centrifuged for 1 minute at 6,000 *g*. This was repeated a second time. The eluted chromosomal DNA was then stored at -20°C.

2.3.2 Isolation of plasmid DNA

Plasmid isolation was performed using a Promega Wizard Plus SV miniprep DNA purification system. 10 ml of LB in a 250 ml conical flask with the appropriate antibiotic added was inoculated with a single tested colony of *E. coli* containing the plasmid, and incubated overnight at 37°C with shaking at 200 rpm. The following day, 3 ml of the overnight culture was pelleted in two micro-centrifuge tubes at 11,000 *g* for 5 minutes. Single pellets were suspended with 250 µl of resuspension solution and then this mixture used to resuspend the remaining pellet. 250 µl of cell lysis solution was then added, and the sample inverted 4 times. 10 µl alkaline protease solution (25 µg/ul) was added, the sample inverted 4 times and incubated at room temperature for 5 min. 350 µl neutralising solution was added, the sample, inverted 4 times and pelleted in a micro-centrifuge tube at 11,000 *g* for 10 min. The supernatant was carefully removed ensuring the pellet was not disturbed, and decanted to a spin column. The spin column was centrifuged at 11,000 *g* for 1 min and the flow through discarded. 750 µl wash solution was added to the spin column and centrifuged at 11,000 *g* for 1 min. The flow through was discarded and the column washed again. The spin column was placed in a sterile micro-centrifuge tube to which 100 µl of nuclease free water and 11 µl of 10 x TE buffer was added and incubated at room temperature for 1 minute. The column was then spun at 14,000 *g* for 1 minute and the eluted plasmid DNA stored at -20°C.

2.3.3 Polymerase chain reaction

The polymerase chain reaction (PCR) was used to amplify specific regions of DNA. In this investigation, PCR was used for both cloning and rapid diagnostic testing. The primers used in PCR reactions were designed to have a G+C content of less than 50%, a length of no less than 20 base pairs and to have melting temperatures (Tm) of no more than 5°C difference between complementing pairs. Primers were designed by hand and checked for compatibility, melting temperature and secondary structures on the Sigma-Aldrich website (www.Sigma-Aldrich.co.uk).

2.3.4 Standard PCR

For the production of genomic material used in the production of knockout (KO) strains, PCR reactions were conducted using *Taq* polymerase (NEB) in a 50 µl reaction mixture (Table 2.5). PCR was conducted over 30 cycles with the reaction conditions specified in Table 2.6.

Table 2.5: PCR reaction mixture

Reagent	Amount (µl per 50 µl)
Forward Primer (20 µM)	5
Reveres primer (20 µM)	5
dNTP (10 mM)	1
B. subtilis chromosomal DNA	5
ddH_20	28
10 x *Taq* buffer	5
Taq polymerase	1

Table 2.6: KO PCR conditions (over 30 cycles)

Phase	Temperature (^OC)		Time (Min)	
	pMUTIN 4	pSG1164	pMUTIN	pSG1164
Initial denaturation	95	95	5	5
Denature	94	94	1.5	1.5
Anneal	52	49	1.5	0.5
Extension	68	72	2	1
Soak	4	4	-	-

For the production of genomic material for heterologous protein production, PCR was conducted using the proof-reading polymerase enzyme Platinum® *Pfx* polymerase (Invitrogen) as per reaction mixture in Table 2.7 and reactions conditions in Table 2.8.

Table 2.7: PCR reaction mixture for the production of heterologous proteins

Reagent	Amount (µl per 50µl)
10 x Platinum® *Pfx* amplification buffer	5
$MgSO_4$ (50 mM)	1
dNTP mixture (10 mM)	1.5
Platinum® *pfx* polymerase	1
B. subtilis chromosomal DNA	5
ddH_20	33
Forward primer (10mM)	0.75
Reverse primer (10mM)	0.75

Table 2.8: Heterologous protein production PCR conditions (over 30 cycles)

Phase	Temperature (^OC)	Time (Min)
Initial denaturation	94	2
Denature	94	15
Anneal	55	30
Extend	68	1 per kb
Soak	4	-

2.3.4.1 Primers for KO experiments

Table 2.9: Primers for knockout experiments

Primer Name	Sequence 5'- 3'	Complementary
yacL KO forward	5'-CGC AAG CTT GCG GCA ATA GGA GCA ATT ATC-3'	*B. subtilis* 168
yacL KO reverse	5'-CAA GGA TCC GGA ATC ACG ATA ACT CCC TCT-3'	*B. subtilis* 168
yacL reverse	5'-ACT TGT GAC GAG TAC ATC AAT AT-3'	*B. subtilis* 168
fbaB KO forward	5'-GAT TCT AGA CGA AGG CGA TTT TGC AGG CGG-3'	*B. subtilis* 168
fbaB KO reverse	5'-TAC TCT AGA GAT ATC CGC ATA GCG GAC CCC-3'	*B. subtilis* 168
fbaB reverse	5'-ATT TCC TTA AAT CCG AGA TTC-3'	*B. subtilis* 168
ywdH KO forward	5'-CGG AAG CTT CAG GAT CTC CAT AAA TCC GAG-3'	*B. subtilis* 168
ywdH KO reverse	5'-ATA GGA TCC TTT CCT ACA GCT ACA CTG CCG-3'	*B. subtilis* 168
ywdH reverse	5'-TTG AGT ATT TTT CGA ATC ATT-3'	*B. subtilis* 168
yndH KO forward	5'-CGT AAG CTT GAA AGA GGC AAA GAC ATT CCA-3'	*B. subtilis* 168
yndH KO reverse	5'-CGT CCG CGG TTT CGC AAA CTT CGG ATG TAC-3'	*B. subtilis* 168
yndH reverse	5'-AAG AAT TGT ATT TTG CAC ATG-3'	*B. subtilis* 168
ybfS KO forward	5'-GAA TCT AGA CGG CCT TCA GCC TCC GTA TGA-3'	*B. subtilis* 168
ybfS KO reverse	5'-ACA TCT AGA GCC GGT TAC GGT GTT GCC TGT-3'	*B. subtilis* 168
ybfS reverse	5'-TTC ATT TGA ATG TGC TTT AGA T-3'	*B. subtilis* 168
ycdH KO forward	5'-CTG TCT AGA CTG TTA ATC CCA TCT TCC GTT-3'	*B. subtilis* 168
ycdH KO reverse	5'-CTG TCT AGA GTA TTC TTT TGC CAG ATA GCC-3'	*B. subtilis* 168
ycdH reverse	5'-GAT TTA ACC AAT AGT GAA TCT T-3'	*B. subtilis* 168
ywgQ KO forward	5'-TTA ACT AGC GGC CGC CAC CAC GGC AGT TAT TTG CAG-3'	*B. subtilis* 168
ywgQ KO reverse	5'-CGC GGA TCC TCA TGA TAT GCT CTC GTC CGC-3'	*B. subtilis* 168
ywgQ reverse	5'-TCG TCA ATA ATC TCT TCA ATC T-3'	*B. subtilis* 168

Table 2.9 (continued): Primers for knockout experiments

Primer Name	Sequence 5'- 3'	Complementary
luxS KO forward	5'-CTG TCT AGA GCC AAA TAA ACA GGC GAT GAA-3'	*B. subtilis* 168
luxS KO reverse	5'-CTG TCT AGA CGC AGC AGG TAT TTC TGT AAT-3'	*B. subtilis* 168
luxS reverse	5'-AAT ACT TTT AGC AAT TCT TCT TTA T-3'	*B. subtilis* 168
cheB KO forward	5'-CGT TCT AGA ACT CTT GAT GTT GAA ATG CCG-3'	*B. subtilis* 168
cheB KO reverse	5'-CGT TCT AGA GTC AGA TAA ATG ATT CAG CCG-3'	*B. subtilis* 168
cheB reverse	5'-GAT ATC TTC CAC ATG TTT AAT C-3'	*B. subtilis* 168
licT KO forward	5'-CGT AAG CTT GAT ATA CCG ATC GAG TGT ATG-3'	*B. subtilis* 168
licT KO reverse	5'-CGT CCG CGG GTG AAG CGA TTC TTC GTT GAA-3'	*B. subtilis* 168
licT reverse	5'-CTT GTT TAA CTA CCC TTT CTA T-3'	*B. subtilis* 168
yjcH KO forward	5'-CGT TCT AGA TTG CTG TCG AAA CGC GAA ATA-3'	*B. subtilis* 168
yjcH KO reverse	5'-CGT TCT AGA CGG TTC TGT AAA GTC CAG AAT-3'	*B. subtilis* 168
yjcH reverse	5'-TTC TGT AAA GTC CAG AAT GTT-3'	*B. subtilis* 168
abnA KO forward	5'-CGT AAG CTT TGG TCC AAT TAT GTG CCG AAT-3'	*B. subtilis* 168
abnA KO reverse	5'-CGT CCG CGG TAA GAG TAG GAG CTT CTA AGG-3'	*B. subtilis* 168
abnA reverse	5'-GGC CAG CCC GAG CTC CAA TTC A-3'	*B. subtilis* 168
Pmut	5'-TCC TAA CAG CAC AAG AGC-3'	pMUTIN4
pxyl	5'-TAT ATC TAA GAT CTT CAT GAA AAA C-3'	pSG1164

2.3.4.2 Primers for arginine and pyrimidine (uracil) biosynthetic system investigation

Table 2.10: Primers for arginine and pyrimidine (uracil) biosynthetic system investigation

Primer Name	Sequence 5'- 3'	Complementary
argF forward	5'-TTT TTT GAA TTC ATG CAC ACA GTG ACG CAA ACC-3'	*B. subtilis* 168
argF reverse	5'-TTT TTT GTC GAC TCA GCA GTT TTT TGA TGA TTC CCC-3'	*B. subtilis* 168
carA forward	5'-TTT TTT GAA TTC ATG GAA GGT TAT TTA GTG TTA GAA GAT GGG ACA-3'	*B. subtilis* 168
carA reverse	5'-TTT TTT GTC GAC TCA GGC ATG CGC GAT TTC TCT CCT-3'	*B. subtilis* 168
carB forward	5'-TTT TTT GAA TTC ATG CCT AAA GAC ACC AGT ATT TCA AGC-3'	*B. subtilis* 168
carB reverse	5'-TTT TTT GTC GAC TCA CTG TGT GCA TGA TGC CAC TTC CTT-3	*B. subtilis* 168
pyrAA forward	5'-TTT TTT GAA TTC ATG AAG AGA CGA TTA GTA CTG GAA AAC GG-3'	*B. subtilis* 168
pyrAA reverse	5'-TTT TTT GTC GAC TCA CGC GTT TTG GCA TAC CGC TTC-3'	*B. subtilis* 168
pyrAB forward	5'-TTT TTT TCT AGA ATG CCA AAA CGC GTA GAC ATT AAC-3'	*B. subtilis* 168
pyrAB reverse	5'-TTT TTT GTC GAC TCA TAT AGT GAC TGC CGC CTC CTG ATT-3'	*B. subtilis* 168
argF sequencing #2	5'-GCT GAC ATT CCG GTG ATC AAT GGA C-3'	*B. subtilis* 168
carA sequencing #2	5'-AAT GTG GCA GAG CAG GCT TCC GCC-3'	*B. subtilis* 168
pyrAA sequencing #2	5'-AAG AGG CAA ACG CAT TGT CTT GGT-3'	*B. subtilis* 168
pyrAB sequencing #2	5'-CAT CGC CGG CTA TAA AGA AAT CGA G-3'	*B. subtilis* 168
pyrAB sequencing #3	5'-CGG GCG ATA CAG ATG TGC TGA GA-3'	*B. subtilis* 168
pyrAB sequencing #4	5'-GGT TAT CCG GTA CTG GTA CGC CC-3'	*B. subtilis* 168
pyrAB sequencing #5	5'-TGA AAT CAA CAG GTG AAG TCA TGG G-3'	*B. subtilis* 168
carB sequencing #2	5'-TGA AGT CAT GCG TGA CAG CAA TAA-3'	*B. subtilis* 168
carB sequencing #3	5'-ATG GAA GCA GGG AGC GAT CTC TCT-3'	*B. subtilis* 168
carB sequencing #4	5'-ATC GGC GGA ATG GGC ATG ATC ATT-3'	*B. subtilis* 168
malE forward	5'-GGT CGT CAG ACT GTC GAT GAA GCC-3'	pMAL-p2X

2.3.5 Colony PCR

A rapid diagnostic PCR method was used for the validation of construct or mutant status. A single colony from an LB plate was resuspended in 20 µl of sterile double deionised water in a 1.5ml micro-centrifuge tube and placed in a 100°C water bath for 10 minutes. Samples were then centrifuges at 11,000 *g* for 5 minutes and the supernatant used in the *Taq* PCR reaction in place of genomic DNA**.**

2.3.6 Gel electrophoresis

The visualisation and sizing of PCR products and plasmids was performed by gel electrophoresis. Samples were mixed with 6 x loading dye (Promega) and loaded onto 0.8% agarose gels (100 ml 1 x TBE, 0.8g agarose and 5µl of ethidium bromide) against Promega 1 kb and 100 bp standards. The gel was electrophoresed at 70V for 50 minutes.

2.3.7 Purification of DNA

DNA was purified from agarose gel using the Qiagen QIAquick PCR purification Kit. DNA containing gel fragments were excised from the agarose gels and weighed within a micro-centrifuge tube. Three gel volumes of GC buffer were added to the fragment and heated to 50°C until the fragment had melted. The samples were vortexed intermittently as per the manufacturer's instructions. Once melted, a single gel volume of isopropanol was added to the gel mix and vortexed. The mixture was transferred to a DNA binding column and centrifuged at 11,000 *g* for 1 minute, The flow through was discarded, 500 µl of QG buffer added and the spin column centrifuged at 11,000 *g* for 1 minute. The flow through was again discarded, 750 µl of PE buffer added and the spin column centrifuged at 11,000 *g* for 1 minute. Following the removal of the flow through, the spin column was dried by centrifuging at 11,000 *g* for 1 minute and transferred to a fresh micro-centrifuge tube where 50 µl of EB buffer was added, the sample incubated for 1 min and centrifuged for 1 min. The eluted DNA was stored at -20°C.

2.3.8 Cloning

2.3.8.1 Restriction digest

Restriction digests of both plasmids and amplified chromosomal DNA was designed to generate overlapping ends. The reactions were conducted sequentially using restriction endonucleases (NEB) in a final volume of 40 µl at 37°C for 3 hours. Digestion reactions involving chromosomal DNA were inactivated by heating (where a suitable restriction endonuclease had been used) to 80°C for 20 min. Digested DNA not suitable for heat inactivation together with all plasmid digestions were subjected to gel electrophoresis, and the required fragments purified from the gels by the DNA purification procedure (2.3.7). The composition of the digestion mixture is shown in Table 2.11.

Table 2.11: Restriction digest reaction mixture

Reagent	Volume (µl)
Restriction enzyme 1 (20 Units/µl)	1
Restriction enzyme 2 (20 Units/µl)	1
DNA (30 ng/µl)	20
Buffer	4
10 x Bovine serum albumen (BSA)	4
Sterile double deionised water	10

2.3.8.2 Ligation reaction

Purified linearised plasmid and insert DNA were quantitated and ligated together in a vector to insert ratio of 1:4 with T4 DNA Ligase (Promega). The reaction mixture was incubated overnight at 4°C. The composition of the ligation mixture is shown in Table 2.12.

Table 2.12: T4 DNA ligase reaction mixture

Reagent	Volume (µl)
T4 DNA Ligase (2 units/µl)	1
10 x T4 DNA Ligase buffer	1
Insert	Calculated according to size
Vector	100 ng
Sterile double deionised water	To a final reaction mixture volume of 10µl

2.3.9 *E. coli* transformation

2.3.9.1 Generation of chemically competent *E. coli* cells

E. coli was treated chemically to induce competency. 100 ml of LB in a 250 ml conical flask was inoculated from an overnight culture and grown to mid-exponential phase (OD_{600nm} 0.5). Cells were pelleted for 5 minutes at 3,000 *g* and 4 ^{o}C and the supernatent discarded. Cells were resuspended in 25 ml of ice cold 0.1M $MgCl_2$ and left to incubate on ice for 10 minutes. Cells were then pelleted at 3,000 *g* and 4 ^{o}C for 5 minutes, and the supernatent discarded. Cells were resuspended with 12.5 ml of ice cold $CaCl_2$ and incubated on ice for 30 minutes. Cells were pelleted at 3,000 *g* and 4 ^{o}C for 5 minutes before being resuspended in 2 ml of ice cold $CaCl_2$ containing 15% (v/v) glycerol. This

solution was distributed into 100µl aliquots in sterile micro-centrifuge tubes, snap frozen in liquid nitrogen and stored at -80^{o}C.

2.3.9.2 Transformation of chemically competent *E. coli* cells

Frozen competent *E. coli* cells (100 µl) were thawed on ice. Once thawed the 10 µl of ligation mixture was added to the cells, mixed gently and returned to ice for a further 30 min. The mixture was heat shocked at 42°C for 45 seconds and returned to the ice for a further 2 min. 900 µl of LB broth was added to each aliquot of transformed cells and incubated for 90 min at 37°C with shaking at 200 rpm, before being plated on to LB media containing a selective antibiotic.

2.3.10 *B. subtilis* transformation

2.3.10.1 Generation of competent *B. subtilis* cells (Bron, 1990)

B. subtilis becomes naturally competent towards the end of exponential growth phase in response to nutrient starvation. Cells from a single colony on a standard LB plate were inoculated in to 10 ml of minimal medium in a 250 ml conical flask and grown overnight with shaking at 200 rpm at 37^{o}C for 18 hours. The following day 1.4 ml of overnight culture was added to a 250 ml conical flask containing 10 ml of fresh minimal medium. This was grown at 37^{o}C with shaking at 200 rpm for 3 hours after which time 11 ml of starvation medium was added. The culture was grown for a further 3 hours after which 2.49 ml of 80% (v/v) glycerol was added and 1 ml aliquots removed and distributed to sterile micro-centrifuge tubes, snap-frozen in liquid nitrogen and stored at -80^{o}C.

2.3.10.2 Transformation of competent *B. subtilis* cells (Bron, 1990)

10µl of plasmid DNA was added to 400 µl of competent *B. subtilis* cells and incubated for 1 hour at 37^{o}C with shaking at 200 rpm, before being plated on to selective antibiotic LB plates.

2.4 Screening and validation of transformants

2.4.1 Validation of pMUTIN4 and pSG1164 KO mutants

Chromosomal DNA was used in a diagnostic PCR to confirm the authenticity of all putative KO's. This was done when the single mutations were generated with either pMUTIN4 or pSG1164, and after the single mutations were combined in a single cell to

generated double mutants. The primer pairs used for the validation of mutations involved a forward primer that annealed to either the P_{spac}, or P_{xyl} promoters of the pMUTIN4 or pSG1164 plasmids respectively, and the reverse primer used to generate the insert. Clones positive for this initial screen were then confirmed with a primer pair that annealed to sequences flanking the insert. Each insert, as well as areas outside of insert integration were used to perform PCR reactions on chromosomal DNA extracted from *B. subtilis* mutants confirming the presence of inserts as well as their orientation.

2.5 Analysis of therapeutic targets

2.5.1 Sample preparation for high-throughput analysis

To maximise the reproducibility of data from the high-throughput analyses, the generated mutants were prepared ready to be tested in a high-throughput manner in the following way. Mutant strains were streaked from glycerol stock onto fresh LB agar containing the appropriate inducer (1mM IPTG/1% (w/v) xylose), and grown overnight at 37^{o}C. The following day a single colony from each mutant was inoculated into 3 ml of LB in a test tube with loose fitting lid, containing the appropriate inducer and grown overnight for 18 hours. The following day the cultures were diluted 10-fold and the OD determined at 600 nm. The OD was used to determine the amount of the overnight cultures required to inoculate 20 ml pre-warmed LB containing inducer in 250 ml flasks to an OD of 0.01. Cultures were grown at 37^{o}C with shaking at 200 rpm in a water bath until an OD of 0.5. 12 ml of cultures were removed and added to 4 ml of ice-cold 80 % glycerol (v/v). A multi channel pipette was then used with pre-cooled tips to aliquot 250 µl of culture to pre-cooled PCR tubes and stored at -80 ^{o}C.

2.5.2 Sample preparation for low-throughput analysis

Mutant strains were streaked from glycerol stock onto an LB agar containing the appropriate inducer (1mM IPTG/1% (w/v) xylose) and grown overnight at 37^{o}C. The following day a single colony from each mutant was used to inoculated 3 ml of LB containing inducer and grown overnight for 18 hours with shaking at 200 rpm at 37^{o}C. Pre-warmed 250 ml conical flasks containing 25 ml of LB broth, and inducer, were inoculated with overnight cultures to an OD of 0.01, and the cultures grown to an OD of 0.5. 5ml were then added to 1ml of ice-cold 80% glycerol (v/v). The suspension was mixed by vortexing, divided into 600 µl aliquots and frozen at -80^{o}C.

2.5.3 High-throughput mutant testing

96 well microplates were prepared within a 37°C environment. All items of equipment were pre-warmed and the transport of materials and samples to and from the preparatory environment to the pre-warmed microplate reader was done using insulated boxes. The inducers xylose (1% w/v) and IPTG (1mM) were added to 2 x 30 ml of pre-warmed LB media and a third 30 ml of pre-warmed LB containing both inducers all in 250 ml conical flasks. From these 140 µl aliquots were added to each well of a 96 well micro plate with the exception of the wells in row H (control). To these wells 150 µl of LB was added. Prepared suspensions of the mutants were removed from the -80°C freezer and warmed to 37°C for 20 minutes in a water bath with samples being gently vortexed every 5 minutes. 10 µl aliquots of the mutant suspension were added to each well using a multi-channel pipette, with each row containing a different mutant and ensuring that no air bubbles were present. A gas permeable membrane was applied to the microplate with a brayer, and transferred to a pre-warmed (37°C) microplate reader.

For the analysis of mutant responses to various stressors, samples were stopped at the end of cycle 12, corresponding to the mid-exponential growth phase. The microplate was removed from the reader, transferred to the 37°C incubator and the gas permeable membrane removed. A 50 µl solution comprising pre-warmed LB broth containing inducer and stressor at 4 x the required final concentration was added to each of the wells including the control. A new gas permeable membrane was applied and the plate returned to the microplate reader. This procedure was conducted within 8 minutes. In the case of addition of the stressor heat to the samples, 50 µl of LB with appropriate inducer was added to each well during the second stage before being returned to a 45 °C microplate reader. In the case of NaCl stressor addition, 50 µl of 1, 5, 9, 13, 17, 21 and 25 % (w/v) pre-warmed NaCl LB with appropriate inducer was added, to produce the final concentrations 1, 2, 3, 4, 5, 6, 7 % (w/v) repectivly in a final 200 µl volume of LB.

2.5.4 Low-throughput mutant testing

A 600 µl aliquot of the previously prepared and frozen mutant culture was removed from the -80°C freezer and thawed in a 37°C water bath for 10 minutes. After incubation, the aliquot was added to 25 ml of pre-warmed LB in a 250 ml conical flask. The culture was incubated at 37°C in a shaking water bath. At an OD of 0.3 (mid-

exponential phase), the designated stressor was applied to the appropriate concentration and growth monitored regularly until the sample entered stationary phase.

2.5.5 Microscopy

2.5.5.1 Slide preparation and mounting

Poly-L-lysine (35µl 0.01%) was applied to each well of a glass slide, and incubated for 2 minutes. The poly-L-lysine was aspirated and slides allowed to air dry. 15 µl of culture was transferred to the surface of the slide, left in place for 1 minute, allowing the attachment of cells after which the remaining liquid was aspirated. The slide was air dried and then 2 µl of 50% glycerol (v/v) was applied to the surrounding area of each well and a cover slide applied.

2.5.5.2 DAPI staining

25 ml LB broth in a 250 ml conical flask was pre-warmed to 37ºC in a shaking water bath at 200 rpm and inoculated with 600 µl of previously thawed cells (2.6.4) and grown to an OD of 0.3. A 100 µl sample was removed immediately prior to the addition of stressor and then again every 10 min for 40 min after addition. 500 µl of ice-cold phosphate buffered saline (PBS) was added to each sample and the cells pelleted by centrifugation (13,000 *g*, 5 min, 4ºC). The supernatant was removed, and the pellet resuspended in 100 µl of ice cold PBS and 1 µl DAPI stain (100 µg/ml) added. The mixture was incubated for 5 minutes on ice, after which time the cells were mounted onto a poly-L-lycine-treated microscope slide (2.5.5.1).

2.5.5.3 Vancomycin staining

25 ml LB broth in a 250 ml conical flask was pre-warmed to 37ºC in a shaking water bath at 200 rpm and inoculated with 600 µl previously thawed cells and grown to an OD of 0.3. Fluorescent vancomycin was added to a concentration of 1 µg/ml combined with standard vancomycin (1 µg/ml) to produce a final concentration of 2 µg/ml.

2.6 Production of heterologous proteins

2.6.1 pMAL expression system

The pMAL expression vector is designed for the high-level production of recombinant proteins fused to a maltose binding protein (MBP)-derived affinity tag. The fusion of target proteins to MBP has been shown to increase their solubility (Kapust & Waugh, 1999) and for this reason the pMAL system was chosen to express components of the arginine and pyrimidine biosynthetic pathways. Gene sequences were inserted into the pMAL-p2X vector, down stream and in frame with the *malE* gene encoding the MBP. The resulting expressed protein would be fused to the MBP affinity tag (di Guan *et al.*, 1988; Maina *et al.*, 1988), allowing for affinity purification on an amylose resin column. Following affinity purification, the protein of interest would be cleaved from the MBP using the protease factor Xa (Kellermann & Ferenci, 1982) which targets a recognition site located on the linker between the expressed protein and the MBP.

E. coli cells containing recombinant expression vectors were grown overnight in 5 ml LB in tubes with loose fitting lids at 37^{o}C with ampicillin (100 µg/ml) and shaking at 200 rpm. 1 ml from this culture was used to inoculated 100 ml LB with ampicillin (100 µg/ml) in a 250 ml conical flask. The culture was grown at 37^{o}C to 0.5 OD with shaking at 200 rpm. Fresh ampicillin was added (100 µg/ml) to the culture at reaching an OD of 0.5 and before being split into 10 ml aliquots in tubes with loose fitting lids. IPTG was added to each individual culture to a final concentration of 1 mM, 0.3mM or 0mM respectively and cells grown at 37^{o}C with shaking. Two 1ml aliquots were removed before the addition of IPTG and every 2 hours thereafter for 4 hours. Cells were centrifuged at 13,000 *g* for 5 min, the supernatant removed and samples frozen at -20^{o}C.

2.6.2 Screening and validation of heterologous proteins

2.6.2.1 SDS-PAGE gel preparation (Laemmli, 1970)

10 % (w/v) polyacrylamide resolving gels were cast from the following 20 ml solution. 5 ml 4 x SDS-PAGE running buffer, 5 ml 40 % (w/v) acrylamide, 10 ml deionised water, 100 µl ammonium persulphate (10 % w/v) and 10 µl TEMED. The solution was pipetted between two 1 mm glass plates and overlaid with 1 ml isopropanol. The gels were left to polymerise for 2 hours, then washed with deionised water removing the isopropanol. The resolving gel was overlaid with 2 ml of stacking gel produced from the

following (10 ml). 2.5 ml 4 x SDS stacking gel buffer, 1.25 ml 40 % acrylamide (w/v), 6.25 ml deionised water, 60 µl ammonium persulphate (APS) (10 % w/v) and 30 µl TEMED. Combs were inserted and the gels allowed to polymerise for a further 2 hours.

2.6.2.2 SDS-PAGE sample preparation

Each 1 ml aliquot of frozen culture pellet samples was resuspended in 1ml ice cold 1 x SDS-PAGE running buffer, and sonicated for 10 seconds before incubating on ice for 30 seconds. The sonication treatment was repeated and the sample stored on ice. A 20 µl sample of sonicated suspension was removed from each time-point sample and added to 4 µl of loading dye and labelled "whole cell extract". The remaining sonicated mixture was centrifuged at 14,000 *g* for 10 minutes at 4^{o}C and 20 µl of the resulting supernatant added to 4 µl of loading dye and labelled "soluble cell extract".

2.6.2.3 Running and visualisation of SDS-PAGE gels

Protein samples for SDS-PAGE analysis were loaded as a time series with the whole cell extracts being run along side the equivalent soluble cell extracts. A Bio-Rad precision plus protein standard was included for size estimations. Samples were electrophoresed in 1 x SDS-PAGE running buffer at 80 V until the loading dye had reached the bottom of the stacking gel at which point the voltage was increased to 150 V and electrophoresis continued until the marker dye had reached the bottom of the resolving gel. The gel was immersed in Coomasie blue stain overnight to fix and stain the proteins. The following day the gel was immersed in de-staining solution which was regularly changed until the bands were visible against a clear background.

Chapter 3

Therapeutic Target Selection and Experimental Testing Approaches

3. Therapeutic target selection

At the start of this investigation, discussions were held with our industrial partner, e-Therapuetic's, with respect to potential pairs of genes that could act as anti-infective targets when targeted together. These pairs of genes were identified using in-house data driven computational approaches. This chapter describes these pairs and an iterative experimental approach that was designed to test the validity of their predictions. In parallel with this, *in silico* models were developed to also allow computational analysis (Chapter 5).

3.1 Introduction to the e-Therapeutics network

Traditionally, many antimicrobial drugs have been designed to target a limited number of essential components within the metabolism of micro-organisms. Because cellular metabolism contains critical steps, biological systems have developed mechanisms to resist these compounds and improve the robustness of their metabolic pathways. The resulting mechanisms limit the use of pathway-blocking compounds, reducing or abolishing their effectiveness.

More recently, *in silico* Systems Biology approaches that allow a data-driven approached to biological research have been developed. This has allowed the identification of multiple, seemingly non-essential components within metabolic systems, that when targeted together are capable of generating effective system wide responses. Using such an approach to develop therapeutic compounds produces more focused responses, expands the number of potential targets available, while at the same time reducing the side effects associated with traditional approaches of compound development. An added benefit of using this approach is a reduced likelihood of an organism developing resistance (Kitano, 2002a) to such developed compounds. As there is such a large repertoire of potential targets, the organism must take time to develop methods to negate the effects of targeting them.

e-Therapeutics plc is a drugs discovery and development company, based in the United Kingdom and India. Using an *in silico* Systems Biology approach, the company has developed new technologies for identifying new drug targets and predicting their effects in a range of organisms. In particular, they have analysed the potential effects of existing licensed drugs, both individually and in combination, to identify novel

applications. Adopting this approach, targets and therapeutic compounds can be discovered, tested and optimised in considerably less time than conventional screening approaches, with reduced costs and licensing implications. Currently, this approach has led to the development of compounds to treat asthma, a novel-mechanism antidepressant, antibiotics to treat MRSA, and a cancer chemotherapy treatment that works at safe doses in a very short period of time.

e-Therapeutics construct and simulate complete biological networks, and apply network analysis to determine how best to perturb the network and achieve particular biological behaviours. From these simulations, non-essential components that significantly contribute to these behaviours, either individually or in combination, can be identified. Therapeutic compounds that act upon these components, both new and novel, are discovered, developed and optimised before continuous and rigorous multi-stage testing and (re)licensing for the new application.

3.2 Proposed therapeutic target gene candidates

Using the above described approach and proprietary confidential algorithms, e-Therapeutics have predicted the following non-essential (Kobayashi *et al.*, 2003) gene pairs in *B. subtilis* that when targeted together could produce notable effects on the growth of *B. subtilis*. These could then have the potential to be used as new anti-infective drug targets:

Pair 1

- ***ybfS****;* also known as *gamP* and *yzfA* (Lechat *et al.*, 2008), encodes a suspected glucosamine-specific enzyme IICBA component of the phosphotransferase system (PTS) (The universal protein resource (UniProt) 2009; Reizer *et al.*, 1999).
- ***ywdH****;* also known as *ipa-58r* (Lechat *et al.*, 2008) encodes a broad spectrum aldehyde dehydrogenase (The universal protein resource (UniProt) 2009; Kanehisa & Goto, 2000).

Pair 2

- ***yvgQ****;* also known as *cysL* (Lechat *et al.*, 2008) encodes the *ß*-subunit of a sulphite reductase (The universal protein resource (UniProt) 2009; Kanehisa & Goto, 2000; Lechat *et al.*, 2008; van der Ploeg *et al.*, 2001).

- ***luxS;*** also known as *ytjB* (The universal protein resource (UniProt) 2009; Lechat *et al.*, 2008), encodes an S-ribosylhomocysteine lyase, involved in the production of auto-inducer 2 protein (The universal protein resource (UniProt) 2009; Kanehisa & Goto, 2000; Lechat *et al.*, 2008).

Pair3

- ***ycdH;*** also known as *adcA* (The universal protein resource (UniProt) 2009; Lechat *et al.*, 2008), encodes a probable high-affinity zinc ATP binding cassette (ABC) transporter that is transcriptionally repressed by zinc (The universal protein resource (UniProt) 2009; Lechat *et al.*, 2008).

- ***yndH;*** encodes an uncharacterised hypothetical protein (Kanehisa & Goto, 2000; Lechat *et al.*, 2008). Basic local alignment search tool (BLAST) indicates a very close similarity to other uncharacterised hypothetical genes within other *Bacillus* species (Lechat *et al.*, 2008).

Pair 4

- ***yacL;*** encodes an uncharacterised protein (The universal protein resource (UniProt) 2009; Kanehisa & Goto, 2000) that has been shown to contain the protein domains TRAM, HIN and PIN (Kanehisa & Goto, 2000). BLAST analysis indicates a close similarity to similar uncharacterised genes within *Bacillus* species.

- ***fbaB;*** also known as *iol J* and *yxdL* (Lechat *et al.*, 2008) encodes a 6-phospho-5-dehydro-2-deoxy-D-gluconate aldolase (The universal protein resource (UniProt) 2009; Kanehisa & Goto, 2000; Lechat *et al.*, 2008).

Pair 5

- ***yjcH***; encodes a hydrolase (Lechat *et al.*, 2008). It has also been shown to posses the protein domains associated with two alpha/beta hydrolase folds and esterase domain (Kanehisa & Goto, 2000). BLAST analysis indicates a close similarity to other uncharacterised genes as well as esterase's within *Bacillus* species.

- ***abnA***; encodes an arabinan-endo 1,5-alpha-L-arabinase (The universal protein resource (UniProt) 2009; Kanehisa & Goto, 2000; Lechat *et al.*, 2008).

Pair 6

- ***cheB***; also known as *cheL* (The universal protein resource (UniProt) 2009; Lechat *et al.*, 2008), encodes a methyl-accepting chemotaxis protein (MCP). It is a glutamate methylesterase (The universal protein resource (UniProt) 2009; Kanehisa & Goto, 2000; Lechat *et al.*, 2008).
- ***licT;*** encodes a transcriptional anti-terminator of the BglG family (The universal protein resource (UniProt) 2009; Kanehisa & Goto, 2000; Lechat *et al.*, 2008).

3.3 Development of a testing method for validating selected targets

3.3.1 Approach

The aim of this investigation was to determine the influence of inactivating target genes of *B. subtilis* individually and in combination. Initial experiments would involve inactivating genes individually and determining any changes to the phenotype produced under a range of different stress conditions. Following this, the target genes would be inactivated in combination and their phenotypes would be tested under the same stress conditions. This approach would then enable the comparison of the effects of both single and double gene targets inactivation and assess their potential as therapeutic drug targets.

3.3.2 KO mutants

The inactivation of target genes in *B. subtilis* was achieved by the interruption of gene sequences using an integration vector, producing KO mutants. Vectors were designed to integrate into the target gene sequence *via* a single crossover homologous recombination event. Two compatible integration vectors were used to generate the KO mutants, pMUTIN4 and pSG1164. These vectors encode different antibiotic markers to facilitate the selection of developed KO mutants in later experiments.

Most *B. subtilis* genes are organised into polycistronic transcriptional units (Vagner *et al.*, 1998), and the interruption of an upstream gene would separate downstream genes from their natural promoter. This could have the potential to produce undesirable effects and phenotypes through downstream polar effects that could affect this investigation. Consequently, both vectors encode different inducible promoters that can be used to ensure the expression of genes downstream of the target gene.

3.3.2a pMUTIN4

The pMUTIN integration plasmid was developed by Valérie Vagner and colleagues at the French National Institute of Agricultural Research (INRA) in the late 1990s (see appendix). Its inception was due to the specific requirements of the international functional analysis consortium (18 European and 12 Japanese laboratories) aimed at characterising all genes encoded by the *B. subtilis* chromosome (Vagner *et al.*, 1998). The groups within the consortium used pMUTIN to inactivate each gene systematically and under a range of different growth conditions (Vagner *et al.*, 1998) study its effect on cell fitness. Based on pBR322 (Kaltwasser *et al.*, 2002), a well-established *E. coli* cloning vector (Bolivar *et al.*, 1977), pMUTIN is not able to replicate autonomously in *B. subtilis* and can only be stably maintained in this bacterium following integration *via* a homologous recombination event at the site of the target gene.

Four versions of the pMUTIN vector have been constructed (pMUTIN 1-4), ranging in size from 8.3 to 8.6 Kbp. They differ with respect to their internal terminator sequences, the sequences of the multiple cloning site and repressor/operator sites used to control the expression of a controllable promoter. All versions encode:

- An *E. coli*-active ColE1 origin of replication
- A *ß*-lactamase gene (Ap^R) for selection in *E. coli*
- An erythromycin resistance gene (Em^R) for selection in *B. subtilis*
- A modified *E. coli lacZ* reporter gene with a *B. subtilis* optimised ribosome binding site allowing transcriptional fusions and the monitoring of gene expression
- A P_{spac} IPTG-inducible promoter (Jana *et al.*, 2000) and associated "tight" repressor encoded by *lacI*.

3.3.2b pSG1164Δ

The pSG1164 integration plasmid, was developed by Peter Lewis and Adele Marston, at the Sir William Dunn School of Pathology, University of Oxford (Lewis & Marston, 1999) (see appendix). Developed to replace existing integration vectors that had been used to fuse fluorescent tags to *B. subtilis* proteins for microscopic analysis. pSG1164 encodes a *gfpmut*1 gene that produces a high fluorescence version of the standard green

fluorescent protein (GFP) (Cormack *et al.*, 1996). pSG1164 is a 5.5 kb plasmid, based on the pRD96 integration vector (Daniel *et al.*, 1998). It encodes:

- A xylose-inducible P_{xyl} promoter
- An *E. coli*-active ColE1 origin of replication
- A *ß*-lactamase gene (Ap^R) for selection in *E. coli*
- A chloramphenicol acetyltransferase gene (Cm^R) for selection in *B. subtilis*

In this investigation the *gfpmut1* gene was surplus to requirements and its expression may have produced unwanted side effects. As the gene was bordered by two *Xba*I restriction endonuclease sites, these were used to generate a version of pSG1164 (pSG1164Δ - 4.8 Kbp) that had the gene removed (Figure 3.1).

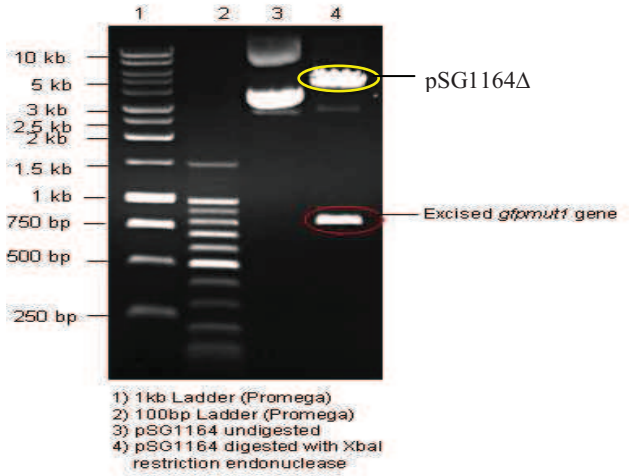


Figure 3.1: An agarose gel showing the digestion and removal of the *gfpmut1* gene from the pSG1164 plasmid. Lane 1, 1kb ladder. Lane 2, 100bp ladder. Lane 3, Undigested pSG1164. Lane 4, Digested pSG1164 (now linearised) using XbaI restriction endonuclease. Also present within this lane is the excised *gfpmut1* gene.

3.3.3 KO mutant construction and validation

The target genes were each designated to one of the integration vectors and used to generate the respective KO mutant (Table 3.1).

Table 3.1: Table indicating target genes and destination integration vector

pSG1164	pMUTIN4	pSG1164/pMUTIN4
ybfS	*ywdH*	*ybfS/ywdH*
luxS	*yvgQ*	*luxS/yvgQ*
ycdH	*yndH*	*ycdH/yndH*
fbaB	*yacL*	*fbaB/yacL*
yjcH	*abnA*	*yjcH/abnA*
cheB	*licT*	*cheB/licT*

A DNA sequence of approximately 500 bp, homologous to the 5' end of each candidate gene was amplified by PCR with terminal restriction sites. These were ligated into appropriately digested integration vectors. The target sequence of each clone was generated using primer pair A (Figure 3.2). The resulting recombinant plasmids were transformed in to *E. coli* and selected for by plating onto LB containing ampicillin. Plasmid DNA was isolated from several recombinant colonies and the presence and orientation of the insert in the vector determined using primer pair B (Figure 3.2). In this case the forward primer was specific for the integration vector promoter, while the reverse primer was insert specific and the same as that used to generate the insert fragment. A single verified clone was used to transform *B. subtilis,* selecting either with erythromycin/lincomycin (pMUTIN4) or chloramphenicol (pSG1164) antibiotics. The putative recombinant clones were checked for integration into the correct candidate gene using a vector-specific forward primer and a target-specific reverse primer located outside the original amplified insert- primer pair C (Figure 3.2). After the construction and confirmation of single KO mutants (Figure 3.3), double mutants were constructed. Chromosomal DNA was isolated from the pMUTIN4 generated KO mutant, and transformed into the appropriate pSG1164 generated single KO mutant to generate a double KO mutant (Figure 3.3). Double mutants were selected for using lincomycin/erythromycin and retested with primer pairs C (Figure 3.2) to confirm the presence of both plasmids (Figure 3.3).

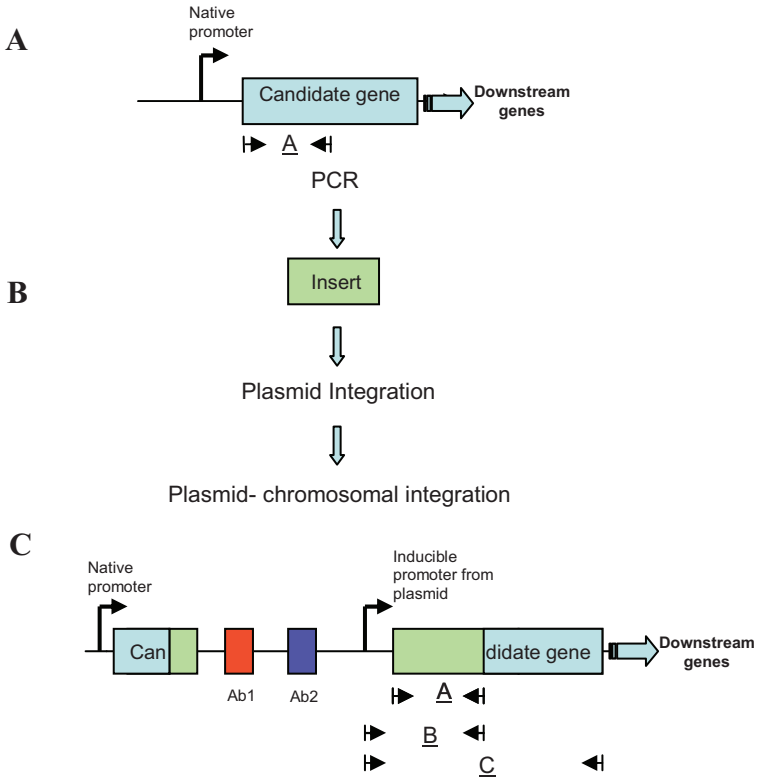


Figure 3.2: A graphical representation of the construction of KO mutants. A, Target gene on the bacterial chromosome together with its native promoter and downstream genes. B, Cloning of a 500 bp complementary fragment of the target gene and its integration into the integration vector. C, Disruption of the target gene with the integration vector, containing antibiotic selection markers (Ab1+Ab2) and inducible promoter.
The primer pair locations required to validate the produced KO mutants are Primer pair A, A forward primer for the beginning and reverse primer for the end of the inserted sequence. Primer pair B, A forward primer for the inducible promoter contained within the integration vector (P_{xyl}/P_{mut}), and a reverse primer for inserted sequence. Primer pair C, A forward primer for the inducible promoter contained within the integration vector and a reversible primer for a region outside of the integration site.

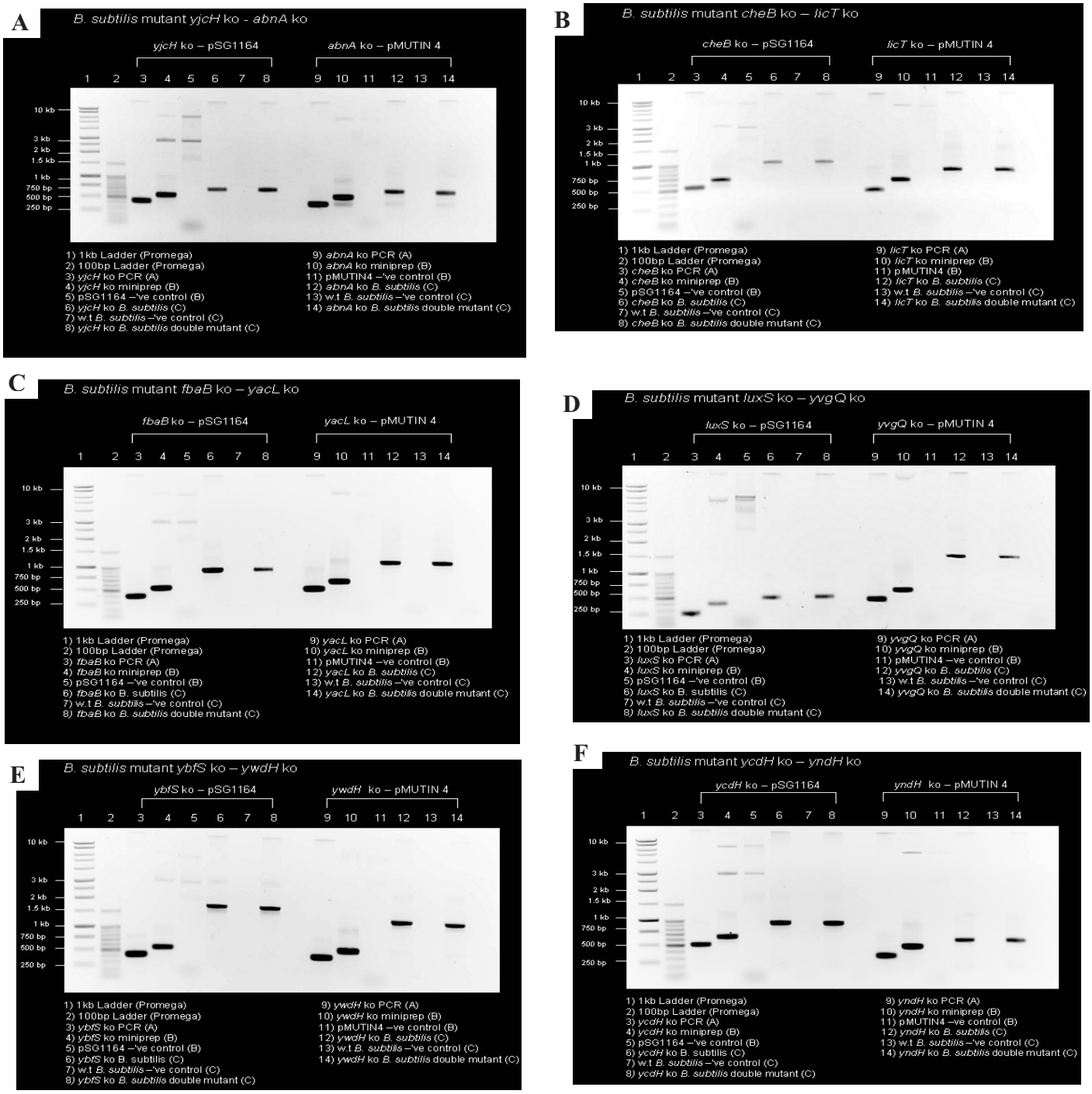


Figure 3.3: Agarose gels showing the stages in the construction of single and double KO mutants. Primer pairs are represented as A, B or C under each image. Primer pair A represents insert sequence. Primer pair B represents forward primer for the inducible promoter contained within the integration vector, and a reverse primer for inserted sequence. Primer pair C represents forward primer for the inducible promoter contained within the integration vector and a reversible primer for a region outside of the integration site represents. Gel A, *yjcH* KO-*abnA* KO. Gel B, *cheB* KO-*licT* KO. Gel C, *fbaB* KO-*yacL* KO. Gel D, *luxS* KO-*yvgQ* KO. Gel E, *ybfS* KO-*ywdH* KO. Gel F, *ycdH* KO-*yndH* KO.

3.4 Phenotype testing

e-Therapeutics supplied combinations of target genes, which were predicted to combine synergistically to increase susceptibility to stress. These combinations were to be subjected to a range of different stress's to test these hypotheses.

3.4.1 Stressors

To determine the effect of inactivating candidate gene, both singly and in combination, the growth phenotypes of the various *B. subtilis* mutants were to be compared with that of the wild-type under a range of different stress conditions. The aim was to determine whether the absence of a target gene function, either on its own or in combination with its cognate partner, could result in an increased susceptibility to a particular stress. The stresses to which the KO mutants were to be exposed covered a broad range of metabolic processes (Table 3.2).

Table 3.2: The areas of metabolism affected by each stressor

Stressor	Area of metabolism tested
Growth with/without inducer	Downstream polar effects
Nalidixic acid	DNA replication
Rifampicin	mRNA synthesis
Kanamycin	Protein synthesis
Streptomycin	Protein synthesis
Tetracycline	Protein synthesis
Vancomycin	Cell wall synthesis
Paraquat	Oxidative stress
Heat	Heat shock
NaCl	Osmotic shock

3.4.2 Stressor concentration

In order to conduct a meaningful comparative investigation, it was necessary to determine an appropriate working concentration of the various stressors. Using data from previous investigations, a range of stressor concentrations was established. These were tested against the wild-type grown to mid exponential phase (OD 0.3) in LB medium. The aim was to identify a stressor concentration that would affect the normal growth profile noticeably without leading to growth inhibition or cell death, and allow any increased sensitivity of the mutant to a stress to be observed.

The suggested concentrations used in experiments to inhibit growth varied (Table 3.3), and was dependent on strain, growth and testing conditions. The list described is not exhaustive. Table 3.4 describes the final stressor concentrations used in the final investigation.

Table 3.3: Suggested growth inhibiting concentrations of different stressors

Stressor	Suggested growth inhibiting concentration
Nalidixic acid	2 µg/ml (Rodriguez-Martinez *et al.*, 2008) 25-75 µg/ml (Schujman *et al.*, 2001) 300 µg/ml + (Vazquez-Ramos & Mandelstam, 1981)
Rifampicin	0.06 µg/ml (Bandow *et al.*, 2002) 0.125 µg/ml (Hutter *et al.*, 2004)
Kanamycin	1.25 µg/ml (Goldthwaite *et al.*, 1970) 2.5 µg/ml (Schirner *et al.*, 2009) 8 µg/ml (Rahman *et al.*, 2007)
Streptomycin	1000 µg/ml (Goldthwaite *et al.*, 1970) 12.5 µg/ml (Balasubramanian *et al.*, 2006)
Tetracycline	1-2 µg/ml (Andrews & Wise, 2002) 6.5 µg/ml (Coonrod *et al.*, 1971)
Vancomycin	0.1-0.5 µg/ml (Mota-Meira *et al.*, 2000) 0.4-0.5 µg/ml (Schirner *et al.*, 2009) 1 µg/ml (Mascher *et al.*, 2004)
Paraquat	0.4 mM (personal communication S. Pohl)
NaCl	6% (w/v)(Hoper *et al.*, 2006) 4% (w/v)(Hecker *et al.*, 1988)

The range of stressor concentrations (Table 3.3) was used to establish a starting point from which further investigations would be conducted to refine this value, to those that would be used in this investigation (Table 3.4).
The way in which wild type *B. subtilis* reacted to the different stressors varied both with respect to the cell growth profile and the time taken for the stressor to have an effect.

3.4.2.1 Nalidixic acid

The concentration of nalidixic acid chosen for this investigation was 300 µg/ml (Figure 3.4A). This concentration resulted in a decreased culture density of wild type *B. subtilis* during both exponential and early stationary phase when compared to untreated wild type *B. subtilis*. Increasing the concentration of nalidixic acid to 600 µg/ml resulted in a steady decline in culture density.

3.4.2.2 Rifampicin

The concentration of rifampicin chosen for this investigation was 0.06 µg/ml (Figure 3.4B). This concentration resulted in a decrease in culture density of wild type *B. subtilis* during both exponential and early stationary phase when compared to untreated wild type *B. subtilis* culture density.

3.4.2.3 Kanamycin

The concentration of kanamycin chosen for this investigation was 10 µg/ml. The effect of concentrations below this on wild type *B. subtilis*, were not clearly distinguishable from the untreated wild type *B. subtilis* (Figure 3.4C) culture density.

3.4.2.4 Streptomycin

The concentration of streptomycin chosen for this investigation was 80 µg/ml (Figure 3.4D). Lower concentrations produced wild type *B. subtilis* culture densities approximating, unstressed wild type *B. subtilis*. The 80 µg/ml concentration reduced the growth rate and total yield.

3.4.2.5 Tetracycline

The concentration of tetracycline chosen for this investigation was 2.5 µg/ml, a compromise between the culture densities generated by applying a concentration of between 1.25 µg/ml and 5 µg/ml to wild type *B. subtilis*, both of which appeared similar

(Figure 3.4E). The culture densities produced using a concentration of 10 µg/ml tetracycline was significantly different from the previous concentrations, and was not used. Using a final concentration of 2.5 µg/ml, there was a decline in culture density in the transition into stationary phase, but with a final culture density in stationary phase comparable to untreated *B. subtilis.*

3.4.2.6 Vancomycin

The concentration of vancomycin chosen for this investigation was 2 µg/ml. The kinetics of inhibition by vancomycin was different from that of the other antibiotics in that growth was initially inhibited, but growth resumed after cycle 45-50 (Figure 3.4F). Lower concentrations of vancomycin (e.g. 1 µg/ml) produced no discernable differences in growth profile when compared to the untreated control. A concentration of 4 µg/ml caused an irreversible decline in culture density when compared to untreated wild type *B. subtilis*.

3.4.2.7 Paraquat

The concentration of paraquat chosen for this investigation was 2.4 mM paraquat (Figure 3.4G), (despite being suggested a concentration of 0.4mM). This concentration caused the growth profile of wild type *B. subtilis* to enter stationary phase at an OD that was lower than that of untreated wild type *B. subtilis*. The concentration of 4.8mM caused the growth profile to decline after its addition while lower concentrations e.g. 1.2 mM had only minimal effects on growth profile.

3.4.2.8 Temperature

The temperature to which wild type *B subtilis* would be exposed was 45 ^{o}C, the maximum temperature achievable using the microplate reader. A slight overall difference in culture density while in stationary phase was seen between wild type *B. subtilis* grown at 37^{o}C to that grown at 45 ^{o}C (Figure 3.4H).

3.4.2.9 NaCl

The NaCl concentration chosen for this investigation was 5% (w/v). NaCl concentrations between 2% and 5% caused a notable difference in the growth profile as the cells entered stationary phase, while higher concentrations of 6% and 7% (w/v) showed similar but more profound affects (Figure 3.4I).

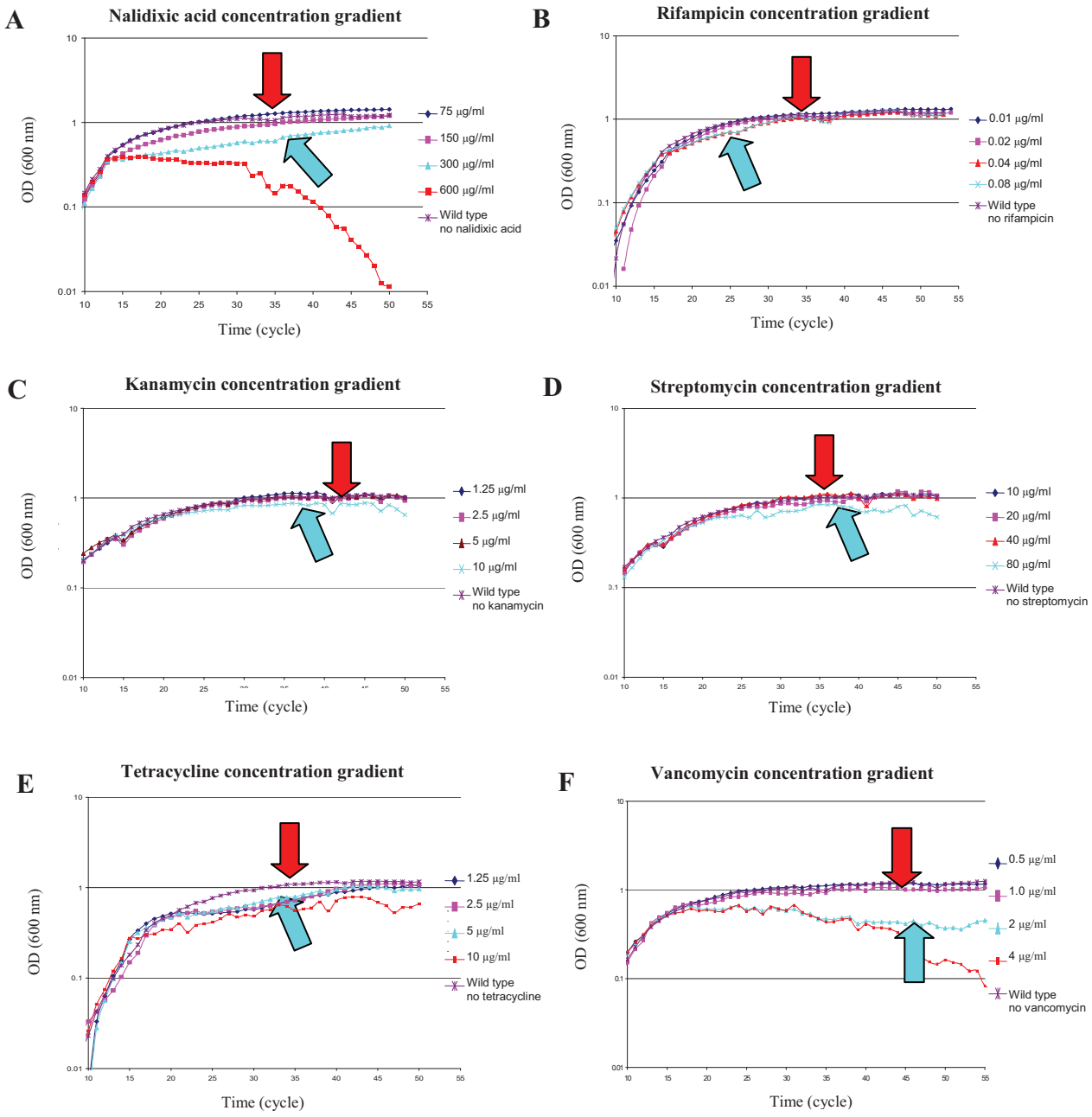


Figure 3.4: Graphs to show the determination of stressor concentrations to be used in this investigation. Red arrow indicates the growth profile of wild type *B. subtilis* with no stressor applied. The blue arrow indicates the concentrations of stressor selected for this investigation. A, Nalidixic acid. B, Rifampicin. C, Kanamycin. D, Streptomycin. E, Tetracycline. F, Vancomycin.

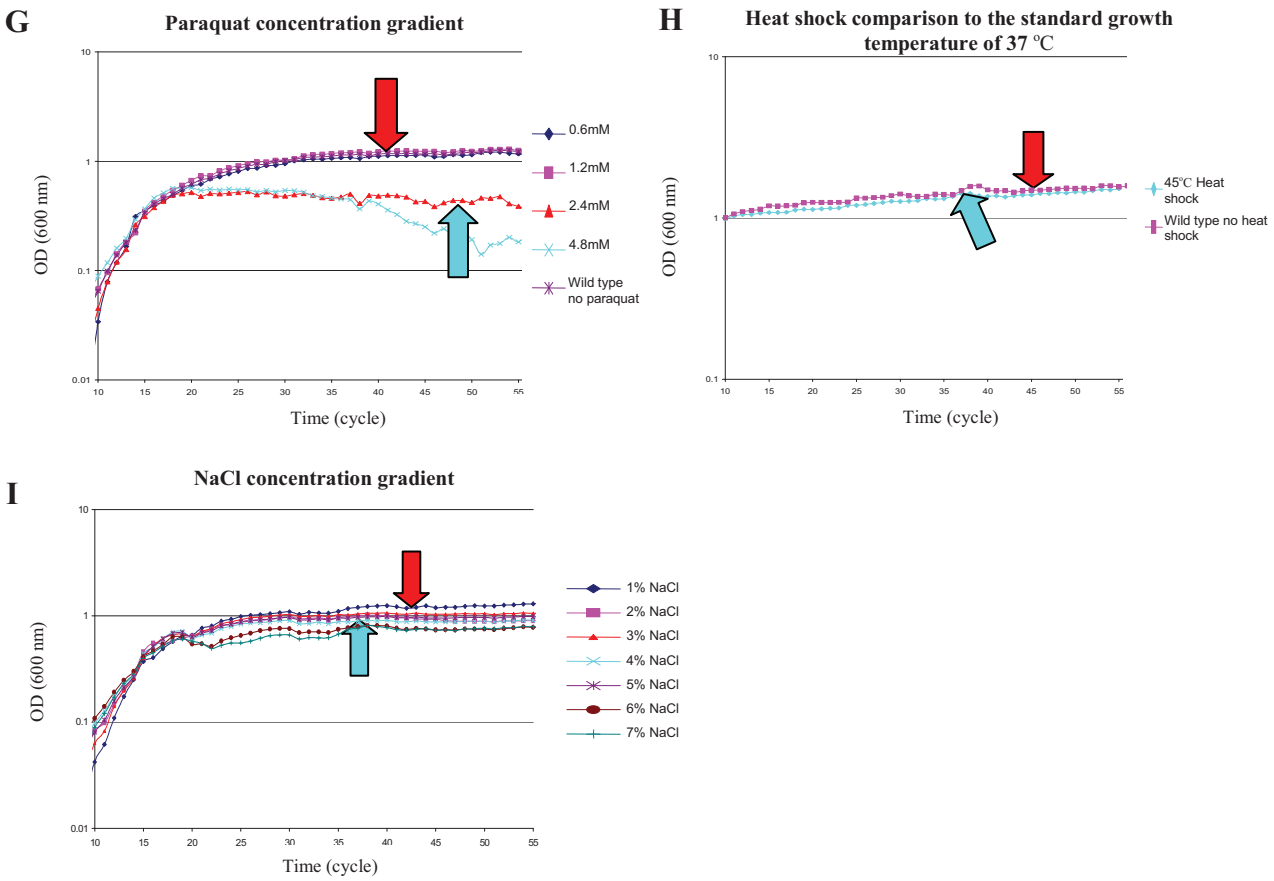


Figure 3.4 (continued): Graphs to show the determination of stressor concentrations to be used in this investigation. Red arrow indicates the growth profile of wild type *B. subtilis* with no stressor applied. The blue arrow indicates the concentrations of stressor selected for this investigation. G, Paraquat. H, Heat shock. I, NaCl.

Table 3.4: The stressors concentrations determined for this investigation.

Stressor	Concentration
Nalidixic acid	300 µg/ml
Rifampicin	0.06 µg/ml
Kanamycin	10 µg/ml
Streptomycin	80 µg/ml
Tetracycline	2.5 µg/ml
Vancomycin	2 µg/ml
Paraquat	2.4 mM
Heat	45^{o}C
NaCl	5 %

3.4.3 Experimental approach

The targeting of a single or a few essential components within a biological system as most modern therapeutic compounds aim to do, results in definable phenotypic behaviours, that may not be directly attributed to the therapeutic compound, but rather a side effect of its use. The phenotypic behaviour produced using Systems Biology approaches and the targeting of non-essential genes, produces a cumulative phenotypic effect. The greater the number of targets the more visible the phenotype produced (Figure 3.5). Because this investigation relies on the targeting of two targets simultaneously particularly sensitive methods are required to analyse the phenotypic properties produced.

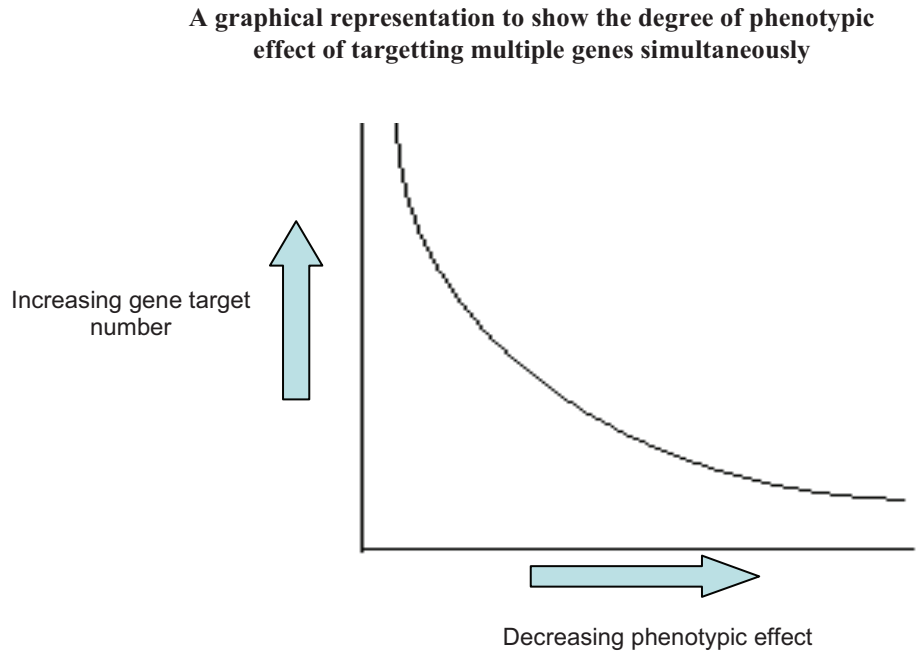


Figure 3.5: The phenotypic effects of cumulative gene targeting. A graphical representation showing the scale of phenotypic effects seen when multiple genes are simultaneous targeted by stressor.

Due to the number of mutants involved, and the time required to analyse the influence of each mutation both alone and in combination, a three-stage approach was developed to screen for altered growth phenotypes (Figure 3.6). This approach could be scaled up if necessary to test thousands of potential gene combinations. In stage one, a high-throughput computer-controlled approach was used. The influence of a particular stressor on the growth kinetics of all candidate genes of one vector type was determined in a 96 well plate in a microplate reader at 37°C with shaking. Once the influence of the genes was determined individually, the influence of the combined mutations was determined. Samples were analysed in triplicate and the data averaged. In stage two, samples that deviated significantly from the control were reanalysed, along with their cognate partners, using a low-throughput manual approach. In stage three, samples that continued to show differences to control samples were analysed in further detail.

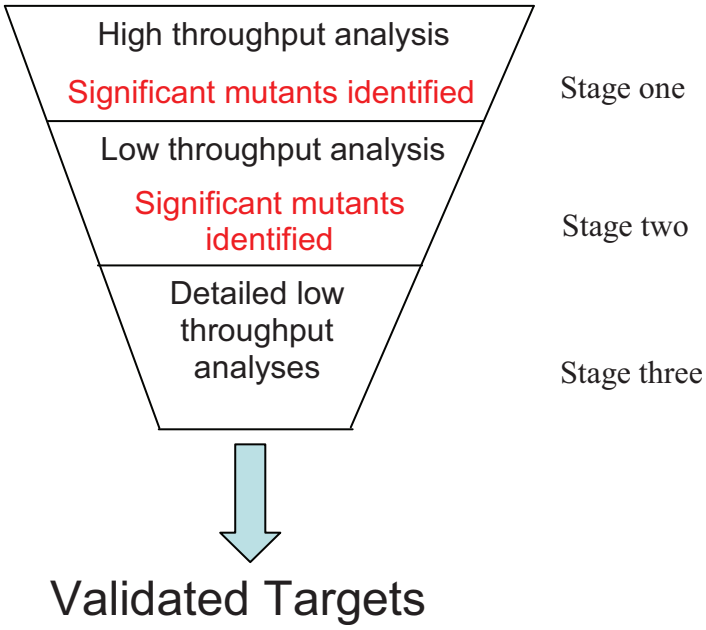


Figure 3.6: The multi-stage focusing experimental approach. The focusing approach adopted in this investigation to identify and validate potential therapeutic target candidates.

Chapter 4
Experimental Analysis of Gene/Protein Interaction

4. Therapeutic targets and arginine and pyrimidine (uracil) biosynthetic system analysis

This chapter discusses the accuracy and validity of using an automated high-throughput approach to test candidate genes as potential therapeutic targets. Details of how the high-throughput approach was tailored to these investigations are discussed, as well as the results obtained. Candidates identified by high-throughput analysis were subsequently analysed by further low-throughput approaches, mimicking the conditions of the high-throughput stage, together with more specific phenotypic testing.

The results obtained through this testing are discussed and evaluated in the context of the identification of successful therapeutic targets.

This chapter also describes the experimental and computational approaches taken to investigate the arginine and pyrimidine biosynthetic pathways, their genetic organisation and potential interaction with one another through a hypothesis driven approach to experimentation.

4.1 High-throughput therapeutic target testing

4.1.1 Microplate reader validation

4.1.1.1 Determination of data consistency with respect to well location

The consistency and accuracy of measurements taken across a 96-well microplate by a computer controlled microplate reader was validated prior to its use. To conduct this validation, multiple 96-well plates were completely filled, with each well containing 200 µl of sterile LB from the same batch (unpublished protocol). The optical density (OD) for each well across the plate was measured and repeated several times, with the values being averaged, compiled and then plotted on a graph (Figure 4.1) against position. This allowed the identification of any position upon the plate that was producing inconsistent results, and could be avoided during further testing.

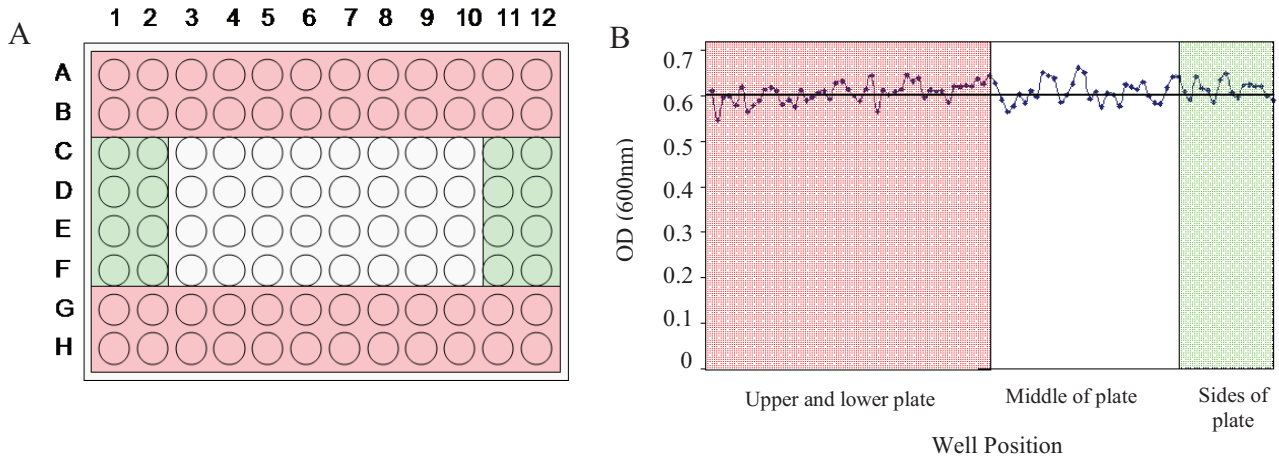


Figure 4.1: A graphical representation of the distinct areas and consistency of measurement of a 96 well plate using a microplate reader. A, Distinct areas of 96 well plate. B, The plotted results of OD consistency, testing the distinct regions of the 96 well plate with a microplate reader. A line of best fit has been applied, and indicated samples were achieving an OD of 0.6 +/- 0.05, with no anomalous results occurring from any specific region of the plate.

Repeated experimentation determined that there were no significant variations in observed measurements across the plate, with only small OD variations of +/- 0.05, which could be attributable to minor changes in aliquot volume or physical limitations associated with measurements taken by the microplate reader. Sample repetition, and averaging of the results obtained would reduce these variations.

4.1.1.2 Sample repetition

To provide an accurate representation of the behaviour of the mutants, and to avoid the minor variations identified tests on each mutant pair with a specific stressor were all carried out on the same plate, and then duplicated on a separate plate with samples loaded at different locations.

The 96-well plate was split into thirds (Figure 4.2). Each double mutant and its constituent single mutants were tested in parallel on the same row. Each mutant was loaded into four adjacent wells, and the data for each mutant/growth combination was averaged. Also on the plate was a control of wild type *B. subtilis* and a row of blanks.

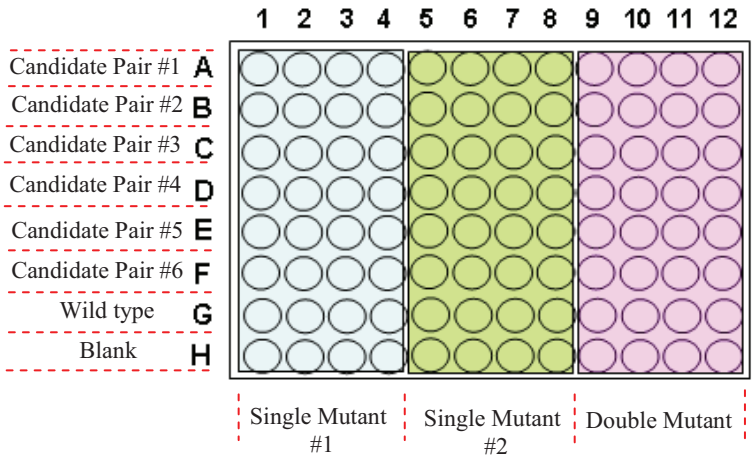


Figure 4.2: Mutant layout on 96 well plates. A graphical representation of the mutant layout on 96 well plates. Blue represents a single mutant of a pair. Green indicates the second single mutant of the pair. Pink indicates the combined double mutant. Each candidate combination was tested in four wells from which an average was taken, before being repeated on a separate plate, and in different well positions.

4.1.1.3 The influence of using a semi-permeable sealing membrane

All investigations with the microplate reader were conducted over a period of eight hours at a temperature of 37°C. Due to the small volume used in the 96 well plates of 200 µl, and the time taken to conduct each experiment, it was necessary to establish the extent of evaporation that was likely to occur during this time. A preliminary investigation was performed to determine the loss of culture volume during a typical experiment. All 96-wells were filled with 200 µl of LB and subjected to an eight hour microplate reader cycle at 37°C. On completion 20 random wells were selected, and the volume of LB remaining within them measured and averaged.

The average volume loss was 9.25%. The experiment was repeated after covering the 96 well plate with a Breatheasy® membrane, - a breathable membrane designed to prevent evaporation. The average volume loss per well was reduced to 2.75%, consequently all following experiments were performed using a Breatheasy® membrane.

4.1.1.4 Microplate reader accuracy

To determine the comparability between growth curves obtained from high and low-throughput analyses, the ODs obtained from the microplate reader (high-throughput analysis) and from a conventional spectrophotometer (low-throughput analysis) were compared. An overnight culture of *B. subtilis* was diluted with LB media in 20% increments. The OD of a single sample was measured using two independent Biochrom ultrospec II spectrophotometers, before being distributed as 200 µl aliquots onto a 96 well plate which was measured using the FLUOstar OPTIMA microplate reader. The values generated were plotted and compared (Figure 4.3).

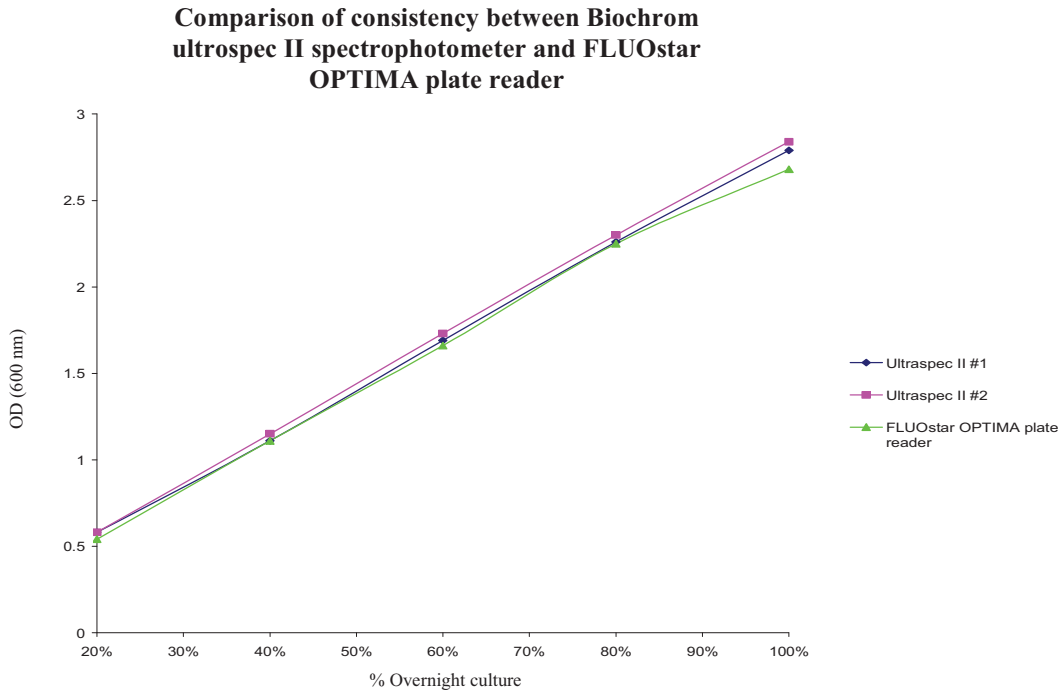


Figure 4.3: A graph to show the comparison of 2 Biochrom Ultraspec II spectrophotometers against a FLUOstar OPTIMA microplate reader. Various dilutions of overnight culture were measured and compared in the ultrospec II #1 spectrophotometer (Blue). Ultrospec # 2 spectrophotometer (Red) and FLUOstar OPTIMA plate reader (Green), for consistency of result.

The analyses indicated that the OD measurements obtained from the microplate reader and ultrospec spectrophotometer were comparable. Values at the lowest dilutions varied slightly between the equipment, but were not deemed to be significant, and not in the ranges, required to conduct this investigation.

4.1.1.5 Determination of the optimal time of stressor addition

The timing of the application of stressor to the mutant strains to maximise its effect was confirmed by experimentation. The point of mid-exponential phase- an OD of 0.3, was chosen as the point at which stressors would be applied to have maximum effect.

To produce comparative data, all experiments were started from an identical OD of 0.01. This OD was chosen because it would allow sufficient generations of growth of the mutants to allow them to adapt to the growth conditions prior to the addition of stressor. The point at which mid exponential point occurred within the microplate reader program was determined by plotting the growth of wild type *B. subtilis* grown from an inoculum with a starting OD of 0.01 in the microplate reader (Figure 4.4). This point was found to occur at approximately 1 hour and 33 minutes after the start of the program, at the end of microplate reader cycle 12.

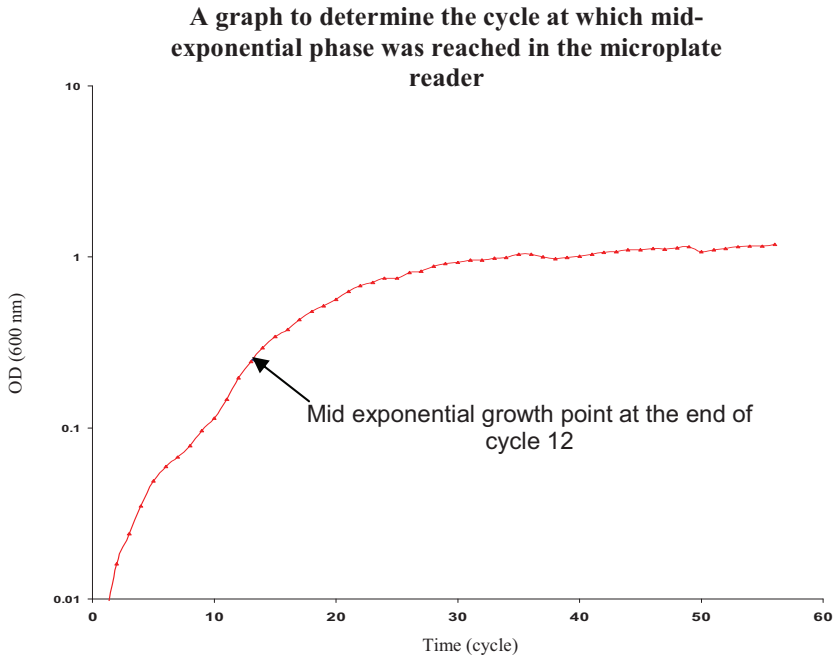


Figure 4.4: A graph to determine the cycle number at which mid-exponential point is reached in the microplate reader. A *B. subtilis* culture was grown for 8 hours, being measured continuously to determine the mid exponential point. This was reached at the end of cycle 12 (0.3 OD).

4.2 High-throughput analysis of potential therapeutic targets

The growth profiles of the single and double mutants of the potential drug targets suggested by e-Therapeutics were analysed using the high-throughput testing approach and the results discussed below.

4.2.1 Growth with and without inducer

The growth profiles of mutants were analysed both in the presence and absence of the inducers, xylose or IPTG, required for the activation of their inducible promoters (Figure 4.5). The inducers were added at mid exponential phase, equivalent to the time at which the stressors would be added. This test was designed to identify any downstream polar effects in the single mutants that could potentially influence the behaviour of the double mutants. This test also ensured that both sets of mutants created with either pSG1164 or pMUTIN4, were not adversely affected by the inducer of each other (Figure 4.6), a situation that would occur in the testing of double mutants.

The control experiments designed to establish the influence of inducers on the mutants (Figure 4.5) showed that neither had a significant influence on their growth profiles. This indicated that the mutants were not being affected by the removal of the function of any genes downstream and in the same operon as the target genes. The testing of growth of single mutants in opposite inducers (Figure 4.6), indicated that there was no influence on growth profiles also.

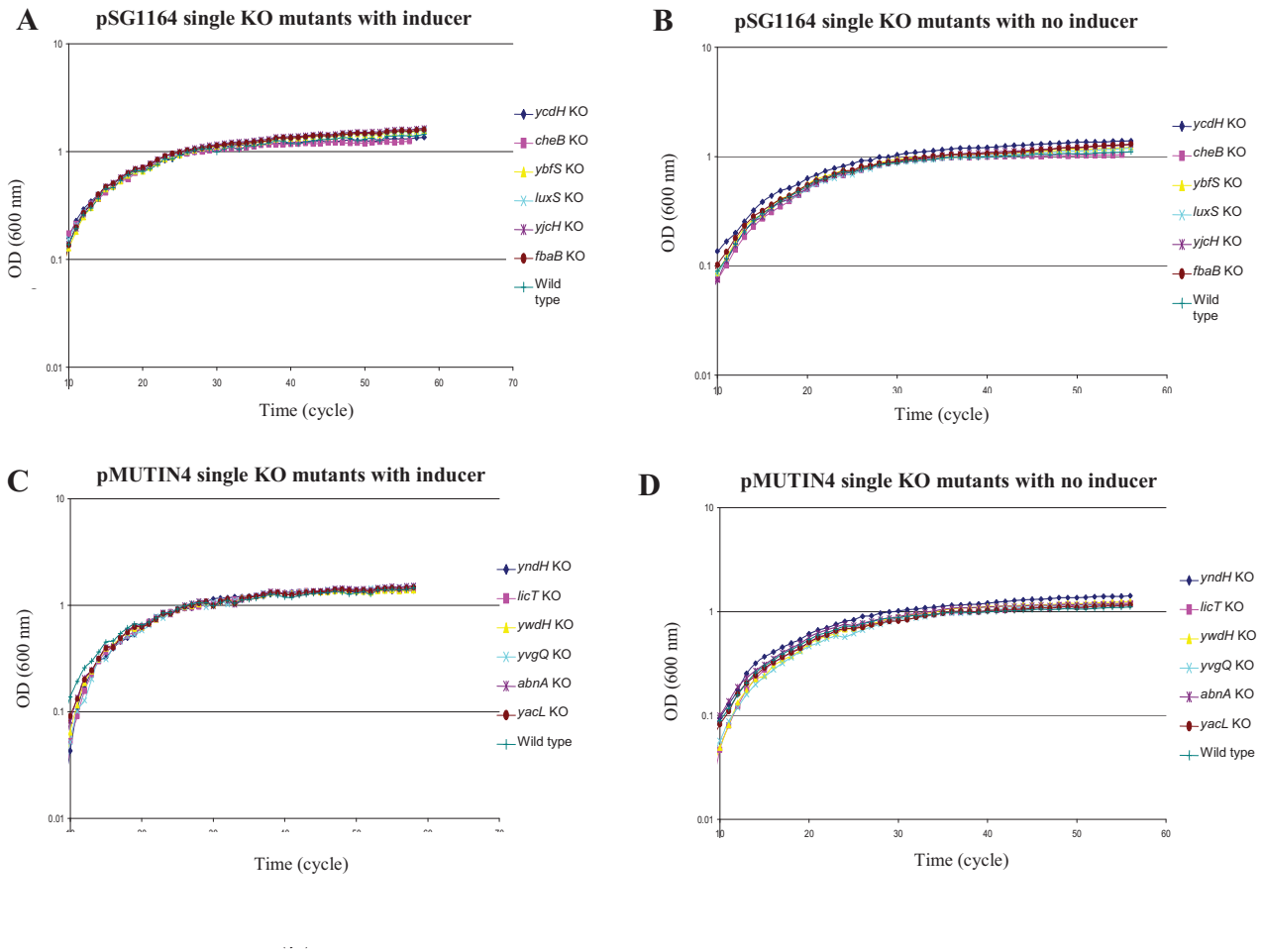


Figure 4.5: Graphs to show the growth profile of pSG1164/pMUTIN4 and pSG1164-pMUTIN4 mutants with and without inducer. A,pSG1164 mutants with xylose (1% w/v). B, pSG1164 mutants with no xylose. C, pMUTIN4 KO mutants with IPTG (1mM). D, pMUTIN4 KO mutants with no Inducer.

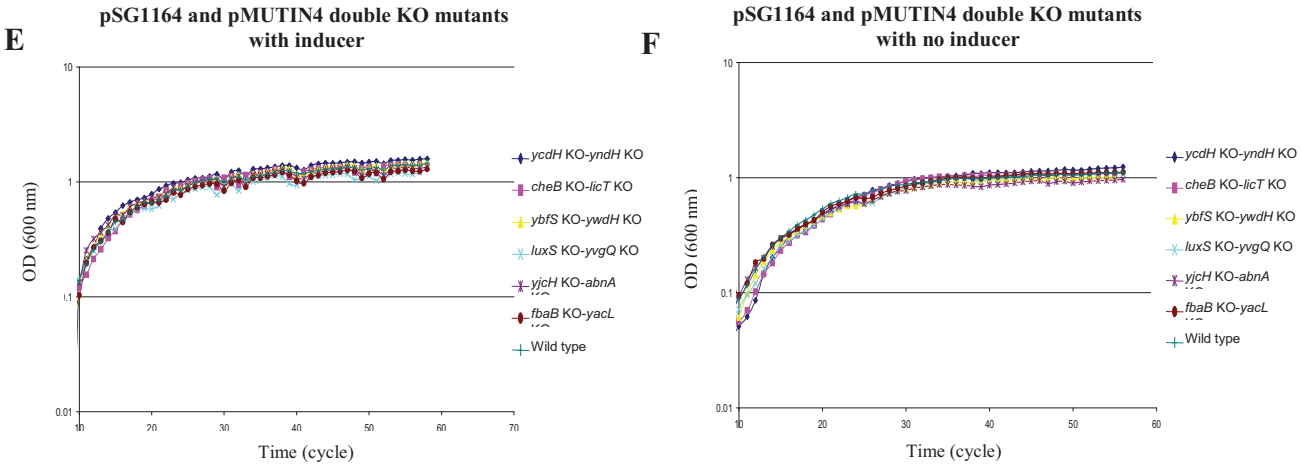


Figure 4.5 (continued): Graphs to show the growth profile of pSG1164/pMUTIN4 and pSG1164-pMUTIN4 mutants with and without inducer. E, pSG1164-pMUTIN4 KO mutants with xylose (1% w/v). F, pSG1164-pMUTIN4 KO mutants with no inducer.

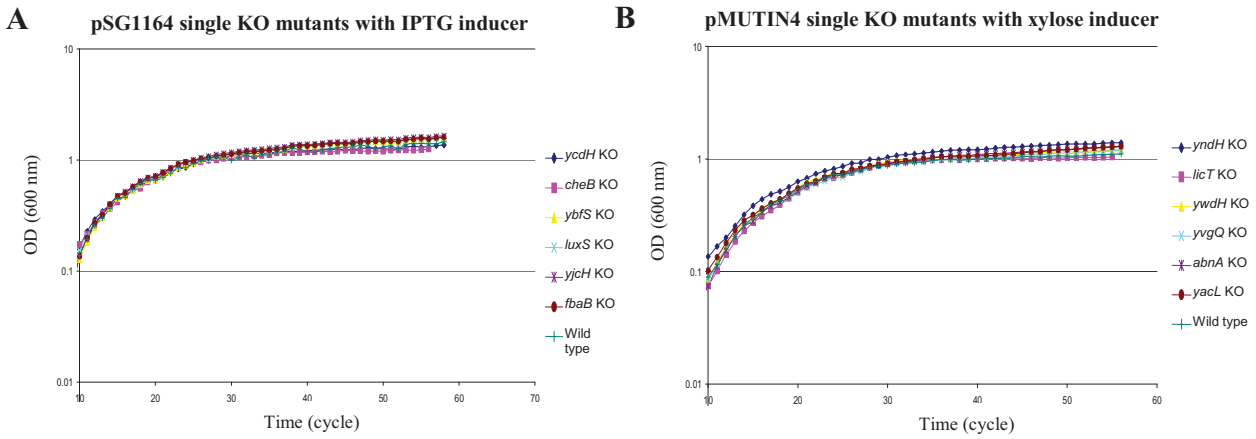


Figure 4.6: Graphs to show the growth profile of pSG1164 and pMUTIN4 KO mutants grown in opposite inducer. A, pSG1164 mutants treated with IPTG (1mM). B, pMUTIN4 mutants treated with xylose (1% w/v).

4.2.2 Growth kinetics following a challenge with nalidixic acid

The growth profiles of each of the mutants (single and double) were determined and analysed after the addition, during mid-exponential phase, of the stressor nalidixic acid at a sub-inhibitory concentration of 300 µg/ml (Figure 4.7).

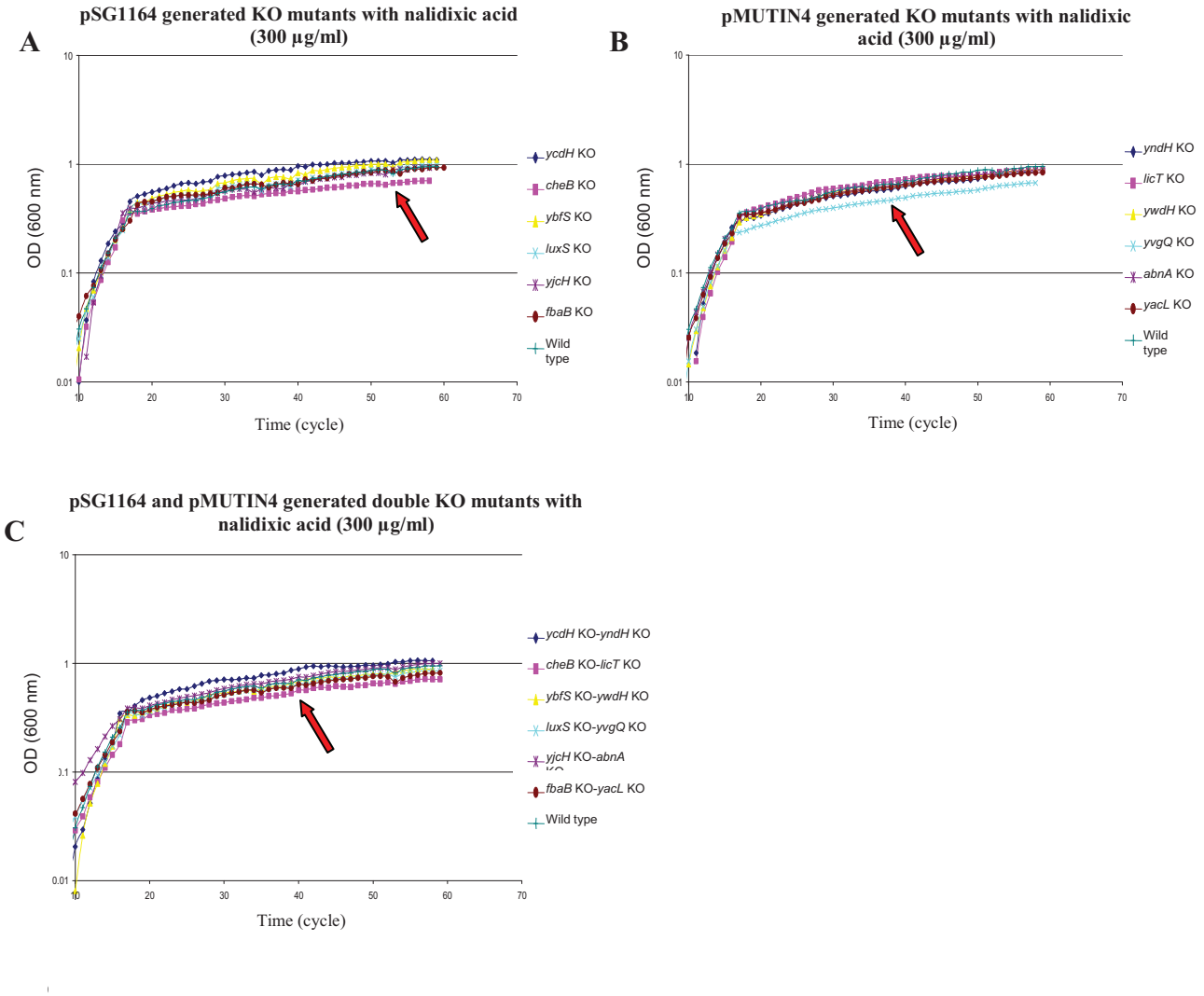


Figure 4.7: Graphs to show the growth profiles of pSG1164/pMUTIN4 KO and pSG1164-pMUTIN4 KO mutants exposed to the stressor nalidixic acid. Nalidixic acid was added to mutants at the mid-exponential growth phase to a final concentration of 300 µg/ml. A, Single KO mutants generated using the integration vector pSG1164. The red arrow indicates abnormal *cheB* KO growth profile under these conditions. B, Single KO mutants generated using the integration vector pMUTIN4. The red arrow indicates an abnormal growth profile for *yvgQ* KO under these conditions. C, Double KO mutants generated using both the pSG1164 and pMUTIN4 integration systems. The red arrow indicates a slight abnormal growth profile in *cheB* KO-*licT* KO mutant under these conditions.

The growth profiles of each of the *cheB* KO and *ycdH* KO mutants exhibited a slight reduction in the OD in the stationary phase when compared to wild type *B. subtilis* treated in the same way. This behaviour was not observed by their respective double mutants. The growth profiles of the other mutants tested were found to be indistinguishable from the wild type *B. subtilis*.

4.2.3 Growth kinetics following a challenge with rifampicin

The growth profiles of each of the mutants (single and double) were determined and analysed after the addition, during mid-exponential phase, of the stressor rifampicin at a sub-inhibitory concentration of 0.06 µg/ml (Figure 4.8).

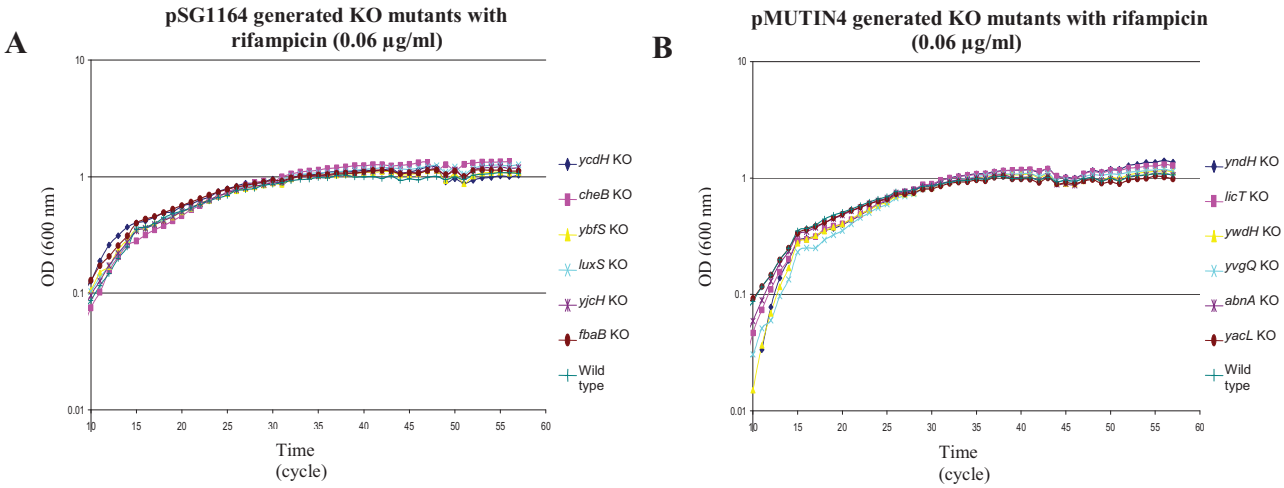


Figure 4.8: Graphs to show the growth profiles of pSG1164/pMUTIN4 KO and pSG1164-pMUTIN4 KO mutants exposed to the stressor rifampicin. Rifampicin was added at mid-exponential growth phase to a final concentration of 0.06 µg/ml. A, Single KO mutants generated using the integration vector pSG1164. B, single KO mutants generated using the integration vector pMUTIN4. C, Double KO mutants generated using both the pSG1164 and pMUTIN4 integration systems.

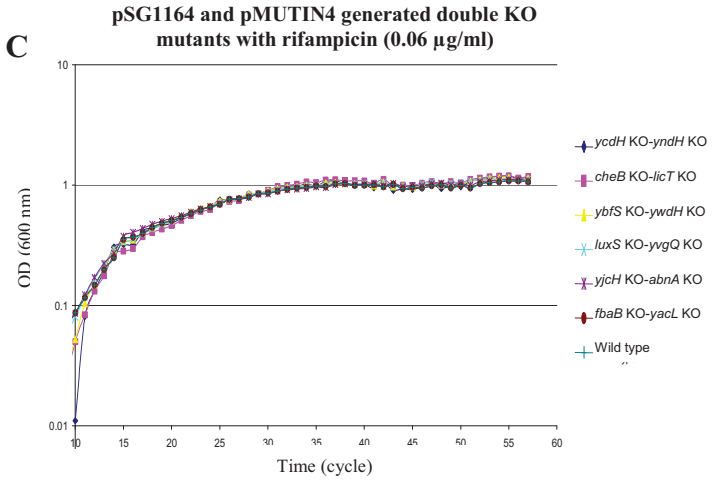


Figure 4.8 (continued): Graphs to show the growth profiles of pSG1164/pMUTIN4 KO and pSG1164-pMUTIN4 KO mutants exposed to the stressor rifampicin. Rifampicin was added at mid-exponential growth phase to a final concentration of 0.06 µg/ml. C, Double KO mutants generated using both the pSG1164 and pMUTIN4 integration systems.

The growth profiles of all mutants tested were found to be indistinguishable from wild type *B. subtilis*.

4.2.4 Growth kinetics following a challenge with kanamycin

The growth profiles of each of the mutants (single and double) were determined and analysed after the addition, during mid-exponential phase, of the stressor kanamycin at a sub-inhibitory concentration of 10 µg/ml (Figure 4.9).

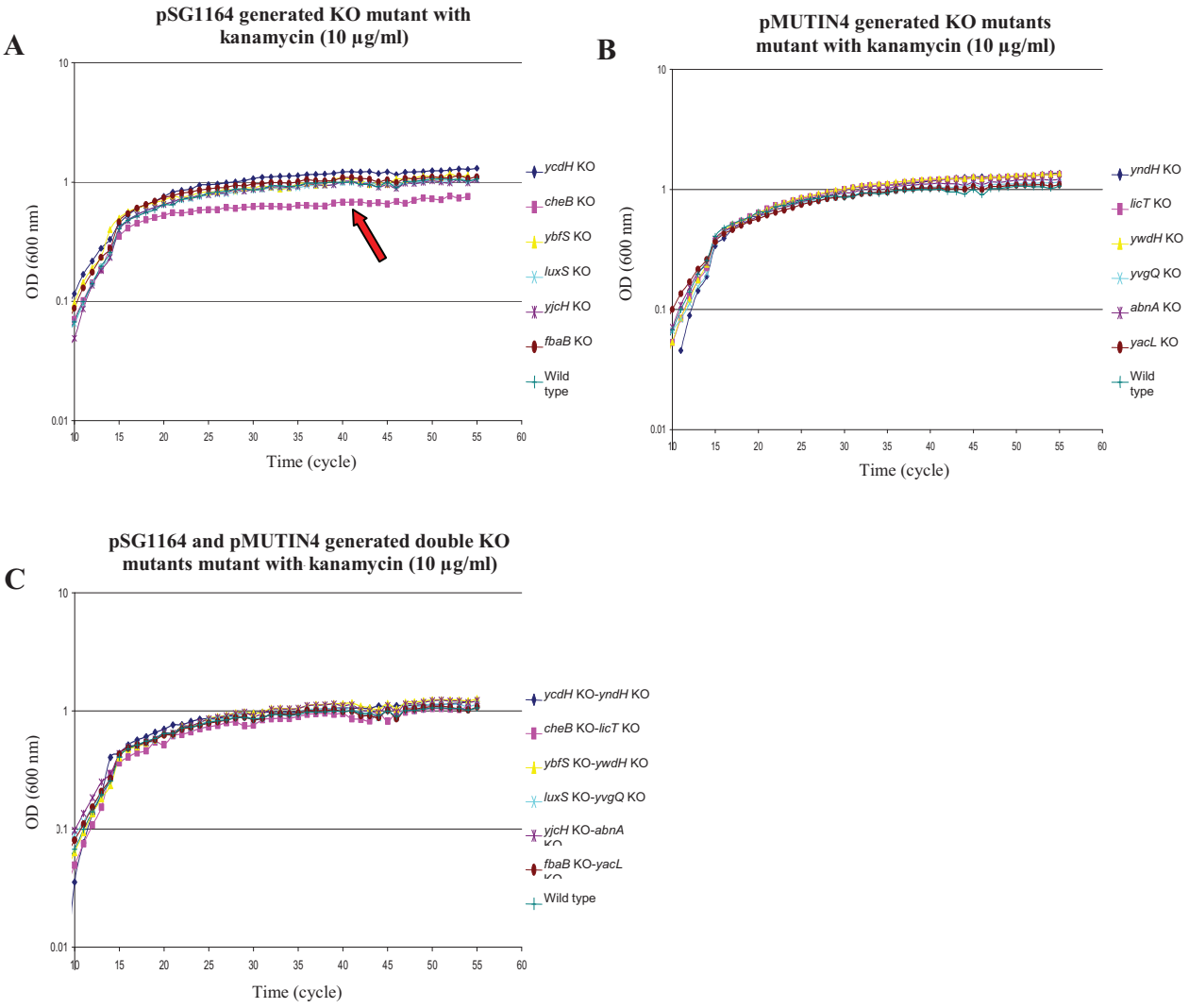


Figure 4.9: Graphs to show the growth profiles of pSG1164/pMUTIN4 KO and pSG1164-pMUTIN4 KO mutants exposed to the stressor kanamycin. Kanamycin was added at the mid-exponential growth phase to a final concentration of 10 µg/ml. A, Single KO mutants generated using the integration vector pSG1164. The red arrow indicates an abnormal growth profile of *cheB* KO under these conditions. B, Single KO mutants generated using the integration vector pMUTIN4. C, Double KO mutants generated using both the integration vectors pSG1164 and pMUTIN4.

The growth profile of the *cheB* KO mutant exhibited a variation between single mutant phenotype and wild type *B. subtilis* when treated in the same way. The culture density of the *cheB* KO mutant was lower in stationary phase with an OD of 0.7 compared to wild type *B. subtilis* with an OD of 1.0, a difference that remained throughout the experiment. Interestingly, this behaviour was not seen in the case of the double mutant or any of the other mutants, which were all found to be indistinguishable from wild type *B. subtilis*.

4.2.5 Growth kinetics following a challenge with streptomycin

The growth profiles of each of the mutants (single and double) were determined and analysed after the addition, during mid-exponential phase, of the stressor streptomycin at a sub-inhibitory concentration of 80 µg/ml (Figure 4.10).

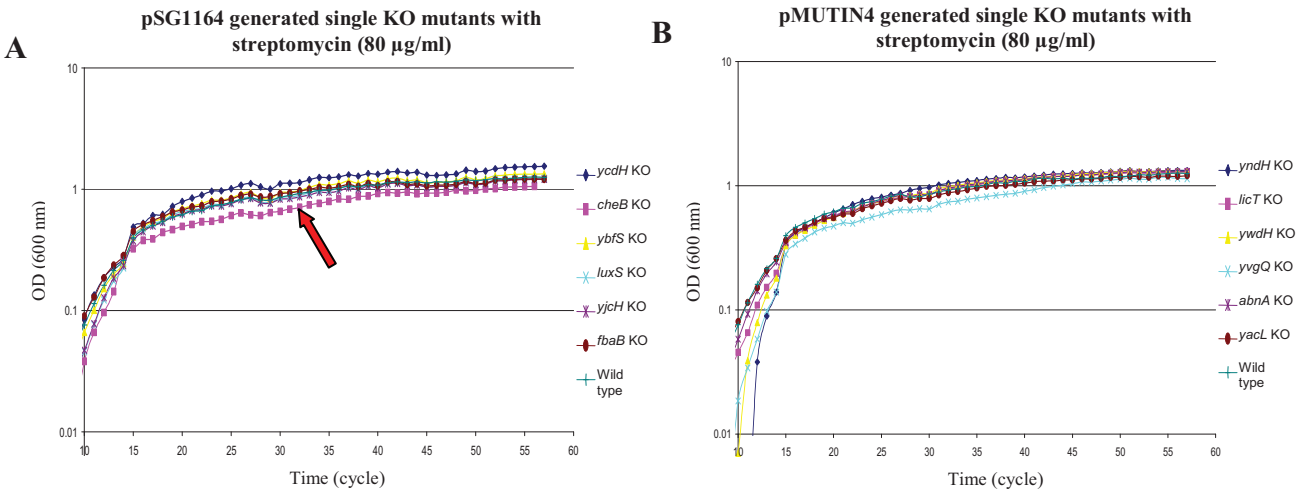


Figure 4.10: Graphs to show the growth profiles of pSG1164/pMUTIN4 KO mutants exposed to the stressor streptomycin. Streptomycin was added at mid-exponential growth phase to a final concentration of 80 µg/ml. A, Single KO mutants generated using the integration vector pSG1164. The red arrow indicates abnormal growth profile for *cheB* KO under these conditions. B, Single KO generated mutants generated using the integration vector pMUTIN4.

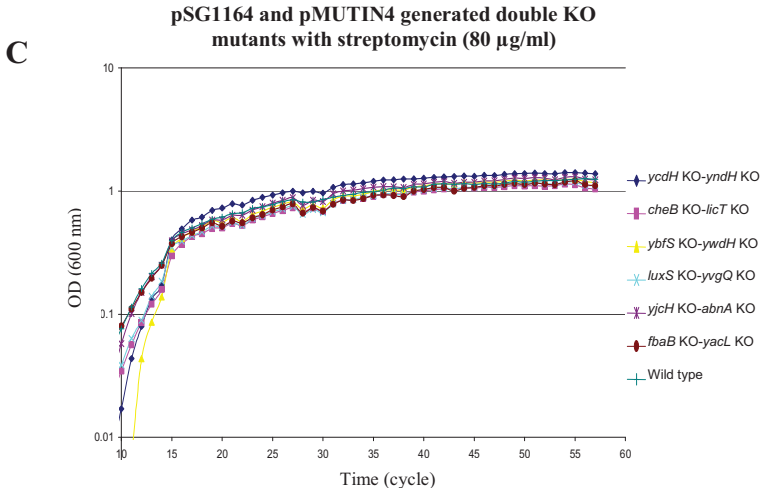


Figure 4.10 (continued): Graph to show the growth profile of pSG1164-pMUTIN4 KO mutants exposed to the stressor streptomycin. Streptomycin added at mid-exponential growth phase to a final concentration of 80 µg/ml. C, Double KO mutants generated using both pSG1164 and pMUTIN4 integration vectors.

The growth profile of the *cheB* KO mutant exhibited a slight reduction in OD in the stationary phase when compared to wild type *B. subtilis* treated in the same way. This behaviour was not observed in the respective double mutant. The growth profiles of the other mutants tested were found to be indistinguishable from wild type *B. subtilis*.

4.2.6 Growth kinetics following a challenge with tetracycline

The growth profiles of each of the mutants (single and double) were determined and analysed after the addition, during mid-exponential phase, of the stressor tetracycline at a sub-inhibitory concentration of 2.5 µg/ml (Figure 4.11).

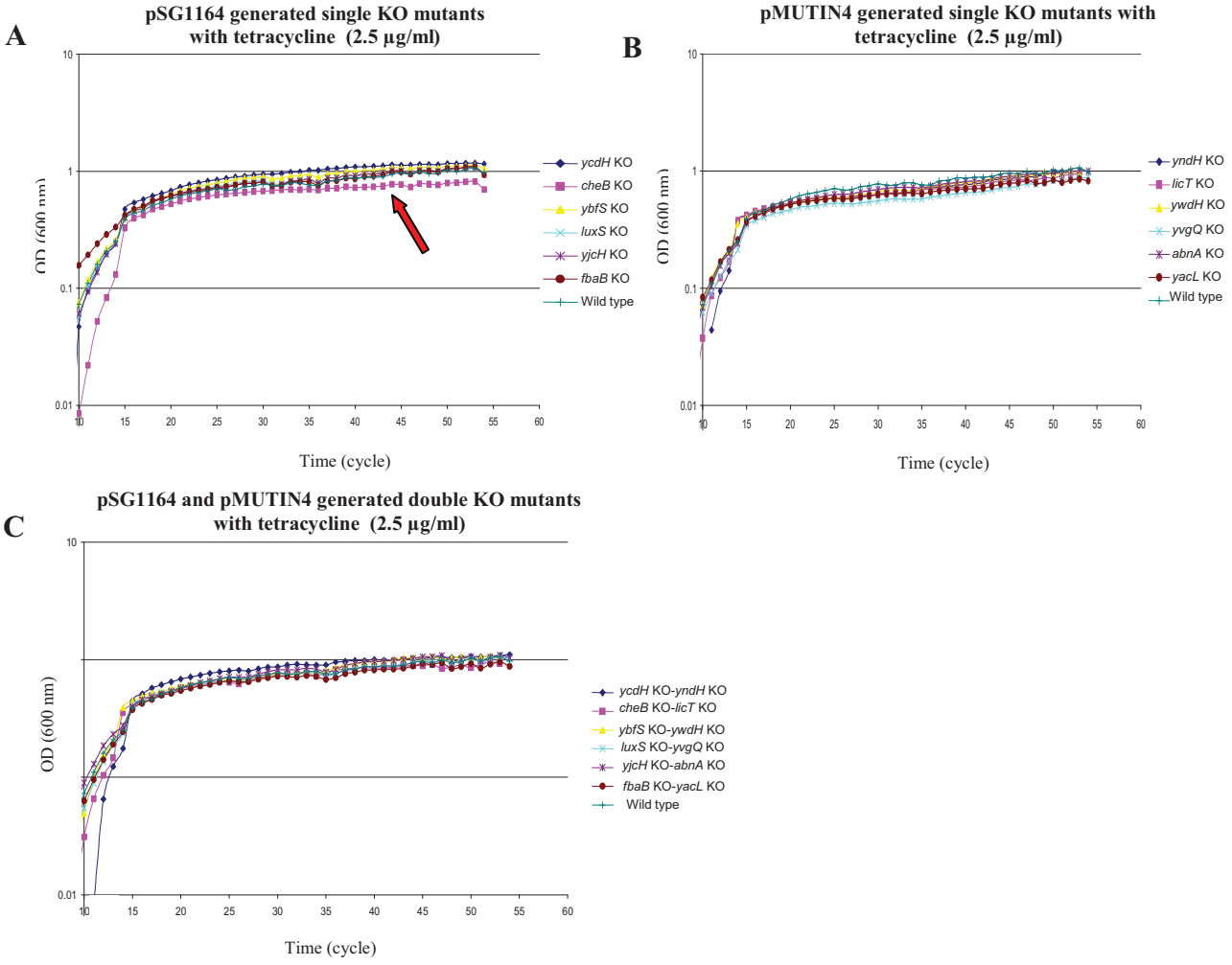


Figure 4.11: Graphs to show the growth profiles of pSG1164/pMUTIN4 KO and pSG1164-pMUTIN4 KO mutants exposed to the stressor tetracycline. Tetracycline added at mid-exponential growth phase to a final concentration of 2.5 µg/ml. A, Single KO mutants generated using the integration vector pSG1164. The red arrow indicates an abnormal growth profile for *cheB* KO under these conditions. B, Single KO mutants generated using the integration vector pMUTIN4. C, Double KO mutants generated using the integration vector pSG1164 and pMUTIN4.

The growth profiles of the *cheB* KO exhibited a slight reduction in the OD in the stationary phase when compared to wild type *B. subtilis* treated in the same way. This behaviour was not observed in the respective double mutant. The growth profiles of the other mutants tested were found to be indistinguishable from wild type *B. subtilis*.

4.2.7 Growth kinetics following a challenge with vancomycin

The growth profiles of each of the mutants (single and double) were determined and analysed after the addition, during mid-exponential phase, of the stressor vancomycin at a sub-inhibitory concentration of 2 µg/ml (Figure 4.12).

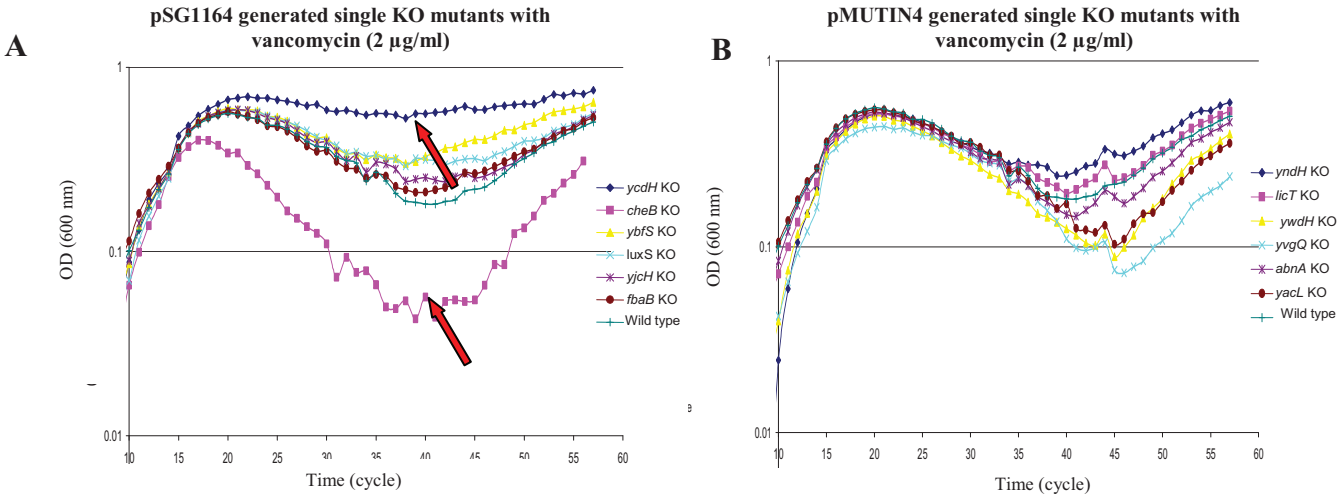


Figure 4.12: Graphs to show the growth profiles of pSG1164/pMUTIN4 KO mutants exposed to the stressor vancomycin. Vancomycin added at mid- exponential growth phase to a final concentration of 2 µg/ml. A, Single KO mutants generated using the integration vector pSG1164. The red arrow indicated an abnormal growth profile for *cheB* KO and *ycdH* KO. B, Single KO mutants generated using the integration vector pMUTIN4.

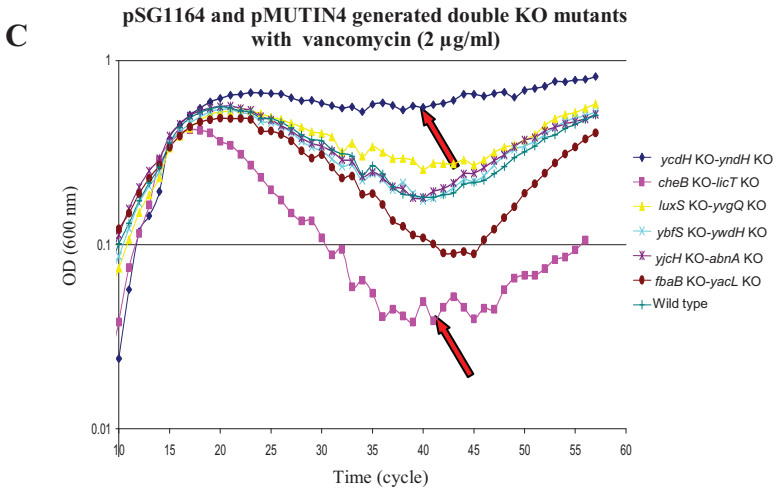


Figure 4.12(continued): Graph to show the growth profile of pSG1164-pMUTIN4 KO mutants exposed to the stressor vancomycin. Vancomycin added at mid-exponential growth phase to a final concentration of 2 µg/ml. C, Double mutants generated using both pSG1164 and pMUTIN4. The red arrow indicates an abnormal growth profile for *cheB* KO-*licT* KO and *ycdH* KO-*yndH* KO.

The growth profiles of the mutants indicated that the single *cheB* KO and *ycdH* KO mutants showed significant differences to wild type *B. subtilis* when treated with vancomycin. The culture density of wild type and remaining mutants continued to increase after the addition of vancomycin to an OD of 0.7, before declining rapidly to an OD of 0.4, and eventually re-establishing growth at a slower rate. The OD of the *cheB* KO single and double mutant continued to increase after the addition of vancomycin to an OD of 0.6 before declining rapidly and more severely than the other mutants to an OD of 0.06 before re-establishing growth at a considerably reduced growth rate. In contrast to the other mutants *ycdH* KO single and double mutants, does not show a change in OD following the addition of vancomycin, instead its growth kinetics were similar to the wild type without the addition of this antibiotic.

These behaviours of the double mutants reflected those seen in the single mutants to broadly the same extent, indicating that the behaviours seen were due to the dominant single mutants.

4.2.8 Growth kinetics following a challenge with paraquat

The growth profiles of each of the mutants (single and double) were determined and analysed after the addition, during mid-exponential phase, of the stressor paraquat at a sub-inhibitory concentration of 2.4 mM (Figure 4.13).

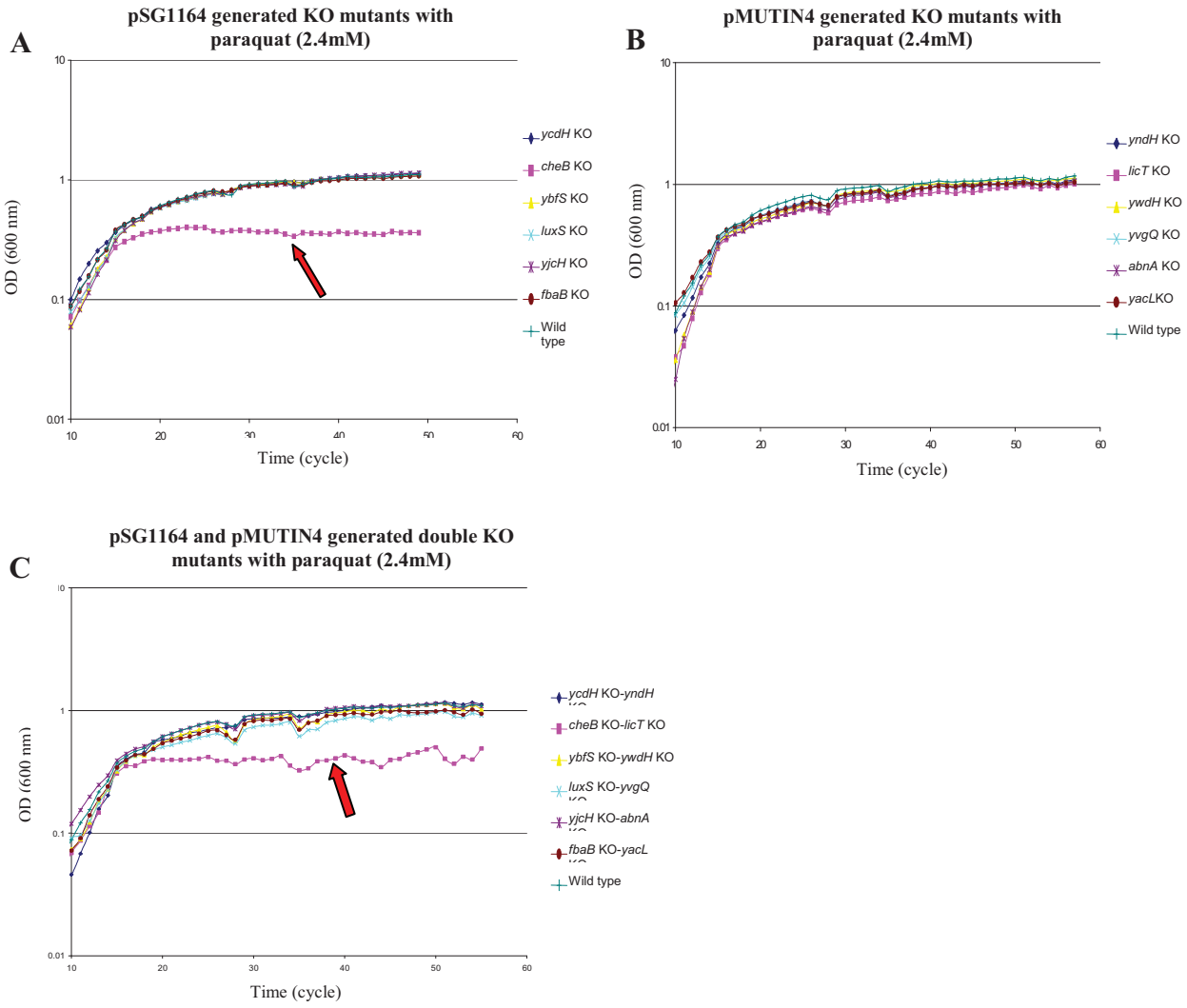


Figure 4.13: Graphs to show the growth profiles of pSG1164/pMUTIN4 KO and pSG1164-pMUTIN4 KO mutants exposed to the stressor paraquat. Paraquat added at mid-exponential growth phase to a final concentration of 2.4mM. A, Single KO mutants generated using the integration vector pSG1164. The red arrow indicates an abnormal growth profile for *cheB* KO under these conditions. B, Single KO mutants generated using the integration vector pMUTIN4. C, Double mutants generated using the integration vectors pSG1164 and pMUTIN4. The red arrow indicates an abnormal growth profile for *cheB* KO-*licT* KO.

The growth profile of the *cheB* KO mutant exhibited a significant reduction in culture density in the stationary phase from an OD of 1.0 for wild type *B. subtilis* treated in the same way, to an OD of 0.8. This behaviour was also observed in the respective double mutant. The growth profiles of the other mutants tested were found to be indistinguishable from wild type *B. subtilis*.

4.2.9 Growth kinetics following a challenge with heat shock

The growth profiles of each of the mutants (single and double) were determined and analysed after subjection to the stressor heat shock-increasing the temperature from 37°C to 45°C at mid exponential phase (Figure 4.14).

The growth profiles of each of the single *cheB* KO and *ycdH* KO mutants exhibited a slight reduction in culture density when compared to wild type *B. subtilis* treated in the same way. This behaviour was not observed by their respective double mutants. The growth profiles of the other mutants tested were found to be indistinguishable from wild type *B. subtilis*.

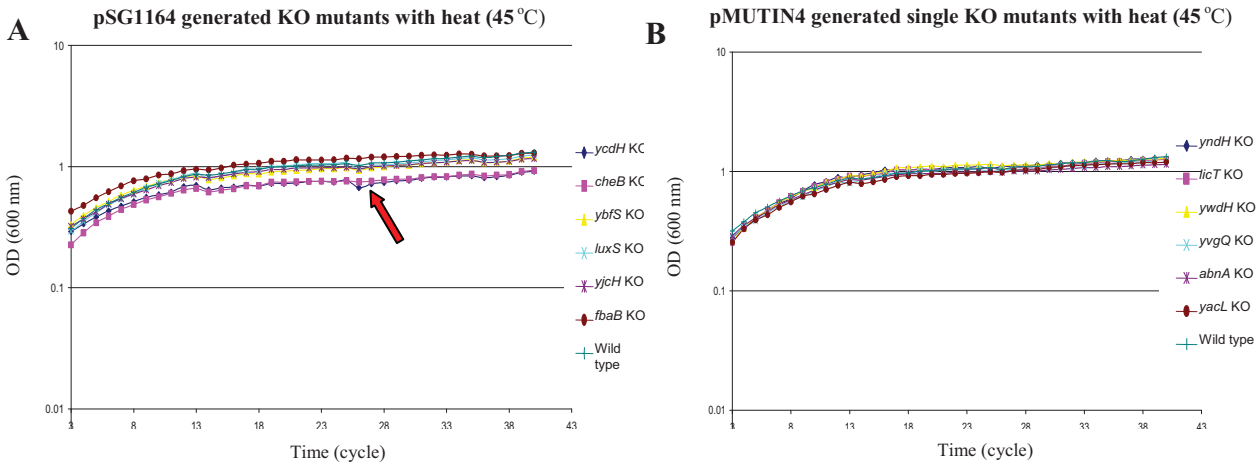


Figure 4.14: Graphs to show the growth profiles of pSG1164/pMUTIN4 KO and pSG1164-pMUTIN4 KO mutants exposed to the stressor heat. Heat to a final temperature of 45°C was added at mid-exponential growth phase from a normal growth temperature of 37°C. A, Single KO mutants generated using the integration vector pSG1164. The red arrow indicates an abnormal growth profile for both *cheB* KO and *ycdH* KO. B, Single KO mutants generated using the integration vector pMUTIN4.

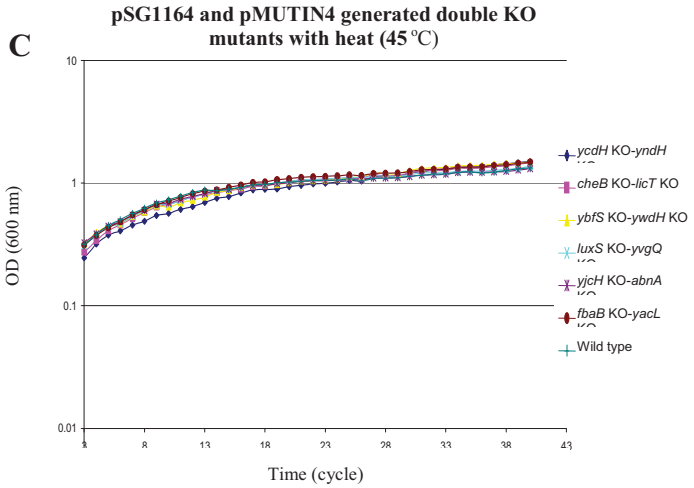


Figure 4.14 (continued): Graphs to show the growth profiles of pSG1164/pMUTIN4 KO and pSG1164-pMUTIN4 KO mutants exposed to the stressor heat. C, Double mutants generated using the integration vectors pSG1164 and pMUTIN4.

4.2.10 Growth kinetics following a challenge with high salt concentrations

The growth profiles of each of the mutants (single and double) were determined and analysed after the addition, during mid-exponential phase, of the stressor NaCl to a sub-inhibitory concentration of 5% (w/v) (Figure 4.15).

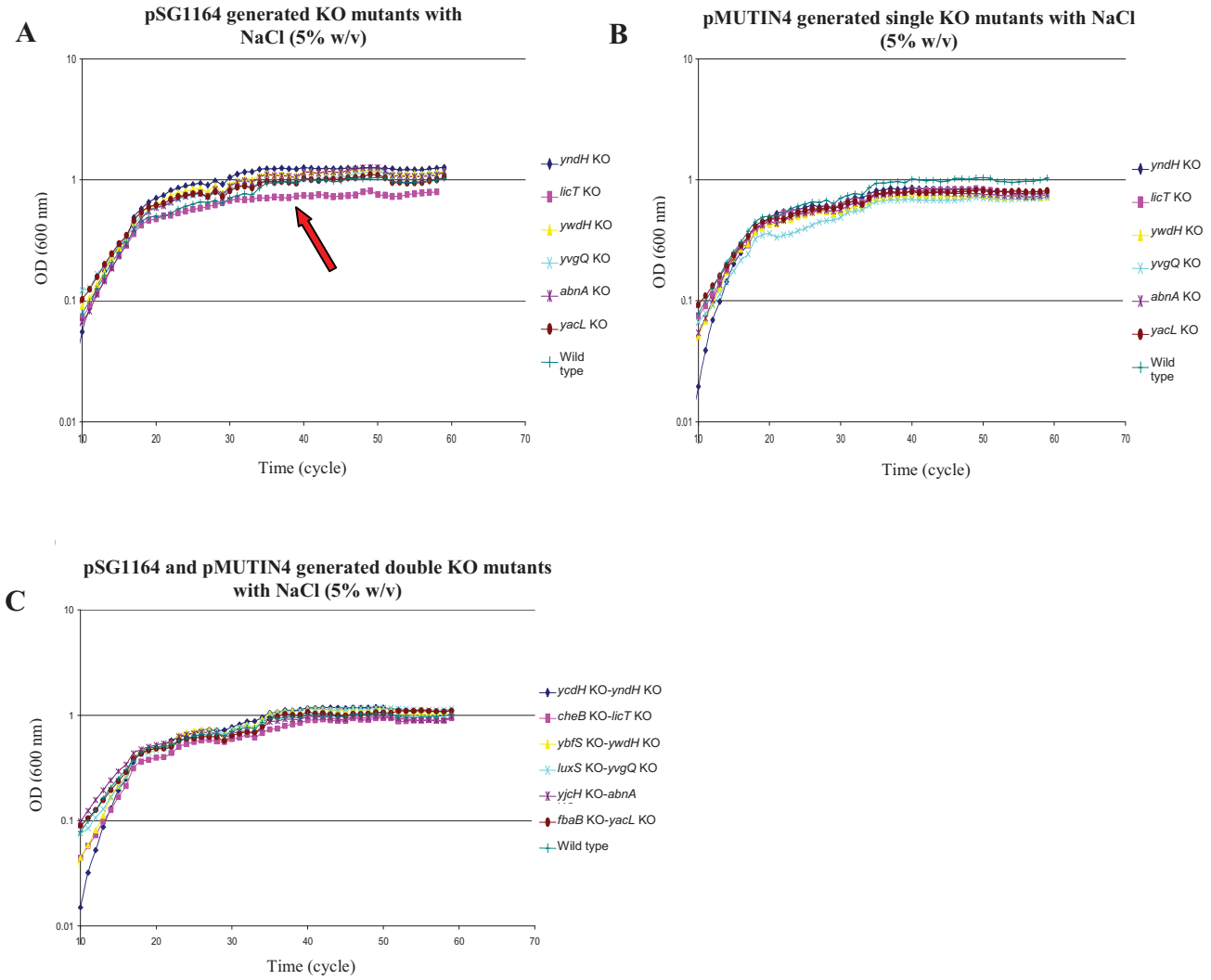


Figure 4.15: Graphs to show the growth profiles of pSG1164/pMUTIN4 KO mutants exposed to the stressor NaCl, NaCl added at mid-exponential phase to a final concentration of 5% (w/v). A, Single KO mutants generated using the integration vector pSG1164. The red arrow indicates an abnormal growth profile for *cheB* KO. Graph B, KO mutants generated using the integration vector pMUTIN4. C, Double KO mutants generated using the integration vectors pSG1164 and pMUTIN4.

The growth profiles of the single *cheB* KO exhibited a significant reduction in culture density in the stationary phase (OD 0.8) when compared to wild type *B. subtilis* treated in the same way (OD 1.0). This behaviour was not observed in the respective double mutant. The growth profiles of the other mutants tested were found to be indistinguishable from wild type *B. subtilis*.

4.2.11 High-throughput analysis of therapeutic target summary

The use of the high-throughput approach produced clear, consistent and unambiguous data. Only a few of the mutants exhibited a difference in their growth profile when treated with the various stressors compared to wild type *B. subtilis* (Table 4.1). The mutant with the most notable and consistent differences was the *cheB* KO with differences also being observed in its double mutant counterpart. The response of the *cheB* mutant was particularly noticeable in response to the stressor vancomycin. The single mutant *ycdH* KO and its double mutant also displayed variation in growth profile when exposed to the stressor vancomycin and heat. Double mutants that displayed variation could be attributed to variations displayed in its single constituent mutants

From these results in contradiction to the predictions of e-Therapeutics, there were no indications that the combination of mutations resulted in growth profiles, or any other behaviour indicative of the discovery of a potential therapeutic target. The behaviour of those single KO mutants described however would require further analysis.

Table 4.1: Summary grid of all mutants both single KO and double KO and their response to stressor. ✓ Denotes a difference seen to wild type under the same stressor. X indicates no difference to wild type under the same stressor.

Stressor	Mutant(s) (KO) A/B (Single mutants) A-B (Double mutant)					
	ybfS/ywdH *ybfS-ywdH*	*luxS/yvgQ* *luxS-yvgQ*	*ycdH/yndH* *ycdH-yndH*	*fbaB/yacL* *fbaB-yacL*	*yjcH/abnA* *yjcH-abnA*	*cheB/licT* *cheB-licT*
Inducer/no inducer	X	X	X	X	X	X
Nalidixic acid	X	✓	X	X	X	✓
Rifampicin	X	X	X	X	X	X
Kanamycin	X	X	X	X	X	✓
Streptomycin	X	X	X	X	X	✓
Tetracycline	X	X	X	X	X	✓
Vancomycin	X	X	✓	X	X	✓
Paraquat	X	X	X	X	X	✓
Heat	X	X	✓	X	X	✓
NaCl	X	X	X	X	X	✓

4.3 Low-throughput therapeutic target results

The mutants identified as having a different growth profile to that of the wild type *B. subtilis* (Table 4.1) were retested in a low-throughput approach, using the same stressors and concentrations and inducers. The mutants were grown in 25 ml aliquots within 250 ml conical flasks in a shaking water bath at 37ºC. The cultures remained in the water bath during sampling to avoid temperature fluctuations which could have affected the investigation. The mutants were grown from a starting OD of 0.01 and the stressor applied at an OD of 0.3. Growth and sampling continued until the overall response of the cultures was determined. This was repeated three times for each mutant, and the data averaged.

4.3.1 Growth kinetics following a challenge with nalidixic acid

Strains with the *luxS* and *yvgQ* mutations, singly and in combination (Figure 4.16A), and the *cheB* and *licT* mutations, single and in combination (Figure 4.16B), were stressed with nalidixic acid to a final concentration of 300 µg/ml. This low-throughput testing and analysis revealed that there were no differences between the growth profiles of the *yvgQ* and *luxS* single KO mutants, or the *luxS-yvgQ* double KO mutant in response to nalidixic acid, when compared to wild type *B. subtilis* treated in the same way. The same observations were made with the single c*heB* KO and *licT* KO mutants as well as the *cheB-licT* KO double mutant.

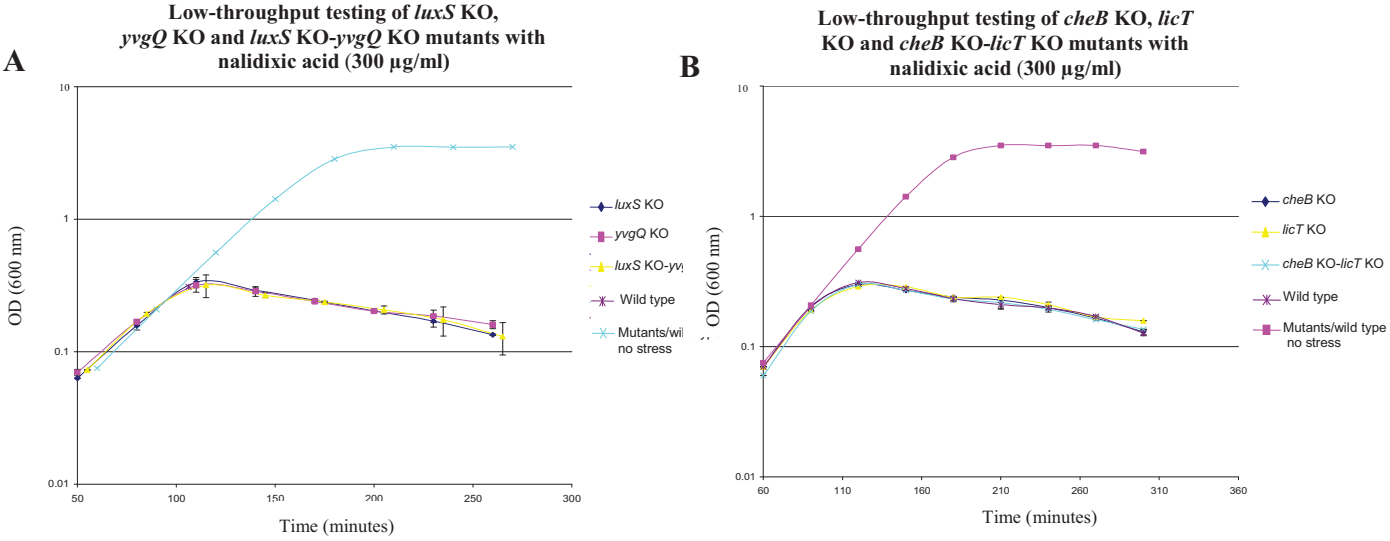


Figure 4.16: Low-throughput testing of single and double mutants with the stressor nalidixic acid. A, The growth profile of single KO mutants *luxS*, *yvgQ* and double KO mutant *luxS* KO-*yvgQ* KO in comparison to wild type *B. subtilis* when subjected to the stressor nalidixic acid (300 µg/ml). Also displayed are unstressed mutants/ wild type. B, Growth profile of single KO mutants *cheB, licT* and double mutant *cheB* KO-*licT* KO in comparison to wild type *B. subtilis* when subjected to the stressor nalidixic acid (300 µg/ml). Also displayed is a sample growth curve for the unstressed mutants and wild type.

4.3.2 Growth kinetics following a challenge with kanamycin

Strains with the *cheB* and *licT* mutations, singly and in combination (Figure 4.17), were stressed with kanamycin to a final concentration of 10 µg/ml. This low-throughput testing and analysis revealed that there were no differences between the growth profiles of either the single *cheB* KO and *licT* KO mutants or the double *cheB* KO-*licT* KO mutants when compared to wild type *B. subtilis* treated in the same way.

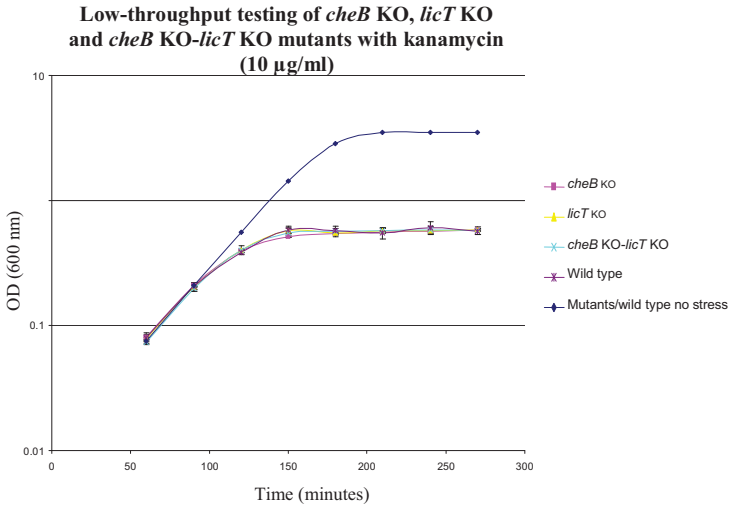


Figure 4.17: Low-throughput testing of single and double mutants with the stressor kanamycin. The growth profile of the single KO mutants *cheB* and *licT* and the double KO mutant *cheB* KO –*licT* KO in comparison to wild type *B. subtilis* when subjected to the stressor kanamycin (10 µg/ml). Also displayed is a sample growth curve for the unstressed mutants and wild type.

4.3.3 Growth kinetics following a challenge with streptomycin

Strains with the *cheB* KO and *licT KO* mutations, singly and in combination (Figure 4.18), were stressed with streptomycin to a final concentration of 80 µg/ml. This low-throughput testing and analysis revealed that there were no differences between the growth profiles of either the single *cheB* KO and *licT* KO mutants or the double *cheB* KO-*licT* KO mutants when compared to wild type *B. subtilis*.

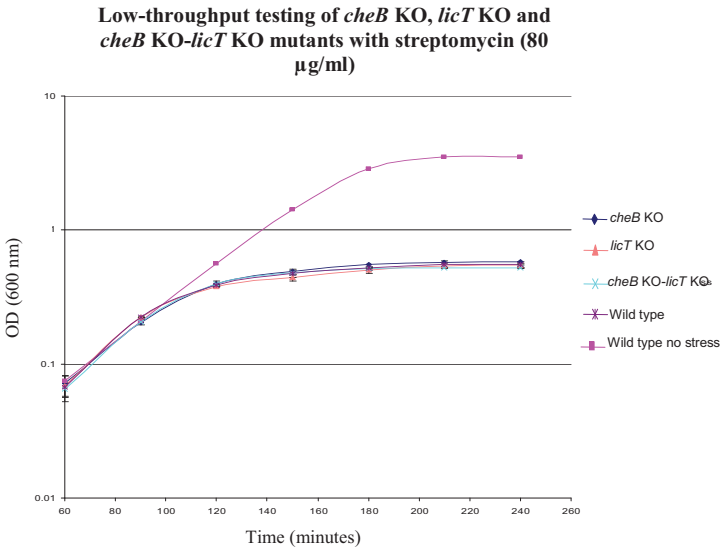


Figure 4.18: Low-throughput testing of single and double mutants with the stressor streptomycin. The growth profile of the single KO mutants *cheB* and *licT* and the double KO mutant *cheB* KO-*licT* KO in comparison to wild type *B. subtilis* when subjected to the stressor streptomycin (80 µg/ml). Also displayed is a sample growth curve for the unstressed mutants and wild type.

4.3.4 Growth kinetics following a challenge with tetracycline

Strains with the *cheB* KO and *licT KO* mutations, singly and in combination (Figure 4.19), were stressed with tetracycline to a final concentration of 2.5 µg/ml. This low-throughput testing and analysis revealed that there were no differences between the growth profiles of either the single *cheB* KO and *licT* KO mutants or the double *cheB* KO-*licT* KO mutants when compared to wild type *B. subtilis* treated in the same way.

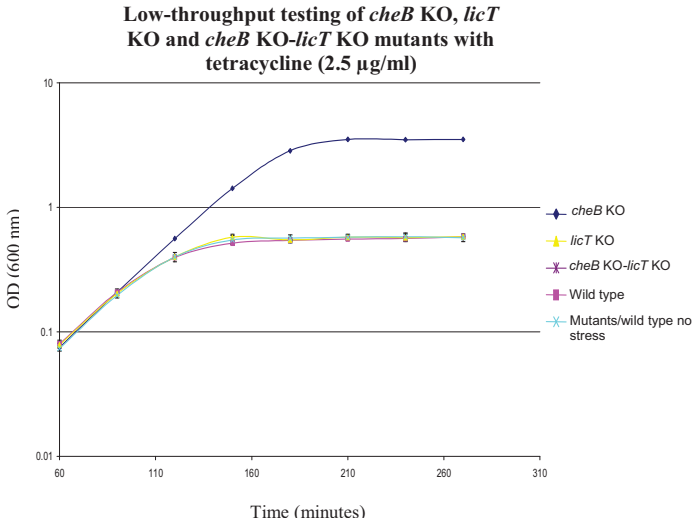


Figure 4.19: Low-throughput testing of single and double mutants with the stressor tetracycline. The growth profile of the single KO mutants *cheB* and *licT* and the double KO mutant *cheB* KO-*licT* KO in comparison to wild type *B. subtilis* when subjected to the stressor tetracycline (2.5 µg/ml). Also displayed is a sample growth curve for the unstressed mutants and wild type.

4.3.5 Growth kinetics following a challenge with vancomycin

The Single KO mutants, *ycdH* and *yndH* together with the double KO mutant *ycdH-yndH,* were stressed with 2 µg/ml final concentration of vancomycin (Figure 4.20A). This experiment was also repeated for *cheB* KO, its partner *licT* KO and the double mutant *cheB* KO-*licT* KO (Figure 4.20B). The low-throughput testing and analysis of the *ycdH* KO mutant, constitutive partner *yndH* KO and double mutant *ycdH KO-yndH KO* to the stressor vancomycin, revealed no difference in the growth profile when compared to wild type *B. subtilis* treated in the same way,

The growth profiles of the single *cheB* KO mutant, and its *cheB* KO–*licT* KO double mutant were found to be significantly different from that of wild type *B. subtilis* and the *licT* KO single mutant, both of which showed identical growth profiles. The OD of the single and double *cheB* KO mutants declined more steeply than the other mutants in response to vancomycin. The differences seen between the single *cheB* KO and the

double *cheB* KO–*licT* KO was minimal, leading to the conclusion that the removal of the function of *cheB* is not influenced by the presence of the *licT* lesion. These observations were similar to those made with the same mutants in the high-throughput analysis.

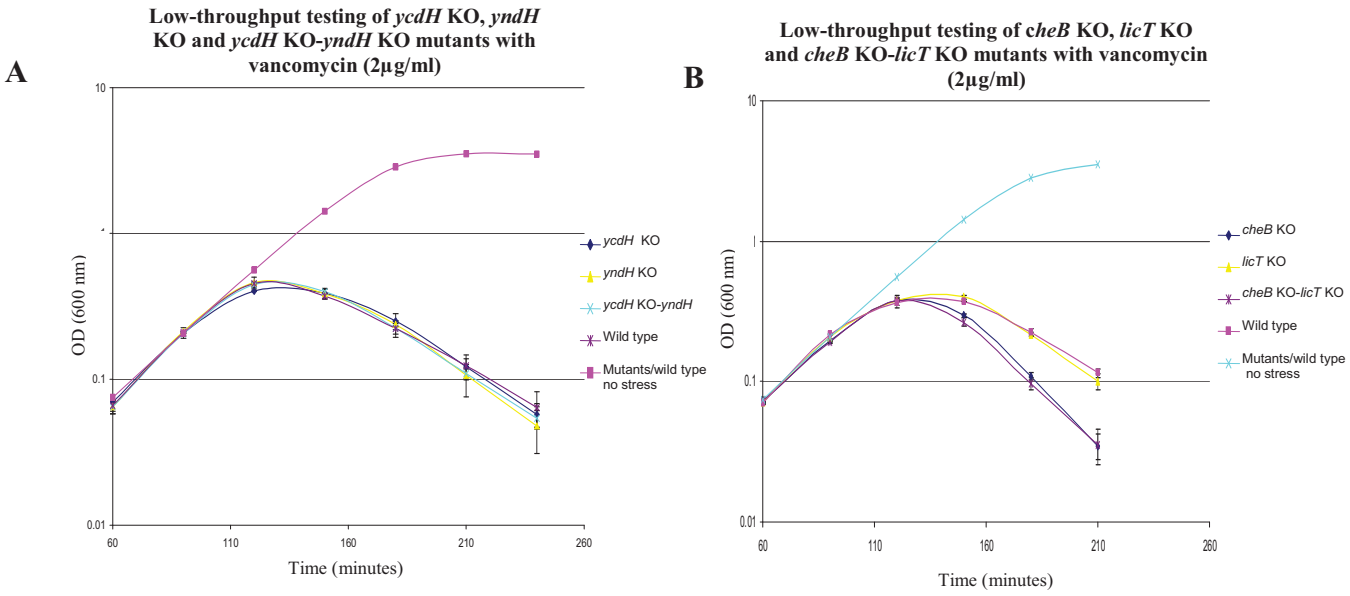


Figure 4.20: Low-throughput testing of single and double mutants with the stressor vancomycin. A, The growth profile of single KO mutants *ycdH*, *yndH* and double KO mutant *ycdH* KO-*yndH* KO in comparison to wild type *B. subtilis* when subjected to the stressor vancomycin (2 µg/ml). Also displayed are unstressed mutants/ wild type. B, The growth profile of single KO mutants *cheB, licT* and double KO mutant *cheB* KO-*licT* KO in comparison to wild type *B. subtilis* when subjected to the stressor vancomycin (2 µg/ml). Also displayed is a sample growth curve for the unstressed mutants and wild type.

The single and double *cheB* KO and *licT* KO mutants were retested with the stressor vancomycin (2µg/ml) but without the addition of inducers to activate the genes downstream of *cheB* and *licT*. This allowed the influence of the downstream genes to be determined (Figure 4.20B and 4.21).

The presence or absence of inducer had no influence on the growth profiles of both the single mutants *cheB* KO and *licT* KO or the double mutant *cheB* KO-*licT* KO, indicating that the genes downstream did not produce the observed phenotypes.

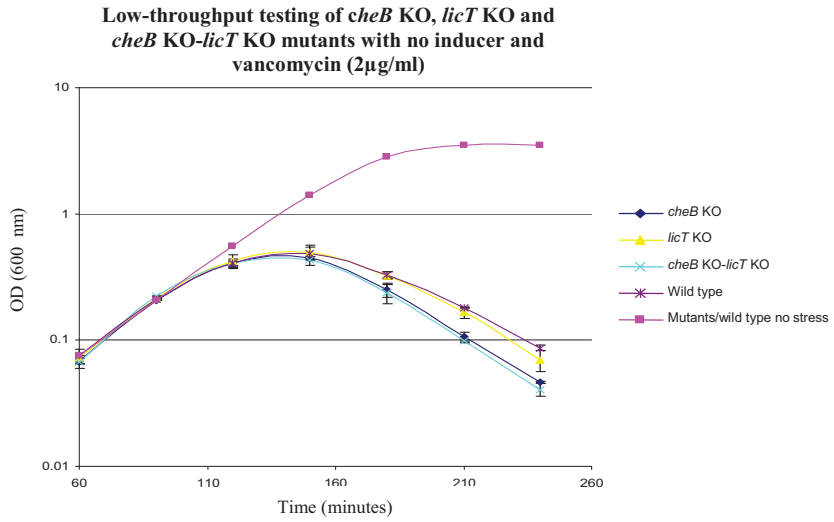


Figure 4.21: Low-throughput testing without inducer of single and double mutants with the stressor vancomycin. The growth profile of single KO mutants *cheB, licT* and double KO mutant *cheB* KO-*licT* KO without inducer in comparison to wild type *B. subtilis* when subjected to the stressor vancomycin (2 µg/ml). Also displayed is a sample growth curve for the unstressed mutants and wild type.

.

4.3.6 Growth kinetics following a challenge with paraquat

An initial low-throughput analysis of mutants with paraquat at the same concentration used in the high-throughput experiments (2.4mM), produced a rapid decline in growth, that was now unsuitable for the investigation. Because of this, the concentration was reduced to 0.4mM, allowing the investigation to continue.

Strains with the *cheB* KO and *licT KO* mutations, singly and in combination (Figure 4.22), were stressed with paraquat at a final concentration of 0.4mM. This revealed that there were no differences between the growth profiles of either the single *cheB* KO and *licT* KO mutants or the double *cheB* KO-*licT* KO mutants when compared to wild type *B. subtilis*.

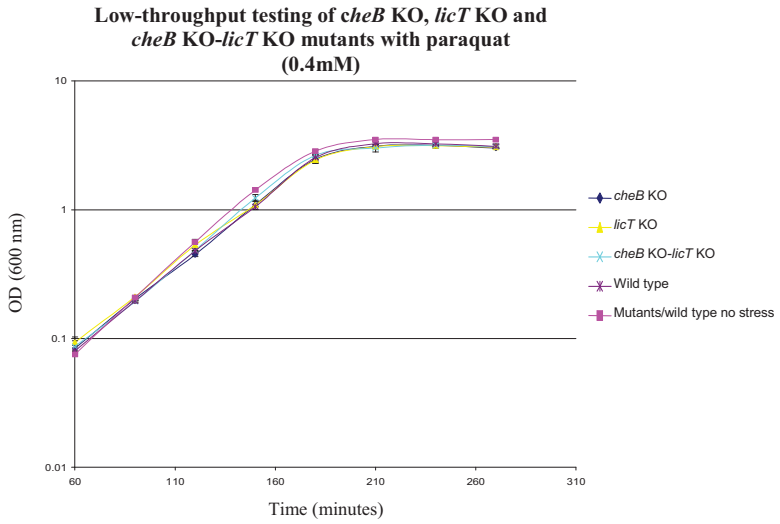


Figure 4.22: Low-throughput testing of single and double mutants with the stressor paraquat. The growth profile of the single KO mutants *cheB* and *licT* and the double KO mutant *cheB* KO-*licT* KO in comparison to wild type *B. subtilis* when subjected to the stressor paraquat (0.4mM). Also displayed is a sample growth curve for the unstressed mutants and wild type.

.

4.3.7 Growth kinetics following a challenge with heat

Strains with the *cheB* KO and *licT* KO mutations, singly and in combination (Figure 4.23A), together with strains *ycdH* KO and *yndH* KO mutations, singly and in combination (Figure 4.23B) were stressed with heat to a final temperature of 45ºC. This low-throughput testing and analysis revealed that there was no difference between the growth profiles of either the single mutants or double mutants when compared to wild type *B. subtilis* treated in the same way.

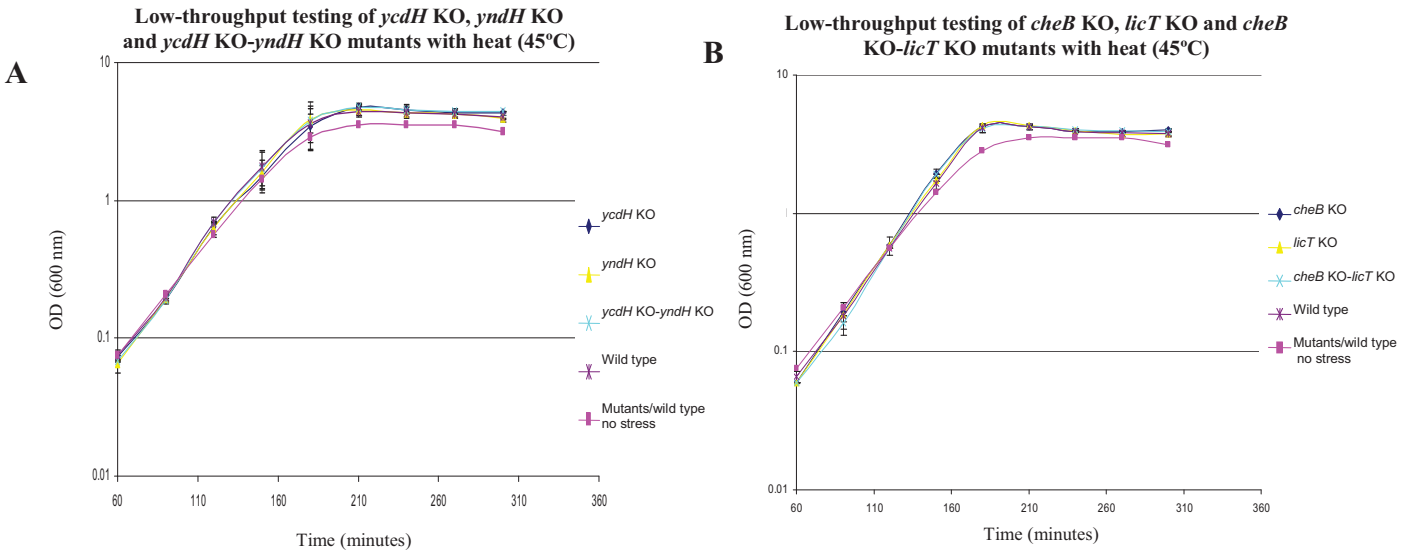


Figure 4.23: Low-throughput testing of single and double mutants with the stressor heat. A, The growth profile of the single KO mutants *ycdH*, *yndH* and double KO mutants *ycdH* KO-*yndH* KO when subjected to 45ºC. Also displayed are unstressed mutants/wild type. B, The growth profile of the single KO mutants *cheB, licT* and *cheB-licT* double KO mutant when subjected to 45ºC. Also displayed is a sample growth curve for the unstressed mutants and wild type.

4.3.8 Growth kinetics following a challenge with NaCl

Strains with the *cheB* KO and *licT* KO mutations, singly and in combination (Figure 4.24), were stressed with NaCl to a final concentration of 5% (w/v). This low-throughput testing and analysis revealed that there were no difference between the growth profiles of either the single *cheB* KO and *licT* KO mutants or the double *cheB* KO-*licT* KO mutants when compared to wild type *B. subtilis* treated in the same way.

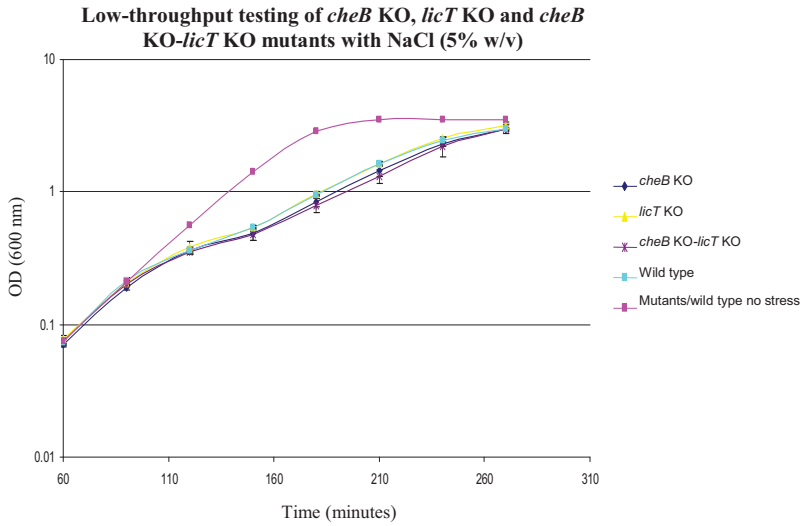


Figure 4.24: Low-throughput testing of single and double mutants with the stressor NaCl. The growth profile response of the single KO mutants *cheB* and *licT* and the *cheB* KO-*licT* KO double mutant in comparison to wild type *B. subtilis* when subjected to the stressor NaCl (5% w/v). Also displayed is a sample growth curve for the unstressed mutants and wild type.

4.3.9 Low-throughput analysis of therapeutic target summary

The mutant that consistently displayed a growth profile that was different from the wild type in both high- and low-throughput analyses was *cheB* KO (Table 4.2). Analysis of the single and double mutants indicate that the altered profiles are the result of the *cheB* lesion, since the double mutant behaves in a manner that is identical to that of *cheB* rather than the *licT* single mutants. Moreover, the potential influence of genes downstream of *cheB* was ruled out by determining the growth profiles in the presence and absence of inducer, xylose (*cheB)* and IPTG (*licT*) indicating that the altered behaviour was due to the *cheB* (Figure 4.19B and Figure 4.20).

Table 4.2: Summary of the results of low-throughput testing on mutants identified in the high-throughput testing stage. X indicates no difference to wild type *B. subtilis* exposed to the same stressor. The appearance of gene name beside a stress indicates the gene that is responsible for the susceptible of *B. subtilis* to the stressor.

Stressor	Mutant(s) (KO) A/B (Single mutants) A-B (Double mutant)					
	ybfS/ywdH *ybfS-ywdH*	*luxS/yvgQ* *luxS-yvgQ*	*ycdH/yndH* *ycdH-yndH*	*fbaB/yacL* *fbaB-yacL*	*yjcH/abnA* *yjcH-abnA*	*cheB/licT* *cheB-licT*
Nalidixic acid	X	X	X	X	X	X
Kanamycin	X	X	X	X	X	X
Streptomycin	X	X	X	X	X	X
Tetracycline	X	X	X	X	X	X
Vancomycin	X	X	X	X	X	*cheB*
Paraquat	X	X	X	X	X	X
Heat	X	X	X	X	X	X
NaCl	X	X	X	X	X	X

4.4 Additional small scale analyses

The effects of vancomycin on the growth profile of the *cheB* mutant warranted further investigation, to determine the underlying mechanism for the results seen. This was performed as part of an additional program of low-throughput experimentation.

4.4.1 General characteristics.

The growth profile of *cheB* mutant was identical to that of wild type *B. subtilis,* in the absence of vancomycin when grown in LB. This suggested that the absence of the CheB protein has little or no affect in nutrient medium. The motility of the *cheB* KO mutant was compared with that of wild type *B. subtilis* through exponential and stationary phase using light microscopy. In comparison with the wild type, the swimming behaviour of the mutant was not altered during late exponential phase – like the wild

type, the mutant swam with a similar bias between tumbles and straight runs. These results are consistent with the previously described phenotype of *cheB* mutants (Kirsch *et al.*, 1993; Zimmer *et al.*, 2002).

4.4.2 Wall synthesis

Bacterial cell walls are composed primarily of peptidoglycan, a polymer of alternating N-acetylmuramic acid and N-acetylglucosamine. Attached to N-acetylmuramic acid is a pentapeptide chain, that facilitates the cross linking with the glycan chains using the enzyme transpeptidase to give the cell wall structural integrity. Vancomycin functions by binding to the terminal D-ala-D-ala residues of the pentapeptide chain preventing the transpeptidase enzyme from cross linking chains (Walsh, 2000). As such the cell wall becomes susceptible to osmotic lysis.

To determine whether the behaviour of the *cheB* mutant was the result of a general effect on cell wall synthesis, an investigation was carried out to determine the effect of adding ampicillin as a stressor. Ampicillin functions by preventing peptidoglycan cross linking through the inhibition of the transpeptidase enzyme directly rather than in the case of vancomycin acting upon its substrate. If the *cheB* mutation was affecting cell wall synthesis, the behaviour observed with vancomycin might have been expected to apply to ampicillin.

Preliminary tests were conducted to establish the optimum concentration of ampicillin to add to wild type *B. subtilis* to perturb but not inhibit growth Concentration of ampicillin between 0.1 and 0.4 µl/ml were applied Increasing concentrations of the antibiotic had increasing significant influences on the growth profile. It was concluded that the optimum concentration for use would be 0.1 µg/ml (Figure 4.25).

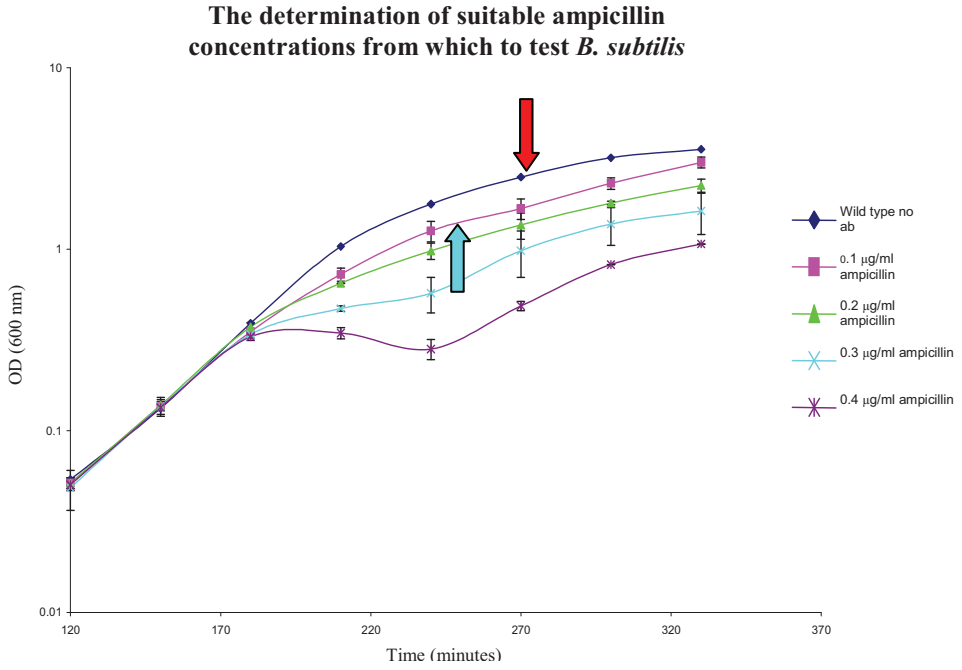


Figure 4.25: The influence of various concentrations of ampicillin on *B. subtilis*. Dark blue, No ampicillin (red arrow). Pink, 0.1 µg/ml. Green, 0.2 µg/ml. Turquoise, 0.3 µg/ml. Purple, 0.4 µg/ml. The blue arrow indicates the selected antibiotic concentration for use in further investigations.

When 0.1 µg /ml ampicillin was used to stress the *cheB* KO*, licT* KO and *cheB-licT* KO mutants their growth profiles were identical to the wild type, indicating that the influence of vancomycin on the *cheB* mutant was due to a specific rather than a general effect on cell wall synthesis (Figure 4.26).

The addition of ampicillin to growing cultures of both *cheB* KO and wild type *B. subtilis* produces a slight reduction in both the growth rate and eventual stationary phase culture density when compared to untreated wild type *B. subtilis*. These results suggest that the mutation does not have an influence on general cell wall biosynthesis, suggesting that the *cheB* KO mutation is affecting another cellular function

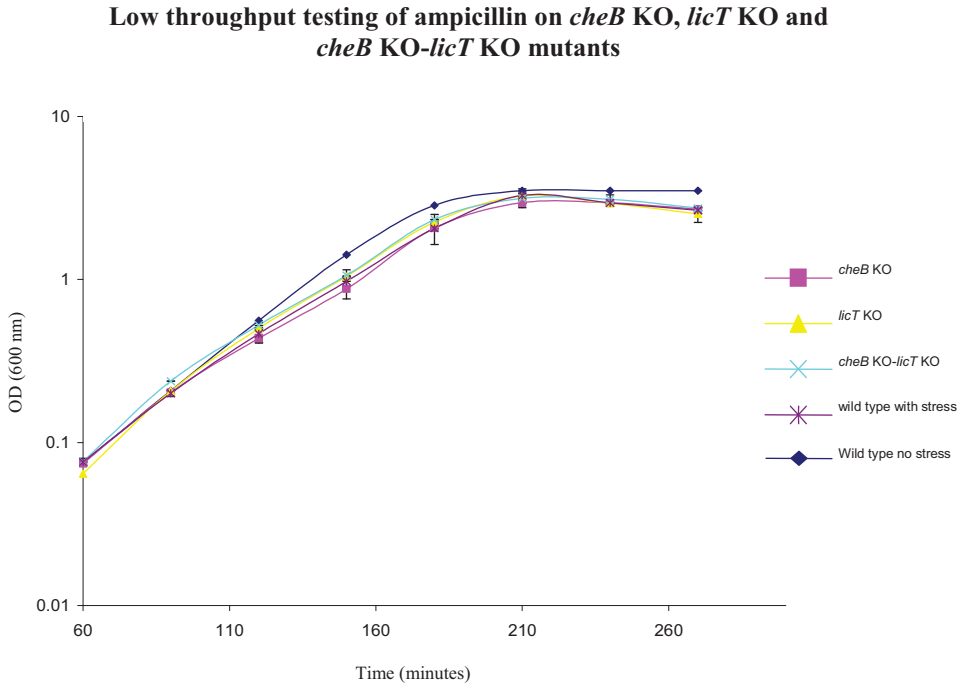


Figure 4.26: Low-throughput testing of single and double mutants with the stressor ampicillin. The growth profile response of the single KO mutants *cheB* and *licT* and the *cheB* KO-*licT* KO double mutant in comparison to wild type *B. subtilis* when subjected to the stressor ampicillin (0.1µg/ml). The growth profiles of the unstressed wild type and mutants are also included.

.

4.4.3 Microscopy

The influence of vancomycin on the growth profile of the *cheB* KO mutant and wild type was noted in section 4.3.5. To attempt to understand the underlying basis for these observations, morphological studies were conducted using phase contrast and fluorescence microscopy. Wild type *B. subtilis* and *cheB* KO mutants were grown in LB at 37ºC and samples removed every 10 minute over a period of 40 minutes, following the addition of vancomycin during mid exponential growth (Figure 4.27). Samples were mounted and studied using a range of microscopy techniques, designed to identify any physical differences that may be present between the samples.

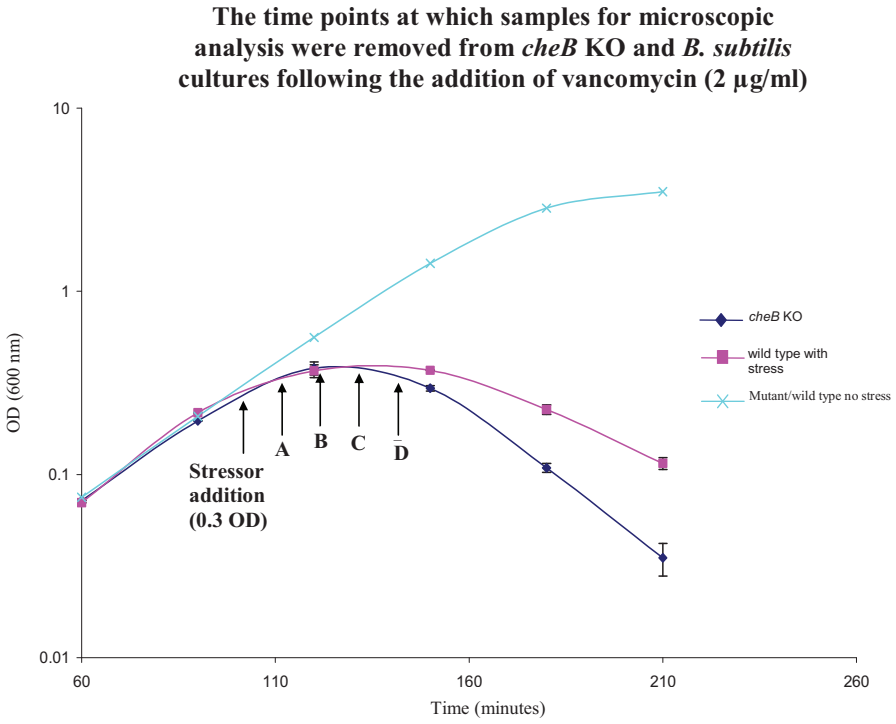


Figure 4.27: A graph indicating sample time points for microscopy analysis. Indicated on the graph are the time points, that samples were removed for microscopic analysis after the addition of 2 µg/ml vancomycin. Sample A, 10 minutes post stress addition. B, 20 minutes post stress addition. C, 30 minutes post stress addition. Sample D, 40 minutes post stress addition.

4.4.3.1 General morphology

The appearance of the wild type *B. subtilis* and *cheB* KO mutant cells was analysed for differences in cell size, morphology and timing of cell division following the addition of vancomycin. Cell size and morphology were similar, but there were indications that the timing of cell division was delayed by approximately 10 minutes in the case of the wild type (Figure 4.28). The *cheB* KO mutant stopped forming dividing chains at time point A after the addition of vancomycin, with decreasing numbers of bacteria dividing and forming chains at time points C and D after the addition of vancomycin.

These observations are consistent with the growth profiles observed for these sample sets, a reduction in growth by approximately 20 minutes after the addition of vancomycin before a prominent decline in growth rate occurred.

4.4.3.2 DAPI staining of chromosomal DNA

Following morphological analyses by phase contrast microscopy, samples were stained with 4'-6-diamidino-2-phenylindole (DAPI) DNA stain to identify the location and condition of the nuclear material when visualised with fluorescence microscopy. The results obtained for both sample sets taken at 30 and 40 minutes were typical of normally cells preparing to divide with two regions of nuclear material per cell (Figure 4.28).

4.4.3.3 Fluorescent vancomycin staining

With no differences seen on the addition of vancomycin to sample sets, fluorescent vancomycin was used in replacement in a ratio of 50:50 with standard vancomycin to a final concentration of 2 μg/ml, This enabled the distribution and action of the compound to the sample sets to be visualised, indicating any potential differences that may be occurring (Figure 4.28). Analysis has indicated the comparable distribution of fluorescent vancomycin attachment between *cheB* KO and wild type *B. subtilis* sample sets. Time points A and B indicated a general binding of fluorescent vancomycin to the cell wall, with a larger incorporation of fluorescent vancomycin occurring at time points C and D, the point of greatest peptidoglycan synthesis and monomer cross-linking.

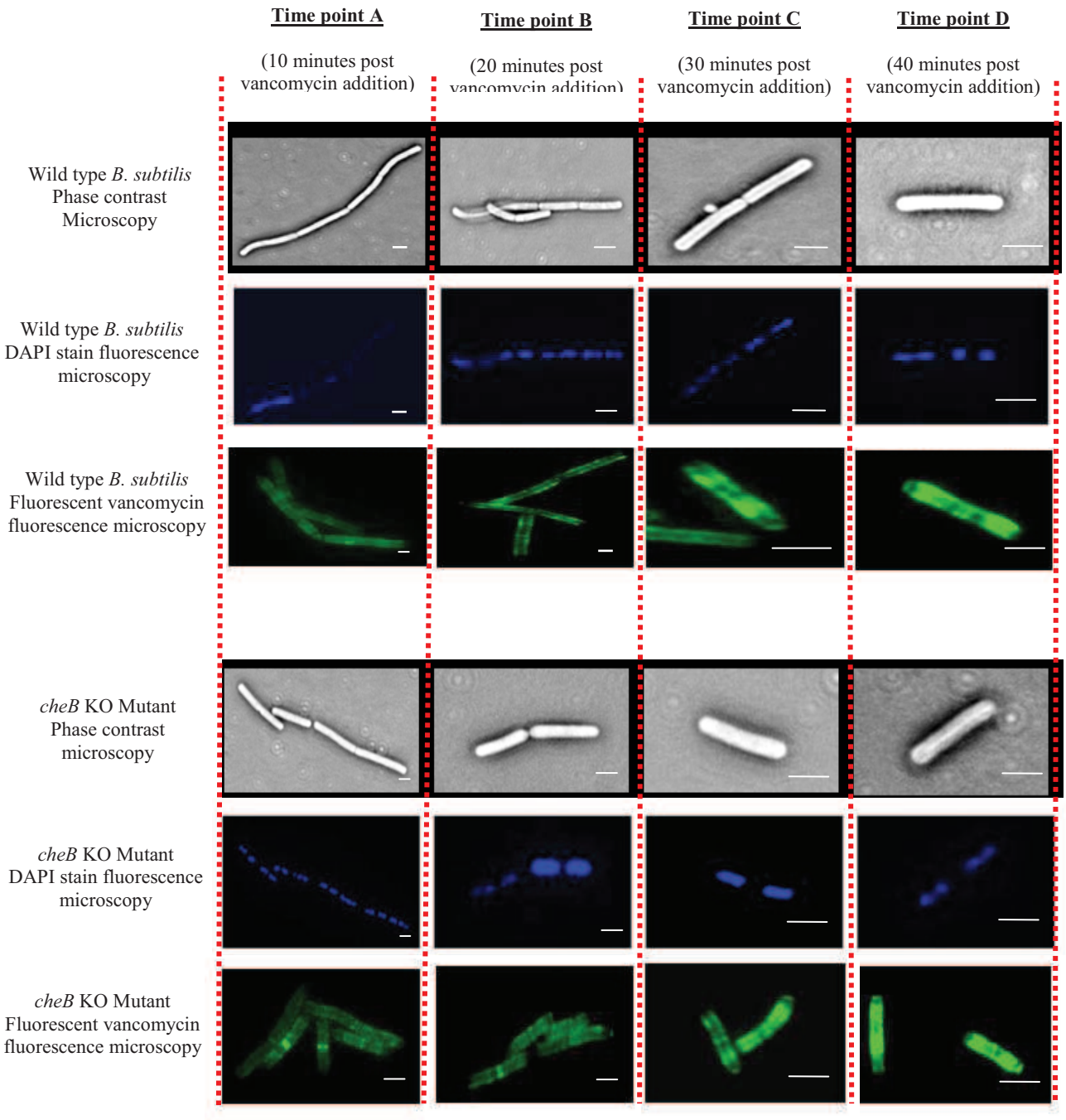


Figure 4.28: Microscopy comparison and analysis of the *cheB* KO mutant with wild type *B. subtilis*. *cheB* KO samples taken at the time points A-D as indicated in Figure 4.26 after the addition of 2 µg/ml vancomycin. Samples were initially visualised by phase contrast microscopy (grey) after which samples were stained with DAPI staining (blue) and visualised using fluorescence microscopy. Samples in a separate investigation were treated with a 50:50 mixture of fluorescent vancomycin and standard vancomycin to a final concentration of 2 µg/ml (green) and visualised using fluorescence microscopy. Scale bars indicate 2µm length.

4.4.4 Small scale analysis summary

To determine whetheher the effect of vancomycin was specific to the compound or was a generic response to cell wall synthesis acting antibiotics, experimentation was repeated using ampicillin. It was found that all cells tested behaved in the same way as wild type *B. subtilis*, indicating that the effects of vancomycin were specific to the compound. Microscopic investigation techniques were then used to investigate for any unique properties specific to the *cheB* KO mutant when compared to wild type *B. subtilis*.

The results of such investigations found only a single difference, this being a reduction in time taken by *cheB* KO mutant cells to elongate and divide. This difference was initially small, but became more prominent as exponential growth continued. Additional observations using DAPI and fluorescent vancomycin staining of the two cell types, indicated no difference in nuclear or peptidoglycan chain localisation.

4.5 The arginine and pyrimidine (uracil) biosynthetic systems

4.5.1 Introduction

The arginine biosynthetic pathway shares a common intermediate-carbamoyl phosphate with the pyrimidine biosynthetic pathway. It was of interest to discover that the genes for key steps in both pathways were clustered together on the chromosome and with each pathway having a seemingly dedicated carbamoyl phosphate synthetase (CPS) enzyme, where most prokaryotes utilise just one. This organisation and presence of two CPS suggested the possibility that pathway intermediates from each individual system may be being channelled through distinct multi enzyme complexes to other components within the same system, preventing them from entering alternative utilising systems. This could explain the clustered organisation of biosynthetic pathway genes and the duplication of the CPS enzyme. The aim of this component of the project was to express and purify key components of both biosynthetic systems to determine if there could be any interrelationship between them.

4.5.2 Approach

To study this, a structural and interaction analysis was initiated. Recombinant proteins of all key components of both biosynthetic systems were to be isolated in preparation for crystallisation and structural analyses as well as protein interaction studies.

The study was initiated by cloning the genes encoding both the arginine and pyrimidine specific CPS's as well as the *argF* gene encoding an ornithine transcarbamylase enzyme into the pMAL expression vector to generate affinity tagged fusion proteins. After expression in *E. coli*, these proteins would be purified by affinity chromatography and analysed by SDS-PAGE, before being subjected to crystallisation, structural and protein interaction analyses.

4.5.3 pMAL expression system

The pMAL system has been designed to give controlled expression at high levels of up to 100 mg/litre of soluble recombinant proteins within *E. coli*. Proteins are induced and expressed using the strong P_{tac} promoter and emerge fused to a maltose binding protein (MBP) (di Guan *et al.*, 1988; Maina *et al.*, 1988; Riggs, 2001) (see appendix). MBP-tagged proteins can then be purified using a one-step affinity purification process, after which the tag can be cleaved from the protein using a protease specific for the linker region located between the recombinant protein and the MBP.

The ability to perform crystallographic and interaction studies is dependant on the production of soluble proteins that can precipitate out to form crystals. Two versions of pMAL allow for synthesis and folding of the target protein in either the cytoplasm (c) or the periplasm (p) of *E. coli*. The translocation and subsequent folding of the target protein in the periplasm increases the likelihood of the protein being folded into its native configuration. The inclusion of MBP within the recombinant protein, not only ensures the ability to easily purify expressed proteins easily, but also by its inclusion reduces the likelihood of producing insoluble proteins (Kapust & Waugh, 1999).

4.5.4 Expression vector construction

Upstream of the multiple cloning site (MCS) of pMAL is the *malE* gene encoding a MBP. The target gene is cloned in-frame with the MBP, separated by a cleavage site on the linker between the two sequences. High level expression of the target recombinant protein is achieved by using the IPTG-regulated P_{tac} promoter via the $lacI^q$ encoded lactose repressor within the plasmid. A *ß*-lactamase gene has also been included to allow selection of pMAL transformed *E. coli* cells.

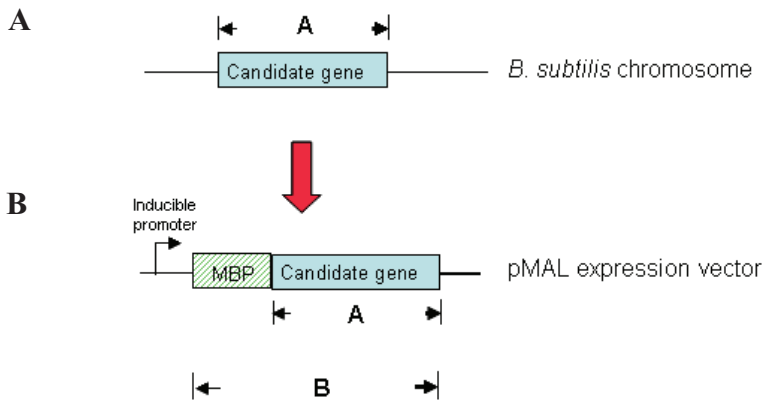


Figure 4.29: pMAL expression vector construction. A, Candidate gene to be expressed located on the *B. subtilis* chromosome. Primer pair A, forward and reverse primers for the candidate gene. B, Candidate gene integrated into the pMAL expression vector. Primer Pair B, forward primer for the MBP and reverse primer for the candidate gene

PCR primer pairs spanning the beginning and end of the target gene were designed (primer pair A) (Figure 4.29), with each having a restriction site that was complimentary to a restriction site in the MCS of the plasmid. The forward primer was designed to ensure that the target gene was in-frame with that of the MBP. The reverse primer included an additional stop codon at the end of the gene to ensure minimal translational read-through. The PCR was performed using the proof-reading polymerase *Pfx* (invitrogen) to produce full length copies of the target genes together with the additional features described (Figure 4.30).

The genes encoding the proteins ArgF, CarA (carbamoyl phosphate synthetases – small subunit) and CarB (carbamoyl phosphate synthetases – large subunit) of the arginine biosynthetic system were cloned into pMAL together with the genes encoding PyrAA (carbamoyl phosphate synthetases – small subunit) and PyrAB (carbamoyl phosphate synthetases – large subunit) of the pyrimidine biosynthetic system.

The pMAL plasmid and amplified genes were digested with cognate restriction endonucleases and ligated, before transforming and selecting in *E. coli*. Putative transformants were subject to a second PCR, this time using the forward primer for the *malE* gene and the reverse primer for the inserted gene (primer pair B), to confirm the presence and orientation of insert (Figure 4.29, Figure 4.3).

A separate set of sequencing primers ca. 800 bp apart were designed to allow the entire gene sequence of the tagged target protein to be confirmed.

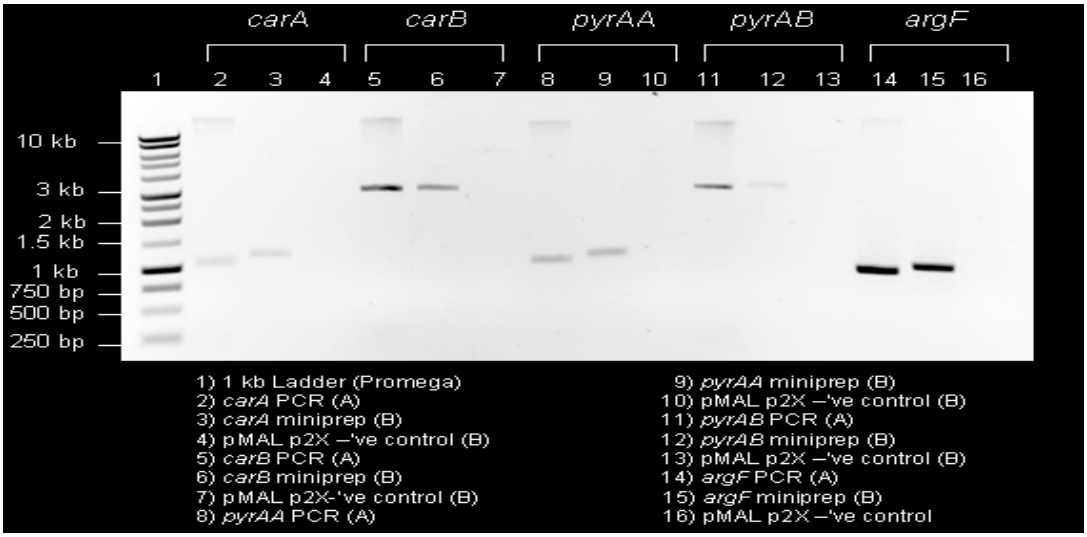


Figure 4.30: Construction stages of pMAL expression strains for the production of key arginine and pyrimidine biosynthetic system components. An agarose gel photograph of the stages of construction of the various expression strains, used to produce recombinant proteins for the analysis of the arginine and pyrimidine biosynthetic systems. Lane contents and primer pairs (A/B) are represented under the gel photograph.

4.6 Analysis of the arginine and pyrimidine biosynthetic systems

4.6.1 Protein expression

The analysis of potential interactions between components of the arginine and pyrimidine biosynthetic systems were to be conducted using the biophysical techniques of surface plasmon resonance (SPR) (Boozer *et al.*, 2006; Torreri *et al.*, 2005) and isothermal titration calorimetry (ITC)(Velazquez Campoy & Freire, 2005). To perform these techniques as well as subsequent crystallisation and structural analyses required relatively large quantities of protein, to which the pMAL system was well suited.

Before these investigations could begin small-scale studies were first performed in the *E. coli* host strain TB1 to verify the presence of recombinant proteins, their expression profile and solubility. Each of the constructed recombinants was grown at 30^{o}C in the presence of the inducer IPTG (1mM). The inducer was added at mid exponential phase and the cultures harvest at two hours and four hours post induction. Samples were

centrifuged and the pellet resuspended in gel filtration buffer, before being sonicated. A sample of the sonicated mixture was taken as a representation of the whole-cell including both soluble and insoluble fractions. The remaining mixture was centrifuged and a sample taken of the supernatent representative of the soluble cell fraction. Both these samples were then checked for the presence of the recombinant protein by running on SDS-PAGE gels, and compared against a protein standard of known size. Uninduced samples taken at the same time points were also processed for comparative purposes.

The recombinant proteins CarA, PyrAA and ArgF were expected to produced bands in the size range of between 76 and 81 kDa, while the larger PyrAB and CarB proteins were expected to produce bands in the size range of 154 to 160 kDa. These sizes included the fused MBP affinity tag of 42 kDa.

The recombinant proteins CarA, CarB, PyrAA and PyrAB failed to be over expressed when compared to both soluble and whole cell fractions of their uninduced counterparts (Figure 4.31). The lack of initial evidence for recombinant protein over-expression from four of the constructs prompted further investigation. All expression constructs were sequenced to ensure the presence, correct orientation and required reading frame of the target gene within the pMAL vector and deemed correct. Induction of the recombinant plasmids was repeated at lower temperatures (25^{o}C) and with lower inducer concentration (0.3 mM IPTG), as well as attempts to over-express the proteins carB and pyrAB in an alternative host strain of *E. coli* (BL21). An attempt was also made to express protein carB in the alternative expression vector pGEX but all with no success.

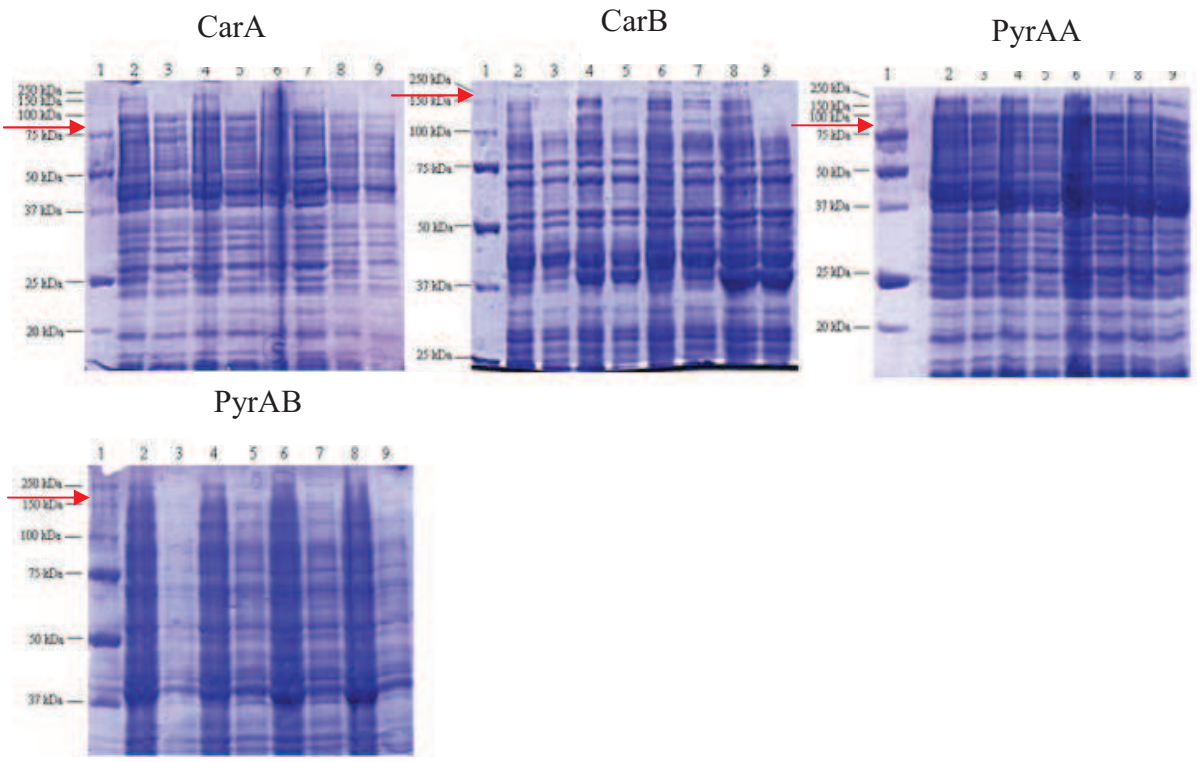


Figure 4.31: SDS-PAGE analysis of key proteins in the arginine and pyrimidine biosynthetic systems. Expressed recombinant target proteins CarA, CarB, PyrAA, PyrAB from the pMAL expression vector in *E. coli* strain TBI. Lane 1, BioRad molecular size ladder. Lane 2, 2 hours post induction with no IPTG, whole cell fraction. Lane 3, 2 hours post induction with no IPTG, soluble fraction. Lane 4, 2 hours post induction with 1mM IPTG, whole cell fraction. Lane 5, 2 hours post induction with 1mM IPTG, soluble cell fraction. Lane 6, 4 hours post induction with no IPTG, whole cell fraction. Lane 7, 4 hours post induction with no IPTG, soluble cell fraction. Lane 8, 4 hours post induction with 1mM IPTG, whole cell fraction. Lane 9, 4 hours post induction with 1mM IPTG, soluble cell fraction. The expected size of product is indicated on lane 1 of the molecular size ladder with a red arrow.

The clone expressing the ArgF protein showed evidence of a recombinant protein in the whole cell fraction, but a corresponding band was not detected in the soluble or non-induced cell fraction (Figure 4.32). Attempts were made to produce soluble protein by using a variety of growth temperatures from 16-37^{o}C, various concentrations of inducer from 0.1-1mM IPTG, and alternative *E.coli* host strains, namely Bl21 instead of TB1. All attempts were unsuccessful, with only the insoluble form of ArgF being produced.

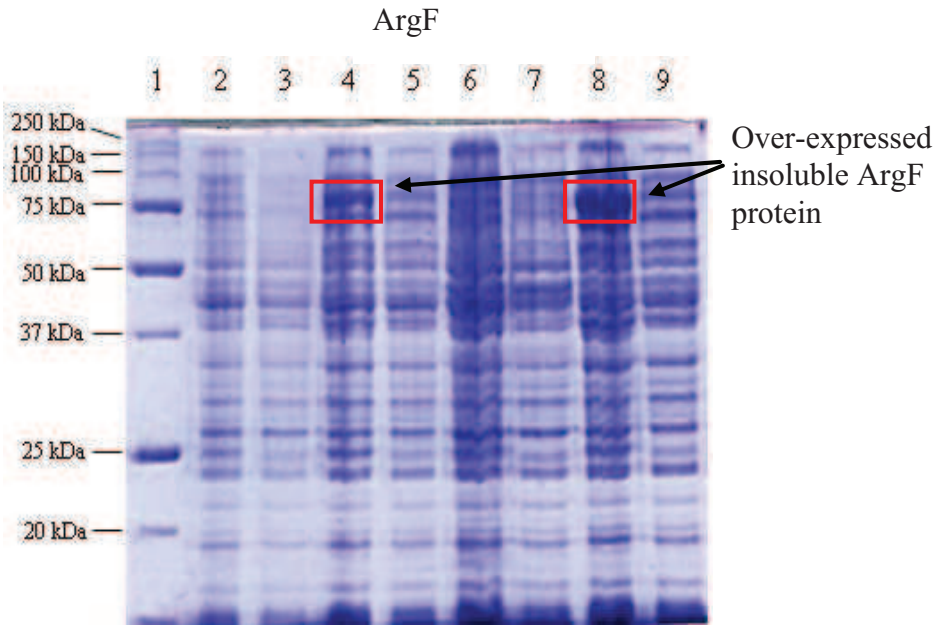


Figure 4.32: SDS-PAGE analysis of the expressed recombinant protein ArgF. Expressed recombinant ArgF protein produced by the pMAL expression vector in *E. coli* strain TBI. Lane 1, BioRad molecular size ladder. Lane 2, 2 hours post induction with no IPTG, whole cell fraction. Lane 3, 2 hours post induction with no IPTG, soluble fraction. Lane 4, 2 hours post induction with 1mM IPTG, whole cell fraction. Lane 5, 2 hours post induction with 1mM IPTG, soluble cell fraction. Lane 6, 4 hours post induction with no IPTG, whole cell fraction. Lane 7, 4 hours post induction with no IPTG, soluble cell fraction. Lane 8, 4 hours post induction with 1mM IPTG, whole cell fraction. Lane 9, 4 hours post induction with 1mM IPTG, soluble cell fraction.

4.6.2 Summary

Attempts to express proteins of the arginine and pyrimidine biosynthetic pathways in *E. coli* using the pMAL expression system proved unsuccessful, despite sequencing and the use a variety of growth and induction conditions together with host strains. These findings could indicate that the problems encountered may be attributed to the

recombinant proteins being unstable, toxic to the host, or requiring the co-expression of other protein(s). Attempts at producing a soluble form of the only recombinant protein, to be produced, ArgF was also unsuccessful, despite attempts to optimise its expression. Although insoluble, ArgF could have been denatured to produce a soluble form, using urea or guanidine hydrochloride, although it would not be clear as to whether or not it would then be possible to re-nature it to an active form.

The lack of expression of potential interacting partners to ArgF and difficulties associated with its solubilisation led to the abandonment of all experiments to investigate protein interactions, crystallisation and structural analysis between members of the arginine and pyrimidine pathways, and instead emphasis the analysis was shifted towards a computational analysis approach. .

4.7 e-Therapeutics and the arginine and pyrimidine biosynthetic system discussion

4.7.1 e-Therapeutics target testing

A temperature controlled shaking microplate reader was evaluated for the high-throughput screening of a number of mutants constructed for the evaluation of e-Therapeutics predicted therapeutic target genes. The reader provided data that was sufficiently accurate and consistent for such as screening procedure. Appropriate growth and stress induction parameters were established, and all of the mutants constructed were analysed to determine their response, relative to wild-type *B. subtilis*, to a range of different stressors, each designed to target a different area of cellular metabolism.

The implementation of this approach did not identify any gene targets that when targeted together as predicted by e-Therapeutics could act as a therapeutic target, although it did identify several mutants whose behaviour was significantly differently to that of wild type *B. subtilis* warranting further investigation.

The *cheB* KO mutant produced consistently different growth profiles to wild-type *B. subtilis* in the presence of the majority of stressors in the high-throughput investigation stage, the main exception being with the stressor rifampicin. The *cheB* KO mutant behaviour was independent of the presence or absence of its cognate partner gene (*licT* KO) mutation.

The *luxS* single and double KO mutants, produced different growth profiles to wild type *B. subtilis* when the stressor nalidixic acid was applied, behaviour that was shared also

by the single and double KO mutants of *ycdH* in response to the stressors heat and tetracycline.

These initial findings were evaluated using a low-throughput approach. Growth experiments were repeated in larger culture volumes using mutants and the stressors that produced abnormal growth profiles. Low- throughput analysis failed to confirm the high-throughput findings for the *ycdH* and *luxS* KO mutants.

Low-throughput analysis did, however, confirm the abnormal response of the *cheB* KO mutant to vancomycin. This unexpected behaviour led to a series of more comprehensive low-throughput investigations aimed at identifying the underlying mechanisms responsible for the behaviour observed. The *cheB* KO mutant was subjected to sub-inhibitory concentrations of ampicillin, an antibiotic like vancomycin,- inhibiting cell wall synthesis and allowing the determination of whether the effects seen were specific to the vancomycin compound or a more general response by the cell to cell wall stress.

Since the growth profile of the *cheB* KO mutant was found to be indistinguishable from that of wild-type *B. subtilis* in the presence of ampicillin, it was concluded that the response generated was specific to vancomycin.

Following treatment with vancomycin, the *cheB* KO mutant underwent microscopic analysis, which identified a reduction in the rate of cell division, placing samples approximately 10 minutes behind wild-type *B. subtilis* treated in the same way. There were no differences noted in morphology, localisation of nuclear material or the distribution of peptidoglycan within the mutants. As expected (Kirsch *et al.*, 1993; Zimmer *et al.*, 2002), deletion of *cheB* did not result in a change in observable swimming behaviour. In conclusion, virtually all of mutants exhibited a growth profile in response to the various stresses that was indistinguishable from the wild-type. The main exception was the response of the *cheB* mutation to vancomycin. The high-throughput and subsequent low-throughput methodology used in this investigation was successful in identifying changes in mutants compared to wild type *B. subtilis*, and would be suitable to be applied to larger industrial and research applications that use a Systems Biology approach in their investigations.

To further the research conducted in this investigation, a series of reporter gene transcriptional fusion mutants with an inducible producer could be produced for target genes, allowing their transcriptional activity to be modulated and effects studied.

The identification of the *cheB* KO mutant, with altered behaviour, identified using wet lab approaches, can now be complemented by computer-based modelling and analysis to try to determine the underlying biological mechanism for its behaviour.

4.7.2 Arginine and pyrimidine biosynthetic systems discussion

The inability to produce soluble recombinant proteins for key components of the arginine and pyrimidine biosynthetic pathways to use in biophysical interaction studies has prevented the completion of the wet lab studies. During small scale experiments to test the production and solubility of recombinant proteins, only a single candidate, argF could be expressed but in an unusable insoluble form. Attempts were made to solubilise the protein using different biological methods, but without success. All constructs were also sequenced and shown to be correct.

As a result of the problems encountered expressing recombinant proteins, a bacterial two-hybrid approach could be adopted in the future for protein interaction studies which could prove to be more successful. However, to validate any identified interactions using this method would still require the production of recombinant proteins for further biophysical analyses and so alternative recombinant protein production strategies, possibly using different expression systems could be explored.

Since it was not possible to conduct experimental investigations on the arginine and pyrimidine biosynthetic systems, an alternative computational approach was applied instead (Chapter 5).

Chapter 5
SubtilNet2 Compilation

5. SubtilNet2 Compilation

This chapter describes the methods used to produce a new *B. subtilis* functional interaction network, based on the architecture of a previously generated *B. subtilis* functional interaction network, SubtilNet, as well as the analytical programs that would be used for its analysis.

This new network-SubtilNet2 was applied to computationally investigate the predicted therapeutic targets supplied by e-Therapeutics, to try and identify any functional associations that may occur between them that could make them potential therapeutic targets. This analysis follows traditional laboratory analysis of the targets, where mutants devoid of target function were produced, combined and tested under a range of different stress conditions. As well as SubtilNet2 identifying any potential functional associations that may occur between targets, the new network was also used to try and discover the molecular basis for the laboratory results obtained.

In a separate investigation, SubtilNet2 was used to investigate potential interactions and their significance that may occur between components of the arginine and pyrimidine biosynthetic systems, after the failure of traditional wet lab based approaches.

5.1 SubtilNet2

Because of the original SubtilNet network having been developed to investigate the range of expressed proteins from multiple *Bacillus* species, modifications were needed to adapt it to the requirements of this investigation. Since the development of SubtilNet in 2005, considerably larger high-throughput data sets from which the network is constructed have become have become available. This justified the development of an entirely new network, adopting many of the concepts and methods developed for the original network. This new network, SubtilNet2, was designed specifically so that it could be used to investigate e-Therapeutics predicted targets as well as the arginine and pyrimidine biosynthetic systems.

The method of weighting an interaction used in the original SubtilNet, which was based on the probability of their being a functional association, has the potential to overlook very weak but potentially important interactions for which the evidence was insufficient for the assignment of an interaction probability. This was acceptable for the

uses to which SubtilNet was applied, namely the analysis of a multiple candidates in a wide-reaching collection of multi-species networks.

In this investigation, where only a few genes and their products were to be studied from a single *Bacillus* species, functional association weighting was removed. This allowed for the identification and investigation of all possible associations. With this being the case, all identified functional interactions had to be validated by returning to their original source.

5.2 SubtilNet2 data Sources

The data sources, compiled to produce SubtilNet2, consisted of multiple different data types, each representing different analytical aspects of *B. subtilis* physiology. This allowed for the production of a diverse functional interaction network. A decision was made to remove many of the original data sets used in SubtilNet, to reduce the amount of data that was not specific to the investigation, and that would require extra computing and analysis power. In addition to this and according to the "Peaking" effect (Trunk, 1979), the addition of large volumes of "noisy" data can result in a decrease in the accuracy of observed results, this would then have required additional manual validation.

The data sources that were selected for use in this investigation were as follows:

- Pathway data from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000). This database contains details genes and the pathways to which they are associated in *B. subtilis.* The data is a collection of manually curated and computer compiled data, containing 39,154 interactions.

- The Database of Transcriptional regulation in *Bacillus subtilis* DBTBS (Sierro *et al.*, 2008). This database contains a collection of experimentally validated gene regulatory relations and the corresponding transcription factor binding sites upstream of *B. subtilis* genes. In total it details 1399 interactions.

- Co-citation data from PubMed (www.ncbi.nlm.nih.gov/pubmed/). This data set is the result of the analysis of abstract data from the PubMed online journal site. 23,624 abstracts for articles containing the term *subtilis* were downloaded, and "mined" for the co-occurrence of the names of all 4,100 genes of *B. subtilis*. This task was performed using a Perl computer script developed by Dr Matthew

Pocock of the Integrative Systems Biology group at Newcastle University. The result being the identification of 4472 node associations within *B. subtilis*.

- The Microbial Protein Interaction Database (MPIDB) (Goll *et al.*, 2008) details a total of 99 *B. subtilis* genetic interactions. The data within MPIDB has been manually curated from literature, imported from databases (IntAct (Aranda *et al.*), DIP (Xenarios *et al.*, 2000), BIND (Bader *et al.*, 2001), and MINT (Zanzoni *et al.*, 2002) and experimental evidence. Data is further supported by evidence on interaction conservation, protein complex membership, and 3D domain contact (iPfam (Finn *et al.*, 2008), 3did (Stein *et al.*, 2005)).

These data sets were integrated together, to produce the SubtiNet2 functional interaction network.

SubtilNet2 was computationally analysed using Cytoscape, an open source Java-based software package developed by Shannon *et al* (Shannon *et al.*, 2003), to facilitate the visualisation and analysis of biological networks. Datasets were combined into a unified, conceptual framework that had the benefit of allowing individual data sets to be added or removed (Figure 5.1). This feature allowed researchers to investigate functional associations based on particular datasets, or a compilation of them all.

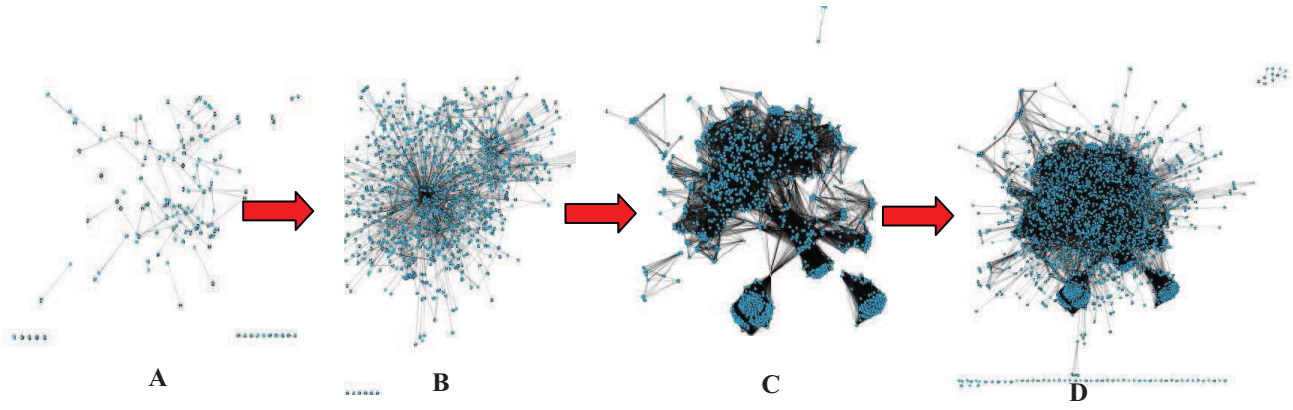


Figure 5.1: Stepwise data compilation of SubtilNet2. The stepwise integration of data sets used to construct SubtilNet2, a functional interaction network for *B. subtilis*. A, MPIDB dataset. B, MPIDB and DBTBS datasets. C, DBTBS, MPIDB and KEGG datasets. C, Complete functional interaction network model- DBTBS, MPIDB, KEGG and co-citation datasets.

5.3 The network analysis of generated models

Using SubtilNet2, various network statistical analyses could be performed within the Cytoscape environment. This was achieved using a range of plug-in programs specifically developed for this purpose. This investigation used the following:

5.3.1 PeSca 2.0

The PeSca2.0 Cytoscape plug-in has been developed by Petterlini & Scardoni (Petterlini & Scardoni, 2008), to analyse networks for the shortest paths between any two nodes. The shortest path between two nodes may indicate the most efficient and likely route of transfer between them (Raman & Chandra, 2008; Zhu *et al.*, 2007). Determining the exact route that the shortest path takes can identify intermediates that may interact with the candidate nodes.

5.3.2 Molecular complex detection algorithm (MCODE)

The molecular complex detection algorithm (MCODE) (Bader & Hogue, 2003) finds highly interconnected regions of nodes (clusters) within generated interaction networks. The presence of clusters could give an indication of nodes that are likely to be functionally associated, and could potentially be components of larger biological systems.

Based on a node weighting scheme, the connectivity of an individual node is analysed and compared to surrounding node connectedness, after which a score is applied.

After weighting, MCODE focuses on high scoring nodes and recursively moves outwards, analysing connected nodes and identifying those that continue to remain above a set weight threshold, potentially indicating membership to a highly connected region and possible cluster. Clusters within a network are scored and then ranked. The score is calculated by multiplying the connectivity of the cluster by the number of nodes within the cluster (Bader & Hogue, 2003) . The larger, denser, clusters score higher than smaller, more sparse, clusters (Bader & Hogue, 2003). Scores within the SubtilNet2 network range from 86.902 for the largest, most highly connected cluster to 1 for the smallest and least connected cluster. For this investigation the settings by which MCODE assigns nodes to clusters- the node score cutoff, was left at its default setting, giving a compromise between the speed and accuracy of cluster identification.

5.3.3 BiNGO

BiNGO (Maere *et al.*, 2005), a Java-based Cytoscape plug-in, analyses and visualises the over representation of gene ontology (GO) designation within selected groups of nodes, giving indications as to the possible biological function of the node groups.

Analysis is performed using the hypergeometric test – a statistical test that compares the proportions of gene ontologies from the analysed network with that of a reference set. Conducted at a significance level of 0.05, any ontologies that score a p-value below this threshold are deemed to be significantly overrepresented in the analysed network and are not simply occurring by chance. As a visual representation of this, the lower the p-value below this threshold the darker orange the gene ontology term becomes (Figure 5.2). In order to conduct accurate BiNGO analysis, the latest GO terms were obtained for *B. subtilis* from the European Bioinformatics Institute (EBI) (www.ebi.ac.uk).

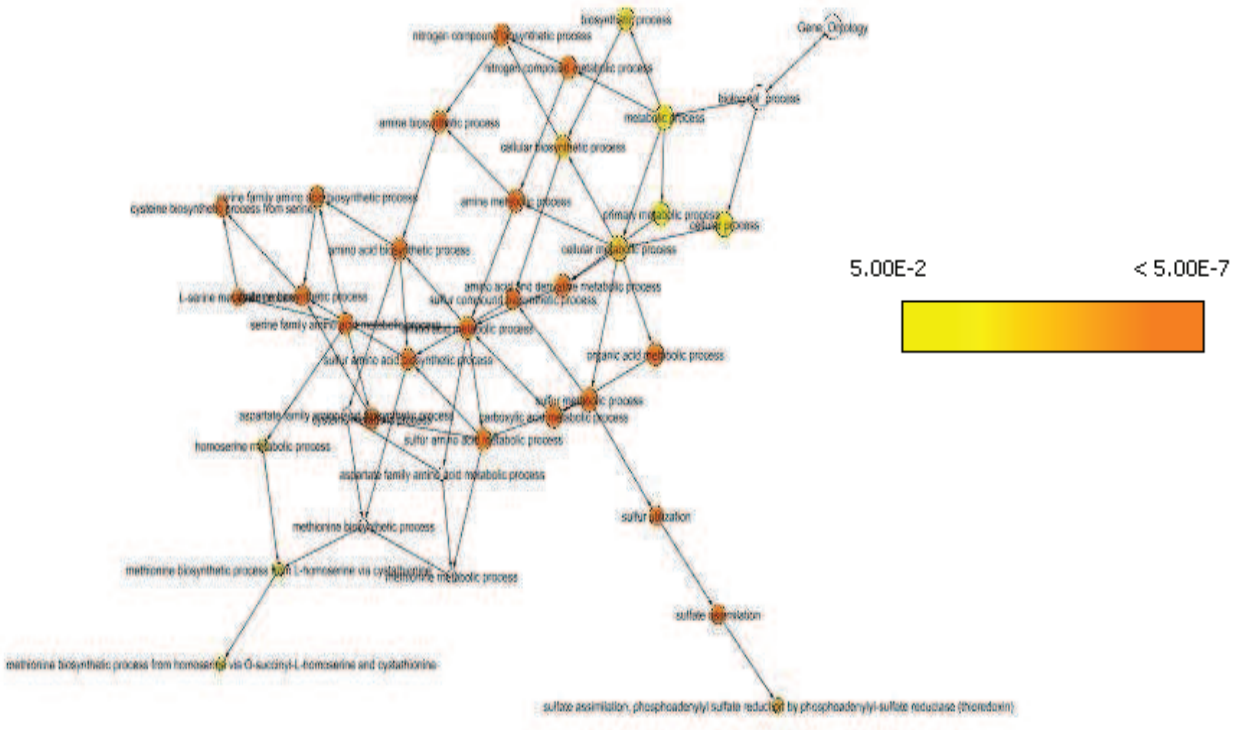


Figure 5.2: A visual representation of the gene ontology of networks generated by the BiNGO plug-in. A generated ontology tree and significance scale, for a single cluster selected from SubtilNet2 and processed with BiNGO. The darker the orange colour associated with gene ontology term, the lower its p-value and greater the significance of overrepresentation within the tree.

5.3.4 Network analyzer

Network analyzer (Assenov *et al.*, 2008) is a Cytoscape plug-in that produces a comprehensive set of topological parameters for generated networks. These include the total number of nodes and edges within a network, a networks diameter, the average number of neighbours for each node, and characteristic path length.

5.3.5 System-wide and local node analysis

Individual nodes of interest can be identified within large interaction networks and, from this, the above-described plug-ins can be used to determine any unusual or interesting associations that occur. This analysis can be repeated on a smaller and more specific scale, where first-neighbour networks of nodes of interest can be produced. Analysis of these networks can result in more specific insights into the localised properties of the candidate nodes. In the current investigation both approaches were combined to produce more comprehensive analyses.

5.4 SubtilNet2 summary

The decision to develop SubtilNet2, a new functional interaction network, rather than modifying the existing SubtilNet network was the result of many mitigating factors, the primary one being its age. Since being constructed in 2005, a number of new and appropriate data sets for this investigation have been made publicly available, allowing the construction of a more accurate and comprehensive network, on which to conduct analyses. This new network would be analysed using the same approaches used in the analysis of the original SubtilNet network, and applied to investigate both the predicted e-Therapeutic candidates, with respect to their likely functional associations and basis for the experimental results seen. Together with this investigation will also be performed a SubtilNet2 investigation of the arginine and pyrimidine biosynthetic systems, to try to determine any functional associations that may occur between the two systems and the significance that this may have. This investigation follows on from experimental analysis attempts.

Analysis in both investigations will initially be conducted using a wide ranging non targeted approach, analysing the entire SubtilNet2 network, and looking for general properties and features within it. This will be followed by a more focused approach that will analyse specific nodes in detail, the systems to which they belong and their likely

functional associations. By combining both these approaches, a more comprehensive understanding of the nodes and systems being analysed can be made.

Chapter 6
SubtilNet2 Application

6. SubtilNet2 Application

This chapter initially describes the generation and exploration of the SubtilNet2 network, using three randomly chosen nodes from different biological systems selected from SubtilNet2. These were subjected to various network analysis techniques and plug-in tools, with the aim of identifying functional associations with other nodes and characteristics that could potentially represent underlying properties and biological systems.

Following the analysis of the three random nodes, combinations of candidate genes supplied by e-Therapeutics that were predicted to act as potential therapeutic candidates when targeted simultaneously, were investigated using SubtilNet2 and network analysis. Potential, possible and plausible functional association occurring between each candidate of each pair were suggested, analysed and reconstructed, determining their likely validity.

The laboratory results obtained for each candidate pair were also investigated using SubtilNet2 and network analysis techniques- identifying potential associating nodes and likely systems to which they belong, and how these could potentially account for the laboratory result obtained.

SubtilNet2 was used in an attempt to analyse the investigation of the organisation and functioning of the arginine and pyrimidine biosynthetic systems to determine if carbamoyl phosphate produced by each system is directed towards specific reactions associated with each system or is used more generally in multiple systems irrespective of the producing system.

6.1 The generated SubtilNet2 network for *B. subtilis*

SubtilNet2 was generated from KEGG pathways, PubMed co-citation, DBTBS, and the MPIDB databases and displayed within the Cytoscape platform (Figure 6.1A). The resulting network was subjected to statistical analysis using the Network analyzer plug-in. The network was determined to have 2466 nodes, with 45,124 edges, connected in a degree distribution profile that suggested a scale free topology (Figure 6.1B) typically expected for a biological network. The network was found to have a diameter of 8 nodes, and a characteristic path length of 3.4 edges (Figure 6.1C), with each node

having an average of 34.6 degrees. Cluster analysis using the default settings of MCODE identified a total of 99 clusters within the entire SubtilNet2 network.

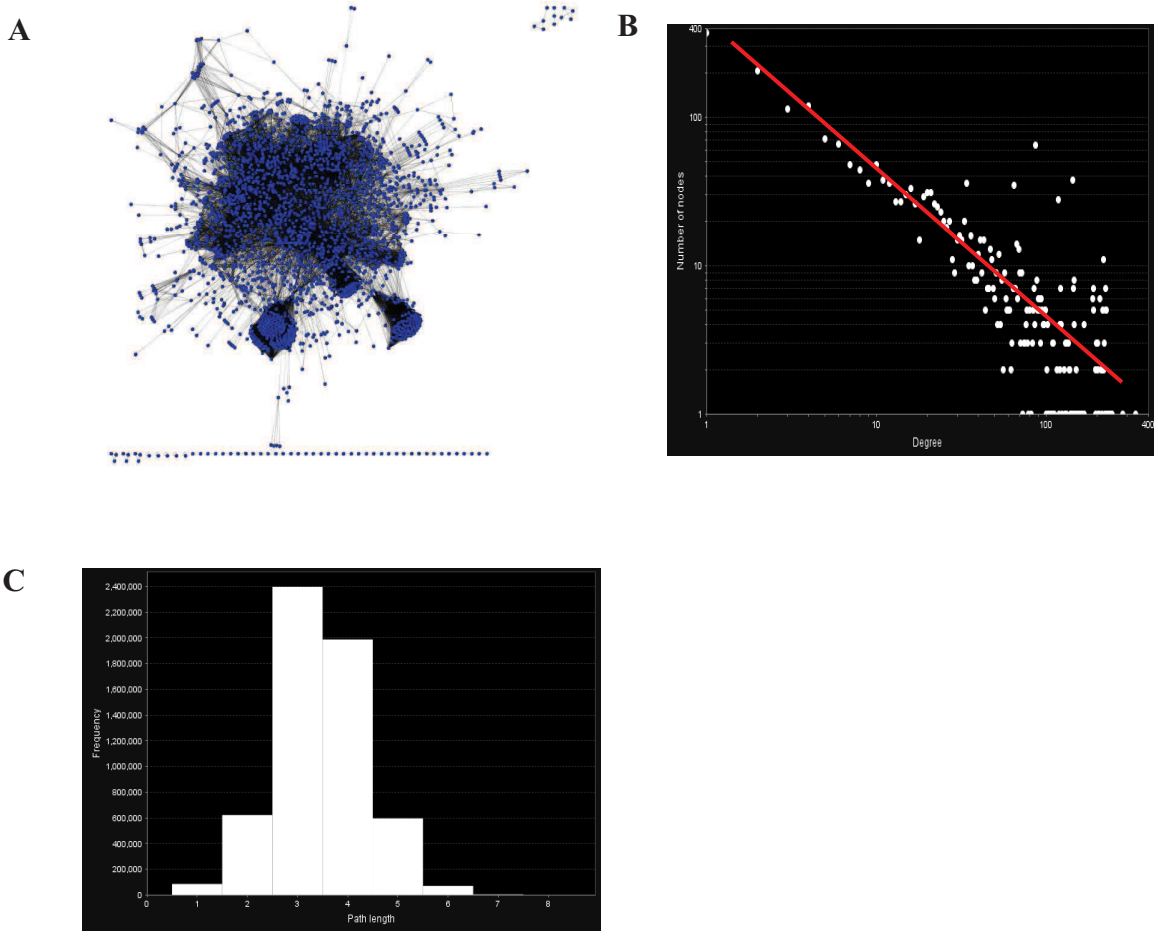


Figure 6.1: SubtilNet2 statistics. A, The generated SubtilNet2 network, visualised using Cytoscape. B, A graph to indicate the degree distribution profile for the generated network, with a line of best fit plotted (red). C, A Graph indicating the characteristic path lengths within the network and the frequencies at which they appear.

6.2 SubtilNet2 exploration

The SubtiliNet2 functional association network together with Cytoscape plug-in tools were explored prior to the adoption of SubtilNet2 for the analysis of e-Therapeutics predicted candidates and a detailed investigation of the arginine and pyrimidine biosynthetic systems.

Cluster analysis was performed on SubtilNet2 and three random nodes were selected. From these nodes, associating clusters and first neighbour networks were identified produced and analysed using the Cytoscape plug-ins- Network analyzer and BiNGO.

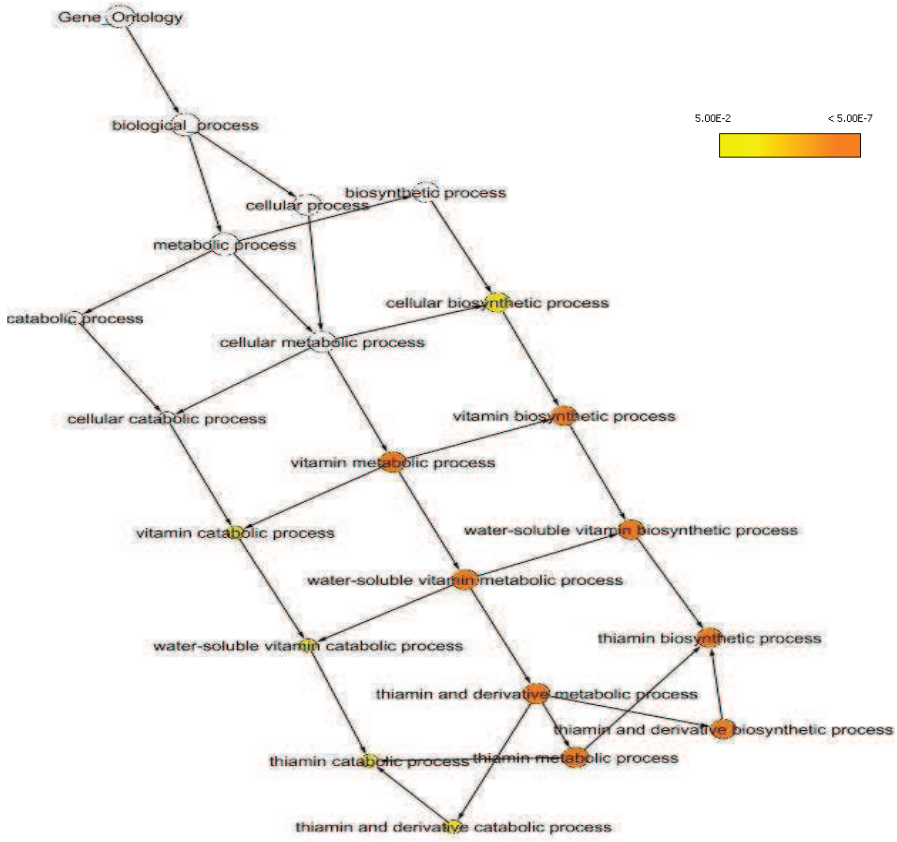
6.2.1 Example 1: Thiamine metabolic pathway

The node *thiE* encoding a thiamine-phosphate pyrophosphorylase, involved in thiamine metabolism (Kanehisa & Goto, 2000; Lechat *et al.*, 2008) was randomly selected from SubtilNet2. Cluster analysis, performed on the entire network found *thiE* to be present within a single fully connected (co-efficient 1.0) cluster, containing an additional 10 nodes. Originating predominantly from KEGG pathways data, all 10 nodes were found to be functionally associated with the thiamine metabolism pathway (Kanehisa & Goto, 2000). BiNGO analysis of this cluster (Figure 6.2) revealed an over representation of the term thiamine and derivative metabolic process (p-value $8.4228E^{-16}$) with 7 nodes being associated with the term out of a possible 12 nodes present within the entire SubtilNet2 network.

Analysis of the generated first neighbour network of *thiE* identified 12 nodes. BiNGO analysis identified 8 of the 12 nodes as being associated with the ontology term for thiamine and derivative metabolic process (p-value $7.0329E^{-18}$). When comparisons were made as to the nodes present in the first neighbour network and cluster identified identical nodes as well as a single extra.

The conclusion from these finding is that nodes from the identified cluster and first neighbour network of a randomly chosen node, have functional associations and over representation of gene ontologies that can both be affiliated to the randomly chosen node and also make biological sense, in this case the thiamine metabolic pathway. These findings can be confirmed, through existing described biological knowledge known about a number of the identified cluster and first neighbour nodes (Kanehisa & Goto, 2000; Lechat *et al.*, 2008).

A



B

Description	p-val	Total freq in network	Number present in cluster	%	Gene
Thiamin and derivative metabolic process	8.42E-16	12	7	58	*thiE,thiG,thiL,thiM,thiC,tenA,thiD*
Thiamin metabolic process	8.42E-16	12	7	58	*thiE,thiG,thiL,thiM,thiC,tenA,thiD*
Thiamin biosynthetic process	1.38E-13	10	6	60	*thiE,thiG,thiL,thiM,thiC,thiD*
Thiamin and derivative biosynthetic process	3.03E-13	11	6	55	*thiE,thiG,thiL,thiM,thiC,thiD*
Water soluble vitamin metabolic process	3.04E-10	58	7	12	*thiE,thiG,thiL,thiM,thiC,tenA,thiD*
Vitamin metabolic process	6.99E-10	65	7	11	*thiE,thiG,thiL,thiM,thiC,tenA,thiD*
Water soluble vitamin biosynthetic process	6.69E-09	47	6	13	*thiE,thiG,thiL,thiM,thiC,thiD*
Vitamin biosynthetic process	1.59E-08	54	6	11	*thiE,thiG,thiL,thiM,thiC,thiD*
Thiamin catabolic process	4.04E-03	1	1	100	*tenA*
Thiamin and derivative catabolic process	4.04E-03	1	1	100	*tenA*
Water soluble vitamin catabolic process	4.04E-03	1	1	100	*tenA*
Vitamin catabolic process	4.04E-03	1	1	100	*tenA*
Cellular biosynthetic process	4.40E-03	463	6	1.3	*thiE,thiG,thiL,thiM,thiC,thiD*

Figure 6.2: BiNGO analysis of the *thiE* cluster. A, A Bingo generated ontology tree for the *thiE* associating cluster, indicating an overrepresentation towards gene ontologies relating to thiamine biosynthetic and related process. The darker the colourisation of the orange label, the statistically more significant the linkage is. B, BiNGO statistical output of generated graphical tree indicating gene presence and frequency in cluster.

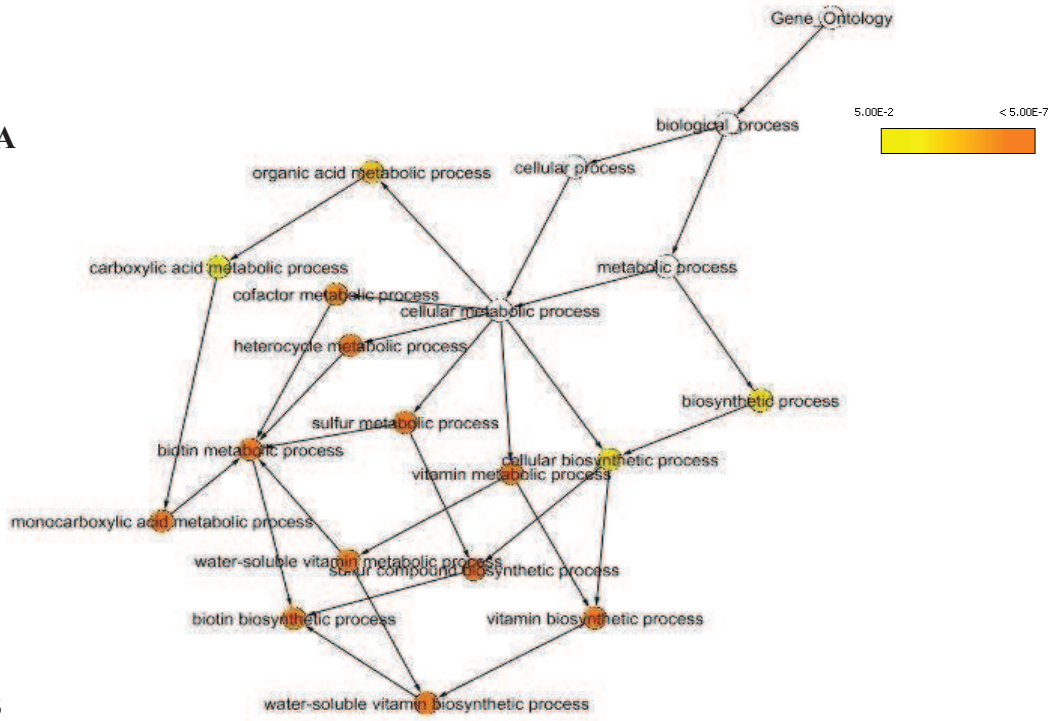
6.2.2 Example 2: Biotin metabolic pathway

The node *bioA* encoding a lysine-8-amino-7-oxononanoate aminotransferase involved in biotin metabolism (Lechat *et al.*, 2008), was randomly selected from the SubtilNet2 network. *bioA* was found to associate with a single cluster containing 6 nodes, 21 edges and which was found to be fully connected (co-efficient 1.0). The origin of these produced functional associations came from the KEGG pathways database and co-citation data. Of the 7 nodes, 6 were found to be associated with biotin metabolism, and represented all 6 nodes known to associate with biotin metabolism. Subsequent BiNGO analysis (Figure 6.3) identified the overrepresentation of 5 nodes within this cluster associated with the term biotin metabolism (p-value $4.6595E^{-14}$).

Generation and analysis of the first neighbour network for *bioA* identified 7 nodes and 28 edges, with 6 of these nodes being associated with the same gene ontology term (p-value 2.2007^{E-17}). Comparing these nodes with those from cluster analysis, all were found to be identical with the extra node identified in the first neighbour network discounted through further investigation.

The findings show that the nodes from cluster and first neighbour network of the randomly selected node *bioA* have both an overrepresentation of the gene ontology as well as functional associations that can also be applied to the randomly selected node. This is very similar to the biologically determined characteristics of *bioA*, and cluster and first neighbour nodes.

A



B

Description	p-val	Total freq in network	Number present in cluster	%	*Gene*
Biotin metabolic process	4.66E-14	6	5	83	*bioA,bioB,bioD,bioW,bioF*
Biotin biosynthetic process	4.66E-14	6	5	83	*bioA,bioB,bioD,bioW,bioF*
Suphur compound biosynthetic process	1.10E-09	30	5	17	*bioA,bioB,bioD,bioW,bioF*
Sulphur metabolic process	5.05E-09	40	5	13	*bioA,bioB,bioD,bioW,bioF*
Water soluble vitamin biosynthetic process	1.17E-08	47	5	11	*bioA,bioB,bioD,bioW,bioF*
Vitamin biosynthetic process	2.42E-08	54	5	9	*bioA,bioB,bioD,bioW,bioF*
Water-soluble vitamin metabolic process	3.50E-08	58	5	8	*bioA,bioB,bioD,bioW,bioF*
Monocarboxylic acid metabolic process	3.82E-08	59	5	8	*bioA,bioB,bioD,bioW,bioF*
Vitamin metabolic process	6.29E-08	65	5	7	*bioA,bioB,bioD,bioW,bioF*
Heterocycle metabolic process	1.71E-07	79	5	6	*bioA,bioB,bioD,bioW,bioF*
Cofactor metabolic process	8.76E-07	109	5	4.5	*bioA,bioB,bioD,bioW,bioF*
Carboxylic acid metabolic process	7.16E-05	263	5	1.9	*bioA,bioB,bioD,bioW,bioF*
Organic acid metabolic process	7.43E-05	265	5	1.8	*bioA,bioB,bioD,bioW,bioF*
Cellular biosynthetic process	1.14E-03	463	5	1	*bioA,bioB,bioD,bioW,bioF*
Biosynthetic process	1.47E-03	837	6	0.7	*bioA,bioB,bioD,bioW,bioF,birA*

Figure 6.3: BiNGO analysis of the *bioA* cluster. A**,** A BiNGO generated ontology tree for the *bioA* associating cluster, indicating an overrepresentation towards gene ontologies relating to biotin biosynthetic process. The significance level of ontology terms increases from yellow to orange, the darker the orange colouration of the label, the statistically more significant the linkage is. B, BiNGO statistical output of the generated graphical tree indicating gene presence and frequency in cluster.

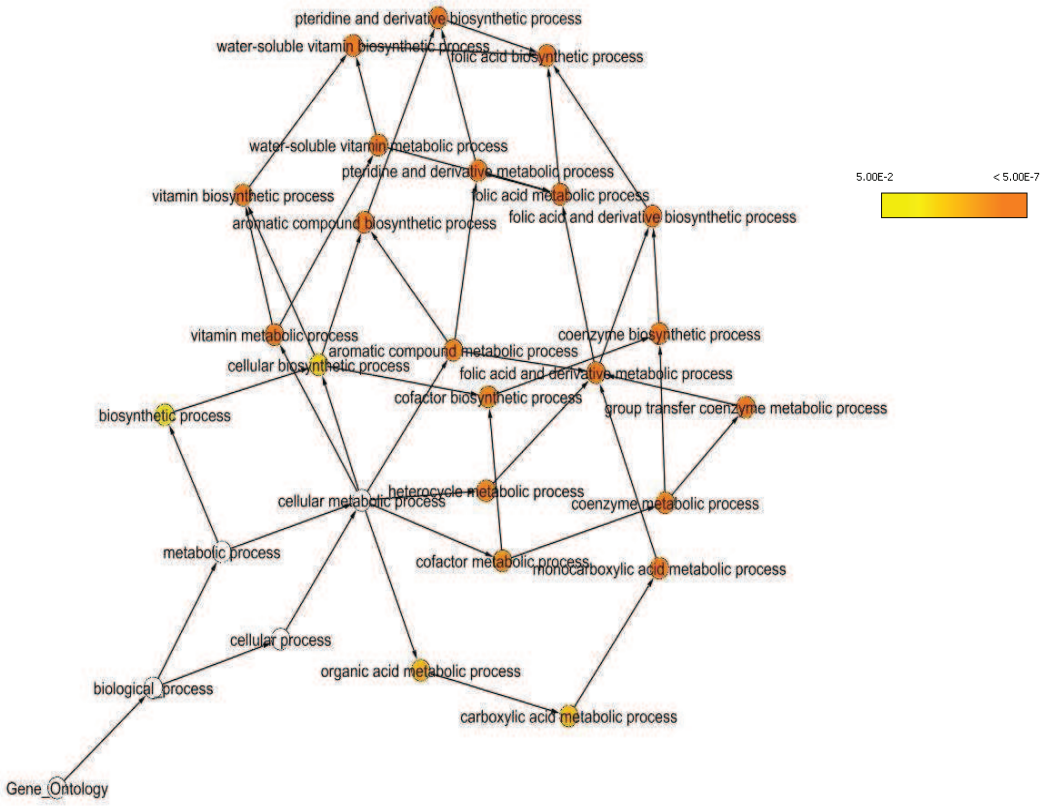
6.2.3 Example 3: Folic acid metabolic pathway

The node *folK* representing a 2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase involved in folic acid biosynthesis (Lechat *et al.*, 2008) was randomly selected from the SubtilNet2 network. The cluster to which *folk* associated contained 7 nodes, and was shown to be fully connected (co-efficient 1.0). Analysis as to the origins of the functional associations identified indicated the exclusive use of the KEGG pathways database where nodes were found to associate with folate biosynthesis. BiNGO analysis of this cluster (Figure 6.4) identified 4 nodes with functional associations to folate biosynthesis (p-value $9.6018E^{-12}$), with a total of 10 nodes with this same ontology being present in SubntilNet2.

First neighbour network analysis of *folK* identified 9 nodes, with an overrepresentation of the ontology term folic acid and derivative biosynthetic process associated with 7 nodes (p-value $3.8408E^{-17}$). Included within theses 7 nodes were 3 not found within the original cluster.

These finding confirm the findings of both the first and second randomly picked node examples studied.

A



B

Description	p-val	Total freq in network	Number present in cluster	%	Gene
Folic acid biosynthetic process	9.60E-12	6	4	66.6	folB,folK,pabC,pabB
Folic acid metabolic process	9.60E-12	6	4	66.6	folB,folK,pabC,pabB
Folic acid and derivative biosynthetic process	1.34E-10	10	4	40	folB,folK,pabC,pabB
Folic acid and derivative metabolic process	1.34E-10	10	4	40	folB,folK,pabC,pabB
Pteridine and derivative metabolic process	1.17E-09	16	4	25	folB,folK,pabC,pabB
Pterdine and derivative biosynthetic process	1.17E-09	16	4	25	folB,folK,pabC,pabB
Group transfer coenzyme metabolic process	3.10E-09	20	4	20	folB,folK,pabC,pabB
Aromatic compound biosynthetic process	4.23E-08	37	4	10.8	folB,folK,pabC,pabB
Water-soluble vitamin biosynthetic process	1.14E-07	47	4	8.5	folB,folK,pabC,pabB
Coenzyme biosynthetic process	1.36E-07	49	4	8.2	folB,folK,pabC,pabB
Vitamin biosynthetic process	2.02E-07	54	4	7.4	folB,folK,pabC,pabB
Water-soluble vitamin metabolic process	2.72E-07	58	4	6.8	folB,folK,pabC,pabB
Monocarboxylic acid metabolic process	2.91E-07	59	4	6.7	folB,folK,pabC,pabB
Vitamin metabolic process	4.33E-07	65	4	6.1	folB,folK,pabC,pabB
Coenzyme metabolic process	6.22E-07	71	4	5.6	folB,folK,pabC,pabB
Aromatic compound metabolic process	7.37E-07	74	4	5.4	folB,folK,pabC,pabB
Cofactor biosynthetic process	8.21E-07	76	4	5.2	folB,folK,pabC,pabB
Heterocycle metabolic process	9.62E-07	79	4	5	folB,folK,pabC,pabB
Cofactor metabolic cycle	3.56E-06	109	4	3.6	folB,folK,pabC,pabB
Carboxylic acid metabolic process	1.25E-04	263	4	1.5	folB,folK,pabC,pabB
Organic acid metabolic process	1.29E-04	265	4	1.5	folB,folK,pabC,pabB
Cellular biosynthetic process	1.21E-03	463	4	0.9	folB,folK,pabC,pabB
Biosynthetic process	1.30E-02	837	4	0.5	folB,folK,pabC,pabB

Figure 6.4: BiNGO analysis of the *folK* cluster. A, A BiNGO generated ontology tree for the *folK* associating cluster, indicating an overrepresentation towards gene ontologies relating to folic acid metabolic processes. The significance level of respective ontologies increases from yellow to orange, the darker the orange colourisation of the label, the statistically more significant the linkage is. B, BiNGO statistical output of the generated graphical tree indicating gene presence and frequency in cluster.

6.2.4 SubtilNet2 exploration summary

The generation and visualisation of the SubtilNet2 model using the selected datasets and the Cytoscape platform has produced a network from which to explore both the predicted e-Therapeutics therapeutic candidates and the arginine and pyrimidine biosynthetic systems.

In preparedness for this task the network and the analysis plug-in tools that were to be used were explored. Three random nodes were selected and investigated in the following ways:

- Cluster association and to what degree of connectedness.
- The generation of first neighbour networks.
- The functional associations present between nodes of clusters and first neighbour networks.
- Potential overrepresentation of gene ontologies in clusters and first neighbour networks.

By gathering this data, indications of any underlying functional associations that the random node may have could potentially be identified. Clusters and first neighbour networks that have been analysed in this way, have been found to identify functional associations that make biological sense, when compared to existing biological knowledge of the nodes contained within them.

6.3 SubtilNet2 application to e-Therapeutics candidate analysis

The six candidate gene pairs supplied by e-Therapeutics were analysed using SubtilNet2 in an attempt to ascertain potential functional associations between each target in the pair, as well as to aid in explain the experimental results seen during testing.

6.3.1 *ywdH- ybfS* interaction

6.3.1.1 *ywdH* background

The gene *ywdH* is thought to encode a putative broad spectrum aldehyde dehydrogenase (ALDH) enzyme that is used in the oxidisation of aldehydes in multiple biological pathways (Kanehisa & Goto, 2000). ALDH are present in all living organisms, and are part of a large superfamily of enzymes catalysing the oxidation of aldehydes (Sophos *et al.*, 2001). Some ALDHs are specific for individual substrates while others have broad

substrate specificity, with all requiring the cofactors NAD or NADP (Perozich *et al.*, 1999). *B. subtilis,* to-date, has a recorded 11 ALDH (Sophos & Vasiliou, 2003) that show a broad spectrum in their specificity (Sophos *et al.*, 2001).

6.3.1.2 *ybfS* background

The gene *ybfS* encodes a glucosamine specific enzyme IICBA, of the phosphotransferase system (PTS), a complex of proteins allowing the transport, phosphorylation, and sensing of various carbohydrates (Postma *et al.*, 1993). *ybfS* also has a secondary function in metabolic and transcriptional regulation of a variety of processes (Saier, 2001). Several variations of the PTS mechanism exist, dependant on substrate specificity. The system of which *ybfS* is a component is suggested to take up glucosamine from the environment (Reizer *et al.*, 1999).

The PTS consists of three domains: IIA, IIB, and IIC (Figure 6.5). The first domain, IIA, forms a cytoplasmic protein, while the second domain, IIB, forms a hydrophilic domain that is fused to IIC, a membrane-spanning hydrophobic domain. The combination of the two domains produces a channel through which extracellular substrates can pass into the interior of the cell. Powering this system is the transfer of a phosphate group provided by the dephosphorylation of phosphoenol pyruvate in a stepwise manner through the components of the PTS system. This begins with the enzyme I (EI), followed by the phosphocarrier protein (HPr), cytoplasmic protein IIA, IIB domain and finally the membrane component IIC. IIC complexes the phosphate group with glucosamine to produce a concentration gradient between the inside and outside of the cell, which draws glucosamine into the channel.

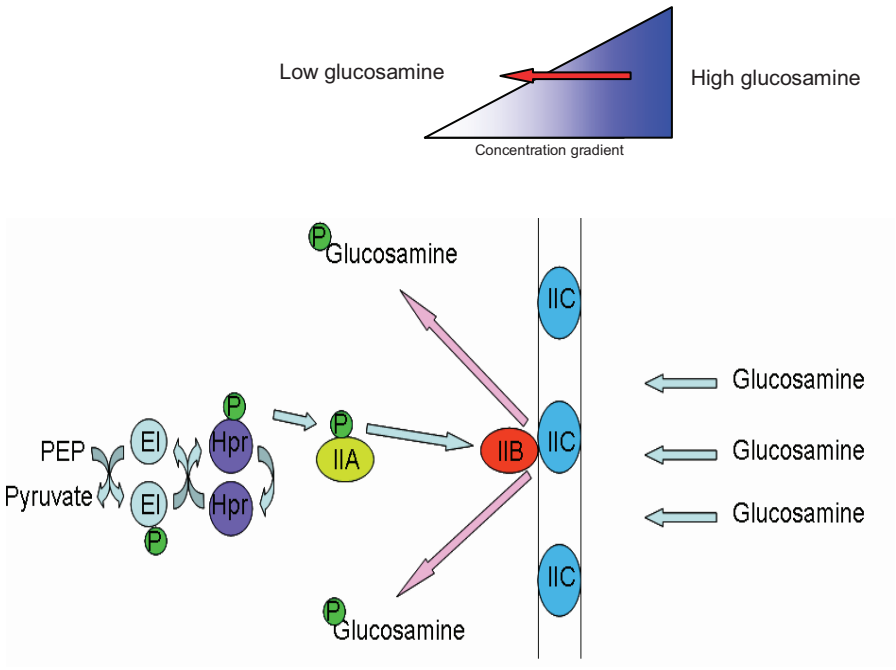


Figure 6.5: The IICBA components of the phosphotransferase system (PTS). The PTS system facilitates the diffusion of glucosamine into the cell. IIA, a cytoplasmic protein. IIB, contains a hydrophilic domain that fuses with IIC, a membrane spanning protein, forming a hydrophilic channel through which extracellular substrates can pass. Phosphate groups donated by phosphoenol pyruvate and transported by EI and HPr power the system.

6.3.1.3 Network analysis of *ywdH-ybfS* using SubtilNet2

The nodes *ywdH* and *ybfS* were identified in the SubtilNet2 network, after which cluster analysis was performed. No single cluster was identified in which both nodes were present.

Table 6.1: Network analysis properties obtained for cluster, sub-cluster and first neighbour network of the *ywdH* node.

***ywdH* node**	**Cluster**	**Sub-cluster**	**First neighbour network**
MCODE cluster ranking	6/99	2/2	N/A
Functional association data source	KEGG Co-citation	KEGG	KEGG Co-citation
Clustering Co-efficient	0.944	0.910	N/A

The node *ywdH* was identified to a highly connected cluster (Table 6.1) containing 58 nodes. This cluster contained an overrepresentation of functionally associated nodes of the valine, leucine and isoleucine degradation pathway (13 nodes), as well as the glycolysis/ gluconeogenesis pathways (10 nodes). The analysis of gene over-representation within this cluster, identified 32 nodes with the oxidation reduction process term (p-value $1.0620E^{-18}$).

Further cluster analysis of the identified cluster, associated the *ywdH* node to second less highly connected cluster of 36 nodes (Figure 6.6). This cluster had functional associations to the valine, leucine and isoleucine degredation pathways (11 nodes) and an overrepresentation of 19 nodes with ontology terms associated with oxidation and reduction reactions (p-value $4.9721E^{-11}$). The existing biological data known about the nodes contained within these clusters supported these findings.

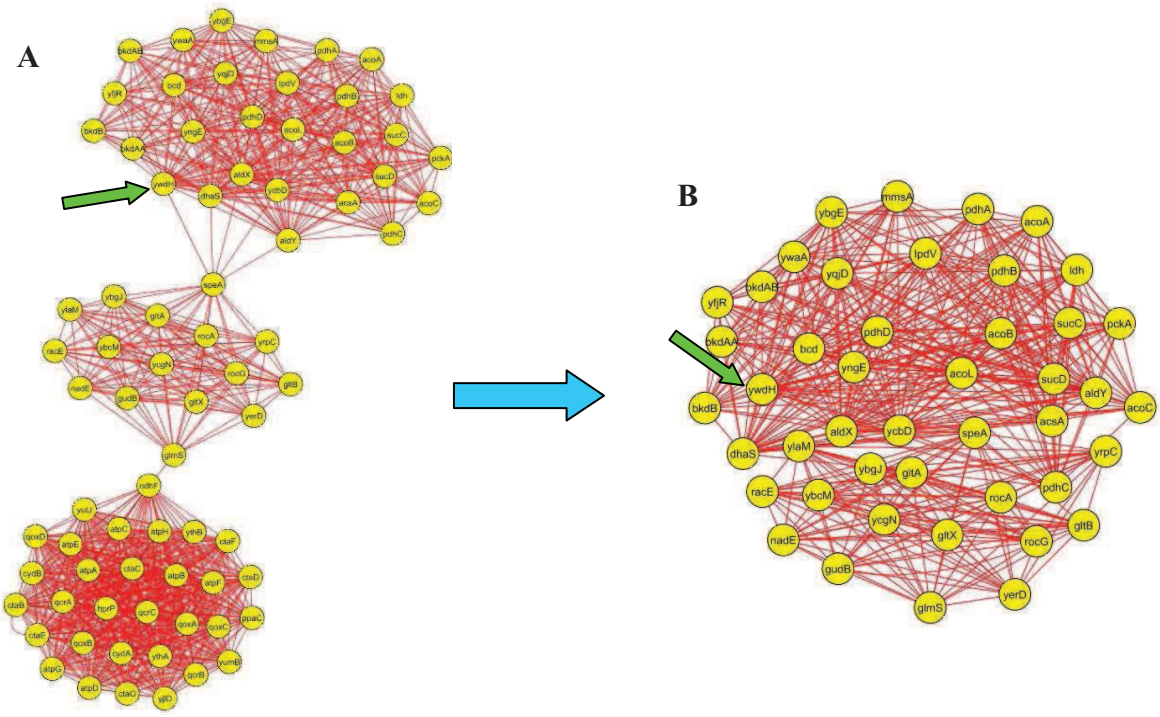


Figure 6.6: Clustering of the *ywdH* node. A, The identified *ywdH* cluster within the whole SubtilNet2 network. B, The sub-cluster to which the *ywdH* node has been identified to. The green arrow indicates the presence of the *ywdH* node within both clusters.

Table 6.2: Network analysis properties obtained for cluster, sub-cluster and first neighbour network of the *ybfS* node.

***ybfs* node**	**Cluster**	**Sub-cluster**	**First neighbour network**
MCODE cluster ranking	7/99	3/7	N/A
Functional association data source	KEGG	KEGG	KEGG
Clustering Co-efficient	0.907	0.985	N/A

The second half to the selected target pair *ybfS (gamp),* associates with a large, 135 node highly connected cluster (Table 6.2) with a significant proportion of nodes having functional associations to the amino acid metabolism pathways, of cysteine and methionine (18 nodes) and, arginine and proline (16 nodes) as well as the phosphotransferase pathway.

Analysis of the representation of gene ontologies within this cluster found that 36 nodes were represented with the term amino acid and derivative metabolic process (p-value $3.8538E^{-12}$). Further cluster analysis, reducing the size of the initial cluster to a highly connected cluster containing 42 nodes (Figure 6.7), was shown to as having some functional association with the phosphotransferase systems (18 nodes). Ontology analysis of this cluster revealed the 18 nodes had an overrepresentation of the term phosphoenol pyruvate-dependant sugar phosphotransferase system p-value ($8.1498E^{-26}$). The existing biological data known about the nodes contained within these clusters again supported these findings.

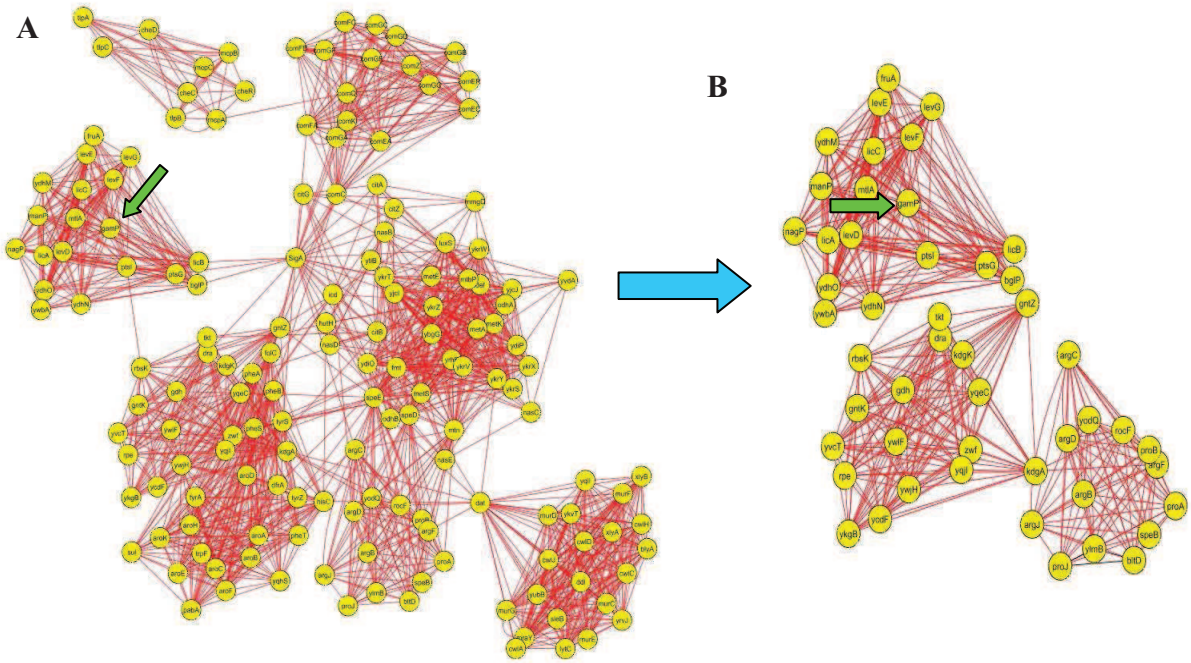


Figure 6.7: Clustering of the *ybfS* node. A, The identified *ybfS* cluster within the whole SubtilNet2 network. B, The sub-cluster to which the *ybfS* node has been identified to. The green arrow indicates the presence of the *ybfS* node within both clusters.

Potential functional associations between the pair of candidate nodes was determined by analysing the nodes within each cluster for any co-association and then looking at the first neighbour network of each target, and doing the same. This identified *ybfS* to be present within the network of *ywdH*, suggesting a close functional association. The node *ypqE*, encoding a glucose specific phosphotransferase system IIA component that allows the uptake of glucose, maltose, N-acetyl-muramic acid, trehalose and arbutin (Kanehisa & Goto, 2000), was also found to be co-associated with both networks. Shortest path analysis, using the first neighbour networks of *ywdH* and *ybfS* nodes, found a path with a length of two (smaller than the average calculated for the SubtilNet2 network (3.4)) connecting the two nodes via the additional nodes *malP* or *yyzE* or *ypqE* (Figure 6.8A).

Further manual validation of the intermediate nodes revealed an interchangeable use of the node names *ypqE*, and *yyzE*. This was also observed for the node name *malP* which was being used interchangeably with *ybfS*. After compiling these node synonyms, the shortest path between the *ywdH* and *ybfS* node was found only to be via *ypqE*. (Figure 6.8B).

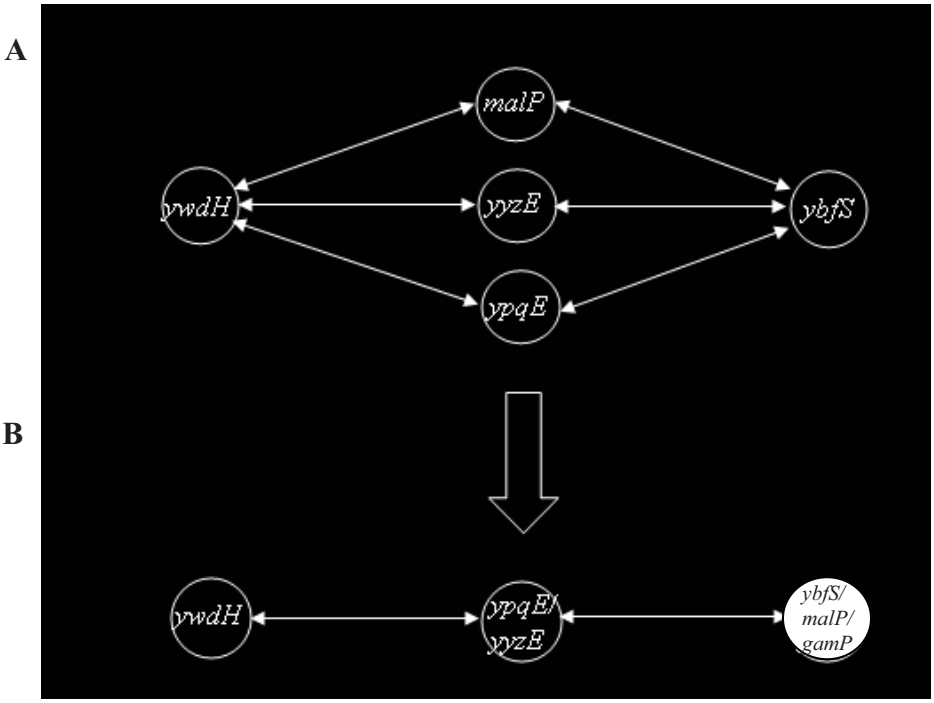


Figure 6.8: A graphical representation of the SubtilNet2 determined shortest paths between *ywdH* and *ybfS* nodes. A, Shortest path before manual verification. B, Shortest path after manual verification.

Both the *ywdH* and *ybfS* nodes can be associated with different pathways of a multi stage amino sugar metabolism process (Kanehisa & Goto, 2000), *ybfS* being involved in the sensing and uptake of glutamine from the environment, and *ywdH* with the production of acetate and acetaldehyde from acetyl-CoA. Visualising these stages (Figure 6.9), together with knowledge of the only identified shortest path intermediate *ypqE*, made it apparent that the fructose metabolism pathway could be used to functionally associate the two stages together in a biologically plausible way. Returning to the first neighbour networks of the *ywdH* and *ybfS* nodes, identified a number of nodes found to be functionally associated with the fructose metabolism pathway, potentially providing further evidence to support this conclusion.

The functional association between *ywdH* and *ybfS* nodes could proceeds as follows using the shortest path intermediate *ypqE*. Glucosamine taken up by the cell using *ybfS*, is converted to D-glucosamine-6-P in the early stages of the multi stage amino sugar metabolism pathway. While this is occurring, N-acetyl-muramic acid is up taken by *ypqE* and also converted to D-glucosamin-6-P in addition to glucose being up taken. Both D-glucosamine-6-P and glucose enter the fructose metabolism pathway where they are converted to D-fructose-6-P and eventually glyceraldehyde-3-P, which is passed onto the latter stages of the amino sugar metabolism pathway where the *ywdH* encoded aldehyde dehydrogenase functions.

6.3.1.4 Conclusions of laboratory and SubtilNet2 based analysis of *ywdH-ybfS* targets

The insights obtained using SubtilNet2 into the potential functional associations between the *ywdH* and *ybfS* nodes may provide explanations as to why no discernable differences were detected in the growth of mutants removed of these nodes functions and subjected to different stresses. When the function of the *ybfS* node is removed, mutants are simply able to uptake glucosamine using an alternative uptake system. It has been noted that *E. coli,* which shares a PTS system specifically for the uptake of glucosamine (Tchieu *et al.*, 2001) like *B. subtilis,* can also utilize the mannose PTS system when required (Tchieu *et al.*, 2001). This could be occurring in this investigation. Alternatively, *B. subtilis* has the ability to produce D-glucosamine-6-P, the product of *ybfS,* by up taking and converting N-acetyl muramic acid and fructose respectively, both systems utilising distinct uptake pathways. The removal of the *ywdH*

node from the system may not have had an affect due to the presence of 11 paralogues known to exist within the cell, that could function in replacement of *ywdH*.

SubtilNet2 has identified potential functional interactions of both target nodes to associated pathways of nodes belonging to the fructose metabolism pathway that could provide the functional link between the *ybfS* and *ywdH* nodes. This analysis, together with literature has provided several plausible explanations for the results obtained in experimental testing of mutants removed of these node functions.

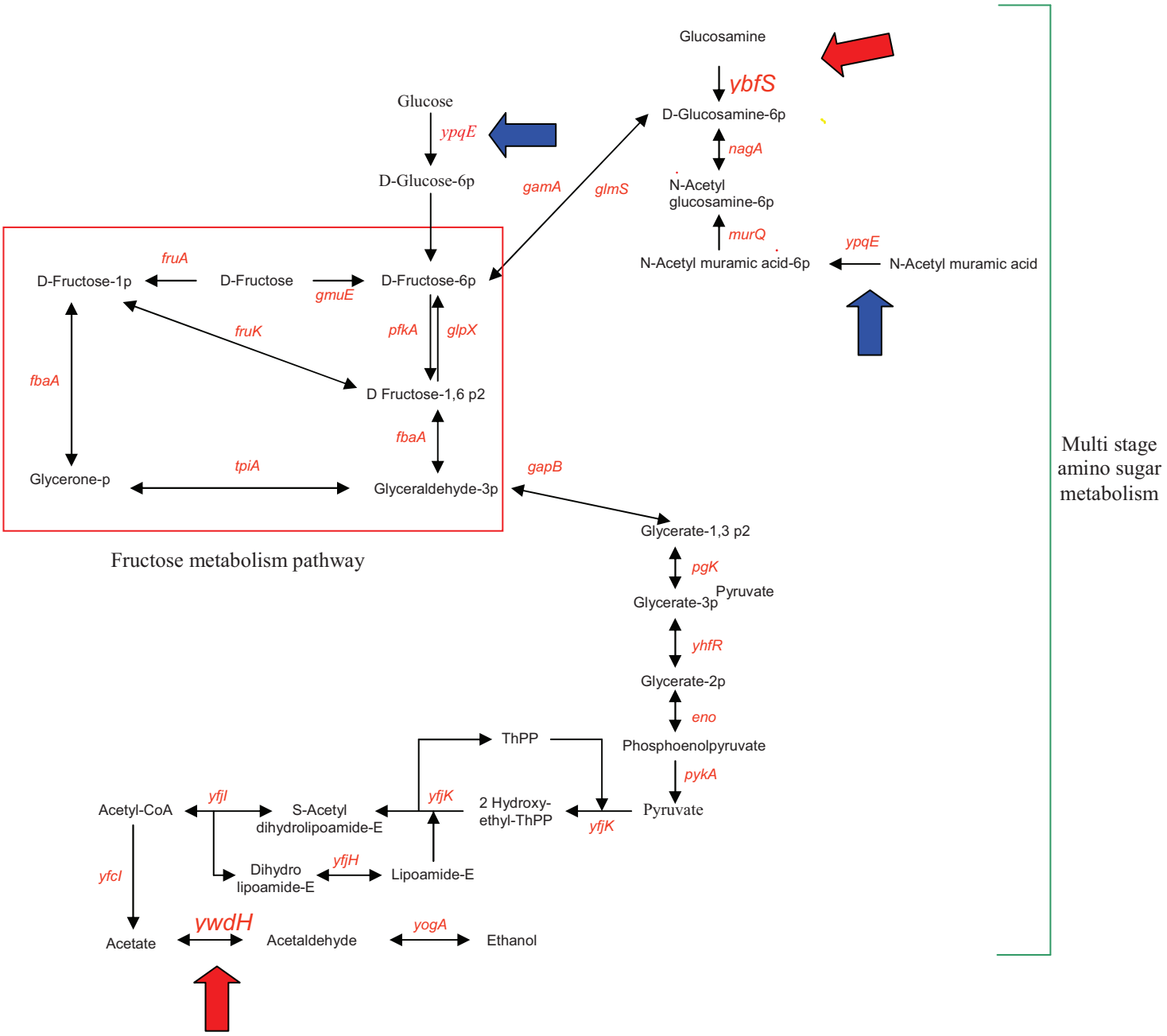


Figure 6.9: A schematic diagram of the candidate nodes *ybfS* and *ywdH* and the pathways that could functionally associate them. Generated using the KEGG pathways database, the above schematic diagram describes the possible functional association of the *ywdH* and *ybfS* nodes via the fructose metabolism pathway. Targets highlighted with red arrows indicates candidate nodes *ywdH* and *ybfS*. Blue arrows indicate the nodes predicted to be within the shortest path.

6.3.2 *yvgQ-luxS* interaction

6.3.2.1 *yvgQ* background

The gene *yvgQ* encodes a sulphite reductase (The universal protein resource (UniProt) 2009; Lechat *et al.*, 2008; van der Ploeg *et al.*, 2001) that is used in the conversion of sulphur into the amino acids cysteine, methionine and *S*-adenosylmethionine (AdoMet). In *B. subtilis*, sulphur, a vital element for the synthesis of proteins and cofactors is acquired from both organic and inorganic sources.

Inorganic sulphate ions are taken up by the cell and activated using the enzymes sulphate permease (*cysP*) and ATP sulfurylase (*sat*) before being reduced using the enzymes APS kinase (*cysG*) and phosphoadenosine phosphosulphate (PAPS) reductase (*cysH*) to produce sulphite (Albanesi *et al.*, 2005). Organic sulphur containing compounds are taken up and converted to sulphite by the sulphonate uptake and degradation system *ssuBACD* (Albanesi *et al.*, 2005).

Sulphite is reduced to sulphide by the action of sulphite reductase, (*yvgQ)*, before O-acetylserine (OAS) enters the system and with the action of O-acetylserinethiol lyase (*cysK*) and cysteine synthase (*yrhA*) to catalyze the conversion of sulphides to L-cysteine (van der Ploeg *et al.*, 2001).

L-cysteine can be converted to methionine by the reversible transsulfuration pathway (Hullo *et al.*, 2007). The first step converts L-cysteine to L-cystathionine using cystathionine synthase (*metI*), which is then converted to homocysteine using cystathionine lyases, (*metC*) and (*patB*)(Auger *et al.*, 2002; Auger *et al.*, 2005; Hullo *et al.*, 2007), before a final methylation reaction to methionine by methionine synthase (*metE*). AdoMet, a methyl donor to numerous reactions (Chiang *et al.*, 1996) and precursor to polyamine synthesis, can then be synthesised from methionine by an AdoMet synthase (*metK*)(Hullo *et al.*, 2007).

6.3.2.2 *luxS* background

The gene *luxS* encodes an *S*-ribosylhomocysteine hydrolase that converts *S*-ribosylhomocysteine (SRH) to homocysteine, as well as the auto-inducer 2 (AI2) used in cell to cell communication (Winzer *et al.*, 2002).

B. subtilis has the ability to catabolise methionine, using it as a sulphur source if it encounters such limitation within its environment. To enable this to happen, methionine

is converted to S-adenosyl-L-methionine (AdoMet) via one of two separate pathways (Hullo *et al.*, 2007). The first reproduces methionine and ultimately more AdoMet by the initial degradation of AdoMet to methylthioadenosine (MTA) by the action of spermidine synthase *(sped)* and AdoMet decarboxylase (*speE)*. This is then degraded to adenine and methylthioribose (MTR) by AdoHey/MTA nucleosidase, (*mtnN)*. MTR is phosphorylated by methylthioribose kinase (*mtnK)*, to produce MTRP which is converted to keto methylthiobutyric acid (KMBA), and then methionine by an aminotransferase (*mtnE).* This pathway is known as the MTR recycling pathway and can feed a second pathway, the AdoMet recycling pathway (Hullo *et al.*, 2007).

AdoMet is converted to homocysteine in a three step process. The initial step removes the methyl group from AdoMet and converts it to AdoHey. The second step, catalysed by AdoHey/MTA nucleosidase (*mtn),* converts AdoHey to SRH. The final step, catalysed by *S*-ribosylhomocysteine hydrolase (*luxS)*, converts SRH to homocysteine (Hullo *et al.*, 2007).

Homocysteine can then be used to produce cysteine by feeding sulphur into the forward transsulfuration pathway, entering as sulphide using cystathionine lyase and homocysteine lyase (*yrhB*) (Hullo *et al.*, 2007). It can also produce cysteine using the reverse transsulfuration pathway and the enzymes cystathionine synthase, *yrhA*, cystathionine lyase and homocysteine lyase *yhrB* as well as the intermediate cystathionine (Hullo *et al.*, 2007).

6.3.2.3 *yvgQ- luxS* network analysis

The nodes *yvgQ* and *luxS* were identified in SubtilNet2 and cluster analysis determined that they were not found to be present within the same cluster.

Table 6.3: Network analysis properties obtained for cluster, sub-cluster and first neighbour network of the *yvgQ* node

***yvgQ* node**	**Cluster**	**Sub-cluster**	**First neighbour network**
MCODE cluster ranking	21/99	N/A	N/A
Functional association data source	KEGG	N/A	KEGG
Clustering Co-efficient	1.0	N/A	N/A

The node *yvgQ* was found to associate with a highly connected cluster (Table 6.3) containing eight nodes (Figure 6.10), with three nodes showing functional associations with the cysteine and methionine metabolic pathways and three nodes with an association with other sulphur metabolic pathways.

The representation of gene ontologies within the cluster showed 8 nodes with a bias to the term cysteine biosynthetic processes (p-value 7.1278 E^{-20}). Separate biological analysis of these associating cluster nodes confirmed these findings. Attempts to disclose any further sub clusters within the original identified cluster resulted in non being found.

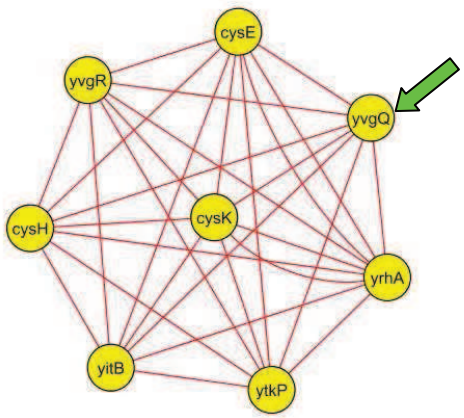


Figure 6.10: Clustering of the *yvgQ* node.

Table 6.4: Network analysis properties obtained for cluster, sub-cluster and first neighbour network of the *luxS* node

***luxS* node**	**Cluster**	**Sub-cluster**	**First neighbour network**
MCODE cluster ranking	7/99	1/7	N/A
Functional association data source	KEGG	N/A	KEGG
Clustering Co-efficient	0.907	1.0	N/A

The node *luxS* was found to associate with a highly connected cluster (Table 6.4) which had 135 nodes with predominant functional associations to metabolism pathways of the amino acids, proline (8 nodes) arginine (7 nodes) and methionine (11 nodes). An overrepresentation of gene ontologies associated with cell communication was found to be present on 47 of the cluster nodes (p-value $1.5944E^{-21}$). Further clustering analysis of the original *luxS* cluster found it to be connected to a second fully connected cluster with 11 nodes and functional associations to cysteine and methionine metabolism pathways (Figure 6.11). This cluster had 5 nodes over represented by the ontology term of amino acid and derivative metabolic process (p-value $8.3622E^{4)}$. These findings provide evidence to existing biological properties known about the nodes contained within the clusters.

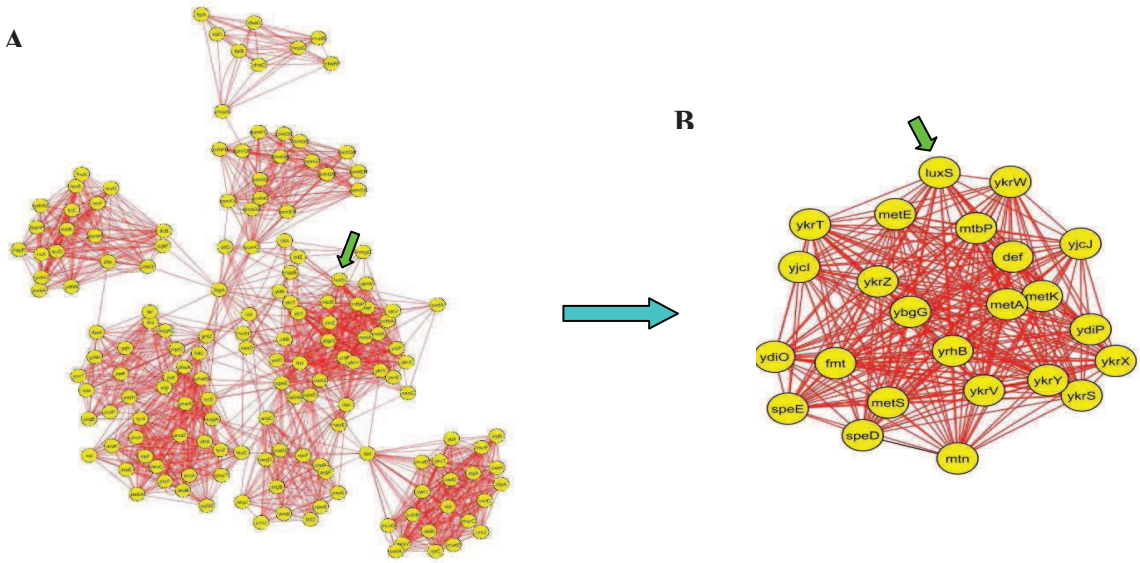


Figure 6.11: Clustering of the *luxS* node. A, The identified cluster of the *luxS* node within the whole SubtilNet2 network. B, The sub-cluster to which the *luxS* node has been identified to. The green arrow indicates the presence of the *luxS* node within both clusters.

The first neighbour networks of both *yvgQ* and *luxS* were compiled and compared in search of co-associating nodes. Three were found, *metA*, *metC* and *yrhB*. On comparing these nodes with those identified in the shortest path analysis between the *yvgQ* and *luxS* nodes, all three were re-identified, together with an additional four nodes *yrhA*, *cysE*, *metI* and *cysK*. (Figure 6.12). These nodes were shown to contribute in producing a shortest path length of two.

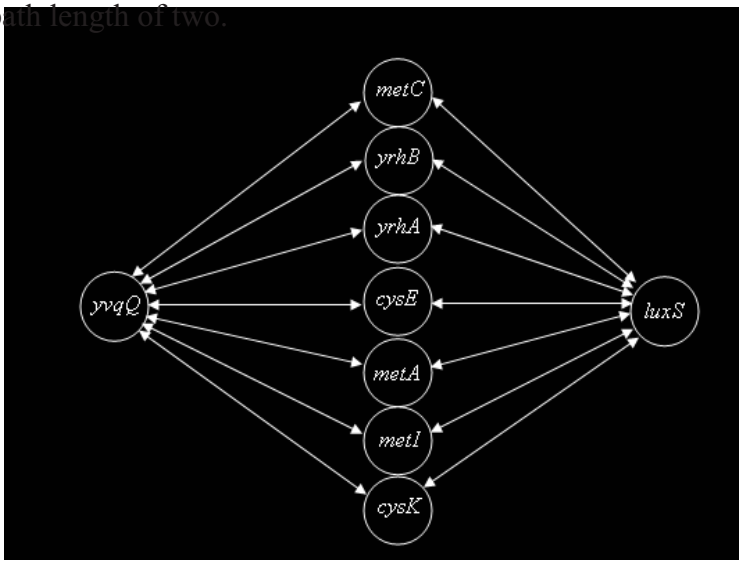


Figure 6.12: A graphical representation of SubtilNet2 determined shortest paths between the *yvgQ* and *luxS* nodes.

The seven identified intermediates nodes present within the shortest pathway, encoded the following:

- *metC*: The enzyme encoded by this gene has both a cystathionine *ß*-lyase, activity-converting L-cystathionine to homocysteine, as well as a cysteine desulfhydrase activity, converting cysteine to sulphide (The universal protein resource (UniProt) 2009; Hullo *et al.*, 2007). This gene can link the activity of *luxS* with that of *yvgQ* in that it processes substrates into products also produced by *luxS* and *yvgQ.*

- *yrhB*: The enzyme encoded by this gene is a cystathionine γ-lyase (Finn *et al.*, 2008), and has homocysteine γ-lyase activity as well as a cysteine desulfhydrase activity in vitro (Hullo *et al.*, 2007). It is a member of the transsulfuration pathway and also the cysteine synthesis pathway, both of which link the products of *yvgQ* and *luxS.*

- *yrhA:* The enzyme represented by this node is a cystathionine β-synthase (The universal protein resource (UniProt) 2009; Hullo *et al.*, 2007), producing cystathionine from homocysteine and involved in the transsulfuration pathway (Hullo *et al.*, 2007). This enzyme has a low OAS -lyase activity in vitro (Hullo *et al.*, 2007), and converts o-acetyl serine into cysteine using sulphides. This gene links the activities of the node *luxS* with *yvgQ*, in that it processes the products produced by each.

- *cysE:* The enzyme represented by this node is a serine O-acetyltransferase (The universal protein resource (UniProt) 2009; Hullo *et al.*, 2007), that transfers the acetyl group from acetyl-CoA to serine, to produce OAS. OAS is combined with the products from the reactions of both *luxS* and *yvgQ* nodes.

- *metA*: The enzyme encoded by this gene is a homoserine O-succinyltransferase that catalyses the production of O-acetyl-L-homoserine from acetyl-CoA and L-homoserine (The universal protein resource (UniProt) 2009; Kunst *et al.*, 1997) . This enzyme does not act upon any direct products or reactions of *luxS* or *yvgQ,* but is involved within the pathway between the conversion of homocysteine and cysteine to which both *luxS* and *yvgQ* contribute.

- *metI:* The enzyme encoded by this gene is a cystathionine β-synthase (Hullo *et al.*, 2007), catalysing the production of cystathionine from cysteine. This

enzyme does not act upon any direct products or reactions of *luxS* or *yvgQ* but is involved within the pathway between the conversion of homocysteine and cysteine to which both *luxS* and *yvgQ* contribute.

- *cysK: The* enzyme encoded by this gene is an OAS lyase that converts OAS into cysteine (Finn *et al.*, 2008). This only affects the products of *yvgQ* and not directly those of *luxS*, but it is a member of a pathway to which both are involved.

All seven of the intermediate nodes identified have been shown to be components of either the cysteine synthesis pathway to which *yvgQ* participates, the AdoMet/ methionine recycling pathway to which *luxS* participates or a transsulfuration pathway that converts cysteine to homocysteine and vice versa, linking the products of both pathways together. This provides evidence of very plausible biological associations that could link the nodes of *yvgQ* and *luxS*.

6.3.2.4 Conclusions of laboratory and SubtilNet2 based analysis of *yvgQ-luxS* targets

The experimental testing of individual and combinations of KO mutants of *yvgQ* and *luxS* genes resulted in no discernable differences in growth and stress response when compared to wild type *B. subtilis* treated in the same way. SubtilNet2 provided potential explanations for these results. Using intermediate nodes, identified in shortest path analysis, the target nodes could be functionally associated to one another in a very plausible biological manner using the transsulfuration pathway (Figure 6.13).

The removal of the function of the *yvgQ* node, prevented the conversion of sulphides to sulphites, and eventually cysteine. However, cysteine can also be assimilated from the environment, with several uptake systems characterised (Burguiere *et al.*, 2004). Cysteine can also be produced from methionine, homocysteine or cystathionine taken from the environment and able to function via the transsulfuration pathway.

In relation to *luxS*, the removal of this nodes function may not have produced phenotypic effects, as the product produced by it, homocysteine can be up-taken from the environment, or supplied from cystathionine via the transsulfuration pathway.

The insights given by SubtilNet2 allowed the construction of a pathway schematic for both *yvgQ* and *luxS* nodes.

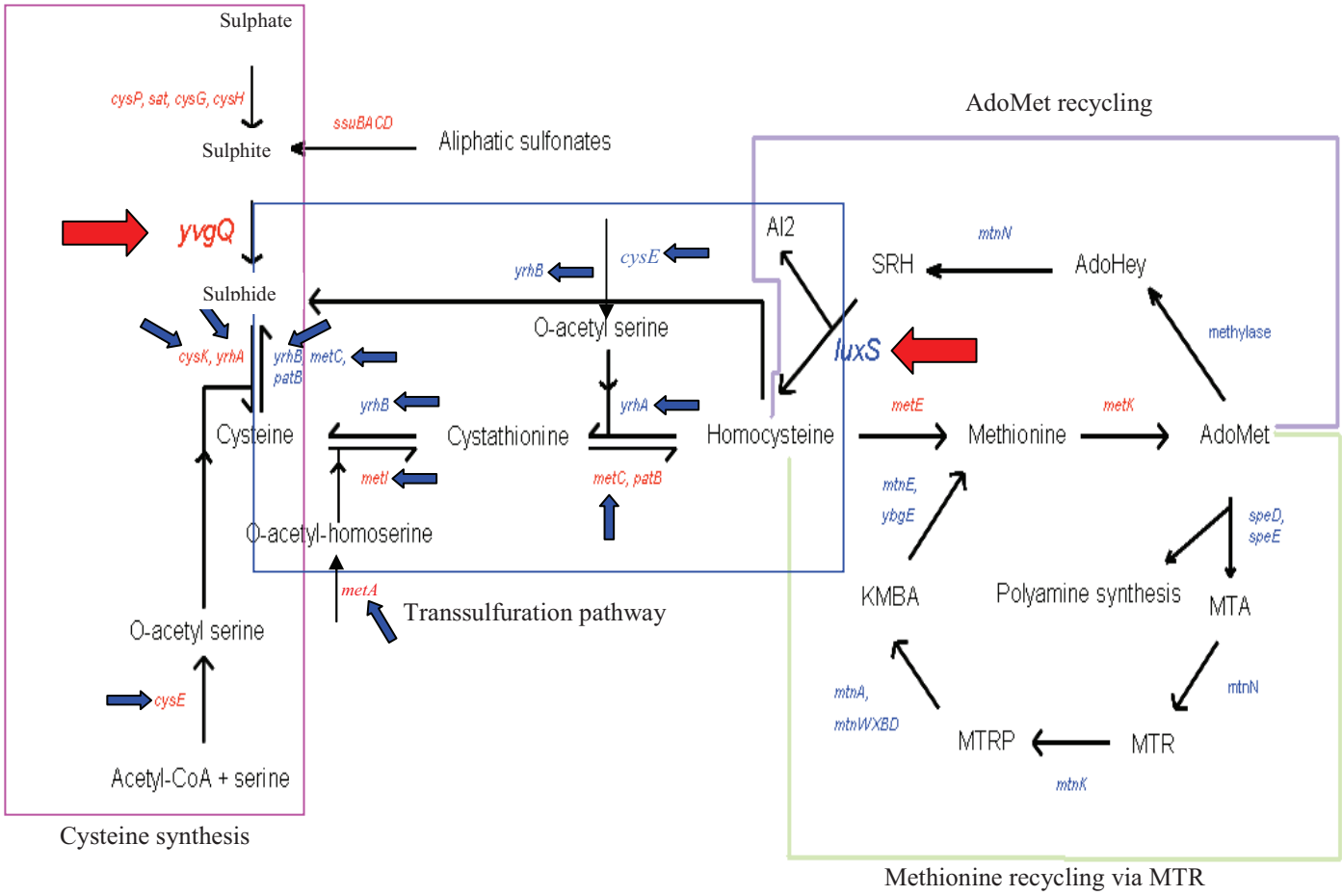


Figure 6.13: A schematic diagram of the candidate nodes *yvgQ* and *luxS* and the pathways that could functionally associate them. Generated using the KEGG pathways database, the above schematic diagram describes the possible functional association of the *yvgQ* and *luxS* nodes via the, transsulfuration pathway. Targets highlighted with red arrows indicates the candidate nodes *yvgQ* and *luxS*. Blue arrows indicate the nodes predicted to be within the shorted path.

6.3.3 *licT–cheB* interaction

6.3.3.1 *licT* background

The gene *licT* encodes a transcription anti-terminator that is used to allow *B. subtilis* to utilise aryl-β-glucosides as a carbon source in conditions of nutrient limitation(Kruger & Hecker, 1995; Tobisch *et al.*, 1997). This transcription anti-terminator binds to

specific sequences within transcribed (m)RNA termed ribonucleic anti-terminator (RAT) sequences (Schnetz *et al.*, 1996), allowing transcription to proceed, beyond a terminator sequence and allow the production of components necessary for a PTS system to uptake aryl-β-glucosides (Schnetz *et al.*, 1996).

When *B. subtilis* is in an environment with preferred carbohydrate sources, the action of *licT* is repressed, a process known as catabolite repression (Gorke & Stulke, 2008; Lindner *et al.*, 2002). This ensures the metabolism of preferred carbon sources first, before switching to alternatives. When in an environment of preferred carbon sources, no binding of anti-terminator to the RAT sequences occurs and the transcription of components that allow the PTS to transport and utilise aryl-β-glycoside is turned off(Schnetz *et al.*, 1996).

The repression of *licT* occurs as two regulatory domains known as phosphotransferase regulated domains (PRD) become phosphorylated by donated phosphates from the components PEP, HPr and enzyme I of the PTS which are taking up preferred carbohydrates (Lindner *et al.*, 2002) (Figure 6.14).

The *licT* gene is found within and regulates the *bglPH* operon (Le Coq *et al.*, 1995) containing *bglS*, which encodes an extracellular β-glucanase (Murphy *et al.*, 1984) , *bglP,* encoding an aryl-β-glucoside specific enzyme IICBA and regulator of *licT* (Lindner *et al.*, 2002), and *bglH* encoding a 6-P-β-glucosidase (Le Coq *et al.*, 1995).

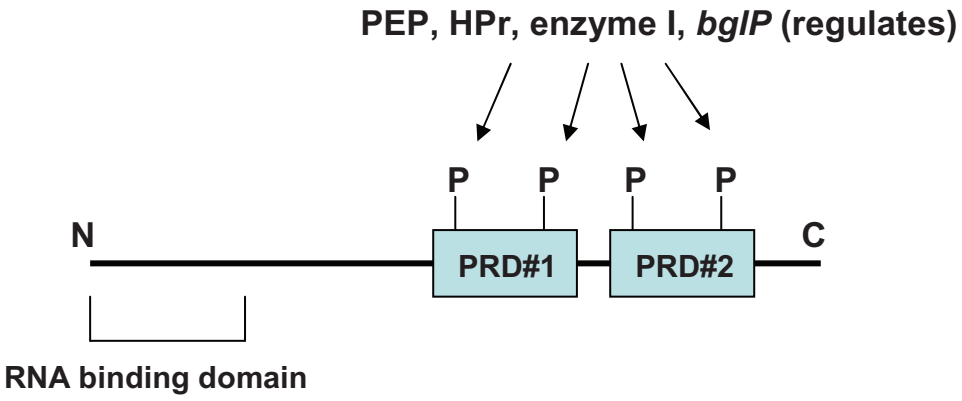


Figure 6.14: A graphical representation of the transcription anti-terminator protein LicT, and its various domains.

6.3.3.2 *cheB* background

Bacteria use chemotaxis to move within an environment towards gradients of attractants and down gradients of repellents (Sierro *et al.*, 2008). In the absence of stimulant *B. subtilis* swim in a tumbling motion, beating its flagella in a clockwise direction and allowing it to sample its immediate vicinity. Upon sensing a stimulant/repellent, the direction of flagella rotation changes to an anticlockwise movement, causing the bacteria to swim in a smooth and focused way towards the stimulant

cheB is encoded within a 26kb *fla/che* operon (Lechat *et al.*, 2008; Lindner *et al.*, 2002), that contains an additional 30 genes encoding chemotaxis, and flagella hook basal body assembly complex components (Finn *et al.*, 2008). The gene encodes a glutamate methylesterase (Kanehisa & Goto, 2000). This is involved in the adaptation response of chemotaxis receptors, to ensue that *B. subtilis* can continue to sense increasing concentrations of stimulant, even when in environments in which they may already be high. This is achieved by *cheB* aiding to reset the chemotaxis receptor (Saulmon *et al.*, 2004; Wadhams & Armitage, 2004).

The mechanism of chemotaxis relies upon a two component signal transduction system, (Fabret *et al.*, 1999). The environment is sensed by transmembrane methyl chemotactic proteins (MCP)(Zimmer *et al.*, 2000) located upon the poles of the bacteria (Gestwicki *et al.*, 2000). In response to triggering of the receptor by stimulant, a conformational change in receptor shape occurs (Saulmon *et al.*, 2004). These changes are detected by an auto-phosphorylating histidine kinase, CheA, connected through coupling proteins CheW and CheV to the receptor. CheA up-regulates its behaviour, phosphorylating a response regulator, CheY, that then acts upon the flagella motor mechanism, to alter the direction of flagella movement (Saulmon *et al.*, 2004). This changes from a default clockwise direction (Bischoff & Ordal, 1992) that causes a tumbling swimming motion allowing bacteria to sense surrounding environments, to an anticlockwise direction, causing smooth and straight swimming and allowing bacteria to move towards or away from stimulant (Figure 6.15).

For *B. subtilis* to continue to sense and respond to changes in stimulant concentrations even while still within the presence of a stimulant, its receptors must be “reset”, reducing the amount of CheA-P and CheY-P to pre-stimulus levels (Saulmon *et al.*,

2004; Wadhams & Armitage, 2004). This process, known as adaptation occurs in *B. subtilis* by three proposed mechanisms (Rao *et al.*, 2008):

- The methylation and demethylation of selected glutamic acid residues (Saulmon *et al.*, 2004) on MCP that has potential effects on CheA activity (Rao *et al.*, 2008).
- The phosphorylation of the coupling protein CheV, suspected to couple reactions between the MCP and CheA (Rosario *et al.*, 1994).
- The application of the phosphatase CheC (Rao *et al.*, 2008) and FliY (Szurmant *et al.*, 2003) to reduce the levels of CheA-P and CheY-P.

CheB participates in the demethylation component of the adaptation response (Rao *et al.*, 2008), after its first activated by CheA-P. It rapidly (Rao *et al.*, 2008) demethylates selected glutamate residues on MCP (Saulmon *et al.*, 2004), CheD a receptor deamidase (Kristich & Ordal, 2002) contributes. A suspected complex pattern of shuffling and addition of methyl group to the receptor then occurs using the methyltransferase, CheR, resetting the receptor (Rao *et al.*, 2008). During this process CheD also deamidates glutamine residues on the MCP (Kristich & Ordal, 2002), enabling specific MCP's to begin functioning and others to sense attractants at specific concentrations (Kirsch *et al.*, 1993).

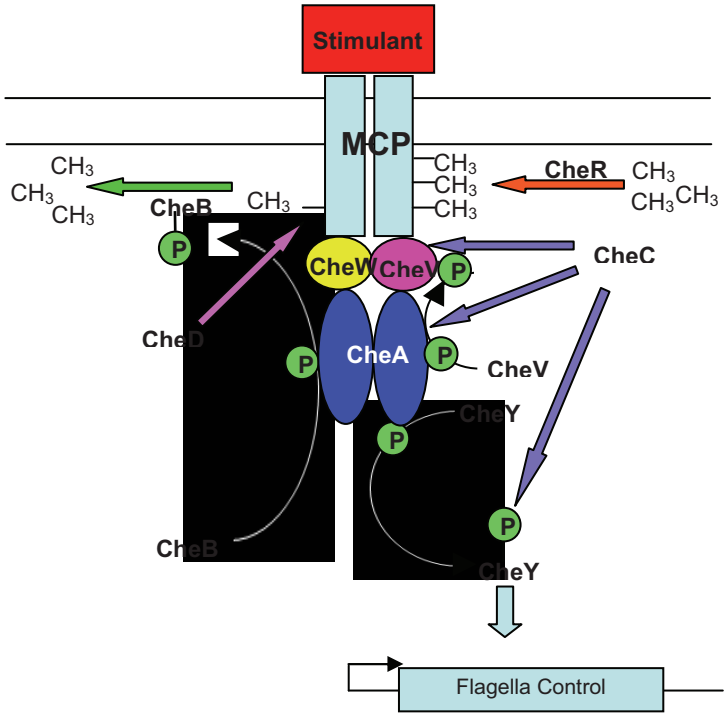


Figure 6.15: A graphical representation of the bacterial chemotaxis sensory system. MCP, Methyl accepting chemotactic protein. CheY, Response regulator. CheV, Coupling protein. CheC, Receptor deamidase/ phosphatase. CheR, Methyltransferase. CheA, Histidine kinase. CheW, Coupling protein. CheB, Glutamate methylestaerase. CheD, Receptor deamidase.

6.3.3.3 *licT-cheB* network analysis

The initial investigation of *licT* and *cheB* using SubtilNet2 and cluster analysis found that the nodes were not present within the same cluster.

Table 6.5: Network analysis properties obtained for cluster, sub-cluster and first neighbour network of the *licT* node

***licT* node**	**Cluster**	**Sub-cluster**	**First neighbour network**
MCODE cluster ranking	N/A	N/A	N/A
Functional association data source	N/A	N/A	KEGG Co-citation DBTBS
Clustering Co-efficient	N/A	N/A	N/A

The analysis of *licT* found it not to be associated with a single cluster identified in SubtilNet2 (Table 6.5).

Table 6.6: Network analysis properties obtained for cluster, sub-cluster and first neighbour network of the *cheB* node.

***cheB* node**	**Cluster**	**Sub-cluster**	**First neighbour network**
MCODE cluster ranking	2/99	1/3	N/A
Functional association data source	KEGG	KEGG	KEGG
Clustering Co-efficient	0.959	0.99	N/A

The node *cheB* with a very highly connected cluster of 134 nodes (Table 6.6) has functional associations to pathways containing two component systems (81 nodes). Within this cluster, the gene ontology term signal transduction was overrepresented by 53 nodes (p-value $2.2399E^{-43}$). On further cluster analysis (Figure 6.16), a smaller cluster of 88 nodes was produced containing *cheB* that was more highly connected than the previous network (Table 6.6). These nodes showed a functional association with pathways involving the ribosome and had 56 nodes over representing the gene ontology for translation ($7.2788E^{-62}$).

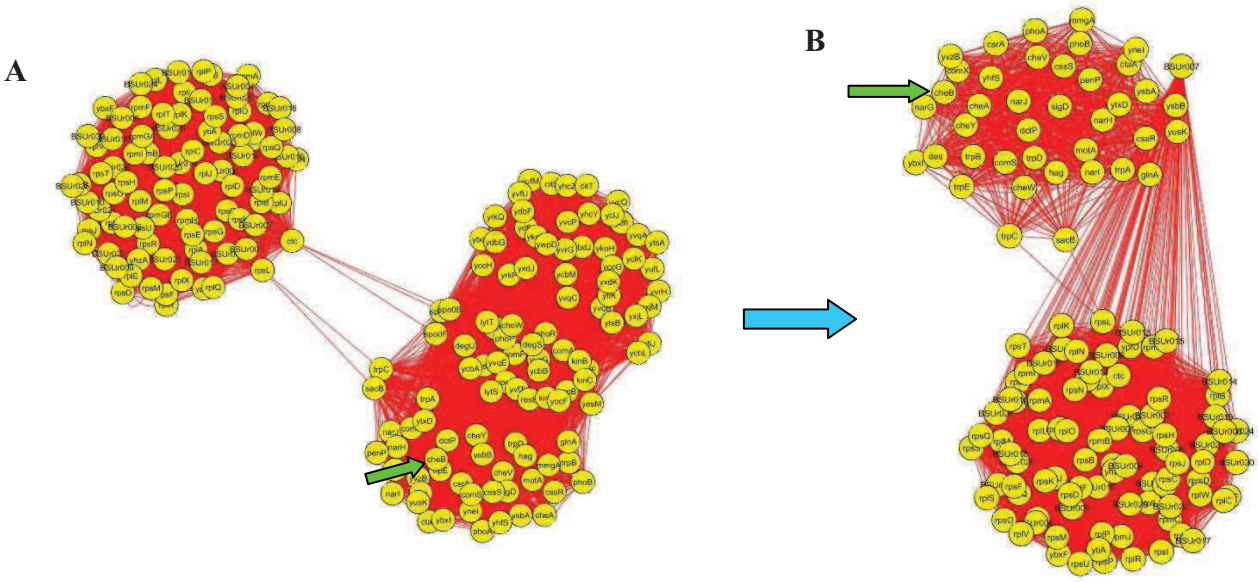


Figure 6.16: Clustering of the *cheB* node. A, The identified *cheB* cluster within the whole SubtilNet2 network. B, The sub-cluster to which the *cheB* node has been identified to. The green arrow indicates the presence of the *cheB* node within both clusters.

First neighbour networks generated for each candidate revealed no co-associated nodes being present. The shortest path between the *cheB* and *licT* node was investigated, and determined to contain 5 possible intermediate nodes, *penP*, *bglS*, *bglP*, *des* and *sacB* generating a path length of between two and three (Figure 6.17):

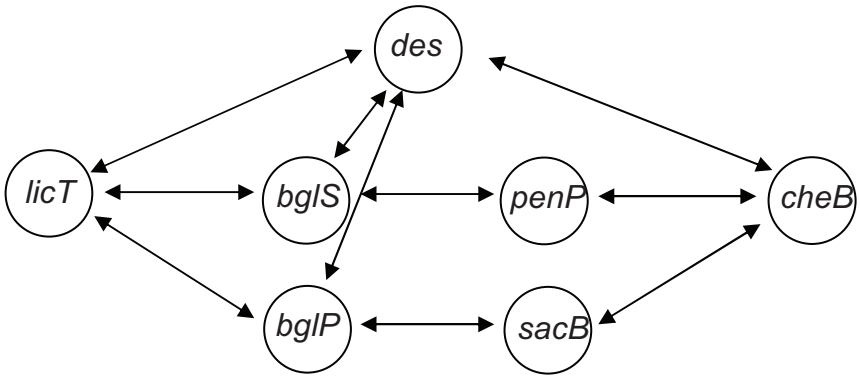


Figure 6.17: A graphical representation of SubtilNet2 determined shortest paths between the *licT* and *cheB* nodes.

The intermediate nodes predicted were shown to encode the following:

- *penP:* Encodes a β-lactamase precursor (Finn *et al.*, 2008) that is subsequently used to hydrolyse β-lactam rings.
- *bglS:* Encodes an extracellular endo-beta-1,3-1,4 glucanase enzyme that hydrolyses the cell walls of lichen (Lechat *et al.*, 2008).
- *bglP:* Encodes a phosphotransferase system (PTS) β -glucoside-specific enzyme IIBCA component (Lechat *et al.*, 2008).
- *des:* Encodes a fatty acid desaturase, used in controlling the synthesis of unsaturated fatty acids from saturated phospholipid precursors to allow the bacterium membrane to adapt and remain fluid during reductions in external temperature (Aguilar *et al.*, 1998).
- *sacB:* Encodes a levansucrase that converts sucrose to levan, a fructose polymer (Ortiz-Soto *et al.*, 2008).

None of the intermediate nodes within the shortest path could be found to completely functionally associate with both *cheB* and *licT*. There were confirmed functional associations between the *cheB*, *penP* and *sacB* nodes, as well as *licT*, *bglS* and *bglP* nodes, but no functional association between *des*, *bglS* and *penP* nodes or *bglP* and *sacB* nodes for any of the candidates.

The node *des*, as well as representing a legitimate node within SubtilNet2, also represents an artefact from the mining of co-citation data used to produce the SubtilNet2

network. The titles and abstracts of papers in this case have contained the names Laboratoire de Génétique "des" Microorganismes and also *licT*, *bglS* and *bglP*(Le Coq *et al.*, 1995). The mining program interpreted this as a fuctional association and linked them as such, when in actual fact they represent no functional association. This has also been found to occur with the nodes *bglS* with *penP* and *bglP* with *sacB,* in which described functional association are simply the result of the use of the gene names within abstracts that describe the similarities in structures with one another (Hess & Graham, 1990; Kruger & Hecker, 1995).

When data sets are mined computationally, errors such as these are likely to occur. As automatic mining programs develop and the possibilities to take context into account, these incidences are likely to be reduced. This method, despite its errors, allows the incorporation of large amounts of existing valuable data, which otherwise would need to be manually entered into the network or excluded all together. As the number of data sources increase and integration methods develop these common errors will produce fewer noticeable results.

It has been noted that the PTS system can control the *fla/che* system in *E. coli* and *B. subtilis*, guiding the cell towards carbohydrate sources that it can utilise (Bachem & Stulke, 1998; Deutscher *et al.*, 2006). Given the need to search for nutrients and the known activity of *licT* and its control and regulation of itself and other components required for carbohydrate uptake, it is reasonable to assume that there could be a functional association with *cheB*. Further specific analysis of SubtilNet2 for this possibility did not demonstrate any such associations, although bearing in mind that the network does not represent all functional associations this could remain a possibility.

6.3.3.4 Conclusions of laboratory and SubtilNet2 based analysis of *licT-cheB* targets

The testing of combinations of single and combined *licT* and *cheB* KO mutants when exposed to different combinations of stressors showed only a single difference in response when exposed to vancomycin compared to wild type *B. subtilis* treated in the same way. This was attributed to the single *cheB* KO mutant with its behaviour not found to be influenced by its combination with the *licT* KO, and despite this behaviour re-appearing on testing of the double KO mutant. Using SubtilNet2 the observed results for all mutants were analysed, and there were found to be no functional association

between the two target nodes. This example emphasises the importance of manual validation of data from which functional associations are predicted, and some of the potential limitations of using such an approach.

The removal of function of the *cheB* node within laboratory investigations may not have produced other observable results, due to the conditions of growth. Its removal may also have had little direct effect upon the underlying systems involved in the stress response. The removal of *licT* node function which also produced no observable effects could be attributable to other mechanisms known to exist within *B. subtilis* that are able to able to utilise β-glucosides (Kruger & Hecker, 1995), replacing the function of the *licT* node. In addition to this, the testing of mutants in complex media containing carbohydrate sources more preferential than β-glucosides would cause the natural inactivation of *licT* by the catabolite repression process (Lindner *et al.*, 2002).

6.4.13 Additional *cheB* analysis-using microarray techniques.

The removal of the node *cheB'*s function and the resulting mutants response to vancomycin, was unusual, and unexpected. Further laboratory investigation failed to uncover an explanation for these results. First neighbour network analysis was conducted on the *cheB* node in an attempt to uncover likely underlying mechanisms responsible for these behaviours, that could be further explored.

The *cheB* node belongs to a highly connected cluster containing 81 nodes (Figure 6.18), with the majority of functional associations identified from the KEGG pathways database with a few instances from MPIDB and co-citation data. A total of 64 nodes from the cluster were found to be functionally associated with two component systems. BiNGO analysis identified 35 nodes with an overrepresentation of the ontology term associated with signal transduction (p-value $3.8023E^{-28}$). Further clustering of the first neighbour network identified *cheB* as associating with a cluster of 67 nodes with functional association with signal transduction. BiNGO analysis identified a large proportion of these nodes, 28, being functionally associated to pathways involving two component signal transduction systems (p-value $3.0221E^{-30}$).

Discovering these facts has led to the use of microarray analysis from which to further study the *cheB* node and its associating systems. Currently ongoing, *cheB* KO mutants together with wild type *B. subtilis* were exposed to 2 μg/ml vancomycin from which

samples were obtained pre and at 10 and 60 minutes post addition of stressor. These were then analysed for differences in gene expression using microarray technology.

Identified nodes of interest from the original clusters will be used as a starting point in this investigation and specifically monitored for differences in gene expression between wild type *B. subtilis* and the *cheB* KO mutant. These results will be analysed using SubtilNet2, increasing the speed of analysis and helping to realise any underlying functional associations as well as their potential significance. In addition to this, any other interesting results displayed by additional non-cluster specific nodes will also be investigated in the same manner. The results and potential insights obtained using this approach will be used to determine the course of any potential future investigations.

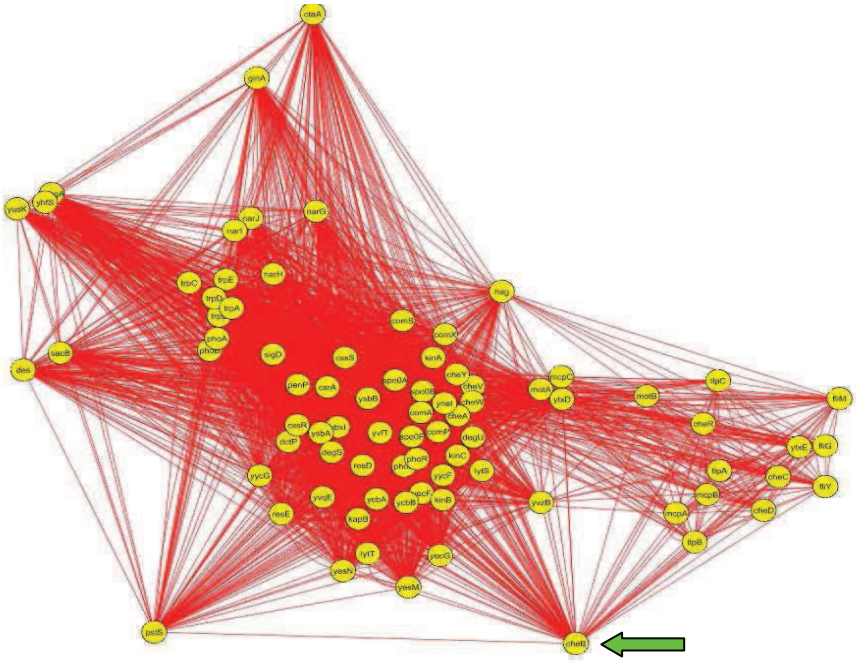


Figure 6.18: The first neighbour network of the *cheB* node. The first neighbour network of the *cheB* node consisting of 81 associating nodes. The green arrow indicates the *cheB* node.

6.3.4 *fbaB-yacL* interaction

6.3.4.1 *fbaB* background

The gene *fbaB* encodes a 6-phospho-5-dehydro-2-deoxy-D-gluconate aldolase enzyme that converts 6-phospho-5-dehydro-2-deoxy-D-gluconate to dihydroxyacetone phosphate and 3-oxopropanoate (Yoshida *et al.*, 2008). This enzyme catalyses step six of seven steps in the degradation of myo-inositol. The products produced as part of this pathway are acetyl-CoA, produced from 3-oxopropanoate and glyceraldehydes 3-phosphate produced from dihydroxyacetone phosphate (Finn *et al.*, 2008).

6.3.4.2 *yacL* background

The gene *yacL* encodes a putative membrane protein that shows similarities to those from other *Bacillus* species. It possesses PIN, HIN and TRAM protein domains (Kanehisa & Goto, 2000) and could function within a pili retraction system (Lechat *et al.*, 2008):

- PIN domains are named after their homology with the N- terminal domain of the pili biogenesis protein (Wall & Kaiser, 1999). They are part of a large family of proteins found in over 300 eukaryotes, bacteria and archaea, and initially thought to function in signalling (Noguchi *et al.*, 1996). Recent bioinformatics analysis has suggested that the domain has an exonuclease function (Clissold & Ponting, 2000).
- HIN domains have no known function. They are found in one or two copies per protein, and are found to follow the PAAD/DAPIN domain (Finn *et al.*, 2008; Liu *et al.*, 2003).
- TRAM protein domains are predicted to be RNA-binding domains (Anantharaman *et al.*, 2001) and may perform a nucleic acid binding role (Finn *et al.*, 2008).

6.3.4.3 *fbaB-yacL* network analysis

Cluster analysis of SubtilNet2 did not identify any cluster to which both the *fbaB* node and *yacL* node were associated.

Table 6.7: Network analysis properties obtained for cluster, sub-cluster and first neighbour network of the *fbaB* node

***fbaB* node**	**Cluster**	**Sub-cluster**	**First neighbour network**
MCODE cluster ranking	31/99	1/2	N/A
Functional association data source	KEGG Co-citation	KEGG	KEGG Co-citation
Clustering Co-efficient	0.933	1.0	N/A

fbaB was found to associate with a highly connected 7 node cluster (Table 6.7) containing 6 nodes with a predominate functional association to carbohydrate metabolism pathways, specifically inositol phosphate (The universal protein resource (UniProt) 2009; Kanehisa & Goto, 2000; Lechat *et al.*, 2008).

Attempts to investigate this cluster with the BiNGO plug-in were unsuccessful. Further clustering analysis identified a second sub cluster to which *fbaB* node was found to be fully connected containing 6 nodes and 15 edges (Figure 6.19). All 6 nodes showed functional associations with the Inositol phosphate pathway. Attempts to BiNGO analyse the sub cluster also failed.

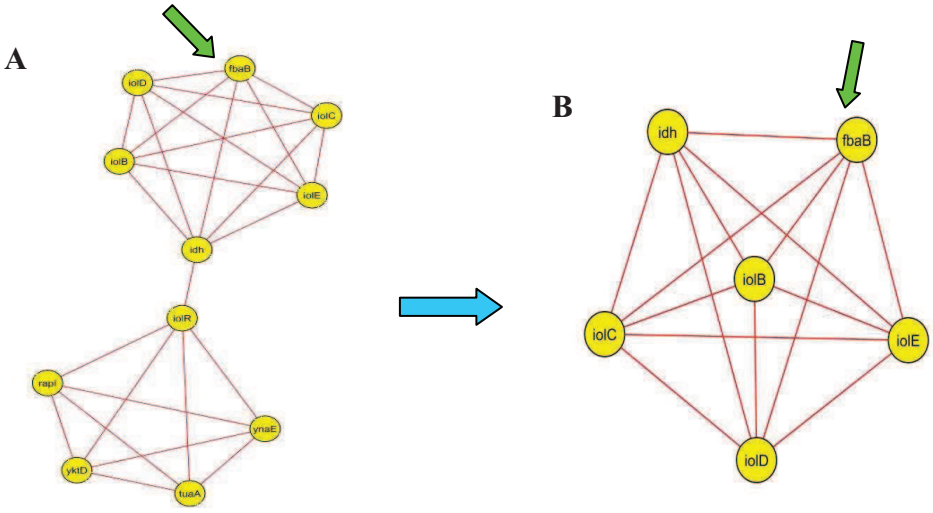


Figure 6.19: Clustering of the *fbaB* node. A, The identified *fbaB* cluster within the whole SubtilNet2 network. B, The sub-cluster to which the *fbaB* node has been identified to. The green arrow indicates the presence of the *fbaB* node within both clusters.

Table 6.8: Network analysis properties obtained for cluster, sub-cluster and first neighbour network of the *yacL* node.

***yacL* node**	**Cluster**	**Sub-cluster**	**First neighbour network**
MCODE cluster ranking	N/A	N/A	N/A
Functional association data source	N/A	N/A	KEGG
Clustering Co-efficient	N/A	N/A	N/A

The node, *yacL* was found not to be present within any cluster in SubtilNet2 (Table 6.8).

A generated first neighbour network for both nodes did not reveal any co-associated functionally associated nodes. Shortest path analysis from SubtilNet2 identified a shortest path length of 4 nodes with the nodes, *mmsA*, *tpiA*, *aldY* amd *sigB* acting as intermediates in the path connecting the *fbaB* and *yacL* nodes (Figure 6.20).

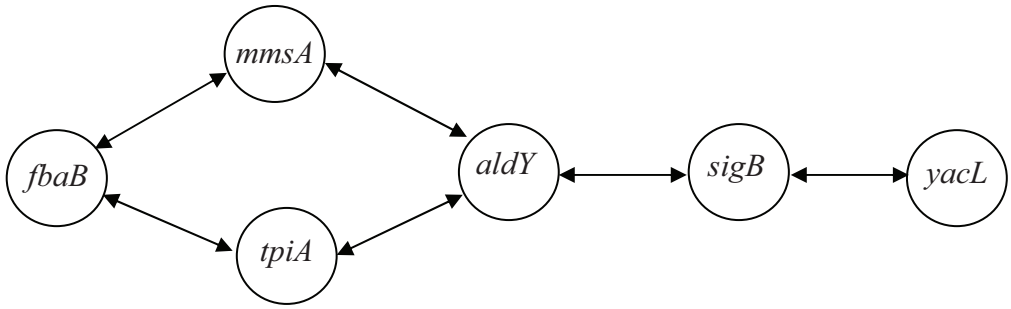


Figure 6.20: A graphical representation of SubtilNet2 determined shortest paths between the *fbaB* and *yacL* nodes.

With an unknown characterised function for the *yacL* node, speculation as to its functional association to *fbaB* were be made. Due to a lack of cluster membership, its first neighbour network was instead analysed. This was found to contain only a single node *sigB* that was functionally associated via transcriptional regulation. *SigB* is used in widespread stress responses in *B. subtilis* suggesting that *yacL* may also be part of a stress response.

Shortest path analysis between the *yacL* and *fbaB* nodes identified the following intermediate nodes associated with a path length of 4:

- *mmsA:* Encodes a methylmalonate-semialdehyde dehydrogenase that is involved in the final step (7/7) of the metabolism of myo-inositol, converting 3-oxopropanoate into acetyl-CoA (Yoshida *et al.*, 2008).
- *tpiA:* Encodes a triosephosphate isomerase, that converts dihydroxyacetone phosphate to glyceraldehyde 3 phosphate (Kanehisa & Goto, 2000).
- *aldY*: Encodes a putative aldehyde dehydrogenase (Lechat *et al.*, 2008). In sequence comparisons with closely related species- *B. anthracis* and *B. cereus* there was found to be 45.7 % and 45.1 % identity from sequence alignment with a glyceraldehyde-3-phosphate dehydrogenase (Finn *et al.*, 2008).
- *sigB*: Encodes an RNA polymerase sigma factor that is induced under conditions of general stress. It promotes the attachment of transcription factors to promoter sites, inducing systems to negate the effects of stress (Finn *et al.*, 2008) such as the node *aldY*.

Functional associations have been found between the *fbaB* node and *mmsA* and *tpiA* nodes. This is due to membership of the inositol phosphate metabolism pathway. In this pathway *fbaB* converts 6-phospho-5-dehydro-2-deoxy-D-gluconate to either malonic semialdehyde or dihydroxacetone phosphate (Finn *et al.*, 2008). These products are then converted by either *mmsA* or *tpiA* to acetyl-CoA and glyceraldehyde-3-P respectively (Kanehisa & Goto, 2000). These products enter the glycolysis and gluconeogenesis pathway, where *aldY* is thought to function under the control of *sigB* in conditions of ethanol and other such stresses (Petersohn *et al.*, 1999), with *yacL* found to be co-transcribed with *sigB* (Sierro *et al.*, 2008).

6.3.4.3 Conclusions of laboratory and SubtilNet2 based analysis of *fbaB- yacL* targets

Existing biological knowledge, together with network analysis could not identify any functional associations between the target nodes or any likely underlying biological mechanisms that could have been responsible for a lack of phenotypic difference seen with mutants devoid of node function and wild type *B. subtilis* tested under different stress conditions.

B. subtilis can use myo-inositol as a carbon source (Kanehisa & Goto, 2000), and will induce the enzyme *fbaB* amongst others required for its degradation in conditions that require it. During this investigation, the growth and stress conditions may not have been suitable for its induction, and so any phenotype due to its removal may not have been seen. Should the investigation have tested the mutant in conditions using myo-inositol as a sole carbon source, the experimental phenotype that may have resulted, together with that when combined with a mutant devoid of *yacL* function may have been different. Exact speculation as to the combined mutant phenotype, has been hampered, with no function, or functional associations known about the *yacL* node, apart from being co-transcribed with *sigB*.

No significantly different phenotypes were obtained when *yacL* function was removed from *B. subtilis* and compared to standard wild type *B. subtilis* under different stress conditions. This could suggest that either the system(s) to which *yacL* associates, has the potential to use a redundant system, or the function of the node may not produce a phenotypic effect under the testing conditions used.

6.3.5 *abnA-yjcH* interaction

6.3.5.1 *abnA* background

The gene *abnA* encodes an arabinan-endo 1,5-alpha-L-arabinase. This enzyme catalyzes the hydrolysis of the alpha-1,5-linked L-arabinofuranoside backbone polysaccharides found in plant cell walls (Finn *et al.*, 2008). The transcription of *abnA* is repressed in the presence of glucose and induced by arabinose and arabinan in the growth environment (Raposo *et al.*, 2004).

6.3.5.2 *yjcH* background

The gene *yjcH* encodes an as yet uncharacterised putative hydrolase (Lechat *et al.*, 2008).

6.3.5.3 *abnA-yjcH* network analysis

The *abnA* node was not found to be within any clusters when SubtilNet2 was subjected to cluster analysis (Table 6.9).

Table 6.9: Network analysis properties obtained for cluster, sub-cluster and first neighbour network of the *abnA* node.

***abna* node**	**Cluster**	**Sub-cluster**	**First neighbour network**
MCODE cluster ranking	N/A	N/A	N/A
Functional association data source	N/A	N/A	KEGG
Clustering Co-efficient	N/A	N/A	N/A

Table 6.10: Network analysis properties obtained for cluster, sub-cluster and first neighbour network of the *yjcH* node.

***yjcH* node**	**Cluster**	**Sub-cluster**	**First neighbour network**
MCODE cluster ranking	N/A	N/A	N/A
Functional association data source	N/A	N/A	N/A
Clustering Co-efficient	N/A	N/A	N/A

The node *yjcH* was found to be absent from SubtilNet2 entirely (Table 6.10). BLAST searches were conducted to determine if there were any other nodes likely to be confused for it and attempts were made to find potential associating systems using SubtilNet2, but due to its uncharacterised nature and potential to be used in multiple reactions, little network analysis could be conducted upon it. In an additional effort to characterise this node and any potential functional associations that it may have had in particular with *abnA*, the first neighbour network of *abnA* was investigated for the presence of any hydrolase enzymes. Two nodes were found in this network that had this property, *abfA* and *xsa.* These encoded two alpha-L-arabinofuranosidase that are involved in the hydrolysis of non-reducing alpha-L-arabinofuranoside residues in alpha-L-arabinosides (Finn *et al.*, 2008). BLAST searches conducted upon their coding sequences revealed that they were not mistaken for the *yjcH* node.

6.3.5.4 Conclusions of laboratory and SubtilNet2 based analysis of *abnA-yjcH* targets

With a lack of biological knowledge and characterisation of the *yjcH* node and a lack of presence within the SubtilNet2 network, analysis could not be conducted as to the effects that its removal would have. As such speculation as to the potential underlying biological systems that could be affected could not be known. The effect of combining the removal of this node together with the removal of the function of the *abnA* node could also not be known. Despite this, the experimental testing of both single and

combinations of mutants lacking their functions under a range of different stresses did not reveal any phenotypic differences when compared with wild type *B. subtilis* treated in the same way. The use of complex growth media during stress testing of mutants could have potentially suppressed any phenotypic effects that would have been seen by the removal of the *abnA* node. The lack of phenotypic difference of the removal of the *yjcH* node function both individually and when combined with *abnA* could suggest, the use of redundant system, a lack of node requirement during a stress response, or simply being a member of a non critical system.

6.3.6 *yndH-ycdH* interaction

6.3.6.1 *yndH* background

The gene *yndH* encodes a conserved hypothetical protein (Lechat *et al.*, 2008). The UniProt database (Finn *et al.*, 2008) identifies it an as an as yet uncharacterised protein. BLAST searches indicate a link to uncharacterised hypothetical proteins with no functional annotation.

6.3.6.2 *ycdH* background

The gene *ycdH* has been suggested to encode a Zn(II)-binding lipoprotein (Lechat *et al.*, 2008), a solute binding component of an ATP binding cassette, associated with Zn (II) uptake (Gaballa & Helmann, 1998) and one of three such systems with *B. subtilis* (Gaballa *et al.*, 2002). It has also been suggested that the gene encodes a Mn (II) solute binding protein involved in the ATP binding cassette for Mn (II) uptake (Bunai *et al.*, 2004).

There is some evidence that *ycdH* is controlled by the zinc uptake repressor (Zur), being repressed in the presence of zinc (II) (Gaballa *et al.*, 2002), giving greater weight to the speculation that this gene is involved with the uptake of zinc from the environment.

6.3.6.3 *yndH-ycdH* network analysis

From the global SubtilNet2 cluster analysis there were no clusters identified in which both candidate nodes were present together.

Table 6.11: Network analysis properties obtained for cluster, sub-cluster and first neighbour network of the *yndH* node.

***yndH* node**	**Cluster**	**Sub-cluster**	**First neighbour network**
MCODE cluster ranking	N/A	N/A	N/A
Functional association data source	N/A	N/A	N/A
Clustering Co-efficient	N/A	N/A	N/A

The node *yndH* was not present within the SubtilNet2 network (Table 6.11). BLAST searches of the coding sequence of the node did not reveal any other node that was likely to be associated with it.

Table 6.12: Network analysis properties obtained for cluster, sub-cluster and first neighbour network of the *ycdH* node.

***ycdH* node**	**Cluster**	**Sub-cluster**	**First neighbour network**
MCODE cluster ranking	1/99	N/A	N/A
Functional association data source	KEGG	N/A	KEGG
Clustering Co-efficient	0.945	N/A	N/A

The node *ycdH* was identified in the SubtilNet2 network, and was found to be within a highly connected cluster (Table 6.12) containing 114 nodes with 100 nodes (Figure 6.21) representing a functional association to multiple pathways using ATP Binding cassettes. Gene ontology analysis of the cluster identified 111 nodes as being associated with the ontology terms associated with transport (p-value $1.0408E^{-}76$). These findings

suggest that the *ycdH* belong to the systems speculated for it. Further clustering analysis identified no additional sub-clusters.

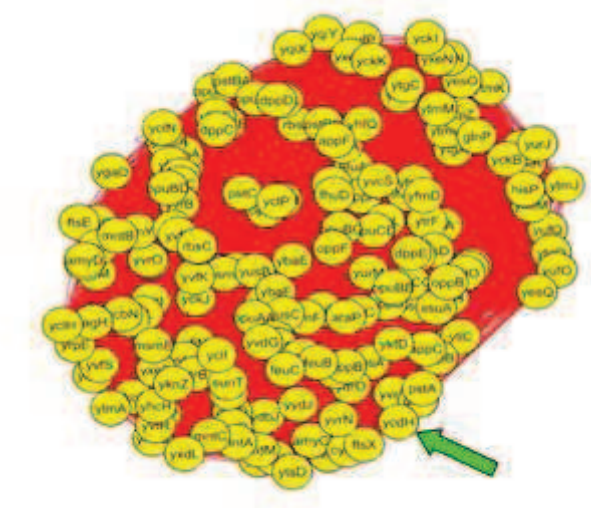


Figure 6.21: Clustering of the *ycdH* node. The identified *ycdH* cluster within the whole SubtilNet2 network.

6.3.6.4 Conclusions of laboratory and SubtilNet2 based analysis of *yndH-ycdH* targets

A lack of data available for the nodes *yndH* and *ycdH*, prevented their local properties as well as potential functional associations from being investigated. This prevented the understanding of behaviours and potential functional associations exhibited during laboratory experimentation. Cluster analysis did show that the cluster to which the *ycdH* node associated with did have functional associations with ATP binding cassettes, adding weight to speculations of involvement in the Zn uptake system. As our knowledge of these nodes increases, and their functional associations better defined, it should be possible to improve our understanding of the experimental data.

6.4 Arginine and pyrimidine (uracil) biosynthetic system analysis using SubtilNet2

6.4.1 Arginine and pyrimidine biosynthetic systems

The arginine and pyrimidine biosynthetic system both require CP. The majority of organisms utilise a single CP producing enzyme CPS, however *B subtilis* utilises a separate CPS for each systems designated CPS-A supplying the arginine biosynthetic system and CPS-P, supplying the pyrimidine biosynthetic system. Analysis of the gene organisation of both systems indicates the clustered organisation of genes encoding

systems associated with each CPS (see 1.1.4.4). In organisms possessing a single CPS, these genes are found to be distributed widely amongst the chromosome.

An investigation was begun to determine if CP produced by each CPS containing system was specific to individual reactions within the cell, or could be used by the opposite system. If it couldn't, this could potentially suggest the presence of a macromolecular complex to channel CP, together with other substrates in each biosynthetic system, preventing them from becoming available to enter other pathways. This could account for the unusual arrangement of genes found to surround each biosynthetic systems CPS.

Attempts were made in the laboratory to express both the large and small subunits, that form the CPS from each biosynthetic system, to try and combine them in an opposite configuration (Figure 6..22) and to see if they would produce CP and to which system if any it would be targeted to. CLUSTAL W sequence alignment analysis (Thompson *et al.*, 1994) had already revealed each subunit to share 55% homology to its partner in the opposite biosynthetic system. As this was being explored attempts were also made to express a few genes found within the gene clusters surrounding each CPS, to see if they had associations with specific CPS and add evidence to the theory of the presence of a macromolecular complex.

Both of these attempts failed, due to either a failure to produce the required protein or the protein being produced in an unsuitable form. Instead the investigation was conducted using SubtilNet2 in a computational approach. Potential functional associations with each CPS producing system were identified, allowing the destination of each systems produced CP to be predicted and determining if there was any specificity to them.

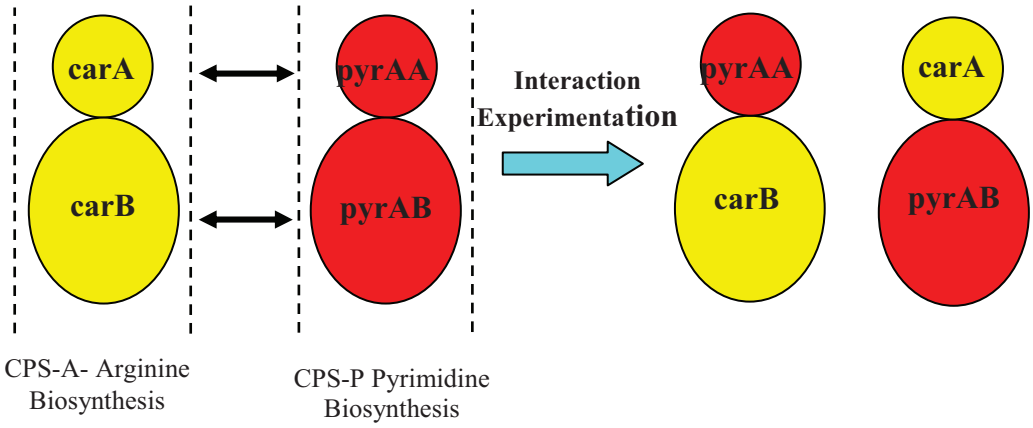


Figure 6.22: A graphical representation of the laboratory analysis of CPS originating from the arginine and pyrimidine biosynthetic systems.

6.4.2 SubtilNet2 analysis

The *carA* and *carB* nodes were identified in SubtilNet2, from which first neighbour networks were produced, and network analysis conducted. All nodes within the two networks associated with these two nodes were then cross checked with one another, checking for overlaps. This was repeated for the *pyrAA* and *pyrAB* nodes after which the node lists from both systems were checked against one another. This would allow the identification of functional associations either unique to each system, suggesting specificity, or more generalised shared functional associations.

Network analysis of the first neighbour network of the *carA* node, representing the small subunit of the CPS for the arginine biosynthetic systems contained 66 nodes that were determined to be functional associated (KEGG and co-citation data). 46 nodes were found to have an association with the pyrimidine metabolism pathway. BiNGO analysis of the first neighbour network nodes showed an overrepresentation of the ontology term for nucleobase, nucleoside and nucleotide metabolism (p-value $3.2208E^{-21}$). Cluster analysis indicated the presence of two clusters, the first containing 42 nodes had functional associations with the pyrimidine metabolism pathway (40 nodes). BiNGO analysis revealed 40 nodes associated with the term nucleobase, nucleoside and nucleotide metabolism (p-value $1.0461E^{-23}$). The second identified cluster containing 22 nodes, had 18 nodes functionally associated to the pathway of alanine, aspartate and

glutamate metabolism and 14 nodes associated to the ontology term linked to amino acid and derivative metabolic processes (p-value 3.3184 E^{-11}).

Network analysis of the first neighbour network of the *carB* node, representing the large subunit of the CPS for the arginine biosynthetic system, contained 65 nodes associated through the pyrimidine metabolic pathway with data taken from KEGG. Subsequent cluster analysis identified two clusters, the first containing 41 nodes, had 40 nodes functionally associated to pyrimidine metabolic pathways. BiNGO analysis identified 40 nodes with the gene ontology term associated to nucleobase, nucleoside and nucleotide (p-value 1.0461E^{-23}). The second cluster containing 23 nodes had an overrepresentation of 14 nodes associated to the gene ontology term amino acid and derivative metabolic processes (p-value 3.3184 E^{-11}).

The analysis of both first neighbour node networks for *carA* and *carB* identified the presence of almost identical nodes, varying only by the additional node *argF* in the *carA* first neighbour network. The results of clustering also produced identical results, with two sub networks that contained the identical number of nodes and overrepresentation of the same gene ontology, although this represented pyrimidine biosynthesis rather than the expected arginine biosynthesis. Despite this, these findings suggest biologically plausible functional association relationships between the two given the existing knowledge of *carA* and *carB* nodes.

The first neighbour network analysis of *pyrAA* contained 64 nodes, functionally associated to pyrimidine metabolism pathways (46 nodes) by KEGG pathways and co-citation data. BiNGO analysis showed the first neighbour network of *pyrAA* to have an overrepresentation (44 nodes) of the ontology term nucleobase, nucleoside and nucleotide metabolic process (p-value 7.1612E^{-15}). Cluster analysis of the generated first neighbour network identified 2 further sub-networks. The first containing 42 nodes, with functional associations to pyrimidine metabolism (40 nodes) and the overrepresentation (40 nodes) of the ontology term associated with nuclease, nucleoside and nucleotide metabolic processes (p-value 1.0461E^{-23}). The second cluster with 22 nodes had 20 nodes associated with alanine, aspartate and glutamate metabolism and an overrepresentation (20 nodes) associated with the term amino acid and derivative metabolic process (p-value 3.3184 E^{-11}).

The first neighbour network of *pyrAB* contained *64* nodes, with 36 nodes functionally associated to the pyrimidine metabolic pathway by KEGG pathways and co-citation data. This first neighbour network had an over representation (27 nodes) of the ontology associated with nucleobase, nucleoside and nucleotide metabolic processes (p-value $1.1872E^{-21}$). Cluster analysis of the generated first neighbour network identified two further sub networks. The first, 42 nodes in size had functional associations with the pyrimidine metabolism pathway (36 nodes) and an over representation (40 nodes) with the ontology nucleobase, nucleoside and nucleotide metabolic process (p-value $1.0461E^{-23}$). The second 22 node cluster, had overrepresentation (14 nodes) of the ontology term amino acid and derivative metabolic process (p-value $3.3184\ E^{-11}$).

The overlap of shared nodes present in the first neighbour networks of *pyrAA* and *pyrAB* was identical, as found to occur with the *carA* and *carB* networks. The clustering behaviour and overrepresentation of gene ontologies within these networks was also found to be identical. These finding together could indicate a functional association between *pyrAA* and *pyrAB* that is biologically plausible and would agree with previously described functional associations of the *pyrAA* and *pyrAB* nodes.

When the nodes present within the first neighbour networks of *carA* and *carB,* were compared with those of the *pyrAA* and *pyrAB* networks they were found to be almost identical with the exception of the *argC* and *argF* nodes missing from the *pyrAA* and *pyrAB* first neighbour networks. This initially suggested a biologically plausible association between the components and associated nodes of both the arginine and pyrimidine biosynthetic systems, indicating the likely sharing of produced CP. However further analysis of the implied functional interactions evidence, generated predominantly from KEGG pathways data, has identified the use of the CPS enzymes interchangeably in both described systems. This is as a result of the way in which data is integrated into the KEGG pathways database, which relies upon enzyme commission number (E.C) that describes the enzymes activity and substrate.

Both CPS components of the arginine and pyrimidine biosynthetic system perform the same function, with the same substrate and so share the same E.C designation number (E.C.6.3.5.5 (Kanehisa & Goto, 2000)). KEGG distinguishes between enzymes of a system based upon this E.C number, and so when it is shared, errors such as those seen occur. This issue has also occurred for the assignment of gene ontologies which is based upon associated biological processes, cellular components and molecular functions in

a species independent manner (Ashburner *et al.*, 2000). This occurrence has made the separation and identification of systems that may be associated to individual CPS impossible, and comes as a result of efforts to try and adapt and integrate existing data to new systems biology applications.

6.4.3 Arginine and pyrimidine biosynthetic system summary

A laboratory investigation was initiated to determine wether CP produced by the arginine biosynthetic system could be shared with the pyrimidine biosynthetic system. If it couldn't, this could suggest that the unusual cluster of genes found to surround each CPS may be being used to encode a macromolecular complex structure to channel CP and other substrates within a biosynthetic system.

To test this hypothesis we attempted to identify interactions within and between pathway components using biophysical techniques such as SPR and ITC. This required the production of microgram quantities of the target proteins for key components of the arginine and pyrimidine biosynthectic pathways

Attempts to express the subunits failed, and so a computer investigation was launched using the newly developed SubtilNet2 network. This approach focused on the analysis of closely associating systems for each CPS. From this, any associating systems common to both biosynthetic systems could be identified that could indicate the sharing of CP or it may instead indicate a lack of associating systems, suggesting the potential for a macromolecular complex.

The first neighbours of each subunits of each CPS were generated and cluster analysis performed. Each first neighbour network produced identical clusters that had over representation of the same ontologies nucleobase, nucleoside and nucleotide metabolic process and amino acid and derivative metabolic processes and a predominate number of nodes associated with pathways involved in pyrimidine metabolism. Analysis of the exact nodes present within these first neighbour networks for individual CPS subunits identified identical nodes as being present with the addition of *argF* in the arginine biosynthetic system. Comparisons made between both CPS systems identified almost identical nodes with the absence of *argC* and *argF* from the pyrimidine biosynthetic system. These findings, as well as suggesting potential functional associations amongst components of the same biosynthetic system also suggest a lack of unique systems

associated with either biosynthetic system, potentially indicating the sharing of CP irrespective of the system producing it.

The unusual gene ontology overrepresentation produced in the cluster analysis of the arginine biosynthetic system was further analysed. Looking into the exact nature of functional associations predicted by SubtilNet2, revealed the predominate use of the KEGG pathways database. Further analysis of individual associations within the database particularly for *carA* and *carB,* revealed their interchangeable use with *pyrAA* and *pyrAB*. This can be attributed to the way in which KEGG designates enzyme to pathways, and is based upon E.C number, which both CarA/B and PyrA/B components share (E.C.6.3.5.5 (Kanehisa & Goto, 2000)). The same was also found for the gene ontology applied to both *carA* and *carB,* which was identical to that of *pyrAA* and *pyrAB*. Because of this, the generated associating systems are inaccurate and the conclusion originally made before this was known using the SubtilNet2 network must be discounted.

The adoption of greater number of data sources, manual checking and better assignment of gene ontology terms in the future, should eliminate these inaccuracies from occurring, an indication of what needs to be achieved in the future to adopt systems biology approaches, and the problems associated with using existing data sources.

6.5 SubtilNet2 discussion

6.5.1 General discussion

The initial exploration of the generated SubtilNet2 functional network to test the visualisation and analysis properties of the plug-in tools, called for the selection of three random nodes, with the intention of applying network analysis techniques to analyse them, determining likely biological properties and functional associations.

Cluster analysis was first performed on each randomly selected node, identifying associating clusters and the exact extent of its association. Nodes associating with the randomly selected node and within the cluster were identified and there functional association with one another analysed. This was followed by the analysis of gene ontology representation of the individual clusters identifying occurrences of overrepresentation that could provide clues to potential underlying biological function/property associated with the cluster and its nodes.

Following the analysis of these clusters first neighbours of the randomly chosen nodes were made into their own network (first neighbour networks) and subjected to the same analysis.

This has demonstrated that all three randomly chosen nodes were fully connected to individual clusters, which showed significant overrepresentation in gene ontologies associated with the randomly selected node. Functional association investigations of the nodes present within these cluster, found them to share similarities to known functional associations of the randomly selected node. Similar results were obtained when analysing the first neighbour networks of random nodes in this way.

6.5.2 e-Therapeutics candidate targets

Of all six provided e-Therapeutics candidate pairs target pairs experimentally tested, non were found to act as potential therapeutic targets when combined. Using SubtilNet2, they were investigated, to determine any probable functional interactions between one another using a network analysis approach, and if this could be used to explain the biological behaviour observed during experimental testing.

None of the predicted target nodes could be associated with the same cluster, instead associating with either separate clusters or none at all. First neighbour network analysis of both target nodes, found the majority have common associating nodes, which were commonly found to be included within the shortest path analysis. Shortest path analysis identified several biologically plausible functional associations between nearly all targets nodes upon which it could be conducted, however in some cases this represented only a small association and could also have been the result of coincidence, and completely unrelated. Not forgetting that this investigation has focused upon the shortest path analysis as the most likely association between nodes however, there are likely to be other longer pathways through which functional associations could occur and which have not been listed here.

The following are biologically plausible functional associations between candidate pairs, predicted by SubtilNet2:

- *T*he *ywdH-ybfS* candidate pair are both components of a multistage amino sugar metabolism pathway and can be functionally associated with one another in a

biologically plausible way using the fructose metabolism pathway, with the suggested *ypqE* intermediate node identified from shortest path analysis.

- The *yvgQ-luxS* candidate pair components of either the cysteine synthesis pathway or methionine and AdoMet recycling pathway can be linked with one another using the sulphur transsulfuration pathway. Nodes from both first neighbour networks of *yvgQ* and *luxS* have been identified as belonging to these pathways, suggesting a functional associations between the two candidate nodes.
- The *licT-cheB* candidate pair components of the allowing the utilisation of aryl-β-glucosides as a carbon source under conditions of carbohydrate limitation, and chemotaxis towards attractants respectively. These systems could not be functionally associated with one another, using the SubtilNet2 predicted intermediates. These intermediates were discovered to be a result of how the network had been produced, and were not relevant. However further analysis into the systems of *licT* and *cheB* suggested a strong possible functional association through the combined use of a PTS.
- The *fbaB-yacL* candidates have been shown to have a functional association through co-transcription of nodes present with the first neighbour networks of *yacL* and membership to the inositol phosphate pathway of intermediate nodes present in the first neighbour network of *fbaB*.

The target nodes, *yjcH-abnA* and *yndH-ycdH* could not be functionally associated. This is not to suggest that there is no functional association, but rather a current inability to analyse the target nodes, through poor functional association characterisation or a lack of node presence within SubtilNet2. Attempts made to speculate any potential functional interaction based upon the limited information available or closely associating nodes, could still not be made.

During the network analysis of the functional associations between the target nodes, several suggestions as to associating systems, properties and their underlying biological functions that could account for the experimental phenotype observed when removing their function singly and in combination could be made. It was noted on several occasions, that the conditions under which some nodes and their associated systems would have been active were not compatible with the growth conditions used. This could have prevented any experimental effect being seen on the removal of the node

function. Future investigations would take this into account and conduct experimentation with these conditions in mind.

6.5.3 Further analysis of *cheB*

Following the experimental identifying of the mutant *cheB* KO as acting differently to wild type following the addition of vancomycin stressor. Network analysis using the newly developed SubtilNet2 network was used to direct and focus a new round of experimentation involving the testing of mutant and wild type gene expression after the addition of vancomycin using microarrays, to attempt to discover the underlying biological basis for the behaviours exhibited.

Finding insights into how candidate nodes interact with one another and the likely associating systems has made predicting in some cases the causes for the experimental results obtained easier, identifying associating and redundant systems to the modifications made.

The first neighbour network identifying the most closely functionally associating nodes to *cheB* was constructed and nodes contained within it identified. This was subjected to network analysis identifying potential clusters and likely associated biological systems. These identified nodes were then monitored in the microarray experiments. All results obtained were analysed using the SubtilNet2 network to try and identify the underlying biological basis for the results obtained.

6.5.4 Arginine and pyrimidine biosynthesis

Following the attempted laboratory analysis of interactions between components of the arginine and pyrimidine biosynthetic system in attempts to identify the destination or produced CP, a computational analysis approach was adopted using the newly developed SubtilNet2 and network analysis, to attempt to track the utilising biological systems. The results of network analysis initially indicated the probable sharing of produced CP by both the arginine and pyrimidine biosynthetic systems, as the network analysis identified identical associating nodes, and clusters to both systems. Further analysis of the suggested functional interactions and there suggested source indicated a problem with this interpretation of the KEGG pathways, and gene ontology data used. Because of the way in which it they are compiled the identification of specific associations with additional systems unique to themselves systems could not be

identified. This resulted in a premature conclusion that, both the arginine and pyrimidine biosynthetic systems were likely sharing their produced CP, which would negate a need for a mechanism to channel CP between these systems.

These findings have again identified the problem of poor characterisation of functional association/ lack of node presence within the SubtilNet2 network, as has been found with the testing of the therapeutic candidates. This problem will continue to be present as existing data sources, which were never intended to be used in such a way continue to be used for Systems Biology. The immediate implications of this with regard to this investigation is that it doesn't allow the identification of potential macromolecular complexes or otherwise within the biosynthetic systems of arginine and pyrimidine. This represents a wider problem associated with the generation of accurate and representative biological models from limited numbers of data sources.

The almost exclusive use of the KEGG pathways database to functionally associate nodes together continues to remain an issue, as by using multiple data sources, the limitations encountered may have been avoided. With the lack of other high-throughput, high coverage data sources available for *B. subtilis* this was taken into account before conclusions of any biologically plausible functional associations were made.

In the interval of waiting for research and technology to develop to allow better integration of existing datasets, developments need to be made into new technologies and standards specifically for use in Systems Biology approaches, allowing predictions to become potentially more accurate and reliable. In the mean time SubtilNet2 and other such developed systems and approaches should be used as a guide only.

Given the limitations discussed, SubtilNet2 has made the analysis of functional associations in *B.subtilis* more rapid and provided focus for subsequent microarray experimentation, while also providing valuable insights and starting points for further analysis. As more data sets become integrated into the network, together with developments in integration and analysis techniques of existing data, the problems encountered in this investigation should become more infrequent, and it is hoped that SubtilNet2 will be developed to be able to predict functional associations, based on surrounding nodes.

Chapter 7
Conclusions

7. Conclusions

This chapter summarises and integrates the results of the experimental and computational analyses of e-Therapeutics candidates and the arginine and pyrimidine biosynthetic systems. It also includes a more general summary of the field and its potential benefits and limitations over a range of time scales.

7.1 e-Therapeutics candidates

The simultaneous targeting of multiple non-essential genes within a biological system, looking for a synergistic therapeutic effect, has the potential to become an established method from which future therapies may develop. Producing therapeutic compounds that function in this way, could replace traditional approaches that focus on very specific interactions with a few essential biological components, often exhibit side effects and, in the case of antimicrobials, have the potential to develop resistance. In contrast, the adoption of the specific targeting of multiple components within a biological system holds out the prospect of delivering effective drugs with better efficacy and reduced side effects and resistance.

This investigation has focused upon the validation of gene combinations, generated using data driven methods, selected as having the potential to be used as therapeutic targets within the model Gram positive bacteria *B. subtilis*. These predictions were made by our industrial collaborators at e-Therapeutics, after producing and testing a proprietary *in silico B. subtilis* interactome model.

These predictions were tested using a strategy that involved the production of a series of knockout mutants. The mutants were subjected to a range of sub-inhibitory stresses, each designed to target a different areas of cellular metabolism. The aim was to determine whether the disruption of combinations of genes would make them more susceptible to the stress than observed with single gene disruptions, and thereby indicating if the targeting of combinations of genes could identify novel potential therapeutic targets.

A multi-staged approach was developed to test the candidate genes. Initial tests used a high throughput approach, with potential growth affects confirmed by re-testing using low throughput approaches. This multi-stage approach was designed in such a way as to ensure that candidate genes could be tested individually and in combination, as well as

avoiding downstream polar effects. This technique could be complemented in future investigations with the creation of transcriptional fusion mutants where the transcription of a gene of interest can be varied. This approach would complemet the standard knockout analysis used to identify potential candidates for further study.

In total 4500 separate growth curves were produced and analysed throughout the course of this investigation. As technology develops, it is anticipated that more rapid, sensitive and highly automated phenotypic assays systems with be developed to facilitate the testing of vast numbers of potential candidates combinations, an approach that could have applications in other Systems Biology projects.

With these new technologies and the information gathered, it is anticipated that methods for large-scale automated and continuous model will facilitate *in silico* predictions that can be rapidly evaluated in the laboratory. The results will then be fed back into the model for further and more accurate *in silico* predictions. Such model development/experimental interactions will continue to a point at wich in silico predictions rival the accuracy of laboratory experimentation.

As efforts are under way to develop these new approaches so too are efforts to improve the integration and modelling techniques for the vast amounts of heterogeneous data that is currently being made available. This will facilitate the production of more comprehensive, accurate and representative biological models, and should overcome many of the limitations encountered in this investigation regarding poorly represented or absent data.

The high throughput stage identified eight potential candidate target genes that, when inactivated, showed variations in their growth phenotypes when compared with the wild type. These candidates were subjected to a low throughput analysis stage where only a single target gene, *cheB*, continued to demonstrate a significantly different phenotype. Throughout the high throughput testing, the *cheB* mutant produced consistently different growth phenotypes when exposed to most of the applied stresses. However during the low throughput testing, the only stress that produced a consistently different growth phenotype was vancomycin. It was established that this behaviour was specific to the *cheB* lesion and was not affected by its cognate partner, *licT*, or any genes downstream and in the same transcriptional unit as *cheB*.

In conclusion the laboratory experiments established that none of the combinations of predicted gene targets exhibited synergistic effects with respect to a variety of stresses and are therefore unlikely to be suitable targets for the development of novel therapeutic /antimicrobial compounds. The phenotypic testing did reveal that the *cheB* mutant exhibited an unexpected altered growth profile in the presence of sub-inhibitory concentrations of vancomycin. The influence of vancomycin on *cheB*, a gene involved in demethylating chemotactic proteins, warrants further investigation, and to this end DNA array experiments are currently being undertaken.

7.2 Development and application of SubtilNet2 for the verification of e-Therapeutics drug candidates

In an alternative approach to the validation of the e-Therapeutics targets, a computer-based approach was adopted. In collaboration with the Integrated Bioinformatics Group at Newcastle University, an in-house functional interaction model, SubtilNet2, was developed. In the absence of the original proprietary algorithms developed by e-Therapeutics, it was not possible to compare these two models in terms of their structures, only in terms of their outcomes.

SubtilNet2 was constructed using a combination of four distinct data sources. These were used to model potential functional interactions between *B. subtilis* genes and their products. Before the application of SubtilNet2 and associated statistical tools to the potential therapeutic candidates, exploration of the network identified several nodes that were consistent with already well understood biologically systems. SubtilNet2 was then used to explore associations between the therapeutic targets predicted by e-Therapeutics in attempt to explain the experimental results, as well as acting as a point from which to continue further investigations.

Computational investigation of the supplied therapeutic targets with SubtilNet2 and associated statistical tools suggested possible associations between the individual targets within each of the pairs *ywdH-ybfS*, *yvgQ-luxS* and *fbaB-yacL*. However these targets were only found to associate *via* long, multiple-step, biosynthetic pathways or processes. In contrast, it was not possible to analyse three candidate pair combinations, *abnA-yjcH*, *yndH-ycdH* and *cheB-licT,* due to insufficient information on the functional associations between nodes.

Following the experimental approach, SubtilNet2 provided an alternative data-driven approach with which to validate the therapeutic candidates predicted by e-Therapeutics. Both approaches came to the same conclusion, namely that the proposed candidate pairs did not have the capacity to act as therapeutic targets. The behaviour of one of the gene candidates, namely *cheB* encoding a chemotactic demethylating protein, did prove to exhibit an unexpected growth phenotype when exposed to vancomycin, indicating a greater susceptibility to this compound. Investigations to determine the exact cause of this behaviour have proved to be inconclusive and we are awaiting the analysis of DNA array data. SubtilNet2 identified 81 genes/gene products that are associated with *cheB* and these will be used as the starting point in the analysis of the microarray data. It is expected that the availability of SubtilNet2 will reduce the time needed to complete this analysis.

To continue this investigation, the data sources utilised by SubtilNet 2 will need to be updated to improve its predictability and accuracy, and this is likely to be possible using data currently being generated by the EU-funded BaSysBio project and the multinational Bacell SysMo projects. Once this is done it would be more appropriate to adopt a probabilistic method for producing the functional network, circumventing current problems associated with data coverage.

7.3 Arginine and pyrimidine (uracil) biosynthetic system analysis

The second component to this study was to investigate aspects of the arginine and pyrimidine biosynthetic system in relation to hypothesis driven questions as to its organisation and function. Carbamoyl phosphate is a central component in the biosynthesis of arginine and pyrimidines. In most organisms a single carbamoyl phosphate synthetases supplies CP for both systems. *B. subtilis* and close relatives encode two separate CP producing enzymes, one associated with the arginine biosynthetic pathway and the other with the pyrimidine pathway. Moreover, the genes associated with these pathways are clustered on the chromosome, together with the genes encoding their cognate CPS. The presence of two CPS enzymes, and the clustering of the genes encoding the system components has led us to propose that the enzymes in these pathways may form a multi-enzyme complex that channels pathway substrates. Such a cellular organisation would have a significant influence on current areas of our understanding of systems modelling and would have implications for the

likely efficiency of pathways developed along synthetic biology principles. It would also reveal new insights into the evolutionary pressure that have maintained operon structures in prokaryotic organisms.

The challenge of these studies was to determine whether components of the arginine biosynthetic system could interact with themselves and/or with components of the pyrimidine biosynthetic system. The CPSs of both systems were used as the starting point. CPSs are heterodimeric enzymes and therefore it was interesting to determine whether active enzymes could be reconstituted from subunits of the different pathways. Our approach was to purify the four proteins and to analyse all-against-all interactions by surface plasmon resonance and isothermal titration calorimetry. However, attempts to isolate active CPS subunits and associated pathway enzymes were largely unsuccessful, despite rigorous extensive modifications to the expression protocols. ArgF, encoding ornithine carbamoyl phosphatase was successfully expressed but was insoluble. Attempts to solubilise this protein were unsuccessful and it is possible that it may need to be co-expressed with other pathway components to produce a soluble version of this protein. In retrospect, it would have been useful to have adopted a two-hybrid approach to the analysis of these pathways.

In the absence of purified components on which to carry out physical interaction studies, a computational approach was undertaken using the newly generated SubtilNet2. The SubtilNet2 analysis showed that the arginine and pyrimidine pathway components formed clusters within a single node that included the two CPS enzymes. However, further analysis revealed that, because of the reliance in the KEGG and BiNGO databases of EC enzyme nomenclature for identifying pathway enzymes, the CPS were being used interchangeably even though they share only 55% identity. As a result it was not possible to determining potential cross talk between the pathways with any confidence.

Once again, the use of a computational approach for the analysis of hypothesis driven research has identified some of the potential problems that face Systems Biology. In this component of the investigation, the source of data providing most of the associations was that of the KEGG pathways database. During the manual validation of the associations suggested by SubtilNet2, it was noted that the data source did not distinguish between common components within individual systems. This results in the inaccurate representation of exactly which genes associate with which system, and for

this investigation makes it impossible to clearly distinguish between components and their products. This highlights the need to validate all of the data suggested by the network and represent a fundamental problem of network analysis that need to be addressed by annotators. This problem can only be overcome by improving the quality of the supplied data as well as a greater and more diverse number of data sources applied to the model, removing the dependency inherent when only limited numbers of data sources are available.

In a separate investigation conducted in the Harwood group, a two-hybrid approach was used to investigate potential interactions between key components of the arginine and pyrimidine biosynthetic pathways. Initial results have shown that there is interaction between the components CarA and CarB of the arginine systems and PyrAA and PyrAB of the pyrimidine system, which was to be expected. However, an important observation was the absence of interactions between CarA/B or PyrAA/AB. Interestingly there was an interaction between ArgF and PyrB, tow enzymes with a common evolutuionary history. Such observations are still preliminary and require further investigation using biophysical methodologies.

7.4 The major benefits and limitations of Systems Biology

The emerging field of Systems Biology has the potential to revolutionise many aspects of bacteriology including drugs discovery and metabolic engineering. Systems Biology holds out the prospect of using large datasets to develop a comprehensive understanding of the complex network of interacting systems that control the behaviour of bacteria. With this improved understanding comes the prospect of manipulating these systems for the benefit of human kind. The long term objective is to develop an *in silico* model that accurately predicts bacterial behaviour in response to changes in their environment, reducing the need to conduct time-consuming laboratory investigations and speeding up the acquisition of new knowledge. In the shorter term, this developing technology is able to complement traditional approaches by suggesting important points to be considered within individual investigations.

In order to meet the eventual goal of using Systems Biology as an *in silico*-based approach to experimentation, the complexities and problems associated with this task, need to be understood and solved in a progressive manner. As part of the progressive development of this technology, current attempts to generate *in silico* models are based

on using existing data collected over the last decade and in databases that were created for other purposes. This data is usually non-standardised both with respect to their collection and analytical approaches, causing difficulties when compiling the data into forms suitable for network analysis. It is only very recently that dedicated Systems Biology databases have been established with appropriate metadata.

Rather than discounting the diverse sources of existing data, methods are continually being developed for processing and integrating this data. However, in light of these data issues, current models provide only a limited representation of system behaviour and are limited to being used to complement traditional hypothesise-driven approaches. This should be complemented by a progressive shift in experimentation to one focused on the development of systems rather than component-based knowledge. Simple predictive models should be constructed and validated experimentally, with the results being reapplied to the original model in an iterative manner to improve the predictability of the model. Once predictions from the model are shown to be accurate, and the fundamental principles of the approach established, the need for experimental validation will be considerably reduced.

In the current work, an experimental approach failed to verify the initial predictions of potential therapeutic targets. However, it should be born in mind that the original predictions were made in 2006 when both the available technology and data for the construction of networks were limited. During the course of these investigations, the re-application of a Systems Biology approach to the existing data did provided a starting point for analysing potential relationships between the original gene candidates. The network analysis also allowed us to predict the reasons for the failure of these candidates to influence growth.

7.5 The future of Systems Biology

In the short term, the developing field of Systems Biology is likely to provide informed choices for traditional hypothesis-driven laboratory-based experimentation, incorporating currently available data using developing integration methods. The incorporation of the resulting data, and other data sources, with improve the accuracy with which the predictions can be made. In the medium term, a shift will occur to a more system-wide experimental approach in which standardised experimentation and integration methods will be driven by modellers. This will increase the complexity of

the models, and interaction between the experimental and modelling approaches should lead to the ultimate longer term goal of establishing a truly functional *in silico* model that is highly predictive of actual bacterial behaviour.

As Systems Biology moves the medium to the very long term long term, it will begin influence a wide variety of practical applications. One of the most promising is that of the design of combinatorial therapeutic drugs that are able to target of multiple gene targets. Such drugs would be expected to be highly effective in inhibiting the bacterium, have reducing the rates of resistance and less pronounced side effects. Adopting this approach could provide a wider range of potential drug targets, limited only by the number of genes within the organism.

Looking at the wider applications of this technology, it provides the prospect of a more complete understanding of the fundamental principles that underpin biology and with it the ability to profoundly influence many aspects of biology The principles and discoveries made using a systems' approach are already beginning to be used to help establish new fields of biology such as Synthetic Biology. For example, the importance of chromosomal gene context, gained from insights in the arginine/pyrimidine biosynthetic pathways of *B. subtilis,* could help determine the success of such approaches.

Bibliography

Lengeler J. (1999). *Biology of the Prokaryotes*: Thieme publishing group

The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res* **37**, D169-174.

Aguilar, P. S., Cronan, J. E., Jr. & de Mendoza, D. (1998). A Bacillus subtilis gene induced by cold shock encodes a membrane phospholipid desaturase. *J Bacteriol* **180**, 2194-2200.

Ahn, A. C., Tewari, M., Poon, C. S. & Phillips, R. S. (2006). The limits of reductionism in medicine: could systems biology offer an alternative? *PLoS Med* **3**, e208.

Albanesi, D., Mansilla, M. C., Schujman, G. E. & de Mendoza, D. (2005). Bacillus subtilis cysteine synthetase is a global regulator of the expression of genes involved in sulfur assimilation. *J Bacteriol* **187**, 7631-7638.

Albert, R. & Barabasi, A. L. (2002). Statistical mechanics of complex networks. *Review of Modern Physics* **74**, 47-97.

Albert, R. (2005). Scale-free networks in cell biology. *J Cell Sci* **118**, 4947-4957.

Albert, R. (2007). Network inference, analysis, and modeling in systems biology. *Plant Cell* **19**, 3327-3338.

Alm, E. & Arkin, A. P. (2003). Biological networks. *Curr Opin Struct Biol* **13**, 193-202.

Anantharaman, V., Koonin, E. V. & Aravind, L. (2001). TRAM, a predicted RNA-binding domain, common to tRNA uracil methylation and adenine thiolation enzymes. *FEMS Microbiol Lett* **197**, 215-221.

Andrews, J. M. & Wise, R. (2002). Susceptibility testing of Bacillus species. *J Antimicrob Chemother* **49**, 1040-1042.

Aranda, B., Achuthan, P., Alam-Faruque, Y. & other authors The IntAct molecular interaction database in 2010. *Nucleic Acids Res* **38**, D525-531.

Ashburner, M., Ball, C. A., Blake, J. A. & other authors (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29.

Assenov, Y., Ramirez, F., Schelhorn, S. E., Lengauer, T. & Albrecht, M. (2008). Computing topological parameters of biological networks. *Bioinformatics* **24**, 282-284.

Auger, S., Yuen, W. H., Danchin, A. & Martin-Verstraete, I. (2002). The metIC operon involved in methionine biosynthesis in Bacillus subtilis is controlled by transcription antitermination. *Microbiology* **148**, 507-518.

Auger, S., Gomez, M. P., Danchin, A. & Martin-Verstraete, I. (2005). The PatB protein of Bacillus subtilis is a C-S-lyase. *Biochimie* **87**, 231-238.

Bachem, S. & Stulke, J. (1998). Regulation of the Bacillus subtilis GlcT antiterminator protein by components of the phosphotransferase system. *J Bacteriol* **180**, 5319-5326.

Bader, G. D., Donaldson, I., Wolting, C., Ouellette, B. F., Pawson, T. & Hogue, C. W. (2001). BIND--The Biomolecular Interaction Network Database. *Nucleic Acids Res* **29**, 242-245.

Bader, G. D. & Hogue, C. W. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4**, 2.

Balasubramanian, S., Aridoss, G., Parthiban, P., Ramalingan, C. & Kabilan, S. (2006). Synthesis and biological evaluation of novel benzimidazol/benzoxazolylethoxypiperidone oximes. *Biol Pharm Bull* **29**, 125-130.

Bandow, J. E., Brotz, H. & Hecker, M. (2002). Bacillus subtilis tolerance of moderate concentrations of rifampin involves the sigma(B)-dependent general and multiple stress response. *J Bacteriol* **184**, 459-467.

Barabasi, A. L. & Albert, R. (1999). Emergence of scaling in random networks. *Science* **286**, 509-512.

Barabasi, A. L. & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**, 101-113.

Barabasi, A. L. (2009). Scale-free networks: a decade and beyond. *Science* **325**, 412-413.

Bassingthwaighte, J., Hunter, P. & Noble, D. (2009). The Cardiac Physiome: perspectives for the future. *Exp Physiol* **94**, 597-605.

Bassingthwaighte, J. B. (2000). Strategies for the physiome project. *Ann Biomed Eng* **28**, 1043-1058.

Bergman, N. H., Passalacqua, K. D., Hanna, P. C. & Qin, Z. S. (2007). Operon prediction for sequenced bacterial genomes without experimental information. *Appl Environ Microbiol* **73**, 846-854.

Bhalla, U. S. & Iyengar, R. (1999). Emergent properties of networks of biological signaling pathways. *Science* **283**, 381-387.

Bischoff, D. S. & Ordal, G. W. (1992). Bacillus subtilis chemotaxis: a deviation from the Escherichia coli paradigm. *Mol Microbiol* **6**, 23-28.

Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H. & Falkow, S. (1977). Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* **2**, 95-113.

Boozer, C., Kim, G., Cong, S., Guan, H. & Londergan, T. (2006). Looking towards label-free biomolecular interaction analysis in a high-throughput format: a review of new surface plasmon resonance technologies. *Curr Opin Biotechnol* **17**, 400-405.

Bringel, F., Frey, L., Boivin, S. & 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.

Bron, S. (1990). Plasmids. In *Molecular Biological Methods for Bacillus* pp. 75-174. Edited by C. R. a. C. Harwood, S.M Chichester: John Wiley and Sons

Brouwer, R. W., Kuipers, O. P. & van Hijum, S. A. (2008). The relative value of operon predictions. *Brief Bioinform* **9**, 367-375.

Bruggeman, F. J. & Westerhoff, H. V. (2007). The nature of systems biology. *Trends Microbiol* **15**, 45-50.

Bunai, K., Ariga, M., Inoue, T., Nozaki, M., Ogane, S., Kakeshita, H., Nemoto, T., Nakanishi, H. & Yamane, K. (2004). Profiling and comprehensive expression analysis of ABC transporter solute-binding proteins of Bacillus subtilis membrane based on a proteomic approach. *Electrophoresis* **25**, 141-155.

Burguiere, P., Auger, S., Hullo, M. F., Danchin, A. & Martin-Verstraete, I. (2004). Three different systems participate in L-cystine uptake in Bacillus subtilis. *J Bacteriol* **186**, 4875-4884.

Butcher, E. C., Berg, E. L. & Kunkel, E. J. (2004). Systems biology in drug discovery. *Nat Biotechnol* **22**, 1253-1259.

Butland, G., Peregrin-Alvarez, J. M., Li, J. & other authors (2005). Interaction network containing conserved and essential protein complexes in Escherichia coli. *Nature* **433**, 531-537.

Caldovic, L. & Tuchman, M. (2003). N-acetylglutamate and its changing role through evolution. *Biochem J* **372**, 279-290.

Callaway, D. S., Newman, M. E., Strogatz, S. H. & Watts, D. J. (2000). Network robustness and fragility: percolation on random graphs. *Phys Rev Lett* **85**, 5468-5471.

Camazine, S., Deneubourg, J. L., Franks, N. R., Sneyd, J., G.Theraulaz & Bonabeau, E. (2001). *Self-Organisation in Biological systems* Princeston University Press.

Carballido-Lopez, R. & Errington, J. (2003). The bacterial cytoskeleton: in vivo dynamics of the actin-like protein Mbl of Bacillus subtilis. *Dev Cell* **4**, 19-28.

Carroll, S. F. (2007). *Drug and Biological development - From Molecule to Product and Beyond*: Springer.

Cases, I. & de Lorenzo, V. (2005). Promoters in the environment: transcriptional regulation in its natural context. *Nat Rev Microbiol* **3**, 105-118.

Casteele, M., Demarez, M., Legrain, C., Glansdorff, N. & Pierard, A. (1990). Pathways of arginine biosynthesis in extreme thermophilic archaeo- and eubacteria. . *J GEN MICROBIOL* **136**, 1177-1183.

Celis, T. F., Rosenfeld, H. J. & Maas, W. K. (1973). Mutant of Escherichia coli K-12 defective in the transport of basic amino acids. *J Bacteriol* **116**, 619-626.

Chen, K. C., Csikasz-Nagy, A., Gyorffy, B., Val, J., Novak, B. & Tyson, J. J. (2000). Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol Biol Cell* **11**, 369-391.

Chiang, P. K., Gordon, R. K., Tal, J., Zeng, G. C., Doctor, B. P., Pardhasaradhi, K. & McCann, P. P. (1996). S-Adenosylmethionine and methylation. *Faseb J* **10**, 471-480.

Claus, D. a. F., D. (1989). *Taxonomy of Bacillus* Plenum: New York

Clissold, P. M. & Ponting, C. P. (2000). PIN domains in nonsense-mediated mRNA decay and RNAi. *Curr Biol* **10**, R888-890.

Coffey, D. S. (1998). Self-organization, complexity and chaos: the new biology for medicine. *Nat Med* **4**, 882-885.

Commichau, F. M., Gunka, K., Landmann, J. J. & Stulke, J. (2008). Glutamate metabolism in Bacillus subtilis: gene expression and enzyme activities evolved to avoid futile cycles and to allow rapid responses to perturbations of the system. *J Bacteriol* **190**, 3557-3564.

Coonrod, J. D., Leadley, P. J. & Eickhoff, T. C. (1971). Antibiotic susceptibility of Bacillus species. *J Infect Dis* **123**, 102-105.

Cormack, B. P., Valdivia, R. H. & Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33-38.

Craddock, T. (2008). Integrating distributed post-genomic data to infer the molecular basis of bacterial phenotypes In *Integrative Bioninformatics group* pp. 271. Newcastle: Newcastle University

Csete, M. E. & Doyle, J. C. (2002). Reverse engineering of biological complexity. *Science* **295**, 1664-1669.

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

Dandekar, T., Snel, B., Huynen, M. & Bork, P. (1998). Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci* **23**, 324-328.

Daniel, R. A., Harry, E. J., Katis, V. L., Wake, R. G. & Errington, J. (1998). Characterization of the essential cell division gene ftsL(yIID) of Bacillus subtilis and its role in the assembly of the division apparatus. *Mol Microbiol* **29**, 593-604.

Davidov, E., Holland, J., Marple, E. & Naylor, S. (2003). Advancing drug discovery through systems biology. *Drug Discov Today* **8**, 175-183.

Deutscher, J., Francke, C. & Postma, P. W. (2006). How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol Mol Biol Rev* **70**, 939-1031.

di Guan, C., Li, P., Riggs, P. D. & Inouye, H. (1988). Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. *Gene* **67**, 21-30.

Diestel, R. (2005). *Graph Theory* 3rd- Electronic version 2005 edn: Springer-Verlag Heidelberg

DiMasi, J. A., Hansen, R. W. & Grabowski, H. G. (2003). The price of innovation: new estimates of drug development costs. *J Health Econ* **22**, 151-185.

Dubnau, D. (1991). Genetic competence in Bacillus subtilis. *Microbiol Rev* **55**, 395-424.

Fabret, C., Feher, V. A. & Hoch, J. A. (1999). Two-component signal transduction in Bacillus subtilis: how one organism sees its world. *J Bacteriol* **181**, 1975-1983.

Ferrell, J. E., Jr. (2009). Q&A: systems biology. *J Biol* **8**, 2.

Finn, R. D., Tate, J., Mistry, J. & other authors (2008). The Pfam protein families database. *Nucleic Acids Res* **36**, D281-288.

Finnigan, J. (2005). The Science of Complex Systems *Australasian Science* 32-34.

Gaballa, A. & Helmann, J. D. (1998). Identification of a zinc-specific metalloregulatory protein, Zur, controlling zinc transport operons in Bacillus subtilis. *J Bacteriol* **180**, 5815-5821.

Gaballa, A., Wang, T., Ye, R. W. & Helmann, J. D. (2002). Functional analysis of the Bacillus subtilis Zur regulon. *J Bacteriol* **184**, 6508-6514.

Galperin, M. Y. & Ellison, M. J. (2006). Systems biology: sprint or marathon? *Current Opinion in Biotechnology* **17**, 437-439.

Gestwicki, J. E., Lamanna, A. C., Harshey, R. M., McCarter, L. L., Kiessling, L. L. & Adler, J. (2000). Evolutionary conservation of methyl-accepting chemotaxis protein location in Bacteria and Archaea. *J Bacteriol* **182**, 6499-6502.

Goldthwaite, C., Dubnau, D. & Smith, I. (1970). Genetic mapping of antibiotic resistance in markers Bacillus subtilis. *Proc Natl Acad Sci U S A* **65**, 96-103.

Goll, J., Rajagopala, S. V., Shiau, S. C., Wu, H., Lamb, B. T. & Uetz, P. (2008). MPIDB: the microbial protein interaction database. *Bioinformatics* **24**, 1743-1744.

Gorke, B. & Stulke, J. (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol* **6**, 613-624.

Halinnan, J. S., Pocock, M., Addinall, S., Lydall, D. A. & Wipat, A. (2009). Clustering incorporating shortest paths identifies relevant modules in functional interaction networks In *Computational Intelligence in Bioinformatics and Computational Biology* Nashville TN.

Hallinan, J. S. & Wipat, A. (2007). Motifs and Modules in Fractured Functional Yeast Networks In *Computational Inteligligence and Bioinformatics and Computational Biology* Honolulu HI.

Hansen, C. H., Endres, R. G. & Wingreen, N. S. (2008). Chemotaxis in Escherichia coli: a molecular model for robust precise adaptation. *PLoS Comput Biol* **4**, e1.

Hartl, D. & Jones, E. (2008). *Genetics- Analysis of genes and genomes*, 7 edn: Jones and Bartlett.

Harwood, C. R. (1992). Bacillus subtilis and its relatives: molecular biological and industrial workhorses. *Trends Biotechnol* **10**, 247-256.

Harwood, C. R. & Wipat, A. (1996). Sequencing and functional analysis of the genome of Bacillus subtilis strain 168. *FEBS Lett* **389**, 84-87.

Harwood, C. R. & Cranenburgh, R. (2008). Bacillus protein secretion: an unfolding story. *Trends Microbiol* **16**, 73-79.

Hecker, M., Heim, C., Volker, U. & Wolfel, L. (1988). Induction of stress proteins by sodium chloride treatment in Bacillus subtilis. *Arch Microbiol* **150**, 564-566.

Hess, G. F. & Graham, R. S. (1990). Efficiency of transcriptional terminators in Bacillus subtilis. *Gene* **95**, 137-141.

Holden, H. M., Thoden, J. B. & Raushel, F. M. (1999). Carbamoyl phosphate synthetase: an amazing biochemical odyssey from substrate to product. *Cell Mol Life Sci* **56**, 507-522.

Hood, L. & Perlmutter, R. M. (2004). The impact of systems approaches on biological problems in drug discovery. *Nat Biotechnol* **22**, 1215-1217.

Hoper, D., Bernhardt, J. & Hecker, M. (2006). Salt stress adaptation of Bacillus subtilis: a physiological proteomics approach. *Proteomics* **6**, 1550-1562.

Hu, Z., Mellor, J., Wu, J., Yamada, T., Holloway, D. & Delisi, C. (2005). VisANT: data-integrating visual framework for biological networks and modules. *Nucleic Acids Res* **33**, W352-357.

Hullo, M. F., Auger, S., Soutourina, O., Barzu, O., Yvon, M., Danchin, A. & Martin-Verstraete, I. (2007). Conversion of methionine to cysteine in Bacillus subtilis and its regulation. *J Bacteriol* **189**, 187-197.

Hutter, B., Fischer, C., Jacobi, A., Schaab, C. & Loferer, H. (2004). Panel of Bacillus subtilis reporter strains indicative of various modes of action. *Antimicrob Agents Chemother* **48**, 2588-2594.

Hyde, C., Ahmed, S. A., Padlan, E. A., Miles, E. W. & Davies, D. R. (1988). Three-dimensional Structure of the Tryptophan Synthasea alpha2 Beta2 Multienzyme Complex from Salmonella typhimurium. *The Journal of Biological Chemistry* **263**, 17857-17871.

Jacob, F. & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **3**, 318-356.

Jana, M., Luong, T. T., Komatsuzawa, H., Shigeta, M. & Lee, C. Y. (2000). A method for demonstrating gene essentiality in Staphylococcus aureus. *Plasmid* **44**, 100-104.

Janecka, I. P. (2007). Cancer control through principles of systems science, complexity, and chaos theory: a model. *Int J Med Sci* **4**, 164-173.

Joyce, A. R. & Palsson, B. O. (2006). The model organism as a system: integrating 'omics' data sets. *Nat Rev Mol Cell Biol* **7**, 198-210.

Kaltwasser, M., Wiegert, T. & Schumann, W. (2002). Construction and application of epitope- and green fluorescent protein-tagging integration vectors for Bacillus subtilis. *Appl Environ Microbiol* **68**, 2624-2628.

Kanehisa, M. & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30.

Kapust, R. B. & Waugh, D. S. (1999). Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci* **8**, 1668-1674.

Kellermann, O. K. & Ferenci, T. (1982). Maltose-binding protein from Escherichia coli. *Methods Enzymol* **90 Pt E**, 459-463.

Kim, J. & Raushel, F. M. (2001). Allosteric control of the oligomerization of carbamoyl phosphate synthetase from Escherichia coli. *Biochemistry* **40**, 11030-11036.

Kim, J., Howell, S., Huang, X. & Raushel, F. M. (2002). Structural defects within the carbamate tunnel of carbamoyl phosphate synthetase. *Biochemistry* **41**, 12575-12581.

Kim, Y., Cho, J. Y., Kuk, J. H., Moon, J. H., Cho, J. I., Kim, Y. C. & Park, K. H. (2004). Identification and antimicrobial activity of phenylacetic acid produced by Bacillus licheniformis isolated from fermented soybean, Chungkook-Jang. *Curr Microbiol* **48**, 312-317.

Kirsch, M. L., Peters, P. D., Hanlon, D. W., Kirby, J. R. & Ordal, G. W. (1993). Chemotactic methylesterase promotes adaptation to high concentrations of attractant in Bacillus subtilis. *J Biol Chem* **268**, 18610-18616.

Kitano, H. (2000). Perspectives on systems biology. *New generation computing* **18**, 199-216.

Kitano, H. (2002a). Computational systems biology. *Nature* **420**, 206-210.

Kitano, H. (2002b). Systems biology: a brief overview. *Science* **295**, 1662-1664.

Klug, W. S., Cummings, M. R. & Spencer, C. A. (2005). *Concepts of Genetics*, 8 edn: Pearson Education.

Kobayashi, K., Ehrlich, S. D., Albertini, A. & other authors (2003). Essential Bacillus subtilis genes. *Proc Natl Acad Sci U S A* **100**, 4678-4683.

Krahn, J. M., Kim, J. H., Burns, M. R., Parry, R. J., Zalkin, H. & Smith, J. L. (1997). Coupled formation of an amidotransferase interdomain ammonia channel and a phosphoribosyltransferase active site. *Biochemistry* **36**, 11061-11068.

Kristich, C. J. & Ordal, G. W. (2002). Bacillus subtilis CheD is a chemoreceptor modification enzyme required for chemotaxis. *J Biol Chem* **277**, 25356-25362.

Kruger, S. & Hecker, M. (1995). Regulation of the putative bglPH operon for aryl-beta-glucoside utilization in Bacillus subtilis. *J Bacteriol* **177**, 5590-5597.

Kunst, F., Ogasawara, N., Moszer, I. & other authors (1997). The complete genome sequence of the gram-positive bacterium Bacillus subtilis. *Nature* **390**, 249-256.

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

Le Coq, D., Lindner, C., Kruger, S., Steinmetz, M. & Stulke, J. (1995). New beta-glucoside (bgl) genes in Bacillus subtilis: the bglP gene product has both transport and regulatory functions similar to those of BglF, its Escherichia coli homolog. *J Bacteriol* **177**, 1527-1535.

Lechat, P., Hummel, L., Rousseau, S. & Moszer, I. (2008). GenoList: an integrated environment for comparative analysis of microbial genomes. *Nucleic Acids Res* **36**, D469-474.

Lee, I., Date, S. V., Adai, A. T. & Marcotte, E. M. (2004). A probabilistic functional network of yeast genes. *Science* **306**, 1555-1558.

Lewis, P. J. & Marston, A. L. (1999). GFP vectors for controlled expression and dual labelling of protein fusions in Bacillus subtilis. *Gene* **227**, 101-110.

Li, S., Armstrong, C. M., Bertin, N. & other authors (2004). A map of the interactome network of the metazoan C. elegans. *Science* **303**, 540-543.

Lindner, C., Hecker, M. S., #4; Vidal, #3; , #14; Anthony Reddy1, 1998 #11; Baudot, 2008 #16; Goll, 2006 #7; Graille, 2005 #29; Ito, 2001 #18; JS, 2004 #13; Li, 2004 #20; Lin, 2008 #15; Lindner, 1999 #30; Lindner, 2002 #31; Lukashin, 2003 #27; Maxwell, 2003 #12; Mountain, 1986 #6; Ng, 2006 #9; Ng, 2006 #24; Parrish, 2006 #8; Pieroni, 2008 #23; Przulj, 2004 #26; Sanchez, 1999 #17; Schnetz, 1996 #33; Shoemaker, 2007 #28; Soon-Hyung Yook1, 2004 #22; Stelzl, 2005 #21; Thattai, 2003 #32; Tsumoto, 2003 #10; Uetz, 2000 #19; Vidal, 2005 #25}, Le Coq, D. & Deutscher, J. (2002). Bacillus subtilis mutant LicT antiterminators exhibiting enzyme I- and HPr-independent antitermination affect catabolite repression of the bglPH operon. *J Bacteriol* **184**, 4819-4828.

Little, J. W., Shepley, D. P. & Wert, D. W. (1999). Robustness of a gene regulatory circuit. *Embo J* **18**, 4299-4307.

Liu, T., Rojas, A., Ye, Y. & Godzik, A. (2003). Homology modeling provides insights into the binding mode of the PAAD/DAPIN/pyrin domain, a fourth member of the CARD/DD/DED domain family. *Protein Sci* **12**, 1872-1881.

Lopez-Maury, L., Marguerat, S. & Bahler, J. (2008). Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat Rev Genet* **9**, 583-593.

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

Maas, W. K. (1991). The regulation of arginine biosynthesis: its contribution to understanding the control of gene expression. *Genetics* **128**, 489-494.

Maere, S., Heymans, K. & Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448-3449.

Maina, C. V., Riggs, P. D., Grandea, A. G., 3rd, Slatko, B. E., Moran, L. S., Tagliamonte, J. A., McReynolds, L. A. & Guan, C. D. (1988). An Escherichia coli vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. *Gene* **74**, 365-373.

Mascher, T., Zimmer, S. L., Smith, T. A. & Helmann, J. D. (2004). Antibiotic-inducible promoter regulated by the cell envelope stress-sensing two-component system LiaRS of Bacillus subtilis. *Antimicrob Agents Chemother* **48**, 2888-2896.

Mason, O. & Verwoerd, M. (2007). Graph theory and networks in Biology. *IET Syst Biol* **1**, 89-119.

Matthews, S. L. & Anderson, P. M. (1972). Evidence for the presence of two nonidentical subunits in carbamyl phosphate synthetase of Escherichia coli. *Biochemistry* **11**, 1176-1183.

Meister, A. (1989). Mechanism and regulation of the glutamine-dependent carbamyl phosphate synthetase of Escherichia coli. *Adv Enzymol Relat Areas Mol Biol* **62**, 315-374.

Miles, E. W., Rhee, S. & Davies, D. R. (1999). The molecular basis of substrate channeling. *J Biol Chem* **274**, 12193-12196.

Mitchell M, N. M. (2002). *Complex Systems Theory and Evolution*: New York: Oxford University Press.

Mota-Meira, M., LaPointe, G., Lacroix, C. & Lavoie, M. C. (2000). MICs of mutacin B-Ny266, nisin A, vancomycin, and oxacillin against bacterial pathogens. *Antimicrob Agents Chemother* **44**, 24-29.

Mountain, A., McChesney, J., Smith, M. C. & Baumberg, S. (1986). Gene sequence encoding early enzymes of arginine synthesis within a cluster in Bacillus subtilis, as revealed by cloning in Escherichia coli. *J Bacteriol* **165**, 1026-1028.

Murphy, N., McConnell, D. J. & Cantwell, B. A. (1984). The DNA sequence of the gene and genetic control sites for the excreted B. subtilis enzyme beta-glucanase. *Nucleic Acids Res* **12**, 5355-5367.

Narsingh, D. (2004). *Graph Theory, - with applications to engineering and computer science* Prentice-Hall of India Pvt.Ltd.

Needham, C. J., Bradford, J. R., Bulpitt, A. J. & Westhead, D. R. (2006). Inference in Bayesian networks. *Nat Biotechnol* **24**, 51-53.

Nicholl, D. S. (2002). *Genetic engineering*, 2 edn: Cambridge University Press.

Nicoloff, H., Hubert, J. C. & 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.

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

Noble, D. (2002). Modeling the heart--from genes to cells to the whole organ. *Science* **295**, 1678-1682.

Noble, D. (2005). The heart is already working. *Biochem Soc Trans* **33**, 539-542.

Noble, D. (2008). Computational models of the heart and their use in assessing the actions of drugs. *J Pharmacol Sci* **107**, 107-117.

Noguchi, E., Hayashi, N., Azuma, Y. & other authors (1996). Dis3, implicated in mitotic control, binds directly to Ran and enhances the GEF activity of RCC1. *Embo J* **15**, 5595-5605.

Oda, K., Matsuoka, Y., Funahashi, A. & Kitano, H. (2005). A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol Syst Biol* **1**, 2005 0010.

Oliver, S. (2000). Guilt-by-association goes global. *Nature* **403**, 601-603.

Ortiz-Soto, M. E., Rivera, M., Rudino-Pinera, E., Olvera, C. & Lopez-Munguia, A. (2008). Selected mutations in Bacillus subtilis levansucrase semi-conserved regions affecting its biochemical properties. *Protein Eng Des Sel* **21**, 589-595.

Perozich, J., Nicholas, H., Wang, B. C., Lindahl, R. & Hempel, J. (1999). Relationships within the aldehyde dehydrogenase extended family. *Protein Sci* **8**, 137-146.

Petermann, T. & Rios, P. D. L. (2004). Exploration of scale-free networks *The European Physical Journal B* **38**, 201-204.

Petersohn, A., Bernhardt, J., Gerth, U., Hoper, D., Koburger, T., Volker, U. & Hecker, M. (1999). Identification of sigma(B)-dependent genes in Bacillus subtilis using a promoter consensus-directed search and oligonucleotide hybridization. *J Bacteriol* **181**, 5718-5724.

Petterlini & Scardoni (2008). pp. Path extraction by smallest cost algorithm (PeSca) project.

Poolman, B., Driessen, A. J. & Konings, W. N. (1987). Regulation of arginine-ornithine exchange and the arginine deiminase pathway in Streptococcus lactis. *J Bacteriol* **169**, 5597-5604.

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

Powers, S. G., Meister, A. & Haschemeyer, R. H. (1980). Linkage between self association and catalytic activity of Escherichia coli carbamyl phosphate synthetase. *J Biol Chem* **255**, 1554-1558.

Price, M. N., Arkin, A. P. & Alm, E. J. (2006). The life-cycle of operons. *PLoS Genet* **2**, e96.

Priest, F. G. (1989). *Isolation and identification of anaerobic endospore-forming bacteria* New York:Plenum

Proulx, S. R., Promislow, D. E. & Phillips, P. C. (2005). Network thinking in ecology and evolution. *Trends Ecol Evol* **20**, 345-353.

Provvedi, R. & Dubnau, D. (1999). ComEA is a DNA receptor for transformation of competent Bacillus subtilis. *Mol Microbiol* **31**, 271-280.

Purcarea, C., Herve, G., Cunin, R. & Evans, D. R. (2001). Cloning, expression, and structure analysis of carbamate kinase-like carbamoyl phosphate synthetase from Pyrococcus abyssi. *Extremophiles* **5**, 229-239.

Rahman, M. M., Alam, A. H., Sadik, G., Islam, M. R., Khondkar, P., Hossain, M. A. & Rashid, M. A. (2007). Antimicrobial and cytotoxic activities of Achyranthes ferruginea. *Fitoterapia* **78**, 260-262.

Rain, J. C., Selig, L., De Reuse, H. & other authors (2001). The protein-protein interaction map of Helicobacter pylori. *Nature* **409**, 211-215.

Raman, K. & Chandra, N. (2008). Mycobacterium tuberculosis interactome analysis unravels potential pathways to drug resistance. *BMC Microbiol* **8**, 234.

Rao, C. V., Glekas, G. D. & Ordal, G. W. (2008). The three adaptation systems of Bacillus subtilis chemotaxis. *Trends Microbiol* **16**, 480-487.

Raposo, M. P., Inacio, J. M., Mota, L. J. & de Sa-Nogueira, I. (2004). Transcriptional regulation of genes encoding arabinan-degrading enzymes in Bacillus subtilis. *J Bacteriol* **186**, 1287-1296.

Rawlins, M. D. (2004). Cutting the cost of drug development? *Nat Rev Drug Discov* **3**, 360-364.

Reizer, J., Bachem, S., Reizer, A., Arnaud, M., Saier, M. H., Jr. & Stulke, J. (1999). Novel phosphotransferase system genes revealed by genome analysis - the complete complement of PTS proteins encoded within the genome of Bacillus subtilis. *Microbiology* **145 (Pt 12)**, 3419-3429.

Riggs, P. (2001). Expression and purification of maltose-binding protein fusions. *Curr Protoc Mol Biol* **Chapter 16**, Unit16 16.

Rodriguez-Martinez, J. M., Velasco, C., Briales, A., Garcia, I., Conejo, M. C. & Pascual, A. (2008). Qnr-like pentapeptide repeat proteins in gram-positive bacteria. *J Antimicrob Chemother* **61**, 1240-1243.

Rosario, M. M., Fredrick, K. L., Ordal, G. W. & Helmann, J. D. (1994). Chemotaxis in Bacillus subtilis requires either of two functionally redundant CheW homologs. *J Bacteriol* **176**, 2736-2739.

Saier, M. H., Jr. (2001). The bacterial phosphotransferase system: structure, function, regulation and evolution. *J Mol Microbiol Biotechnol* **3**, 325-327.

Sakanyan, V., Charlier, D., Legrain, C., Kochikyan, A., Mett, I., Pierard, A. & Glansdorff, N. (1993). Primary structure, partial purification and regulation of key

enzymes of the acetyl cycle of arginine biosynthesis in Bacillus stearothermophilus: dual function of ornithine acetyltransferase. *J Gen Microbiol* **139**, 393-402.

Sakanyan, V., Petrosyan, P., Lecocq, M., Boyen, A., Legrain, C., Demarez, M., Hallet, J. N. & Glansdorff, N. (1996). Genes and enzymes of the acetyl cycle of arginine biosynthesis in Corynebacterium glutamicum: enzyme evolution in the early steps of the arginine pathway. *Microbiology* **142 (Pt 1)**, 99-108.

Sakanyan, V., Kochikyan, A, Mett A, et al (1992). A re-examination of the pathway for ornithine biosynthesis in a thermophilic and two mesophilic Bacillus species. *Journal of General Microbiology* **138**, 125-130.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). Molecular Cloning, a laboratory manual (2nd edition). Coldspring Harbour Laboratory Press.

Sauers, K. J., P. Groh, S. (1975). Alcohol-bicarbonate-water system. Structure-reactivity studies on the equilibriums for formation of alkyl monocarbonates and on the rates of their decomposition in aqueous alkali. *J Am Chem Soc* **97**, 5546-5553.

Saulmon, M. M., Karatan, E. & Ordal, G. W. (2004). Effect of loss of CheC and other adaptational proteins on chemotactic behaviour in Bacillus subtilis. *Microbiology* **150**, 581-589.

Schirner, K., Marles-Wright, J., Lewis, R. J. & Errington, J. (2009). Distinct and essential morphogenic functions for wall- and lipo-teichoic acids in Bacillus subtilis. *Embo J* **28**, 830-842.

Schnetz, K., Stulke, J., Gertz, S., Kruger, S., Krieg, M., Hecker, M. & Rak, B. (1996). LicT, a Bacillus subtilis transcriptional antiterminator protein of the BglG family. *J Bacteriol* **178**, 1971-1979.

Schoeberl, B., Eichler-Jonsson, C., Gilles, E. D. & Muller, G. (2002). Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat Biotechnol* **20**, 370-375.

Schujman, G. E., Choi, K. H., Altabe, S., Rock, C. O. & de Mendoza, D. (2001). Response of Bacillus subtilis to cerulenin and acquisition of resistance. *J Bacteriol* **183**, 3032-3040.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504.

Short, B. (2009). Cell biologists expand their networks. *J Cell Biol* **186**, 305-311.

Sierro, N., Makita, Y., de Hoon, M. & Nakai, K. (2008). DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information. *Nucleic Acids Res* **36**, D93-96.

Sindelar, D. J. A. C. a. R. D. (2002). *Pharmaceutical Biotechnology- An introduction for Pharmacists and Pharmaceutical scientists*: Taylor and Francis

Sophos, N. A., Pappa, A., Ziegler, T. L. & Vasiliou, V. (2001). Aldehyde dehydrogenase gene superfamily: the 2000 update. *Chem Biol Interact* **130-132**, 323-337.

Sophos, N. A. & Vasiliou, V. (2003). Aldehyde dehydrogenase gene superfamily: the 2002 update. *Chem Biol Interact* **143-144**, 5-22.

Stein, A., Russell, R. B. & Aloy, P. (2005). 3did: interacting protein domains of known three-dimensional structure. *Nucleic Acids Res* **33**, D413-417.

Strogatz, S. H. (2001). Exploring complex networks. *Nature* **410**, 268-276.

Szurmant, H., Bunn, M. W., Cannistraro, V. J. & Ordal, G. W. (2003). Bacillus subtilis hydrolyzes CheY-P at the location of its action, the flagellar switch. *J Biol Chem* **278**, 48611-48616.

Tatonetti, N. P., Liu, T. & Altman, R. B. (2009). Predicting drug side-effects by chemical systems biology. *Genome Biol* **10**, 238.

Tchieu, J. H., Norris, V., Edwards, J. S. & Saier, M. H., Jr. (2001). The complete phosphotranferase system in Escherichia coli. *J Mol Microbiol Biotechnol* **3**, 329-346.

Thoden, J. B., Holden, H. M., Wesenberg, G., Raushel, F. M. & Rayment, I. (1997). Structure of carbamoyl phosphate synthetase: a journey of 96 A from substrate to product. *Biochemistry* **36**, 6305-6316.

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

Tobisch, S., Glaser, P., Kruger, S. & Hecker, M. (1997). Identification and characterization of a new beta-glucoside utilization system in Bacillus subtilis. *J Bacteriol* **179**, 496-506.

Torreri, P., Ceccarini, M., Macioce, P. & Petrucci, T. C. (2005). Biomolecular interactions by Surface Plasmon Resonance technology. *Ann Ist Super Sanita* **41**, 437-441.

Trunk, G. V. (1979). A problem of dimensionality: a simple example. *IEEE Transaction on Pattern Analysis and Machine Inteliligence* **1**, 306-307.

Vagner, V., Dervyn, E. & Ehrlich, S. D. (1998). A vector for systematic gene inactivation in Bacillus subtilis. *Microbiology* **144 (Pt 11)**, 3097-3104.

van der Ploeg, J. R., Barone, M. & Leisinger, T. (2001). Functional analysis of the Bacillus subtilis cysK and cysJI genes. *FEMS Microbiol Lett* **201**, 29-35.

Van Regenmortel, M. H. (2004). Reductionism and complexity in molecular biology. Scientists now have the tools to unravel biological and overcome the limitations of reductionism. *EMBO Rep* **5**, 1016-1020.

van Riel, N. A. (2006). Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. *Brief Bioinform* **7**, 364-374.

Vazquez-Ramos, J. M. & Mandelstam, J. (1981). Oxolinic acid-resistant mutants of Bacillus subtilis. *J Gen Microbiol* **127**, 1-9.

Velazquez Campoy, A. & Freire, E. (2005). ITC in the post-genomic era...? Priceless. *Biophys Chem* **115**, 115-124.

Vogel, H. J. (1953). Path of Ornithine Synthesis in Escherichia Coli. *Proc Natl Acad Sci U S A* **39**, 578-583.

Wadhams, G. H. & Armitage, J. P. (2004). Making sense of it all: bacterial chemotaxis. *Nat Rev Mol Cell Biol* **5**, 1024-1037.

Wall, D. & Kaiser, D. (1999). Type IV pili and cell motility. *Mol Microbiol* **32**, 1-10.

Walsh, C. (2000). Molecular mechanisms that confer antibacterial drug resistance. *Nature* **406**, 775-781.

Wang, T. T., Bishop, S. H. & Himoe, A. (1972). Detection of carbamate as a product of the carbamate kinase-catalyzed reaction by stopped flow spectrophotometry. *J Biol Chem* **247**, 4437-4440.

Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature* **393**, 440-442.

Winzer, K., Hardie, K. R. & Williams, P. (2002). Bacterial cell-to-cell communication: sorry, can't talk now - gone to lunch! *Curr Opin Microbiol* **5**, 216-222.

Wissenbach, U., Six, S., Bongaerts, J., Ternes, D., Steinwachs, S. & Unden, G. (1995). A third periplasmic transport system for L-arginine in Escherichia coli: molecular characterization of the artPIQMJ genes, arginine binding and transport. *Mol Microbiol* **17**, 675-686.

Xenarios, I., Rice, D. W., Salwinski, L., Baron, M. K., Marcotte, E. M. & Eisenberg, D. (2000). DIP: the database of interacting proteins. *Nucleic Acids Res* **28**, 289-291.

Yoshida, K., Yamaguchi, M., Morinaga, T., Kinehara, M., Ikeuchi, M., Ashida, H. & Fujita, Y. (2008). myo-Inositol catabolism in Bacillus subtilis. *J Biol Chem* **283**, 10415-10424.

Zanzoni, A., Montecchi-Palazzi, L., Quondam, M., Ausiello, G., Helmer-Citterich, M. & Cesareni, G. (2002). MINT: a Molecular INTeraction database. *FEBS Lett* **513**, 135-140.

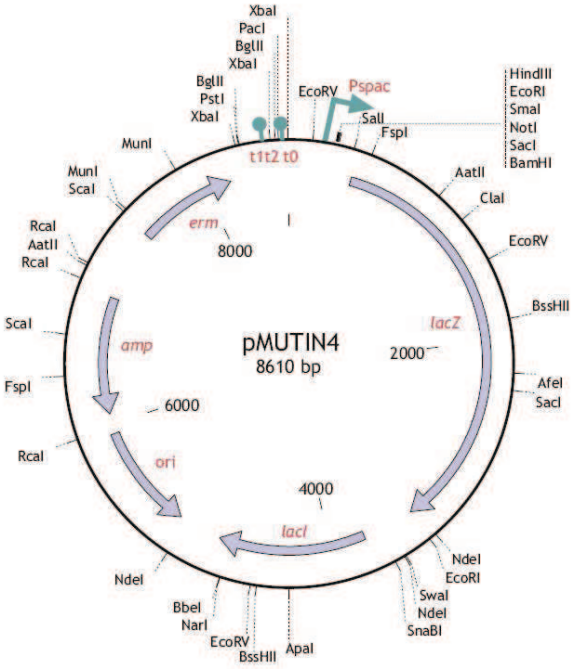
Zhu, X., Gerstein, M. & Snyder, M. (2007). Getting connected: analysis and principles of biological networks. *Genes Dev* **21**, 1010-1024.

Zimmer, M. A., Tiu, J., Collins, M. A. & Ordal, G. W. (2000). Selective methylation changes on the Bacillus subtilis chemotaxis receptor McpB promote adaptation. *J Biol Chem* **275**, 24264-24272.

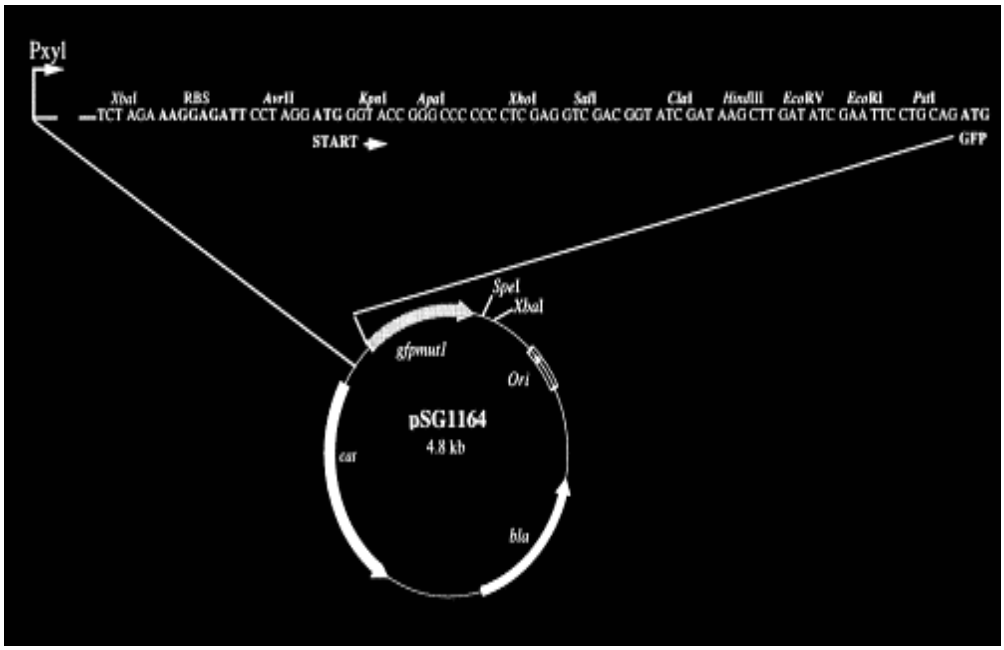
Zimmer, M. A., Szurmant, H., Saulmon, M. M., Collins, M. A., Bant, J. S. & Ordal, G. W. (2002). The role of heterologous receptors in McpB-mediated signalling in Bacillus subtilis chemotaxis. *Mol Microbiol* **45**, 555-568.

Zweers, J. C., Barak, I., Becher, D., Driessen, A. J., Hecker, M., Kontinen, V. P., Saller, M. J., Vavrova, L. & van Dijl, J. M. (2008). Towards the development of Bacillus subtilis as a cell factory for membrane proteins and protein complexes. *Microb Cell Fact* **7**, 10.

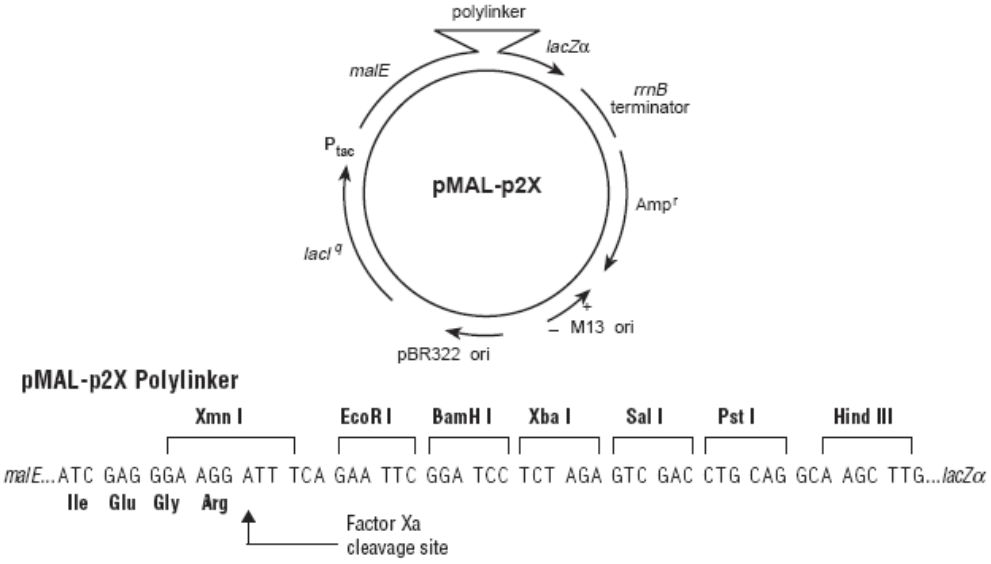
Appendix



pMUTIN4 plasmid map (Vagner *et al.*, 1998)



pSG1164 plasmid map (Lewis & Marston, 1999)



pMAL-p2X plasmid map (NEB)