Comparative Flux Control through the Cytoplasmic Phase of Cell Wall Biosynthesis

Author: Daniel James Bearup

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Mathematical Biology and Biophysical Chemistry

University of Warwick, Molecular Organisation and Assembly in Cells

Doctoral Training Centre

March, 2010

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my supervisors Michael Chappell and Christopher Dowson for their help and guidance throughout my time as a PhD student at Warwick. I greatly appreciate the help David Roper, Adrian Lloyd, Anita Catherwood, Andrew Quigley and Tom Clarke provided while I was getting used to lab work and indeed ever since. I am also grateful for the help provided by Neil Evans, Markus Kirkilionis and David Mond with the theoretical and modelling matters arising throughout this work. Tim Bugg's assistance in the acquisition of a number of chemical inhibitors of the enzymes used was also much appreciated. Finally, at least for those who helped directly with this work, I would like to thank Francois Boulier for taking time out of his holiday to explain a relatively abstruse concept in differential algebra.

It goes without saying, and yet nonetheless I feel compelled to say it, that the support of the MOAC community past and present has been pivotal to my successful completion of this work. Special thanks go to Alison Rodger, Dorothea Mangels and Monica de Lucena for handling my various administrative problems with remarkable ease. I appreciate the companionship provided by the many members of my lab in biology, C010, especially those who were forced to put up with my inadequate ball skills on the football field and my extreme smugness on the cricket pitch. And of course I must thank my wife for putting up with the back of my head over the last few months in which I have been writing this work.

Finally I must express my thanks to MOAC and more specifically the EPSRC for providing the money which made this work possible.

DECLARATION

All material contained in this thesis is my own work. No part of this work has been submitted previously for another degree. As yet no part of this work has been published in a research journal or any other format.

ABSTRACT

The introduction of antibacterial drugs in the middle of the last century heralded a new era in the treatment of infectious disease. However the parallel emergence of antibiotic resistance and decline in new drug discovery threatens these advances. The development of new antibacterials must therefore be a high priority.

The biosynthesis of the bacterial cell wall is the target for several clinically important antibacterials. This extracellular structure is essential for bacterial viability due to its role in the prevention of cell lysis under osmotic pressure. Its principal structural component, peptidoglycan, is a polymer of alternating *N*-acetyl-glucosamine (GlcNAc) and *N*-acetyl muramic acid (MurNAc) residues crosslinked by peptide bridges anchored by pentapeptide stems attached to the MurNAc moieties. The biosynthesis of peptidoglycan proceeds in three phases. The first, cytoplasmic, phase is catalysed by six enzymes. It forms a uridine diphosphate (UDP) bound MurNAc residue from UDP-GlcNAc and attaches the pentapeptide stem. This phase is a relatively unexploited target for antibacterials, being targeted by a single clinically relevant antibacterial, and is the subject of this thesis.

The *Streptococcus pneumoniae* enzymes were kinetically characterised and *in silico* models of this pathway were developed for this species and *Escherichia coli*. These models were used to identify potential drug targets within each species. In addition the potentially clinically relevant interaction between an inhibitor of and feedback loops within this pathway was investigated.

The use of direct parameter estimation instead of more traditional approaches to kinetic characterisation of enzymes was found to have significant advantages where it could be successfully applied. This approach required the theoretical analysis of the models used to determine whether unique parameter vectors could be determined. Such an analysis has been completed for a broad range of biologically relevant enzymes. In addition a relatively new approach to such analysis has been developed.

LIST OF ABBREVIATIONS

Standard notation for chemical elements and units of measurement is used throughout. Note the use of: b, Kb to indicate bases and kilobases when referring to the length of DNA strands; $\times g$ to indicate a number of (Earth) gravities; AU to denote absorbance units; and Da, Daltons, a unit of molecular mass. Mathematical notation is defined where used.

Abbreviation	Full term
A_x	absorbance at $x$nm
BDF	backward differentiation function
ADP	adenosine diphosphate
Alr	an alanine racemase encoded by *alr*
AmpD	N-acetylmuramoyl-L-alanine amidase
AmpG	a peptidoglycan permease encoded by *ampG*
AP4	adenosine tetraphosphate
APS	ammonium persulphate
ATP	adenosine triphosphate
D-Ala	D-alanine
D-Ala-D-Ala	D-alanyl-D-alanine
D-Glu	D-glutamate
DacB	a muramoyltetrapeptide carboxypeptidase encoded by *dacB*
Ddl	D-alanine-D-alanine ligase
dH_2O	deionised water
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EPUG	uridine diphosphate *N*-acetylglucosamine-enolpyruvate

Abbreviation	Full term
FAD	flavin adenine dinucleotide
FPLC	fast protein liquid chromatography
FTA	Fundamental Theorem of Algebra
GI	globally identifiable
GlcNAc	*N*-acetyl glucosamine
HMM	high molecular mass
HPLC	high performance liquid chromatography
IDH	isocitrate dehydrogenase
IPTG	isopropyl β-D-1-thiogalactopyranoside
L-Ala	L-alanine
L-DA	L-diaminoacid
L-Lys	L-lysine
LB	lysogeny broth (Luria-Bertani)
LDH	lactate dehydrogenase
LI	locally identifiable
Lipid I	undecaprenyl diphosphate *N*-acetylmuramoyl-pentapeptide
Lipid II	undecaprenyl diphosphate *N*-acetylmuramoyl-pentapeptide-*N*-acetylglucosamine
LMM	low molecular mass
Mpl	UDP-*N*-acetylmuramate: L-alanyl-γ-D-glutamyl-meso-diaminopimelate ligase
MRSA	multi-drug resistant *Staphylococcus aureus*
mDAP	2,6-diaminopimelic acid
MESG	*N*-methyl-2-thioguanosine
MMK	Michaelis-Menten kinetics
MraY	phospho-*N*-acetylmuramoyl-pentapeptide-transferase
MTG	methylthioguanine
MurA	UDP-*N*-acetylglucosamine enolpyruvyl transferase 1
MurB	UDP-*N*-acetylenolpyruvylglucosamine reductase
MurC	UDP-*N*-acetylmuramate-L-alanine ligase
MurD	UDP-*N*-acetylmuramoyl-L-alanine-D-glutamate ligase
MurE	UDP-*N*-acetylmuramoyl-L-alanyl-D-glutamate-L-lysine ligase
MurF	UDP-N-acetylmuramoyl-tripeptide-D-alanyl-D-alanine ligase

Abbreviation	Full term
MurG	undecaprenyldiphospho-muramoylpentapeptide β-*N*-acetylglucosaminyl-transferase
MurI	a glutamate racemase encoded by *murI*
MurJ	mouse virulence factor N, posited Lipid II flippase (alternative name of MviN)
MurNAc	*N*-acetyl muramic acid
MurZ	UDP-*N*-acetylglucosamine enolpyruvyl transferase 2
MviN	mouse virulence factor N, posited Lipid II flippase
N-terminus	the amine end of a polypeptide
NAD^+	oxidised form of nicotinamide adenine dinucleotide also denoted: NAD
NADH	reduced form of nicotinamide adenine dinucleotide
$NADP^+$	oxidised form of nicotinamide adenine dinucleotide phosphate also denoted: NADP
NADPH	reduced form of nicotinamide adenine dinucleotide phosphate
NagZ	a β-*N*-acetylglucosaminidase encoded by *nagZ*
ODE	ordinary differential equation
P^+	inorganic phosphate
PBP	penicillin binding protein
PCR	polymerase chain reaction
PEP	phosphoenolpyruvate
PK	pyruvate kinase
PMSF	phenylmethanesulfonylfluoride
PNP	purine nucleoside phosphorylase
PP	ping-pong
PRT	Pohjanpalo's Jacobian rank test
QSS	quasi-steady state
QSSA	quasi-steady state assumption
RMS	root mean square
RMSE	root mean square error
RNA	ribose nucleic acid
S. pneumoniae	*Streptococcus pneumoniae*
SD	standard deviation
SDS	sodium dodecyl sulfate

Abbreviation	Full term
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SGI	structurally globally identifiable
SI	structurally indistinguishable
SLI	structurally locally identifiable
SO	simple ordered
TAE	tris(hydroxymethyl)aminomethane-acetate-ethylenediaminetetraacetic acid
TEMED	tetramethylethylenediamine
tRNA	transfer ribonucleic acid
U1P	uridine diphosphate *N*-acetyl-muramoyl-L-alanine also denoted: UDP-MurNAc-L-Ala, UDP-MurNAc-monopeptide
U2P	uridine diphosphate *N*-acetyl-muramoyl-L-alanine-D-glutamate also denoted: UDP-MurNAc-L-Ala-D-Glu, UDP-MurNAc-dipeptide
U3P	uridine diphosphate *N*-acetyl-muramoyl-L-alanine-D-glutamate-L-diaminoacid (either L-lysine or 2,6-diaminopimelic acid) also denoted: UDP-MurNAc-L-Ala-D-Glu-L-DA, UDP-MurNAc-tripeptide
U5P	uridine diphosphate *N*-acetyl-muramoyl-L-alanine-D-glutamate-L-diaminoacid-D-alanyl-D-alanine also denoted: UDP-MurNAc-L-Ala-D-Glu-L-DA-D-Ala-D-Ala, UDP-MurNAc-pentapeptide
UDP	uridine diphosphate
UDP-intermediate	the UDP bound intermediates of the cytoplasmic phase of the peptidoglycan biosynthesis pathway
UDPPEE	uridine diphosphate *N*-acetylglucosamine-enolpyruvate
UGP	uridine diphosphate *N*-acetylglucosamine also denoted UDP-GlcNAc
UK	United Kingdom
UMN	uridine diphosphate *N*-acetylmuramic acid also denoted: UDP-MurNAc

Abbreviation	Full term
UMP	uridine monophosphate
UppP	undecaprenyl pyrophosphate phosphatase
URL	uniform resource locator
USA	United States of America
USD	United States Dollars
UV	ultra-violet
v/v	volume by volume
w/v	weight by volume

1. INTRODUCTION

1.1 Antibacterials and antibiotic resistance

The introduction of antibacterial drugs to the clinical arsenal in the middle of the last century heralded a new era in the treatment of infectious disease [1, 2]. However the parallel emergence of antibiotic resistance and decline in new drug discovery threatens these advances [3, 4, 5]. Consequently making progress on the development of new antibacterials must be seen as a critical priority for the academic and pharmaceutical communities and is the subject of this thesis.

The development of antibiotic resistance is inevitable once a drug enters widespread use [6]. The prevalence of these drugs creates selective pressure encouraging the evolution of resistant strains [7]. Furthermore, where broad spectrum antibiotics are used, the elimination of competing bacterial flora provides an ideal environment for the proliferation of such strains [7]. While resistant strains were initially encountered in nosocomial infections they are increasingly observed in community acquired infections [8]. Worldwide reports of emerging multidrug-resistant *Mycobacterium tuberculosis*, the bacterium responsible for tuberculosis, provides a worrying example of the scope of this problem [9, 10, 11].

Historically the majority of new antibacterials have been developed by modification, or synthesis of analogues, of natural products [7, 12]. Despite this, still rich, source of antimicrobials, new drug discovery has declined in the last thirty years [5, 12]; there was a gap of almost four decades before the introduction of a new class of antimicrobials (oxazolidinones) after 1962

(quinolones and streptgramins) [2]. The causes of this trend are by no means simple, however the economics of drug discovery must be considered a significant factor [1]. With the cost of drug development rising four fold between 1987 and 2000 [1]; antibacterials typically requiring only short treatment periods; and high attrition within the antibacterial development pipeline, this area of research is an increasingly unattractive proposition for the pharmaceutical industry [13].

Little can be done about the effectiveness of antibacterial therapy; however reducing the cost and attrition in antibiotic development may be achievable. The use of *in silico* models to design inhibitors based on crystal structures may reduce the cost of drug discovery [14]. Furthermore the ability to model increasingly complex biological systems may allow the effectiveness of a particular inhibitor to be assessed, thus further reducing cost and attrition [15, 16]. The focus of this thesis is the development of an *in silico* model of the cytoplasmic phase of the peptidoglycan biosynthesis pathway; a potential antibacterial target [17].

1.1.1 Antibacterial mode of action

Antibacterials are drugs which interfere with a process or structure that is essential for bacterial viability without compromising the eukaryotic host. There are presently three main antibacterial targets [7]: bacterial DNA transcription and replication; bacterial protein synthesis; and cell wall biosynthesis. These areas are briefly discussed in the following section. Other potential modes of action, such as the disruption of cell membranes [18], exist but are not considered here.

1.1.1.1 Bacterial DNA transcription and replication

In order to allow transcription and replication of DNA, the two strands of the double helix must be separated [19]. This process is mediated by topoisomerases which cleave and then recombine one or both DNA strands to release tension when the DNA is uncoiled. Bacterial cells possess

unique topoisomerases, not found in eukaryotes, such as DNA gyrase and topoisomerase IV which are the targets of the fluoroquinolone family of antibacterials [19, 20]. These drugs form a complex with cleaved DNA and the topoisomerase inhibiting religation and causing a fatal accumulation of double-strand breaks [7]. Transcription of DNA is also inhibited using the rifamycin family of drugs [21]. These drugs bind to the β subunit of bacterial RNA polymerase. The nitroimidazole group of drugs is capable of oxidizing DNA and proteins causing damage to both [22, 23].

1.1.1.2 Bacterial protein synthesis

The bacterial protein synthesis mechanisms are sufficiently different from their eukaryotic counterparts to provide numerous targets for antibacterials. The focal point of protein synthesis is the ribosome, a ribonucleoprotein assembly of two subunits, 30S and 50S, which associate upon initiation of protein synthesis [24]. Inhibitors of protein synthesis target either distinct subunits of the ribosome or the various steps in protein synthesis. Examples include: the macrolides, preventing translocation; tetracyclines, which compete with tRNA; and aminoglycosides which cause misreading of mRNA [7].

1.1.1.3 Cell wall biosynthesis

The bacterial cell wall is essential for preventing cell lysis under osmotic pressure and thus for bacterial survival. The principal source of mechanical strength in the structure is a polymer of alternating glycan residues crosslinked by short peptide bridges called peptidoglycan [25]. It is associated to the extracellular face of the plasma membrane. The biosynthesis of this macromolecule is described in detail in Section 1.2. The major antibacterials used in human medical treatment known to target this process are presented in Table 1.1.

Antibacterial	Mode of action
Fosfomycin [26]	The *N*-acetylmuramic acid subunit of peptidoglycan is derived from *N*-acetylglucosamine by the addition of a lactic acid substituent derived from phosphoenolpyruvate. The pyruvyl transferase enzyme involved is MurA and is inhibited by fosfomycin.
D-Alanine analogues [27, 28]	The terminus of the pentapeptide sidechain is a dipeptide of two D-Ala moieties. This dipeptide is produced by racemisation of L-Ala to D-Ala and subsequent ligation by a D-Ala-D-Ala ligase, Ddl. The racemisation reaction is inhibited by a number of D-Ala analogues. Both reactions can be inhibited by D-cycloserine, an antibiotic used in the treatment of tuberculosis.
Ramoplanin and analogues [29]	Ramoplanin is a lipoglycodepsipeptide active against Gram-positive bacteria. It inhibits the uptake of *N*-acetylglucosamine by growing cells with a resultant accumulation of UDP-MurNAc-pentapeptide (U5P). Inhibition is of *N*-acetylglucosaminyl-transferase, MurG, that adds *N*-acetylglucosamine to the undecaprenyl-muramyl-pentapeptide.
Bacitracin [30]	The lipid carriers involved in transporting the disaccharide subunits across the cell membrane are C_{55} isoprenyl phosphates. In the transport process the lipid acquires an additional phosphate moiety and must be dephosphorylated to continue transport. Bacitracin binds to the isoprenyl pyrophosphate preventing dephosphorylation.
β-Lactams [31]	β-Lactams inhibit the transpeptidase domain of penicillin binding proteins, PBPs, by forming an acyl-enzyme complex which is very slow to degrade. This inactivates the PBPs preventing the crosslinking reactions which provide rigidity to peptidoglycan.
Glycopeptides[32]	These molecules bind with high affinity to the uncrosslinked D-Ala-D-Ala termini of the MurNAc-pentapeptide moieties. This steric hindrance prevents crosslinking with similar effect to that of the β-lactams.

Table 1.1: Medically relevant classes of antibacterials that inhibit cell wall synthesis and their mode of action.

1.1.2 Mechanisms of antibacterial resistance

Bacteria have evolved a variety of antibacterial resistance mechanisms broadly categorised into the following groups: reduction of antibacterial concentration in the cytoplasm; inactivation or modification of the antibacterial; and alteration of the antibacterial target [7]. Validated responses to these antibacterial resistance strategies are also discussed where they have been developed.

1.1.2.1 Reduction of antibacterial concentration in the cytoplasm

The control of metabolite concentrations within the bacterial cell is necessary for viability. Naturally it requires mechanisms for influx and efflux of metabolites. Given that the majority of antibacterials target intracellular processes these mechanisms are a natural area in which antibacterial resistance can evolve [33].

Gram-negative bacteria have two cell membranes encasing the bacterial cell wall, Figure 1.1. The outer membrane is composed principally of lipopolysaccharides which carry a net negative charge. This membrane significantly reduces penetration by hydrophobic molecules in comparison to inner phospholipid bilayers, providing a passive defence against certain antibiotics, such as the glycopeptides [34]. Influx of hydrophobic compounds to these bacteria is dependent on specialized porins. Where antibiotics take advantage of these paths bacteria may adapt by reducing or eliminating formation of the porins used [35]. Alternatively the structure of the porin can be altered preventing efflux of the challenging antibacterial [36].

Most Gram-positive organisms are less well protected from penetration by antibiotics, Figure 1.1, [37]. The peptidoglycan macromolecule does not typically pose a barrier to molecules smaller than 50000Da. However modifications to the cell wall structure can reduce penetration by antibiotics; as demonstrated by the intermediate vancomycin-resistance phenotype of *Staphylococcus aureus*. Here increasing the cell wall thickness and incorporating unamidated

glutamine allows the cell wall to bind more vancomycin, preventing it from reaching its target [38].

Efflux pumps, ATP or proton motive force dependent membrane proteins, are found in all bacteria and are used for the transport of lipophilic and amphiphatic molecules [39]. They also prevent the accumulation of fatal concentrations of antimicrobials in antimicrobial producing species [40]. Proteins of this type are implicated in numerous cases of clinically relevant antibacterial resistance [41, 42]. Inhibitors of these pumps, intended to reverse this resistance, are currently under development [40, 43].

1.1.2.2 Inactivation or modification of the antibacterial

The two classical examples of this strategy are hydrolysis of the β-lactams by β-lactamases and enzymatic modification of the aminoglycosides [7]. The β-lactamases may represent a modification of the PBPs, sharing as they do similar functional structure and binding affinity to the β-lactams [44]. However these enzymes hydrolyze the β-lactam ring much more rapidly than the PBPs inactivating these drugs [45]. The spread of β-lactamase genes via mobile genetic elements, such as plasmids, has allowed the rapid spread of this form of resistance and its integration with other resistance strategies, yielding multidrug resistant strains [46]. The use of β-lactamase inhibitors to restore the efficacy of β-lactams has yielded some success and is of great potential interest [47].

The aminoglycosides are not subjected to hydrolysis. They are instead modified by the addition of acetyl, nucleotidyl and phospho groups by appropriate transferases decreasing their affinity to their target, ribosomal RNA, [48]. The addition of these groups is necessary to control processes within the cell indicating a route by which this resistance strategy may have developed [49].

While inactivation of the antibiotic is sufficient for antibacterial resistance, it is unlikely that this completes the metabolism of the drug. It has been shown that some species of soil bacteria

can subsist using certain antibiotics as their sole carbon source [50]. While soil bacteria have historically been a source of antibacterials, they are also likely to be a source of antibacterial resistance genes.

1.1.2.3 Alteration of the antibacterial target

Alterations to the protein machinery necessary to the viability of the cell may be risky however it has proven effective in introducing antibacterial resistance. Mutation of PBPs producing enzymes with low affinity to β-lactams is another source of resistance to these drugs [7]. The MRSA phenotype has developed by this route [51].

Vancomycin resistance can also be attributed to this type of strategy. Vancomycin binds to the D-Ala-D-Ala terminus of the MurNAc pentapeptide moiety within peptidoglycan. When an alternative dipeptide is used, D-Ala-D-Lac, the hydrogen bonding network required by vancomycin is disrupted drastically reducing the effectiveness of this drug [52].

1.1.3 The cost of antibacterial resistance

The cost of antibacterial resistance from the economic and health care perspectives has been much discussed, for examples see [3, 45, 53, 54]. It has been observed that antibiotic resistant infections result in higher mortality, prolonged hospitalisation, and higher health care costs relative to antibiotic susceptible bacteria [55]. Three studies, summarised in [54], show significant increases in these measures, a five fold increase in mortality in one case, where infections were resistant. The broader economic cost of antibacterial resistance to the USA economy was estimated at between 0.1-30 billion USD in 1989 [56]; follow-up of this study would appear to be lacking. Nonetheless what data are available indicates that the development of new antibacterials is justified economically and socially.

1.2 Construction of the bacterial cell wall

There are at present six major classes of antibacterials targeting the construction of the bacterial cell wall, Table 1.1. However as yet many of the stages within the pathway are not targeted by effective antibacterials [28]. Of particular interest, in the context of this thesis, are the inhibitors of cytoplasmic phase enzymes which have been discovered, but as yet not developed as antibacterials [57]. Some additional inhibitors of the amino acid ligases have been developed since that review, for example [58, 59, 60, 61, 62, 63].

Several of these inhibitors, specifically targeting MurC or MurD, were shown to have antibacterial effect *in vitro* [62, 63]. These inhibitors demonstrate that inhibition of this section of the pathway can have antibacterial effect, but raises the question of why other inhibitors have as yet proved ineffective. Passage through the cell membrane may be a problem in the case of the phosphinate inhibitors, for example [64]. Competitive inhibition can be overcome by sufficient accumulation of the substrates with which the inhibitor competes. This effect has been termed metabolic resistance and may also be a factor [65, 66, 67]. It should be possible to study this effect through *in silico* modelling of the pathway.

The structure of the bacterial cell wall is reviewed briefly in the next section. In the following sections the peptidoglycan biosynthesis pathway is covered. The cytoplasmic phase, being the focus of this thesis, is covered in the most detail. The remaining membrane bound phases and remodelling and recycling phases are then covered briefly.

1.2.1 Cell wall structure

Bacteria are broadly classified as Gram-positive or Gram-negative based on their reaction to a particular iodine-dye complex. Staining with this reagent is dependent on cell wall structure. Where Gram-positive bacteria have a thick (20 - 80nm) peptidoglycan layer which reacts strongly with Gram's reagent, Gram-negative bacteria have a relatively thin layer of pepti-

doglycan surrounded by a second cell membrane which excludes the dye, Figure 1.1 [68, 69]. As observed above, Section 1.1.2.1, these differences significantly affect the bacterial response to antibiotics. However, the changes in the peptidoglycan biosynthesis pathway necessary to support these differences are as yet uncharacterised.

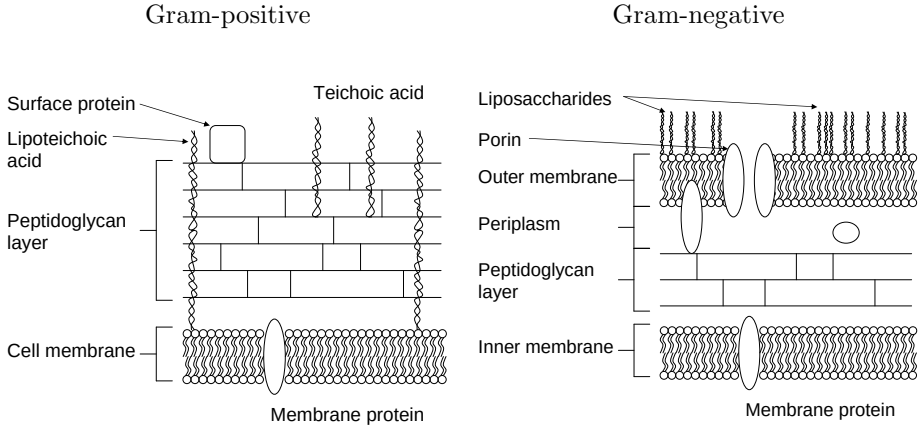


Figure 1.1: Schematic diagram of the Gram-positive and Gram-negative extracellular structure. In both species a peptidoglycan layer forms the structural basis of the cell wall, a thicker layer being present in Gram-positive species. Gram-negative species have two cell membranes, Gram-positives a single membrane. Lipoteichoic acids are found only in Gram-positive species. They are lipolyated polysaccharides which extend through the peptidoglycan layer to the cell surface; they act as antigenic determinants. Lipopolysaccharides are found only in Gram-negative species and act both as a structural component of the outer membrane and as antigens.

Living cells concentrate metabolites and cellular machinery within a confined volume. This results in a significant pressure differential across the cell membranes. Without structural support this pressure would result in osmotic lysis and consequently cell death. In bacterial cells, the cell wall, specifically the peptidoglycan layer, is responsible for satisfying this need [70]. In Gram-positive bacteria it also provides an anchoring point for a variety of cell surface molecules [71].

Peptidoglycan is composed of a linear glycan backbone of alternating *N*-acetylglucosamine, GlcNAc, and *N*-acetylmuramic acid, MurNAc, subunits. This backbone is common to all bacterial peptidoglycans and has only minor variations between species. In *S. aureus* up to half of the MurNAc moieties have an *O*-acetyl group on the sixth carbon which is known to

provide resistance to lysozyme and is implicated in resistance to β-lactams [72]. The glycan strands found in a given cell wall are of varying lengths, average lengths vary between species [73, 74, 75].

Pentapeptide stems extend from the carboxyl group of each MurNAc moiety. These stems are crosslinked either directly or by short peptide bridges, dependent on bacterial species [71]; transforming the glycan strands into a rigid mesh. The typical peptide sequence of this stem is L-Ala-D-Glu-X-D-Ala-D-Ala; where X is an L-amino acid with an amino sidechain. The alternating stereochemistry of the amino acids is unique and prevents attacks by proteases. The third peptide in the stem is typically L-Lys in Gram-positive species and 2,6-diaminopimelic acid, mDAP, in Gram-negative species [76]. Variation of the other peptides is less common but is implicated in a number of resistance strategies. For example alteration of the terminal D-Ala to a D-serine or D-lactate is associated with weak or strong vancomycin resistance respectively [77, 78].

Crosslinks are typically formed between the third and fourth peptides of two stems. In general if the third peptide is mDAP this link is direct, if it is L-Lys the link is indirect [76]. The degree of crosslinking can vary substantially depending on which penicillin binding proteins (PBPs) are present. In certain cocci, *S. aureus* for example, more than 90% of the peptide side chains are cross-linked, while in bacilli only 20-50% are cross-linked [79].

1.2.2 The cytoplasmic phase of cell wall biosynthesis

The first committed steps in the cell wall biosynthesis pathway occur in the cytoplasm, Figure 1.2 [17]. Two enzymes, MurA and MurB, convert UDP-GlcNAc (UGP) into UDP-MurNAc (UMN). The pentapeptide stem is then constructed by four amino acid ligases, MurC, MurD, MurE and MurF. The substrates of this pathway are used in a variety of metabolic functions with the exception of D-Glu and D-Ala-D-Ala which are formed by MurI, glutamate racemase, and Alr, alanine racemase, and DdlA/B, the D-alanine ligases.

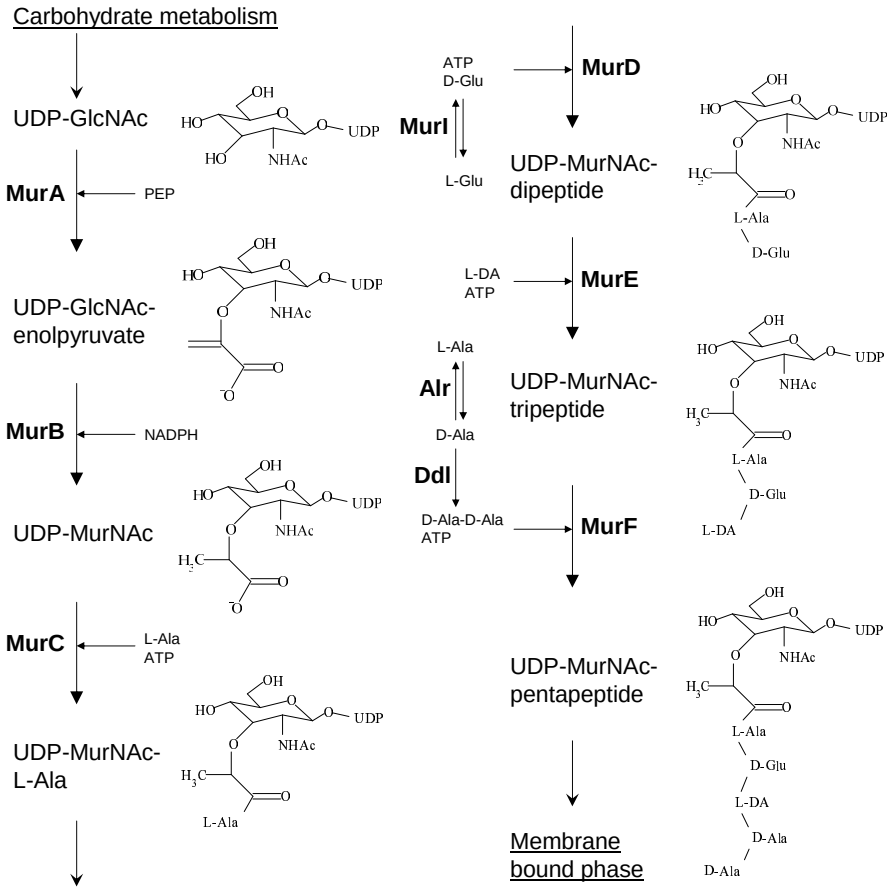


Figure 1.2: Schematic diagram of the cytoplasmic phase of cell wall biosynthesis, [17]. L-DA denotes an L-diamino acid.

1.2.2.1 Formation of UDP-MurNAc

The hexosamine biosynthesis pathway, a subsection of carbohydrate biosynthesis, converts D-fructose-6-phosphate to UDP-GlcNAc, which is then incorporated into several macromolecules: peptidoglycan, lipopolysaccharides, and teichoic acids in bacteria [80]. The first committed step of peptidoglycan biosynthesis is the addition of the enolpyruvate group from phosphoenolpyruvate, PEP, to UDP-GlcNAc to give UDP-GlcNAc-enolpyruvate (UDPPEE). This reaction is catalysed by UDP-*N*-acetylglucosamine enolpyruvyl transferase, MurA and MurZ [81, 82]. The *E. coli* and *S. pneumoniae* forms of this enzyme have been kinetically characterised on steady state and pre-steady state time scales for a range of experimental conditions [82, 83, 84].

The reduction of the vinyl bond of UDPPEE, catalysed by UDP-*N*-acetylenolpyruvylglucosamine reductase, MurB, yields UDP-MurNAc [85]. The reaction proceeds via a hydride transfer mechanism transferring electrons from NADPH to FAD, a tightly bound cofactor of MurB, and finally to the unsaturated substrate. The *E. coli* and *S. pneumoniae* forms of this enzyme have been kinetically characterised on steady state and pre-steady state time scales [85, 86]. *In vitro* reconstructions of the peptidoglycan biosynthesis pathway suggest that it is possible to produce an unreduced form, UDPPEE-pentapeptide, by omitting MurB.

1.2.2.2 Structure and catalytic mechanism of MurA and MurB

Crystal structures of MurA from *E. coli* and *Enterobacter cloacae* reveal two globular domains connected by a double stranded linker [87, 88]. The folds of these two domains are very similar, three parallel α-helices enclosed in three further helices and three β-sheets. This structure closely resembles that of 5-enolpyruvylshikimate-3-phosphate synthase, the only other known enolpyvuyl transferase.

UGP is constrained at the catalytic site by hydrophobic associations and hydrogen bonds between the uridine ring and regions of these two domains. The result is highly specific binding to the uridine ring. However since the cytoplasmic phase intermediates all contain this moiety some may also be bound by MurA with possibly inhibitory consequences. A covalent adduct can be formed between PEP and an active site cysteine [84]. This is exploited by the antibiotic fosfomycin [89, 90]. The antibiotic has no effect on *M. tuberculosis* MurA which lacks an active site cysteine [91, 92].

The reaction proceeds by ordered formation of a tetrahedral intermediate, with UGP binding first followed by PEP [84]. The subsequent elimination of phosphate completes the reaction. The binding order proposed is substantiated by the significant conformation changes observed on binding of UGP. The resulting closed state facillitates the binding of PEP [93].

Crystal structures of MurB show it to be composed of three domains of mixed $\alpha + \beta$ structure [94]. In the absence of UDPPEE the third domain swings closer to the other domains, consequently leaving the enzyme in a more closed conformation. However this movement opens the channel through which the substrate accesses the active site.

Kinetic studies suggest that the mechanism should be described by a ping-pong model [95]. This observation is substantiated by NMR studies which indicate that both substrates use the same binding pocket [96]. Furthermore these results are consistent with the bound FAD cofactor mediating the two electron transfer from NADPH to UDPPEE. The overall reaction is composed of two stages, first the NADPH substrate binds and is oxidised, reducing FAD to $FADH_2$. The release of NAD^+ then allows binding of UDPPEE which is reduced to UMN while the bound cofactor is oxidised returning it to its original state.

Substrate inhibition has been observed for both substrates. The inhibitory effect created by UDPPEE is much stronger than that produced by NADPH [95]. Given the mechanism described above, this likely arises from the formation of a nuisance complex containing the wrong substrate for a given oxidation state of FAD.

1.2.2.3 Construction of the pentapeptide stem

The pentapeptide stem is appended to the lactyl group of UDP-MurNAc by the sequential action of four amino acid ligases: MurC, MurD, MurE, and MurF, Figure 1.2 [57]. These enzymes require a Mg^{2+} cofactor and use ATP. The amino acid sequences of the *E. coli* enzymes share 10-20% sequence identity, suggesting that these enzymes may be evolutionarily related [70]. A comparison of these enzymes across seven bacterial species found them to share a common ATP binding sequence and six separately conserved amino acids [97]. The reactions catalysed by these enzymes share a common mechanism [57].

Variability of the pentapeptide stem between and within species has been observed. Specifically

in Gram-positive species typically L-Lys is found in the third position, while in Gram-negative species m-DAP is usually found here. This is specified by the form of MurC encoded by the species. As has been noted previously alternatives to the classical D-Ala-D-Ala terminus are found in vancomycin resistant *S. aureus*. MurF has been shown to utilise a number of alternative dipetides so alternative forms of MurF are not needed in this case.

Each of the *E. coli* ligases has been kinetically characterised for steady state time scales [98, 99, 100, 101, 102]. The results are summarised in Table 7.4.

1.2.2.4 Structure and catalytic mechanism of the amino-acid ligases

Crystallisation of MurC has shown that it exists in equilibrium between monomeric and dimeric states. No kinetic differences have been observed between these two forms [103]. As noted above the amino-acid ligases show a high degree of sequence homology. These similarities are reflected in the crystal structures obtained for MurC, MurD, MurE and MurF [104].

Each of the amino acid ligases are composed of three globular domains each of which contains the binding pocket for a specific substrate. The ATP binding pocket is located on the central domain and is closely associated with a conserved glycine rich loop which resembles the canonical mononucleotide P-loop observed in ATP and GTP binding proteins. This domain also contains Mg^{2+} binding motifs, indicating that it is responsible for coordinating these essential ions. The central domain is the mostly closely conserved of the three domains across the amino-acid ligases.

The C-terminal domains are also essentially topologically conserved and have a Rossmann-type fold. However these domains lack the dinucleotide binding motif usually associated with such folds and are believed to be responsible for binding the amino acids. They also contribute some residues to securing ATP.

Despite these commonalities each enzyme is highly specific to an amino acid/UDP-intermediate

combination. This specificity is thought to be determined by the N-terminal domains of each ligase which display much greater sequence variety [17]. The N-terminal domain is responsible for binding the UDP-intermediate. The structure of this domain takes two forms, one in MurC and MurD and a second in MurE and MurF. This reflects alternative means of securing the UDP-intermediate. MurC and MurD rely on securing the uridine ring by means of hydrophobic regions and hydrogen bonds, in a manner similar to that employed by MurA. However the increased length of the amino-acid stem associated with the substrates of MurE and MurF may render this approach intractable. Regardless for these enzymes the UDP-intermediate appears to be secured by a hydrogen bonding loop which directs the end of the stem towards the binding pockets for ATP and the amino acid. The end of the stem appears to form associations with domains 2 and 3 for these enzymes [104].

The four enzymes appear to take a relatively open conformation in the absence of substrates. This is believed to contract to a closed form upon substrate binding to shelter the catalytic site. In MurF this is particularly pronounced [105].

The reaction catalysed by each ligase is believed to proceed as follows. An acylphosphate intermediate is formed by ATP and the nucleotide substrate stabilised by the enzyme. Nucleophilic attack by the amino group of the condensing amino acid or dipeptide follows releasing phosphate and forming the peptide bond [106, 107]. In MurD the binding order has been shown to be: ATP, UDP-MurNAc-L-Ala, D-Glu [108]. MurD and MurE derived from *E. coli* and *S. pneumoniae* have been shown to produce adenosine tetraphosphate in the absence of amino acid substrate [109, 110]; MurC does not.

1.2.2.5 Production of D-amino acids

D-alanine and D-glutamic acid are directly synthesised from their respective L- forms by the racemases Alr and MurI [111, 112]. An alternative route for D-Glu formation has been observed catalysed by a D-Ala aminotransferase [113]. The D-alanine dipeptide is formed by an ATP

dependent D-Ala:D-Ala ligase, Ddl [114]. Inhibitors of these enzymes have been discovered which have antibacterial activity, D-cycloserine being clinically used antibiotic [27, 115]. Kinetic characterisations of these enzymes have yet to be undertaken.

1.2.2.6 Data available for pathway modelling

As noted above each of the enzymes MurA to MurF from *E. coli* have been kinetically characterised. However the characterisations were undertaken under a variety of different experimental conditions; the most significant difference being the range of temperatures from room to 37^{o}C. These differences are likely to have a profound effect on the observed maximal rates of reaction catalysed by each enzyme (denoted k_{cat}). As such these kinetic constants are not ideally suited for use in a predictive *in silico* model of this pathway; parameters collected under more consistent conditions would be preferred.

Concentrations of many of the substrates required by this pathway have been estimated *in vivo* for *E. coli* for a variety of growth conditions [116, 117, 118]. As yet similar estimates of enzyme concentrations are unavailable. These concentrations will eventually be necessary for the validation of *in vivo* models of pathway dynamics.

1.2.3 Membrane bound phases

The final two phases of peptidoglycan biosynthesis are membrane anchored; either to the intra- or extracellular face of the inner membrane, Figure 1.3 [119]. On the intracellular face the phospho-MurNAc-pentapeptide moiety of UDP-MurNAc-pentapeptide (U5P) is transferred to a lipid carrier, undecaprenyl-phosphate, producing Lipid I and releasing uridine mono-phosphate, UMP. This reaction is catalysed by translocase I, MraY, an integral membrane protein with ten membrane spanning helices [122]. A condensation reaction, catalysed by translocase II, MurG, transfers the GlcNAc moiety from UDP-GlcNAc to the C4 hydroxyl group of Lipid I forming Lipid II. MurG is not an intrinsic membrane protein; instead a hydrophobic patch on

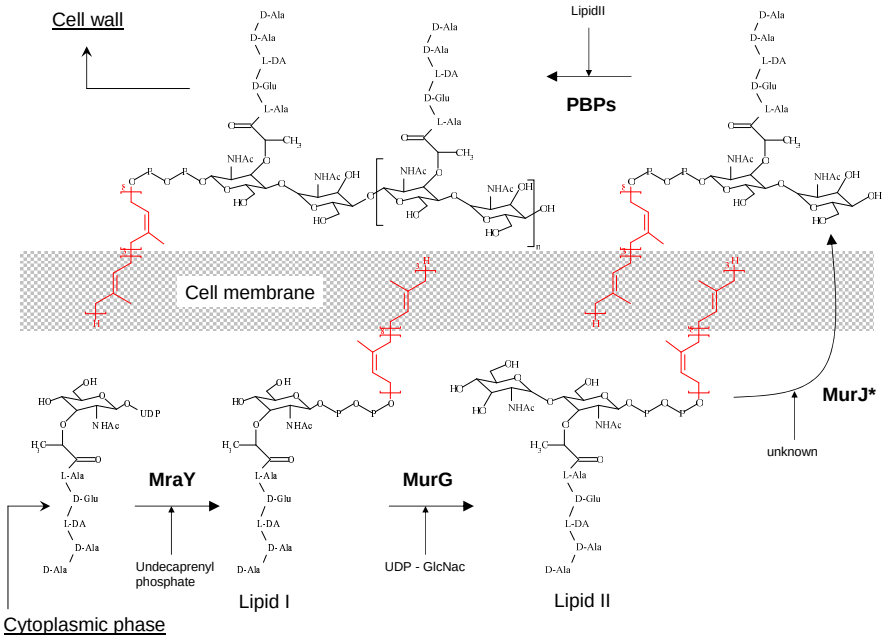


Figure 1.3: Schematic diagram of the membrane bound phases of cell wall biosynthesis, [119]. The putative flippase MurJ is as yet uncharacterised [120, 121]

its N-terminal domain anchors it to the cell membrane [123]. Lipid II is then transferred to the extracellular face of the cell membrane. The membrane protein MviN, renamed MurJ by some authors, is thought to be a flippase responsible for this transfer [120, 121].

On the extracellular face PBPs are responsible for forming the mature peptidoglycan [124]. Further condensation reactions, catalysed by transglycosylase domains of the PBPs, form the glycan backbone from multiple molecules of Lipid II. The undecaprenyl-diphosphate released is dephosphorylated and returned to the intracellular face of the cell membrane by UppP [119]; this reaction is inhibited by bacitracin. Transpeptidase domains catalyse the crosslinking of adjacent pentapeptide sidechains in a wide variety of conformations [76]; these steps are inhibited by the β-lactam family of antibacterials.

PBPs can be classified into two groups: high molecular mass (HMM) and low molecular mass (LMM). HMM PBPs consist of a cytoplasmic tail, a transmembrane anchor and two extracellular domains responsible for peptidoglycan biosynthesis or remodelling [124]. LMM PBPs are involved in cell separation, peptidoglycan maturation or recycling.

Kinetic characterisations of MraY, MurG and a number of class A PBPs have been undertaken for enzymes from a number of species [125, 126, 127, 128, 124]. The ratios of UDP-MurNAc-pentapeptide to Lipid I and Lipid II were estimated at 300:1:3 and 140:1:2.7 in *E. coli* [129, 117]. As such based on the estimated levels of UDP-MurNAc-pentapeptide the concentrations of Lipid I and Lipid II are quite low *in vivo* [116]. Given the high k_{cat}s reported and the relative abundance of undecaprenyl phosphate this suggests that the utilisation of these intermediates is quite efficient [119]. However the k_{cat}s reported for PBPs are quite low suggesting that either these enzymes are present at significantly higher concentrations than MraY and MurG or that other classes of PBPs are more active.

It should be noted that the kinetic characterisations undertaken assume typical Michaelis-Menten kinetics for these reactions. This assumption is likely to be inappropriate given that, for each reaction, one substrate and the catalyzing enzyme are membrane bound and are thus less mobile than the second substrate. This may significantly alter the appropriate model for the kinetics of these reactions [130].

1.2.4 Remodelling and recycling of peptidoglycan

The expansion of the bacterial cell wall, prior to replication for example, requires the peptidoglycan mesh to be cleaved to provide a location in which to expand [131]. Enzymes that catalyse the breakdown of peptidoglycan are found in all bacteria with cell walls and are referred to as murein hydrolases [132], some are also LMM PBPs [124]. There is a murein hydrolase for almost every bond of *E. coli* peptidoglycan: some of which accepting only the intact high molecular weight peptidoglycan while others are active only on soluble degradation products [131]. Some of these enzymes are potentially autolytic, disruption of their tight regulation processes is a potential target for antibacterials [133].

The activity of these processes results in a large number of potential peptidoglycan precursors. In *E. coli*, and presumably other Gram-negative organisms, the disaccharides released are

subject to an efficient recycling pathway [134]: losing only 6-8% per cell cycle in contrast to Gram-positives which lose between 25-50% [135]. Uptake of these molecules is via the AmpG permease; recycling then proceeds through a number of cytoplasmic steps. The AmpD amidase releases the tri-peptide (L-Ala-D-Glu-L-DA) which can then be attached to UDP-MurNAc by the Mpl ligase [136]; the specificity of AmpD is sufficient to ensure it does not degrade the UDP-MurNAc-tripeptide and downstream products which are part of the cytoplasmic phase of biosynthesis [131]. The disaccharide is hydrolysed by the *N*-acetylglucosaminidase NagZ to GlcNAc and 1,6-anhydroMurNAc [137]; these residues are then fed back into the hexosamine biosynthesis pathway. Inhibition of NagZ is being investigated as a possible means to reverse β-lactam resistance [138].

1.3 In silico modelling of metabolic pathways

While *in vitro* reconstruction of large metabolic pathways is possible obtaining detailed quantitative data becomes increasingly difficult as more reactions, and thus species to be monitored, are added. For sufficiently complex systems it is preferable to construct an *in silico* model of the pathway which can then be validated against specific *in vitro* and *in vivo* experiments to ensure that it remains representative of experimental pathway dynamics. Such a model provides the additional benefit of providing a low cost means to identify interesting interactions between inhibitors, knockouts, and other means by which a pathway can be challenged which can then be validated experimentally [16].

In order to construct such a model a variety of data are required dependent on the complexity of the desired model. To fully understand a metabolic pathway it is necessary to model everything from levels of gene expression and transcription; to the dynamics of the species within the pathway; to uptake and efflux of the raw materials and final products. Thus each of these stages needs to be modelled and characterised so that they can be interpreted mathematically

[15].

Such a model of the peptidoglycan biosynthesis pathway would dramatically exceed the scope of this thesis. Instead a model of the species dynamics of the cytoplasmic phase of this pathway will be developed.

1.3.1 Mathematical modelling of metabolism

Constructing a mathematical model of any system always involves a trade-off between amount of detail and ease of analysis. The mathematical model can only ever be an approximation of the true system. It is necessary to decide how detailed, and thus how complex, the mathematical model must be in order to provide the information required. Furthermore for the model to be useful it must be possible to analyse it, either symbolically or through numerical solutions. The difficulty of this analysis is directly related to model complexity.

In mathematical modelling of metabolism, an inherently complex and multiscale system of interactions, this trade-off is of particular importance. It is feasible to model the metabolism for a whole cell under certain simplifying assumptions, specifically that the dynamics of internal processes establish a steady state very rapidly. It is then possible to consider various phenotypes, characterised by a specific steady state, and how they are affected by changes in external variables, i.e. the introduction of a drug. Such models are broadly categorised as stoichiometric models and have been used for a wide variety of purposes [139].

Such models are however limited in that they do not consider the internal dynamics, the transient behaviour, of the system. Consequently they neglect the control structures that regulate these processes. Models constructed by coupling the kinetics of individual processes provide a means to study these details, however they are inherently more complex than stoichiometric models. Consequently they are typically used to model pathways within the overall metabolism rather than the global metabolic state [140]. Determining parameters for such models experi-

mentally is more complex than it is for stoichiometric models, which can be constructed based on observed levels of substrates within the cell. Typically the kinetics of individual processes are characterised in isolated, *in vitro* experiments. Consequently the results must be treated with caution given that *in vivo* behaviour may differ from that observed *in vitro*. For example spatial effects are typically ignored, the system is simply assumed to be well mixed. While in experiments this is often true, *in vivo* we would expect compartmentalisation to play a part in most aspects of metabolism.

Use of this approach must also overcome the problem of modelling of fluxes into and out of the pathway of interest. Such fluxes arise from other pathways yet it is not feasible to construct such detailed models of all the possible pathways that may feed into a pathway of interest. A possible solution is the introduction of cybernetic components, relatively simple functions which are designed with a particular goal which is thought to be a function of that pathway [140].

It is this sort of model which is most appropriate in this case. The subsection of the pathway to be considered is relatively small and its dynamics and regulatory processes are of particular interest. It is these details which would be expected to most directly influence the effect of an inhibitor on a pathway, assuming of course that the inhibitor will be taken up *in vivo*.

Models of the individual reactions could be constructed from deterministic or stochastic differential equations [141, 142]. Stochastic models are most appropriate for modelling processes involving very small amounts, in the tens of particles, of substrate or enzyme [143]. As such, within the context of peptidoglycan biosynthesis, stochastic techniques could be best applied in modelling the expression of enzymes and possibly in the membrane bound phases. The cytoplasmic phase reactions, which have been observed to have high substrate pool levels, are best modelled using deterministic models [141].

Thus the key requirements for construction of this model are kinetic characterisations of the enzymes MurA to MurF for given species of bacteria; *in vitro* methods to monitor species within

the pathway; and ultimately estimates of the *in vivo* levels of these species and enzymes.

1.3.2 Kinetic characterisation of enzymes

The steady state kinetic characterisation of enzymes by traditional techniques is time consuming, requiring the measurement of initial rates for a range of combinations of substrate concentrations [144]. Time courses measured on pre-steady state time scales are typically analysed differently, by direct estimation of parameters for a selection of exponential forms, see for example [145]. The resulting parameters can then interpreted in the context of the reaction mechanism.

In light of improving computational resources it is natural to extend this approach to parameter estimation using a full recorded time course in combination with a mechanistic model of the reaction. Such a technique has been demonstrated with some success [146, 147]. However it has not entered general usage in the biochemistry community: see for example [82, 83, 85, 86], [98] - [102], all of which have been published since the works cited above. Nonetheless this approach makes more efficient use of the data obtained and consequently is less time consuming. As such, given the need for large scale kinetic characterisation as a precursor to pathway modelling, the development of this approach seems appropriate. The parameter estimation techniques used are outlined in Section 3.4.

If the results obtained from this approach are to be meaningful it is necessary to determine whether the parameters determined are unique and furthermore whether the model used is appropriate. Structural identifiability and indistinguishability are concerned with determining whether two or more sets of parameters or different candidate models can give rise to the same output [148, 149, 150]. These fields and the tools deriving from them are described in Sections 3.2 and 3.3.

1.3.3 Numerical solution of differential equations

The use of numerical methods to solve differential equations is well developed. Typical approaches rely on the interpolation of an approximation of the differential equations describing the system over time or spatial steps. The size of these steps, and the accuracy of the approximation, determine the accuracy of these methods. As such the method that should be used is dependent on the complexity of the system and the degree of accuracy required. A brief outline of the techniques is given in Section 3.5.

1.4 Goals

The principal goal of this work is to investigate the fluxes through the cytoplasmic phase of the peptidoglycan biosynthesis pathway; in order to identify potential drug targets and investigate the effects of and interaction between pathway inhibitors. Differences in these dynamics between Gram-positive and Gram-negative pathway are of additional interest. This is achieved by the construction of an *in silico* model of this pathway.

The enzymes, MurC to MurF, are as yet kinetically uncharacterised for a Gram-positive species; thus these enzymes must be characterised prior to construction of the model. In addition since, for the purposes of pathway modelling, it is best that kinetics be obtained under a consistent set of initial conditions MurA and MurB will also be characterised. The enzymes used derive from *S. pneumoniae*.

Structural identifiability and indistinguishability analyses of the relevant reaction models will also be undertaken to determine in which cases direct parameter estimation is appropriate. This approach will be attempted in conjunction with more traditional approaches in kinetic characterisation; providing basis for comparison between the techniques.

1.4.1 Outline of the remainder of the thesis

The experimental and theoretical methods used are described briefly in Chapters 2 and 3. Chapter 2 also describes the provenance of the various substrates and enzymes used.

The results obtained are described in Chapters 4-7. Quasi-steady state models for the kinetics of two and three substrate models is the subject of Chapter 4. These models are derived from transient mechanistic models of the ping-pong or simple ordered types. They are then subjected to structural identifiability and indistinguishability analyses for experimentally relevant output structures.

Kinetic characterisations, for the quasi-steady state time scale, of the six *S. pneumoniae* enzymes used are presented in Chapter 5. In addition the use of direct parameter estimation is demonstrated for MurA and MurB. Simulation of candidate models of the MurF catalysed reaction is used to suggest a new approach for the detection of substrate inhibition.

In Chapter 6 structural identifiability analyses of the transient two substrate models is presented. Time courses recorded under pre-steady state conditions for MurB and IDH are then subjected to direct parameter estimation.

The *E. coli* and *S. pneumoniae* cytoplasmic phases are compared using quasi-steady state *in silico* models in Chapter 7. The predictions of the *S. pneumoniae* model are validated against *in vitro* reconstructions of this pathway. Finally the interaction between internal pathway inhibitors and a recently developed inhibitor of MurC is investigated using both *in vitro* and *in silico* methods.

The final chapter, Chapter 8, draws together the results presented within the context of new antibacterial discovery and the development of antibacterial resistance.

2. EXPERIMENTAL METHODS

2.1 Introduction

In order to study the cytoplasmic phase of the peptidoglycan biosynthesis pathway with a minimum of noise the enzymes MurA-F from *Streptococcus pneumoniae* were cloned, over-expressed in host *Escherichia coli* strains and purified. Spectroscopic assays were then used to analyze the steady state kinetics of the enzymes. Pre-steady state kinetics were also considered in some cases. Pathways were also reconstructed *in vitro* and their fluxes assessed by spectroscopic and chromatographic techniques. In this chapter the experimental methods used throughout the project are described.

In the following section the preparation of *E. coli* strains to express the desired enzymes is described. The source and structure of the transformation plasmids and techniques for producing them and assessing their structure are outlined. Then preparation of chemically competent cells and their transformation is described. The long term storage transformed cells is also covered. In the third section expression and purification of enzymes is described. Two expression systems are outlined as is the preparation of crude cell lysates. Chromatographic purification techniques are then outlined. Analysis and quality control of the process is also covered, detailing SDS-PAGE and assays of protein concentration. Spectroscopic assays are discussed in the fourth section. Determination of substrate concentration is outlined first. Then two continuous assays and two direct assays producing steady state data are described. Assays of reconstructed *in vitro* pathways are covered as are stopped-flow experiments.

2.2 Transformation of E. coli strains with plasmid DNA

E. coli strain	Genotype
TOP10 [151]	F^- mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZ ΔM15 ΔlacX74 deoR recA1 araD139 Δ (ara-leu)7696 galU galK rpsL (STR^R) endA1 nupG
BL21 Star® (λDE3) [152]	F^- ompT hsdS$_B$ (r_B^- m_B^-) gal dcm rne131 (λDE3)

Table 2.1: *E.coli* strains used in this project

2.2.1 Transformation vectors

Transformation vectors were created in a preceding project. The *murA-F* genes, cloned from *S. pneumoniae* R6, were individually incorporated into plasmids using the pET46 EK/Lic vector (Merck Chemicals Ltd, Nottingham, UK). The chosen vector incorporates ampicillin resistance, lactose based expression control and adds 6 N-terminal histidine residues to the resultant protein. Plasmids were preserved *in vivo* in glycerol stocks of *E. coli* TOP10 cells and stored at -80°C.

2.2.2 Preparation of plasmid DNA

A small sample from each glycerol stock was streaked onto sterile lysogeny broth (LB) 10:5:5 agar plates containing 0.1mg/ml ampicillin. The plates were incubated overnight at 37°C. A single colony from each plate was then grown overnight in sterile 10ml LB broth and ampicillin (0.1mg/ml) at 37°C and 180rpm. Plasmid DNA was extracted from these cultures using the Qiagen MiniPrep® extraction kit, according to the manufacturer's instructions. This kit employs the alkaline lysis method, where DNA is purified by adsorption to a silica matrix under high salt conditions.

2.2.3 Validation of plasmid integrity

The gene content of the plasmids was checked in two ways, sequencing and diagnostic polymerase chain reaction (PCR). For sequencing 500ng of DNA (determined by Microdrop spectrometer) were mixed with 5 pmol of forward or reverse primers corresponding to the gene inserted in 10μl dH_2O. The mixture was then submitted to the University of Warwick Sequence Facility. The resulting reverse sequence was complemented in *BioEdit Sequence Alignment Editor* and aligned with the forward sequence (Ibis Therapeutics, Carlsbad, USE). The resulting consensus sequence, excluding sections of unreliable data, was then compared to the published sequences of the gene.

Diagnostic PCR was undertaken to determine whether the gene was correctly inserted into the plasmid. The *Taq* Polymerase kit (Fermentas, York, UK) was used according to the manufacturer's instructions. Primer combinations used were: T7P fwd and rev; gene for and rev; gene for and T7P rev; and T7P for and gene rev. Reactions were carried out in a 25μl volume using 100ng of template DNA in an Eppendorf Mastercycler Gradient thermal cycler with an annealing temperature of 50°C. The resulting DNA was run on a 1% agarose gel, see Section 2.2.4.

2.2.4 Agarose gel electrophoresis

A 1% (w/v) solution of high melting point agarose (Invitrogen Ltd, Paisley, UK), was prepared by melting 1g of agarose per 100ml 1x Tris-acetate-EDTA (TAE) buffer and heating in a microwave oven. The solution was cooled to hand heat and 5μl of a 10mg/ml solution of ethidium bromide added per 100ml agarose solution. The solution was then poured into a cast and left to set. The gel was submerged in a gel tank containing 1x TAE buffer and DNA samples were loaded into wells with 2 μl 6x DNA Loading Buffer (0.25%(w/v) bromophenol blue, 0.25% 9(w/v) xylene cyanol FF and 15% (v/v) Ficoll 400) per 10μl of DNA sample. 3 μl of a DNA

standard 1Kb ladder, premixed with loading buffer, was loaded to allow determination of the size of DNA samples. Electrophoresis was carried out at 120V for 30 minutes and samples visualised under ultraviolet light.

2.2.5 Preparation of competent cells for DNA transformation

Cells from the required strain of *E. coli* were streaked onto a sterile LB agar plate containing the appropriate antibiotic and incubated overnight at 37°C. 10ml of sterile LB broth containing the appropriate antibiotic was inoculated with one colony from the resulting plate and incubated at 37°C with shaking at 180rpm. This culture was used to inoculate 100ml of sterile LB broth which was incubated at 37°C with shaking at 180rpm until the optical density at A_{600}nm reached 0.5. The culture was centrifuged at 4500×g to pellet cells and resuspended in TFB1 buffer (30mM potassium acetate pH 5.8, 10mM calcium chloride, 50mM manganese chloride, 100mM rubidium chloride, 15% glycerol by volume), centrifuged and resuspended in TFB2 buffer (10mM MOPS or PIPES pH 6.5, 75mM calcium chloride, 10mM rubidium chloride, 15% glycerol by volume), following the procedure adapted from [153]. Competent cells were frozen in 200μl aliquots in liquid nitrogen and stored at -80°C.

2.2.6 DNA transformation of E. coli

Chemically competent cells, see Section 2.2.5, were thawed on ice and 20-50 μl of cells incubated with 50-200ng of plasmid DNA on ice for 5 minutes. Cells were heat shocked at 42°C for 30 seconds and then returned to ice to incubate for 2 minutes. 500μl of sterile LB broth was added and the cells were incubated without shaking for one hour. 100μl of transformation mixture was plated using aseptic technique onto sterile LB agar plates containing 0.1mg/ml ampicillin. Plates were incubated overnight at 37°C.

Primer name	Primer sequence
T7P†	TAA TAC GAC TCA CTA TAG GG
T7T†	GCT AGT TAT TGC TCA GCG G
murA1 fwd	GAC GAC GAC AAG ATG AGA AAA ATT GTT ATC AAT GG
murA1 rev	GAG GAG AAG CCC GGT TTA ATC CTC AAC AAG TCT AAT ATC C
murA2 fwd	GAC GAC GAC AAG ATG GAT AAA ATT GTG GTT CAA GG
murA2 rev	GAG GAG AAG CCC GGT TTA TTC ATC TTC ATC ATT TGC C
murB fwd	GAC GAC GAC AAG ATG TCT GTA AGA GAA AAA ATG C
murB rev	GAG GAG AAG CCC GGT CTA CCT CTT GCA GGG AGT AAA ACC
murC fwd	GAC GAC GAC AAG ATG TCA AAG ACA TAT CAT TTT ATC G
murC rev	GAG GAG AAG CCC GGT CTA TTG AAC ATT GCT TGT CAA GTT AGA C
murD fwd	GAC GAC GAC AAG ATG AAA GTA ATA GAT CAA TTT AAA AAT AAG
murD rev	GAG GAG AAG CCC GGT TTA TTC TTT TAA CTC CGC TAC TGT GTC G
murE fwd	GAC GAC GAC AAG ATG ATT AAG ATT GAA ACC G
murE rev	GAG GAG AAG CCC GGT TTA TAA ATA ATT TTC TGC GAC G
murF fwd	GAC GAC GAC AAG ATG AAA TTA ACA ATC CAT GAA ATT GC
murF rev	GAG GAG AAG CCC GGT TCA CTT GTC TTC ATT TTC TAA ACT TTC TAC

Table 2.2: Sequences of the primers used throughout the project. Primers marked † obtained from (Merck Chemicals Ltd, Nottingham, UK).

2.2.7 Preparation of glycerol stocks

10 ml of sterile LB broth was inoculated with a single *E. coli* colony from a fresh transformation, see Section 2.2.6, and grown overnight in the presence of 0.1mg/ml ampicillin with shaking at 37°C. 1 ml of culture was aseptically mixed with 1 ml of analytical reagent grade 100% glycerol by volume in a Corning® cryovial and frozen at -80°C (Sigma-Aldrich, St Louis, USA).

2.3 Expression and purification of MurA-F from Streptococcus pneumoniae

2.3.1 Protein over-expression in E. coli (induction by IPTG)

Proteins were over-expressed in *E. coli* BL21 Star Rosetta (DE3) transformed as described above, Section 2.2.1 and Section 2.2.6. 10ml of sterile LB broth, 0.1mg/ml ampicillin and 0-0.2% (w/v) glucose was inoculated with a single *E. coli* colony from a fresh transformation and incubated overnight at 37°C with shaking at 180rpm. This culture was used to inoculate 750ml of sterile LB broth, 0.1mg/ml ampicillin and 0-0.2% (w/v) glucose which was incubated at 37°C with shaking at 180rpm until the optical density at A_600nm reached 0.7. Cultures were induced with 1mM IPTG and grown at 37°C with shaking for 2-3 hours. 1ml samples of the culture were harvested before induction and after completion of this growth period to check protein expression these cultures were pelleted at 10000rpm and stored at -20°C. The remaining cells were centrifuged at 50000×g using a Beckman JLA 8.1000 rotor and an Avanti® J-20XPI centrifuge, resuspended in 250mM pH 7.6 Hepes, 50mM magnesium chloride, transferred to a Falcon™ tube and centrifuged at 5000×g in a Labofuge® 400R (DJB Labcare Ltd, Newport Pagnell, UK). The resulting pellets were stored at -20°C.

2.3.2 Protein over-expression in E. coli (using autoinduction media)

The procedure above, Section 2.3.1, was followed with the following exceptions. A 1ml sample was taken from the overnight culture to check protein expression. The overnight culture was

then used to inoculate 929ml of sterile LB broth, 50ml sterile 20xNPS (0.5M ammonium sulfate, 1M monopotassium phosphate, 1M disodium phosphate, pH 6.75), 20ml sterile 50x 5052 (20% glycerol by volume, 2.5% (w/v) glucose, 10% (w/v) α-lactose), and 1 ml sterile 1M magnesium sulphate. The cultures were incubated at one of 17, 27, 37°C for between 18 and 48 hours and the cultures were not induced. The balance of glucose to lactose in the expression culture controls expression directly. When the cells exhaust the available glucose supply they start metabolising lactose initiating expression [154].

2.3.3 Preparation of crude cell lysates

Lysates were prepared from previously frozen cells, see Sections 2.3.1 and 2.3.2. Pellets were thawed on ice and resuspended in 20-50ml of nickel column running buffer A (NCRBA) (50mM Hepes pH 7-7.6, 0.5-1M sodium chloride, 1mM magnesium chloride, 5% glycerol by volume, 5mM imidazole, 0.2μM leupeptin, 0.2μM pepstatin, 0.2mM PMSF). Chicken egg white lysozyme (Merck Chemicals Ltd, Nottingham, UK) was added to the cell suspension to a concentration of 2.5mg/ml. The treated cells were sonicated on ice for 2-5 bursts 30 second bursts at 70% power using a Bandelin Sonopuls sonicator and centrifuged in two stages of thirty minutes at 70000$\times g$ to pellet cell membranes and other debris (Bandelin, Berlin, Germany). The supernatant was collected and applied directly to an equilibrated chromatography column.

2.3.4 Affinity chromatography

Proteins were partially purified from cell lysates through binding affinity to nickel. HisTrap® HP, high performance nickel columns were used (GE Healthcare, Little Chalfont, UK). Columns were developed using ACTA™ purifier 100 systems (GE Healthcare, Little Chalfont, UK). They were first washed with 10 column volumes of filtered dH_2O, to remove the 20% ethanol storage solution, then equilibrated with 10 column volumes of filtered NCRBA. Cell lysate was then loaded and flowthrough collected. Columns were washed with filtered NCRB until A_{280} and

A_{258} returned to baseline levels. Protein was eluted by increasing imidazole concentration introduced by mixing filtered NCRBA and filtered NCRBB (as NCRBA but containing 0.5-1M imidazole). A chromatogram was recorded on a linked PC, typically 10ml fractions were collected, however peaks (increases in A_{280} and A_{258}) were manually fractionated where possible to increase purity.

After use affinity columns were either stripped with 5 column volumes of 500mM EDTA pH 8.0 followed by 10 column volumes of water, or washed with 10 column volumes of water. Columns were washed with 10 column volumes of 20% ethanol by volume prior to storage. Stripped columns were recharged by washing with 10 column volumes of water followed by 50mM sodium acetate pH 4.0 containing 10mg/ml nickel chloride to chelate Ni^{2+} to the Sepharose® resin.

2.3.5 Size exclusion chromatography

Proteins were separated on the basis of size using 300ml Superdex™ 200 columns attached to an ACTA™ purifier 100 system (GE Healthcare, Little Chalfont, UK). Columns were equilibrated with a single column volume of gel filtration running buffer (GFRB) (50mM Hepes pH 7.6, 500mM sodium chloride, 1mM magnesium chloride, 10% glycerol by volume, 0.2μM leupeptin, 0.2μM pepstatin, 0.2mM PMSF). Samples were loaded via an injection loop in volumes of less than 10ml to reduce loss of resolution due to sample smearing. Buffer was pumped through the column at 1ml per minute for one column volume and 10 ml fractions were collected. Peaks recorded as an increase in A_{280} and A_{258} were manually fractionated where possible to increase purity.

2.3.6 Anion exchange chromatography

Proteins were separated on the basis of charge using a MonoQ™ HR 5/5 column (10μm beads) attached to an ACTA™ purifier 100 system (GE Healthcare, Little Chalfont, UK). The resin has an attached quaternary amine group, which gives the resin a strong positive charge over a

broad range of pH 2-12. Proteins applied in low salt buffer at least one pH unit above their isoelectric point have an overall negative charge and bind strongly. They can be eluted by a salt gradient with weakly bound proteins eluting at low salt and strongly bound proteins eluting at high salt.

Columns were equilibrated by three washes of 10 column volumes of MonoQ™ running buffer A (MQRBA) (20mM Tris pH 8.2, 1mM magnesium chloride, 2mM β-mercaptoethanol, 0.2μM leupeptin, 0.2μM pepstatin, 0.2mM PMSF) followed by 10 column volumes of MQRBB (20mM Tris pH 8.2, 500mM sodium chloride, 1mM magnesium chloride, 2mM β-mercaptoethanol, 0.2μM leupeptin, 0.2μM pepstatin, 0.2mM PMSF) and a final wash of 10 column volumes of MQRBA. This procedure ensured that no excess salt or other contaminants were bound and that the resin was charged with the appropriate ion. Samples of no more than 1ml were loaded via an injection loop and the column developed by washing with a sodium chloride gradient from 0 to 500mM over 20ml. Fractions of 2ml were collected throughout and the presence of eluted protein was judged by an increase in A_{280} and A_{258}.

2.3.7 Buffer exchange

Buffer exchange was performed by placing the protein solution into 14 or 28 millimetre dialysis tubing, which had previously been boiled in 5mM EDTA and washed in dH_2O. The ends of the tubing were secured and placed in at least 10× its volume of the required exchange buffer chilled to 4°C. Dialysis was performed overnight with stirring at 4°C.

2.3.8 Concentration of protein

Protein solutions were concentrated using Vivaspin™ centrifugal concentrators with a molecular weight cut off (MWCO) of 10kDa. The protein solution was placed in the top chamber of the concentrator and centrifuged at 3500×g in 10-15 minute bursts in a chilled Labofuge® 400R

bench top centrifuge at 4°C until the desired concentration was reached. The concentration of the protein was determined as described in Section 2.3.11.

2.3.9 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE used the Tris-glycine buffer system to separate and visualise polypeptides under denaturing conditions on a discontinuous gel system consisting of a resolving gel at pH 8.8 and a stacking gel at pH 6.5 [155]. The final composition of the resolving gel was 0.371M pH 8.8 Tris-HCl, 10% acrylamide:bis-acrylamide (29:1), 0.01% SDS. Gels were cast using the BioRad Mini-PROTEAN® kit (Bio-Rad Laboratories Ltd, Hemel Hempstead, UK), 22.75ml was polymerised with 0.23ml 10% APS and 20μl TEMED, then overlaid with 100% ethanol to remove the meniscus. The final composition of the stacking gel was 0.123M pH 6.5 Tris-HCl, 10% acrylamide:bis-acrylamide (29:1), 0.01% SDS. 10ml of stacking gel was polymerised with 0.1ml 10% APS and 15μl TEMED and pipetted onto the polymerised resolving gel, to an appropriate thickness, where a plastic comb was inserted to form the wells. Protein samples for analysis were prepared in 4 $\times$ loading buffer (62.5mM pH 6.8 Tris-HCl, 10% glycerol by volume, 2% SDS, 5% 2-β-mercaptoethanol and bromophenol blue to colour) and heat denatured at 95°C in a heat block for 5 minutes. The gel comb was removed and the wells washed with SDS-PAGE running buffer (25mM Tris, 0.192M Glycine, 0.1% SDS pH 8.3) and samples were pipetted into the wells. Gels were run in SDS-PAGE running buffer at 100V for 20 minutes or until the gel front created a thin band at the interface of the stacking and resolving gels. The voltage was then increased to 120V and run until the gel front reached the base of the gel.

2.3.10 Coomassie® staining of gels

Proteins resolved by SDS-PAGE, see Section 2.3.9, were visualised using Coomassie® dye. Gels were stained for up to 1 hour in either 50% methanol, 10% acetic acid, 40% water containing 10mg L^{-1} Coomassie® R-250 dye or a colloidal mix (10% ammonium sulphate, 0.1% Coomassie

R-250 dye, 3% orthophosphoric acid, 20% ethanol). Gels were destained overnight with either the same solution in the absence of Coomassie R-250 or dH_2O respectively. Gels were photographed and then disposed of.

2.3.11 Determination of protein concentration

Protein concentration was determined by colourimetric assay at 595nm (Bio-Rad Laboratories Ltd, Hemel Hempstead, UK). $x\mu$l of protein sample were added to $800 - x\mu$l dH_2O and 200μl BioRad Protein Assay Dye Reagent in a cuvette, mixed and the absorbance measured at 595nm using an Ultrospec™ 2000 UV-visible spectrometer (GE Healthcare, Little Chalfont, UK). Protein concentrations that gave a reading beyond the linear range of the assay (above 0.6 at 595nm) were diluted and the reading repeated. Readings for several x were collected, absorbances were plotted against x and a straight line was fitted to the data using *Microsoft Excel*© (Microsoft Ltd, Reading, UK). The gradient of the best fit line was used to calculate the concentration of protein using the following equation:

$$[\text{protein}] = \text{gradient} \times 19500 \times \text{dilution factor}\,. \tag{2.1}$$

2.4 Biochemical spectrophotometric assays

2.4.1 Determination of concentration of substrates

The concentrations of substrates with a measurable absorbance were determined directly by spectroscopic analysis. A UV-visible quartz cuvette was filled with 200-$x\mu$l of dH_2O and placed in the cell changer of a Cary® 100 spectrometer (Varian Medical Systems Inc, Palo Alto, USA). The wavelength of the instrument was set to that of the absorbance maxima of the substrate and the absorbance of the sample was set to zero. The absorbance of the sample was then followed for 1 minute before $x\mu$l of substrate was added to the cuvette and mixed. The absorbance was

then followed for an additional 5 minutes. The concentration of the substrate, [substrate], was then determined from a Beer-Lambert Law calculation:

$$[\text{substrate}] = \mathrm{A}_y \,/\epsilon_y \times \text{dilution factor}\,. \tag{2.2}$$

The absorbance measured at the desired wavelength is denoted A_y and the extinction coefficient of the substrate at the same wavelength by ϵ_y. Typically four assays were undertaken simultaneously for different values of x and the results averaged.

Where substrates did not have a directly measurable absorbance one of the assays described below, Sections 2.4.2, 2.4.3, and 2.4.4, was used to ascertain its concentration. The components of a reaction which utilised the substrate were assembled, consumable elements in significant excess of that expected of the substrate assayed. The total absorbance change produced by adding a known volume of substrate to this mixture was then determined. A coupled assay was used if necessary. This absorbance change was assumed to result from the total consumption of the assayed substrate and thus its concentration could be determined by a Beer-Lambert Law calculation.

Substrate Name	Abs Max (nm)	Ext Coeff ($\mathrm{M^{-1}cm^{-1}}$)
Adenosine triphosphate (ATP)	260	13400
Uridine diphosphate glucosamine (UGP)	260	10000
Nicotinamide adenine dinucleotide (NADH)	340	6220
Nicotinamide adenine dinucleotide phosphate (NADPH)	340	6220
Phospho-enolpyruvate (PEP)	240	1700

Table 2.3: Absorbance maxima (Abs Max) and extinction coefficients (Ext Coeff) of the substrates used throughout the project. Note that all peptidoglycan precursors have the same absorbance characteristics, as such only one has been listed UGP.

2.4.2 Continuous ADP production assay

This assay links the production of ADP to NADH oxidation, creating an absorbance change which can be measured at 340nm [156]. ADP is phosphorylated by the conversion of phosphoenolpyruvate to pyruvate catalysed by pyruvate kinase regenerating ATP. Pyruvate is then converted to lactate by lactate dehydrogenase requiring the oxidation of NADH to NAD^+ (Figure 2.1). NADH oxidation can be measured spectrophotometrically as the quinone ring of NADH absorbs light at 340nm, extinction coefficient $6220 M^{-1} cm^{-1}$, while the aromatic ring of NAD^+ does not. Therefore NADH oxidation is observed as an absorbance decrease at 340nm which only occurs when ADP is produced by the reaction to be investigated. The assay was used as an indirect measure of Mur ligase activity since these enzymes are ATP dependent. ADP is formed upon ligation of amino acid to the peptidoglycan precursor peptide. When this assay is used ATP concentration is maintained at a constant level throughout the measurement period.

A typical reaction mixture using this assay system contained 50mM HEPES pH 7.6, 10mM magnesium chloride, 1mM dithiothreitol (DTT), 200μM NADH, 2mM PEP, 3.2 units of pyruvate kinase (PK), 4.6 units of lactate dehydrogenase (LDH) (Sigma-Aldrich, St Louis, USA) and the components of the reaction made up to a concentration of 200μl. All components listed above and all but one of the reaction components were added to a UV-visible quartz cuvette, mixed and placed in the cell changer of a Cary® 100 spectrometer (Varian Medical Systems Inc, Palo Alto, USA), with temperature control set to 37°C. The cuvette was allowed to equilibrate to temperature before setting the absorbance at 340nm to zero. Sample absorbance was followed for 1-2 minutes to establish a base rate. The cuvette was then removed from the instrument, the final component of the reaction was added to the reaction mixture which was mixed and the cuvette was quickly returned to the cell changer. The absorbance changes were recorded using Varian kinetics software for 5 to 10 minutes. Sampling rates were dependent on the number of

reactions being monitored in a particular experiment. For a single reaction a measurement can be made every tenth of a second; two reactions allow a measurement every four seconds; and four reactions allow a measurement every ten seconds.

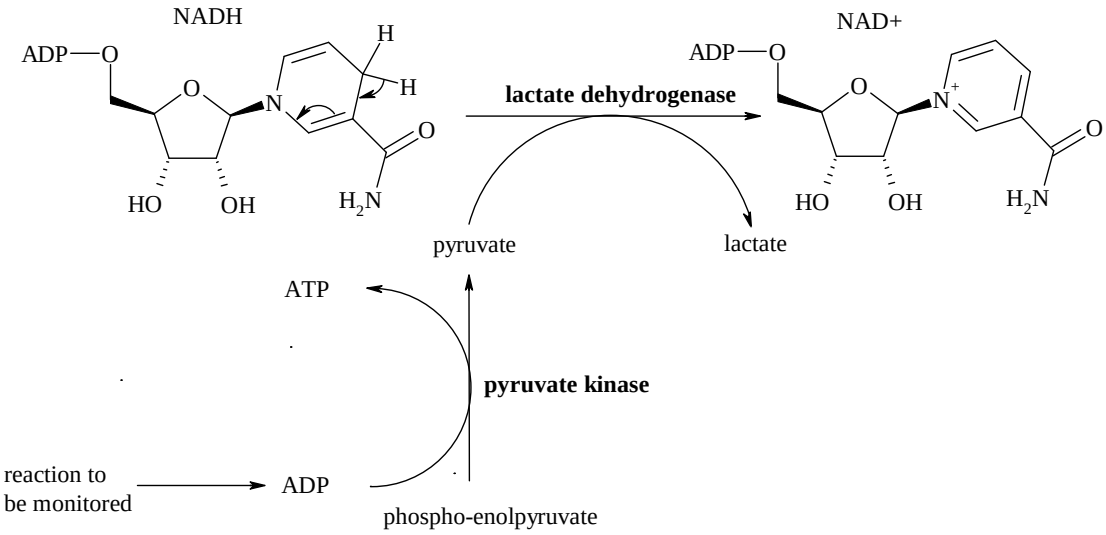


Figure 2.1: Schematic diagram of the basis of the continuous assay for ADP production, coupled to pyruvate kinase and lactate dehydrogenase. NADH oxidation is monitored spectrophotometrically by a decrease in absorbance at 340nm.

2.4.3 Continuous phosphate production assay

This assay uses the production of inorganic phosphate to phosphorylate MESG, creating an absorbance change which can be measured at 360nm [157]. Purine nucleoside phosphorylase catalyses the conversion of MESG and phosphate to methylthioguanine and ribose 1-phosphate see Figure 2.2. The purine ring of MESG and MTG is sensitive to pH. Increasing pH causes the dissociation of hydrogen from the nitrogen atom closest to the sulphur atom inducing a rearrangement of the ring structure. Since MESG and MTG have different pK_a's they exist in different states between pH 6.5 and 8.8, see Figure 2.3. A change in absorbance due to this rearrangement, change in extinction coefficient $11000 M^{-1} cm^{-1}$ at pH 7.6, is observable at 360nm. The assay was used as an indirect measure of Mur ligase and MurA activity. Mur ligases release inorganic phosphate upon ATP hydrolysis and formation of the peptidoglycan

peptide precursor. MurA catalyses the transfer of the enolpyruvate to UDP-GlcNAc releasing phosphate from PEP.

As this assay produces an increase of absorbance from a low initial level, unlike the assay described in Section 2.4.2, noise levels are typically lower over the initial minutes of the assay. This is an advantage when assaying enzyme kinetics based on initial rates. However it is highly sensitive to phosphate contamination of the substrates and as such it is often impractical to use it.

Typically the following components were mixed in a cuvette to a final volume of 200μl:50mM Hepes pH 7.6, 10mM magnesium chloride, 40μl of nominally 1mM MESG, 1mM dithiothreitol, 2.244 units of PNP, and the components of the reaction to be assayed. Assays were carried out in similar fashion to the ADP production assay described in Section 2.4.2 in a Cary® 100 spectrometer at 360nm and 37°C.

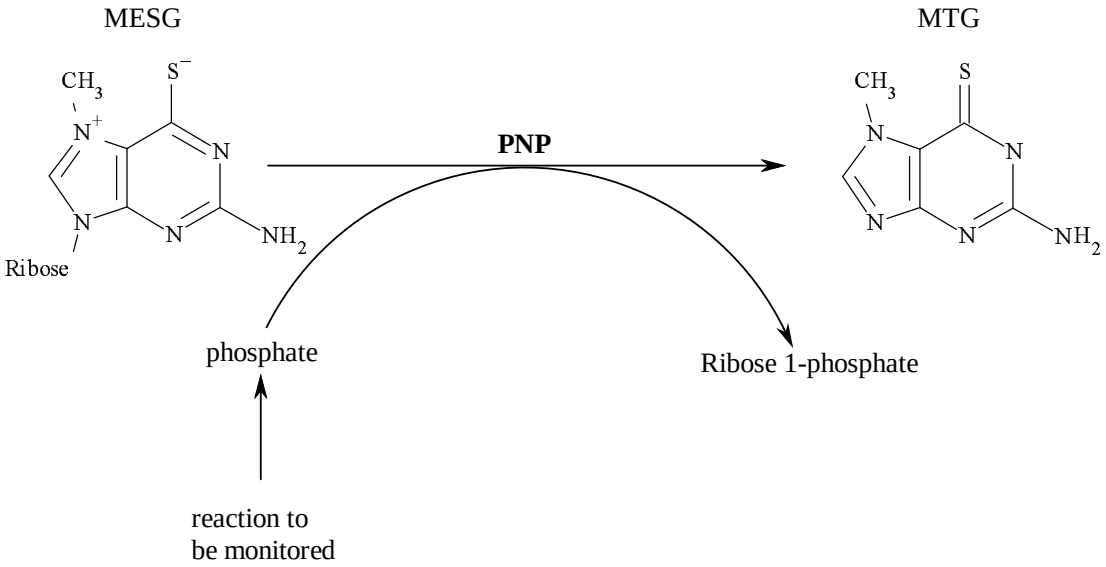


Figure 2.2: Schematic diagram of the basis of the continuous assay for inorganic phosphate production, coupled to purine nucleoside phosphorylase. Conversion of MESG to methylthioguanine is monitored spectrophotometrically by an increase in absorbance at 360nm.

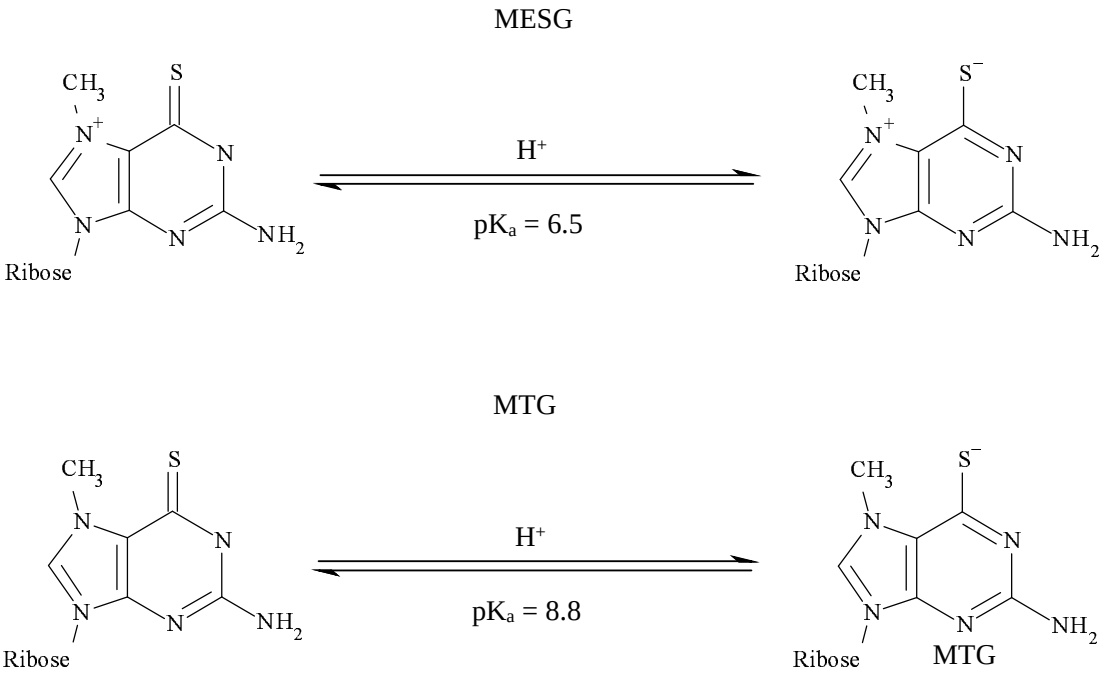


Figure 2.3: Diagram showing the dependence of the absorbance change measured in the phosphate release assay on pH. Dissociation of the nitrogen bound hydrogen causes a rearrangement of the purine ring reducing its absorbance. MESG and MTG have different pK_a's allowing the two structures to be distinguished between pH 6.5 and 8.8.

2.4.4 Continuous assay of MurB activity

MurB activity produces a native absorbance change at 340nm by the oxidation of NADPH [85]. NADPH is oxidized by MurB when catalyzing the reduction of a double bond on the peptidoglycan precursor, see Figure 2.4. As with NADH, Section 2.4.2, NADPH oxidation can be measured spectrophotometically due to the rearrangement of the quinone ring. An absorbance change at 340nm can be measured with extinction coefficient $6220 M^{-1} cm^{-1}$. This assay was typically used as a direct assay of the activity of MurB.

The reaction mixture for this assay typically consisted of: 50mM Hepes pH 7.6, 100mM potassium chloride, 1mM dithiothreitol and 0.096μM MurB made up to 200 μl. NADPH and UDPPEE concentration varied dependent on the purpose of the assay. Assays were carried out in a similar fashion to that described above, Section 2.4.2, in a Cary® 100 spectrophotometer at 340nm and 37°C (Varian Medical Systems Inc, Palo Alto, USA).

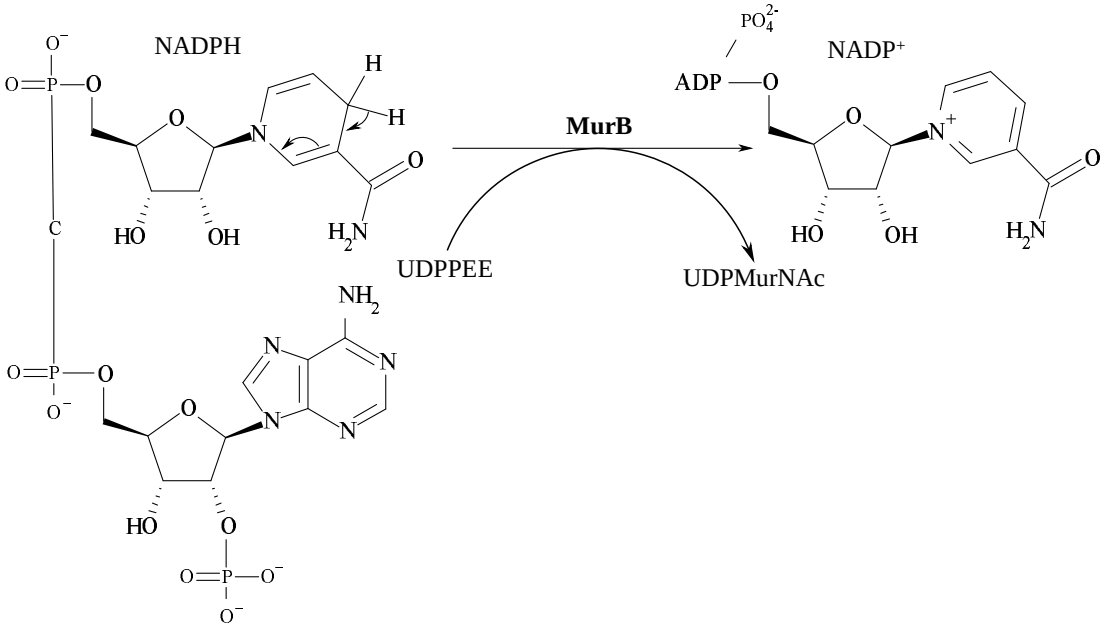


Figure 2.4: Schematic diagram of the direct assay for MurB activity. NADPH oxidation is monitored spectrophometrically by a decrease in absorbance at 340nm.

2.4.5 Continuous assay of Lactate dehydrogenase activity

Lactate dehydrogenase activity produces a direct change in absorbance at 340nm by the oxidation of NADH. NADH is oxidised by LDH when catalyzing the formation of lactate from pyruvate, see Figure 2.1. The details of the absorbance change caused by this reaction are in Section 2.4.2.

2.4.6 Determination of pathway fluxes

Pathways were reconstructed *in vitro*, typically reaction mixtures included: 50mM Hepes pH7.6, 10mM magnesium chloride, 150mM potassium chloride, 1mM dithiotheitol, 1mM ATP, 2mM PEP, 10mM of L-alanine, D-glutamate, L-lysine and D-alanyl-D-alanine. Each enzyme included in the pathway was present at 0.1μM concentration. If MurB was included 100uM NADPH was also included. If the absorbance change caused by this reaction was not to be monitored 0.18 units/ml of isocitrate dehydrogenase and 10mM isocitrate were added to convert the $NADP^+$

produced to NADPH. Typically 10-50μM of the first peptidoglycan precursor required for the pathway was included. The assay system used to monitor the pathway was typically one of the assays described above, Sections 2.4.2 or 2.4.3. Concentrations of NADH and MESG were selected to provide an excess over the expected production of the pathway. The final assay volume was 200μl however typically 4 pathways, or other sets of conditions, were run simultaneously. The mixed reaction components excluding the first enzyme in the pathway were added to UV-visible quartz cuvettes and placed in the cell changer of a Cary® 100 spectrometer, with temperature control set to 37°C. The samples were allowed to equilibrate at 37°C and their absorbances at the appropriate wavelength was set to zero. Sample absorbances were then followed for 5 minutes. Then the final assay component was added, the solution mixed and cuvettes stoppered to prevent evaporation. Sample absorbances were then recorded for an additional 55 minutes.

2.4.7 Pre-steady state kinetics experiments

Stopped-flow experiments were carried out on a Bio-Logic SFM300/400, schematic diagram Figure 2.5, with attached MM450 spectrometer which could measure the change in absorbance at a preset wavelength (Bio-Logic SAS, Claix, France). 8000 data points were collected during time scales from 2.4ms to 8 seconds. Only direct assays of enzyme activity were used for these experiments and as a consequence only MurB and lactate dehydrogenase could be analyzed. The activity of these enzymes oxidises NADH or NADPH, see Sections 2.4.4 and 2.4.5, producing an absorbance change at 340nm. The reaction mixtures for these experiments typically included, 50mM Hepes pH 7.6, 1mM dithiothreitol and either 10mM magnesium chloride in the case of LDH or 100mM potassium chloride in the case of MurB. Protein concentrations were determined by BioRad assay as described in Section 2.3.11. MurB concentration was further checked by observing FAD absorbance at 480nm. Assays were undertaken at a variety of ratios of enzyme to substrate concentration typically 1:10 and 1:1. Enzyme concentration was typically 25μM

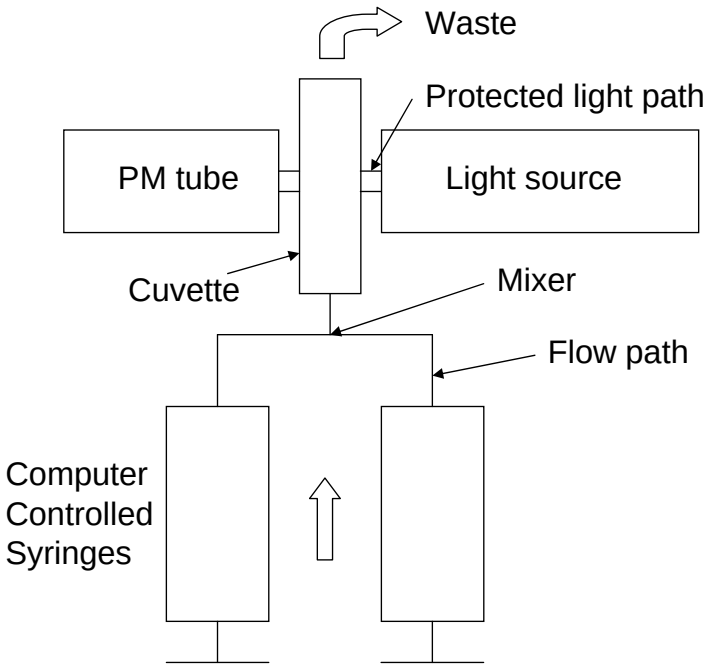


Figure 2.5: Schematic diagram of stopped flow apparatus.

and thus concentrations of the substrates were typically 250μM or 25μM.

The measurement cell was filled with the buffer described above and the absorbance was set to zero. NADH was then added to the appropriate concentration and the resultant absorbance was checked against the expected value. Finally all components were mixed and transient absorbances were measured over appropriate time frames. Assays were performed at either 20 or 37°C. Parameters were estimated as described in Section 3.4.

3. THEORETICAL METHODS

3.1 Introduction

The ultimate goal in modelling any physical system is a quantitative, predictive, understanding under a wide range of conditions. To achieve this the system must be expressed in mathematical terms. A reaction mechanism can be naturally converted to a state space model as follows. The concentrations of the species in the reaction mixture correspond to the state variables of the mathematical model. Transitions between states normally take the form of first order differential equations in the state variables, and can be determined from the reaction mechanism by mass action kinetics. Given appropriate parameter values a reaction, or a system of reactions, can be simulated to obtain predicted behaviour under a range of conditions. It is hoped that the parameters of a model can be estimated from experimental data. In this chapter theoretical and computational approaches to the problems of parameter estimation from experimental data and simulation of reactions and pathways are outlined.

The first section, Section 3.2, introduces structural identifiability, the field in mathematics/systems theory concerned with whether the parameters of a model system can be determined from perfect observations [148]. A formal mathematical basis for this problem is constructed in the language of general state space models. A criterion for determining the structural identifiability of a given model is provided. From this what is meant by an identifiable model is defined. Techniques for determining the identifiability of a model are described and implementations within the computer algebra package *Maple* are given [158]. Certain theoretical aspects of the input-

output relationship approach are further explored to resolve some unstated assumptions in the current understanding of this approach. The second section, Section 3.3, follows naturally from the first, introducing structural indistinguishability, the field concerned with whether perfect observations can be used to distinguish between two model systems [149]. This section follows the general structure of the preceding section. It starts with definitions of what is meant by distinguishability, outlines relevant analysis techniques and gives computational implementations of these techniques. Parameter estimation from experimental data is outlined in Section 3.4. Two computational packages *COPASI* [159] and *Facsimile 4* (MCPA Software, UK) are described and an overall fitting procedure using these programs is given. *COPASI* is also well suited to simulation of metabolic pathways; as described in section 3.5. *Microsoft Excel* was used to manually implement numerical integration of simpler systems.

3.2 Structural identifiability

The parameters of a model can be estimated by numerical fitting techniques. However even for perfect data it is possible that there is more than one parameter vector which will produce the observed output. If this is the case little confidence can be placed in the parameters produced by the estimation process. This problem can be avoided by changing the model such that the parameters that are subsumed into other parameters, for example by reparametrisation [160], or by using additional experiments permitting the identification of more parameters [161]. The effect of these approaches can be predicted by structural identifiability analysis. This field of mathematics/systems theory is concerned with determining whether for a given model system, including descriptions of the processes occurring, of the measurements made and of any inputs to the system, a perfect, that is continuous and noise free, observation will uniquely determine the parameters of the model.

These concepts are now formally defined [162]. The models used in this work have no inputs. It

is therefore unnecessary to include inputs in the definitions given. Define a general state space model, $\sigma(\mathbf{x}, \mathbf{f}, \mathbf{x}_0, \mathbf{y}, \mathbf{C}, \mathbf{p})$, as follows:

$$\mathbf{x}'(t, \mathbf{p}) = \mathbf{f}(\mathbf{x}(t, \mathbf{p}), \mathbf{p}), \tag{3.1}$$

$$\mathbf{x}(0, \mathbf{p}) = \mathbf{x}_0(\mathbf{p}), \tag{3.2}$$

$$\mathbf{y}(t, \mathbf{p}) = \mathbf{C}(\mathbf{p})\mathbf{x}(t, \mathbf{p}), \tag{3.3}$$

where $\mathbf{x}(t, \mathbf{p}) = (x_1(t, \mathbf{p}), \ldots, x_n(t, \mathbf{p}))^T$ is the state vector, containing the state variables $x_i(t, \mathbf{p})$. Let $\mathbf{p} = (p_1, \ldots, p_q)^T$ be a vector of unknown model parameters which lies in some open set, $\Omega \subset \mathbb{R}^q$, of feasible values. Let $\mathbf{x}(t, \mathbf{p})$ lie in a connected open subset $M(\mathbf{p}) \subset \mathbb{R}^n$ and $\mathbf{f}(\cdot, \mathbf{p})$ be analytic on $M(\mathbf{p})$. The vector of model outputs, $\mathbf{y}(t, \mathbf{p}) = (y_1(t, \mathbf{p}), \ldots, y_r(t, \mathbf{p}))^T$, comprises the combination of elements of the state vector which are measured experimentally. This is often described as a linear function of $\mathbf{x}(t, \mathbf{p})$ with an $r \times n$ observation gain matrix $\mathbf{C}(\mathbf{p})$ [162].

For the purposes of a structural identifiability analysis the observations are assumed to perfectly represent the input-output structure of the model. That is they are assumed to be continuous, noise and error free. For a model of the form (3.1) - (3.3) a parameter vector $\overline{\mathbf{p}}$ is **indistinguishable** from $\mathbf{p}$, denoted $\mathbf{p} \sim \overline{\mathbf{p}}$, if they give rise to identical model outputs, that is:

$$\mathbf{y}(t, \mathbf{p}) = \mathbf{y}(t, \overline{\mathbf{p}}), \text{ for all } t \geq 0. \tag{3.4}$$

For generic $\mathbf{p} \in \Omega$ (that is, for all $p \in \Omega$ except for a subset of a closed set of Lebesgue measure zero) a parameter p_i is **locally identifiable** (LI) if there exists a neighbourhood of points around $\mathbf{p}$, $N(\mathbf{p})$, such that if $\overline{\mathbf{p}} \in N(\mathbf{p})$, $\mathbf{p} \sim \overline{\mathbf{p}}$ implies that $\overline{p_i} = p_i$. If no such $N(\mathbf{p})$ exists for p_i it is **unidentifiable**. If $N(\mathbf{p}) = \Omega$ for p_i then it is **globally identifiable** (GI). A model is unidentifiable if any parameter is unidentifiable. It is **structurally locally identifiable** (SLI)

if all parameters are LI and at least one is not GI. It is **structurally globally identifiable** (SGI) if all parameters are GI.

The following subsections outline a number of approaches to structural identifiability analysis. Implementations of these approaches in the computer algebra package *Maple* are also given.

3.2.1 Taylor series approach

To determine whether a model is structurally globally identifiable it must be shown that the observations determine a unique parameter vector. The output function $\mathbf{y}(t, \mathbf{p})$ can be expanded as a Taylor series in t since $\mathbf{f}(\mathbf{x}, \mathbf{p})$ is analytic. Expansion at time, $t = a$, creates a unique expression:

$$\mathbf{y}(t, \mathbf{p}) = \sum_{i=0}^{\infty} \mathbf{y}^{(i)}(a, \mathbf{p}) \frac{(t-a)^i}{i!}, \tag{3.5}$$

in terms of derivatives of the output at that point. Using Equation (3.1) these derivatives can be written in terms of elements of the state vector and elements of $\mathbf{p}$. The coefficients of a Taylor series are unique for a given output, thus by solving these equations for $\mathbf{p}$ it is possible to determine whether there is a unique $\mathbf{p}$ for the output structure used [163]. If a unique solution exists the model is structurally globally identifiable. However this technique provides an infinite number of coefficients. For a linear system, without input (or with a single impulsive input), it is known that at most $2n - 1$ (where n is the state space dimension) independent equations are required to determine the possible solutions for $\mathbf{p}$ [164]. For general non-linear systems no strict upper bound has been determined although a loose upper bound exists [165]. As such it can be difficult to prove unidentifiability using this technique. Moreover, in either case, high order coefficients are very complex making them difficult to solve even using symbolic algebra packages. In some cases it may be possible to construct an inductive argument to describe the form of the Taylor series coefficients and thus prove unidentifiability.

While for some systems it is possible to apply this approach by hand, many systems are too complex for this to be practical. Regardless, in order to avoid errors, it is preferable to implement it using a computer algebra package. In this work it was implemented in *Maple* as follows [158]. This approach is applied to a simple model in Section 3.2.4.

Step 1: Definition of model

Model constants, initial conditions, and differential equations were defined as follows:

```
assume(constant_name, constant):
variable_name(0) := initial_value;
D(variable_name(t), t) := function;
```

The use of the assume function when defining constants ensures that *Maple* will not attempt to differentiate them. Initial values are either zero or a constant that has been defined. There is no intrinsic reason why an initial condition should be used rather than a nonzero time. However for the models analyzed the most appropriate time at which to expand was $t = 0$ since this was the time at which concentrations could be controlled. The `D` function in *Maple* denotes the differential operator and can be applied to a function. The definitions used cause *Maple* to replace variables to which the operator has been applied with the function given. This function should thus be the relevant differential equation in `variable\_names(t)` and constants.
The output function was defined as follows:

```
y := [a_1*variable_name_1(t), ... ,a_n*variable_name_n(t)];
SUBS1 := a_1 = value_1, ... , a_n = value_n;
f_0_i := subs(SUBS1, y)[i];
g_0_i := eval(f_0_i, t=0);
```

For maximum flexibility a matrix A could be defined and multiplied by the state vector x to obtain the output functions. However for the models analyzed outputs did not consist of linear

combinations of species, as such the definition above was sufficient. `SUBS1` determines which variables were measured, the `value_i`s were either 1 or 0 indicating whether the variable was measured. The `f\_0\_i`s are the output functions and the `g\_0\_i`s are the zeroth order Taylor series coefficients.

Step 2: Calculation of Taylor series coefficients

The Taylor series coefficients were calculated as follows:

```
f_i_j := D(f_i-1_j):
g_i_j := eval(f_i_j, t=0):
```

The Taylor series coefficients are determined inductively, starting from the coefficients defined above. The differential operator `D` was applied to the previous derivative of the output function. The `eval` function is then used to determine the corresponding Taylor series coefficient. There being no strict upper limit on the number of Taylor series coefficients needed for a nonlinear system an arbitrary number of coefficients was computed. Calculation of these coefficients in *Maple* requires large amounts of memory and can require significant computational time. The number of coefficients calculated typically depends on these factors.

Step 3: Definition of an alternative parameter vector

Suppose there exists an alternative parameter vector, $\overline{\mathbf{p}}$, which produces the same output as the starting parameters. If it can be shown that $\overline{\mathbf{p}} = \mathbf{p}$ then the model is globally identifiable. If there are countable many solutions the model is locally identifiable. The alternative parameter vector was defined and used to create equations from the Taylor series coefficients as follows:

```
SUBS2 := parameter_name_1 = alt_parameter_name_1,
 ... , parameter_name_m = alt_parameter_name_m;
ga_i_j := subs(SUBS2, g_i_j):
eqn_i_j := g_i_j - ga_i_j:
```

The `subs` function uses the list defined by `SUBS2` to replace names on the left with names on the right. This creates the alternative Taylor series coefficients. Equations are then defined by subtracting these coefficients from the coefficients determined. When these equations are solved *Maple* will assume that they are equal to zero unless otherwise stated.

Step 4: Solving for the alternative parameters

The equations derived above are now solved for the alternative parameters:

```
s:= solve({eqn_1_1, ... , eqn_a_b}, {alt_parameter_1,
 ... , alt_parameter_name_m}):
eqn_1_1 := alt_parameter_1 = solve(eqn_1_1, alt_parameter_1):
eqn_j_k_1 := simplify(subs(s1, ... , si-1, eqn_j_k)):
si := alt_parameter_i = solve(eqn_j_k_1, alt_parameter_i}:
```

There are two possibilities. In some cases *Maple* is able to solve all the equations determined for the alternative parameters. In this case the first line of code is used. The solution can be seen by inputting `s;`. However, this calculation, like the determination of the Taylor series coefficients, consumes memory and computational time. If a solution cannot be obtained within a reasonable time, or without exhausting the available memory, an iterative approach is possible. The first equation is solved for one of the alternative parameters and this solution is used to create a term that can be used in `subs`. Proceeding one equation at a time, further solutions for alternative parameters are determined, by substituting all known solutions into an equation and then solving for an alternative parameter. This approach is more manageable although it may still run out of the available resources.

3.2.2 Pohjanpalo's Jacobian rank test

As observed above symbolic implementations of the Taylor series approach can be computationally expensive. In cases where computing power is insufficient a result, due to Pohjanpalo,

can be used [166]. The derivatives of the output are used as the elements of an infinite vector. The Jacobian matrix of this vector with respect to the unknown parameters is given by:

$$\mathbf{J}(\mathbf{x}) = \begin{pmatrix} & & \\ \frac{\partial y_1(\mathbf{x})}{\partial p_1} & \cdots & \frac{\partial y_1(\mathbf{x})}{\partial p_q} \\ & \ddots & \\ \frac{\partial y_r(\mathbf{x})}{\partial p_1} & \cdots & \frac{\partial y_r(\mathbf{x})}{\partial p_q} \\ \frac{\partial y_1^{(1)}(\mathbf{x})}{\partial p_1} & \cdots & \frac{\partial y_1^{(1)}(\mathbf{x})}{\partial p_q} \\ & \ddots & \\ \frac{\partial y_r^{(1)}(\mathbf{x})}{\partial p_1} & \cdots & \frac{\partial y_r^{(1)}(\mathbf{x})}{\partial p_q} \\ & \ddots & \end{pmatrix} \tag{3.6}$$

A non-regular point, θ, satisfies the condition, $\det(J^k(\theta)) = 0$, $k = 1, \ldots, \infty$ where the superscript denotes the kth selection of q rows. If the set of non-regular points has measure zero in Ω then the model is locally identifiable if and only if there exists some α such that $J(\alpha)$ has rank q. Note that when the elements of the Jacobian are polynomial in p this condition holds. Thus by evaluating the Jacobian matrix at a point it is possible to establish that a model is locally identifiable. This condition does not demonstrate that the model is not globally identifiable [166]. Since differentiation simplifies the coefficients to be used this approach is less computationally demanding than the Taylor series approach.

This test was implemented in *Maple* as follows. This approach is applied to a simple model in Section 3.2.4.

Steps 1 and 2 were implemented as above. Note that in order for the Jacobian matrix to have sufficient rank at least q non-zero Taylor series coefficients are needed.

Step 3: Computation of the rank of a Jacobian matrix

There is a package in *Maple* which will compute the Jacobian matrix of two vectors.

```
\with(VectorCalculus}:
J1 := Jacobian([g_1_1, ... , g_a_b], [parameter_name_1,
 ... , parameter_name_m]):
```

The `VectorCalculus` package should be called only after the Taylor series coefficients have been computed as it changes certain definitions that are used during these calculations. The vector of Taylor series coefficients should include all nonzero coefficients computed. The vector of parameters should list all unknown model parameters. The rank of this Jacobian matrix can then be calculated using another *Maple* package.

```
\with(LinearAlgebra):
\rank(J1);
```

For the models analyzed *Maple* was able to compute a Jacobian matrix with the resources available. If this is not true it is sometimes possible to partially simplify the Jacobian matrix to allow *Maple* to complete the task. No algorithm is provided as this simplification must be done by hand.

3.2.3 Input-output relationship approach

An alternative to the Taylor series approach, an approach, deriving from differential algebra, uses the (input-)output relation associated with the model in question to determine identifiability. The differential equations and equilibrium relations defining the system model are the generators of a radical differential ideal [167]. This differential ideal can be decomposed into an intersection of differential ideals using the Rosenfeld Gröbner algorithm, a generalisation of the division algorithm to differential polynomials [168]. The resulting decomposition corresponds to the general and singular solutions of the differential equations [169]. Given an

appropriate choice of ranking each ideal contains expressions in derivatives of the output (and input) functions only, the (input-)output relation [170]. Given that the models used have no input functions; this relation will be referred to as the output relation. Furthermore since a single output function is used in each case the output relation will consist of a single function denoted $h(y(\mathbf{p},t), y^{(1)}(\mathbf{p},t), \ldots, y^{(i)}(\mathbf{p},t), \mathbf{p})$. The following analysis can be generalised for multiple output functions with a corresponding vector output relation.

The output relation is a generator of an ideal and as such must equal zero. Since the coefficients of this differential polynomial are generically non-zero, they are linear combinations of the elements of the parameter vector, $\mathbf{p}$, the monomials are linearly dependent. It can however be rearranged to give the highest order derivative of $\mathbf{y}(t)$ as a possibly rational function of strictly lower order derivatives:

$$\mathbf{y}^{(i)}(t) = \hat{h}(\mathbf{y}(t), \mathbf{y}^{(1)}(t), \ldots, \mathbf{y}^{(i-1)}(t), \mathbf{p}). \tag{3.7}$$

This relation is used in the analysis of structural identifiability as follows. Two indistinguishable parameter vectors, $\mathbf{p}$ and $\overline{\mathbf{p}}$ satisfy Equation (3.4). Consequently they also satisfy:

$$\mathbf{y}^{(1)}(t, \mathbf{p}) = \mathbf{y}^{(1)}(t, \overline{\mathbf{p}}), \text{ for all } t \geq 0, \tag{3.8}$$

since Equation (3.4) holds for all t and $\mathbf{x}(t, \mathbf{p})$ and thus $\mathbf{y}(t, \mathbf{p})$ are analytic. Furthermore by induction:

$$\mathbf{y}^{(n)}(t, \mathbf{p}) = \mathbf{y}^{(n)}(t, \overline{\mathbf{p}}), \text{ for all } t \geq 0 \text{ and } n \in \mathbb{N}, \tag{3.9}$$

where $\mathbb{N}$ denotes the natural numbers. Thus for indistinguishable parameter vectors:

$$\begin{aligned} &\mathbf{y}^{(i)}(t, \mathbf{p}) - \hat{h}(\mathbf{y}(t), \mathbf{y}^{(1)}(t), \ldots, \mathbf{y}^{(i-1)}(t), \mathbf{p}) \\ &\qquad = \mathbf{y}^{(i)}(t, \overline{\mathbf{p}}) - \hat{h}(\mathbf{y}(t), \mathbf{y}^{(1)}(t), \ldots, \mathbf{y}^{(i-1)}(t), \overline{\mathbf{p}}), \end{aligned} \tag{3.10}$$

which can then be rearranged to obtain:

$$0 = \hat{h}(\mathbf{y}(t), \mathbf{y}^{(1)}(t), \ldots, \mathbf{y}^{(i-1)}(t), \overline{\mathbf{p}}) - \hat{h}(\mathbf{y}(t), \mathbf{y}^{(1)}(t), \ldots, \mathbf{y}^{(i-1)}(t), \mathbf{p}). \tag{3.11}$$

The second expression, Equation (3.11), can now be rearranged to give a differential polynomial. Assuming the monomials of this differential polynomial are linearly independent their coefficients must all be zero. As such these coefficients can be solved for $\overline{\mathbf{p}}$. A single solution implies that the model is SGI; countably many solutions that it is SLI; and uncountably many that the model is unidentifiable.

Thus it is critical to determine whether the monomials of this differential polynomial are linearly independent. If it can be rewritten as a simple polynomial then the Fundamental Theorem of Algebra can be applied. However, where this is not the case an alternative approach is needed; this is discussed in the following section (Section 3.2.3.1).

For exceptionally simple systems it is possible to apply this approach by hand. However in general the algebraic manipulations required are best handled by a computer algebra package. This approach was implemented in *Maple* as follows [158]. It is assumed that a single species is measured and that there are no inputs. This approach is applied to a simple model in Section 3.2.4.

Step 1: Definition of the model

A field extension of the rational numbers and a differential ring are defined as follows:

```
with(diffalg):
F := field_extension(transcendental_elements = CONSANTS):
R := differential_ring(derivations = [t], ranking = RANKING,
field_of_constants = F, notation = diff):
```

The first command indicates that the differential algebra package will be used. `CONSTANTS`

is the set of constants associated with the model, i.e. the elements of **p** and non-zero initial conditions. `RANKING` is an ordered list of the symbols denoting the reaction species. The last element in this list should be the measured species. The model equations are then defined as follows:

```
eqn_mi := diff(x_i, t) - f_i(x):
eqn_ei := Ci - g_i(x):
```

Thus the derivatives of the state vector are rearranged and then denoted by the `eqn_mi`s, while the conservation relations are denoted by the `eqn_ei`s.

Step 2: Derivation of a differential polynomial from the output relation

An autoreduced (see Section 3.2.3.1) ordered subset of the model equations is selected, generally containing all the equilibrium relations, and denoted `GenId`. The Rosenfeld Gröbner algorithm is then used to calculate a decomposition of the ideal generated by these equations [170].

```
P1 := Rosenfeld_Groebner(GenId, R);
```

The generators of the ideals of this decomposition can be output using the following code:

```
rewrite_rules(P1)[i];
```

Where this decomposition contains multiple ideals, ideals describing singular model states, that is cases which could be eliminated experimentally, are eliminated by inspection. The output relation of the remaining ideal, the general solution, is the last generator of the remaining ideal and is denoted `inout`. Let the highest derivative in the output relation have order i. The output relation is converted to a multinomial by substitution of the derivatives by an alternative symbol.

```
SUBSA := diff(y(t), [t$i]) = y_i, diff(y(t), [t$i-1]) = y_i-1, ... , y(t) = y:
inout := subs(SUBSA, inout):
```

The relation is then solved for the highest order derivative to obtain `inout1`. An alternative parameter vector is defined as described in Section 3.2.1 and substituted into `inout1` to obtain `inout2`. The highest order derivative is then eliminated by subtracting `inout2` from `inout1`. The result is converted to a polynomial form.

```
inout1 := solve(inout, y_i):
inout2 := subs(SUBS2, inout1):
outrel := denom(simplify(inout1 - inout2))*simplify(inout1 - inout2):
```

Step 3: Analysis of the coefficients of the differential polynomial obtained

The highest degree of each `y_j` are determined and denoted `o_j`. The coefficients of the monomials of `outrel` are separated from their monomials using the `coeff` function.

```
coeffs_0 := 0:
for k from 0 by 1 to o_0 do
coeffs_0 := coeffs_0, coeff(outrel, y_0, k):
end do:
coeffs_0 := {coeffs_0}:


coeffs_j := 0:
for k from 2 by 1 to nops(coeffs_j-1) do
for l from 0 by 1 to o_j do
coeffs_j := coeffs_j, coeff(coeffs_j[k], y_j, l):
end do:
end do:
coeffs_j := coeffs_j
```

The final set obtained, `coeffs_i-1`, contains all the monomial coefficients. It is then solved for the alternative parameter vector.

```
s:=solve(coeffs_i-1, {alt_parameter_1, ... , alt_parameter_name_m}):
```

A single solution implies that the model is SGI; countably many solutions that it is SLI; and uncountably many that the model is unidentifiable.

3.2.3.1 Establishing linear independence of monomials

The linear independence of the monomials of the differential polynomial derived from Equation (3.11) can be checked by computation of a Wronskian. The Wronskian of the monomials is the determinant of the square matrix constructed by using the monomials as the first row and the i-th derivative of this row as the $i+1$-th row. If there exists a time point at which the Wronskian is non-zero, the monomials are linearly independent [171]. However typically one of the monomials will be the output function $y(t, \mathbf{p})$. As such using this approach is ultimately equivalent in complexity to calculation of $i-1$ Taylor series coefficients, where i is the number of monomials in the input-output relation, and thus to undertaking a Taylor series analysis. An alternative test can be derived using concepts from differential algebra.

For a more formal introduction to these concepts see Ritt [167] or more recently Boulier *et al.* [168]. Define a ranking on a differential ring, $\mathbb{K}\{y_1, \ldots, y_n\}$, to be a total ordering on the ring which is compatible with the differentiations over the alphabet; specifically ranking is preserved by differentiation and higher order derivatives are ranked higher than lower order derivatives. For a differential polynomial, p of $\mathbb{K}\{y_1, \ldots, y_n\}$, and a ranking R, the **leader**, u, of p is the largest derivative with respect to the ranking R which appears in p. Let d be the degree of u in p. The **initial**, I_p, of p is the coefficient of u^d in p. The **separant**, S_p, of p is the initial of all the proper derivatives of p. The differential polynomial q is **partially reduced** with respect to p if no proper derivative of u appears in q. It is **reduced** with respect to p if it is partially reduced with respect to p and its degree in u is less than d.

A set of polynomials A is **triangular** if its elements have different leaders. It is **autoreduced**

if each element is reduced with respect to every other element. It is **orthonomic** if the initials and separants of its elements are in $\mathbb{K}$ [168, 170].

If a set of generators is autoreduced and orthonomic then the ideal generated by them is prime [168] and the corresponding variety is irreducible. As such the generators cannot be factorised and must have a single solution. The output relation is a generator of the differential ideal, albeit restated in an alternative form, and thus shares this property.

So if the leader is eliminated, as described in Section 3.2.3 (Equation (3.11)) the resulting expression can only be identically zero if the coefficients of the remaining monomials are zero; or equivalently the monomials are linearly independent. It is however necessary to check that the initial is generically non-zero since, if this is not the case, the elimination process would involve division by zero.

If the initial is generically zero then it is a generator of the ideal arising from the system model. Thus the ideal defined using the system equations and the initial should be equivalent to that obtained just from the system equations. As such their decompositions under the Rosenfeld Gröbner algorithm should be identical. If this is not the case then the initial is generically non-zero. For the cases considered, in Chapters 4 and 6, the resultant decomposition is a subset of the original decomposition constituent of ideals describing singular reaction states demonstrating that the initial is generically non-zero.

In summary in order to determine whether the monomials of differential polynomial derived from Equation (3.11) are linearly independent it is sufficient to check that the generators of the ideal are orthonomic and autoreduced with respect to some ranking and that the initial of the output relation is generically non-zero.

3.2.4 Application of the identifiability analysis techniques to a simple model

The Taylor series approach, Pohjanpalo's Jacobian rank test and the input-output relationship approach are now applied to the one substrate enzyme catalysed reaction mechanism below:

$$\begin{aligned} E + S &\underset{r_1}{\overset{k_1}{\rightleftharpoons}} ES \\ ES &\overset{k_2}{\rightarrow} E + P \end{aligned} \tag{3.12}$$

where E, S, ES and P represent enzyme, substrate, enzyme-substrate complex and product respectively. The concentrations of these species constitute the model state vector, denoted $\mathbf{x}(t, \mathbf{p})$. It and the initial conditions used are given below:

$$\mathbf{x}(t, \mathbf{p}) = (E(t, \mathbf{p}), S(t, \mathbf{p}), ES(t, \mathbf{p}), P(t, \mathbf{p}))^T, \tag{3.13}$$

$$\mathbf{x}(0, \mathbf{p}) = \mathbf{x}_0 = (E_0, S_0, 0, 0)^T. \tag{3.14}$$

The derivative of the state vector is given by:

$$\frac{d\mathbf{x}(t, \mathbf{p})}{dt} = \begin{pmatrix} -k_1 E(t, \mathbf{p}) S(t, \mathbf{p}) + (r_1 + k_2) ES(t, \mathbf{p}) \\ -k_1 E(t, \mathbf{p}) S(t, \mathbf{p}) + r_1 ES(t, \mathbf{p}) \\ k_1 E(t, \mathbf{p}) S(t, \mathbf{p}) - (r_1 + k_2) ES(t, \mathbf{p}) \\ k_2 ES(t, \mathbf{p}) \end{pmatrix}, \tag{3.15}$$

where k_1 and r_1 are the association and dissociation rate constants for the formation of the complex, and k_2 is the rate of irreversible breakdown of the complex. These constants are the elements of the unknown parameter vector denoted $\mathbf{p}$. The observation considered, denoted $y(t, \mathbf{p})$, will be measurement of the concentration of P:

$$y(t, \mathbf{p}) = P(t, \mathbf{p}). \tag{3.16}$$

The first five coefficients of the Taylor series expansion of Equation (3.16) at $t = 0$ are given below:

$$y(0, \mathbf{p}) = 0, \tag{3.17}$$

$$y^{(1)}(0, \mathbf{p}) = 0, \tag{3.18}$$

$$y^{(2)}(0, \mathbf{p}) = k_1 k_2 E_0 S_0, \tag{3.19}$$

$$y^{(3)}(0, \mathbf{p}) = -k_1 k_2 E_0 S_0 (k_1 (E_0 + S_0) + r_1 + k_2), \tag{3.20}$$

$$\frac{y^{(4)}(0, \mathbf{p})}{k_1 k_2 E_0 S_0} = (k_1^2 (E_0^2 + 4 S_0 E_0 + S_0^2) + 2 k_1 k_2 (E_0 + S_0) + r_1 k_1 (E_0 + 2 S_0) + (r_1 + k_2)^2). \tag{3.21}$$

The rank of the Jacobian matrix of these coefficients with respect to the unknown parameter vector is three. Thus this model is at least locally identifiable by Pohjanpalo's Jacobian rank test. The alternative unknown parameter vector given by:

$$\overline{\mathbf{p}} = (\overline{k_1}, \overline{k_2}, \overline{r_1})^T, \tag{3.22}$$

was used to construct simultaneous equations in the two sets of unknown parameters. For example using Equation (3.19):

$$0 = y^{(2)}(0, \mathbf{p}) - y^{(2)}(0, \overline{\mathbf{p}}) = (k_1 k_2 - \overline{k_1 k_2}) E_0 S_0. \tag{3.23}$$

The equations obtained from the first five Taylor series coefficients were solved for the alternative parameters producing two solutions. The equation obtained from the sixth Taylor series coefficient eliminated one of these solutions leaving $\overline{\mathbf{p}} = \mathbf{p}$, thus this model is structurally globally identifiable.

The model equations are rearranged into the conventional form for generators of a differential

ideal below:

$$E^{(1)}(t,\mathbf{p}) + k_1 E(t,\mathbf{p})S(t,\mathbf{p}) - (r_1 + k_2)ES(t,\mathbf{p}) = 0, \tag{3.24}$$

$$S^{(1)}(t,\mathbf{p}) + k_1 E(t,\mathbf{p})S(t,\mathbf{p}) - r_1 ES(t,\mathbf{p}) = 0, \tag{3.25}$$

$$ES^{(1)}(t,\mathbf{p}) - k_1 E(t,\mathbf{p})S(t,\mathbf{p}) + (r_1 + k_2)ES(t,\mathbf{p}) = 0, \tag{3.26}$$

$$P^{(1)}(t,\mathbf{p}) - k_2 ES(t,\mathbf{p}) = 0, \tag{3.27}$$

$$E(t,\mathbf{p}) + ES(t,\mathbf{p}) - E_0 = 0, \tag{3.28}$$

$$S(t,\mathbf{p}) + P(t,\mathbf{p}) - S_0 + ES(t,\mathbf{p}) = 0. \tag{3.29}$$

The last four of these generators, Equations (3.26)-(3.29), constitute an orthonomic and autoreduced subset of these generators, for any ranking for which the members of the following set, $\{E,\ S\}$, are ranked above the remaining variables, $\{ES,\ P\}$. As such the ideal generated by these expressions is prime by the argument presented in Section 3.2.3.1.

In this case an input-output relationship can be derived by hand. First the conservation relations, Equations (3.28) and (3.29), are used to eliminate $E(t,\mathbf{p})$ and $S(t,\mathbf{p})$ from Equations (3.26) and (3.27), $P(t,\mathbf{p})$ is relabelled $y(t,\mathbf{p})$. The rearrangement then proceeds as follows, from Equation (3.27):

$$y^{(1)}(t,\mathbf{p}) = k_2 ES(t,\mathbf{p}), \tag{3.30}$$

$$ES(t,\mathbf{p}) = \frac{y^{(1)}(t,\mathbf{p})}{k_2}, \tag{3.31}$$

$$ES^{(1)}(t,\mathbf{p}) = \frac{y^{(2)}(t,\mathbf{p})}{k_2}. \tag{3.32}$$

Substituting these expressions into Equation (3.26) gives:

$$\frac{y^{(2)}(t,\mathbf{p})}{k_2} = k_1(E_0 - \frac{y^{(1)}(t,\mathbf{p})}{k_2})(S_0 - \frac{y^{(1)}(t,\mathbf{p})}{k_2} - y(t,\mathbf{p})) - (r_1 + k_2)\frac{y^{(1)}(t,\mathbf{p})}{k_2}. \tag{3.33}$$

The input-output relationship contains the monomials: $z_1 = 1, z_2 = y(t, \mathbf{p})$, $z_3 = y^{(1)}(t, \mathbf{p})$, $z_4 = y(t, \mathbf{p})y^{(1)}(t, \mathbf{p})$, $z_5 = (y^{(1)}(t, \mathbf{p}))^2$, $z_6 = y^{(2)}(\mathbf{p})$. A second expression was created using the alternative parameter vector and subtracted from Equation 3.33 eliminating z_6 to yield:

$$\begin{aligned} 0 = \; & (k_1k_2 - \overline{k_1k_2})E_0S_0z_1 - (k_1k_2 - \overline{k_1k_2})E_0z_2 + (k_1 - \overline{k_1})z_4 + (\tfrac{k_1}{k_2} - \tfrac{\overline{k_1}}{\overline{k_2}})z_5 - \\ & - ((k_1 - \overline{k_1})(E_0 + S_0) + (r_1 - \overline{r_1}) + (k_2 - \overline{k_2}))z_3. \end{aligned} \tag{3.34}$$

Thus the relations to be solved are:

$$\left.\begin{array}{ll} k_1 - \overline{k_1} = 0, & (k_1\overline{k_2} - \overline{k_1}k_2)/(k_2\overline{k_2}) = 0, \\ k_1k_2 - \overline{k_1k_2} = 0, & (k_1 - \overline{k_1})(E_0 + S_0) + (r_1 - \overline{r_1}) + (k_2 - \overline{k_2}) = 0. \end{array}\right\} \tag{3.35}$$

A unique solution to these relations exists, $\overline{\mathbf{p}} = \mathbf{p}$, thus the results of the Taylor series and input-output relationship approaches agree. In this case it is possible to check linear independence of the monomials by computation of a Wronskian[1] as noted in Section 3.2.3.1; the monomials z_1 to z_5 are linearly independent as expected based on the argument presented in that section.

3.3 Structural indistinguishability analysis

Where structural identifiability is used to determined whether two parameter vectors can give rise to the same model output, structural indistinguishability is concerned with whether two models structures do likewise for some feasible pair of parameter vectors. As such structural identifiability is, strictly speaking, a specific case of structural indistinguishability although typically they are treated separately [149, 150]. Similar techniques are used for indistinguishability analysis as have been described in the previous section.

These concepts are now formally defined. Suppose there exist two state space models

[1] The file containing this calculation is available electronically on request and at: http://www2.warwick.ac.uk/fac/sci/moac/currentstudents/2005/daniel_bearup.

$\boldsymbol{\Sigma}(\mathbf{x}, \mathbf{f}, \mathbf{x}_0, \mathbf{y}, \mathbf{C}, \mathbf{p})$ and $\widehat{\boldsymbol{\Sigma}}(\widehat{\mathbf{x}}, \widehat{\mathbf{f}}, \widehat{\mathbf{x}}_0, \widehat{\mathbf{y}}, \widehat{\mathbf{C}}, \widehat{\mathbf{p}})$ as defined previously, Equations (3.1)-(3.3). Analogously to Equation (3.4) the two parameter vectors are said to be indistinguishable if:

$$\mathbf{y}(t, \mathbf{p}) = \widehat{\mathbf{y}}(t, \widehat{\mathbf{p}}), \text{ for all } t \geq 0. \tag{3.36}$$

The models $\boldsymbol{\Sigma}$ and $\widehat{\boldsymbol{\Sigma}}$ are **structurally indistinguishable** if:

$$\left.\begin{array}{r} \text{for all } \mathbf{p} \text{ there exists } \widehat{\mathbf{p}} \text{ such that } \mathbf{p} \sim \widehat{\mathbf{p}}, \\ \text{and for all } \widehat{\mathbf{p}} \text{ there exists } \mathbf{p} \text{ such that } \mathbf{p} \sim \widehat{\mathbf{p}}. \end{array}\right\} \tag{3.37}$$

Otherwise the models are **structurally distinguishable**[172].

The Taylor series and input-output relationship approaches described in Sections 3.2.1 and 3.2.3 can be generalised to indistinguishability analysis; the principles of both techniques remain the same. The Laplace transformation approach, applicable only to linear models, and the similarity transformation approach can also be adapted for indistinguishability analysis [150]. However they are not considered here.

For the Taylor series approach, a Taylor series expansion of the output function of both models is calculated. Taylor series coefficients of the same order are then used to construct simultaneous equations in the two parameter vectors. The resulting equations are then solved, if no solution can be found the models are distinguishable. Note that proving indistinguishability falls prey to the same difficulties as proving unidentifiability. If only a finite number of Taylor series coefficients are used it is possible that the models could be shown to be distinguishable using additional Taylor series coefficients.

Similarly for the input-output relationship approach, an output relation for each model is calculated. In all cases where this was applied the two models had the same leader, which was eliminated by subtraction of one output relation from the other. For the models considered

this yielded a simple, rather than differential, polynomial the monomials of which were linearly independent by the Fundamental Theorem of Algebra. This approach has not yet been generalised for models which do not yield simple polynomials at this stage. The coefficients of the monomials were then used to construct simultaneous equations in the two parameter vectors which were then solved.

For both approaches it was sometimes necessary to subject the solutions to case analysis on the initial conditions to determine whether the solutions were in fact feasible. In some cases experimental constraints prevent the use of initial conditions which would otherwise render the models distinguishable; thus while these models may be structurally distinguishable, they are practically indistinguishable.

3.4 Numerical parameter estimation

This work makes extensive use of numerical parameter estimation in the analysis of experimental data. In this section methods by which parameters may be estimated from experimental data are discussed. Two cases are considered, first the issue of nonlinear curve fitting to data obtained by typical kinetic characterisation techniques, and second the estimation of parameters for a dynamic model from full experimental time courses. In the latter case structural identifiability and indistinguishability analyses are an essential precursor to the numerical techniques. The methods used are described below; they all use least squares minimisation.

The kinetic characterisations undertaken in this work (Chapter 5) involve the measurement of initial rates of reaction for a range of initial substrate concentrations. A single initial substrate concentration is varied while the remaining concentrations are kept constant. The use of mechanical mixing in these experiments allows the reactions to reach the quasi-steady state described by the Michaelis-Menten approximation. As such the data collected can be fitted

using the non-linear equations:

$$v = \frac{V_{max}C}{(C + k_m)}, \tag{3.38}$$

$$v = \frac{V_{max}C}{(C + k_m + \frac{C^2}{k_i})}. \tag{3.39}$$

The variables v and C denote the initial rate of reaction and initial substrate concentration respectively. The parameters V_{max}, k_m and k_i correspond to kinetic constants to be estimated. Equation (3.38) describes uninhibited reaction kinetics while Equation (3.39) describes the effect of substrate inhibition.

Two software packages were used to estimate parameters for these characterisations: *xmgrace* (Grace Development Team) and *Facsimile 4* (MCPA Software, UK). A first set of parameter estimates were obtained in *xmgrace* which implements the Levenburg-Marquardt algorithm, described below. This algorithm converges well regardless of initial parameter values. However the software does not generate particularly detailed statistical analysis of these fits. The fitting algorithm used by *Facsimile 4* converges poorly if initial parameter estimates do not describe the data well. However it provides in-depth statistical analysis of resultant parameter estimates. Thus, in order to obtain this analysis, the estimates obtained using *xmgrace* were used as initial parameter values in *Facsimile 4*. Details of the *Facsimile 4* fitting algorithm can be found in the Facsimile 4 Technical Guide [173].

Confidence intervals for the parameter estimates were estimated in *Facsimile*; assuming a log-normal distribution. The basic technique is described below.[2]

An alternative to this approach was considered, whereby the parameters of the differential equations describing the reaction were estimated directly from time course data. *xmgrace* is unable to estimate parameter values for sets of differential equations. As such an alternative to

[2] Sample code is available electronically on request and at:
http://www2.warwick.ac.uk/fac/sci/moac/currentstudents/2005/daniel_bearup.

this program was required if parameters were to be estimated for time course data. The fitting algorithms in *COPASI* converge adequately regardless of parameter values [159]. They were not used for parameter estimation based on initial rate data since they are not optimised to a system which is not constructed from differential equations. In addition they require sufficient statistical analysis of the parameter estimates obtained to obviate the need for *Facsimile 4*.

The quasi-steady state equations are not implemented in this program and were added as needed, the equations used can be found in Sections 4.2.1 and 4.2.2. Transient models were implemented using the reversible and irreversible mass action functions to describe each complex formation or degradation step; the models used are described in Sections 4.2.1 and 4.2.2. In general the Levenburg-Marquardt algorithm was used [174]; the iteration limit and tolerance being increased to 1000 and 10^{-12} from defaults for best fits. The algorithm proceeds as follows. Given an estimated parameter vector $\mathbf{p}_i$ a new estimate is obtained by adding $\underline{\delta}$ to obtain $\mathbf{p}_{i+1} = \mathbf{p} + \underline{\delta}$. The increment $\underline{\delta}$ is determined from the following equation [174]:

$$(\mathbf{J}^T\mathbf{J} + \lambda \operatorname{diag}(\mathbf{J}^T\mathbf{J}))\underline{\delta} = \mathbf{J}^T(\mathbf{Y} - \mathbf{F}(\mathbf{p}_i)). \tag{3.40}$$

The experimental data and model curve are represented by the vectors, $\mathbf{Y}$ and $\mathbf{F}(\mathbf{p}_i)$, $\mathbf{J}$ is the Jacobian of $\mathbf{F}$ with respect to $\mathbf{p}$. The diagonal of a matrix A is denoted by $\operatorname{diag}(A)$.

For some data, where the response was relatively insensitive to the model parameters, the Levenburg-Marquardt algorithm failed to converge in reasonable time. In these cases an evolutionary programming algorithm was used [175]. This algorithm, using as it does a more random approach to the minimum, allowed a greater exploration of the solution space; allowing results to be obtained from otherwise intractable data. However this increased ability to find a solution also results in less consistent solutions being obtained between runs of the algorithm. As such it was used only as a last resort. The algorithm is briefly described below.

Given a population of 200 parameter vectors, a daughter population of 400 is generated by

adding a mutation, a random modification, of each member of the population. An R^2 value is determined for each member of the population, $\mathbf{p}_j$ using the expression below:

$$R_j^2 = \frac{\sum_{i=1}^{N}(\mathbf{Y}_i - \mathbf{F}_i(\mathbf{p}_j))}{N}. \tag{3.41}$$

The number of data points is denoted N, $\mathbf{Y}_i$ and $\mathbf{F}_i(\mathbf{p}_j)$ denote the i-th elements of the experimental data and model curve vectors respectively. The elements of the population were ranked according to these values; the 200 with the highest R^2 values are eliminated. The process is then repeated; 1000 generations were used.

Confidence intervals were estimated manually for these parameters using the statistical data provided by *COPASI* in *Microsoft Excel* (Microsoft Ltd, Reading, UK): assuming that error was normally distribution. The basic technique was as follows. For a given distribution of error, X, confidence intervals were estimated for each parameter using the following equation:

$$v_{\pm} = v \pm X(p) \cdot se(v). \tag{3.42}$$

Where v is the estimated parameter value, p is a probability such that $1-2p$ gives the required confidence level and $se(v)$ is the standard error (or standard deviation) associated with v. The value denoted by $X(p)$ is the value at which $P(X > x) = p$ for the standardised distribution X.

3.5 Simulation of reaction species concentrations

The concentrations of reaction species were predicted by numerical integration of a model system. The trapezoidal integration method was implemented manually in *Microsoft Excel* and was used to eliminate background production from experimental data. Simulation of time courses for a single reaction were undertaken in *Maple*. Prediction of more complex pathway

dynamics was undertaken using *COPASI*. These three methods are described in detail below.

Experimental observations of pathway dynamics were obtained using the assays described in Section 2.4. Over long time courses background production of the observed species could significantly affect the measured data. Rates of background production at the start and end points of the time course were estimated from experimental data. These data were combined with models of background production to extrapolate rates for all points in the time course. The resulting curve was integrated numerically using a manual implementation of the trapezoidal integration method (Equation (3.43)) in *Microsoft Excel*. The iterative function describing this method is given below:

$$x^{n+1} = x^n + \frac{t^{n+1} - t^n}{2}(f(x^{n+1}, t^{n+1}) + f(x^n, t^n)), \tag{3.43}$$

time steps are enumerated by superscripts; t, x, and f are time, concentration of species, and rate of change of species concentration respectively. This method is rather crude; being prone to accumulation of error and not suitable for stiff systems. However it is adequate to integrate the relatively low curvature functions used to model background production in the models used. The superior algorithms implemented in *Maple* or *COPASI* could have been used. However aligning the resulting time courses to the experimental data would have been a non-trivial problem and as such the simpler solution, a direct implementation in *Microsoft Excel*, was preferred.

Time courses for a single reaction were generated in *Maple* using the `dsolve` function. This uses an algorithm similar to that used by *COPASI* which is described below. *Maple* was used for this case since the subsequent analysis desired was best implemented in this program. Specifically *Maple* was used to plot rates of accumulation of certain reaction species, corresponding to experimental data, and initial rates against initial concentrations of substrates, simulating the initial rate kinetic characterisations described in Sections 2.4 and 5. In order to create equivalent

plots from data simulated using *COPASI* it would have been necessary to export the data to an alternative program.

The differential equations describing model pathways were numerically solved using the time course functions in *COPASI* [159]. This program has a number of features which make it ideal for modelling of metabolic pathways. Models of individual reactions can be selected from a list and populated with relevant parameters. Incorporation of additional reactions to an existing model is relatively simple. Finally it is possible to use stochastic or deterministic simulations, or a hybrid of both.

Time course simulations were deterministic and were produced using a variant of the Livermore Stiff ODE solver [176]. Given that the lowest particle numbers considered were of the order of 10^{11} deterministic simulations are appropriate. Options for the algorithm were left as default. Thus the algorithm was allowed to use mass conservation laws; the relative and absolute tolerances were 10^{-6} and 10^{-12} respectively, significantly higher than can reasonably be expected experimentally; and the maximal internal steps before the next reporting time was 10^5. The algorithm incorporates both stiff, backward differentiation formula (BDF), and non-stiff, Adams-Moulton, integration methods.

The Adams-Moulton and BDF methods of order q are given by the following expressions:

$$\mathbf{Y}_n = \mathbf{Y}_{n-1} + h_n \sum_{j=0}^{q-1} \beta_j \mathbf{f}_{n-j}, \tag{3.44}$$

$$\mathbf{Y}_n = \sum_{j=1}^{q} \alpha_j \mathbf{Y}_{n-j} + h_n \beta_0 \mathbf{f}_n, \tag{3.45}$$

respectively. The interval required in divided into internal, not necessarily regular, increments denoted by the length N vector $\underline{\xi}$. The step size is denoted $h_n (= \underline{\xi}_n - \underline{\xi}_{n-1})$, the difference between the n-th and $(n-1)$-th elements of $\underline{\xi}$. The estimated solution and the derivative of this solution are denoted by $\mathbf{Y}$ and $\mathbf{f}$ respectively. Subscripts denote the position in the vector

corresponding to the mesh point at the same position in $\underline{\xi}$. Thus $\mathbf{f}_n$ is the derivative of $\mathbf{Y}$ at $\underline{\xi}_n$. The coefficients $\{\alpha_j\}$ and $\{\beta_j\}$ are associated with particular orders of these methods. Equations (3.44) and (3.45) are implicit, they require solutions at mesh points which have not yet been simulated, the unknown values are predicted using explicit methods and then corrected iteratively. This approach provides excellent accuracy, that is the difference between estimated and actual values at a given point is small, and stability, that is the accumulated difference between estimated and actual values grows slowly [177].

In the implementation in *COPASI*, the maximal order the Adams method will attempt before switching to the BDF method is 12; the maximal order the BDF method will attempt before switching to smaller internal step sizes is 5.

3.6 Summary

Two primary approaches to structural identifiability and indistinguishability analysis are presented in this chapter, the Taylor series approach and the (input-)output relationship approach. The Taylor series approach is well-known and understood. However the implementation of this approach is often computationally difficult. The theory behind the (input-)output approach is, by contrast, much less developed. The test presented here to determine whether the monomials of the expression analysed are linearly independent has not been previously proposed. The relative effectiveness of, and difficulty of implementing, these two approaches in analysing models of enzyme dynamics will be assessed in the following chapters (Chapters 4 and 6).

4. IDENTIFIABILITY AND INDISTINGUISHABILITY ANALYSIS OF QUASI-STEADY STATE MODELS OF ENZYME REACTIONS

4.1 Introduction

The cytoplasmic phase of the peptidoglycan biosynthesis pathway is composed of six enzyme catalysed reactions. In order to simulate the the behaviour of this pathway it will be necessary to determine the kinetic parameters of each of these reactions. These characterisations will be undertaken using enzymological techniques as discussed in Section 3.4 and implemented in Chapter 5. However an alternative approach, direct estimation of parameters from time course data, may also be viable. It is necessary however to conduct a structural identifiability analysis of the candidate models prior to attempting such an approach.

The reaction mechanisms of the enzymes in this pathway have been determined. However when characterising a new enzyme this information would not generally be available. As such it would be necessary to determine which model for the reaction mechanism is most appropriate during the characteristion process. If the direct estimation approach is to be used this problem, of multiple possible reaction mechanisms, would require a structural indistinguishability analysis to be undertaken.

As such, in this chapter, these two forms of analysis will be undertaken for a variety of two and three substrate reaction mechanisms. The mechanisms in question will first be introduced and converted to state space models using mass action kinetics. Quasi-steady state forms of these models, which are representative of typical experimental conditions, will be derived. The

transient forms of these models will be used later (Chapter 6). Structural identifiability analysis will then used to determine which of these models are suitable for direct parameter estimation. Models will be further analyzed to determine which are distinguishable from quasi-steady state measurements using structural indistinguishability analysis. Throughout both the Taylor series and input-output relationship approaches are used.

While the input-output relationship approach has been known for some time the formal theory behind it appears to be under developed. The basis of this approach, the input-output relation of the system is well understood. However it is unclear from the literature why the monomials within this relation are linearly independent. A proof is presented in Section 3.2.3 however it seems worthwhile, where possible, to check that the two approaches yield the same results. In addition the application of both approaches to each model allows their relative strengths to be assessed. This constitutes a secondary goal for this chapter.

In the next section of the chapter transient and quasi-steady state models for first two and then three substrate enzyme catalysed reactions are derived. The results of structural identifiablity analysis of these models are presented in the subsequent section. In the fourth section structural indistinguishability analysis of the models is undertaken. Finally the last section provides a brief summary of the results presented in this chapter.

4.2 Derivation of quasi-steady state system equations

As the number of substrates involved in a reaction increases the possible reaction mechanisms increases significantly [144]. However the reaction mechanisms required to model the peptidoglycan biosynthesis pathway can be placed into four basic categories since all enzymes are expected to have ordered mechanisms and to catalyse essentially irreversible reactions. The two substrate enzymes, MurA and MurB, utilise different reaction mechanisms; ordered formation of a ternary complex and a ping-pong mechanism respectively [84, 85]. The remaining

enzymes, MurC-F, require three substrates and form a quaternary complex before releasing products [57, 64]. For ease of modelling a three substrate form of the ping-pong mechanism is also considered. Differential equations describing the transient kinetics of these reactions are derived in this section. The experimental conditions and techniques used typically allow the reactions to reach a quasi-steady state before measurements can be started. As such each of the reaction models is simplified using quasi-steady state assumptions to obtain a model corresponding to these experimental conditions.

It will also be necessary to consider how the presence of inhibitors can be incorporated into these models. This question is not covered in this section. Instead the necessary equations are provided where necessary. The derivation of these equations is analogous to that presented in this section.

4.2.1 Two substrate mechanisms

The reaction catalysed by MurB follows the two substrate ping-pong reaction scheme below:

$$\left.\begin{aligned} E + S_1 &\underset{r_1}{\overset{k_1}{\rightleftharpoons}} ES_1 \\ ES_1 &\xrightarrow{k_2} E^* + Q \\ E^* + S_2 &\underset{r_3}{\overset{k_3}{\rightleftharpoons}} E^*S_2 \\ E^*S_2 &\xrightarrow{k_4} E + P \end{aligned}\right\} \tag{4.1}$$

The substrates S_1 and S_2 each bind reversibly to one of two forms of the enzyme, E and E^* respectively. The complex formed from binding of S_1 and E, ES_1, breaks down irreversibly releasing E^* and the first of two products, Q. Similarly the complex formed from binding of S_2 and E^* regenerates E and releases the second product, P. It is assumed that each complex breaks down sufficiently fast that all products involved can be considered to be released simultaneously. The rates of forward and reverse reaction steps are denoted by k_is (M s^{-1}) and

r_is (s^{-1}) respectively.

The reaction scheme above can be represented by the following differential equations obtained using the Law of Mass Action:

$$\frac{dE(t,\mathbf{p})}{dt} = -k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) + r_1ES_1(t,\mathbf{p}) + k_4E^*S_2(t,\mathbf{p}) \tag{4.2a}$$

$$\frac{dS_1(t,\mathbf{p})}{dt} = -k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) + r_1ES_1(t,\mathbf{p}) \tag{4.2b}$$

$$\frac{dES_1(t,\mathbf{p})}{dt} = k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) - (r_1 + k_2)ES_1(t,\mathbf{p}) \tag{4.2c}$$

$$\frac{dQ(t,\mathbf{p})}{dt} = k_2ES_1(t,\mathbf{p}) \tag{4.2d}$$

$$\frac{dE^*(t,\mathbf{p})}{dt} = k_2ES_1(t,\mathbf{p}) - k_3E^*(t,\mathbf{p})S_2(t,\mathbf{p}) + r_3E^*S_2(t,\mathbf{p}) \tag{4.2e}$$

$$\frac{dS_2(t,\mathbf{p})}{dt} = -k_3E^*(t,\mathbf{p})S_2(t,\mathbf{p}) + r_3E^*S_2(t,\mathbf{p}) \tag{4.2f}$$

$$\frac{dE^*S_2(t,\mathbf{p})}{dt} = k_3E^*(t,\mathbf{p})S_2(t,\mathbf{p}) - (r_3 + k_4)E^*S_2(t,\mathbf{p}) \tag{4.2g}$$

$$\frac{dP(t,\mathbf{p})}{dt} = k_4E^*S_2(t,\mathbf{p}). \tag{4.2h}$$

Concentrations of species and rate constants are denoted by the same symbols as used in the reaction mechanism. For a typical experiment the initial concentrations of enzyme and substrates would be known, non-zero, values denoted E_0, S_{10} and S_{20} respectively, while the initial concentrations of the remaining species would be zero.

As previously observed typical experimental data will be representative of a quasi-steady state formed due to the low concentration of enzyme used. This is modelled by assuming that the two complexes and the alternative form of the enzyme rapidly reach equilibrium concentrations. The derivatives of the concentrations of these species can then be considered equal to zero. These quasi-steady state assumptions and a conservation relation on the concentration of E are stated

mathematically below:

$$\frac{dES_1(t,\mathbf{p})}{dt} = \frac{dE^*(t,\mathbf{p})}{dt} = \frac{dE^*S_2(t,\mathbf{p})}{dt} = 0, \tag{4.3}$$

$$E(t,\mathbf{p}) = E_0 - ES_1(t,\mathbf{p}) - E^*(t,\mathbf{p}) - E^*S_2(t,\mathbf{p}). \tag{4.4}$$

With these assumptions Equations (4.2c), (4.2e) and (4.2g) can be solved for $E^*(t,\mathbf{p})$, $ES_1(t,\mathbf{p})$ and $E^*S_2(t,\mathbf{p})$. The results are then substituted into the remaining equations to produce the quasi-steady state system for this model below:

$$\frac{dS_1(t,\mathbf{p})}{dt} = -\frac{V_{max}S_1(t,\mathbf{p})S_2(t,\mathbf{p})}{S_1(t,\mathbf{p})S_2(t,\mathbf{p}) + k_{m,1}S_2(t,\mathbf{p}) + k_{m,2}S_1(t,\mathbf{p})}, \tag{4.5a}$$

$$\frac{dS_2(t,\mathbf{p})}{dt} = -\frac{V_{max}S_1(t,\mathbf{p})S_2(t,\mathbf{p})}{S_1(t,\mathbf{p})S_2(t,\mathbf{p}) + k_{m,1}S_2(t,\mathbf{p}) + k_{m,2}S_1(t,\mathbf{p})}, \tag{4.5b}$$

$$\frac{dQ(t,\mathbf{p})}{dt} = \frac{V_{max}S_1(t,\mathbf{p})S_2(t,\mathbf{p})}{S_1(t,\mathbf{p})S_2(t,\mathbf{p}) + k_{m,1}S_2(t,\mathbf{p}) + k_{m,2}S_1(t,\mathbf{p})}, \tag{4.5c}$$

$$\frac{dP(t,\mathbf{p})}{dt} = \frac{V_{max}S_1(t,\mathbf{p})S_2(t,\mathbf{p})}{S_1(t,\mathbf{p})S_2(t,\mathbf{p}) + k_{m,1}S_2(t,\mathbf{p}) + k_{m,2}S_1(t,\mathbf{p})}. \tag{4.5d}$$

Parameter notation is chosen to correspond with that used by Dixon [144]. The parameters used above can be expressed in terms of the transient state rate constants as follows:

$$V_{max} = \frac{E_0k_2k_4}{k_2+k_4},\ k_{m,1} = \frac{k_4}{k_1}\frac{r_1+k_2}{k_2+k_4},\ k_{m,2} = \frac{k_2}{k_3}\frac{r_3+k_4}{k_2+k_4}, \tag{4.6}$$

a reduction in the number of rate constants describing the model from six to three. Note that the right hand sides of the differential equations above, Equations (4.5a)-(4.5d), are equivalent up to a change of sign. The expression:

$$f_1(S_1(t,\mathbf{p}),S_2(t,\mathbf{p}),\mathbf{p}) = \frac{V_{max}S_1(t,\mathbf{p})S_2(t,\mathbf{p})}{S_1(t,\mathbf{p})S_2(t,\mathbf{p}) + k_{m,1}S_2(t,\mathbf{p}) + k_{m,2}S_1(t,\mathbf{p})} \tag{4.7}$$

is analogous to that which describes one substrate Michaelis-Menten type kinetics. The quasi-

steady state model can thus be described using the new mechanism shown in Figure 4.1.

A similar derivation is now undertaken for the reaction scheme followed by the MurA catalysed reaction. This scheme, referred to in this work as the simple ordered model, is given below:

$$\left.\begin{array}{rcl} E + S_1 & \underset{r_1}{\overset{k_1}{\rightleftharpoons}} & ES_1 \\ ES_1 + S_2 & \underset{k_2}{\overset{r_2}{\rightleftharpoons}} & ES_1S_2 \\ ES_1S_2 & \xrightarrow{k_3} & E + P \end{array}\right\} \tag{4.8}$$

The substrates S_1 and S_2 reversibly bind to the enzyme, E, in order forming complexes, ES_1 and ES_1S_2. The ternary complex, ES_1S_2, then breaks down irreversibly releasing enzyme and product. As for the ping-pong model it is assumed that complex breakdown is sufficiently fast that all products can be considered to be released simultaneously.

As for the ping-pong model this scheme can be used to derive differential equations describing the transient kinetics of this reaction, see below:

$$\frac{dE(t,\mathbf{p})}{dt} = -k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) + r_1ES_1(t,\mathbf{p}) + k_3ES_1S_2(t,\mathbf{p}), \tag{4.9a}$$

$$\frac{dS_1(t,\mathbf{p})}{dt} = -k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) + r_1ES_1(t,\mathbf{p}), \tag{4.9b}$$

$$\begin{aligned}\frac{dES_1(t,\mathbf{p})}{dt} = {} & k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) - r_1ES_1(t,\mathbf{p}) - k_2ES_1(t,\mathbf{p})S_2(t,\mathbf{p}) + \\ & + r_2ES_1S_2(t,\mathbf{p}),\end{aligned} \tag{4.9c}$$

$$\frac{dS_2(t,\mathbf{p})}{dt} = -k_2ES_1(t,\mathbf{p})S_2(t,\mathbf{p}) + r_2ES_1S_2(t,\mathbf{p}), \tag{4.9d}$$

$$\frac{dES_1S_2(t,\mathbf{p})}{dt} = k_2ES_1(t,\mathbf{p})S_2(t,\mathbf{p}) - (r_2 + k_3)ES_1S_2(t,\mathbf{p}), \tag{4.9e}$$

$$\frac{dP(t,\mathbf{p})}{dt} = k_3ES_1S_2(t,\mathbf{p}). \tag{4.9f}$$

Concentrations and rate constants are denoted as previously described. The typical experimental initial conditions would be as described previously, that is initial concentrations of enzyme and substrates would be known, non-zero, values denoted E_0, S_{10} and S_{20} respectively, while

the initial concentrations of the remaining species would be zero.

Following the same reasoning as used above assume that an equilibrium concentration of the two complexes forms. The corresponding quasi-steady state assumptions and conservation relation are stated below:

$$\frac{dES_1(t,\mathbf{p})}{dt} = \frac{ES_1S_2(t,\mathbf{p})}{dt} = 0, \tag{4.10}$$

$$E(t,\mathbf{p}) = E_0 - ES_1(t,\mathbf{p}) - ES_1S_2(t,\mathbf{p}), \tag{4.11}$$

Using these assumptions Equations (4.9c) and (4.9e) can be solved for $ES_1(t,\mathbf{p})$ and $ES_1S_2(t,\mathbf{p})$ as before. The following quasi-steady state system for this model can then be derived:

$$\frac{dS_1(t,\mathbf{p})}{dt} = -\frac{k_3E_0S_1(t,\mathbf{p})S_2(t,\mathbf{p})}{S_1(t,\mathbf{p})S_2(t,\mathbf{p}) + k_{m,1}S_2(t,\mathbf{p}) + k_{m,2}S_1(t,\mathbf{p}) + k_{m,2}k_{s,12}}, \tag{4.12a}$$

$$\frac{dS_2(t,\mathbf{p})}{dt} = -\frac{k_3E_0S_1(t,\mathbf{p})S_2(t,\mathbf{p})}{S_1(t,\mathbf{p})S_2(t,\mathbf{p}) + k_{m,1}S_2(t,\mathbf{p}) + k_{m,2}S_1(t,\mathbf{p}) + k_{m,2}k_{s,12}}, \tag{4.12b}$$

$$\frac{dP(t,\mathbf{p})}{dt} = \frac{k_3E_0S_1(t,\mathbf{p})S_2(t,\mathbf{p})}{S_1(t,\mathbf{p})S_2(t,\mathbf{p}) + k_{m,1}S_2(t,\mathbf{p}) + k_{m,2}S_1(t,\mathbf{p}) + k_{m,2}k_{s,12}}. \tag{4.12c}$$

Parameter notation is largely chosen to correspond with that used by Dixon [144] with the exception of V_{max}, the maximum rate of the reaction, which is denoted here by k_3E_0. The parameters used above can be expressed in terms of the transient state rate constants as follows:

$$k_{m,1} = \frac{k_3}{k_1},\ k_{m,2} = \frac{r_2 + k_3}{k_2},\ k_{s,12} = \frac{r_1}{k_1}, \tag{4.13}$$

an overall reduction in the number of rate constants, from five to four. Again note that the right hand sides of the differential equations above, Equations (4.12a)-(4.12c), are equivalent up to a change of sign, and the expression governing their behaviour is:

$$f_2(S_1,S_2,\mathbf{p}) = \frac{k_3E_0S_1(t,\mathbf{p})S_2(t,\mathbf{p})}{S_1(t,\mathbf{p})S_2(t,\mathbf{p}) + k_{m,1}S_2(t,\mathbf{p}) + k_{m,2}S_1(t,\mathbf{p}) + k_{m,2}k_{s,12}}. \tag{4.14}$$

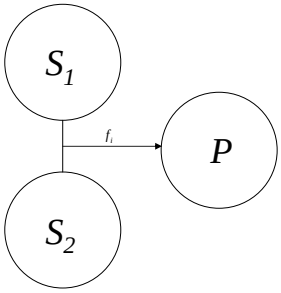


Figure 4.1: Diagram of the general two substrate quasi-steady state reaction mechanism. The rate of conversion of substrates to products is given by a function, f_i, derived as above. For the ping-pong model, $i = 1$, and Equation (4.7) is used, while for the simple ordered model, $i = 2$, and Equation (4.14) is used. All products released are represented by P, its associated state variable would be a vector of the concentrations of the various products released. If all initial product concentrations are equal all subsequent product concentrations will be equal.

Aside from an additional constant term in the denominator $k_{m,2}k_{s,12}$ the remainder of the expression is identical in form to $f_1(S_1(t,\mathbf{p}), S_2(t,\mathbf{p}), \mathbf{p})$ as defined above, Equation (4.7). By rearranging $f_2(S_1(t,\mathbf{p}), S_2(t,\mathbf{p}), \mathbf{p})$ as follows:

$$f_2(S_1(t,\mathbf{p}), S_2(t,\mathbf{p}), \mathbf{p}) = \frac{k_3 E_0}{1 + \frac{k_{m,1}}{S_1(t,\mathbf{p})} + \frac{k_{m,2}}{S_2(t,\mathbf{p})}\left(1 + \frac{k_{s,12}}{S_1(t,\mathbf{p})}\right)} \tag{4.15}$$

it becomes clear that this additional term causes an increase in the effective k_m with respect to the second substrate. This is similar to competitive inhibition of the second stage of the reaction [144]; however in this case the inhibitory effect is inversely related to the concentration of the first substrate. When $S_1 = k_{s,12}$, $k_{m,2}$ is doubled, thus twice as much substrate S_2 is required to achieve the rate of a ping-pong mechanism with equivalent parameters. Thus $k_{s,12}$ determines the degree to which the second stage of the reaction is dependent on the first. As $k_{s,12}$ approaches zero the model approaches a ping-pong mechanism. As in the case of the ping-pong mechanism the quasi-steady state form of this mechanism can be rewritten as shown in Figure 4.1.

4.2.2 Three substrate mechanisms

Reaction mechanisms for three substrate enzymes are now considered. None of the three substrate enzymes studied use the ping-pong mechanism however it is a useful approximation of their kinetics and is easier to determine experimental parameters for than the correct simple ordered mechanism. As such it will be the first mechanism considered, a schematic is given below:

$$\left.\begin{aligned} E + S_1 &\underset{r_1}{\overset{k_1}{\rightleftharpoons}} ES_1 \\ ES_1 &\xrightarrow{k_2} E^* + R \\ E^* + S_2 &\underset{r_3}{\overset{k_3}{\rightleftharpoons}} E^*S_2 \\ E^*S_2 &\xrightarrow{k_4} E^\bullet + Q \\ E^\bullet + S_3 &\underset{r_5}{\overset{k_5}{\rightleftharpoons}} E^\bullet S_3 \\ E^\bullet S_3 &\xrightarrow{k_4} E + P \end{aligned}\right\} \tag{4.16}$$

The substrates S_1, S_2 and S_3 each bind reversibly to one of three forms of the enzyme, E, E^*, $E^\bullet$ respectively. The complex formed from binding of S_1 and E, ES_1, breaks down irreversibly releasing E^* and the first of three products, R. Similarly the complex formed from binding of S_2 and E^* releases the product Q and $E^\bullet$. Finally E is regenerated by the breakdown of the complex formed from S_3 and $E^\bullet$ and the final product P is released. Each release stages is assumed to be sufficiently fast that all products involved can be considered to be released simultaneously.

Proceeding as before the corresponding differential equations for the transient system were found and are presented below:

$$\frac{dE(t,\mathbf{p})}{dt} = -k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) + r_1ES_1(t,\mathbf{p}) + k_6E^\bullet S_3(t,\mathbf{p}), \tag{4.17a}$$

$$\frac{dS_1(t,\mathbf{p})}{dt} = -k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) + r_1ES_1(t,\mathbf{p}), \tag{4.17b}$$

$$\frac{dES_1(t,\mathbf{p})}{dt} = k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) - (r_1 + k_2)ES_1(t,\mathbf{p}), \tag{4.17c}$$

$$\frac{dR(t,\mathbf{p})}{dt} = k_2 ES_1(t,\mathbf{p}), \tag{4.17d}$$

$$\frac{dE^*(t,\mathbf{p})}{dt} = k_2 ES_1(t,\mathbf{p}) - k_3 E^*(t,\mathbf{p})S_2(t,\mathbf{p}) + r_3 E^* S_2(t,\mathbf{p}), \tag{4.17e}$$

$$\frac{dS_2(t,\mathbf{p})}{dt} = -k_3 E^*(t,\mathbf{p})S_2(t,\mathbf{p}) + r_3 E^* S_2(t,\mathbf{p}), \tag{4.17f}$$

$$\frac{dE^*S_2(t,\mathbf{p})}{dt} = k_3 E^*(t,\mathbf{p})S_2(t,\mathbf{p}) - (r_3 + k_4)E^* S_2(t,\mathbf{p}), \tag{4.17g}$$

$$\frac{dQ(t,\mathbf{p})}{dt} = k_4 E^* S_2(t,\mathbf{p}), \tag{4.17h}$$

$$\frac{dE^\bullet(t,\mathbf{p})}{dt} = k_4 E^* S_2(t,\mathbf{p}) - k_5 E^\bullet(t,\mathbf{p})S_3(t,\mathbf{p}) + r_5 E^\bullet S_3(t,\mathbf{p}), \tag{4.17i}$$

$$\frac{dS_3(t,\mathbf{p})}{dt} = -k_5 E^\bullet(t,\mathbf{p})S_3(t,\mathbf{p}) + r_5 E^\bullet S_3(t,\mathbf{p}), \tag{4.17j}$$

$$\frac{dE^\bullet S_3(t,\mathbf{p})}{dt} = k_5 E^\bullet(t,\mathbf{p})S_3(t,\mathbf{p}) - (r_5 + k_6)E^\bullet S_3(t,\mathbf{p}), \tag{4.17k}$$

$$\frac{dP(t,\mathbf{p})}{dt} = k_6 E^\bullet S_2(t,\mathbf{p}). \tag{4.17l}$$

Concentrations and rate constants are denoted as previously described. For a typical experiment initial concentrations of enzyme and substrates would be known, nonzero, values denoted E_0, S_{10}, S_{20} and S_{30} respectively, concentrations of the remaining species would be zero.

Following the reasoning outlined previously assume that the three complexes and the two alternative enzyme forms rapidly reach equilibrium concentrations. These quasi-steady state assumptions and a conservation relation on the concentration of E are stated below:

$$\frac{dES_1(t,\mathbf{p})}{dt} = \frac{dE^*(t,\mathbf{p})}{dt} = \frac{dE^*S_2(t,\mathbf{p})}{dt} = \frac{dE^\bullet(t,\mathbf{p})}{dt} = \frac{dE^\bullet S_2(t,\mathbf{p})}{dt} = 0, \tag{4.18}$$

$$E(t,\mathbf{p}) = E_0 - ES_1(t,\mathbf{p}) - E^*(t,\mathbf{p}) - E^*S_2(t,\mathbf{p}) - E^\bullet(t,\mathbf{p}) - E^\bullet S_2(t,\mathbf{p}), \tag{4.19}$$

With these assumptions Equations (4.17c), (4.17e), (4.17g), (4.17i) and (4.17k) can be solved for $ES_1(t,\mathbf{p})$, $E^*(t,\mathbf{p})$, $E^*S_2(t,\mathbf{p})$, $E^\bullet(t,\mathbf{p})$ and $E^\bullet S_2(t,\mathbf{p})$. The results can then be combined with the remaining equations to give the quasi-steady state form of the model below:

$$\frac{dS_1(t,\mathbf{p})}{dt} = \frac{dS_2(t,\mathbf{p})}{dt} = \frac{dS_3(t,\mathbf{p})}{dt} = -f_3(S_1(t,\mathbf{p}), S_2(t,\mathbf{p}), S_3(t,\mathbf{p}), \mathbf{p}), \tag{4.20a}$$

$$\frac{dR(t,\mathbf{p})}{dt} = \frac{dQ(t,\mathbf{p})}{dt} = \frac{dP(t,\mathbf{p})}{dt} = f_3(S_1(t,\mathbf{p}), S_2(t,\mathbf{p}), S_3(t,\mathbf{p}), \mathbf{p}), \tag{4.20b}$$

$$f_3(S_1, S_2, S_3, \mathbf{p}) = \frac{V_{max} S_1 S_2 S_3}{S_1 S_2 S_3 + k_{m,1} S_2 S_3 + k_{m,2} S_1 S_3 + k_{m,3} S_1 S_2}. \tag{4.20c}$$

Parameter notation is chosen to correspond with that used by Dixon [144]. The parameters used above can be expressed in terms of the transient state constants as follows:

$$\begin{aligned} V_{max} &= \frac{E_0 k_2 k_4 k_6}{k_2 k_4 + k_4 k_6 + k_2 k_6}, \qquad k_{m,1} = \frac{k_4 k_6}{k_1} \frac{r_1 + k_2}{k_2 k_4 + k_4 k_6 + k_2 k_6}, \\ k_{m,2} &= \frac{k_2 k_6}{k_3} \frac{r_3 + k_4}{k_2 k_4 + k_4 k_6 + k_2 k_6} \quad k_{m,3} = \frac{k_2 k_4}{k_5} \frac{r_5 + k_6}{k_2 k_4 + k_4 k_6 + k_2 k_6}. \end{aligned} \tag{4.21}$$

The number of rate constants describing the model is reduced from nine to four. These parameters are a generalisation of those derived for the two substrate ping-pong mechanism as is the quasi-steady state relation obtained. As for the two substrate case the model can be expressed by a new mechanism (Figure 4.2).

The mechanism of the amino-acid ligases MurC, MurD, MurE and MurF is known to involve the ordered formation of a quaternary complex. This mechanism will be referred to as the simple ordered model and is given by the following schematic:

$$\left.\begin{aligned} E + S_1 &\underset{r_1}{\overset{k_1}{\rightleftharpoons}} ES_1 \\ ES_1 + S_2 &\underset{r_2}{\overset{k_2}{\rightleftharpoons}} ES_1S_2 \\ ES_1S_2 + S_3 &\underset{r_3}{\overset{k_3}{\rightleftharpoons}} ES_1S_2S_3 \\ ES_1S_2S_3 &\underset{k_4}{\longrightarrow} E + P \end{aligned}\right\} \tag{4.22}$$

Diagram of the three substrate reaction mechanism in which a ternary complex is formed. The substrates S_1, S_2 and S_3 reversibly bind to the enzyme, E, in sequence forming complexes, ES_1, ES_1S_2 and $ES_1S_2S_3$. The quaternary complex, $ES_1S_2S_3$, then breaks down irreversibly releasing enzyme and product, P. All products are again assumed to be released simultaneously.

The differential equations describing the transient kinetics of this system are given below:

$$\frac{dE(t,\mathbf{p})}{dt} = -k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) + r_1ES_1(t,\mathbf{p}) + k_4ES_1S_2S_2(t,\mathbf{p}), \tag{4.23a}$$

$$\frac{dS_1(t,\mathbf{p})}{dt} = -k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) + r_1ES_1(t,\mathbf{p}), \tag{4.23b}$$

$$\frac{dES_1(t,\mathbf{p})}{dt} = k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) - r_1ES_1(t,\mathbf{p}) - k_2ES_1(t,\mathbf{p})S_2(t,\mathbf{p}) + \\ + r_2ES_1S_2(t,\mathbf{p}), \tag{4.23c}$$

$$\frac{dS_2(t,\mathbf{p})}{dt} = -k_2ES_1(t,\mathbf{p})S_2(t,\mathbf{p}) + r_2ES_1S_2(t,\mathbf{p}), \tag{4.23d}$$

$$\frac{dES_1S_2(t,\mathbf{p})}{dt} = k_2ES_1(t,\mathbf{p})S_2(t,\mathbf{p}) - r_2ES_1S_2(t,\mathbf{p}) - \\ - k_3ES_1S_2(t,\mathbf{p})S_3(t,\mathbf{p}) + r_3ES_1S_2S_3(t,\mathbf{p}), \tag{4.23e}$$

$$\frac{dS_3(t,\mathbf{p})}{dt} = -k_3ES_1S_2(t,\mathbf{p})S_3(t,\mathbf{p}) + r_3ES_1S_2S_3(t,\mathbf{p}), \tag{4.23f}$$

$$\frac{dES_1S_2S_3(t,\mathbf{p})}{dt} = k_3ES_1S_2(t,\mathbf{p})S_3(t,\mathbf{p}) - (r_3+k_4)ES_1S_2S_3(t,\mathbf{p}), \tag{4.23g}$$

$$\frac{dP(t,\mathbf{p})}{dt} = k_4ES_1S_2(t,\mathbf{p}). \tag{4.23h}$$

Concentrations and rate constants are denoted as previously described. Typical experimental initial conditions would be as described for the ping-pong model above.

Proceeding as before to produce a quasi-steady state model the following assumptions and conservation rules are used:

$$\frac{dES_1(t,\mathbf{p})}{dt} = \frac{dES_1S_2(t,\mathbf{p})}{dt} = \frac{dES_1S_2S_3(t,\mathbf{p})}{dt} = 0, \tag{4.24}$$

$$E(t,\mathbf{p}) = E_0 - ES_1(t,\mathbf{p}) - ES_1S_2(t,\mathbf{p}) - ES_1S_2S_3(t,\mathbf{p}). \tag{4.25}$$

With these assumptions Equations (4.23c), (4.23e) and (4.23g) can be solved for $ES_1(t,\mathbf{p})$, $ES_1S_2(t,\mathbf{p})$ and $ES_1S_2S_3(t,\mathbf{p})$. The solutions can then be combined with the remaining dif-

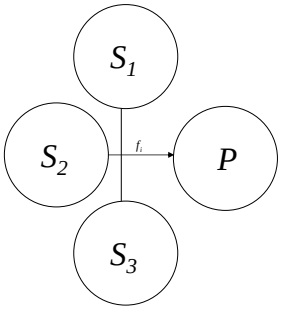


Figure 4.2: Diagram of the general three substrate quasi-steady state reaction mechanism. The rate of conversion of substrates to products is given by a function, f_i, derived as above. For the ping-pong model $i = 3$ while for the simple ordered model $i = 4$. All products released are represented by P, its associated state variable would be a vector of the concentrations of the various products released. If all initial product concentrations are equal all subsequent product concentrations will be equal.

ferential equations to obtain the following quasi-steady state model:

$$\frac{dS_1(t,\mathbf{p})}{dt} = \frac{dS_2(t,\mathbf{p})}{dt} = \frac{dS_3(t,\mathbf{p})}{dt} = -f_4(S_1(t,\mathbf{p}), S_2(t,\mathbf{p}), S_3(t,\mathbf{p}), \mathbf{p}), \tag{4.26a}$$

$$\frac{dP(t,\mathbf{p})}{dt} = f_4(S_1(t,\mathbf{p}), S_2(t,\mathbf{p}), S_3(t,\mathbf{p}), \mathbf{p}), \tag{4.26b}$$

$$f_4(S_1, S_2, S_3, \mathbf{p}) = \\ = \frac{V_{max}}{1 + \frac{k_{m,1}}{S_1} + \frac{k_{m,2}}{S_2}\left(1 + \frac{k_{s,12}}{S_1}\right) + \frac{k_{m,3}}{S_3}\left(1 + \frac{k_{s,23}}{S_2}\left(1 + \frac{k_{s,12}}{S_1}\right)\right)}. \tag{4.26c}$$

Relations between the parameters used in this system and those of the transient system are shown below:

$$\begin{aligned} V_{max} &= -k_4 E_0, \quad k_{m,1} = \frac{k_4}{k_1}, \quad k_{m,2} = \frac{k_4}{k_2}, \\ k_{m,3} &= \frac{r_3 + k_4}{k_3}, \quad k_{s,12} = \frac{r_1}{k_1}, \quad k_{s,23} = \frac{r_2}{k_2}. \end{aligned} \tag{4.27}$$

The number of rate constants required is reduced from seven to six. The second step of the reaction is coupled to the first as observed for the two substrate simple ordered mechanism, Equation (4.15). The third step is coupled to both the first and second steps with $k_{s,12}$ acting on $k_{s,23}$ in the same way as it acts on $k_{m,2}$, thus if S_1 and S_2 are equal to their respective coupling constants $k_{m,3}$ would be increased by a factor of four. The use of quasi-steady state

approximations reduces the system to the mechanism shown in Figure 4.2.

4.3 Structural identifiability analysis

A structural identifiability analysis is now undertaken for each of the models derived in order to determine for which models, and hence enzymes, it will be possible to estimate parameters directly from time course data. For each model Pohjanpalo's Jacobian rank test (Section 3.2.2) was used to determine if the model could be shown to be at least locally identifiable. Full analyses were then undertaken using the Taylor series and (input-)output relationship approaches (Sections 3.2.1 and 3.2.3). A summary of the results of this analysis are presented in Table 4.1. Since the kinetic parameters of the ping-pong models are structurally identifiable it would be appropriate to use direct parameter estimation where reactions follow this mechanism. Unfortunately the majority of reactions in the pathway follow a simple ordered mechanism for which only V_{max} proved structurally identifiable. As such direct parameter estimation will not be useful for the majority of the analysis to be undertaken.

The use of both Taylor series and (input-)output relationship approaches allows the relative effectiveness of the two approaches in analysing these types of models to be assessed. Where both approaches produce a result the results agree providing validation for the less developed (input-)output relationship approach. Implementation of the Taylor series approach proved too computationally intensive for the majority of the models considered. By way of contrast analysis using the (input-)output relationship approach was generally successful, failing only for the most complex models.

The model equations for these reaction mechanisms were derived in Sections 4.2.1 and 4.2.2, Equations (4.5), (4.12), (4.20) and (4.26). For a given mechanism the Taylor series expansions of any state variable are equivalent up to a change of sign for powers of $t \geq 1$ since the derivatives of the state variables are equivalent up to a change of sign. Thus, under the Taylor series approach,

Model	Parameters	Taylor series		Input-output relationship		
		KSC	USC	KSC	USC	UIC
2 PP	V_{max}	global	CI	global	global	global
	$k_{m,i}$s	global	CI	global	local	local
	ICs	-	CI	-	local	unident
2 SO	V_{max}	global	CI	global	global	-
	$k_{m,i}$s	unident	CI	unident	unident	-
	ICs	-	CI	-	local	-
3 PP	V_{max}	CI	-	global	global	global
	$k_{m,i}$s	CI	-	global	local	CI
	ICs	-	-	-	local	CI
3 SO	V_{max}	CI	-	global	global	-
	$k_{m,i}$s	CI	-	unident	unident	-
	ICs	-	-	-	CI	-

Table 4.1: Summary of the structural identifiability analysis results for two and three substrate quasi-steady state models. Models are denoted by a number indicating the number of substrates and an abbreviation, PP indicating ping-pong and SO indicating simple ordered. Three cases were considered for each model; known substrate concentrations (KSC), unknown substrate concentrations (USC) and unknown initial conditions (UIC). Results are either: global, structurally globally identifiable; local, structurally locally identifiable; unident, structurally unidentifiable; CI, analysis was computationally intractable; or -, not analysed.

the only difference between non-trivial measurement schemes is the information obtained from the first Taylor series coefficient. Given an expansion about $t = 0$ the first coefficient, for y^0, will be determined by the initial conditions of the experiment and as such provides no useful information if the initial conditions are known. As such without loss of generality the analyses undertaken assume that product concentration is measured.

Note finally that in all cases there are more transient parameters than quasi-steady state parameters. As such it is impossible to solve these parameters to obtain unique solutions for all the transient parameters. Thus the transient model is unidentifiable under the quasi-steady state assumptions, although a subset of its parameters may be identifiable.

4.3.1 Two substrate ping-pong mechanism

A Taylor series analysis of the two substrate ping-pong mechanism is presented first. The mechanism was stated in the form introduced in Section 3.2:

$$\left.\begin{aligned} &\mathbf{x}(t,\mathbf{p}) = (S_1(t,\mathbf{p}), S_2(t,\mathbf{p}), P(t,\mathbf{p}))^T, \\ &\frac{d\mathbf{x}(t,\mathbf{p})}{dt} = f_1(S_1(t,\mathbf{p}), S_2(t,\mathbf{p}), \mathbf{p})(-1,-1,1)^T, \\ &\mathbf{x}(0,\mathbf{p}) = \mathbf{x}_0 = (S_{10}, S_{20}, 0)^T, \\ &\mathbf{y}(t,\mathbf{p}) = (0,0,1)\cdot\mathbf{x}(t,\mathbf{p}) = P(t,\mathbf{p}), \end{aligned}\right\} \tag{4.28}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the state vector and observation vectors respectively. The vector of unknown parameters is denoted, $\mathbf{p}$, and is given by:

$$\mathbf{p} = (V_{max}, k_{m,1}, k_{m,2}). \tag{4.29}$$

The first four Taylor series coefficients were then calculated by differentiating $\mathbf{y}(t,\mathbf{p})$ with respect to t and evaluating at $t = 0$.

$$\mathbf{y}(0,\mathbf{p}) = 0 \tag{4.30}$$

$$\mathbf{y}^{(1)}(0,\mathbf{p}) = \frac{V_{max}S_{10}S_{20}}{S_{10}S_{20} + k_{m,1}S_{20} + k_{m,2}S_{10}} \tag{4.31}$$

$$\mathbf{y}^{(2)}(0,\mathbf{p}) = \frac{-V_{max}^2 S_{20}S_{10}(k_{m,2}S_{10}^2 + k_{m,1}S_{20}^2)}{(S_{20}S_{10} + k_{m,2}S_{10} + k_{m,1}S_{20})^3} \tag{4.32}$$

$$\begin{aligned} \mathbf{y}^{(3)}(0,\mathbf{p})(S_{20}S_{10} + k_{m,2}S_{10} + k_{m,1}S_{20})^5 &= V_{max}^3 S_{20}S_{10}(k_{m,2}^2 S_{10}^4 + 6k_{m,2}S_{10}^2 k_{m,1}S_{20}^2 \\ &- 2k_{m,2}S_{10}^4 S_{20} - 2k_{m,2}S_{10}^3 k_{m,1}S_{20} + k_{m,1}^2 S_{20}^4 - 2k_{m,1}S_{20}^4 S_{10} - 2k_{m,1}S_{20}^3 k_{m,2}S_{10}) \end{aligned} \tag{4.33}$$

The Jacobian matrix of these coefficients with respect to the elements of $\mathbf{p}$ has rank three. Therefore the mechanism is at least locally identifiable by Pohjanpalo's Rank Test (PRT) (Section 3.2.2). Following the approach described in Section 3.2.1 an alternative parameter vector was defined as follows:

$$\overline{\mathbf{p}} = (\overline{V_{max}}, \overline{k_{m,1}}, \overline{k_{m,2}}). \tag{4.34}$$

The equations $\mathbf{y}^{(i)}(0, \mathbf{p}) - \mathbf{y}^{(i)}(0, \overline{\mathbf{p}}) = 0$ for $1 \leq i \leq 3$ were then solved for the elements of the alternative parameter vector yielding the following solution:

$$\overline{\mathbf{p}} = \mathbf{p}. \tag{4.35}$$

As such the unknown parameters are globally identifiable when initial conditions are known.

The input-output relationship approach is now applied to the same problem. The following functions generate a differential ideal equivalent to the system defined above:

$$S_1(t, \mathbf{p}) - S_{10} + P(t, \mathbf{p}) = 0, \tag{4.36a}$$

$$S_2(t, \mathbf{p}) - S_{20} + P(t, \mathbf{p}) = 0, \tag{4.36b}$$

$$\frac{dP(t, \mathbf{p})}{dt}(S_1(t, \mathbf{p})S_2(t, \mathbf{p}) + k_{m,1}S_2(t, \mathbf{p}) + k_{m,2}S_1(t, \mathbf{p})) - V_{max}S_1(t, \mathbf{p})S_2(t, \mathbf{p}) = 0, \tag{4.36c}$$

$$y(t, \mathbf{p}) - P(t, \mathbf{p}) = 0. \tag{4.36d}$$

The first two functions define conservation relations inherent in the model, the third is a rearrangement of $P^{(1)}(t, \mathbf{p}) = f_1(S_1(t, \mathbf{p}), S_2(t, \mathbf{p}), \mathbf{p})$ as a differential polynomial, and the fourth restates that P is the measured variable. The elimination ordering, $y(t) < P(t) < S_1(t) < S_2(t)$, was chosen to obtain an output relation in terms of $y(t)$:

$$y^{(1)}(t) = \frac{V_{max}(y(t)^2 - y(t)(S_{10} + S_{20}) + S_{20}S_{10})}{(y(t)^2 - y(t)(S_{10} + S_{20} + k_{m,1} + k_{m,2}) + S_{20}S_{10} + k_{m,2}S_{10} + k_{m,1}S_{20})}. \tag{4.37}$$

Note that an alternative choice of $y(t)$, i.e. $y(t) - S_i(t, \mathbf{p})$, will result in a different output relation. The transformations used to create these output relations are invertible, as such they define an isomorphism between the possible output relations. As such only one output relation must be analysed in order to determine the identifiability of measuring any single species.

The alternative parameter vector, $\overline{\mathbf{p}}$, was used as described in Section 3.2.3 to eliminate the leader of Equation (4.37), $y^{(1)}(t)$:

$$y^{(1)}(t, \mathbf{p}) = y^{(1)}(t, \overline{\mathbf{p}}) \tag{4.38}$$

$$\frac{V_{max}(y(t)^2 - y(t)(S_{10} + S_{20}) + S_{10}S_{20})}{(y(t)^2 - y(t)(S_{10} + S_{20} + k_{m,1} + k_{m,2}) + S_{10}S_{20} + k_{m,2}S_{10} + k_{m,1}S_{20})} = \frac{\overline{V_{max}}(y(t)^2 - y(t)(S_{10} + S_{20}) + S_{10}S_{20})}{(y(t)^2 - y(t)(S_{10} + S_{20} + \overline{k_{m,1}} + \overline{k_{m,2}}) + S_{10}S_{20} + \overline{k_{m,2}}S_{10} + \overline{k_{m,1}}S_{20})} \tag{4.39}$$

Let: $C = y(t)^2 - y(t)(S_{10} + S_{20}) + S_{10}S_{20}$

$$0 = C\Big(V_{max}(y(t)^2 - y(t)(S_{10} + S_{20} + \overline{k_{m,1}} + \overline{k_{m,2}}) + S_{10}S_{20} + \overline{k_{m,2}}S_{10} + \overline{k_{m,1}}S_{20}) - \\ - \overline{V_{max}}(y(t)^2 - y(t)(S_{10} + S_{20} + k_{m,1} + k_{m,2}) + S_{10}S_{20} + k_{m,2}S_{10} + k_{m,1}S_{20})\Big) \tag{4.40}$$

The result is a degree four polynomial in $y(t)$, p, $\overline{p}$ and the initial conditions. As such the monomials are linearly independent by the Fundamental Theorem of Algebra (FTA). The five nonzero coefficients of the polynomial were solved showing that:

$$\mathbf{p} = \overline{\mathbf{p}}, \tag{4.41}$$

as expected.

So far it has been assumed that the initial concentrations of S_1 and S_2 are known. However the quasi-steady state approximation applies only once constant, non-zero levels of intermediate complexes have been established, consuming some of each substrate. As such it may be more appropriate to model the initial concentrations of the substrates as unknown parameters.

S_{10}, S_{20} are now treated as elements of the unknown parameter vector, $\mathbf{p}$. This can be done using the Taylor series approach. However the additional coefficients required to determine the new unknown parameters proved computationally intractable. The rank of the Jacobian matrix of the computed coefficients was five, equal to the number of unknown parameters, thus by PRT the model is at least locally identifiable.

The addition of S_{10} and S_{20} to the unknown parameter vector does not significantly increase the complexity of the analysis using the input-output relationship approach. In order to account for known nonzero initial P, $P(0)$ was subtracted from the first two generators of the differential ideal, Equations (4.37). This had no effect on the subsequent analysis. The parameter and alternative parameter vectors were extended to:

$$\mathbf{p} = (V_{max}, k_{m,1}, k_{m,2}, S_{10}, S_{20}) \tag{4.42}$$

$$\overline{\mathbf{p}} = (\overline{V_{max}}, \overline{k_{m,1}}, \overline{k_{m,2}}, \overline{S_{10}}, \overline{S_{20}}) \tag{4.43}$$

respectively and the analysis described above was repeated. This model was shown to be locally identifiable with the following solutions:

$$\left\{\overline{V_{max}} = V_{max},\ \overline{k_{m,1}} = k_{m,1},\ \overline{k_{m,2}} = k_{m,2},\ \overline{S_{10}} = S_{10},\ \overline{S_{20}} = S_{20}\right\}, \tag{4.44a}$$

$$\left\{\overline{V_{max}} = V_{max},\ \overline{k_{m,1}} = k_{m,2},\ \overline{k_{m,2}} = k_{m,1},\ \overline{S_{10}} = S_{20},\ \overline{S_{20}} = S_{10}\right\}. \tag{4.44b}$$

Note that the second solution, where $\overline{\mathbf{p}} \neq \mathbf{p}$, swaps the $k_{m,i}$ values and the S_{i0} values. Thus the second solution above can be eliminated practically by choosing one initial substrate concentration to be significantly higher than the other; any solution in which an estimated initial substrate concentration is higher than the known concentration of substrate in the reaction mixture can be rejected.

If it is assumed that an unknown concentration of P is reached prior to reaching the steady

state it is necessary to modify the differential ideal, (4.36), as follows:

$$y(t) - P(t) - P(0) = 0, \tag{4.45}$$

and to add $P(0)$ to the unknown parameter vector. Analysis using the input-output relationship approach shows that V_{max} remains globally identifiable and $k_{m,1}$ and $k_{m,2}$ remain locally identifiable as in Equations (4.44). The initial concentrations of substrates and product are unidentifiable. As such the model is unidentifiable. However since the complexes formed require enzyme, their total concentration cannot exceed that of the initial free enzyme concentration, E_0. Thus the possible range in which the observed initial conditions can lie is determined by the following inequality:

$$S_{n0,act} - E_0 - P(0) < S_{n0,obs} < S_{n0,act} - P(0), \tag{4.46}$$

$S_{n0,obs}$ being the unknown concentration of substrate remaining after complex formation while $S_{n0,act}$ is the known concentration of substrate initially used. It can be assumed that $P(0)$ and $E(0)$ are small. Using these constraints it is possible to practically eliminate solutions which predict $S_{n0,obs}$ values which are significantly different from the $S_{n0,act}$ values. As such if the $S_{n0,act}$ values are significantly different, as above, it would be possible to determine which of the local solutions for the $k_{m,i}$s is appropriate.

4.3.2 Two substrate simple ordered mechanism

The two substrate simple ordered mechanism is now considered. The state space model for this mechanism is identical to that for the ping-pong mechanism, see Equations (4.28), with the following change to the derivative of the state vector:

$$\frac{d\mathbf{x}(t,\mathbf{p})}{dt} = f_2(S_1(t,\mathbf{p}), S_2(t,\mathbf{p}), \mathbf{p})(-1,-1,1)^T, \tag{4.47}$$

and the following addition to the unknown parameter vector:

$$\mathbf{p} = (k_3, k_{m,1}, k_{m,2}, k_{s,12}). \tag{4.48}$$

A Taylor series expansion of the observation $\mathbf{y}$ at $t = 0$ was undertaken. The first four Taylor series coefficients of the expansion are presented below:

$$\mathbf{y}(0, \mathbf{p}) = 0, \tag{4.49}$$

$$\mathbf{y}^{(1)}(0, \mathbf{p}) = \frac{k_3 E_0 S_{10} S_{20}}{(S_{10} S_{20} + k_{m,1} S_{20} + k_{m,2} S_{10} + k_{s,12} k_{m,2})}, \tag{4.50}$$

$$\mathbf{y}^{(2)}(0, \mathbf{p}) = \frac{-k_3^2 E_0^2 S_{10} S_{20} (k_{m,1} S_{20}^2 + k_{m,2} S_{10}^2 + k_{s,12} k_{m,2} S_{20} + k_{s,12} k_{m,2} S_{10})}{(S_{10} S_{20} + k_{m,1} S_{20} + k_{m,2} S_{10} + k_{s,12} k_{m,2})^3}, \tag{4.51}$$

$$\begin{aligned}
&\mathbf{y}^{(3)}(0, \mathbf{p})(S_{10} S_{20} + k_{m,1} S_{20} + k_{m,2} S_{10} + k_{s,12} k_{m,2})^5 = \\
&- k_3^3 E_0^3 S_{20} S_{10} (2 k_{s,12} k_{m,2} S_{10}^2 k_{m,1} S_{20} - 6 k_{m,1} S_{20}^2 k_{m,2} S_{10}^2 - 2 k_{m,1} S_{20}^3 k_{s,12} k_{m,2} \\
&+ 2 k_{m,1} S_{20}^3 k_{m,2} S_{10} - 4 k_{m,2}^2 S_{10}^2 k_{s,12} S_{20} + 2 k_{m,2} S_{10}^3 k_{m,1} S_{20} + 2 k_{m,1} S_{20}^4 S_{10} \\
&- 2 k_{m,2}^2 S_{10}^3 k_{s,12} + 2 k_{m,2} S_{10}^4 S_{20} - k_{s,12}^2 k_{m,2}^2 S_{20}^2 - k_{s,12}^2 k_{m,2}^2 S_{10}^2 - k_{m,1}^2 S_{20}^4 \\
&- 4 k_{m,1} S_{20}^2 k_{s,12} k_{m,2} S_{10} - 4 k_{s,12}^2 k_{m,2}^2 S_{20} S_{10} + 2 k_{s,12} k_{m,2} S_{20}^3 S_{10} \\
&+ 2 k_{s,12} k_{m,2}^2 S_{20}^2 S_{10} + 2 k_{s,12} k_{m,2} S_{10}^2 S_{20}^2 + 2 k_{s,12} k_{m,2} S_{10}^3 S_{20} - k_{m,2}^2 S_{10}^4),
\end{aligned} \tag{4.52}$$

The fifth coefficient was also calculated but is not presented for brevity. The Jacobian matrix of these coefficients was calculated and found to have a rank of three. This is insufficient to establish local identifiability by PRT, however it does not conclusively establish unidentifiability since it is possible that inclusion of higher order Taylor series coefficients would increase the rank of the infinite Jacobian matrix.

An alternative parameter vector $\overline{\mathbf{p}} = (\overline{k_3}, \overline{k_{m,1}}, \overline{k_{m,2}}, \overline{k_{s,12}})$ was used as described previously. The following relations between $\mathbf{p}$ and $\overline{\mathbf{p}}$ were calculated from the Taylor series coefficients,

Equations (4.49)-(4.52):

$$\left.\begin{aligned} \overline{k_3} &= k_3 \\ \overline{k_{m,1}} + \overline{k_{m,2}} &= k_{m,1} + k_{m,2} \\ \overline{k_{m,2}}(S_{20} - S_{10} - \overline{k_{s,12}}) &= k_{m,2}(S_{20} - S_{10} - k_{s,12}) \end{aligned}\right\} \tag{4.53}$$

This shows that k_3 is globally identifiable and that more coefficients are required to determine the identifiability of $k_{m,1}$, $k_{m,2}$ and $k_{s,12}$. However it will now be shown that no further information can be acquired from additional Taylor series coefficients. The proof is inductive and proceeds as follows.

First determine expressions for $k_{m,1}$ and $k_{s,12}$ in terms of k_3 and $k_{m,2}$ from the second and third Taylor series coefficients, Equations (4.50) and (4.51), yielding:

$$k_{m,1} = \frac{y^{(2)}(0)k_3E_0S_{10}S_{20} - (y^{(1)}(0))^3(S_{10} + S_{20} + k_{m,2}) + (y^{(1)}(0))^2k_3E_0(S_{10} + S_{20})}{(y^{(1)}(0))^3}, \tag{4.54}$$

$$k_{s,12} = -\frac{(S_{10} - S_{20})(y^{(1)}(0))^3k_{m,2} + E_0S_{20}^2(S_{10}y^{(2)}(0) + y^{(1)}(0))^2)k_3 - S_{20}^2(y^{(1)}(0))^3}{(y^{(1)}(0))^3k_{m,2}}. \tag{4.55}$$

Substitution of these solutions into the fourth Taylor series coefficient yields an expression in k_3 but not $k_{m,2}$.

Proposition 4.1: *The n-th derivative of $y(t, \mathbf{p})$ for $n > 1$ has the following form:*

$$y^{(n)} = \frac{y^{(n)}D - 2k_3E_0y^{(n-1)}y - k_3E_0y^{(n-1)}(S_1 + S_2) + k_3E_0(y^2)^{(n-1)} - (y^{(1)}D)^{(n-1)}}{D}, \tag{4.56}$$

where D denotes the denominator of $f_2(S_1(t, \mathbf{p}), S_2(t, \mathbf{p}), \mathbf{p})$, i.e.

$$S_1(t, \mathbf{p})S_2(t, \mathbf{p}) + k_{m,1}S_2(t, \mathbf{p}) + k_{m,2}S_1(t, \mathbf{p}) + k_{m,2}k_{s,12}.$$

(Note that $y^{(n)}$ is expressed solely in terms of lower order derivatives of y and D since the first two coefficients are included for convenience of notation, they eliminate terms of the fifth and fourth coefficients respectively.)

Proof: Determine $y^{(2)}$ from $f_2(S_1(t,\mathbf{p}), S_2(t,\mathbf{p}), \mathbf{p})$:

$$\begin{aligned} y^{(2)} &= \frac{d}{dt}y^{(1)} \\ &= \frac{D(k_3E_0S_1^{(1)}S_2 + k_3E_0S_1S_2^{(1)}) - D^{(1)}k_3E_0S_1S_2}{D^2} \\ &= \frac{-k_3E_0y^{(1)}(S_1+S_2) - D^{(1)}y^{(1)}}{D} \end{aligned}$$

and Equation (4.56):

$$\begin{aligned} y^{(2)} &= \frac{y^{(2)}D - 2k_3E_0y^{(1)}y - k_3E_0y^{(1)}(S_1+S_2) + k_3E_0(y^2)^{(1)} - (y^{(1)})^{(1)}}{D} \\ &= \frac{y^{(2)}D - 2k_3E_0y^{(1)}y - k_3E_0y^{(1)}(S_1+S_2) + 2k_3E_0y^{(1)}y - y^{(2)}D - y^{(1)}D^{(1)}}{D} \\ &= \frac{-k_3E_0y^{(1)}(S_1+S_2) - y^{(1)}D^{(1)}}{D} \end{aligned}$$

proving the proposition holds for $n = 2$.

Assuming it holds for n:

$$\begin{aligned} y^{(n+1)} &= \frac{d}{dt}y^{(n)} \\ &= \frac{\begin{array}{l} D(y^{(n)}D - 2k_3E_0y^{(n-1)}y - k_3E_0y^{(n-1)}(S_1+S_2) + k_3E_0(y^2)^{(n-1)} - (y^{(1)}D)^{(n-1)})^{(1)} \\ \qquad - D^{(1)}Dy^{(n)} \end{array}}{D^2} \\ &= \frac{\begin{array}{l} -2k_3E_0(y^{(n)}y + y^{(n-1)}y^{(1)}) - k_3E_0(y^{(n)}(S_1+S_2) - y^{(n-1)}2y^{(1)}) \\ \qquad + k_3E_0(y^2)^{(n)} - (y^{(1)}D)^{(n)} - D^{(1)}y^{(n)} + y^{(n+1)}D + y^{(n)}D^{(1)} \end{array}}{D} \end{aligned}$$

$$= \frac{y^{(n+1)}D - 2k_3E_0y^{(n)}y - k_3E_0y^{(n)}(S_1 + S_2) + k_3E_0(y^2)^{(n)} - (y^{(1)}D)^{(n)}}{D}$$

as required for $n+1$. Thus the proposition holds for all $n > 1$ by induction. □

Since Equation (4.56) expresses $y^{(n)}$ in terms of lower order derivatives, by induction $y^{(n)}(0)$ contains no $k_{m,2}$ terms if $D(0)$ and $D^{(i)}(0)$ contain no $k_{m,2}$ terms for $i < n$. Substituting the solutions above for $k_{m,1}$ and $k_{s,12}$ into $D(S_1(t,\mathbf{p}), S_2(t,\mathbf{p}), \mathbf{p})$ results in the following equations:

$$D(S_{10}, S_{20}, \mathbf{p}) = \frac{E_0S_{10}S_{20}k_3}{y^{(1)}(0)} \tag{4.57}$$

$$\begin{aligned} D^{(1)}(S_1(t,\mathbf{p}), S_2(t,\mathbf{p}), \mathbf{p}) &= -y^{(1)}(t)(S_1(t) + S_2(t)) - \\ &- y^{(1)}(t)\frac{E_0S_{10}S_{20}y^{(2)}(0)k_3 - (S_{10} + S_{20})((y^{(1)}(0))^3 - (y^{(1)}(0))^2E_0k_3)}{(y^{(1)}(0))^3}. \end{aligned} \tag{4.58}$$

From Equation (4.57) $D(0)$ contains no $k_{m,2}$ terms. It is clear that $D^{(i)}(0)$, $i \geq 1$, can be expressed in terms of derivatives of y of order less than or equal to i but greater than 0 since the derivatives of S_1 and S_2 are equal to $-y^{(1)}$. Moreover the only $k_{m,2}$ terms that could arise would occur in these derivatives. Since the only derivatives of D in Equation (4.56) are of order less than n by induction no $k_{m,2}$ terms arise in $y^{(n)}(0)$.

Thus further Taylor series coefficients will not determine $k_{m,2}$ and $k_{m,1}$, $k_{m,2}$ and $k_{s,12}$. As such the model is unidentifiable for a quasi-steady state measurement. The transient and quasi-steady state parameter k_3 is however globally identifiable.

Using the input-output relationship approach, the generators of the differential ideal are as

follows:

$$S_1 - S_{10} + P = 0, \tag{4.59a}$$

$$S_2 - S_{20} + P = 0, \tag{4.59b}$$

$$P^{(1)}(S_1S_2 + k_{m,1}S_2 + k_{m,2}S_1 + k_{s,12}k_{m,2}) - V_{max}S_1S_2 = 0, \tag{4.59c}$$

$$y - P = 0, \tag{4.59d}$$

with $V_{max} = E_0k_3$, from Equations (4.13). These equations are similar to those obtained for the ping-pong model, Equations (4.36), having a single additional term in the third equation, $k_{s,12}y^{(1)}(t)$. Explicit statements of dependence on t and $\mathbf{p}$ are omitted for brevity. The output relation obtained under the elimination ordering $y < P < S_1 < S_2$ is:

$$y^{(1)} = \frac{V_{max}(y^2 - y(S_{10} + S_{20}) + S_{10}S_{20})}{y^2 - y(S_{10}y + S_{20} + k_{m,1} + k_{m,2}) + S_{10}S_{20} + k_{m,2}S_{10} + k_{m,1}S_{20} + k_{m,2}k_{s,12}}. \tag{4.60}$$

An alternative parameter vector $\overline{\mathbf{p}}$ was introduced and the leader eliminated. This produces a degree four polynomial in y, $\mathbf{p}$, $\overline{\mathbf{p}}$ and the initial conditions. The monomials of this expression are linearly independent by the FTA. The five non-zero coefficients were solved obtaining the same relations between $\mathbf{p}$ and $\overline{\mathbf{p}}$ as were found using the Taylor series approach, Equations (4.53). This is sufficient to show that the model is unidentifiable, with k_3 remaining globally identifiable.

Extending the analysis to the case where the initial concentrations of S_1 and S_2 are unknown,

as in Section 4.3.1, the following sets of relations were obtained:

$$\left\{ \begin{array}{c} \overline{k_3} = k_3 \ , \overline{k_{m,1}} + \overline{k_{m,2}} = k_{m,1} + k_{m,2}, \\ \\ \overline{k_{m,2}}(\overline{k_{s,12}} - S_{20} + S_{10}) = k_{m,2}(k_{s,12} - S_{20} + S_{10}), \\ \\ \overline{S_{10}} = S_{10} \ , \overline{S_{20}} = S_{20}, \end{array} \right\} \tag{4.61a}$$

$$\left\{ \begin{array}{c} \overline{k_3} = k_3 \ , \overline{k_{m,1}} + \overline{k_{m,2}} = k_{m,1} + k_{m,2}, \\ \\ \overline{k_{m,2}}(\overline{k_{s,12}} - S_{10} + S_{20}) = k_{m,2}k_{s,12} + k_{m,1}S_{20} - S_{10}k_{m,1}, \\ \\ \overline{S_{10}} = S_{20} \ , \overline{S_{20}} = S_{10}. \end{array} \right\} \tag{4.61b}$$

The model remains unidentifiable. However k_3 remains globally identifiable and the initial conditions are locally identifiable and can be estimated when one substrate concentration is significantly higher than the other.

4.3.3 Deconvolution of parameters obtained by conventional means

Typically when determining quasi-steady state parameters for two substrate enzymes with this mechanism a series of assays are undertaken in which the initial concentration of one substrate is held constant while the other is varied. The initial rates are then plotted against the varied initial concentration and the results fitted to the one substrate Michaelis-Menten equation given by [144]:

$$\frac{dP(t)}{dt} = \frac{V_{max}}{1 + \dfrac{k_m}{S}}. \tag{4.62}$$

Typical plots can be found in Chapter 5 (Figure 5.2). The apparent kinetic constants obtained from this approach will be denoted $V_1^{app}, V_2^{app}, k_{m,1}^{app}, k_{m,2}^{app}$ where numbers indicate which of the two initial concentrations was varied throughout the assay series. By comparing f_2 and (4.62)

the following relations between the actual and the apparent parameters were obtained:

$$\left.\begin{aligned} V_1^{app} &= \frac{k_3 E_0 S_{20}}{S_{20}+k_{m,2}}, \quad k_{m,1}^{app} = \frac{k_{m,1}S_{20}+k_{m,2}k_{s,12}}{S_{20}+k_{m,2}}, \\ V_2^{app} &= \frac{k_3 E_0 S_{10}}{S_{10}+k_{m,1}}, \quad k_{m,2}^{app} = k_{m,2}\frac{S_{10}+k_{s,12}}{S_{10}+k_{m,1}}. \end{aligned}\right\} \tag{4.63}$$

Given that k_3 is globally identifiable it should be possible to estimate its value from a progress curve. It is then possible to deconvolute the relations above to estimate the actual parameters, $k_{m,1}$, $k_{m,2}$ and $k_{s,12}$ in terms of the apparent values.

4.3.4 *Three substrate ping-pong mechanism*

An analysis of the three substrate ping-pong mechanism is presented next. The model was stated in the form given in Section 3.2:

$$\left.\begin{aligned} &\mathbf{x}(t,\mathbf{p}) = (S_1(t,\mathbf{p}), S_2(t,\mathbf{p}), S_3(t,\mathbf{p}), P(t,\mathbf{p}))^T, \\ &\frac{d\mathbf{x}(t,\mathbf{p})}{dt} = f_3(S_1(t,\mathbf{p}), S_2(t,\mathbf{p}), S_3(t,\mathbf{p}), \mathbf{p})(-1,-1,-1,1)^T, \\ &\mathbf{y}(t,\mathbf{p}) = (0,0,0,1)\cdot\mathbf{x}(t,\mathbf{p}) = P(t,\mathbf{p}), \\ &\mathbf{x}(0,\mathbf{p}) = \mathbf{x}_0 = (S_{10}, S_{20}, S_{30}, 0)^T, \end{aligned}\right\} \tag{4.64}$$

where $\mathbf{x}$ and $\mathbf{y}$ denote the state and observation vectors respectively. The unknown parameter vector is denoted by $\mathbf{p}$ and given by:

$$\mathbf{p} = (V_{max}, k_{m,1}, k_{m,2}, k_{m,3}). \tag{4.65}$$

Pohjanpalo's Jacobian rank test shows this model to be at least locally identifiable. However solving the Taylor series coefficients necessary to complete a Taylor series analysis proved computationally intractable. An analysis using the input-output relationship approach yielded

complete results.

The generators of the differential ideal, analogous to those derived for the two substrate cases, are as follows:

$$S_1 - S_{10} + P = 0, \tag{4.66a}$$

$$S_2 - S_{20} + P = 0, \tag{4.66b}$$

$$S_3 - S_{30} + P = 0, \tag{4.66c}$$

$$P^{(1)}(S_1S_2S_3 + S_1S_3k_{m,2} + S_2S_3k_{m,1} + S_1S_2k_{m,3}) - V_{max}S_1S_2S_3 = 0, \tag{4.66d}$$

$$y - P = 0. \tag{4.66e}$$

Dependence on t and $\mathbf{p}$ is omitted for brevity. An output relation was obtained using the elimination ranking $y < P < S_1 < S_2 < S_3$:

$$y^{(1)} = \frac{V_{max}(y^3 - y^2(S_{10} + S_{20} + S_{30}) + y(S_{10}S_{20} + S_{20}S_{30} + S_{30}S_{10}) - S_{10}S_{20}S_{30})}{\Big(y^3 - y^2(k_{m,1} + k_{m,3} + k_{m,2} + S_{10} + S_{20} + S_{30}) + y\Big(k_{m,2}(S_{10} + S_{30}) + \\ + (S_{10} + k_{m,1})(S_{20} + S_{30}) + S_{20}S_{30} + k_{m,3}(S_{10} + S_{20})\Big) - \\ - S_{10}S_{20}S_{30} - k_{m,1}S_{20}S_{30} - k_{m,2}S_{10}S_{30} - k_{m,3}S_{10}S_{20}\Big)}. \tag{4.67}$$

As previously, an alternative parameter vector, $\overline{\mathbf{p}} = (\overline{V_{max}}, \overline{k_{m,1}}, \overline{k_{m,2}}, \overline{k_{m,3}})$, was used to create two versions of the output relation. The second relation was subtracted from the first eliminating $y^{(1)}$. The resulting expression was rearranged to produce a degree six polynomial in $y(t)$, the monomials of which are linearly independent by the FTA. The seven nonzero coefficients were solved to yield $\overline{\mathbf{p}} = \mathbf{p}$, indicating that the model is globally identifiable.

The case where the initial substrate concentrations are unknown was also analysed using the

input-output relationship approach. The unknown parameter vector was modified as follows:

$$\mathbf{p} = (V_{max}, k_{m,1}, k_{m,2}, k_{m,3}, S_{10}, S_{20}, S_{30}), \tag{4.68}$$

and a corresponding alternative parameter vector was introduced:

$$\overline{\mathbf{p}} = (\overline{V_{max}}, \overline{k_{m,1}}, \overline{k_{m,2}}, \overline{k_{m,3}}, \overline{S_{10}}, \overline{S_{20}}, \overline{S_{30}}). \tag{4.69}$$

The leader of the output relation was eliminated as previously described to obtain a degree six polynomial in $y(t)$. The coefficients of this polynomial could not be simultaneously solved for the alternative parameters in *Maple*, however it was possible to solve them iteratively using *Maple* functions. Six solutions were obtained, in all cases $\overline{V_{max}} = V_{max}$:

$$\left\{\overline{k_{m,1}} = k_{m,1}, \overline{k_{m,2}} = k_{m,2}, \overline{k_{m,3}} = k_{m,3}, \overline{S_{10}} = S_{10}, \overline{S_{20}} = S_{20}, \overline{S_{30}} = S_{30}\right\}, \tag{4.70a}$$

$$\left\{\overline{k_{m,1}} = k_{m,1}, \overline{k_{m,2}} = k_{m,3}, \overline{k_{m,3}} = k_{m,2}, \overline{S_{10}} = S_{10}, \overline{S_{20}} = S_{30}, \overline{S_{30}} = S_{20}\right\}, \tag{4.70b}$$

$$\left\{\overline{k_{m,1}} = k_{m,2}, \overline{k_{m,2}} = k_{m,1}, \overline{k_{m,3}} = k_{m,3}, \overline{S_{10}} = S_{20}, \overline{S_{20}} = S_{10}, \overline{S_{30}} = S_{30}\right\}, \tag{4.70c}$$

$$\left\{\overline{k_{m,1}} = k_{m,2}, \overline{k_{m,2}} = k_{m,3}, \overline{k_{m,3}} = k_{m,1}, \overline{S_{10}} = S_{20}, \overline{S_{20}} = S_{30}, \overline{S_{30}} = S_{10}\right\}, \tag{4.70d}$$

$$\left\{\overline{k_{m,1}} = k_{m,3}, \overline{k_{m,2}} = k_{m,2}, \overline{k_{m,3}} = k_{m,1}, \overline{S_{10}} = S_{30}, \overline{S_{30}} = S_{10}, \overline{S_{20}} = S_{20}\right\}, \tag{4.70e}$$

$$\left\{\overline{k_{m,1}} = k_{m,3}, \overline{k_{m,2}} = k_{m,1}, \overline{k_{m,3}} = k_{m,1}, \overline{S_{10}} = S_{30}, \overline{S_{30}} = S_{20}, \overline{S_{20}} = S_{10}\right\}. \tag{4.70f}$$

It would be possible to practically eliminate solutions as described for the two substrate case (Section 4.3.1) by not using solutions where the estimated substrate concentration is higher than the known concentration in the reaction mixture.

If the initial concentration of product is also unknown the output relation still yields a degree six polynomial. As such at least one of the eight unknown parameters must be unidentifiable. While further analysis of this case is computationally intractable, it is possible to show that V_{max}

remains globally identifiable. By analogy with the two substrate case it can be hypothesised that the $k_{m,i}$s will be locally identifiable while the substrate and product concentrations will be unidentifiable. However it is currently impossible to confirm this supposition. If it does hold then it would be possible to eliminate inappropriate solutions as described for the two substrate case in Section 4.3.1.

4.3.5 Three substrate simple ordered mechanism

The three substrate simple ordered mechanism is now analysed. The model is similar to that for the three substrate ping-pong mechanism with the following alteration to the function governing the derivative of the state vector:

$$\frac{d\mathbf{x}(t,\mathbf{p})}{dt} = f_4(S_1(t,\mathbf{p}), S_2(t,\mathbf{p}), S_3(t,\mathbf{p}), \mathbf{p})(-1,-1,-1,1)^T. \tag{4.71}$$

The unknown parameter vector for this model is given by:

$$\mathbf{p} = (V_{max}, k_{m,1}, k_{m,2}, k_{m,3}, k_{s,12}, k_{s,23}). \tag{4.72}$$

For the Taylor series coefficients calculated the Jacobian matrix with respect to the unknown parameter vector was rank deficient; having a rank of four. Computational resources proved insufficient to carry out a complete Taylor series analysis. Since Taylor series based approaches were inconclusive, the input-output relationship approach was used.

The generators of the differential ideal, similar to those for the three substrate ping-pong model,

are as follows:

$$S_1 - S_{10} + P = 0, \tag{4.73a}$$

$$S_2 - S_{20} + P = 0, \tag{4.73b}$$

$$S_3 - S_{30} + P = 0, \tag{4.73c}$$

$$P^{(1)}\Big(S_1S_2S_3 + S_1S_3k_{m,2} + S_2S_3k_{m,1} + S_1S_2k_{m,3} + S_3k_{m,2}k_{s,12} + \\ + S_1k_{m,3}k_{s,23} + k_{m,3}k_{s,12}k_{s,23}\Big) - V_{max}S_1S_2S_3 = 0 \tag{4.73d}$$

$$y - P = 0. \tag{4.73e}$$

Dependence on t and $\mathbf{p}$ is again omitted for notational convenience. An output relation was obtained using the elimination ranking $y < P < S_1 < S_2 < S_3$:

$$y^{(1)} = \frac{V_{max}(y^3 - y^2(S_{10} + S_{20} + S_{30}) + y(S_{10}S_{20} + S_{20}S_{30} + S_{30}S_{10}) - S_{10}S_{20}S_{30})}{\Big[y^3 - y^2(S_{10} + S_{20} + S_{30} + k_{m,1} + k_{m,2} + k_{m,3}) + y\Big[(S_{10} + k_{m,1})(S_{20} + S_{30}) \\ + S_{20}S_{30} + k_{m,2}(k_{s,12} + S_{10} + S_{30}) + k_{m,3}(S_{10} + S_{20} + k_{s,23})\Big] - \Big[S_{10}S_{20} \\ + k_{s,23}(k_{s,12} + S_{10})\Big]k_{m,3} - k_{m,2}S_{30}(S_{10} + k_{s,12}) - (S_{10} + k_{m,1})S_{20}S_{30}\Big]}. \tag{4.74}$$

The alternative parameter vector, $\overline{\mathbf{p}} = (\overline{V_{max}}, \overline{k_{m,1}}, \overline{k_{m,2}}, \overline{k_{m,3}}, \overline{k_{s,12}}, \overline{k_{s,23}})$, was used to create two versions of the output relation. Once $y^{(1)}$ was eliminated the resulting expression was rearranged to obtain a degree six polynomial in y. As such the FTA guaranteed linear independence of the monomials. The seven nonzero coefficients were solved in *Maple* [158] to obtain relations between $\mathbf{p}$ and $\overline{\mathbf{p}}$. V_{max} and hence k_4 are globally identifiable. However there are insufficient relations to obtain solutions for $\overline{k_{m,3}}$ or $\overline{k_{s,23}}$ in terms of the elements of $\mathbf{p}$ so the model is unidentifiable.

If the initial substrate concentrations are treated as unknown parameters it is again possible to construct a degree six polynomial in y from the output relation. Since the total number of

unknown parameters, nine, exceeds the number of coefficients, seven, at least two parameters are unidentifiable. It is possible to show that V_{max} remains globally identifiable. Given that for the two substrate case the initial substrate concentrations are locally identifiable it is hypothesised that they are also locally identifiable. However analysis of this supposition proved computationally intractable.

Deconvolution of quasi-steady state parameters using knowledge of V_{max} as described in Section 4.3.3 is not possible for this model. The constraints created by the apparent parameters and initial conditions are insufficient to determine the actual parameters.

4.4 Structural indistinguishability analysis

The reaction mechanisms of the enzymes in the cytoplasmic phase of the peptidoglycan biosynthesis pathway are well characterised. However when studying a new enzyme this would not be the case. In this case a structural indistinguishability analysis would be required to determine whether a specific reaction mechanism could be identified from the data that could be collected. Such an analysis is presented in this section for the models introduced earlier in this chapter.

This analysis was carried out using the input-output relationship approach (Section 3.3) for all pairs of models. An additional analysis was undertaken using the Taylor series approach for the two models using two substrates. As explained in Section 4.3 it is unnecessary to consider alternative output structures for each pair of models.

A summary of the results are presented in Table 4.2. The ping-pong and simple ordered models for a given number of substrates are indistinguishable; indicating that either model will adequately describe a single experimental time course. As such more complex experiments would be needed to differentiate between these mechanisms. If mechanisms require a different number of substrates they are distinguishable. However distinguishing between such mechanisms is relatively simple in any case so this particular result is not of much interest.

		Result	
Model 1	Model 2	KSC	USC
2PP	2SO	indist	indist
	3PP	dist	dist
	3SO	dist	indist
2SO	3PP	dist	dist
	3SO	dist	indist
3PP	3SO	indist	CI

Table 4.2: Summary of the structural indistinguishability analysis results for two and three quasi-steady state substrate models. Models are denoted as described in Table 4.1. Two cases were considered for each pair of models; known substrate concentrations (KSC) and unknown substrate concentrations (USC). Results are either: indist, the models are indistinguishable; dist, the models are distinguishable; or CI, analysis was computationally intractable.

4.4.1 Two substrate ping-pong and simple ordered models

The models have been described previously in Sections 4.3.1 and 4.3.2. The variables and parameters of the simple ordered model were relabelled as: $\widehat{\mathbf{x}} = (\widehat{S_1}(t), \widehat{S_2}(t), \widehat{P}(t))^T$ and $\widehat{\mathbf{p}} = (\widehat{V_{max}}, \widehat{k_{m,1}}, \widehat{k_{m,2}}, \widehat{k_{s,12}})$. The first six Taylor series coefficients for each model yielded the following relations between the two unknown parameter vectors:

$$\left\{ \begin{array}{l} \widehat{V_{max}} = V_{max},\ \widehat{k_{m,1}} + \widehat{k_{m,2}} = k_{m,1} + k_{m,2}, \\ \\ \widehat{k_{m,2}}((S_{20} - S_{10}) + \widehat{k_{s,12}}) = k_{m,2}(S_{20} - S_{10}). \end{array} \right\} \tag{4.75}$$

Analysis of further Taylor series coefficients did not produce further relations. This does not demonstrate that the models are indistinguishable for the same reason that a model cannot be shown to be unidentifiable from analysis of a finite number of Taylor series coefficients. It would be necessary to extend the inductive argument presented in Section 4.3.2 to show that these models are indistinguishable using the Taylor series approach. Instead the input-output relationship approach was used.

The two output relations described in Sections 4.3.1 and 4.3.2 were used, relabelling the parameters of the simple ordered model as described above. The leader of these relations was eliminated by subtraction of one from the other and the expression rearranged to obtain a degree four polynomial in y. The coefficients of the monomials, being linearly independent by the FTA yielded five nonzero coefficients. These coefficients were solved in *Maple* to obtain three relations between the two unknown parameter vectors, equivalent to those above, (4.75). If it is assumed that $S_{10} = S_{20}$ then $\widehat{k_{m,2}}\widehat{k_{s,12}} = 0$ and thus $\widehat{k_{m,2}} = 0$ or $\widehat{k_{s,2}} = 0$.

Since parameters must be generically nonzero these solutions do not lie in the open set of feasible parameter values and can be rejected. (Such solutions will referred to as invalid.) Thus in this case the models are distinguishable. However it is impossible to guarantee that this assumption holds experimentally, in general the substrate concentrations will differ. As such two cases must be considered. If $S_{10} < S_{20}$ then any value of $\widehat{k_{s,12}} > 0$ results in a value of $\widehat{k_{m,2}} > 0$. Thus any solution is valid. If $S_{20} < S_{10}$ then if $\widehat{k_{s,12}} \geq |S_{20} - S_{10}|$ then $\widehat{k_{m,2}} \leq 0$, an invalid solution. However if $0 < \widehat{k_{s,12}} < |S_{20} - S_{10}|$ then $\widehat{k_{m,2}} > 0$ which is valid. Thus a valid solution exists in either experimentally practical case thus the models are indistinguishable.

For unknown initial concentrations of substrates two solutions were obtained from the coefficients of the polynomial obtained from the output relations. The first is equivalent to that above, the second is shown below:

$$\left.\begin{array}{c} \widehat{V_{max}} = V_{max},\ \widehat{k_{m,1}} + \widehat{k_{m,2}} = k_{m,1} + k_{m,2}, \\ \widehat{k_{m,2}}(\widehat{k_{s,12}} + (S_{10} - S_{20})) = k_{m,1}(S_{10} - S_{20}), \\ \widehat{S_{10}} = S_{20},\ \widehat{S_{20}} = S_{10}. \end{array}\right\} \tag{4.76}$$

This second solution is comparable to the additional solution obtained in the structural identifiablity analysis of the two substrate ordered model under the same assumptions. A case study similar to that above shows that a singular solution exists where the models are distinguishable

if $S_{10} = S_{20}$. Otherwise either $k_{s,12}$ or $k_{m,2}$ can be chosen freely, unless $S_{10} - S_{20} < 0$, in which case $k_{s,12}$ must be less than $|S_{10} - S_{20}|$. The existence of either of these solutions is sufficient to show that the models remain indistinguishable.

4.4.2 Two and three substrate ping-pong models

The models have been described previously in Sections 4.3.1 and 4.3.4. The variables and parameters of the three substrate ping-pong model were relabelled as: $\widehat{\mathbf{x}} = (\widehat{S_1}(t), \widehat{S_2}(t), \widehat{S_3}(t), \widehat{P}(t))^T$ and $\widehat{\mathbf{p}} = (\widehat{V_{max}}, \widehat{k_{m,1}}, \widehat{k_{m,2}}, \widehat{k_{m,3}})$. The two output relations described in Sections 4.3.1 and 4.3.4 were used. The leader of these relations was eliminated, as previously described, and the expression rearranged to obtain a degree five polynomial in y. The coefficients of the monomials, being linearly independent by the FTA yielded six nonzero coefficients. These coefficients were solved in *Maple* to obtain the following four relations between the two unknown parameter vectors:

$$\widehat{V_{max}} = V_{max},\ \widehat{k_{m,1}} = k_{m,1},\ \widehat{k_{m,2}} = k_{m,2},\ \widehat{k_{m,3}} = 0 \tag{4.77}$$

Since the fourth relation is invalid, the models are distinguishable.

For unknown initial concentrations of substrates twelve solutions were obtained from the coefficients of the polynomial derived from the output relations. Six solutions require at least one unknown parameter to equal zero, like the solution above. As such they are invalid. The remaining six take forms similar to that below:

$$\left\{ \begin{array}{c} \widehat{V_{max}} = V_{max},\ \widehat{k_{m,1}} = -\widehat{k_{m,3}} + k_{m,2},\ \widehat{k_{m,2}} = k_{m,1}, \\ \\ \widehat{k_{m,3}} = \widehat{k_{m,3}},\ \widehat{S_{10}} = S_{20},\ \widehat{S_{20}} = S_{10},\ \widehat{S_{30}} = S_{20}, \end{array} \right\} \tag{4.78}$$

where the initial concentrations of each of the three substrates from the three substrate reaction must equal that of one of the two from the two substrate reaction. These solutions can be

practically eliminated by choosing three significantly different substrate concentrations. As such the models are distinguishable.

4.4.3 Two substrate ping-pong and three substrate simple ordered models

The models have been described previously in Sections 4.3.1 and 4.3.5. The variables and parameters of the three substrate simple ordered model were relabelled as: $\widehat{\mathbf{x}} = (\widehat{S_1}(t), \widehat{S_2}(t), \widehat{S_3}(t), \widehat{P}(t))^T$ and $\widehat{\mathbf{p}} = (\widehat{V_{max}}, \widehat{k_{m,1}}, \widehat{k_{m,2}}, \widehat{k_{m,3}}, \widehat{k_{s,12}}, \widehat{k_{s,23}})$. The two output relations described in Sections 4.3.1 and 4.3.5 were used. The leader of these relations was eliminated, as previously described, and the expression rearranged to obtain a degree five polynomial in y. The coefficients of the monomials, being linearly independent by the FTA yielded six nonzero coefficients. These coefficients were solved in *Maple* to obtain two possible sets of relations between the two unknown parameter vectors:

$$\left\{ \begin{array}{l} \widehat{V_{max}} = V_{max}, \widehat{k_{m,1}} + \widehat{k_{m,2}} = k_{m,2} + k_{m,1}, \widehat{k_{m,2}} = \widehat{k_{m,2}}, \\ \\ \widehat{k_{m,3}} = 0,\ \widehat{k_{m,2}}(\widehat{k_{s,12}} + (S_{10} - S_{20})) = (S_{10} - S_{20})k_{m,2} \end{array} \right\} \tag{4.79}$$

$$\left\{ \begin{array}{l} \widehat{V_{max}} = V_{max},\ \widehat{k_{s,23}}(\widehat{k_{s,12}} + (S_{10} - S_{30})) = -(S_{10} - S_{30})(S_{20} - S_{30}), \\ \\ (\widehat{k_{m,1}}(S_{10} - S_{30}) - \widehat{k_{m,2}}\widehat{k_{s,12}})(\widehat{k_{s,12}} + (S_{10} + S_{20})) = \\ \qquad = k_{m,1}(\widehat{k_{s,12}} + (S_{10} - S_{20}))(S_{10} - S_{30}) + k_{m,2}\widehat{k_{s,12}}(S_{20} - S_{30}), \\ \\ \widehat{k_{m,3}}((S_{10} - S_{30})(\widehat{k_{s,12}} + (S_{10} - S_{20}))) + \\ \qquad + \widehat{k_{m,2}}(\widehat{k_{s,12}}(\widehat{k_{s,12}} + 2S_{10} - S_{20} - S_{30}) + (S_{10} - S_{20})(S_{10} - S_{30})) = \\ \qquad = k_{m,2}(S_{10} - S_{20})(\widehat{k_{s,12}} + (S_{10} - S_{30})). \end{array} \right\} \tag{4.80}$$

The first solution above is invalid, since $\widehat{k_{m,3}} = 0$. The second relation in the second solution yields a set of experimental conditions that allow the models to be distinguished, specifically

$S_{10} > S_{30}$ and $S_{20} > S_{30}$. Under these conditions:

$$\widehat{k_{s,23}}(\widehat{k_{s,12}} + (S_{10} - S_{30})) = -(S_{10} - S_{30})(S_{20} - S_{30}) < 0$$

$$\widehat{k_{s,23}}(S_{10} - S_{30}) + \widehat{k_{s,23}}\widehat{k_{s,12}} < 0.$$

Since $k_{s,23}$ must be greater than zero:

$$\widehat{k_{s,23}}\widehat{k_{s,12}} < 0. \tag{4.81}$$

However this requires that $\widehat{k_{s,12}} < 0$ which is invalid. Thus the models are distinguishable given an appropriate, and practical, choice of initial concentrations of the substrates.

For unknown initial concentrations of substrates sixteen solutions were obtained from the coefficients of the polynomial derived from the output relations. Eight solutions require at least one unknown parameter to equal zero, like Equations (4.79) above. As such they are invalid. The remaining eight are less simple. The solution below however:

$$\left\{ \begin{array}{l} \widehat{V_{max}} = V_{max},\ \widehat{k_{m,1}} + \widehat{k_{m,3}} + \widehat{k_{m,2}} = k_{m,1} + k_{m,2},\ \widehat{k_{s,12}} = \widehat{S_{20}} - S_{10}, \\ \\ \widehat{k_{m,3}}(\widehat{k_{s,23}} + (S_{10} - S_{20})) = k_{m,2}(S_{10} - S_{20}),\ \widehat{S_{10}} = S_{10},\ \widehat{S_{30}} = S_{20}, \end{array} \right\} \tag{4.82}$$

is sufficient to establish that these models are indistinguishable. A singular set of initial conditions exists for which this solution is distinguishable. If $S_{10} = S_{20}$ then either $\widehat{k_{m,3}}$ or $\widehat{k_{s,23}}$ must equal zero. As previously observed in Section 4.4.1 a solution of this type is experimentally impractical. Otherwise a case analysis identical to those in Section 4.4.1 shows that the fourth relation above allows one of the parameters, $\widehat{k_{m,3}}$ and $\widehat{k_{s,23}}$, to be chosen with some constraints. Thus valid solutions for all unknown parameters of the three substrate ordered model exist and as such the models are indistinguishable.

4.4.4 Two substrate simple ordered and three substrate ping-pong models

The models have been described previously in Sections 4.3.2 and 4.3.4. The variables and parameters of the three substrate ping-pong model were relabelled as described in Section 4.4.2. The two output relations described in Sections 4.3.2 and 4.3.4 were used. The leader of these relations was eliminated, as previously described, and the expression rearranged to obtain a degree five polynomial in y. The coefficients of the monomials, being linearly independent by the FTA yielded six nonzero coefficients. These coefficients were solved in *Maple* to obtain the following four relations between the two unknown parameter vectors:

$$\left\{ \begin{array}{c} \widehat{V_{max}} = V_{max},\ \widehat{k_{m,1}}(S_{10} - S_{20}) = k_{m,1}(S_{10} - S_{20}) - k_{m,2}k_{s,12}, \\ \\ \widehat{k_{m,2}}(S_{10} - S_{20}) = k_{m,2}(k_{s,12} + (S_{10} - S_{20})),\ \widehat{k_{m,3}} = 0. \end{array} \right\} \tag{4.83}$$

Since the fourth relation requires that $\widehat{k_{m,3}} = 0$ the solution is invalid; thus the models are distinguishable.

For unknown initial concentrations of substrates twelve solutions were obtained from the coefficients of the polynomial derived from the output relations. Six solutions require at least one unknown parameter to equal zero, like Equations (4.83) above. As such they are invalid. The remaining six solutions are like those described in Section 4.4.2, Equations (4.78). As such they can be rendered invalid by choosing all initial conditions to be significantly different. Thus these models are distinguishable.

4.4.5 Two and three substrate simple ordered models

The models have been described previously in Sections 4.3.2 and 4.3.5. The variables and parameters of the three substrate simple ordered model were relabelled as described in Section 4.4.3. The two output relations described in Sections 4.3.2 and 4.3.5 were used. The leader of

these relations was eliminated, as previously described, and the expression rearranged to obtain a degree five polynomial in y. The coefficients of the monomials, being linearly independent by the FTA yielded six nonzero coefficients. These coefficients were solved in *Maple* to obtain two possible sets of relations between the two unknown parameter vectors:

$$\left\{ \begin{array}{l} \widehat{V_{max}} = V_{max},\ \widehat{k_{m,1}} + \widehat{k_{m,2}} = k_{m,2} + k_{m,1},\ \widehat{k_{m,3}} = 0, \\ \\ \widehat{k_{m,2}}(\widehat{k_{s,12}} + (S_{10} - S_{20})) = k_{m,2}(k_{s,12} + (S_{10} - S_{20})) \end{array} \right\} \tag{4.84}$$

$$\left\{ \begin{array}{l} \widehat{V_{max}} = V_{max},\ \widehat{k_{s,23}}(\widehat{k_{s,12}} + (S_{10} - S_{30})) = -(S_{10} - S_{30})(S_{20} - S_{30}) \\ \\ (\widehat{k_{m,1}}(S_{10} - S_{30}) - \widehat{k_{m,2}}\widehat{k_{s,12}})(\widehat{k_{s,12}} + (S_{10} - S_{20})) = \\ \qquad = k_{m,2}(k_{s,12}(S_{30} - S_{10}) + \widehat{k_{s,12}}(S_{20} - S_{30}) - \widehat{k_{s,12}}k_{s,12}) + \\ \qquad\quad + k_{m,1}((\widehat{k_{s,12}} + (S_{10} - S_{20}))(S_{10} - S_{30})), \\ \\ \widehat{k_{m,3}}(S_{10} - S_{30})(\widehat{k_{s,12}} - S_{20} + S_{10}) + \\ \qquad + \widehat{k_{m,2}}(\widehat{k_{s,12}}(\widehat{k_{s,12}} + 2S_{10} - S_{20} - S_{30}) + (S_{10} - S_{20})(S_{10} - S_{30})) = \\ \qquad\quad = k_{m,2}((k_{s,12} + (S_{10} - S_{20}))(\widehat{k_{s,12}} + (S_{10} - S_{30}))). \end{array} \right\} \tag{4.85}$$

The first solution is invalid since it requires that $\widehat{k_{m,3}} = 0$. The second relation in the second solution is identical to that obtained in Section 4.4.3, Equations 4.80. As such by reasoning advanced in that section the two models are distinguishable for the initial conditions $S_{10} > S_{30}$ and $S_{20} > S_{30}$.

For unknown initial concentrations of substrates twenty solutions were obtained from the coef-

ficients of the polynomial derived from the output relations. The solution below:

$$\left\{ \begin{array}{l} \widehat{V_{max}} = V_{max},\ \widehat{k_{m,1}} + \widehat{k_{m,2}} + \widehat{k_{m,3}} = k_{m,1} + k_{m,2}, \\ \\ \widehat{k_{s,12}} = \widehat{S_{20}} - S_{10},\ \widehat{S_{10}} = S_{10},\ \widehat{S_{30}} = S_{20}, \\ \\ (\widehat{k_{s,23}} + (S_{10} - S_{20}))\widehat{k_{m,3}} = (k_{s,12} + (S_{10} - S_{20}))k_{m,2}, \end{array} \right\} \tag{4.86}$$

like Equations (4.82) is sufficient to establish that these models are indistinguishable. A case analysis of the sixth relation is necessary. If $S_{10} \geq S_{20}$ there are infinitely many valid solutions for $\widehat{k_{s,23}}$ and $\widehat{k_{m,3}}$. If $S_{10} < S_{20}$ then:

$$\left. \begin{array}{l} 0 < k_{s,12} < |S_{10} - S_{20}| \\ k_{s,12} = |S_{10} - S_{20}| \\ k_{s,12} > |S_{10} - S_{20}| \end{array} \right\} \text{requires} \left. \begin{array}{l} 0 < \widehat{k_{s,23}} < |S_{10} - S_{20}| \\ \widehat{k_{s,23}} = |S_{10} - S_{20}| \\ \widehat{k_{s,23}} > |S_{10} - S_{20}| \end{array} \right\}, \tag{4.87}$$

and there are still infinitely many valid solutions for $\widehat{k_{s,23}}$ and $\widehat{k_{m,3}}$. Thus valid solutions for all unknown parameters of the three substrate ordered model exist and as such the models are indistinguishable.

4.4.6 Three substrate ping-pong and simple ordered models

The models have been described previously in Sections 4.3.4 and 4.3.5. The variables and parameters of the simple ordered model were relabelled as described in Section 4.4.3. The two output relations described in Sections 4.3.2 and 4.3.5 were used. The leader of these relations was eliminated, as previously described, and the expression rearranged to obtain a degree six polynomial in y. The coefficients of the monomials, being linearly independent by the FTA yielded seven nonzero coefficients. These coefficients were solved in *Maple* to obtain three

possible sets of relations between the two unknown parameter vectors:

$$\left\{\begin{array}{l} \widehat{V_{max}} = V_{max}, \widehat{k_{m,1}} + \widehat{k_{m,2}} = k_{m,2} + k_{m,1}, \widehat{k_{m,3}} = k_{m,3}, \\ \\ \widehat{k_{m,2}}(\widehat{k_{s,12}} + (S_{10} - S_{20})) = k_{m,2}(S_{10} + S_{20}), \widehat{k_{s,23}} = 0, \end{array}\right\} \tag{4.88}$$

$$\left\{\begin{array}{l} \widehat{V_{max}} = V_{max},\ \widehat{k_{m,2}} = 0, \\ \\ \widehat{k_{m,1}} S_{10} - S_{30})(\widehat{k_{s,12}} + (S_{10} - S_{20})) = \\ \qquad = k_{m,1}(\widehat{k_{s,12}} + (S_{10} - S_{20}))(S_{10} - S_{30}) + k_{m,2}\widehat{k_{s,12}}(S_{20} - S_{30}), \\ \\ \widehat{k_{m,3}}(S_{10} - S_{30})(\widehat{k_{s,12}} + (S_{10} - S_{20})) = \\ \qquad = k_{m,2}(S_{10} - S_{20})(\widehat{k_{s,12}} + (S_{10} - S_{30})) + \\ \qquad\quad + k_{m,3}(S_{10} - S_{30})(\widehat{k_{s,12}} + (S_{10} - S_{20})), \\ \\ \widehat{k_{s,23}}(k_{m,2}(S_{10} - S_{20})(\widehat{k_{s,12}} + (S_{10} - S_{30})) + \\ \quad + k_{m,3}(\widehat{k_{s,12}} + (S_{10} - S_{20}))(S_{10} - S_{30})) = \\ \qquad = k_{m,2}(S_{20} - S_{10})(S_{20} - S_{30})(S_{10} - S_{30}), \end{array}\right\} \tag{4.89}$$

$$
\left\{
\begin{array}{l}
\widehat{V_{max}} = V_{max}, \\
\\
\widehat{k_{m,1}}(S_{20}-S_{30})((k_{m,3}-\widehat{k_{m,3}})(S_{10}-S_{30})-\widehat{k_{m,3}}\widehat{k_{s,23}})+ \\
\qquad +\widehat{k_{m,3}}^{\,2}((S_{30}-S_{10})(S_{20}-S_{30})-\widehat{k_{s,23}}(\widehat{k_{s,23}}+S_{10}+S_{20}-2S_{30}))= \\
= k_{m,3}\left(\begin{array}{c}(k_{m,3}-2\widehat{k_{m,3}}+k_{m,2}+k_{m,1})(S_{30}-S_{10})(S_{30}-S_{20})+ \\ +\widehat{k_{m,3}}(\widehat{k_{s,23}}(2S_{30}-S_{10}-S_{20}))\end{array}\right)+ \\
\quad +k_{m,2}\widehat{k_{m,3}}(\widehat{k_{s,12}}+(S_{20}-S_{30}))(S_{30}-S_{10})+ \\
\qquad +k_{m,1}\widehat{k_{m,3}}(\widehat{k_{s,23}}+(S_{10}-S_{30}))(S_{30}-S_{20}), \\
\widehat{k_{m,2}}(S_{20}-S_{30})\left(\begin{array}{c}(k_{m,3}-\widehat{k_{m,3}})(S_{10}-S_{30})- \\ -\widehat{k_{m,3}}\widehat{k_{s,23}}\end{array}\right)+\widehat{k_{m,3}}(\widehat{k_{s,23}}+(S_{10}-S_{30}))= \\
\qquad =\widehat{k_{m,3}}\widehat{k_{s,23}}(k_{m,2}(S_{10}-S_{20})+k_{m,3}(S_{10}-S_{30})), \\
\\
\widehat{k_{m,3}}+\widehat{k_{m,3}}(\widehat{k_{s,23}}(S_{10}-S_{30})+(S_{30}-S_{10})(S_{30}-S_{20})+\widehat{k_{s,12}}\widehat{k_{s,23}})= \\
\\
\qquad =k_{m,3}(S_{30}-S_{10})(S_{30}-S_{20}).
\end{array}
\right\}
\tag{4.90}
$$

The first two solutions Equations (4.88) and (4.89) are invalid since they require an unknown parameter to equal zero. The third has a singular invalid solution; when $S_{10} = S_{20} = S_{30}$, $\widehat{k_{s,12}} = 0$. However for non-singular initial substrate concentrations it has infinitely many valid solutions. As such the models are indistinguishable. For unknown initial conditions analysis proved computationally intractable.

4.5 Summary

In this chapter mechanisms for two and three substrate enzyme catalysed reactions have been introduced. These mechanisms were restated as differential equations using mass action ki-

netics. Quasi-steady state assumptions were then used to simplify these models. Structural identifiability and indistinguishability analyses of these quasi-steady state models were then undertaken.

The two and three substrate ping-pong models were shown to be structurally globally identifiable when initial concentrations of the substrates were known and locally identifiable when these values were unknown. The experimental designs described for each case would allow the appropriate local solution to be determined. Thus it is appropriate to estimate parameters from single progress curves for mechanisms following these reaction schemes. The two and three substrate simple ordered mechanisms are structurally unidentifiable. It is possible to use the identifiable parameter of the two substrate simple ordered mechanism, V_{max}, to deconvolute kinetic parameters obtained from typical experimental methods. While V_{max} can be estimated in the three substrate case no deconvolution is possible.

Using the input-output relationship approach it was possible to complete the analysis for almost all models considered. The necessary calculations become too computationally intensive in only three cases. These problems principally arise when larger models are used and where initial conditions are unknown. By contrast the Taylor series approach failed in a number of cases to produce a solution due to higher computational requirements. It was possible to adapt the Taylor series approach, using an inductive argument, to show that the two substrate simple ordered model was unidentifiable. Where the Taylor series and input-output relationship approaches both produced results, those results were equivalent.

For known initial conditions and appropriate experimental design the two and three substrate mechanisms are distinguishable. For unknown initial conditions it becomes impossible to distinguish either of the two substrate mechanisms from the three substrate simple ordered mechanism. However these results are of limited utility. As long as the enzyme in question catalyses only one reaction, it is possible to distinguish two and three substrate mechanisms experimentally by providing only two substrates. If the enzyme requires two substrates a reaction will

occur; if three no reaction will occur. Addition of a third substrate will not (in general) affect the activity of a two substrate enzyme while a rate change should be immediately observable for a three substrate enzyme.

For both two and three substrate cases, ping-pong and simple ordered mechanisms are practically indistinguishable, regardless of whether initial conditions are known. Since the equations governing the simple ordered mechanisms can be obtained from those for the ping-pong mechanisms by addition of new parameters this was expected. However experimental procedures exist for differentiating these mechanisms. As such for each enzyme the appropriate model should be chosen by consulting the literature or carrying out the necessary experiments.

5. STEADY STATE CHARACTERISATION OF ENZYMES MURA-F FROM *S. PNEUMONIAE*

5.1 Introduction

In this chapter the results of kinetic characterisations of cloned *S. pneumoniae* enzymes MurA-D and F are presented. These characterisations were carried out under a consistent array of conditions as a preliminary step in the construction of a model of the cytoplasmic phase of the peptidoglycan biosynthesis pathway. A kinetic characterisation of *S. pneumoniae* MurE had already been undertaken under these conditions [110]. As such further characterisation was unnecessary. A summary of the data obtained is presented in Table 5.1.

Each kinetic characterisation was undertaken using one of the assays described in Section 2.4. For each enzyme a series of assays were undertaken in which the initial concentration of one of the substrates was varied. Typically the concentrations of the other substrates were significantly greater than their respective k_ms to minimise their effect on the kinetics of the reaction. The initial rate of each reaction was measured as described in Section 2.4 and plotted against the concentration of the varied substrate. Parameters were then estimated from the resulting curves using nonlinear curve fitting as described in Section 3.4. The k_{cat}s presented below were calculated from the apparent V_{max} estimated from data obtained when the concentration of the UDP-intermediate, that is the substrate containing a UDP group, was varied. The V_{max}s obtained for the other substrates deviated from this value by only a small amount as such this is a reasonable approximation.

Enzyme	Parameter name	Estimated value	Lower	Upper
MurA†	k_{cat} (s^{-1})	8.07	7.70	8.46
	k_m (PEP) (μM)	4.81	3.35	6.92
	k_m (UGP) (μM)	390	337	451
MurB†	k_{cat} (s^{-1})	3.17	3.02	3.34
	k_m (NADPH) (μM)	20.9	12.9	33.8
	k_m (UDPPEE) (μM)	32.8	25.6	42.1
MurB‡	k_{cat} (s^{-1})	4.46	3.95	5.04
	k_m (NADPH) (μM)	41.2	20.7	82.0
	k_i (NADPH) (μM)	1110	384	3200
	k_m (UDPPEE) (μM)	68.7	53.5	88.3
	k_i (UDPPEE) (μM)	1070	718	1600
MurC†	k_{cat} (s^{-1})	7.58	6.57	8.74
	k_m (ATP) (μM)	22.4	18.5	27.1
	k_m (L-Ala) (μM)	716	563	911
	k_m (UMN) (μM)	202	150	273
MurD†	k_{cat} (s^{-1})	36.5	34.3	38.7
	k_m (ATP) (μM)	38.2	31.5	46.4
	k_m (D-Glu) (μM)	293	243	353
	k_m (U1P) (μM)	26.2	21.5	32.1
MurF†	k_{cat} (s^{-1})	29.2	26.6	32.0
	k_m (ATP) (μM)	12.3	8.86	17.2
	k_m (D-Ala-D-Ala) (μM)	31.7	24.4	41.3
	k_m (U3P) (μM)	3.40	1.68	6.86
MurF‡	k_{cat} (s^{-1})	54.7	45.3	66.1
	k_m (ATP) (μM)	20.8	14.6	29.6
	k_i (ATP) (μM)	1810	949	3460
	k_m (D-Ala-D-Ala) (μM)	31.7	24.4	41.3
	k_m (U3P) (μM)	17.4	11.9	28.4
	k_i (U3P) (μM)	175	117	263

Table 5.1: Summary of the kinetic constants determined for the *S. pneumoniae* enzymes MurA-D and F. Parameters were determined from plots of initial velocity against initial substrate concentration by nonlinear curve fitting as described in Section 3.4. The upper and lower bounds form a 90% confidence interval. The k_{cat} was determined by converting the estimated V_{max} to μMs^{-1} and then normalising with respect to enzyme concentration. Two models were used to fit the data, Michaelis Menten kinetics with‡ or without† substrate inhibition.

Proteins were over-expressed in *E. coli* and purified by FPLC. The conditions used and yields obtained are briefly described in the following section. In the subsequent sections a more detailed examination of each of the enzyme characterisations is presented in alphabetical order. For the two substrate enzymes, MurA and MurB, parameter estimation from individual progress curves was also attempted, informed by the results in Chapter 4. While characterising MurF an effect of substrate inhibition on initial rates of reaction was observed. This effect was simulated and the results were compared to experimental data.

5.2 Protein expression and purification

E. coli BL21 Star Rosetta (DE3) was transformed using plasmids incorporating the *murA*-*murD* and *murF* genes within the pET46 EK/Lic vector, Section 2.2. The best protein expression levels were obtained from cultures grown at 27^oC in autoinduction media, Section 2.3.2. Crude cell lysates were initially purified using immobilised nickel columns, Section 2.3.4. The resulting solutions were shown to contain active enzyme using the assays described in Section 2.4; however SDS-PAGE showed that they contained a number of impurities. Two additional purification steps, using size exclusion and ion exchange columns, Sections 2.3.5 and 2.3.6, proved necessary to achieve the desired purities.

Protein yield (mg/L)				
MurA	MurB	MurC	MurD	MurF
0.83	2.30	25	8.2	6.2

Table 5.2: Yield of each enzyme after completion of the necessary purification steps.

SDS-PAGE indicated that the masses of the purified proteins were as expected (Figure 5.1); the masses of MurA, MurC, MurD and MurF lying in the 50-52kDa range while MurB was lighter, around 36kDa. Final yields of these enzymes are presented in Table 5.2. MurA yields were low compared to that of the other enzymes. Cultures were observed to grow unusually slowly when expressing MurA; it may be toxic to the cells used. Regardless the enzyme requirements of the

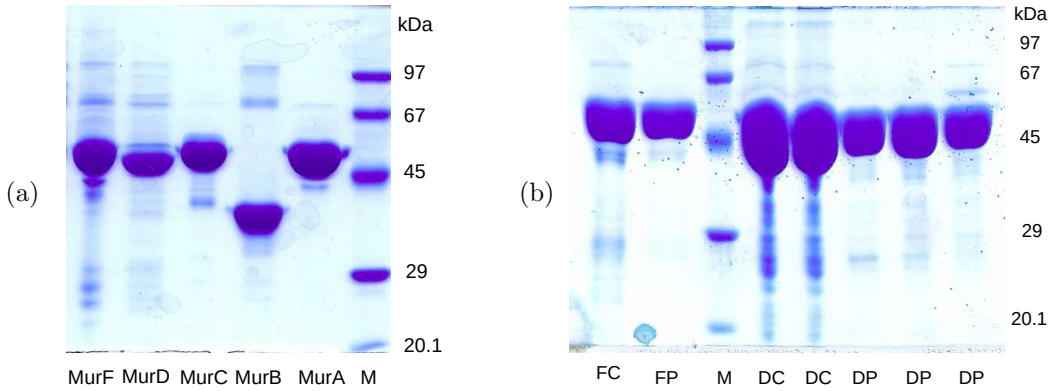


Figure 5.1: 10% SDS-PAGE of Mur enzymes at various stages of purification. Each lane of (a) corresponds to a different protein, labelled on the figure. Gel (b) demonstrates the effects of further purification steps. Lanes DC and FC correspond to protein solutions, containing MurD and MurF respectively, prior to a purification step, DP and FP to samples obtained after a purification step.

planned experimental program could be provided without further refinement of this expression and purification process.

5.3 MurA characterisation

MurA was characterised using the phosphate release assay as described in Section 2.4.3. Two reaction mixtures were prepared as described and monitored simultaneously using the Varian cell changer system. Reactions were initiated with whichever substrate, UGP or PEP, was not varied over the course of the assay. The final concentration of MurA in the reaction mixture was 28.5nM. Three series of assays were completed for each varied substrate and these data were fitted separately and combined for a final estimate of the relevant apparent kinetic parameters, Table 5.3. For determinations with respect to PEP, sixteen PEP concentrations, ranging between 1μM and 1mM, were used. For determinations with respect to UGP, sixteen UGP concentrations, ranging between 50μM and 2.5mM, were used. The concentration of the unvaried substrate was 1mM UGP or 480μM PEP.

Plots of the resulting data (Figure 5.2) show no signs of substrate inhibition which is characterised by a trend of decreasing initial reaction rate for initial substrate concentrations increasing

Series	Parameter Name	For PEP			For UGP		
		Values	Lower	Upper	Values	Lower	Upper
1	V_{max} (AU min^{-1})	0.1124	0.1001	0.1262	0.1513	0.1434	0.1596
1	k_m (μM)	4.51	2.75	7.41	400	342	469
2	V_{max} (AU min^{-1})	0.1137	0.1107	0.1168	0.1582	0.1480	0.1692
2	k_m (μM)	5.60	5.04	6.22	424	349	515
3	V_{max} (AU min^{-1})	0.1164	0.1037	0.1307	0.1456	0.1334	0.1589
3	k_m (μM)	6.01	4.03	8.96	355	271	465
Combined	V_{max} (AU min^{-1})	0.1129	0.1027	0.1242	0.1511	0.1442	0.1583
Combined	k_m (μM)	4.81	3.35	6.92	390	337	451

Table 5.3: Kinetic parameters obtained for MurA for each series of assays undertaken. Each series used sixteen concentrations of the varied substrate ranging between 1μM and 1mM for PEP and 50μM and 2.5mM for UGP. The unvaried substrate concentrations were 480μM PEP and 1mM UGP. The upper and lower bounds determine a 90% confidence interval.

past a threshold value. Instead initial rate appears to converge monotonically to an asymptote as initial substrate concentration increases. Such behaviour is typical of standard Michaelis-Menten type kinetics (described by Equation (3.38)). Kinetic parameters produced for each series of assays are comparable to each other and to the parameters estimated for the combined data. Where measurements were made at the same initial concentration in separate series the measurements were averaged in the combined data. This was necessary due to the limits of the fitting program used, Section 3.4. Note however that the parameters estimated from the combined data are not averages of the parameters of the individual sets and that the confidence intervals are tighter.

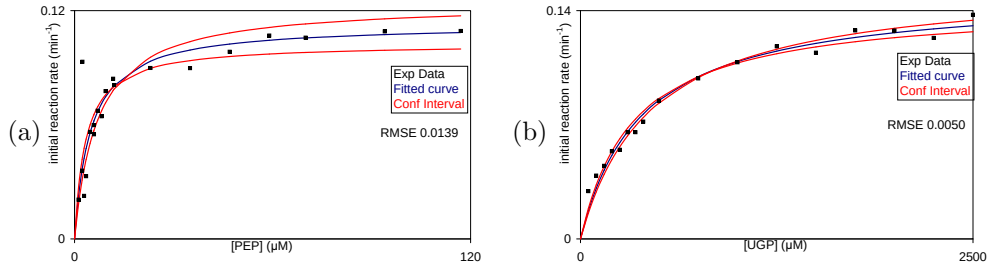


Figure 5.2: Plots of the combined MurA assay data. Graphs (a) and (b) correspond to varying PEP and UGP concentration respectively. Experimental data (Exp Data), the fitted curve, and a 90% confidence interval (Conf Interval) are plotted.

These results correspond in part to those produced by Du *et al.*, in which the enzyme used is referred to as MurA2 [82]. The estimate of the k_m with respect to PEP, 11.1μM, obtained in that work is fairly close to that obtained here. The lower estimate obtained for k_{cat}, 0.78s^{-1}, could be explained by the difference in temperature at which the two characterisations were performed. In previous work a discrepancy of the same order of magnitude was observed between k_{cat} values estimated for the alternative MurA form, MurA1 or MurZ. This further connects the difference between the estimated values to techniques used rather than an intrinsic difference between the purified enzymes. The value estimated by Du *et al* for the k_m with respect to UGP, 119μM, is significantly lower than that obtained in this work. Note that MurA is believed to follow a simple ordered mechanism it is possible therefore that this discrepancy arises from the influence of the dissociation constant $k_{s,UGP}$ ($k_{s,12}$ in the general notation presented previously). This possibility is further discussed in the following section.

5.3.1 Extended characterisation and direct fitting to progress curves

The structural identifiability analysis of the simple ordered model, Section 4.3.2, indicates that it is feasible to identify the actual V_{max} by parameter estimation. This should then allow the deconvolution of the apparent kinetic constants derived above. It is natural to test these theoretical observations using MurA.

A sensitivity analysis of MurB is presented in the following section, Section 5.4.1. Given that only V_{max} is to be estimated the conclusions of that analysis can be reasonably applied in this case. Thus sensitivity to V_{max} is maximised when relatively high initial substrate concentrations are used and the reaction is allowed to proceed to completion. A reaction mixture containing initial concentrations of PEP, UGP and MurA: 820μM, 100μM and 0.1μM respectively, was monitored using the phosphate release assay. The experimental data were first compared to data simulated using the parameters above, Figure 5.3(a). Production of phosphate is similar for the first minute of the reaction. However subsequently the experimental curve slows significantly;

eventually producing about half the phosphate predicted.

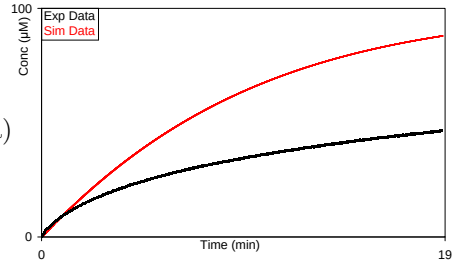
Two explanations for this discrepancy were considered: either the initial substrate and enzyme concentrations were incorrect or product inhibition was occurring. Parameter estimation was used to investigate these two possibilities, Figure 5.3(c). The initial concentrations of PEP and UGP and the value of V_{max} were estimated, using the k_ms above, for an uninhibited model (Equations (4.28) modified by Equation (4.47)). In addition these values and that of an additional inhibition parameter, k_i, were estimated for the following model:

$$\left.\begin{aligned} PEP^{(1)}(t) = UGP^{(1)}(t) &= -f_5(PEP, UGP, t), \\ P^{(1)}(t) = UDPPEE^{(1)}(t) &= f_5(PEP, UGP, t), \\ f_5(PEP, UGP, t) &= \frac{(k_i + UDPPEE(t))V_{max}PEP(t)UGP(t)}{k_i(PEP(t)UGP(t) + k_{m,1}UGP(t) + k_{m,2}PEP(t))}, \end{aligned}\right\} \tag{5.1}$$

with initial conditions as described by Equations (4.28). The best fit, that with the lowest RMS error (the square rooted sum of the squared residuals), was obtained when product inhibition was assumed and V_{max}, k_i and the initial conditions were estimated, Figure 5.3(b). However the improvement of this model over one in which only V_{max} and k_i are estimated is small and the estimates obtained for the initial conditions were experimentally unlikely. Thus the parameter estimation undertaken suggests that MurA is inhibited by UDPPEE. Two assays were undertaken as described in Section 2.4.3 with the initial conditions described above. One reaction mixture also contained 100μM UDPPEE and showed a 91% reduction in initial rate compared to the other assay. This confirms the observation that UDPPEE is an inhibitor of MurA. This interaction has not previously been reported; although a structural basis for this inhibition can be inferred from structural studies of this enzyme [178].

Thus in order to carry out the desired analysis it was necessary to develop an assay of MurA activity which does not allow accumulation of UDPPEE. This was achieved by coupling MurA

(a)



(b)

Estimated parameters				
V_{max}	ICs	k_i	RMSE	SD
yes	no	N/A	4.95	4.95
yes	yes	N/A	2.59	2.59
no	no	yes	1.10	1.10
yes	no	yes	0.45	0.45
yes	yes	yes	0.18	0.18

(c)

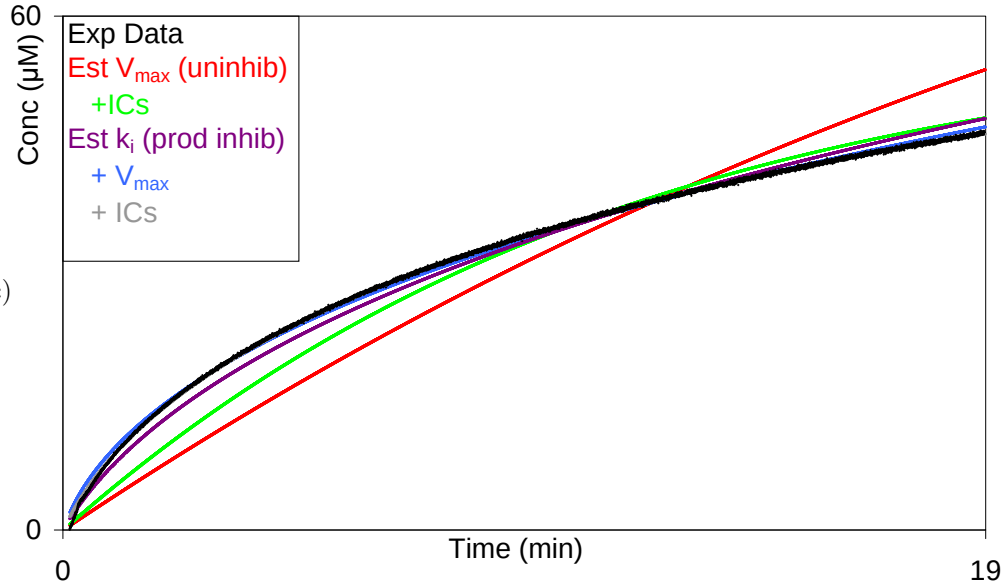


Figure 5.3: Comparison of experimental to simulated time courses for the MurA reaction. Graph (a) plots an experimental time course and a time course simulated for the same initial conditions using the parameters in Table 5.1. Graph (c) plots fits for two models of MurA: Equations (4.28) modified by Equation (4.47) (uninhib) and Equations (5.1) (prod inhib). Table (b) contains the RMSE and estimated standard deviation (SD) for these fits.

to MurB and MurC and using the ADP release assay previously described see Figure 5.4. Isocitrate and isocitrate dehydrogenase (IDH) were used to recycle NADPH to ensure that only ADP production was measured, Figure 5.5(a). MurB and MurC concentrations were 28.6 and 70μM respectively. Pyruvate kinase, lactate dehydrogenase and DTT concentrations were as described in Section 2.4.2. Isocitrate, NADPH and IDH concentrations were 10mM, 100μM and 0.18 units/ml respectively. NADH was provided in excess to ensure that the reaction could be monitored throughout. As such the concentration of NADH was 100μM higher than that of UGP.

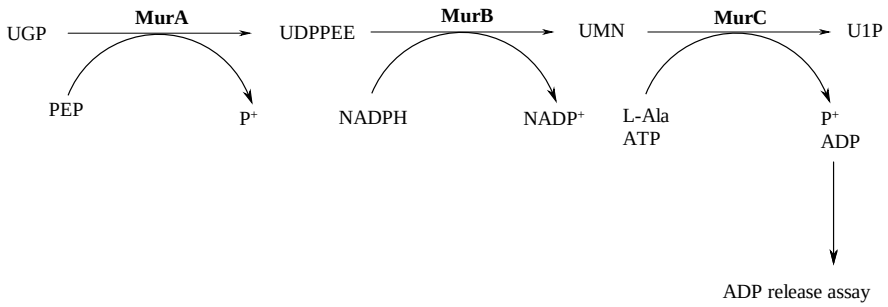
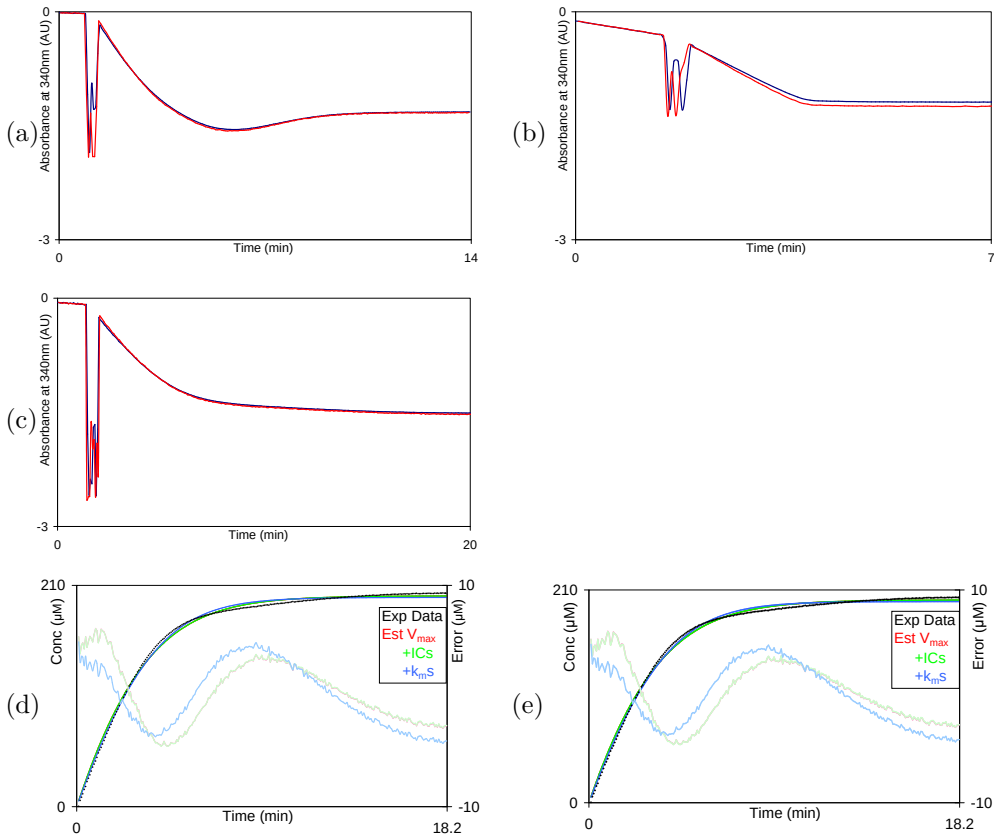


Figure 5.4: Schematic diagram of an alternative assay of MurA activity. MurA activity is coupled to ADP release via MurB and MurC. The ADP release assay is then used.

It is also theoretically possible to use NADPH reduction for a more direct measure of MurA activity. However at the concentration of MurB chosen, the rate of reduction of NADPH by MurB, in the absence of UDPPEE, was very high, Figure 5.5(b). As such it was impossible to accurately distinguish between UDPPEE production and NADPH autoreduction using this assay and so it was not used for the following analysis.

Reaction mixtures were prepared containing the coupling system described above and initial concentrations of PEP, UGP and MurA: 820μM, 200μM and 0.1μM respectively, Figure 5.5(c). Total production of ADP was approximately equal to the initial UGP present. Thus it is reasonable to conclude that this assay is not subject to product inhibition. Three combinations of parameters were estimated from the experimental curves produced, Figure 5.5(d-f). Those parameters which were not estimated were fixed by the stated initial conditions or parameters above, Table 5.1. The estimated time courses were in good agreement with the experimental data. The RMS errors (RMSEs) were lower than obtained for the product inhibited time courses. The estimated standard deviations were close to the RMSEs indicating a relative lack of bias in the fitting procedure. The best fits were obtained when V_{max} and the initial conditions were estimated.

The RMS errors associated with estimates of PEP initial concentration exceed the estimated values of this parameter, Figure 5.5(g). Thus little confidence should be placed in these estimates. The sensitivity of this reaction to the concentration of PEP is probably very low due



(f)

Estimated parameters			RMSE		Standard Deviation	
V_{max}	ICs	k_ms	(d)	(e)	(d)	(e)
yes	no	no	2.83	4.78	2.83	4.79
yes	yes	no	2.83	3.07	2.84	3.09
yes	yes	yes	2.87	3.08	2.89	3.10

(g)

	(d)				(e)			
Parameter	EV	RMSE	Lower	Upper	EV	RMSE	Lower	Upper
V_{max} (μM min^{-1})	173	1.03	175	171	177	5.04	167	187
PEP(0) (μM)	630	1600	-	-	664	8600	-	-
UGP(0) (μM)	200	0.189	200.4	199.6	205	0.244	204.5	205.5

Figure 5.5: Parameter estimation from time courses recorded using the inhibition free assay. Graphs (a), (b) and (c) each plot time courses obtained using the assays described above. Graphs (a) and (c) plot the release of ADP. Graph (a) shows the results of insufficient IDH in the reaction mixture. Graph (b) plots the oxidation of NADPH due to the action of MurB. Graphs (d) and (e) plot fits to the two experimental time courses plotted in Graph (c). Table (f) presents the RMSE and standard deviation for these fits. Table (g) contains the parameter estimates (EV) corresponding to estimation of V_{max} and the initial conditions; the associated RMSE and the bounds of a 90% confidence interval.

to the high concentration, relative to its k_m, of PEP present, see Section 5.4.1. The RMS errors associated with the remaining parameters are three orders of magnitude smaller than the estimated values; as such these parameters are well determined. The initial concentration of UGP corresponds well to the expected value. The estimated V_{max} corresponds to a k_{cat} value of 28.8s^{-1}, much higher than that previously estimated, Table 5.1.

(a)

Parameter	Value	RMSE
V_{max} (min^{-1})	0.127	0.00777
$k_{m,PEP}$ (μM)	0.284	0.165
$k_{m,UGP}$ (μM)	248	36.3
$k_{s,UGP}$ (μM)	808	601

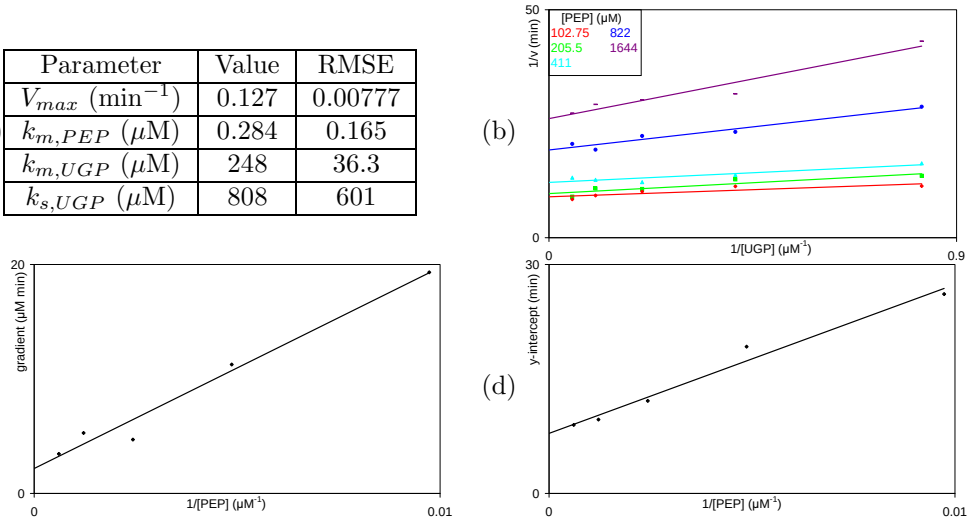


Figure 5.6: Characterisation of MurA using secondary plots. Table (a) summarises the kinetic constants obtained for MurA from Graphs (c)-(d). Graph (b) is the primary plot: 1/v (min) against 1/[UGP] (μM^{-1}) for a range of PEP concentrations as indicated in the key. Graphs (c) and (d) plot the gradient and y-intercept respectively of the lines in Graph (b) against 1/[PEP] (μM^{-1}).

A series of assays were undertaken using the phosphate release assay under the conditions described in Section 5.3 for initial PEP and UGP concentrations 1-20μM and 100-1700μM respectively. The results were analyzed using secondary plots as described in *Enzymes*, [144], to determine actual kinetic parameters for MurA, Figure 5.6. This analysis suggests that UGP is the first substrate to bind since this is the assumption that yields kinetic parameters closest to those determined previously, Table 5.1. The resulting k_{cat} is 6.79s^{-1} and corresponds well to those parameters, but not to the value determined by parameter estimation.

The two assays show a marked difference in initial rate under identical conditions, Figure 5.7. This suggests that the difference between the k_{cat}s obtained from the two assays are caused

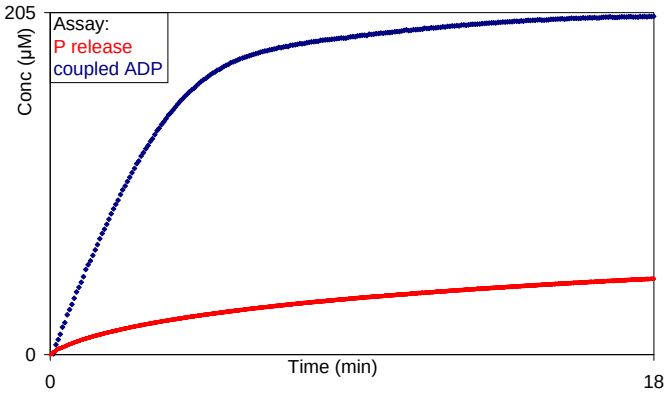


Figure 5.7: Comparison of MurA progress curves using two assays: the phosphate release assay (P release) and the coupled ADP release assay (coupled ADP) as indicated in the key.

by UDPPEE inhibition of MurA in the phosphate release assay. This could be confirmed by undertaking a kinetic characterisation using the alternative assay developed. Deconvolution of the apparent parameters obtained is not feasible given the size of the discrepancy.

Returning to the differences between the results obtained here and those of Du *et al.* note that the technique used by Du *et al* is a stopped rather than continuous assay. Such a technique typically uses measurements taken after a significant concentration of reaction product has accumulated. Consequently the results of this approach are likely subject to greater product inhibition than those obtained using initial rates. Thus just as the phosphate release assay produces lower k_{cat}s than the coupled ADP release assay, the stopped assays produce lower k_{cat}s than continuous assays where initial rates can be obtained.

The complete characterisation undertaken in this section estimates a k_m with respect to UGP of 248μM, which is a somewhat better match to the value obtained by Du *et al.* This improvement appears to arise due to separately estimating the dissociation constant even though the value of this constant is poorly determined. The relatively good match between the estimated values of the k_m with respect to PEP is lost in the complete characterisation, the estimates obtained differ by an order of magnitude. Further kinetic characterisation of MurA is required to determine this parameter, and the dissociation constant, more precisely.

5.4 MurB characterisation

MurB was characterised using the direct assay for MurB activity described in Section 2.4.4. Reaction mixtures were prepared as described and assays were monitored individually. Reactions were initiated with the unvaried substrate, NADPH or UDPPEE. The final concentration of MurB in the assays was 96nM. Three series of assays were undertaken for each substrate and the resulting data were fitted and combined to produce a final estimate of the apparent kinetic constants, see Table 5.4. For determinations with respect to NADPH, twenty NADPH concentrations, ranging between 25-500μM, were used. For determinations with respect to UDPPEE, twenty UDPPEE concentrations, ranging between 10-600μM, were used. The concentration of the unvaried substrate was 200μM in each case. It was noted that a significant concentration of potassium chloride, 100mM, was required to permit maximum activity of MurB as observed by Sylvester *et al.*, [86]. Parameters were initially estimated for Equation (3.38).

		For NADPH			For UDPPEE		
Series	Parameter Name	Values	Lower	Upper	Values	Lower	Upper
1	V_{max} (AU min^{-1})	0.1039	0.0987	0.1095	0.1137	0.1037	0.1246
1	k_m (μM)	29.2	22.3	37.3	66.3	46.2	95.1
2	V_{max} (AU min^{-1})	0.0916	0.0836	0.1003	0.1136	0.1078	0.1198
2	k_m (μM)	14.5	5.7	36.9	27.5	20.8	36.6
3	V_{max} (AU min^{-1})	0.0859	0.0798	0.0924	0.1193	0.1121	0.1270
3	k_m (μM)	16.8	8.73	32.4	31.7	22.7	44.4
Combined	V_{max} (AU min^{-1})	0.0920	0.0863	0.0982	0.1139	0.1084	0.1198
Combined	k_m (μM)	20.9	12.9	33.8	32.8	25.6	42.1

Table 5.4: Kinetic parameters for an uninhibited model of MurB for each series of assays undertaken. Each series used twenty concentrations of the varied substrate ranging between 25 and 500μM for NADPH and 10 and 600μM for UDPPEE. The unvaried substrate concentrations were 200μM NADPH and 200μM UDPPEE. The upper and lower bounds determine a 90% confidence interval.

Estimates of parameters are relatively consistent across series of assays although there is greater deviation in the k_m of UDPPEE than would be preferred. With this proviso the estimated pa-

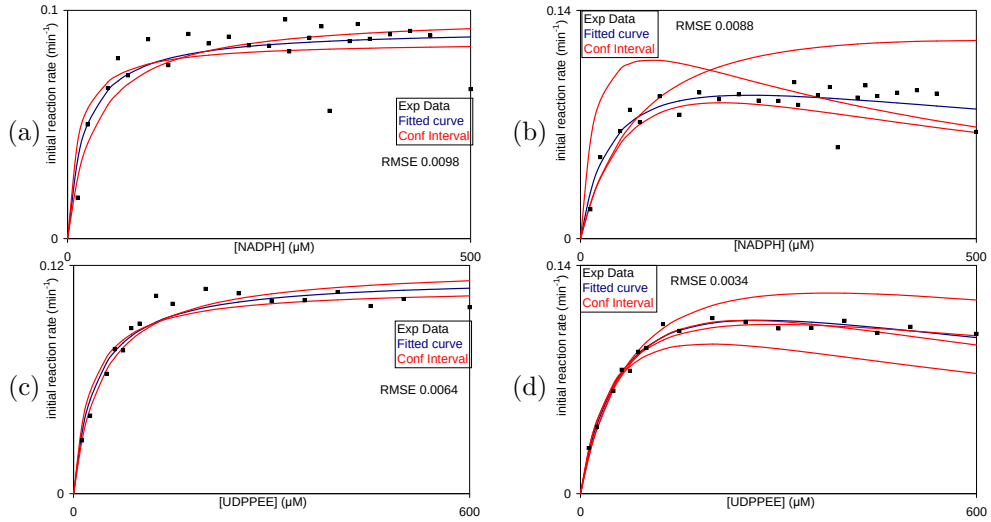


Figure 5.8: Plots of the combined MurB assay data. Graphs (a) and (b) correspond to varying NADPH; (c) and (d) correspond to varying UDPPEE. Graphs (a) and (c) assume standard Michaelis-Menten kinetics; (b) and (d) assume substrate inhibition. Experimental data (Exp Data), the fitted curve, and a 90% confidence interval (Conf Interval) are plotted.

rameters are also consistent with those estimated from the combined data which produce tighter confidence intervals. Plots of the data indicate that both substrates may be substrate inhibitors as expected given [86], see Figure 5.8. As such parameters were estimated for Equation (3.39) which describes this type of inhibition, see Table 5.5. The confidence intervals on these estimates are generally broader than those obtained fitting to Equation (4.62) especially for estimates of k_i. The estimated values of k_i are significantly higher than the range of concentrations over which data were collected; in order to obtain better estimates of k_i a wider range would be required. However the overall fit to the data is improved, the sum of the squared residuals is reduced by a factor of 3.5. This suggests that a substrate inhibition model is more appropriate for this enzyme.

The characterisations undertaken here were carried out at a lower pH than those by Sylvester *et al.*, [86]. The observed k_{cat} is consequentally significantly lower than was determined by that group. However the k_m values are also lower, suggesting that while a higher pH may yield a faster overall reaction rate it may also hamper binding of the substrates.

Series	Parameter Name	For NADPH			For UDPPEE		
		Values	Lower	Upper	Values	Lower	Upper
1	V_{max} (AU min^{-1})	0.1125	0.0952	0.1329	0.1801	0.1090	0.2977
1	k_m (μM)	35.1	22.7	54.0	145	65.0	324
1	k_i (μM)	4110	529	32000	775	209	2880
2	V_{max} (AU min^{-1})	0.1394	0.0941	0.2064	0.1539	0.1330	0.1781
2	k_m (μM)	47.4	18.1	124	55.3	40.1	76.2
2	k_i (μM)	715	237	2150	1200	705	2050
3	V_{max} (AU min^{-1})	0.0903	0.0839	0.0972	0.1687	0.1398	0.2036
3	k_m (μM)	20.0	11.1	36.1	67.7	45.6	101
3	k_i (μM)	NFP	-	-	1050	571	1940
Combined	V_{max} (AU min^{-1})	0.1219	0.0926	0.1604	0.1601	0.1418	0.1808
Combined	k_m (μM)	41.2	20.7	82.0	68.7	53.5	88.2
Combined	k_i (μM)	1110	384	3200	1070	718	1600

Table 5.5: Kinetic parameters for a substrate inhibited model of MurB for each series of assays undertaken. Experiments are as described previously, Table 5.4. NFP indicates that it was impossible to obtain a statistically significant estimate of the parameter in question.

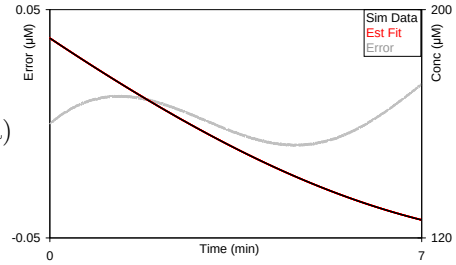
5.4.1 Direct parameter estimation from MurB progress curves

The structural identifiability analysis of the ping-pong model conducted previously, Section 4.3.1, suggests that estimation of kinetic parameters from time course concentration measurements may be feasible. Assuming that the MurB mechanism follows a ping-pong model, it is possible to test this result practically. The results obtained in the previous section are compared to those obtained using parameter estimation software.

It is possible that parameters cannot be estimated from MurB experimental data due to the effects of noise, insufficient parameter sensitivity or inappropriate model choice. To determine whether sensitivity would be a significant factor, parameters were estimated from data simulated in COPASI using the parameters above for an uninhibited model, Figure 5.9. The initial conditions, NADPH 190μM and UDPPEE 75μM, were chosen to correspond to an experiment which had been undertaken. The overall fit was good, RMS error and standard deviation 0.0081,

however the estimate for $k_{m,1}$, the k_m with respect to NADPH, was poor when compared to the actual value. The Fisher information matrix, $\mathbf{\Xi}(\mathbf{p})$ in Figure 5.9, shows that the sensitivity of this parameter is significantly lower than those of the other parameters. An investigation of the sensitivities of the parameters with respect to the controllable experimental conditions was undertaken.

(a)



(b)

	AV	EV	RMSE
V_{max}	19.2	18.3966	0.010397
$k_{m,1}$	20.9	11.1628	0.126894
$k_{m,2}$	32.8	31.9671	0.010795

$$\rho(\mathbf{p}) = \begin{pmatrix} 1 & 0.998 & 0.981 \\ 0.998 & 1 & 0.967 \\ 0.981 & 0.967 & 1 \end{pmatrix}, \qquad \mathbf{\Xi}(\mathbf{p}) = \begin{pmatrix} 14400 & -910 & -3260 \\ -910 & 57.6 & 205 \\ -3260 & 205 & 746 \end{pmatrix}.$$

Figure 5.9: Results of parameter estimation for MurB using simulated data and the evolutionary programming algorithm in COPASI. The fit is plotted in Graph (a). Table (b) compares the actual and estimated values (AV and EV) of each of the parameters. The RMSE of each estimate is also given. The correlation and Fisher information matrices are denoted $\rho(\mathbf{p})$ and $\mathbf{\Xi}(\mathbf{p})$ respectively. $k_{m,1}$ and $k_{m,2}$ refer to k_ms with respect to NADPH and UDPPEE respectively.

The sensitivity of the product concentration to each kinetic parameter was expressed as a differential equation thus:

$$\begin{aligned}
\frac{\partial}{\partial V_{max}}\frac{dP}{dt} &= \frac{dSEN_{V_{max}}}{dt} = \frac{S_1S_2}{(S_1S_2 + k_{m,1}S_2 + k_{m,2}S_1)}, \\
\frac{\partial}{\partial k_{m,1}}\frac{dP}{dt} &= \frac{dSEN_{k_{m,1}}}{dt} = \frac{S_1S_2^2V_{max}}{(S_1S_2 + k_{m,1}S_2 + k_{m,2}S_1)}, \\
\frac{\partial}{\partial k_{m,s}}\frac{dP}{dt} &= \frac{dSEN_{k_{m,2}}}{dt} = \frac{S_1^2S_2V_{max}}{(S_1S_2 + k_{m,1}S_2 + k_{m,2}S_1)}.
\end{aligned}$$

The terms $SEN_{V_{max}}$, $SEN_{k_{m,1}}$ and $SEN_{k_{m,2}}$ denote the sensitivity of product concentration to the parameters V_{max}, $k_{m,1}$ and $k_{m,2}$ respectively. The equations above were numerically solved with the system equations for a range of initial conditions, 10 and 200 μM for each substrate, Figure 5.10. The expressions above are positive definite thus sensitivity is strictly

increasing with respect to time. This is intuitive; the effect of a change in the parameters increases with time. As such the maximal sensitivity to all parameters is reached when the reaction stops due to complete consumption of one or both substrates. The effects of varying the initial concentrations of the two substrates is shown below, Figure 5.10 (c)-(h). Sensitivity to V_{max} is maximised when both initial substrate concentrations are high; it is limited by the lower substrate concentration, Figure 5.10 (c)-(d). The sensitivity of product concentration to the k_m of a particular substrate is maximised when the concentration of the other substrate is high, Figure 5.10 (f)-(g). For these parameters there is a critical concentration of the associated substrate, which is dependent on the concentration of the other substrate, at which its sensitivity is maximised, Figure 5.10 (e) and (h). The relationship between the substrate concentrations was estimated by plotting the concentration of the associated substrate for which the maximal sensitivity was observed against the concentration of the other substrate, Figure 5.10 (i)-(j).

$$\left.\begin{array}{ll} \text{Sensitivity } k_{m,1}: & \log([NADPH]_{max}) = 1.89 + 0.0162[UDPPEE] \\ \text{Sensitivity } k_{m,2}: & 0.0162[UDPPEE]_{max} = \log([NADPH]) - 2.18 \end{array}\right\} \tag{5.2}$$

The choice of an exponential relationship between the two substrates was based on the shapes of the curves, however note that the gradient of these two fits are very similar. This suggests that an appropriate relationship has been chosen.

This analysis confirms the earlier observation that the choice of initial concentrations used was not optimal and further suggests that ideally the measurement duration should be sufficient for the reaction to complete. Data were simulated from initial conditions informed by this relationship and parameters were estimated, results shown in Figures 5.11 and 5.12. Parameters were estimated using the Levenburg-Marquardt algorithm. This algorithm converged well for these data where it had not for the first, relatively insensitive, simulated data set. It is not possible to satisfy both of the expressions above, Equations 5.2. As such two experiments may

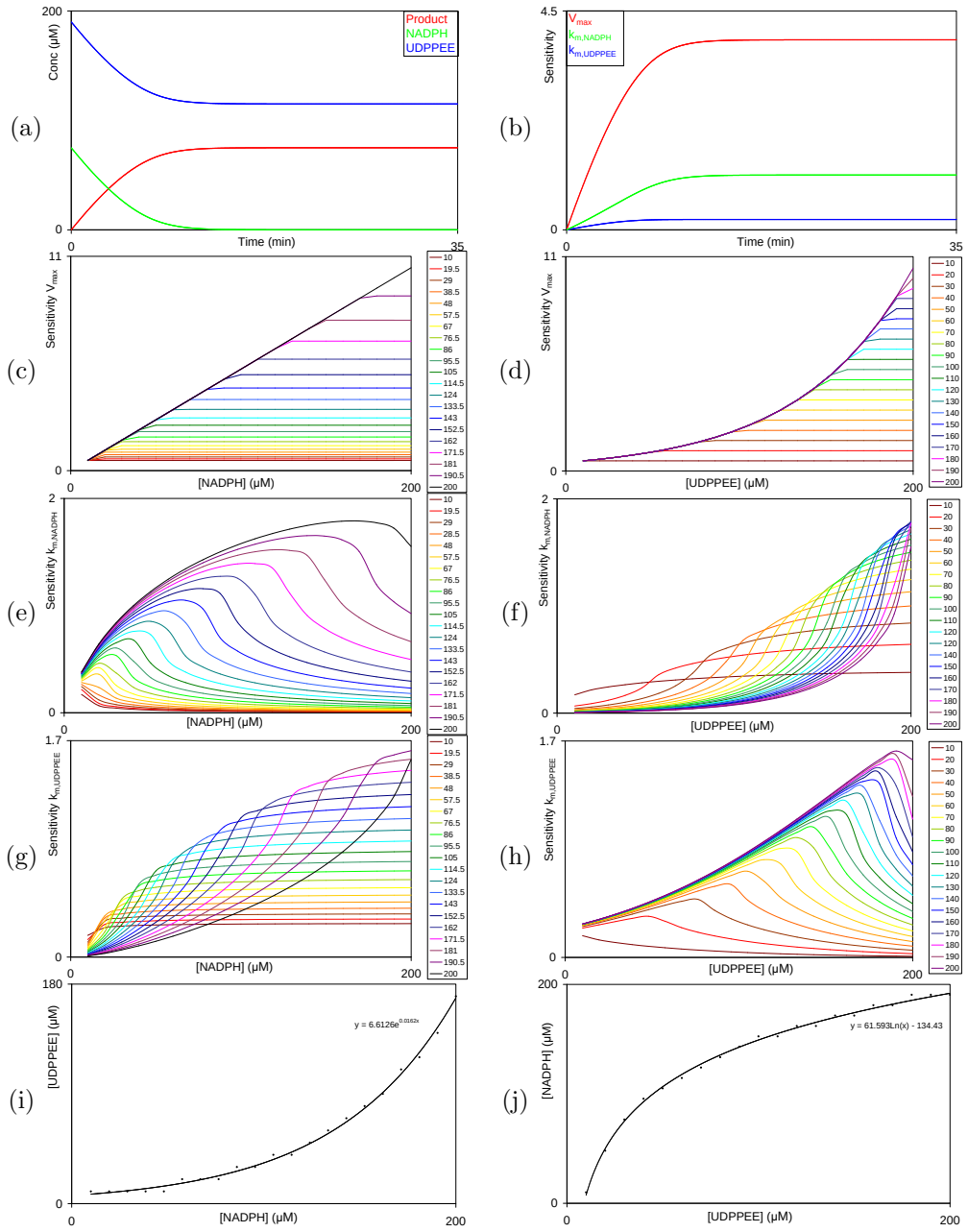
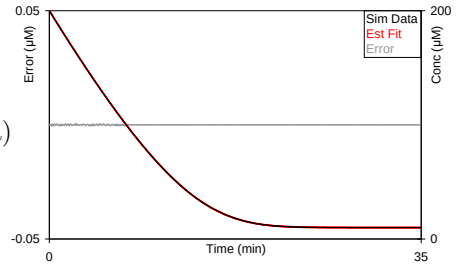


Figure 5.10: Simulation of sensitivities for the MurB reaction using the stated parameters in *COPASI*. Graphs (a) and (b) plot species concentration and sensitivity of the parameters over time for initial concentrations: 200μM NADPH, 80μM UDPPEE and 0.1μM MurB. Graphs (c) to (h) plot final sensitivity against varying initial concentration of the two parameters. The legend of each indicates concentration of the unvaried substrate: a range of 10-200μM, coloured coded from red to violet. Graph (i) plots the concentration of NADPH at which maximum sensitivity to $k_{m,NADPH}$ is observed against UDPPEE concentration. Graph (j) plots the concentration of UDPPEE at which maximum sensitivity to $k_{m,UDPPEE}$ is observed against NADPH concentration.

be required to achieve maximal sensitivity to the two k_m parameters.

(a)



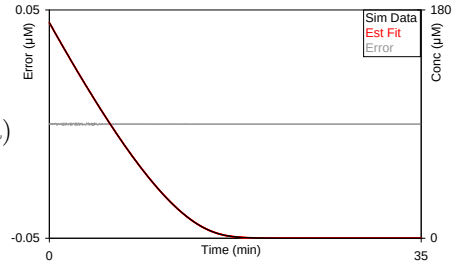
(b)

	AV	EV	RMSE
V_{max}	19.2	19.2	$9.54\text{x}10^{-6}$
$k_{m,1}$	20.9	20.9	$2.05\text{x}10^{-4}$
$k_{m,2}$	32.8	32.9	$1.18\text{x}10^{-4}$

$$\rho(\mathbf{p}) = \begin{pmatrix} 1 & 0.884 & -0.711 \\ 0.884 & 1 & -0.956 \\ -0.711 & -0.956 & 1 \end{pmatrix}, \quad \Xi(\mathbf{p}) = \begin{pmatrix} 17500 & -1940 & -2220 \\ -1940 & 220 & 254 \\ -2220 & 254 & 295 \end{pmatrix}.$$

Figure 5.11: Results of parameter estimation for MurB from simulated data using the Levenburg-Marquardt algorithm in COPASI. Initial conditions were 190 and 200 μM UDPPEE and NADPH respectively. Graphs, tables and matrices are as those in Figure 5.9. $k_{m,1}$ and $k_{m,2}$ refer to k_ms with respect to NADPH and UDPPEE respectively.

(a)



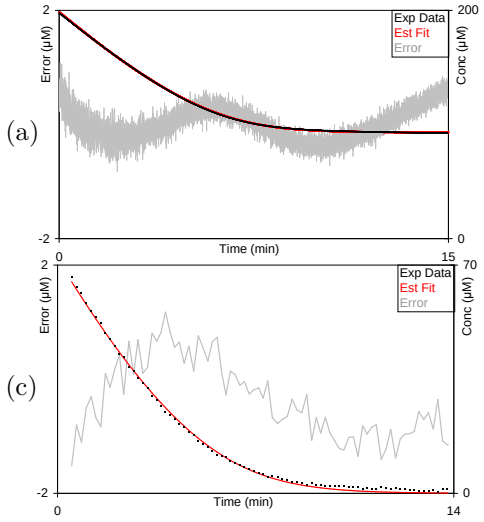
(b)

	AV	EV	RMSE
V_{max}	19.2	19.2	$7.37\text{x}10^{-5}$
$k_{m,1}$	20.9	20.9	$2.66\text{x}10^{-4}$
$k_{m,2}$	32.8	32.8	$6.24\text{x}10^{-4}$

$$\rho(\mathbf{p}) = \begin{pmatrix} 1 & -0.710 & 0.959 \\ -0.710 & 1 & -0.877 \\ 0.959 & -0.877 & 1 \end{pmatrix}, \quad \Xi(\mathbf{p}) = \begin{pmatrix} 13200 & -1930 & -1330 \\ -1930 & 297 & 199 \\ -1330 & 199 & 136 \end{pmatrix}.$$

Figure 5.12: Results of parameter estimation for MurB from simulated data using the Levenburg-Marquardt algorithm in COPASI. Initial conditions were 200 and 170 μM UDPPEE and NADPH respectively. Graphs, tables and matrices are as those in Figure 5.9. $k_{m,1}$ and $k_{m,2}$ refer to k_ms with respect to NADPH and UDPPEE respectively.

Kinetic parameters and initial conditions were estimated from experimental data, Figure 5.13. The estimated curves corresponded well to the experimental data. Results were somewhat worse than those obtained using simulated data: RMS error increasing from 0.283 to 0.757 and standard deviation from 0.283 to 0.766. The first data set allowed the estimation of all parameters except the k_m with respect to NADPH. This was expected given the preceding sensitivity analysis. The estimated values of the other kinetic parameters were fixed and the remaining parameter estimated from a second data set. This approach is somewhat suspect



(b)

Parameter	EV	RMSE	Lower	Upper
V_{max}	24.7	0.110	24.5	24.9
$k_{m,NADPH}$	0.1	0.870	-	-
$k_{m,UDPPEE}$	33.5	0.092	33.3	33.7
$[UDPPEE](0)$	105	0.012	104.98	105.02
$[NADPH](0)$	198	0.011	197.98	198.02

(d)

Parameter	EV	RMSE	Lower	Upper
$k_{m,NADPH}$	37.8	0.381	37.1	38.5
$[NADPH](0)$	71.9	0.199	71.5	72.3

Figure 5.13: Results of parameter estimation for MurB using experimental data and the Levenburg-Marquardt algorithm in COPASI. Parameters were initially estimated for the time course plotted in Graph (a). Initial conditions were approximately 100μM UDPPEE, 200μM NADPH and 0.1μM MurB. Both kinetic parameters and initial substrate concentrations were estimated (EV), Table (b). The undetermined parameter, $k_{m,NADPH}$ and the initial concentration of NADPH were then estimated from the second time course, Graph (c) and Table (d). Initial conditions were approximately 170μM UDPPEE, 80μM NADPH and 0.1μM MurB. The other kinetic parameters were fixed using the parameters estimated from the first data set.

given the high degree of correlation of the parameters observed, for example Figure 5.12(c). However given the sparse data set available this expedient proved necessary.

The estimated value of the k_m with respect to UDPPEE corresponds well to that estimated previously for an uninhibited model, Table 5.1. However those for k_{cat} (4.12s^{-1}) and the k_m with respect to NADPH, correspond well to those estimated for a substrate inhibited model. The parameter values obtained previously have fairly broad confidence intervals especially for the substrate inhibited models. It is possible that the discrepancies observed results from use of an uninhibited rather than substrate inhibited model. The discrepancy in k_{cat} could also arise from an inaccurate determination of the concentration or activity of MurB. Further experimental data and theoretical analysis are required to determine precise causes of these differences.

5.5 MurC characterisation

MurC was characterised using the phosphate release assay described in Section 2.4.3. Reaction mixtures were prepared as described and assays were monitored individually. Assays were initiated with L-Ala to avoid addition of inorganic phosphate to the reaction mixture at this step. Final concentration of MurC in the reaction mixture was 35nM. Three series of twenty assays were undertaken in which the concentration of UMN was varied. For the remaining substrates, ATP and L-Ala, only two series of fifteen assays each were used in order to preserve stocks of UMN. The parameters derived from these experiments are presented in Table 5.6. ATP, L-Ala and UMN concentrations ranged between 5-200μM, 40μM and 10mM, and 5-400μM respectively. The constant ATP, L-Ala and UMN concentrations were 1mM, 10mM and 600μM respectively. Graphs of these data are plotted in Figure 5.14. Kinetic parameters were numerically estimated for Equation (3.38) from these data.

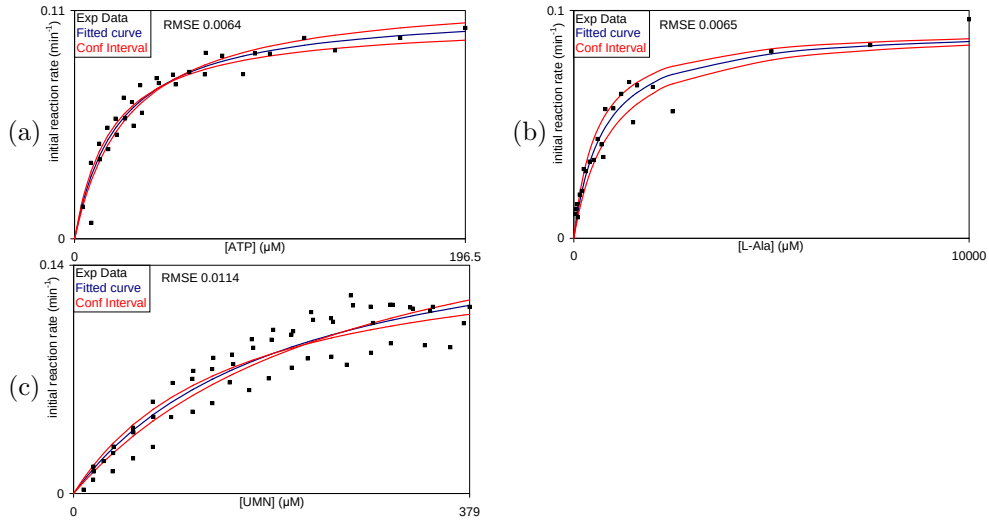


Figure 5.14: Plots of the combined MurC assay data. Graphs (a), (b) and (c) correspond to varying ATP, L-Ala and UMN respectively. Experimental data (Exp Data), the fitted curve, and a 90% confidence interval (Conf Interval) are plotted.

Parameters from individual series of assays are consistent with those derived from combined data. Confidence intervals for the k_m of L-Ala are wider than would be preferred. Some data sets suggest that UMN may be a substrate inhibitor, however this observation was not

Series	Parameter Name	For ATP			For L-Ala			For UMN		
		Values	Lower	Upper	Values	Lower	Upper	Values	Lower	Upper
1	V_{max} (AU min^{-1})	0.1096	0.1055	0.1138	0.1070	0.1004	0.1140	0.1137	0.1037	0.1246
1	k_m (μM)	17.6	15.6	19.8	697	593	819	66.3	46.2	95.1
2	V_{max} (AU min^{-1})	0.1149	0.1062	0.1243	0.0937	0.0880	0.0998	0.1136	0.1078	0.1198
2	k_m (μM)	29.7	23.5	37.4	999	835	1190	27.5	20.8	36.6
3	V_{max} (AU min^{-1})	-	-	-	-	-	-	0.1193	0.1121	0.1270
3	k_m (μM)	-	-	-	-	-	-	31.7	22.7	44.4
Combined	V_{max} (AU min^{-1})	0.1113	0.1047	0.1184	0.0925	0.0852	0.1005	0.1139	0.1084	0.1198
Combined	k_m (μM)	22.4	11.6	27.1	716	563	911	32.8	25.6	42.1

Table 5.6: Kinetic parameters obtained for MurC for each series of assays undertaken. For characterisations with respect to ATP and L-Ala fifteen measurements were made for each series: concentrations ranged between 5 and 200μM for ATP and 40μM and 10mM for L-Ala. For characterisations with respect to UMN twenty measurements were made for each series, concentrations ranged between 5-400μM. The unvaried substrate concentrations were 1mM ATP,10mM L-Ala and 600μM UMN. The upper and lower bounds determine a 90% confidence interval.

sustained by additional assays at higher concentrations of UMN than are shown here. As such an alternative explanation is required for this behaviour.

5.6 MurD characterisation

MurD was characterised using the phosphate release assay described in Section 2.4.3. Reaction mixtures were prepared as described and monitored in pairs using the Varian cell changer system. Either MurD or D-Glu was used to initiate the reaction thus avoiding the addition of inorganic phosphate at this stage. The final concentration of MurD in the reaction mixture was 4.25nM. Three series of sixteen assays were undertaken for each varied substrate: ATP, D-Glu and UDP-MurNAc-L-Ala (U1P). The parameters derived from these experiments are presented in Table 5.7. The concentrations of ATP, D-Glu and U1P ranged between 1-400μM, 50μM and 4mM, and 1-245μM respectively. The constant ATP, D-Glu and U1P concentrations were 1mM, 10mM and 245μM respectively. Kinetic parameters were estimated for Equation (3.38) for these data.

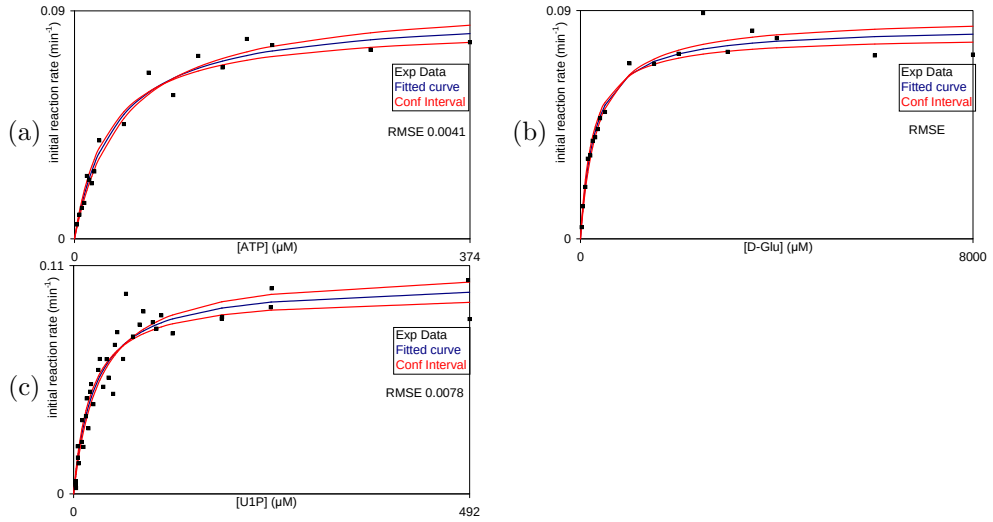


Figure 5.15: Plots of the combined MurD assay data. Graphs (a), (b) and (c) correspond to varying ATP, D-Glu and U1P respectively. Experimental data (Exp Data), the fitted curve, and a 90% confidence interval (Conf Interval) are plotted.

Estimated parameters are consistent with those estimated from combined data and the confi-

Series	Parameter Name	For ATP			For D-Glu			For U1P		
		Values	Lower	Upper	Values	Lower	Upper	Values	Lower	Upper
1	V_{max} (AU min^{-1})	0.0973	0.0855	0.1107	0.0985	0.0933	0.1040	0.1024	0.0942	0.1114
1	k_m (μM)	39.8	27.7	57.2	391	325	470	391	325	470
2	V_{max} (AU min^{-1})	0.0881	0.0810	0.0958	0.0754	0.0721	0.0789	0.0980	0.0914	0.1050
2	k_m (μM)	35.8	27.6	46.4	251	208	303	22.6	17.4	29.4
3	V_{max} (AU min^{-1})	0.0848	0.0796	0.0904	0.0769	0.0738	0.0801	0.1011	0.0926	0.1104
3	k_m (μM)	47.3	39.0	57.3	271	231	318	36.5	27.4	48.6
Combined	V_{max} (AU min^{-1})	0.0893	0.0841	0.0948	0.0836	0.0799	0.0875	0.1023	0.0963	0.1087
Combined	k_m (μM)	38.2	31.5	46.4	293	243	353	26.3	21.5	32.1

Table 5.7: Kinetic parameters obtained for MurD for each series of assays undertaken. Each series used sixteen measurements, concentrations ranged between 1 and 400μM for ATP, 50μM and 4mM for D-Glu and 1 and 245μM for U1P. The unvaried substrate concentrations were 1mM ATP, 10mM D-Glu and 245μM U1P. The upper and lower bounds determine a 90% confidence interval.

dence intervals are tight relative to the size of the parameters. Slight curvature was observed at the beginning of the some of the progress curves measured. While this can be an indicator of substrate inhibition, discussed in Section 5.7.1, it was not the case in this instance. This can be seen from the plots below Figure 5.15. Furthermore increasing the amount of coupling enzyme present eliminated the curvature in this case.

5.7 MurF characterisation

MurF was characterised using the phosphate release assay described in Section 2.4.3. Reaction mixtures were prepared as described and monitored individually. Reactions were initiated with D-Ala-D-Ala to avoid adding inorganic phosphate at this stage. The final concentration of MurF in the reaction mixture was 3.05nM. Three series of fifteen assays were completed for each varied substrate: ATP, D-Ala-D-Ala and UDP-MurNAc-tripeptide (U3P). The estimated parameters are presented in Table 5.8. The concentrations of ATP, D-Ala-D-Ala and U3P ranged between 5-900μM, 5-1000μM, and 5-400μM respectively. The constant ATP, D-Glu and U1P concentrations were 1mM, 10mM and 360μM respectively. Parameters were initially estimated for Equation (3.38).

Parameter estimates for individual series are consistent with those for the combined data. However plots of the data, Figure 5.16, strongly suggest that U3P is a substrate inhibitor. ATP may also be a weak substrate inhibitor. Parameter estimates for a substrate inhibition model, as in Section 5.4, are presented in Table 5.9. The fit is significantly improved for U3P, the root mean standard error decreasing by a factor of 2, and the confidence interval for k_i is tight. For ATP the fit improves only slightly and the confidence interval for the k_i estimate is very broad. Since the estimated k_i is beyond the range of the data it is likely this could be improved by measuring additional rates.

Series	Parameter Name	For ATP			For D-Ala-D-Ala			For U3P		
		Values	Lower	Upper	Values	Lower	Upper	Values	Lower	Upper
1	V_{max} (AU min^{-1})	0.0682	0.0633	0.0735	0.0693	0.0623	0.0772	0.0574	0.0487	0.0676
1	k_m (μM)	21.5	16.6	28.0	26.4	18.0	38.6	3.25	1.00	10.5
2	V_{max} (AU min^{-1})	0.0568	0.0545	0.0593	0.0812	0.0765	0.0862	0.0630	0.0541	0.0733
2	k_m (μM)	11.7	9.15	15.0	27.4	22.1	33.9	3.16	1.07	9.31
3	V_{max} (AU min^{-1})	0.0494	0.0463	0.0527	0.0840	0.0788	0.0895	0.0545	0.0486	0.0610
3	k_m (μM)	18.1	13.1	25.0	36.9	29.8	45.8	4.49	2.32	8.69
Combined	V_{max} (AU min^{-1})	0.0537	0.0507	0.0568	0.0778	0.0723	0.0838	0.0587	0.0535	0.0646
Combined	k_m (μM)	12.3	8.86	17.2	31.7	24.4	41.3	3.40	1.68	6.86

Table 5.8: Kinetic parameters for an uninhibited model of MurF for each series of assays undertaken. Each series used fifteen measurements, concentrations ranged between 5 and 900μM for ATP, 5 and 1000μM for D-Ala-D-Ala and 5 and 400μM for U3P. The unvaried substrate concentrations were 1mM ATP, 10mM D-Ala-D-Ala and 360μM U3P. The upper and lower bounds determine a 90% confidence interval.

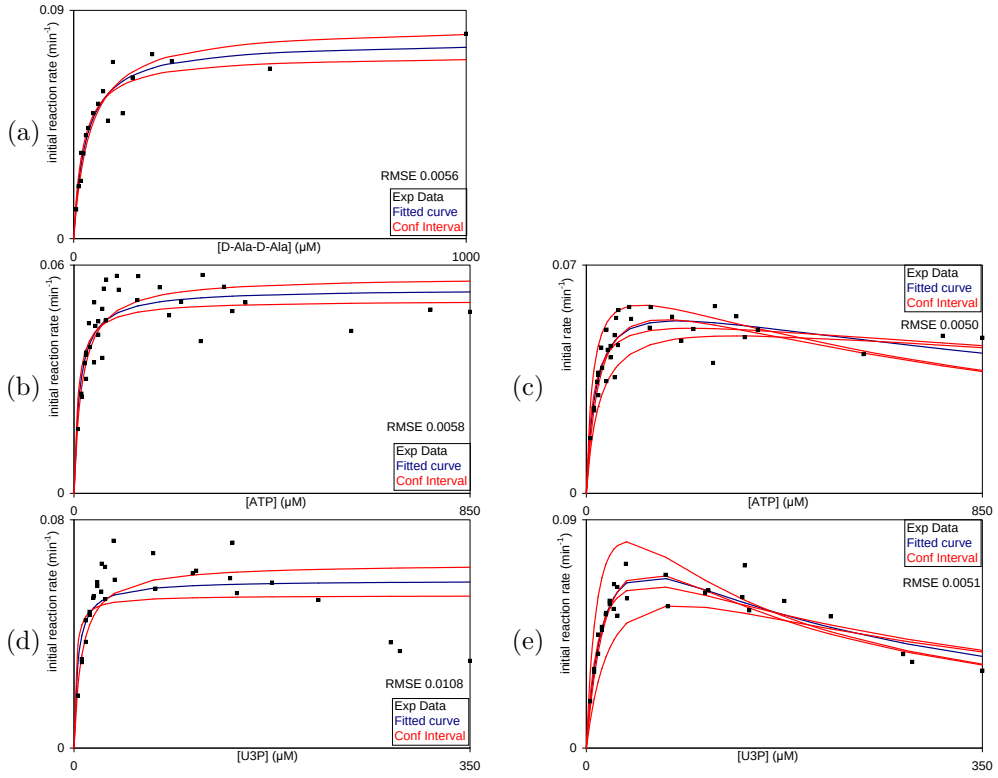


Figure 5.16: Plots of the combined MurF assay data. Graphs (a), (b) and (d) use standard Michaelis-Menten kinetics and correspond to varying D-Ala-D-Ala, ATP and U3P respectively. Graphs (c) and (e) use a substrate inhibition model and correspond to varying ATP and U3P respectively.

5.7.1 Substrate inhibition and progress curve curvature

While undertaking the MurF assays atypical progress curves were observed. For low concentrations of ATP and U3P the progress curve was typically shaped, the rate started at some maximum and then decreased as substrate was exhausted. For high concentrations of these substrates however, the rate increased over the time course measured, see Figure 5.17. It was demonstrated that this behaviour occurred even at significantly greater concentrations of the coupling enzyme PNP. Thus this behaviour was not caused by lack of coupling enzyme.

To investigate this behaviour a substrate inhibited model was simulated for a range of initial substrate concentrations, Figure 5.18, using *Maple* [158]. The reaction mechanism for a

Series	Parameter Name	For ATP			For U3P		
		Values	Lower	Upper	Values	Lower	Upper
1	V_{max} (AU min^{-1})	0.0975	0.0761	0.1250	0.1449	0.1131	0.1856
1	k_m (μM)	40.8	26.9	62.1	26.7	17.7	40.4
1	k_i (μM)	348	158	771	99.0	65.3	150
2	V_{max} (AU min^{-1})	0.0633	0.0568	0.0706	0.1257	0.1020	0.1550
2	k_m (μM)	16.1	11.5	22.6	17.1	11.3	25.8
2	k_i (μM)	2200	823	5870	161	106	246
3	V_{max} (AU min^{-1})	0.0548	0.0475	0.0631	0.0875	0.0723	0.1059
3	k_m (μM)	23.6	15.3	36.2	14.7	9.72	22.2
3	k_i (μM)	4600	1250	16900	246	150	403
Combined	V_{max} (AU min^{-1})	0.0642	0.0572	0.0721	0.1101	0.0911	0.1331
Combined	k_m (μM)	20.8	14.6	29.6	17.4	11.9	25.5
Combined	k_i (μM)	1810	949	3460	175	117	263

Table 5.9: Kinetic parameters for a substrate inhibited model of MurF for each series of assays undertaken. Experiments are as described in Table 5.8. The upper and lower bounds determine a 90% confidence interval.

substrate inhibited single substrate reaction is as follows:

$$\left.\begin{aligned} E + S &\underset{r_1}{\overset{k_1}{\rightleftharpoons}} ES \\ ES &\xrightarrow[k_2]{} E + P \\ ES + S &\underset{ri_1}{\overset{ki_1}{\rightleftharpoons}} ESS \end{aligned}\right\} \tag{5.3}$$

Thus a transient model describing substrate inhibition is given by the following differential

equations:

$$E^{(1)}(t) = -k_1 E(t)S(t) + (r_1 + k_2)ES(t) \tag{5.4a}$$

$$S^{(1)}(t) = -k_1 E(t)S(t) + r_1 ES(t) - ki_1 S(t)ES(t) + ri_1 ESS(t) \tag{5.4b}$$

$$ES^{(1)}(t) = (k_1 E(t) - ki_1 ES(t))S(t) - (r_1 + k_2)ES(t) + ri_1 ESS(t) \tag{5.4c}$$

$$ESS^{(1)}(t) = ki_1 S(t)ES(t) - ri_1 ESS(t) \tag{5.4d}$$

$$P^{(1)}(t) = k_2 ES(t) \tag{5.4e}$$

Typical experimental initial conditions, only substrate and enzyme concentrations greater than zero, were used in these simulations. The parameters, $\mathbf{p} = (k_1, r_1, ki_1, ri_1, k_2)$, of Equations (5.4a)-(5.4e) are related to the steady state parameters as follows:

$$V_{max} = k_2 E_0, \quad k_m = \frac{r_1 + k_2}{k_1}, \quad k_i = \frac{ri_1}{ki_1}.$$

Consequently transient parameters could be partially derived from the kinetic parameters already obtained, Table 5.1. It was necessary to choose a value for two parameters since only three relations exist between the transient and steady state parameters; as such the denominators k_1 and ki_1 were assigned arbitrary values. The relations above were then used to obtain the following parameters for use in the simulations, $\mathbf{p}_{sim,U3P} = (10, 106.5, 0.5, 87.5, 43.5)$ and $\mathbf{p}_{sim,ATP} = (10, 21.1, 0.5, 905, 186.9)$.

Simulations show that after a critical initial substrate concentration, rates of product release increase over a period of at least five minutes, Figure 5.18 (a) and (b); an observable time scale for the technique used. This trend matches that observed in the experimental data. Simulations predict that rates should become positive at concentrations around 80μM U3P and 200μM ATP again observed in the experimental data, Table 5.10. The simulations were also used to construct simulated experimental data, Figure 5.18 (c) - (f), which are comparable

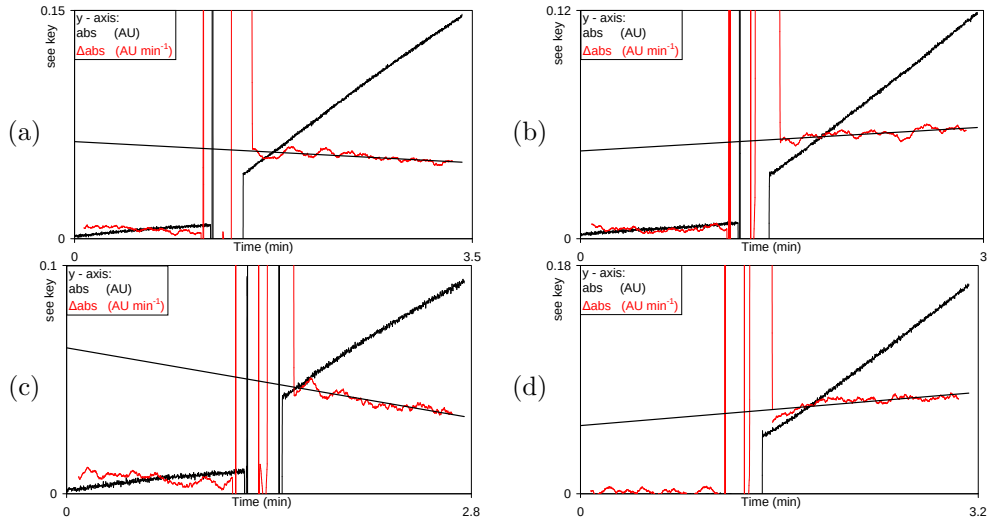


Figure 5.17: Plots of sample MurF progress curves and estimated gradients. Gradients (Δabs) were estimated using the LINEST function in Microsoft Excel for contiguous 100 point subsets of the raw data. A second linear regression was performed to a subset of the estimated gradients. This fit, to the section of the curve after the reaction was initiated, is shown by the black line. Graphs (a) and (b) are time courses from the third ATP characterisation, ATP concentrations 136 and 340μM respectively. Graphs (c) and (d) are time courses from the third U3P characterisation, U3P concentrations 10.32 and 108μM respectively.

[U3P] (μM)	rate change	[ATP] (μM)	rate change
3.44	decreasing	8.5	decreasing
6.88	decreasing	17	decreasing
10.32	decreasing	26	decreasing
13.76	decreasing	34	decreasing
17.2	decreasing	43	decreasing
20.69	increasing*	52	decreasing
24.08	decreasing	61	decreasing
27.52	decreasing	68	decreasing
36†	decreasing	136	decreasing
72†	increasing	204†	decreasing
108	increasing	272†	increasing
144	increasing	340	increasing
216	increasing	595	increasing
288	increasing	765	increasing
360	decreasing*	850	increasing

Table 5.10: Table showing the qualitative effect of increasing substrate concentration on the rate of reaction of MurF. Two unexpected results are indicated by *; the rates measured at these two concentrations were inconsistent with the rest of the data. Rates switch between decreasing and increasing between the concentrations indicated by † these ranges are consistent with those predicted in simulations.

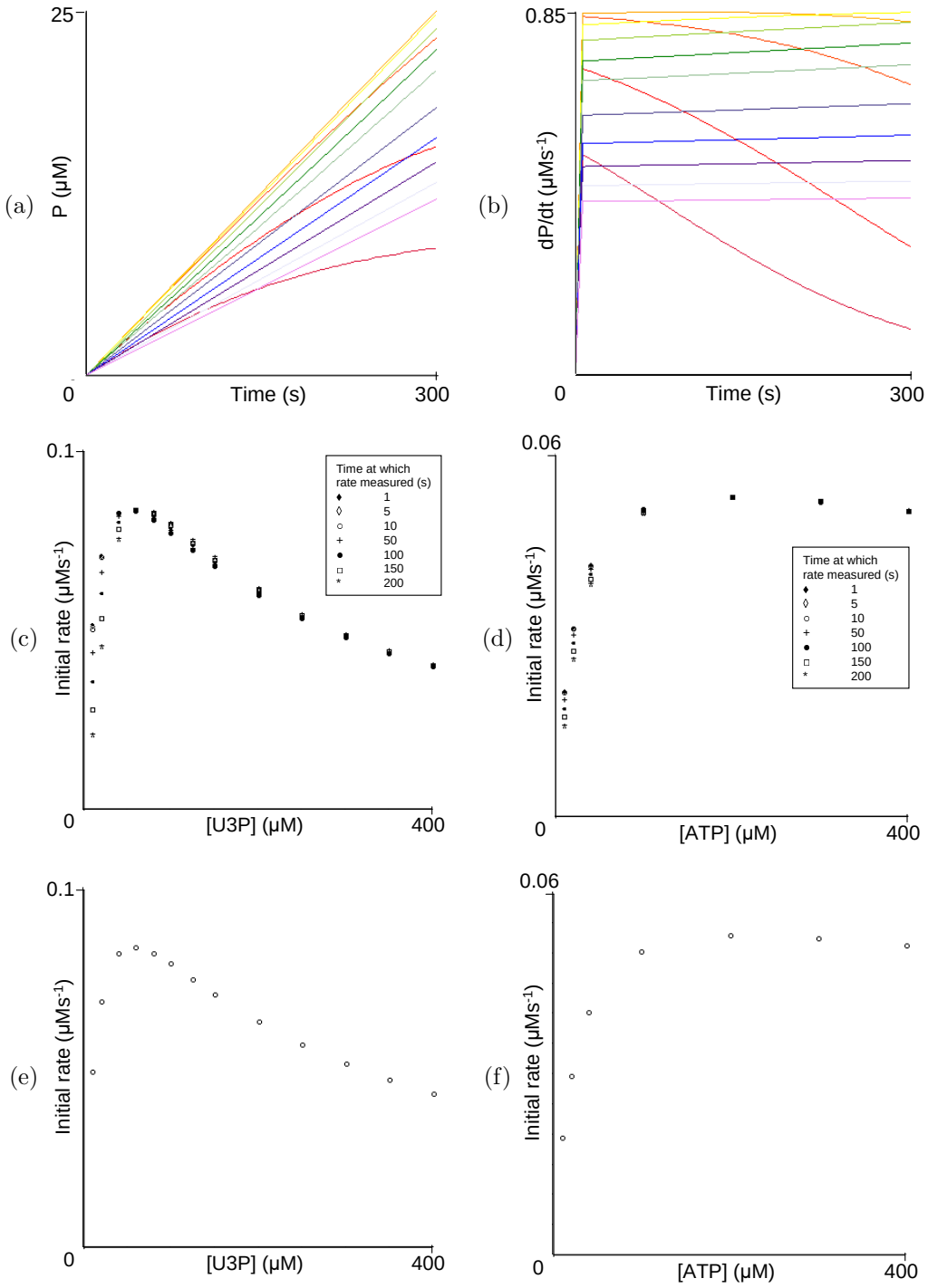


Figure 5.18: Simulations of Equations (5.4a)-(5.4e) using parameters deriving from MurF quasi-steady state parameters in *Maple* [158]. Graphs (a) and (b) plot the release of product and the rate of product release over time, given a range of initial U3P concentrations 10-400μM coloured red to violet. Graphs (c)-(f) plot simulated experimental data, plotting rate of product release against initial substrate concentration. Plots (c) and (d) use rates calculated at specific times indicated by the keys inset. Graphs (e) and (f) use rates calculated at the time when 10% of the initial substrate had been consumed.

with experimental data, Figure 5.16. It is clear from these plots that the shape of the plot is insensitive to the time when the measurement was made.

Note that similar behaviour was not observed for MurB catalysed reactions. However the maximal concentrations of UDPPEE and NADPH used were much lower for those characterisations as such the critical concentration may not have been reached. Furthermore, since NADPH based assays require a high starting concentration of the absorbing compound, the increased noise level may have obscured any small changes in rate. The lower extinction coefficient of NADPH relative to MTG may also have reduced the chance of observing this effect.

5.8 Summary

The kinetic characterisations of the *S. pneumoniae* MurA-F enzymes under consistent conditions provides a fundamental building block in the construction of a model to describe the behaviour of the overall pathway. Using these data it is possible to predict the concentrations of pathway intermediates as is shown in Chapter 7.

It was observed that MurA is strongly product inhibited by UDPPEE. This necessitated the development of an alternative assay of MurA activity to permit parameter estimation. A well determined estimate of the k_{cat} for MurA was obtained using data produced using this alternative coupled assay system. However this estimate differed significantly from those obtained using the phosphate release assay, as such no deconvolution of parameters could be attempted. It is hypothesised that this difference in k_{cat} arises from the effects of product inhibition. Further work is required to confirm this.

The kinetic parameters of MurB were estimated using two data sets as predicted by the sensitivity analysis undertaken. Fits to the experimental data used were good and confidence intervals for the parameters estimated were tight. However of the three parameters estimated only one was close to the value estimated for an uninhibited model. The other parameters were close to

the values expected for a substrate inhibited model. Further experimental work is needed to determine which is the most appropriate model. This may be informed by an indistinguishability analysis of the two models. Nonetheless the results obtained are encouraging evidence that the use of direct parameter estimation may be a viable approach for the characterisation of enzymes.

The effect of substrate inhibition on MurF time courses provides a new tool for the detection of substrate inhibition. Where substrate inhibition occurs increasing substrate concentration should lead to a change in geometry of the progress curve, converting from a parabolic to sigmoidal shape. It should be noted that this change is only observed in simulation when transient model equations are used. The quasi-steady state form of the model does not exhibit this behaviour. However the effect of this apparently transient phenomenon is observable on quasi-steady state time scales.

6. ANALYSIS OF AND PARAMETER ESTIMATION USING TRANSIENT MODELS OF ENZYME REACTIONS

6.1 Introduction

A more complete understanding of the kinetics of a given enzyme can be obtained by considering an unsimplified, transient, model of the expected mechanism. However the experiments previously described in Chapter 5 only collect data over a steady state time scale. In order to collect data over a pre-steady state time frame more advanced mixing techniques, such as stopped flow, are required [179]. Furthermore if reliable parameter estimates are required based on such data, structural identifiability analyses of the possible models should be undertaken prior to experimental design.

In this chapter structural identifiability analyses of the two substrate reaction mechanisms introduced in Section 4.2.1 are presented. A range of possible measurements and product release mechanism are considered in the course of these analyses. In the subsequent section pre-steady state time courses of reactions catalysed by Lactate dehydrogenase and MurB are used for parameter estimation.

6.2 Structural identifiability analysis

Transient models for the two substrate enzyme catalysed mechanisms have been introduced previously (Section 4.2.1). These models can be modified to take into account the effect of reversible product release stages. Structural identifiability analyses of the resulting models

are described in the following sections. For each model all single species measurements were analyzed as were two experimentally relevant sets of initial conditions. Pohjanpalo's Rank Test, the Taylor series approach and the input-output relationship approach were used in each case. Detailed analyses of the models are presented in the following sections. Ping-pong models are considered first. The simple ordered models are then considered in two groups: simultaneous and sequential release of products. For each model a Taylor series based analysis is presented which encompasses all mechanisms of product release considered. Any successful input-output relationship analyses are then presented.

6.2.1 *Ping-pong model*

The two substrate transient ping-pong model previously described in Section 4.2.1 is considered first. Three product release processes were considered; irreversible product release (a), reversible release of the second product (b), and fully reversible product release (c). A reaction scheme for Mechanism (c) is presented below:

$$\left.\begin{aligned} E + S_1 &\underset{r_1}{\overset{k_1}{\rightleftharpoons}} ES_1 \\ ES_1 &\underset{r_2}{\overset{k_2}{\rightleftharpoons}} E^* + Q \\ E^* + S_2 &\underset{r_3}{\overset{k_3}{\rightleftharpoons}} E^*S_2 \\ E^*S_2 &\underset{r_4}{\overset{k_4}{\rightleftharpoons}} E + P \end{aligned}\right\} \tag{6.1}$$

Mechanism (b) is obtained by setting $r_2 = 0$ in Equations (6.1). It will be shown later that for this mechanism can also be applied if the first product release stage is reversible, i.e. $r_4 = 0$, $r_2 \neq 0$ in Equations (6.1). Mechanism (a) is obtained by setting $r_2 = r_4 = 0$ in Equations (6.1). A Taylor series analysis of each of these mechanisms is presented below.

The state vector and initial conditions used were the same for all three models:

$$\left.\begin{aligned} &\mathbf{x}(t,\mathbf{p}) = \\ &= (E(t,\mathbf{p}), S_1(t,\mathbf{p}), ES_1(t,\mathbf{p}), Q(t,\mathbf{p}), E^*(t,\mathbf{p}), S_2(t,\mathbf{p}), E^*S_2(t,\mathbf{p}), P(t,\mathbf{p}))^T, \\ &\mathbf{x}(0,\mathbf{p}) = \mathbf{x}_0 = (E_0, S_{10}, 0, 0, 0, S_{20}, 0, 0)^T. \end{aligned}\right\} \tag{6.2}$$

Each mechanism gives rise to a different derivative of the state vector. For irreversible product release, Mechanism (a), the derivative was as previously stated, Equations (4.2) in Section 4.2.1. Reversible release of the second product, Mechanism (b), requires the following changes to Equations (4.2):

$$\frac{dE(t,\mathbf{p})}{dt} = -k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) + r_1ES_1(t,\mathbf{p}) + k_4E^*S_2(t,\mathbf{p}) - r_4E(t,\mathbf{p})P(t,\mathbf{p}), \tag{6.3a}$$

$$\frac{dE^*S_2(t,\mathbf{p})}{dt} = k_3E^*(t,\mathbf{p})S_2(t,\mathbf{p}) - (r_3+k_4)E^*S_2(t,\mathbf{p}) + r_4E(t,\mathbf{p})P(t,\mathbf{p}), \tag{6.3b}$$

$$\frac{dP(t,\mathbf{p})}{dt} = k_4E^*S_2(t,\mathbf{p}) - r_4E(t,\mathbf{p})P(t,\mathbf{p}). \tag{6.3c}$$

For Mechanism (c), fully reversible product release, the following additional changes to Mechanism (b) were made:

$$\frac{dES_1(t,\mathbf{p})}{dt} = k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) - (r_1+k_2)ES_1(t,\mathbf{p}) + r_2E(t,\mathbf{p})Q(t,\mathbf{p}), \tag{6.4a}$$

$$\frac{dQ(t,\mathbf{p})}{dt} = k_2ES_1(t,\mathbf{p}) - r_2E(t,\mathbf{p})Q(t,\mathbf{p}), \tag{6.4b}$$

$$\frac{dE^*(t,\mathbf{p})}{dt} = k_2ES_1(t,\mathbf{p}) - r_2E^*(t,\mathbf{p})Q(t,\mathbf{p}) - k_3E^*(t,\mathbf{p})S_2(t,\mathbf{p}) + r_3E^*S_2(t,\mathbf{p}). \tag{6.4c}$$

Thus each reaction mechanism requires a different unknown parameter vector, given below:

$$\mathbf{p}_a = (k_1, k_2, k_3, k_4, r_1, r_3), \tag{6.5}$$

$$\mathbf{p}_b = (k_1, k_2, k_3, k_4, r_1, r_3, r_4), \tag{6.6}$$

$$\mathbf{p}_c = (k_1, k_2, k_3, k_4, r_1, r_2, r_3, r_4), \tag{6.7}$$

the subscript indicating to which mechanism the vector corresponds. Eight output functions, corresponding to individual measurement of each reaction species, were analyzed:

$$\left.\begin{array}{l} y_1(t,\mathbf{p}) = E(t,\mathbf{p}),\ y_2(t,\mathbf{p}) = S_1(t,\mathbf{p}),\ y_3(t,\mathbf{p}) = ES_1(t,\mathbf{p}),\ y_4(t,\mathbf{p}) = Q(t,\mathbf{p}), \\ y_5(t,\mathbf{p}) = E^*(t,\mathbf{p}),\ y_6(t,\mathbf{p}) = S_2(t,\mathbf{p}),\ y_7(t,\mathbf{p}) = E^*S_2(t,\mathbf{p}),\ y_8(t,\mathbf{p}) = P(t,\mathbf{p}). \end{array}\right\} \tag{6.8}$$

In order to use Pohjanpalo's Rank Test (PRT) at least as many non-zero Taylor series coefficients as there are unknown parameters must be calculated (Section 3.2.2). For Mechanisms (a) and (b) this was possible for all output functions. For Mechanism (c) eight non-zero Taylor series coefficients were calculated for y_1 to y_3; however this proved computationally intractable for output functions y_4 to y_8, as such the PRT could not be applied for these five observations. For each of the remaining combinations of mechanism and output function the rank of the Jacobian matrix of the Taylor series coefficients with respect to the unknown parameters was calculated. For Mechanism (b) the Jacobian matrix for the expansion of y_8 proved computationally intractable. In each of the remaining cases the rank was equal to the number of unknown parameters present. As such all observations considered for Mechanism (a) are at least locally identifiable. For Mechanism (b) models arising for all observations, except that of P, are at least locally identifiable. Mechanism (c) is at least locally identifiable if concentration of enzyme, first substrate, or the complex of these species is measured.

A full Taylor series analysis for output functions y_5 to y_8 proved computationally intractable for

Mechanisms (a) and (b); moreover it was not possible to show that any individual parameters were globally identifiable for these models. For Mechanism (c) a full Taylor series analysis was computationally intractable for all output functions. However for output functions y_1 to y_4 it was possible to show that k_1, k_2 and r_1 are globally identifiable. The first four Taylor series coefficients used in this analysis were the same as those given below, Equations (6.9)-(6.12).

For observation of species in the first half of the reaction y_1 to y_4, full Taylor series analyses were completed for Mechanisms (a) and (b). The first four Taylor series coefficients for the expansions of these output functions at t=0 are:

$$y_1(0) = E_0,\ y_2(0) = S_{10},\ y_3(0) = 0,\ y_4(0) = 0, \tag{6.9}$$

$$y_1^{(1)}(0) = y_2^{(1)}(0) = -y_3^{(1)}(0) = -k_1 E_0 S_{10},\ y_4^{(1)}(0) = 0, \tag{6.10}$$

$$\left.\begin{aligned} &y_1^{(2)}(0) = y_2^{(2)}(0) = -y_3^{(2)}(0) = k_1 E_0 S_{10}(k_1(E_0 + S_{10}) + r_1), \\ &\mathbf{y_4}^{(2)}(0) = k_1 k_2 E_0 S_{10}, \end{aligned}\right\} \tag{6.11}$$

$$\left.\begin{aligned} &y_1^{(3)}(0) = y_2^{(3)}(0) = -y_3^{(3)}(0) + k_1 k_2 E_0 S_{10}(k_1(E_0 + S_{10}) + r_1 + k_2) = \\ &\quad = -k_1 E_0 S_{10}((k_1(E_0 + S_{10}) + r_1)^2 + 2k_1^2 E_0 S_{10} + k_2 r_1), \\ &y_4^{(3)}(0) = -k_1 k_2 E_0 S_{10}(k_1(E_0 + S_{10}) + k_2 + r_1), \end{aligned}\right\} \tag{6.12}$$

for all three mechanisms considered. The alternative parameter vectors used are given by:

$$\overline{\mathbf{p}}_a = (\overline{k_1}, \overline{k_2}, \overline{k_3}, \overline{k_4}, \overline{r_1}, \overline{r_3})^T, \tag{6.13}$$

$$\overline{\mathbf{p}}_b = (\overline{k_1}, \overline{k_2}, \overline{k_3}, \overline{k_4}, \overline{r_1}, \overline{r_3})^T, \tag{6.14}$$

$$\overline{\mathbf{p}}_c = (\overline{k_1}, \overline{k_2}, \overline{k_3}, \overline{k_4}, \overline{r_1}, \overline{r_3})^T. \tag{6.15}$$

These vectors were used to create simultaneous equations in the unknown parameters from the Taylor series coefficients as described in Section 3.2.1. These equations were then solved in

Maple for the alternative parameters [158]. For Mechanisms (a) and (b) it was shown that $\overline{\mathbf{p}} = \mathbf{p}$. Hence measurement of the concentration of these species results in structurally globally identifiable (SGI) models for these two mechanisms.

The two halves of Mechanism (a) are equivalent up to relabelling as follows:

$$\left.\begin{array}{c} E \leftrightarrow E^*,\ S_1 \leftrightarrow S_2,\ ES_1 \leftrightarrow E^*S_2,\ Q \leftrightarrow P, \\ \\ k_1 \leftrightarrow k_3,\ r_1 \leftrightarrow r_3,\ k_2 \leftrightarrow k_4. \end{array}\right\} \tag{6.16}$$

Consequently it is possible to choose an alternative set of initial conditions, corresponding to a different experimental setup, for which the parameter vector is SGI for y_5 to y_8. Specifically the initial concentrations of E^* and E must be swapped as follows:

$$\mathbf{x}_0 = (0, S_{10}, 0, 0, E_0, S_{20}, 0, 0)^T. \tag{6.17}$$

This may be experimentally viable if the enzyme can be isolated in the required form.

Mechanism (c) also consists of two halves equivalent up to a relabelling described by Equations (6.16) and $r_2 \leftrightarrow r_4$. Thus using the alternative set of initial conditions above, Equations (6.17), k_3, k_4 and r_1 are globally identifiable for output functions y_5 to y_8.

The alternative initial conditions, Equation (6.17), were also considered for Mechanism (b), which is changed by any non-trivial relabelling. For all output functions at least seven non-zero Taylor series coefficients were found. In each case the Jacobian matrix was found to have a rank of seven, equal to the number of unknown parameters, as such all observations produce models which are at least locally identifiable.

For the output functions y_1 to y_4 and y_7 a full Taylor series analysis proved computationally intractable. However for y_7 it was possible to show that k_3, k_4 and r_3 are SGI.

A full Taylor series analysis was completed for each of the models corresponding to measurement

Ping-pong model	Reversible product release stages			
Measurement	None	First	Second	Both
E	SGI	SGI	SGI	ASLI, $\{k_1, r_1, k_2\}$
S_1	SGI	SGI	SGI	ASLI, $\{k_1, r_1, k_2\}$
ES_1	SGI	ASLI, $\{k_1, r_1, k_2\}$	SGI	ASLI, $\{k_1, r_1, k_2\}$
Q	SGI‡	SGI	SGI‡	CI, $\{k_1, r_1, k_2\}$
E^*	SGI†	SGI†	SGI†	ASLI†, $\{k_3, r_3, k_4\}^\dagger$
S_2	SGI†	SGI†	SGI†	ASLI†, $\{k_3, r_3, k_4\}^\dagger$
E^*S_2	SGI†	SGI†	ASLI, $\{k_3, r_3, k_4\}^\dagger$	ASLI†, $\{k_3, r_3, k_4\}^\dagger$
P	SGI‡	SGI‡	SGI†	CI, $\{k_3, r_3, k_4\}^\dagger$

Table 6.1: Summary of the structural identifiability analysis results for the two substrate transient ping-pong models. Results are either: SGI, structurally globally identifiable; ASLI, at least structurally locally identifiable; or CI, analysis was computationally intractable. In addition a subset of globally identifiable parameters are stated in some cases. Results marked by a † were obtained for the alternative initial conditions, Equation (6.17). Results marked by a ‡ can be obtained using the input-output relationship approach, and require only that enzyme and substrate concentrations be initially non-zero.

of E^*, S_2 and P. The first four Taylor series coefficients of these expansions are:

$$y_5(0) = E_0,\ y_6(0) = S_{20},\ y_8(0) = 0, \tag{6.18}$$

$$y_5^{(1)}(0) = y_6^{(1)}(0) = -k_3E_0S_{20},\ y_8^{(2)}(0) = 0, \tag{6.19}$$

$$y_5^{(2)}(0) = y_6^{(2)}(0) = k_3E_0S_{20}(k_3(E_0 + S_{20}) + r_3),\ y_8^{(2)}(0) = k_3k_4E_0S_{20}, \tag{6.20}$$

$$\left.\begin{aligned} y_5^{(3)}(0) = y_6^{(3)}(0) &= -k_3E_0S_{20}((k_3(E_0 + S_{20}) + r_3)^2 + 2k_3E_0S_{20} + k_4r_3), \\ y_8^{(3)}(0) &= -k_3k_4E_0S_{20}(k_3(E_0 + S_{20}) + (r_3 + k_4)). \end{aligned}\right\} \tag{6.21}$$

The Taylor series coefficients were analyzed using the alternative parameter vector, Equation (6.14), as previously described. For each expansion a single solution for the alternative parameter vector was obtained, $\overline{\mathbf{p}} = \mathbf{p}$. Thus the models arising from these three observations are SGI.

Under the relabelling previously defined, Equation (6.16), an alternative model is created de-

scribing reversible release of Q and irreversible release of P, see below.

$$\left.\begin{array}{rcl} E+S_1 & \underset{r_1}{\overset{k_1}{\rightleftharpoons}} & ES_1 \\ ES_1 & \xrightarrow{k_2} & E^*+Q \\ E^*+S_2 & \underset{r_3}{\overset{k_3}{\rightleftharpoons}} & E^*S_2 \\ E^*S_2 & \underset{r_4}{\overset{k_4}{\rightleftharpoons}} & E+P \end{array}\right\} \overset{\text{Relabelling}}{\rightleftharpoons} \left\{\begin{array}{rcl} E^*+S_2 & \underset{r_3}{\overset{k_3}{\rightleftharpoons}} & E^*S_2 \\ E^*S_2 & \xrightarrow{k_4} & E+P \\ E+S_1 & \underset{r_1}{\overset{k_1}{\rightleftharpoons}} & ES_1 \\ ES_1 & \underset{r_2}{\overset{k_2}{\rightleftharpoons}} & E^*+Q \end{array}\right\} \tag{6.22}$$

For consistency of notation r_4 is relabelled r_2. Thus the results obtained above apply to this model as follows. For the initial conditions, Equation (6.2), the models arising from measurement of E, S_1 and Q are SGI. For the initial conditions, Equations (6.17), the models arising from measurement of E^*, S_2, E^*S_2 and P are SGI. All other measurements can be shown to be at least locally identifiable using one of the two sets of initial conditions.

The results of this Taylor series analysis are summarised in Table 6.1. These models were also analyzed using the input-output relationship approach. The results of such an analysis, being less dependent on initial conditions, can be applied more flexibly than those obtained using the Taylor series approach. In addition use of both approaches allows comparisons to be drawn regarding their relative effectiveness for these and similar models. The analysis of Mechanism (c) proved computationally intractable for all output functions considered. The results for Mechanisms (a) and (b) are presented in the following sections.

6.2.1.1 Differential algebra analysis of Mechanism (a) for the ping-pong model

The following expressions are the generators of a differential ideal equivalent to Mechanism (a), explicit dependence on $\mathbf{p}$ is omitted for notational convenience:

$$\frac{dE(t)}{dt} + k_1E(t)S_1(t) - r_1ES_1(t) - k_4E^*S_2(t) = 0, \tag{6.23a}$$

$$\frac{dS_1(t)}{dt} + k_1E(t)S_1(t) - r_1ES_1(t) = 0, \tag{6.23b}$$

$$\frac{dES_1(t)}{dt} - k_1E(t)S_1(t) + (r_1 + k_2)ES_1(t) = 0, \tag{6.23c}$$

$$\frac{dQ(t)}{dt} - k_2ES_1(t) = 0, \tag{6.23d}$$

$$\frac{dE^*(t)}{dt} - k_2ES_1(t) + k_3E^*(t)S_2(t) - r_3E^*S_2(t) = 0, \tag{6.23e}$$

$$\frac{dS_2(t)}{dt} + k_3E^*(t)S_2(t) - r_3E^*S_2(t) = 0, \tag{6.23f}$$

$$\frac{dE^*S_2(t)}{dt} - k_3E^*(t)S_2(t) + (r_3 + k_4)E^*S_2(t) = 0, \tag{6.23g}$$

$$\frac{dP(t)}{dt} - k_4E^*S_2(t) = 0, \tag{6.23h}$$

$$E_0 - E(t) - ES_1(t) - E^*(t) - E^*S_2(t) = 0, \tag{6.23i}$$

$$S_{10} - S_1(t) - ES_1(t) - Q(t) = 0, \tag{6.23j}$$

$$S_{10} - S_1(t) - ES_1(t) - E^*(t) - E^*S_2(t) - P(t) = 0, \tag{6.23k}$$

$$S_{20} - S_2(t) - E^*S_2(t) - P(t) = 0, \tag{6.23l}$$

$$S_{20} - S_2(t) - E^*S_2(t) - E(t) - ES_1(t) - Q(t) = 0. \tag{6.23m}$$

Derivation of output relations for y_1 to y_3 and y_5 to y_7 proved computationally intractable. For y_4 and y_8 the differential ideal was successfully decomposed into two ideals using the `Rosenfeld_Groebner` algorithm in *Maple*. The subset of the generators, Equations (6.23c)-(6.23e), (6.23g)-(6.23i) and (6.23k)-(6.23l), can be made (by appropriate substitutions from the equilibrium relations) orthonomic and autoreduced with respect to any ranking for which the members of the following set, $\{E,\ S_1,\ S_2,\ y_8\}$, are ranked above the remaining variables, $\{ES_1,\ E^*,\ E^*S_2,\ Q,\ P\}$. This subset of the generators was successfully decomposed into the same two ideals as obtained for the full set of generators above. The left hand sides of the

generators of the first ideal are given by:

$$\left.\begin{array}{l} S_2(t),\ E^*S_2(t),\ P(t)-S_{20},\ y_8(t)-S_{20},\ k_2ES_1(t)-E^{*(1)}(t),\ Q^{(1)}(t)-E^{*(1)}(t), \\ k_2(E(t)+E^*(t)-E_0)+E^{*(1)}(t)),\ k_2(S_1(t)+E^*(t)+(S_{20}-S_{10}))+E^{*(1)}(t), \\ k_2E^{*(2)}(t)+k_1k_2^2(E_0(S_{20}-S_{10})+E^*(t)(E_0+S_{10}-S_{20})-E^{*2}(t))- \\ -k_1(E^{*(1)}(t))^2+k_1k_2E^{*(1)}(t)((E_0+S_{10}-S_{20})-2E^*(t))+k_2E^{*(1)}(t)(k_2+r_1), \end{array}\right\} \quad (6.24)$$

(the right hand sides being zero). The first generator above indicates that this is a reaction state in which S_2 has been exhausted. The fifth to ninth generators describe the ongoing first half of the reaction which can continue until the available concentration of E or S_1 is exhausted. This ideal does not describe the required reaction state and can be eliminated where it is known that $S_2(t)$ is non-zero.

Presenting the generators of the second ideal on paper is not feasible due to their size[1]. Several generators contain the denominator:

$$k_2k_3k_4^2(y_8^{(1)}(t)+k_4(y_8(t)-S_{20}))^3, \quad (6.25)$$

If this denominator were zero it could be included as one of the generators of the differential ideal. However in this case the ideal decomposes to a single ideal, Equations (6.24), which does not describe the required reaction state. Thus this denominator can be considered non-zero. Thus the leader of the function, $y^{(4)}$, can be eliminated as described in Section 3.2.3. A second output relation is created by replacing $\mathbf{p}$ with the alternative parameter vector and subtracted from the first output relation, this process is analogous to that presented in Section 4.3.1. Since the denominator of both relations is non-zero, the denominator of the difference of these relations is also non-zero and can be eliminated to yield a differential polynomial. Since the generators

[1] These generators and the following analysis are available on request (and at the following URL) in an electronic *Maple* file: http://www2.warwick.ac.uk/fac/sci/moac/currentstudents/2005/daniel_bearup.

are orthonomic and autoreduced the ideal is prime. Thus, by the argument presented in Section 3.2.3.1, the 265 monomials of this expression are linearly independent. Since the polynomial equals zero its coefficients must all equal zero and these equations can be solved to determine relationships between the unknown parameter vectors. These six coefficients of the polynomial are sufficient to establish that the model is SGI:

$$k_4^2\overline{k_4}^2(k_2k_3\overline{k_1} - k_1\overline{k_2k_3}), \tag{6.26a}$$

$$3k_4^2\overline{k_4}^2(k_2k_3k_4^2\overline{k_1} - k_1\overline{k_2k_3k_4}^2), \tag{6.26b}$$

$$k_3\overline{k_3}(k_2k_4^2\overline{k_1k_3} - k_1k_3\overline{k_2k_4}^2), \tag{6.26c}$$

$$k_2k_3k_4^5\overline{k_2k_3k_4}^5(\overline{k_1k_3} - k_1k_3), \tag{6.26d}$$

$$k_3k_4^4\overline{k_3k_4}^4\begin{pmatrix} 7S_{20}(k_1k_3\overline{k_2k_4} - k_2k_4\overline{k_1k_3}) + 2(k_1k_4\overline{k_2k_3} - k_2k_3\overline{k_1k_4}) + \\ + 2(k_1r_3\overline{k_2k_4} - k_4k_2\overline{k_1}\,\overline{r_3}) + (k_1k_2\overline{k_2k_4} - k_2k_4\overline{k_1k_2}) \end{pmatrix}, \tag{6.26e}$$

$$k_2k_3k_4^5\overline{k_2k_3k_4}^5S_{20}^6\begin{pmatrix} (E_0 + S_{10})(\overline{k_1} - k_1) + (\overline{k_2} - k_2) + S_{20}(\overline{k_3} - k_3) + \\ + (\overline{k_4} - k_4) + (\overline{r_1} - r_1) + (\overline{r_3} - r_3) \end{pmatrix}. \tag{6.26f}$$

They were solved for the alternative parameters in *Maple* to obtain the five solutions below:

$$\{\overline{k_1} = 0, \overline{k_2} = 0\},\ \{\overline{k_1} = 0,\ \overline{k_3} = 0\}, \tag{6.27a}$$

$$\{\overline{k_1} = 0,\ \overline{k_4} = 0\},\ \{\overline{k_3} = 0,\ \overline{k_4} = 0\}, \tag{6.27b}$$

$$\{\overline{k_1} = k_1,\ \overline{k_2} = k_2,\ \overline{k_3} = k_3,\ \overline{k_4} = k_4,\ \overline{r_1} = r_1,\ \overline{r_3} = r_3\}. \tag{6.27c}$$

Note that the same solutions are obtained when all coefficients of the polynomial are solved. The first four solutions above do not lie in the set of feasible parameter values and thus can be eliminated. The only remaining solution (6.27c) indicates that $\overline{\mathbf{p}} = \mathbf{p}$ and thus that the model arising from measurement of P is SGI.

For measurement of Q a different subset of the generators was required. The subset of the generators, Equations (6.23a), (6.23c)-(6.23d), (6.23g)-(6.23j) and (6.23m), can be made (by appropriate substitutions from the equilibrium relations) orthonomic and autoreduced with respect to any ranking for which $\{E^*,\ S_1,\ S_2,\ y_4\} > \{ES_1,\ E,\ E^*S_2,\ Q,\ P\}$. This subset of the generators was successfully decomposed into three ideals. The left hand sides of the generators of the first of these ideals are:

$$\left.\begin{array}{l} S_1(t),\ ES_1(t),\ E^*S_2(t),\ Q(t)-S_{10},\ y_4(t)-S_{10},\ E(t)+S_2(t)+(S_{10}-S_{20}), \\ E^*(t)-S_2(t)-(E_0+S_{10}-S_{20}),\ S_2^2(t)+S_2(t)(E_0+S_{10}-S_{20}),\ P^{(1)}. \end{array}\right\} \quad (6.28)$$

Again this ideal corresponds to a situation where substrate exhaustion, in this case S_1, has occurred. The second ideal also corresponds to such a state but, due to the length of some of the generators this ideal is not presented here[2]. Neither ideal describes the required reaction state, as such they are eliminated.

Presenting the generators of the third ideal on paper is not feasible due to their size[3]. The output relation contains the following denominator:

$$k_1k_2^2k_4(k_2y_4(t)-S_{10}k_2+y^{(1)}(t))^3. \quad (6.29)$$

If this expression is added to the set of generators the only ideals obtained are the two described above corresponding to an inappropriate reaction state. As such this expression can be considered non-zero. As previously argued the 265 monomials of this differential polynomial were

[2] These generators are available on request (and at the following URL) in an electronic *Maple* file: http://www2.warwick.ac.uk/fac/sci/moac/currentstudents/2005/daniel_bearup.

[3] These generators and the following analysis are available on request (and at the following URL) in an electronic *Maple* file: http://www2.warwick.ac.uk/fac/sci/moac/currentstudents/2005/daniel_bearup.

linearly independent. Its coefficients were solved yielding the following five solutions:

$$\{\overline{k_1} = 0,\ \overline{k_2} = 0\},\ \{\overline{k_1} = 0,\ \overline{k_3} = 0\}, \tag{6.30a}$$

$$\{\overline{k_2} = 0,\ \overline{k_3} = 0\},\ \{\overline{k_3} = 0, \overline{k_4} = 0\}, \tag{6.30b}$$

$$\{\overline{k_1} = k_1,\ \overline{k_2} = k_2,\ \overline{k_3} = k_3,\ \overline{k_4} = k_4,\ \overline{r_1} = r_1,\ \overline{r_3} = r_3\} \tag{6.30c}$$

The first four solutions are not in the set of feasible parameter values and thus can be eliminated. The final solution is equivalent to $\overline{\mathbf{p}} = \mathbf{p}$ and thus the model corresponding to measurement of Q is SGI.

The input-output relationship approach shows that measurement of either product is SGI. This result is independent of the initial conditions used, as long as both initial substrate concentrations and the total concentration of enzyme are non-zero; conditions which can be guaranteed experimentally.

6.2.1.2 Differential algebra analysis of Mechanism (b) for the ping-pong model

The generators of the differential ideal equivalent to Mechanism (b) are Equations (6.23b)-(6.23f), (6.23i)-(6.23j) and (6.23l) and the following expressions:

$$\frac{dE(t)}{dt} + k_1E(t)S_1(t) - r_1ES_1(t) - k_4E^*S_2(t) + r_4E(t)P(t) = 0, \tag{6.31a}$$

$$\frac{dE^*S_2(t)}{dt} - k_3E^*(t)S_2(t) + (r_3 + k_4)E^*S_2(t) - r_4E(t)P(t) = 0, \tag{6.31b}$$

$$\frac{dP(t)}{dt} - k_4E^*S_2(t) + r_4E(t)P(t) = 0, \tag{6.31c}$$

$$E^*(t) + E^*S_2(t) + P(t) - Q(t) = 0. \tag{6.31d}$$

Derivation of output relations for y_1 to y_3 and y_5 to y_8 proved computationally intractable. For y_4 the differential ideal was successfully decomposed into two ideals using the `Rosenfeld_Groebner` algorithm in *Maple*. The subset of the generators, Equations (6.31a), (6.23c)-(6.23d), (6.31b),

(6.23i)-(6.23k) and (6.31d), can be made (by appropriate substitutions) orthonomic and autoreduced with respect to any ranking for which $\{E^*,\ S_1,\ S_2,\ y_4\} > \{P\} > \{ES_1,\ E,\ E^*S_2,\ Q\}$. This subset of the generators was successfully decomposed into two ideals obtained using the full set described above. A subset of the left hand sides of the generators of the first of these ideals is:

$$\left.\begin{array}{l} S_1(t),\ ES_1(t),\ Q(t)-S_{10},\ y_4(t)-S_{10},\ E(t)-P(t)+(S_{10}-E_0), \\ k_4(E^*(t)+S_{10}+P(t))+P^{(1)}(t)+r_4P(t)(S_{10}-E_0-P(t)), \\ k_4(S_2(t)-S_{20}+P(t))+P^{(1)}(t)+r_4P(t)(E_0-S_{10}+P(t)), \\ k_4E^*S_2(t)-P^{(1)}(t)+r_4P(t)(S_{10}-E_0-P(t)). \end{array}\right\} \tag{6.32}$$

The final generator, which has the leader $P^{(2)}(t)$, is not presented here due to its length[4]. The generators above are sufficient to show that this ideal does not describe the required reaction state. From the first four generators, S_1 has been exhausted thus only the second half of the reaction is dynamic.

Presenting the generators of the second ideal on paper is not feasible due to their size[5]. The output relation contains the following denominator:

$$k_2^2k_4(k_1(k_2(y(t)-S_{10})+y_4^{(1)}(t)))^3. \tag{6.33}$$

If this expression is added to the set of generators the only ideals obtained are the two described above corresponding to an inappropriate reaction state. As such this expression can be considered non-zero. With this constraint the output relation can be manipulated as previously described to eliminate the leader, $y_4^{(4)}(t)$. The result is a differential polynomial. Since the gen-

[4] This generator is available on request (and at the following URL) in an electronic *Maple* file: http://www2.warwick.ac.uk/fac/sci/moac/currentstudents/2005/daniel_bearup.

[5] These generators and the following analysis are available on request (and at the following URL) in an electronic *Maple* file: http://www2.warwick.ac.uk/fac/sci/moac/currentstudents/2005/daniel_bearup.

erators are orthonomic and autoreduced the ideal is prime. Thus, by the argument presented in Section 3.2.3.1, the 266 monomials of the differential polynomial are linearly independent and their coefficients can be analyzed as previously described. The seven coefficients listed below are sufficient to show this observation is SGI:

$$k_1^2 k_2^2 \overline{k_1}^2 \overline{k_2}^2 (k_1 k_4 \overline{k_3} - k_3 \overline{k_1 k_4}), \tag{6.34a}$$

$$k_1^2 k_2^4 \overline{k_1}^2 \overline{k_2}^4 (k_1 k_2 k_4 \overline{k_3} - k_3 \overline{k_1 k_2 k_4}), \tag{6.34b}$$

$$k_1^3 k_2^5 k_4 \overline{k_1}^3 \overline{k_2}^5 \overline{k_4} E_0 (k_1 k_2 k_3 k_4 - \overline{k_1 k_2 k_3 k_4}), \tag{6.34c}$$

$$2k_1^2 k_2 \overline{k_1}^2 \overline{k_2} (k_1 k_2 k_4 \overline{k_3} (\overline{r_4} - \overline{k_1}) - k_3 \overline{k_1 k_2 k_4} (k_1 - r_4)), \tag{6.34d}$$

$$2k_1 k_2^3 \overline{k_1 k_2}^3 (k_3 \overline{k_1}^2 \overline{k_2}^2 \overline{k_4} (k_1 k_2^2 + r_4) - k_1^2 k_2^2 k_4 \overline{k_3} (\overline{k_1} + \overline{k_3 r_4})), \tag{6.34e}$$

$$k_1^3 k_2^4 \overline{k_1}^3 \overline{k_2}^4 \begin{pmatrix} 4(\overline{k_2}(k_1 - r_4) + k_2(\overline{r_4} - \overline{k_1})) + 5k_4(k_3 \overline{k_2} - k_2 \overline{k_3}) + \\ + 2(k_2 \overline{k_2}(k_3 - \overline{k_3}) + k_2 \overline{k_1}^2 \overline{k_3}\,\overline{r_4}(\overline{r_1} + \overline{k_2}) - \overline{k_2} k_3 k_1^2 r_4 (r_1 + k_2)) + \\ + 3(k_2(k_3 - r_4) + \overline{k_2}(\overline{r_4} - \overline{k_3}) + (k_1 k_2 - \overline{k_1 k_2})) \end{pmatrix}, \tag{6.34f}$$

$$k_1^3 k_2^5 k_4 \overline{k_1}^3 \overline{k_2}^5 \overline{k_4} \begin{pmatrix} (2E_0 + 7S_{10} + S_{20})(k_1 k_3 - \overline{k_1 k_3}) + (k_1 k_2 - \overline{k_1 k_2}) + \\ + (3E_0 + S_{20} + 6S_{10})(\overline{k_3}\,\overline{r_4} - k_3 r_4) + (\overline{r_4 r_1} - r_4 r_1) + \\ + (k_2 k_3 - \overline{k_2 k_3}) + (\overline{r_3 r_4} - r_3 r_4) + (k_1 k_4 - \overline{k_1 k_4}) + \\ + (k_1 r_3 - \overline{k_1}\,\overline{r_3}) + (k_3 k_4 - \overline{k_3 k_4}) + (k_3 r_1 - \overline{k_3}\,\overline{r_1}) + \\ + (\overline{k_2}\,\overline{r_4} - k_2 r_4) \end{pmatrix}. \tag{6.34g}$$

These coefficients were solved for the alternative unknown parameters yielding the following seven solutions:

$$\{\overline{k_1} = 0,\ \overline{k_2} = 0\},\ \{\overline{k_1} = 0,\ \overline{k_3} = 0\},\ \{\overline{k_1} = 0,\ \overline{r_4} = 0\}, \tag{6.35a}$$

$$\{\overline{k_2} = 0,\ \overline{k_3} = 0\},\ \{\overline{k_3} = 0,\ \overline{k_4} = 0\},\ \{\overline{k_1} = \overline{r_4},\ \overline{k_2} = 0,\ \overline{k_4} = \overline{r_1}\}, \tag{6.35b}$$

$$\{\overline{k_1} = k_1,\ \overline{k_2} = k_2,\ \overline{k_3} = k_3,\ \overline{k_4} = k_4,\ \overline{r_1} = r_1,\ \overline{r_3} = r_3,\ \overline{r_4} = r_4\}. \tag{6.35c}$$

Note that the same solutions are obtained when all coefficients of the polynomial are solved. The first six solutions do not lie in the set of feasible parameter values and thus are rejected. The remaining solution (6.35c), $\overline{\mathbf{p}} = \mathbf{p}$, indicates that the model corresponding to measurement of Q alone is SGI. This analysis is not affected by the relabelling, Equation (6.16), although some changes may need to be made to the ranking chosen. Thus when release of Q is reversible, the model corresponding to measurement of P is SGI. Note that these results are independent of initial substrate concentrations; as long as the substrate and enzyme concentrations are non-zero. These conditions can be guaranteed experimentally.

6.2.2 *Simple ordered model with simultaneous product release*

The two substrate transient simple ordered model is now considered. Two product release mechanisms were considered: irreversible product release (a), as previously described in Section 4.2.1, and reversible product release (b). A reaction scheme for Mechanism (b) is given by:

$$\left.\begin{aligned} E + S_1 &\underset{r_1}{\overset{k_1}{\rightleftharpoons}} ES_1 \\ ES_1 + S_2 &\underset{r_2}{\overset{k_2}{\rightleftharpoons}} ES_1S_2 \\ ES_1S_2 &\underset{r_3}{\overset{k_3}{\rightleftharpoons}} E + P \end{aligned}\right\} \tag{6.36}$$

Setting r_3 equal to zero in Equations (6.36) yields Mechanism (a). A Taylor series analysis of these two mechanisms is presented below.

The state vector and initial conditions used were the same for both models:

$$\left.\begin{aligned} &\mathbf{x}(t,\mathbf{p}) = (E(t,\mathbf{p}), S_1(t,\mathbf{p}), S_2(t,\mathbf{p}), ES_1(t,\mathbf{p}), ES_1S_2(t,\mathbf{p}), P(t,\mathbf{p}))^T, \\ &\mathbf{x}(0,\mathbf{p}) = \mathbf{x}_0 = (E_0, S_{10}, S_{20}, 0, 0, 0)^T. \end{aligned}\right\} \tag{6.37}$$

Each mechanism gives rise to a different derivative of the state vector. For irreversible product

release, Mechanism (a), the derivative was as previously stated, Equations (4.9) in Section 4.2.1. Reversible product release requires the following modifications to these equations:

$$\frac{dE(t,\mathbf{p})}{dt} = -k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) + r_1ES_1(t,\mathbf{p}) + k_3ES_1S_2(t,\mathbf{p}) - r_3E(t,\mathbf{p})P(t,\mathbf{p}), \quad (6.38a)$$

$$\frac{dES_1S_2(t,\mathbf{p})}{dt} = k_2ES_1(t,\mathbf{p})S_2(t,\mathbf{p}) - (k_3 + r_2)ES_1S_2(t,\mathbf{p}) + r_3E(t,\mathbf{p})P(t,\mathbf{p}), \quad (6.38b)$$

$$\frac{P(t,\mathbf{p})}{dt} = k_3ES_1S_2(t,\mathbf{p}) - r_3E(t,\mathbf{p})P(t,\mathbf{p}). \quad (6.38c)$$

Thus each reaction mechanism requires a different unknown parameter vector, given below:

$$\mathbf{p}_a = (k_1, k_2, k_3, r_1, r_2)^T, \quad (6.39)$$

$$\mathbf{p}_b = (k_1, k_2, k_3, r_1, r_2, r_3)^T, \quad (6.40)$$

the subscript indicating to which mechanism the vector corresponds. Six output functions, corresponding to individual measurement of each reaction species, were analyzed:

$$\left.\begin{aligned} &y_1(t,\mathbf{p}) = E(t,\mathbf{p}),\ y_2(t,\mathbf{p}) = S_1(t,\mathbf{p}),\ y_3(t,\mathbf{p}) = S_2(t,\mathbf{p}), \\ &y_4(t,\mathbf{p}) = ES_1(t,\mathbf{p}),\ y_5(t,\mathbf{p}) = ES_1S_2(t,\mathbf{p}),\ y_6(t,\mathbf{p}) = P(t,\mathbf{p}). \end{aligned}\right\} \quad (6.41)$$

At least six non-zero Taylor series coefficients were calculated for each output function and mechanism. For Mechanisms (a) and (b) the Jacobian matrices calculated were rank five and six respectively for each observation; equal to the number of unknown parameters for each mechanism. As such by PRT both mechanisms are at least locally identifiable for the observations considered.

Full Taylor series analyses of output functions y_4 to y_6 proved computationally intractable for Mechanism (a); it was not possible to show that any individual parameters were globally identifiable. For Mechanism (b) full analyses proved computationally intractable for y_1 and y_4

to y_6. However if enzyme concentration was observed it was possible to show that k_1 and r_1 are globally identifiable from the Taylor series coefficients below, Equations (6.42)-(6.45).

A full Taylor series analysis of each of the remaining observations, y_1 to y_3 for Mechanism (a) and y_2 to y_3 for Mechanism (b), was completed. The first coefficient of the Taylor series expansions of y_2 and the first four of y_1 and y_4 at $t = 0$ are:

$$y_1(0) = E_0,\ y_2(0) = S_{10},\ y_4(0) = 0 \tag{6.42}$$

$$y_1^{(1)}(0) = -y_4^{(1)}(0) = -k_1 E_0 S_{10} \tag{6.43}$$

$$y_1^{(2)}(0) = -y_4^{(2)}(0) - k_1 k_2 E_0 S_{10} S_{20} = k_1 E_0 S_{10}(k_1(E_0 + S_{10}) + r_1) \tag{6.44}$$

$$\begin{aligned} y_1^{(3)}(0) - k_1 k_2 k_3 E_0 S_{10} S_{20} &= -y_4^{(3)}(0) + k_1 k_2 E_0 S_{10} S_{20}(k_1(E_0 + S_{10}) + k_2 S_{20} + r_1 + r_2) = \\ &= -k_1 E_0 S_{10}((k_1(E_0 + 2k_1 E_0 S_{10} + S_{10}) + r_1)(k_1(E_0 + S_{10}) + r_1) + k_2 r_1 S_{20}) \end{aligned} \tag{6.45}$$

The second, third and fourth Taylor series coefficients for y_2 are equivalent to those listed above for y_1. These coefficients were the same for both mechanisms. The following alternative parameter vectors:

$$\overline{\mathbf{p}}_a = (\overline{k_1}, \overline{k_2}, \overline{k_3}, \overline{r_1}, \overline{r_2})^T, \tag{6.46}$$

$$\overline{\mathbf{p}}_b = (\overline{k_1}, \overline{k_2}, \overline{k_3}, \overline{r_1}, \overline{r_2}, \overline{r_3})^T, \tag{6.47}$$

were used as previously described in Section 3.2.1 to create simultaneous equations in the two sets of parameters from the Taylor series coefficients derived. These equations were then solved in *Maple* for the alternative parameters; showing that $\overline{\mathbf{p}} = \mathbf{p}$. Thus these observations give rise to SGI models.

Where a reaction is known to be reversible it is natural to study the reverse reaction by choosing initial non-zero product concentrations. A structural identifiability analysis of this case was

conducted. The new initial conditions used are:

$$\mathbf{x}_0 = (E_0, 0, 0, 0, 0, P_0)^T. \tag{6.48}$$

An experiment where substrate concentrations were also non-zero could also be considered but has not been in this work. For Mechanism (a) the computable Taylor series expansion at $t = 0$ had only one non-zero coefficient, which contained none of the unknown parameters. As such it was not possible to complete a Taylor series analysis using these initial conditions. However for Mechanism (b) it was possible to complete a Taylor series analysis; the first four Taylor series coefficients at $t = 0$ are:

$$y_6(0) = P_0, \tag{6.49}$$

$$y_6^{(1)} = -r_3 E_0 P_0, \tag{6.50}$$

$$y_6^{(2)} = r_3 E_0 P_0 (r_3(E_0 + P_0) + k_3), \tag{6.51}$$

$$y_6^{(3)} = -k_3 r_3 E_0 P_0 (k_3 + r_2 + r_3(E_0 + P_0)) - r_3^2 E_0 P_0 (k_3 + r_3(E_0 + P_0)) - \\ - 2r_3^3 E_0^2 P_0^2 - r_3^2 E_0^2 P_0 (k_3 + r_3(E_0 + P_0)). \tag{6.52}$$

Note the similarity between the coefficients above and those found for measurement of S_1 under the other set of initial conditions when P_0, r_3, k_3 and r_2 are relabelled as S_1, k_1, r_1 and k_2 respectively. The coefficients were analysed as previously described and it was shown that $\overline{\mathbf{p}} = \mathbf{p}$. As such measurement of product concentration gives rise to a model which is SGI for the initial conditions used for Mechanism (b).

The results of this analysis are summarised in Table 6.2. It was not possible to complete a Taylor series analysis for several output functions, most significantly y_6 for Mechanism (a); measurement of P being a common experimental procedure. As such the input-output relationship approach was attempted for these two mechanisms. Derivation of the output relations

Simple ordered model 1	Reversible product release stages	
Measurement	None	One
E	SGI	ASLI, $\{k_1, r_1\}$
S_1	SGI	SGI
ES_1	SGI	SGI
S_2	ASLI	ASLI
ES_2	ASLI	ASLI
P	SGI‡	SGI†

Table 6.2: Summary of the structural identifiability analysis results for the two substrate transient simple ordered model with simultaneous product release. Results are either: SGI, structurally globally identifiable; ASLI, at least structurally locally identifiable; or CI, analysis was computationally intractable. In addition a subset of globally identifiable parameters are stated in some cases. Results marked by a † were obtained for the alternative initial conditions, Equation (6.48). Results marked by a ‡ can be obtained using the input-output relationship approach, and require only that enzyme and substrate concentrations be initially non-zero.

proved computationally intractable for Mechanism (b); the results for Mechanism (a) in the following section.

6.2.2.1 *Differential algebra analysis for Mechanism (a) for the first simple ordered model*

The following expressions are the generators of a differential ideal equivalent to Mechanism (a):

$$\frac{dE(t,\mathbf{p})}{dt} + k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) - r_1ES_1(t,\mathbf{p}) - k_3EP(t,\mathbf{p}) = 0, \tag{6.53a}$$

$$\frac{dS_1(t,\mathbf{p})}{dt} + k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) - r_1ES_1(t,\mathbf{p}) = 0, \tag{6.53b}$$

$$\frac{dES_1(t,\mathbf{p})}{dt} - k_1E(t,\mathbf{p})S_1 + r_1ES_1 + k_2ES_1(t,\mathbf{p})S_2(t,\mathbf{p}) - r_2ES_1S_2(t,\mathbf{p}) = 0, \tag{6.53c}$$

$$\frac{dS_2(t,\mathbf{p})}{dt} + k_2ES_1(t,\mathbf{p})S_2(t,\mathbf{p}) - r_2ES_1S_2(t,\mathbf{p}) = 0, \tag{6.53d}$$

$$\frac{dES_1S_2(t,\mathbf{p})}{dt} - k_2ES_1(t,\mathbf{p})S_2(t,\mathbf{p}) + (r_2 + k_3)ES_1S_2(t,\mathbf{p}) = 0, \tag{6.53e}$$

$$\frac{dP(t,\mathbf{p})}{dt} - k_3EP(t,\mathbf{p}) = 0, \tag{6.53f}$$

$$E0 - E(t,\mathbf{p}) - ES_1(t,\mathbf{p}) - ES_1S_2(t,\mathbf{p}) = 0, \tag{6.53g}$$

$$S_{10} - S_1(t,\mathbf{p}) - ES_1(t,\mathbf{p}) - ES_1S_2(t,\mathbf{p}) - P(t,\mathbf{p}) = 0, \tag{6.53h}$$

$$S_{20} - S_2(t, \mathbf{p}) - ES_1S_2(t, \mathbf{p}) - P(t, \mathbf{p}) = 0. \qquad (6.53i)$$

Derivation of an output relation for y_1 to y_5 proved computationally intractable. For y_6, corresponding to measurement of product, the differential ideal was successfully decomposed into two ideals using the `Rosenfeld_Groebner` algorithm in *Maple*. The subset of the generators, Equations (6.53c) and (6.53e)-(6.53i), can be made orthonomic and autoreduced with respect to any ranking for which $\{E,\ S_1,\ S_2,\ y_6\} > \{ES_1,\ ES_1S_2,\ P\}$. This subset of the generators was successfully decomposed into two ideals equivalent to those previously found. The left hand side of the generators of the first ideal are:

$$\left.\begin{array}{l} E + ES_1 - E_0,\ S_1 + ES_1 - (S_{10} - S_{20}),\ S_2,\ ES_1S_2,\ y_6 - S_{20}, \\ ES_1^{(1)} + k_1(ES_1(E_0 + (S_{10} - S_{20}) - ES_1) - E_0(S_{10} - S_{20})) + r_1ES_1. \end{array}\right\} \qquad (6.54)$$

From the fourth and fifth elements of this ideal the concentration of S_2 and ES_1S_2 must be zero. The sixth and seventh elements indicate that the concentration of product is always equal to that of the initial concentration of S_2. Based on these observations this ideal describes a model state in which the second substrate has been consumed while some of the first substrate remains. The reversible association and dissociation of the first substrate to and from the enzyme is described by the first, second and seventh elements of the ideal. This ideal does not describe the required reaction state and can be eliminated practically if it is known that the concentration of the second substrate is non-zero. This can be guaranteed experimentally.

Presenting the generators of the second ideal on paper is not feasible due to their size[6]. Several generators contain a denominator:

$$y_6^{(1)} - k_3(S_{20} - y_6) \text{ or equivalently } P^{(1)} = k_3(S_{20} - P), \qquad (6.55)$$

[6] These generators and the following analysis are available on request (and at the following URL) in an electronic *Maple* file: http://www2.warwick.ac.uk/fac/sci/moac/currentstudents/2005/daniel_bearup.

which, if equal to zero, implies that the reaction state is described by the ideal above. As such this denominator can be considered non-zero. With this constraint it is possible to eliminate the leader of the output relation, $y_6^{(3)}$, by substitution of the alternative unknown parameter vector, Equation (6.46), and obtain a differential polynomial as described in Section 6.2.1.1. Since the generators are orthonomic and autoreduced the ideal is prime. Thus, by the argument presented in Section 3.2.3.1, the 34 monomials of this differential polynomial can be considered linearly independent. These five coefficients of the polynomial are sufficient to establish that the model is SGI:

$$k_3^2\overline{k_3}^2(k_2\overline{k_1} - \overline{k_2}k_1), \tag{6.56a}$$

$$k_3^2\overline{k_3}^2(\overline{k_2k_3}(k_2 - k_1) + k_2k_3(\overline{k_1} - \overline{k_2})), \tag{6.56b}$$

$$k_2k_3\overline{k_2k_3}((k_2\overline{k_3} - k_3\overline{k_2}) + 2(\overline{k_3}k_1 - k_3\overline{k_1})), \tag{6.56c}$$

$$\overline{k_2k_3}^3k_2k_3^3S_{20}^2((\overline{k_3} - k_3) + E_0S_{10}(\overline{k_1} - k_1) + S_{20}(\overline{k_2} - k_2) + (\overline{r_1} - r_1) + (\overline{r_2} - r_2)), \tag{6.56d}$$

$$\overline{k_2}k_2\left(\begin{array}{c} S_{20}(\overline{k_3}^3k_1k_2 - k_3^3\overline{k_1k_2}) - k_3^2\overline{k_1k_3}(\overline{k_2}(E_0 + S_{10}) + 2(\overline{k_2}S_{20} + \overline{r_2})) + \\ \\ + k_3\overline{k_3}^2(k_3(2(k_1 - \overline{k_1}) + (k_2 - \overline{k_2})) + k_1(k_2(E_0 + S_{10}) + 2(k_2S_{20} + r_2))) \end{array}\right). \tag{6.56e}$$

These coefficients were solved for the alternative parameter vector using *Maple* to obtain the following four solutions:

$$\{\overline{k_1} = 0,\ \overline{k_2} = 0\},\ \{\overline{k_2} = 0,\ \overline{k_3} = 0\},\ \{\overline{k_1} = 0,\ \overline{k_3} = 0\}, \tag{6.57}$$

$$\{\overline{k_1} = k_1,\ \overline{k_2} = k_2,\ \overline{k_3} = k_3,\ \overline{r_1} = r_1,\ \overline{r_2} = r_2\}. \tag{6.58}$$

Note that the same solutions are obtained when all coefficients of the polynomial are solved simultaneously. The first three solutions do not lie within the set of feasible parameter values, as such they are rejected. Thus a single solution remains, $\overline{\mathbf{p}} = \mathbf{p}$, showing that with this observation the model is SGI. These results are independent of initial substrate concentrations;

as long as initial concentrations of substrate and enzyme concentrations are non-zero. These conditions can be guaranteed experimentally.

6.2.3 Simple ordered model with sequential product release

The release of products from the ternary complex is not necessarily simultaneous. In this section models for sequential product release are analysed. The models used are created by extending the model used previously, described by Equations (4.9) in Section 4.2.1. Four product release mechanisms are considered: fully irreversible (a), reversible for the first product (b), reversible for the second product(c), and fully reversible (d). A reaction scheme for Mechanism (d) is given by:

$$\left.\begin{array}{rcll} E + S_1 & \underset{r_1}{\overset{k_1}{\rightleftharpoons}} & ES_1 & \\ ES_1 + S_2 & \underset{r_2}{\overset{k_2}{\rightleftharpoons}} & ES_1S_2 & E + P \\ ES_1S_2 & \underset{r_3}{\overset{k_3}{\rightleftharpoons}} & EP + Q & \\ EP & \underset{r_4}{\overset{k_4}{\rightleftharpoons}} & & \end{array}\right\} \tag{6.59}$$

Mechanisms (c), (b) and (a) can then be obtained from Equations (6.59) by setting r_3, r_4 or both equal to zero respectively. A Taylor series analysis of each of these mechanisms is presented below.

The state vector and initial conditions are the same for all models:

$$\left.\begin{array}{l} \mathbf{x}(t,\mathbf{p}) = \\ = (E(t,\mathbf{p}), S_1(t,\mathbf{p}), S_2(t,\mathbf{p}), ES_1(t,\mathbf{p}), ES_1S_2(t,\mathbf{p}), Q(t,\mathbf{p}), EP(t,\mathbf{p}), P(t,\mathbf{p}))^T, \\ \\ \mathbf{x}(0,\mathbf{p}) = \mathbf{x}_0 = (E_0, S_{10}, S_{20}, 0, 0, 0, 0, 0)^T. \end{array}\right\} \tag{6.60}$$

However each mechanism gives rise to a different derivative of the state vector. For irreversible product release, Mechanism (a), this derivative is as stated previously, Equations (4.9), with

the following additions and exceptions:

$$\frac{dE(t,\mathbf{p})}{dt} = -k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) + r_1ES_1(t,\mathbf{p}) + k_4EP(t,\mathbf{p}), \tag{6.61a}$$

$$\frac{dEP(t,\mathbf{p})}{dt} = k_3ES_1S_2(t,\mathbf{p}) - k_4EP(t,\mathbf{p}), \tag{6.61b}$$

$$\frac{dQ(t,\mathbf{p})}{dt} = k_3ES_1S_2(t,\mathbf{p}), \tag{6.61c}$$

$$\frac{dP(t,\mathbf{p})}{dt} = k_4EP(t,\mathbf{p}). \tag{6.61d}$$

If the first product release stage is reversible, Mechanism (b), the following additional changes were made:

$$\frac{dES_1S_2(t,\mathbf{p})}{dt} = k_2ES_1(t)S_2(t) - (r_2+k_3)ES_1S_2(t) + r_3EP(t)(t)Q(t), \tag{6.62a}$$

$$\frac{dEP(t,\mathbf{p})}{dt} = k_3ES_1S_2(t,\mathbf{p}) - k_4EP(t,\mathbf{p}) - r_3EP(t)Q(t), \tag{6.62b}$$

$$\frac{dQ(t,\mathbf{p})}{dt} = k_3ES_1S_2(t,\mathbf{p}) - r_3EP(t)S_2(t)Q(t). \tag{6.62c}$$

Similarly if the second product release state is reversible, Mechanism (c), Mechanism (a) was modified as follows:

$$\frac{dE(t,\mathbf{p})}{dt} = -k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) + r_1ES_1(t,\mathbf{p}) + k_4EP(t,\mathbf{p}) - r_4E(t)P(t), \tag{6.63a}$$

$$\frac{dEP(t,\mathbf{p})}{dt} = k_3ES_1S_2(t,\mathbf{p}) - k_4EP(t,\mathbf{p}) + r_4E(t)P(t), \tag{6.63b}$$

$$\frac{dP(t,\mathbf{p})}{dt} = k_4EP(t,\mathbf{p}) - r_4E(t)P(t). \tag{6.63c}$$

Finally for a fully reversible model, Mechanism (d), the following modifications to Mechanism (a) were used:

$$E'(t,\mathbf{p}) = -k_1E(t,\mathbf{p})S_1(t,\mathbf{p}) + r_1ES_1(t,\mathbf{p}) + k_4EP(t,\mathbf{p}) - r_4E(t)P(t), \tag{6.64a}$$

$$ES_1S_2'(t,\mathbf{p}) = k_2ES_1(t)S_2(t) - (r_2+k_3)ES_1S_2(t) + r_3EP(t,\mathbf{p})Q(t,\mathbf{p}), \tag{6.64b}$$

$$EP'(t,\mathbf{p}) = k_3 ES_1 S_2(t,\mathbf{p}) - r_3 EP(t,\mathbf{p})Q(t,\mathbf{p}) - k_4 EP(t,\mathbf{p}) + r_4 E(t)P(t), \tag{6.64c}$$

$$Q'(t,\mathbf{p}) = k_3 ES_1 S_2(t,\mathbf{p}) - r_3 EP(t,\mathbf{p})Q(t,\mathbf{p}), \tag{6.64d}$$

$$P'(t,\mathbf{p}) = k_4 EP(t,\mathbf{p}) - r_4 E(t)P(t). \tag{6.64e}$$

Each reaction mechanism had a different unknown parameter vector:

$$\mathbf{p}_a = (k_1, k_2, k_3, k_4, r_1, r_2)^T, \tag{6.65}$$

$$\mathbf{p}_b = (k_1, k_2, k_3, k_4, r_1, r_2, r_3)^T, \tag{6.66}$$

$$\mathbf{p}_c = (k_1, k_2, k_3, k_4, r_1, r_2, r_4)^T, \tag{6.67}$$

$$\mathbf{p}_d = (k_1, k_2, k_3, k_4, r_1, r_2, r_3, r_4)^T, \tag{6.68}$$

the subscripts indicating to which mechanism the vector corresponds. Eight output functions corresponding to individual measurement of each reaction species, were analysed:

$$\left.\begin{aligned} &\mathbf{y}_1(t,\mathbf{p}) = E(t,\mathbf{p}),\ \mathbf{y}_2(t,\mathbf{p}) = S_1(t,\mathbf{p}),\ \mathbf{y}_3(t,\mathbf{p}) = S_2(t,\mathbf{p}),\\ &\mathbf{y}_4(t,\mathbf{p}) = ES_1(t,\mathbf{p}),\ \mathbf{y}_5(t,\mathbf{p}) = ES_1S_2(t,\mathbf{p}),\ \mathbf{y}_6(t,\mathbf{p}) = Q(t,\mathbf{p}),\\ &\mathbf{y}_7(t,\mathbf{p}) = EP(t,\mathbf{p}),\ \mathbf{y}_8(t,\mathbf{p}) = P(t,\mathbf{p}). \end{aligned}\right\} \tag{6.69}$$

For Mechanisms (a) and (b) all of these observations were shown to be at least locally identifiable by PRT. This test proved computationally intractable for y_7 to y_8 for Mechanism (c) and y_6 to y_8 for Mechanism (d). The remaining observations were shown to be at least locally identifiable by PRT.

Full Taylor series analyses proved computationally intractable for y_3 to y_8 for Mechanisms (a) and (b) and all output functions except y_2 for Mechanisms (c) and (d). For all mechanisms k_1 was shown to be globally identifiable for observation of ES_1 from the first non-zero Taylor

series coefficient of this expansion, equivalent up to a change of sign to Equation (6.71) below. In addition for Mechanisms (c) and (d) it was possible to show that k_1, r_1 and k_2 are globally identifiable for observation of E, using three of the Taylor series coefficients, Equations (6.71)-(6.73), below.

For y_1 to y_2 for Mechanisms (a) and (b) and y_2 for Mechanisms (c) and (d) it was possible to complete a full Taylor series analysis. The first four Taylor series coefficients for each mechanism are:

$$y_1(0) = E_0, y_2(0) = S_{10}, \tag{6.70}$$

$$y_1^{(1)}(0) = y_2^{(1)}(0) = -k_1 E_0 S_{10}, \tag{6.71}$$

$$y_1^{(2)}(0) = y_2^{(2)}(0) = k_1 E_0 S_{10}(k_1(E_0 + S_{10}) + r_1), \tag{6.72}$$

$$\begin{aligned} y_1^{(3)}(0) = y_2^{(3)}(0) = \\ = -k_1 E_0 S_{10}((k_1(E_0 + S_{10}) + r_1)(k_1(E_0 + S_{10}) + r_1) + 2k_1^2 E_0 S_{10} + k_2 r_1 S_{20}). \end{aligned} \tag{6.73}$$

The Taylor series coefficients were analysed as previously described using the alternative unknown parameter vectors:

$$\overline{\mathbf{p}}_a = (\overline{k_1}, \overline{k_2}, \overline{k_3}, \overline{k_4}, \overline{r_1}, \overline{r_2})^T, \tag{6.74}$$

$$\overline{\mathbf{p}}_b = (\overline{k_1}, \overline{k_2}, \overline{k_3}, \overline{k_4}, \overline{r_1}, \overline{r_2}, \overline{r_3})^T, \tag{6.75}$$

$$\overline{\mathbf{p}}_c = (\overline{k_1}, \overline{k_2}, \overline{k_3}, \overline{k_4}, \overline{r_1}, \overline{r_2}, \overline{r_4})^T, \tag{6.76}$$

$$\overline{\mathbf{p}}_d = (\overline{k_1}, \overline{k_2}, \overline{k_3}, \overline{k_4}, \overline{r_1}, \overline{r_2}, \overline{r_3}, \overline{r_4})^T, \tag{6.77}$$

showing that in these cases $\overline{\mathbf{p}} = \mathbf{p}$. As such these observations are SGI.

As for the previous model it is possible to consider an alternative experimental setup designed

to study the reverse reaction. Appropriate initial conditions are as follows:

$$\mathbf{x}_0 = (E_0, 0, 0, 0, 0, Q_0, 0, P_0). \tag{6.78}$$

For Mechanisms (a) and (b) the expansions of y_6 to y_8 yielded at most one non-zero Taylor series coefficient, which contained no unknown parameters. As such no further analysis was possible for these initial conditions. For Mechanism (c) two parameters, k_4 and r_4 were shown to be globally identifiable for observation of EP and P. The Taylor series coefficients used were equivalent up to a change of sign to those below, Equations (6.80)-(6.81). Measurement of Q yielded a single non-zero Taylor series coefficient which contained no unknown parameters. For Mechanism (d) the observations y_6 to y_8 were shown to be at least locally identifiable using PRT. Further analysis proved computationally intractable for y_6. The parameter r_4 was shown to be SGI for y_7 using the second Taylor series coefficient of its expansion, which is equivalent to Equation (6.80).

For measurement of P a full Taylor series analysis was completed. The first four Taylor series coefficients are below:

$$y_8(0) = P_0, \tag{6.79}$$

$$y_8^{(1)} = -r_4 E_0 P_0, \tag{6.80}$$

$$y_8^{(2)} = r_4 E_0 P_0 (r_4(E_0 + P_0) + k_4), \tag{6.81}$$

$$y_8^{(3)}(0) = -r_4 E_0 P_0 (r_4(E_0 + P_0) + k_4)(r_4(E_0 + P_0) + k_4 + 2r_4^2 E_0 P_0 + k_4 r_3 Q_0). \tag{6.82}$$

Note that these coefficients are equivalent to those found for measurement of enzyme and first substrate, Equations (6.70)-(6.73), when relabelled as follows:

$$r_4 \leftrightarrow k_1,\ k_4 \leftrightarrow r_1,\ r_3 \leftrightarrow k_2,\ P_0 \leftrightarrow S_{10},\ Q_0 \leftrightarrow S_{20}. \tag{6.83}$$

The Taylor series coefficients were analysed using the alternative parameter vector defined above, Equation (6.77), as previously described. It was found that $\overline{\mathbf{p}} = \mathbf{p}$; thus with the observation of P this model is SGI.

Simple ordered model 2	Reversible product release stages			
Measurement	None	First	Second	Both
E	SGI	SGI	ASLI, $\{k_1, r_1, k_2\}$	ASLI, $\{k_1, r_1, k_2\}$
S_1	SGI	SGI	SGI	SGI
ES_1	ASLI, $\{k_1\}$	ASLI, $\{k_1\}$	ASLI, $\{k_1\}$	ASLI, $\{k_1\}$
S_2	ASLI	ASLI	ASLI	ASLI
ES_2	ASLI	ASLI	ASLI	ASLI
Q	SGI‡	ASLI	ASLI	ASLI†
EP	ASLI	ASLI	CI, $\{k_4, r_4\}^{\dagger}$	ASLI†, $\{r_4\}^{\dagger}$
P	SGI‡	SGI‡	CI, $\{k_4, r_4\}^{\dagger}$	SGI†

Table 6.3: Summary of the structural identifiability analysis results for the two substrate transient simple ordered model with sequential product release. Results are either: SGI, structurally globally identifiable; ASLI, at least structurally locally identifiable; or CI, analysis was computationally intractable. In addition a subset of globally identifiable parameters are stated in some cases. Results marked by a † were obtained for the alternative initial conditions, Equation (6.78). Results marked by a ‡ can be obtained using the input-output relationship approach, and require only that enzyme and substrate concentrations be initially non-zero.

The results of this analysis are summarised in Table 6.3. It was not possible to determine whether with observation of either product the model was SGI for Mechanisms (a)-(c), as such an input-output relationship analysis was attempted. Mechanism (d) was also analysed to provide basis for comparison. However derivation of the necessary output functions for Mechanisms (c) and (d) proved computationally intractable. The results for Mechanisms (a) and (b) are presented in the following sections.

6.2.3.1 *Differential algebra analysis for Mechanism (a) of the second simple ordered model*

The following expressions are the generators of a differential ideal equivalent to Mechanism (a), dependence on $\mathbf{p}$ is not explicitly noted for notational convenience:

$$\frac{dE(t)}{dt} + k_1E(t)S_1(t) - r_1ES_1(t) - k_4EP(t) = 0, \tag{6.84a}$$

$$\frac{dS_1(t)}{dt} + k_1E(t)S_1(t) - r_1ES_1(t) = 0, \tag{6.84b}$$

$$\frac{dES_1(t)}{dt} - k_1E(t)S_1 + r_1ES_1 + k_2ES_1(t)S_2(t) - r_2ES_1S_2(t) = 0, \tag{6.84c}$$

$$\frac{dS_2(t)}{dt} + k_2ES_1(t)S_2(t) - r_2ES_1S_2(t) = 0, \tag{6.84d}$$

$$\frac{dES_1S_2(t)}{dt} - k_2ES_1(t)S_2(t) + (r_2 + k_3)ES_1S_2(t) = 0, \tag{6.84e}$$

$$\frac{dEP(t)}{dt} - k_3ES_1S_2(t) = 0, \tag{6.84f}$$

$$\frac{dQ(t)}{dt} - k_3ES_1S_2(t) = 0, \tag{6.84g}$$

$$\frac{dP(t)}{dt} - k_4EP(t) = 0, \tag{6.84h}$$

$$E_0 - E(t) - ES_1(t) - ES_1S_2(t) - EP(t) = 0, \tag{6.84i}$$

$$S_{10} - S_1(t) - ES_1(t) - ES_1S_2(t) - EP(t) - P(t) = 0, \tag{6.84j}$$

$$S_{20} - S_2(t) - ES_1S_2(t) - EP(t) - P(t) = 0, \tag{6.84k}$$

$$Q(t) - EP(t) - P(t) = 0. \tag{6.84l}$$

Derivation of an output relation for y_1 to y_4 proved computationally intractable. For y_5 and y_8, the differential ideal was successfully decomposed into two ideals using the `Rosenfeld_Groebner` algorithm in *Maple*. However, for y_5 and y_6, no autoreduced subset of these generators could be decomposed using the algorithm. As such these measurements must also be considered computationally intractable. For y_8 the subset, Equations (6.84c) and (6.84e)-(6.84k), was used. It can be made orthonomic and autoreduced with respect to any ranking such that $\{E,\ S_1,\ S_2,\ y_8\} > \{ES_1,\ ES_1S_2,\ EP,\ P,\ Q\}$. The `Rosenfeld_Groebner` algorithm decomposes it into two ideals equivalent to those previously obtained. The left hand side of the generators

of the first of these ideals are:

$$\left.\begin{array}{l} ES_1S_2(t),\ Q^{(1)}(t),\ EP(t)-(S_{20}-y_8(t)),\ P(t)-y_8(t), \\ \\ y_8^{(1)}(t)-k_4(S_{20}-y_8(t)), \\ \\ ES_1^{(1)}(t)-k_1ES_1(t)^2-ES_1(t)(k_1(2S_{20}-E_0-S_{10}-y_8(t))-r_1)- \\ \qquad -k_1(S_{20}-E_0-y_8(t))(S_{20}-S_{10}). \end{array}\right\} \tag{6.85}$$

This ideal, as that found in Section 6.2.2.1, describes an end stage in the reaction mechanism. From Equation (6.84k) and the first and second expressions above, the concentrations of S_2 and ES_1S_2 are always zero. Thus the ideal corresponds to a situation where S_2 has been exhausted to create EP leaving two dynamic sub-reactions: one in which enzyme and the first substrate combine and dissociate; and another in which EP breaks down to form P. Thus this ideal does not describe the required reaction state and can be eliminated as a possibility where it is known that the concentration of the second substrate is non-zero.

Presenting the generators of the second ideal on paper is not feasible due to their size[7]. Several generators contain the denominator:

$$y^{(2)}(t)+(k_3+k_4)y^{(1)}(t)+k_3k_4(y(t)-S_{20}). \tag{6.86}$$

If this equation equals zero then the reaction state is described by the ideal above, Equation (6.85). This was shown by incorporating it into the set of generators and reapplying the `Rosenfeld_Groebner` algorithm. The single resulting ideal was that above. As such this denominator can be considered non-zero. With this constraint it is possible to eliminate the leader of the output relation, $y_8^{(4)}$, by substitution of the unknown parameter vector, Equation (6.74), and obtain a differential polynomial as described in Section 6.2.1.1. Since the genera-

[7] These generators and the following analysis are available on request (and at the following URL) in an electronic *Maple* file: http://www2.warwick.ac.uk/fac/sci/moac/currentstudents/2005/daniel_bearup.

tors are orthonomic and autoreduced the ideal is prime. Thus, by the argument presented in Section 3.2.3.1, the 79 monomials of this differential polynomial are linearly independent. The six coefficients of this polynomial below are sufficient to show that the observation is SGI:

$$k_2\overline{k_2}(k_3^2k_4^2\overline{k_2k_1} - k_1k_2\overline{k_3}^2\overline{k_4}^2), \tag{6.87a}$$

$$k_3^2k_4^2\overline{k_3}^2\overline{k_4}^2(k_2\overline{k_1} - k_1\overline{k_2}), \tag{6.87b}$$

$$k_2k_3^3k_4^3\overline{k_2k_3}^3\overline{k_4}^3(\overline{k_1k_2} - k_1k_2), \tag{6.87c}$$

$$k_2 * k_3^3 * k_4^3\overline{k_2k_3}^3\overline{k_4}^3(\overline{k_1k_2}(\overline{k_3} + \overline{k_4}) - k_1k_2(k_3 + k_4)), \tag{6.87d}$$

$$k_2k_3^3k_4^3\overline{k_2k_3}^3\overline{k_4}^3S_{20}^2\left(\begin{array}{c}(\overline{k_1} - k_1)(E_0 + S_{10}) + (\overline{k_2} - k_2)S_{20} + (\overline{k_3} - k_3) + \\ + (\overline{k_4} - k_4) + (\overline{r_1} - r_1) + (\overline{r_2} - r_2)\end{array}\right), \tag{6.87e}$$

$$k_2k_3k_4\overline{k_2k_3k_4}\left(\begin{array}{c}k_2\overline{k_3k_4}(k_1(E_0 + S_{10} + 2S_{20}) + k_3 + k_4) + 2k_1\overline{k_3k_4}(k_3 + k_4 + r_2) - \\ - k_3k_4\overline{k_2}(\overline{k_1}(E_0 + S_{10} + 2S_{20}) + \overline{k_3} + \overline{k_4}) - 2k_3k_4\overline{k_1}(\overline{k_3} + \overline{k_4} + \overline{r_2})\end{array}\right) +$$

$$+ k_2\overline{k_2}S_{20}(k_1k_2\overline{k_3}^3\overline{k_4}^3 - \overline{k_1k_2}k_3^3k_4^3). \tag{6.87f}$$

These coefficients were solved for the alternative parameter vector using *Maple* to obtain the following six solutions:

$$\{\overline{k_1} = 0, \overline{k_2} = 0\},\ \{\overline{k_1} = 0,\ \overline{k_3} = 0\},\ \{\overline{k_1} = 0,\ \overline{k_4} = 0\}, \tag{6.88a}$$

$$\{\overline{k_2} = 0,\ \overline{k_3} = 0\},\ \{\overline{k_2} = 0,\ \overline{k_4} = 0\}, \tag{6.88b}$$

$$\{\overline{k_1} = k_1,\ \overline{k_2} = k_2,\ \overline{k_3} = k_3,\ \overline{k_4} = k_4,\ \overline{r_1} = r_1,\ \overline{r_2} = r_2\}. \tag{6.88c}$$

Note that the same solutions are obtained when all coefficients of the polynomial are solved simultaneously. The first five solutions do not lie within the set of feasible parameter values, as such they are rejected. Thus a single solution remains, $\overline{\mathbf{p}} = \mathbf{p}$, showing that with this measurement the model is SGI.

For $\mathbf{y}_7$, denoted y_7, the subset, Equations (6.84c), (6.84e)-(6.84g) and (6.84i)-(6.84l), were used. They can be made orthonomic and autoreduced with respect to any ranking such that $\{E,\ S_1,\ S_2,\ P,\ y_8\} > \{ES_1,\ ES_1S_2,\ EP,\ Q\}$ since $-EP(t) - P(t)$ can be replaced with $Q(t)$ in Equations (6.84j) and (6.84k). The `Rosenfeld_Groebner` algorithm decomposes this subset into three ideals. The left hand sides of the generators of the first two ideals are:

$$\left.\begin{array}{l} ES_1(t),\ ES_1S_2(t),\ k_4EP(t) - P^{(1)}(t),\ Q(t) - S_{10}, \\ P^{(2)}(t) + k_4P^{(1)}(t),\ y_7(t) - S_{10} \end{array}\right\} \tag{6.89}$$

$$\left.\begin{array}{l} ES_1S_2(t),\ k_4EP(t) - P^{(1)}(t),\ Q(t) - S_{20},\ P^{(2)}(t) + k_4P^{(1)}(t),\ y_7(t) - S_{20}, \\ k_4ES_1^{(1)}(t) - k_1ES_1^2(t) + k_1(k_4E_0S_{20} + P^{(1)}(t)(S_{10} - S_{20})) + \\ + S_{20}ES_1(t)(k_1(k_4(E_0 + S_{10} - S_{20} - E_0S_{10}) - k_1P^{(1)}(t)) + r_1k_4). \end{array}\right\} \tag{6.90}$$

These ideals, as that above, describe possible end stages in the reaction mechanism. The first relation in both cases requires that the concentration of enzyme substrate complex is always zero. As such these ideals do not describe the required reaction state and can be eliminated as previously observed.

Presenting the generators of the third ideal on paper is not feasible due to their size[8]. The output relation contains a factorisable denominator; the two components of which are below:

$$y_7^{(1)}(t) + k_3(y_7(t) - S_{20}), \tag{6.91}$$

$$k_1k_2k_3\left(\begin{array}{c} S_{10}S_{20}k_2k_3^2 - k_2k_3^2y_7(t)(S_{10} + S_{20}) + k_3^2k_2y_7(t)^2 + 2k_2k_3y_7(t)y_7^{(1)}(t) + \\ + k_2(y_7^{(1)})^2 - k_3y_7^{(1)}(t)(k_2(S_{10} + S_{20}) + k_3 + r_2) - k_3y_7^{(2)} \end{array}\right). \tag{6.92}$$

When one or both of these expressions equal zero the reaction state is described by one of the

[8] These generators and the following analysis are available on request (and at the following URL) in an electronic *Maple* file: http://www2.warwick.ac.uk/fac/sci/moac/currentstudents/2005/daniel_bearup.

ideals above. This was shown by incorporating each into the set of generators and reapplying the `Rosenfeld_Groebner` algorithm resulting in one or both of the ideals above being obtained. As such these factors of the denominator can be considered non-zero. With this constraint it is possible to eliminate the leader of the output relation, $y_7^{(4)}(t)$, as described in Section 6.2.1.1. Since the generators are orthonomic and autoreduced the ideal is prime. Thus, by the argument presented in Section 3.2.3.1, the such the 387 monomials of the output relation are linearly independent. The coefficients of the polynomial were solved to obtain the following four solutions for the alternative parameters:

$$\{\overline{k_1} = 0,\ \overline{k_2} = 0\},\ \{\overline{k_1} = 0,\ \overline{k_3} = 0\},\ \{\overline{k_2} = 0,\ \overline{k_3} = 0\}, \tag{6.93a}$$

$$\{\overline{k_1} = k_1, \overline{k_2} = k_2, \overline{k_3} = k_3, \overline{k_4} = k_4, \overline{r_1} = r_1, \overline{r_2} = r_2\} \tag{6.93b}$$

The first three solutions do not lie on the set of feasible parameter values, as such they are rejected. Thus a single solution remains, $\overline{\mathbf{p}} = \mathbf{p}$, showing that with this observation the model is SGI.

Thus using the input-output relationship approach it is possible to show that observation of either product is SGI. This result is independent of initial conditions, as long as enzyme and substrate concentrations are non-zero. This can be guaranteed experimentally.

6.2.3.2 Differential algebra analysis of Mechanism (b) for the second simple ordered model

The generators of the differential ideal corresponding to Mechanism (b) are Equations (6.84a)-(6.84d), (6.84h)-(6.84l) and the following:

$$\frac{ES_12(t)}{dt} - k_2 ES_1(t)S_2(t) + (r_2 + k_3)ES_1S_2(t) - r_3 EP(t)Q(t) = 0, \tag{6.94a}$$

$$\frac{P(t)}{dt} - k_3 ES_1S_2(t) + r_3 EP(t)Q(t) = 0, \tag{6.94b}$$

$$\frac{dQ(t)}{dt} - k_3 ES_12(t) + r_3 EP(t)Q(t) = 0. \tag{6.94c}$$

Derivation of an output relation for y_1 to y_7 proved computationally intractable. For y_8, the differential ideal was successfully decomposed into two ideals using the `Rosenfeld_Groebner` algorithm in *Maple*. For y_8 the subset, Equations (6.84c), (6.94a)-(6.94b) and (6.84h)-(6.84l), was used. It can be made orthonomic and autoreduced with respect to any ranking such that $\{E,\ S_1,\ S_2,\ Q,\ y_8\} > \{ES_1,\ ES_1S_2,\ EP,\ P\}$. The `Rosenfeld_Groebner` algorithm decomposes it into two ideals equivalent to those previously obtained. The left hand side of the generators of the first of these ideals are:

$$\left.\begin{array}{c} E(t) + ES_1(t) - E_0,\ S_1(t) + ES_1(t) + (S_{20} - S_{10}),\ S_2(t),\ ES_1S_2(t),\ EP(t), \\ Q(t) - S_{20}, P(t) - S_{20}, y_8(t) - S_{20}, \\ ES_1^{(1)}(t) + k_1E_0(S_{20} - S_{10}) + ES_1(t)(k_1(E_0 + S_{10} - S_{20}) + r_1 - k_1ES_1(t)), \end{array}\right\} \quad (6.95)$$

From the third generator above, this ideal corresponds to a reaction state where the second substrate has been exhausted. The non-constant generators above indicate association and dissociation of the first substrate from the enzyme. As such this ideal does not describe the required reaction state and can be eliminated as previously observed in Section 6.2.2.1.

Presenting the generators of the third ideal on paper is not feasible due to their size[9]. Several generators have the following denominator:

$$k_2k_3k_4^2(k_4y_8^{(2)}(t) + k_4y_8^{(1)}(t)(k_3(1 - S_{20}) + k_4 + r_3y_8(t)) + r_3(y_8^{(1)}(t))^2 + k_3k_4^2y_8(t)) \quad (6.96)$$

When this equation was included as one of the generators of the differential ideal the `Rosenfeld_Groebner` algorithm generated only one ideal, that shown above Equations (6.95). As such this equation can be considered non-zero. With this constraint it is possible to eliminate the leader of the output relation, $y_8^{(4)}$, as described in Section 6.2.1.1. Since the generators are

[9] These generators and the following analysis are available on request (and at the following URL) in an electronic *Maple* file: http://www2.warwick.ac.uk/fac/sci/moac/currentstudents/2005/daniel_bearup.

orthonomic and autoreduced the ideal is prime. Thus, by the argument presented in Section 3.2.3.1, the 171 coefficients of the differential polynomial are linearly independent. The seven coefficients of the polynomial listed below are sufficient to show that with this observation the model is SGI:

$$k_3 k_4^6 \overline{k_3}^2 \overline{k_4}^6 (k_2 \overline{k_1} - \overline{k_2} k_1), \tag{6.97a}$$

$$k_2 k_4^4 \overline{k_2 k_4}^4 (k_3^2 k_4^2 \overline{k_1 k_2} - k_1 k_2 \overline{k_3}^2 \overline{k_4}^2), \tag{6.97b}$$

$$k_2 r_3 \overline{k_2}\,\overline{r_3} (k_3^2 k_4^5 \overline{k_1 k_2}\,\overline{r_3}^3 - k_1 k_2 r_3^3 \overline{k_3}^2 \overline{k_4}^5), \tag{6.97c}$$

$$k_2 k_3^3 k_4^7 \overline{k_2 k_3}^3 \overline{k_4}^7 E_0 (k_1 k_2 k_3 k_4 - \overline{k_1 k_2 k_3 k_4}), \tag{6.97d}$$

$$k_2 k_4^4 r_3 \overline{k_2 k_4}^4 \overline{r_3} (k_3^2 k_4^2 \overline{k_1 k_2}\,\overline{r_3}^3 - k_1 k_2 r_3^3 \overline{k_3}^2 \overline{k_4}^2), \tag{6.97e}$$

$$k_2 k_3 k_4^7 \overline{k_2 k_3}^3 \overline{k_4}^7 S_{20}^2 \left(\begin{array}{c} (E_0 + S_{10})(\overline{k_1} - k_1) + S_{20}(\overline{k_2} - k_2) + (\overline{k_3} - k_3) + \\ + (\overline{k_4} - k_4) + (\overline{r_1} - r_1) + (\overline{r_2} - r_2) \end{array} \right), \tag{6.97f}$$

$$k_2 k_4^7 k_3^3 \overline{k_2 k_4}^7 \overline{k_3}^3 \left(\begin{array}{c} (k_1 r_2 r_3 - \overline{k_1}\,\overline{r_2 r_3}) + (3(E_0 + S_{20}) + S_{10})(k_1 k_2 r_3 - \overline{k_1 k_2}\,\overline{r_3}) + \\ + (\overline{k_1 k_2 k_3} - k_1 k_2 k_3) + (\overline{k_1 k_2 k_4} - k_1 k_2 k_4) \end{array} \right) +$$

$$+ k_2 k_3^2 k_4^6 \overline{k_2 k_3}^2 \overline{k_4}^6 E_0 (k_1 k_2 k_3^2 k_4^2 \overline{r_3} - r_3 \overline{k_1 k_2 k_3}^2 \overline{k_4}^2). \tag{6.97g}$$

These coefficients were solved for the alternative parameter vector in *Maple* to obtain the following seven solutions:

$$\{\overline{k_1} = 0,\ \overline{k_2} = 0\},\ \{\overline{k_1} = 0,\ \overline{k_3} = 0\},\ \{\overline{k_1} = 0,\ \overline{k_4} = 0\}, \tag{6.98a}$$

$$\{\overline{k_2} = 0,\ \overline{k_3} = 0\},\ \{\overline{k_2} = 0,\ \overline{k_4} = 0\},\ \{\overline{k_4} = 0,\ \overline{r_3} = 0\}, \tag{6.98b}$$

$$\{\overline{k_1} = k_1,\ \overline{k_2} = k_2,\ \overline{k_3} = k_3,\ \overline{k_4} = k_4,\ \overline{r_1} = r_1,\ \overline{r_2} = r_2,\ \overline{r_3} = r_3\}. \tag{6.98c}$$

Note that the same solutions are obtained when all coefficients of the polynomial are solved. The first six solutions do not lie in the set of feasible parameter values and are rejected. The

sole remaining solution is $\overline{\mathbf{p}} = \mathbf{p}$ showing that with observation of P the model is SGI. The result is independent of initial conditions under the conditions stated in the previous section.

6.3 Parameter estimation using pre-steady state data

In Sections 5.3.1 and 5.4.1 parameters were estimated from time course data collected over a steady state time frame. The analysis above considers transient models and as such is most applicable to parameter estimation based on data collected over a pre-steady state time scale. Using rapid mixing of reaction components specifically stopped-flow (Section 2.4.7) such time courses were collected for the reactions catalysed by MurB and Lactate dehydrogenase (LDH). It was not feasible to collect similar data for the reactions catalysed by MurA and MurC-F since they are monitored using coupled assays (Sections 5.3 and 5.5-5.7) and the MurE work of Anne Blewett [110]. Deconvolution of the required reaction kinetics from those of the coupled assays may be possible, but was not considered in this project. The reactions catalysed by MurB (Figure 2.4) and LDH (Figure 2.1) were monitored by measurement of the change of absorbance due to oxidation of NADPH or NADH respectively. Since these species are substrates of their respective enzymes no deconvolution was required in this case.

The kinetics of LDH are not directly relevant to the overarching problems investigated in this project. It is however available in large amounts and as such was used to test the viability of the proposed techniques. The results of this work are presented in the next section. The results of experiments using MurB are then presented. These results are not comprehensive. The scarcity of MurB limited the range of initial conditions which could be considered. Furthermore the absorbance change in the LDH reaction is not ideally placed for application of this technique.

6.3.1 Lactate dehydrogenase

Lactate dehydrogenase has been extensively studied and is fully kinetically characterised [180]. As such it is an ideal test-bed for development of direct parameter estimation from stopped flow data, since the mechanism is already well understood and results can be compared to those previously obtained.

LDH catalyses the conversion of NADH and pyruvate to lactate and NAD^+ via an ordered bi bi mechanism [181] given by:

$$\mathrm{E+NADH} \leftrightharpoons \mathrm{E^{NADH}} + \mathrm{Pyruvate} \leftrightharpoons \mathrm{E^{NADH}}_{\mathrm{Pyr}} \leftrightharpoons \mathrm{E^{NAD^+}}_{\mathrm{Lac}} \leftrightharpoons \mathrm{E^{NAD^+}} + \mathrm{Lactate} \leftrightharpoons \mathrm{E+NAD^+}$$

This mechanism is slightly more complex than any that have been modelled here however the simple ordered mechanism with sequential reversible product release is a reasonably close approximation. The primary characterisation techniques used were stopped flow absorbance and fluorescence spectroscopy [180], although a number of other techniques including isothermal calorimetry were also used [182]. Characterisation proceeded by dissection of the mechanism, that is experimental conditions were chosen so as to limit which reaction steps would influence the measured data [183, 184].

Lactate dehydrogenase was obtained from Sigma-Aldrich (Sigma-Aldrich, St Louis, USA). Three sets of initial conditions were used: 15μM NADH, pyruvate and LDH (1); 20μM NADH and pyruvate, 2.5μM LDH (2); and 140μM NADH and pyruvate, 2.5μM LDH(3). For each set of initial conditions between eight and fifteen individual time courses were collected; the mean concentration at each time point was calculated from these sets. Reactions were monitored by following change of absorbance due to NADH oxidation. Initially data collection was undertaken at 37°C. However, under these conditions, measurements indicated that all NADH had been oxidised within the 2.5ms dead time of the instrument. The temperature was reduced to 20°C and viable time courses were obtained. Parameter estimation using these time courses and the

Levenburg-Marquardt algorithm, in *Copasi* [159], was attempted as described in Section 3.4. Typical results are presented in Figure 6.1.

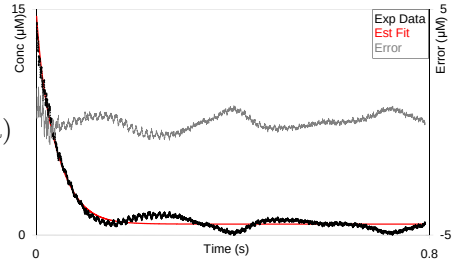
It is known that the absorbance change associated with NADH occurs at the central step of the reaction mechanism [180]. This corresponds to measurement of the concentration of free NADH and all enzyme-NADH complexes. This is not one of the measurement possibilities which has been analysed. Consequently parameter estimation was undertaken assuming that substrate concentration could be directly measured for the model with sequential product release. Since this model is SGI, parameter estimation can be expected to yield meaningful results.

However the results indicate that none of the time courses measured reliable estimation of a significant number of the parameters (Figure 6.1). In all cases the standard errors associated with most of the parameter estimates is greater than the estimated value of the parameter. Which parameters can be estimated to a reasonable degree of accuracy appears to be partially dependent on the initial conditions used. The early parameters being best determined when similar concentrations of enzyme and substrate are used, for example.

Time courses simulated using parameters obtained from the available literature did not correspond well to the experimental data obtained [180, 184, 185]. Simulations of parameters based on the fits obtained in this work suffered from a similar problem. As such this discrepancy is most likely the result of assuming that the absorbance change occurs at the first step of the mechanism. Consequently it would be useful to construct a model corresponding to the experimental case, where the concentration of NADH and all complexes containing it are monitored as one value. This has not, as yet, been attempted.

The model used here assumes that concentration of free NADH can be measured. Given that a fluorescence change occurs on binding of NADH to LDH this is viable [180]. As such this case was studied further using simulated data. Time courses were simulated using the following parameter vector $\mathbf{p} = (k_1, r_1, k_2, r_2, k_3, r_3, k_4, r_4)^T =$

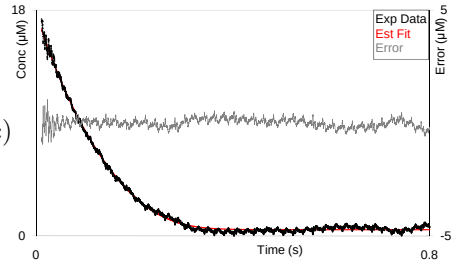
(a)



(b)

Parameter	EV	RMSE	Fit Stats
k_1 ($\mu M^{-1}s^{-1}$)	2.23	0.060	RMSE
r_1 (s^{-1})	403	2700	0.336
k_2 ($\mu M^{-1}s^{-1}$)	10000	29000	SD
r_2 (s^{-1})	344	2700	0.336
k_3 (s^{-1})	64.5	450	
r_3 ($\mu M^{-1}s^{-1}$)	4.26	5.2	
k_4 (s^{-1})	69.1	490	
r_4 ($\mu M^{-1}s^{-1}$)	0.620	2.8	

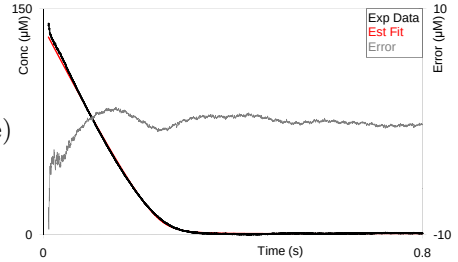
(c)



(d)

Parameter	EV	RMSE	Fit Stats
k_1 ($\mu M^{-1}s^{-1}$)	125	100	RMSE
r_1 (s^{-1})	321	675	0.199
k_2 ($\mu M^{-1}s^{-1}$)	10000	8800	SD
r_2 (s^{-1})	75.3	137	0.199
k_3 (s^{-1})	58.5	2.26	
r_3 ($\mu M^{-1}s^{-1}$)	37.7	17	
k_4 (s^{-1})	536	270	
r_4 ($\mu M^{-1}s^{-1}$)	39.1	32	

(e)

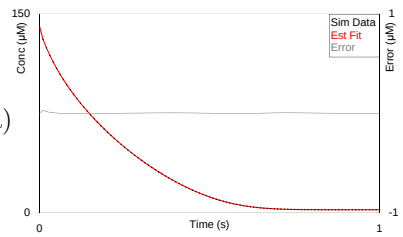


(f)

Parameter	EV	RMSE	Fit Stats
k_1 ($\mu M^{-1}s^{-1}$)	1000	1400	RMSE
r_1 (s^{-1})	3660	9300	1.00
k_2 ($\mu M^{-1}s^{-1}$)	1160	1700	SD
r_2 (s^{-1})	91.4	640	1.00
k_3 (s^{-1})	264	9.5	
r_3 ($\mu M^{-1}s^{-1}$)	0.1	0.52	
k_4 (s^{-1})	2580	861	
r_4 ($\mu M^{-1}s^{-1}$)	632	920	

Figure 6.1: Results of numerical fitting to pre-steady state time courses of the LDH catalysed reaction. The graphs (a), (c), and (e) plot experimental data (Exp Data), the fitted curve (Est Fit) and the associated error for the three sets of initial conditions used. The initial concentrations (1), (2) and (3) correspond to Graphs (a), (c) and (e) respectively. Tables (b), (d), and (f) correspond to Graphs (a), (c), and (e) respectively. The root mean square error, RMSE, and estimated standard deviation (SD) of the overall fit, the estimated parameter values (EV) and the RMSEs associated with these estimates are presented.

(a)



(b)

Parameter	EV	RMSE	Fit Stats
k_1 ($\mu M^{-1}s^{-1}$)	281	8.9	RMSE
r_1 (s^{-1})	95.5	1.3	4.37×10^{-3}
k_2 ($\mu M^{-1}s^{-1}$)	1760	47	SD
r_2 (s^{-1})	605	20	4.56×10^{-3}
k_3 (s^{-1})	834	25	
r_3 ($\mu M^{-1}s^{-1}$)	37	0.38	
k_4 (s^{-1})	555	11	
r_4 ($\mu M^{-1}s^{-1}$)	28.0	0.92	

Figure 6.2: Results of numerical fitting to simulated time courses of the LDH catalysed reaction. The graph (a) plots simulated data (Sim Data), the fitted curve (Est Fit) and the associated error. Initial conditions (3) were used. The RMSE and estimated standard deviation (SD) of the overall fit, the estimated parameter values (EV) and the RMSEs associated with these estimates are presented in Table (b).

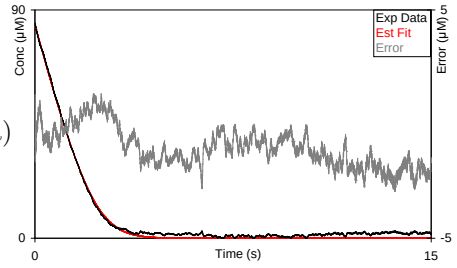
$(400, 400, 1000, 75, 1000, 40, 500, 40)^T$ and the initial conditions above. The resulting data were used for parameter estimation, an example of the results is presented in Figure 6.2. The results show a marked improvement in the reliability of the parameters estimated. However the estimated parameters differ, in some cases significantly, from the parameters used to generate the time course. This suggests that the time courses obtained may be insensitive to variations in some of the parameters.

The results presented here support no clear conclusions regarding the applicability of direct parameter estimation in the study of LDH. It is possible that if the correct model can be used absorbance data would provide sufficient information to characterise the enzyme. Further it is possible that appropriate initial conditions would improve the sensitivity of the model parameters to fluorescence time courses. Further work with this enzyme, and model, would provide further insight into the potential problems arising from using direct parameter estimation in the analysis of pre-steady state data.

6.4 MurB

MurB was over-expressed in *E. coli* and purified as previously described (Sections 2.3 and 5.2). Initial concentrations of NADPH and UDPPEE were both 90μM, that of MurB was 10μM, time courses were collected at 37°C of the change in absorbance due to NADPH. Four individual time courses were collected and a mean concentration was calculated for each time point. Parameter estimation using this time course and the Levenburg-Marquardt algorithm, in *Copasi* [159], was attempted. It is believed that a ping-pong model is the most appropriate for modelling MurB [85]. As such the ping-pong model with irreversible product release was used and it was assumed that NADPH is bound first. Since this model is SGI parameter estimation can be expected to yield meaningful results. The results are presented below (Figure 6.3).

(a)



(b)

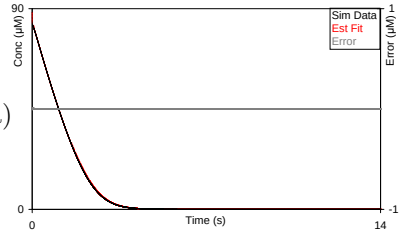
Parameter	EV	RMSE	Fit Stats
k_1 ($\mu M^{-1}s^{-1}$)	1.86	0.654	RMSE
r_1 (s^{-1})	61.7	27.4	1.33
k_2 (s^{-1})	9.16	1.08	SD
k_3 ($\mu M^{-1}s^{-1}$)	6.65	33.8	1.33
r_3 (s^{-1})	0.135	0.814	
k_4 (s^{-1})	10.7	1.15	

Figure 6.3: Results of numerical fitting to pre-steady state time courses of the MurB catalysed reaction. Graph (a) plots experimental data (Exp Data), the fitted curve (Est Fit) and the associated error. Table (b) contains the RMSE and estimated standard deviation (SD) of the overall fit, the estimated parameter values (EV) and the RMSEs associated with these estimates.

The RMSEs associated with two of the parameters, k_3 and r_3, exceed the estimated values of these parameters; as such these estimates are unreliable. The estimates for k_1 and r_1 are better, although lower relative RMSEs, such as those for k_2 and k_4 would be preferred. The steady state parameters k_{cat}, $k_{m,NADPH}$ and $k_{m,UDPPEE}$ can be estimated from these parameters using Equations (4.6) from Section 4.2.1. The resulting values of k_{cat} and $k_{m,NADPH}$ (4.94s^{-1} and 20.5μM respectively) are close to those estimated for the uninhibited model of MurB in Section 5.1. The estimated value of $k_{m,UDPPEE}$ (0.75μM) does not correspond well to these

results. However given that this parameter is dependent on k_3 and r_3, the two parameters which least well determined, this discrepancy is to be expected.

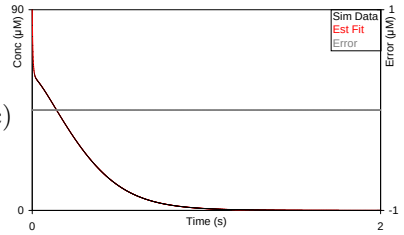
(a)



(b)

Parameter	EV	RMSE	Fit Stats
k_1 (μM^{-1}s^{-1})	2.01	1.55x10^{-4}	RMSE
r_1 (s^{-1})	61.1	6.47x10^{-3}	7.40x10^{-3}
k_2 (s^{-1})	9.10	9.69x10^{-4}	SD
k_3 (μM^{-1}s^{-1})	21.6	3.38x10^{-2}	7.40x10^{-3}
r_3 (s^{-1})	22.4	4.48x10^{-2}	
k_4 (s^{-1})	10.9	1.16x10^{-3}	

(c)



(d)

Parameter	EV	RMSE	Fit Stats
k_1 (μM^{-1}s^{-1})	2.00	2.90x10^{-6}	RMSE
r_1 (s^{-1})	60.0	1.40x10^{-4}	5.34x10^{-5}
k_2 (s^{-1})	9.00	1.37x10^{-5}	SD
k_3 (μM^{-1}s^{-1})	7.04	1.18x10^{-3}	5.35x10^{-5}
r_3 (s^{-1})	0.863	5.37x10^{-4}	
k_4 (s^{-1})	11.0	3.49x10^{-4}	

Figure 6.4: Results of numerical fitting to simulated time courses of the MurB catalysed reaction. Graphs (a) and (c) plot simulated data (Sim Data), the fitted curve (Est Fit) and the associated error for two sets of initial conditions. For Graph (a) initial concentrations of substrates and enzyme were 90μM and 10μM respectively. For Graph (b) the concentration of enzyme was increased to 45μM. Tables (b) and (d) correspond to Graphs (a) and (c) respectively and contain the RMSE and SD of the overall fits, the estimated parameter values (EV) and the RMSEs associated with these estimates.

Time courses were simulated using the parameter vector $\mathbf{p} = (k_1, r_1, k_2, k_3, r_3, k_4)^T = (2, 60, 9, 7, 1, 11)^T$. This parameter vector was obtained from the quasi-steady state estimates using Equations (4.6). Two sets of initial conditions were used, one as previously described, the other with initial enzyme concentration increased to 45μM. The parameter estimates based on these data are summarised in Figure 6.4. For both sets of initial conditions the RMSEs of the estimated values are low. For the initial conditions used experimentally the estimated values of k_3 and r_3 are not close to the expected values although the remaining parameters are (Figure 6.4(b)). Note that these parameters are also poorly determined from the experimental data.

For the time course corresponding to increased enzyme concentration all parameters are close to their expected values. Thus further experiments, using a similar substrate to enzyme ratio, may allow the estimation of these parameters.

6.5 Summary

In this chapter structural identifiability analyses of several transient two substrate models were presented. The models used were derived from those previously introduced in Section 4.2.1 by modification of the product release stages. All single species measurements were considered. In general the input-output relationship approach, Section 3.2.3, was ineffective with these models. The exceptions, where an input-output relationship analysis could be completed, occurred when the species measured was irreversibly released from the reaction. In all other cases it was impossible to derive an output relation using the differential algebra package in *Maple*.

By contrast the Taylor series approach produced solutions for a number of measurements. Successful analyses were typically obtained where the measured species either had a non-zero initial concentration or was one reaction step from species with this property. In addition for more complex reaction schemes the first derivative of the output function was relatively simple. The first of these observations informed the choice of alternative initial conditions used in a number of cases where analysis proved otherwise intractable. Use of non-zero initial concentrations of species in the proximity of the measurement allowed additional progress in some cases. It may be possible, with further study, to create a rigorous scheme using these considerations to predict which combinations of measurement functions and initial conditions are likely to prove computationally intractable.

Pre-steady state time courses for MurB and LDH were measured using absorbance spectroscopy and stopped flow techniques. Parameter estimation based on these time courses yielded some success, although additional work is necessary in both cases. The MurB results are especially

encouraging as some of the parameters estimated correspond well to the steady state parameters previously determined in Chapter 5. In both cases simulated data were used to investigate which initial conditions were more likely to allow estimation of all reaction parameters. In MurB it was found that similar concentrations of MurB and substrates yielded the best parameter estimation results. This is supported by typical approaches to this sort of analysis.

The work undertaken using LDH highlights the potential problems faced in using the direct parameter estimation approach in this sort of analysis. Parameter estimates are likely to be inaccurate if the modelling of the measurement is not correct. Furthermore in more complex models the sensitivity of the time course to some parameters may be low for certain initial conditions. As such it is necessary to determine time course sensitivity to estimated parameters to ensure that these estimates are meaningful.

Analysis of the reactions catalysed by MurA and MurC-F is not currently possible due to obfuscating effects of the coupling reactions needed to monitor these reactions. However with further work it may be possible to analyze such measurements. The transient reaction kinetics of the coupling enzymes could be determined in isolation using the techniques described. These results could then be used, in conjunction with numerical integration software, to deconvolute the contributions of the coupling system from the kinetics of the enzyme under investigation.

As yet no indistinguishability analysis of the two substrate models has been attempted. Such an analysis is ultimately necessary if the techniques above are to be used more widely since it is necessary to know which model to use when undertaking parameter estimation. At present it is necessary to determine which model to use by additional experiments or consultation of the literature. It would be preferable to choose the model based on the experimental data used to estimate parameters. Thus investigation of this problem would be an interesting area for future work.

7. EXPERIMENTAL AND COMPUTATIONAL INVESTIGATION OF THE STEADY STATE PATHWAY FLUXES

7.1 Introduction

An understanding of the dynamics of the whole peptidoglycan biosynthesis pathway would be an invaluable tool in the development of antibiotics targeting this pathway. Measuring the concentrations of every species within the pathway would likely prove impractical, given the number of measurements required and, in the case of most species, the lack of a direct continuous assay. Instead it is preferable to develop an *in silico* model which accurately predicts the concentrations of species that can be more easily measured directly. Such a model could then be used to identify reactions in the pathway as targets for antibiotics and to predict the effects or combinations of inhibitors on the pathway.

In this chapter a model is developed of the first six steps of the peptidoglycan biosynthesis pathway. An uninhibited model is developed first and its predictions are compared to experimental data obtained using *S. pneumoniae* enzymes. It is also used to compare the dynamics of *E. coli* and *S. pneumoniae* pathways. Then two inhibitors of enzymes in the pathway are kinetically characterised. The model is then adapted to take these inhibitors into account and its predictions compared to experimental data. It is then used to investigate the effects of inhibition on concentrations of intermediates within the pathway and the interactions between the inhibitors.

7.2 Modelling of the uninhibited peptidoglycan biosynthesis pathway

The section of the peptidoglycan biosynthesis pathway investigated consists of six reactions catalysed by the enzymes MurA to MurF (Figure 7.1). Each of the reactions catalysed by a Mur enzyme was assumed to reach a quasi-steady state by the same reasoning as was presented in Section 4.2. Furthermore each reaction is approximated by a ping-pong mechanism since kinetic constants for more complex mechanisms have yet to be obtained. Note that the ping-pong model can be considered an upper bound for the behaviour of more complex mechanisms since the additional terms all act to slow the reaction rate. Certain cybernetic components were also incorporated, specifically the recycling of $NADP^+$ to NADPH and ADP to ATP. The resulting system of differential equations is given below:

$$\frac{d[\text{UGP}]}{dt} = \frac{d[\text{PEP}]}{dt} = -f_1^A([\text{PEP}],[\text{UGP}]), \tag{7.1a}$$

$$\frac{d[\text{UDPPEE}]}{dt} = f_1^A([\text{PEP}],[\text{UGP}]) - f_1^B([\text{NADPH}],[\text{UDPPEE}]), \tag{7.1b}$$

$$\frac{d[\text{NADPH}]}{dt} = -\frac{d[\text{NADP}^+]}{dt} = v_{\text{IDH}}[\text{NADP}^+] - f_1^B([\text{NADPH}],[\text{UDPPEE}]), \tag{7.1c}$$

$$\frac{d[\text{UMN}]}{dt} = f_1^B([\text{NADPH}],[\text{UDPPEE}]) - f_3^C([\text{ATP}],[\text{L-Ala}],[\text{UMN}]), \tag{7.1d}$$

$$\frac{d[\text{L-Ala}]}{dt} = -f_3^C([\text{ATP}],[\text{L-Ala}],[\text{UMN}]), \tag{7.1e}$$

$$\frac{d[\text{U1P}]}{dt} = f_3^C([\text{ATP}],[\text{L-Ala}],[\text{UMN}]) - f_3^D([\text{ATP}],[\text{D-Glu}],[\text{U1P}]), \tag{7.1f}$$

$$\frac{d[\text{D-Glu}]}{dt} = -f_3^D([\text{ATP}],[\text{D-Glu}],[\text{U1P}]), \tag{7.1g}$$

$$\frac{d[\text{U2P}]}{dt} = f_3^D([\text{ATP}],[\text{D-Glu}],[\text{U1P}]) - f_3^E([\text{ATP}],[\text{L-DA}],[\text{U2P}]), \tag{7.1h}$$

$$\frac{d[\text{L-DA}]}{dt} = -f_3^E([\text{ATP}],[\text{L-DA}],[\text{U2P}]), \tag{7.1i}$$

$$\frac{d[\text{U3P}]}{dt} = f_3^E([\text{ATP}],[\text{L-DA}],[\text{U2P}]) - f_3^F([\text{ATP}],[\text{D-Ala-D-Ala}],[\text{U3P}]), \tag{7.1j}$$

$$\frac{d[\text{U5P}]}{dt} = -\frac{d[\text{D-Ala-D-Ala}]}{dt} = f_3^F([\text{ATP}],[\text{D-Ala-D-Ala}],[\text{U3P}]), \tag{7.1k}$$

$$\frac{d[\text{ATP}]}{dt} = -\frac{d[\text{ADP}]}{dt} = v_{\text{PK/LDH}}[\text{ADP}] - \sum_{X\in\{C,D,E,F\}} f_3^X(Y_X), \tag{7.1l}$$

$$\frac{d[\mathrm{P}^+]}{dt} = f_1^A([\mathrm{PEP}],[\mathrm{UGP}]) + \sum_{X\in\{C,D,E,F\}} f_3^X(Y_X). \tag{7.1m}$$

The subscript associated with each f denotes the form of the expression used, while the superscript denotes which kinetic parameters are used, for example f_1^A indicates that expression f_1 (from Section 4.2) is used in conjunction with kinetic parameters for MurA. The parameter V_{max} was determined by multiplying the enzyme concentration by the k_{cat} for each enzyme. Each of the four amino-acid ligases has a slightly different combination of substrates; this combination is denoted Y_X in the last two equations for brevity. Abbreviations for substrates and products are defined as stated in the List of Abbreviations. The rates $v_{\mathrm{PK/LDH}}$ and v_{IDH} are used instead of mechanistic models of these enzymes. Values were chosen to ensure the recycling reactions are sufficiently fast. Initial conditions were chosen to correspond to those used experimentally.

Kinetic constants for MurE were obtained previously by Anne Blewett and are presented in Table 7.1 [110].

Enzyme	Parameter name	Value
MurE	k_{cat} (s^{-1})	73.5
	k_m (ATP) (μM)	191.2
	k_m (L-Lys) (μM)	498.6
	k_m (U2P) (μM)	13.4

Table 7.1: Kinetic data for *S. pneumoniae* MurE. These data were obtained from the thesis submitted by Anne Blewett [110].

The pathway can also be reconstructed *in vitro* using the purified enzymes used in Chapter 5 and monitored using the assays described in Section 2.4. In the following section experimental and simulated data are compared. Simulations using the constants found in Tables 5.1 and 7.1 did not correspond well to the experimental data obtained. Alternative k_{cat}s for MurA to MurF were estimated directly from time courses obtained at the enzyme concentrations used in the pathway reconstruction. The results for MurC to MurF are presented in the next section, those for MurA and MurB were presented in Sections 5.3.1 and 5.4.1. In the subsequent section

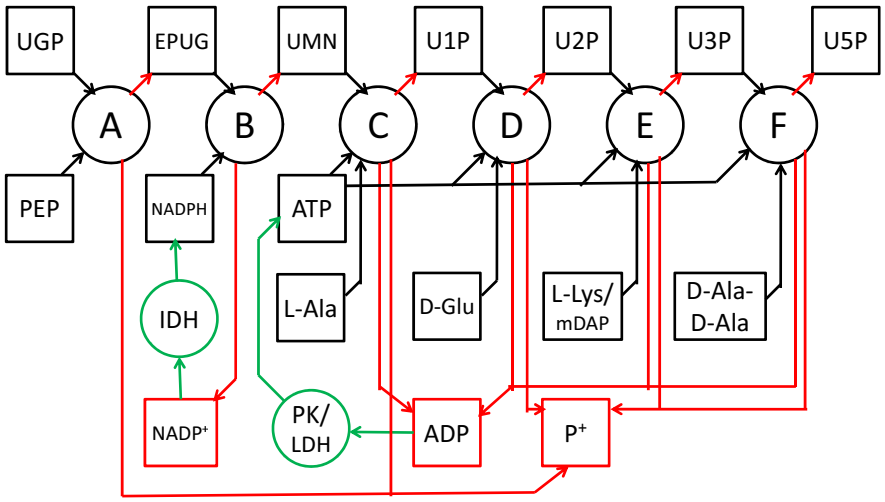


Figure 7.1: Schematic diagram of cytoplasmic phase of peptidoglycan biosynthesis. Enzyme catalysed reactions are denoted by circles, substrates and products are denoted by squares. Black arrows connect substrates to processes, red arrows connect processes to products. Green circles and arrows indicate a recycling process.

the results of simulations using these k_{cat}s are presented. They were found to compare well with experimental data. Finally using constants derived from the literature the dynamics of the *E. coli* pathway were simulated and compared to those obtained using the *S. pneumoniae* parameters.

7.2.1 Comparison of data simulated using parameters from Chapter 5 to experimental results

The peptidoglycan biosynthesis pathway was reconstructed *in vitro* using *S. pneumoniae* enzymes and monitored using an ADP release assay as described in Section 2.4.2. The concentration of UGP used was 50μM, so the full pathway would be expected to produce 200μM of ADP. The concentration of NADH used was increased to 300μM to ensure that it would remain in excess throughout the reaction. Each of the four amino acids were added to a concentration of 10mM. Initial concentrations of PEP, ATP and NADPH were 1.23mM, 200μM and 100μM respectively. Isocitrate and isocitrate dehydrogenase (IDH) were added to concentrations of 10mM and 0.18 units/ml respectively to ensure that NADPH oxidation was not measured.

Potassium chloride was also added to a final concentration of 100mM to provide the necessary cation to MurB [86].

Four *in vitro* pathways were constructed by truncating the natural pathway (Figure 1.2) at each of the amino-acid ligase steps. Thus the pathways used included each of the *S. pneumoniae* enzymes from MurA to MurC, MurA to MurD, MurA to MurE, and MurA to MurF. Initial enzyme concentrations were 0.1μM. MurA was added after an initial baseline had been recorded initiating the reaction pathway; a time course of approximately 40 minutes was then measured. The absorbance at 340nm of each reaction mixture was monitored in a Cary® 100 spectrometer. Three time courses were recorded for each pathway, the results were in close agreement.

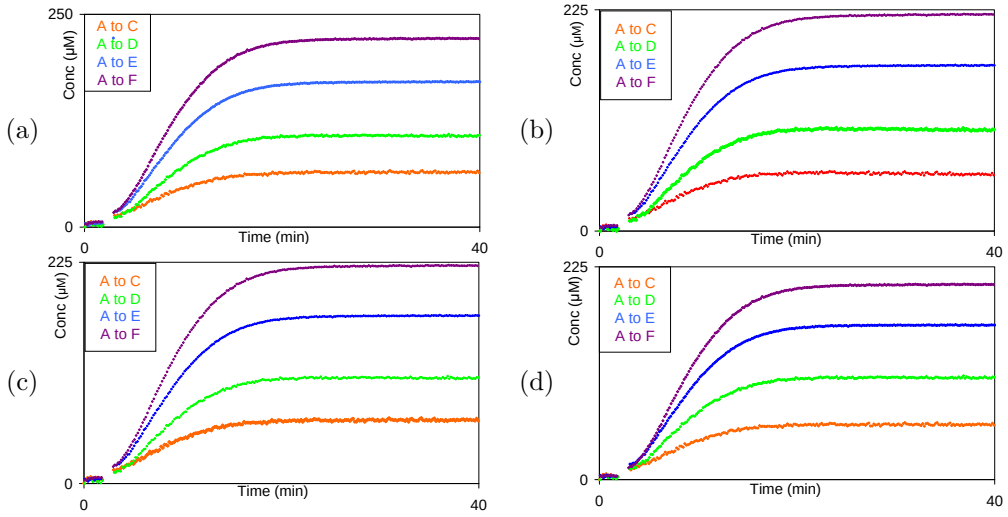
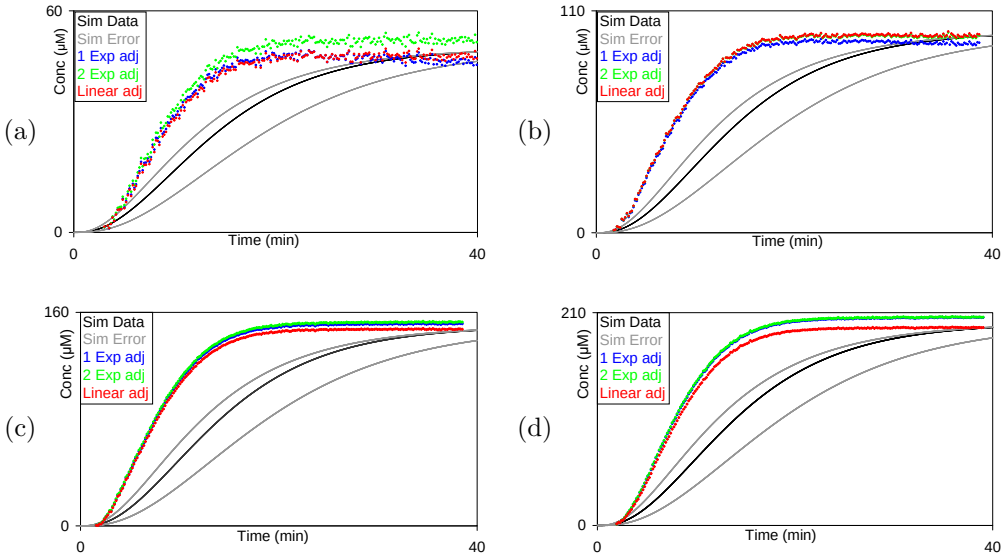


Figure 7.2: Production of NAD^+ by the reconstructed *S. pneumoniae* pathways. Concentration of NAD^+ produced (Conc) is plotted against time in minutes. In each graph four curves corresponding to the pathways MurA to MurC, MurA to MurD, MurA to MurE and MurA to MurF are plotted; they are colour coded according to the keys given. Graph (a) plots the raw data obtained. NAD^+ concentration was adjusted to account for base rate production using three models, single exponential, double exponential and linear. The results are plotted in Graphs (b)-(d) respectively.

It was observed that final NAD^+ concentration exceeded expected production, 50μM$\times$(number of amino acid ligases), for each pathway suggesting that some NAD^+ production occurs independent of the reaction pathway (Figure 7.2(a)). This supposition is reinforced by the non-zero rate of NAD^+ production observed before addition of MurA and after apparent exhaustion of



(e)

	RMS Error (μM)			
Adjust type	A-C	A-D	A-E	A-F
Single Exponential (1 Exp adj)	9.02	19.5	34.9	46.0
Double Exponential (2 Exp adj)	11.2	21.0	35.9	46.7
Linear (Linear adj)	8.54	21.1	32.3	39.4

Figure 7.3: Comparison of experimental time courses to those simulated using Model 7.2.1.1. Graphs (a)-(d) plot concentration of NAD^+ produced (Conc) against time in minutes. Graphs (a), (b), (c) and (d) correspond to the pathways A to C, A to D, A to E and A to F respectively. The adjusted experimental time courses are plotted in colour according to the keys provided. The simulated time course is plotted in black the grey curves indicating a confidence interval based on the confidence intervals of the individual parameters. Table (e) presents the root mean square error (RMS Error) for each of the pathways and adjustment functions used.

the UDP-intermediates. Three models of this additional NAD^+ production were considered, single exponential, double exponential and linear. These models are described in more detail in Appendix A. NAD^+ production predicted by these models was subtracted from the raw data (Figure 7.2(b)-(d)). Note that the apparent final NAD^+ concentration still exceeds expected production. However a jump in absorbance is observed in each time course during the interval where MurA was added which is not accounted for in the models of NAD^+ production used. This most likely results from mechanical error, i.e. movement of the cuvette resulting in a slightly modified absorbance profile, and was eliminated manually. The resulting time courses

are compared to simulated time courses in Figure 7.3.

Time courses were simulated in COPASI using differential equations above, Equations (7.1), and the parameters and initial conditions given in Table C.1, Appendix C. This is referred to as Model 7.2.1.1. Subpathways were modelled by the simple expedient of setting terms in Equations 7.1 corresponding to enzymes that were not in the subpathway to zero. Time courses were also simulated using the upper bounds of the k_{cat}s and lower bounds of k_ms and vice versa to form a confidence interval for the simulated time course.

The graphs presented in Figure 7.3(a)-(d) show that experimental NAD^+ production is significantly quicker than would be expected based on the simulated data. The RMSEs of these comparisons reflect this discrepancy (Figure 7.3(e)). They do not indicate that any particular model of background NAD^+ production should be preferred.

Since the structure of the mathematical model used, Equations (7.1), is a one to one representation of the processes known to be happening biologically, it is unlikely that this discrepancy arises from an issue in the model structure. The discrepancy can thus reasonably be attributed to the parameters used. A sensitivity analysis was undertaken for Model 7.2.1.1 in COPASI. The effect of a change to a single parameter on the concentrations of all reaction intermediates was determined numerically.

Since only $[NAD^+]$ is measured it is natural to focus its sensitivity, Table 7.2. Note it is most sensitive to changes in the V_{max}s of the various enzymes, especially to those in the early stages of the pathway. Sensitivity to k_m values tends to be relatively low although they become more significant in the later stages of the pathway where concentrations of the UDP intermediates are likely to be lower. Guided by these results new k_{cat} values were estimated for each of the enzymes used. The results are presented in Sections 5.3.1, 5.4.1 and 7.2.2.

V^A_{max}	$k^A_{m,\mathrm{UGP}}$	$k^A_{m,\mathrm{PEP}}$	v_{IDH}
1	8.6×10^{-4}	4.6×10^{-2}	5.5×10^{-6}
V^B_{max}	$k^B_{m,\mathrm{UDPPEE}}$	$k^B_{m,NADPH}$	$v_{\mathrm{PK/LDH}}$
0.64	4.4×10^{-4}	1.3×10^{-3}	4.9×10^{-5}
V^C_{max}	$k^C_{m,\mathrm{UMN}}$	$k^C_{m,\mathrm{ATP}}$	$k^C_{m,\mathrm{L\text{-}Ala}}$
0.31	1.5×10^{-3}	0.020	3.5×10^{-4}
V^D_{max}	$k^D_{m,\mathrm{U1P}}$	$k^D_{m,\mathrm{ATP}}$	$k^D_{m,\mathrm{D\text{-}Glu}}$
0.14	0.023	0.013	4.7×10^{-4}
V^E_{max}	$k^E_{m,\mathrm{U2P}}$	$k^E_{m,\mathrm{ATP}}$	$k^E_{m,\mathrm{L\text{-}Lys}}$
0.018	0.022	2.1×10^{-3}	7.2×10^{-4}
V^F_{max}	$k^F_{m,\mathrm{U3P}}$	$k^F_{m,\mathrm{ATP}}$	$k^F_{m,\mathrm{D\text{-}Ala\text{-}D\text{-}Ala}}$
0.018	0.058	0.044	6.6×10^{-3}

Table 7.2: Local sensitivity of NAD^+ concentration to model parameters calculated for Model 7.2.1.1. Sensitivities were calculated in COPASI. Absolute values were then taken and normalised relative to the highest value. The resulting values are presented below the parameter varied.

7.2.2 Estimation of k_{cat}s for MurC to MurF from progress curves

The discrepancies between experimental and simulated pathway dynamics necessitate new estimates for the kinetic parameters of the enzymes used. The principal observable error is one of overall rate and as such the discrepancy is most likely to be due to differences in the k_{cat}s. These parameters are estimated for MurC to MurF by numerical fitting at the concentrations of the enzymes used for the pathway assays described above. Estimates of this type for MurA and MurB are described in Sections 5.3 and 5.4. The experimental time courses used for these estimates were collected concurrently with the work described here.

Reaction mixtures were prepared containing the coupling components previously described, Section 2.4.2, for each of the enzymes MurC to MurF. The concentrations of UDP intermediates used were as follows: 105, 75, 80 and 65μM of UMN, U1P, U2P and U3P respectively. Concentrations of ATP and enzyme were 1mM and 0.1μM respectively. Reactions were initiated by addition of the required amino acid to a concentration of 10mM. Time courses were recorded as described in Section 2.4.2.

As previously noted the reactions were modelled using the quasi-steady state form of the three substrate ping-pong model, Equations 4.20. This is the same model as was used for the kinetic

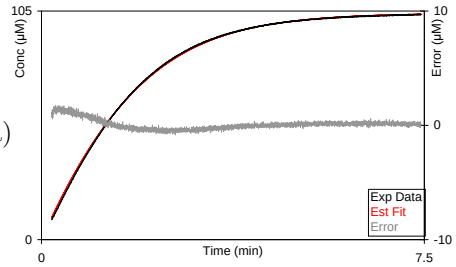
characterisations undertaken previously in Chapter 5. The initial concentration of the relevant UDP intermediate and k_{cat} of the reaction were estimated using the Levenburg-Marquardt algorithm in COPASI [159]. The remaining kinetic parameters, the k_ms with respect to the three substrates, were fixed to the values stated in Table 5.1 and Table 7.1. The structural identifiability analysis presented previously, Section 4.3.4, indicates that this model is structurally globally identifiable. As such parameter estimation can be expected to yield meaningful results. The results are presented below in Figure 7.4.

For MurC and MurF the fitted curves closely match the shape of the experimental data (Figure 7.4(a) and (g)), this is reflected by the low R^2 values for these fits. Thus the model used describes these reactions despite the lack of a complete kinetic characterisation. However the fits for MurD and MurE are less close (Figure 7.4(c) and (e)). For these two enzymes the k_ms used do not fully describe the dynamics of the reactions catalysed at the substrate concentrations used. A complete kinetic characterisation as outlined in Dixon [144] should allow these discrepancies to be eliminated. However, for the purposes of this analysis, the data and models will be as previously described.

The k_{cat}s estimated for MurD, MurE and MurF are lower than those previously obtained, Tables 5.1 and 7.1. The overall loss of activity is relatively small and probably results from time in storage. The effects of these, relatively minor, changes on pathway dynamics are noted in the following section. They cause an overall reduction in rate and as such are not responsible for the discrepancies observed in the previous section.

The estimated k_{cat} for MurC is a factor of four greater than that previously obtained. The enzyme solutions used here and in the initial characterisation, described in Section 5.5, were derived from the same stock, as such no immediate explanation for this discrepancy is forthcoming. It is possible that the initial characterisation was flawed; for example by loss of activity after dilution of the stock solution due to enzyme instability. Regardless this change is incorporated into the pathway model in the following section. It partially explains the observed

(a)

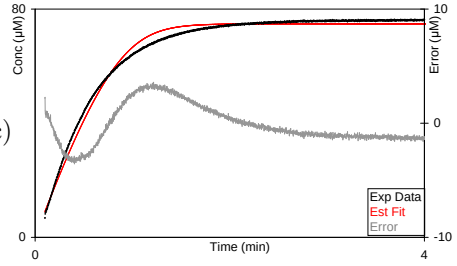


(b)

Fit Stats	RMSE	SD
	0.456	0.456

Parameter	EV	RMSE
k_{cat} (s^{-1})	26.6	$6.94x10^{-3}$
[UMN](0) (μM)	104	$8.58x10^{-3}$

(c)

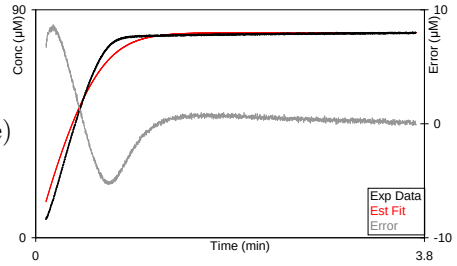


(d)

Fit Stats	RMSE	SD
	1.69	1.69

Parameter	EV	RMSE
k_{cat} (s^{-1})	21.5	$3.33x10^{-2}$
[U1P](0) (μM)	74.8	$2.96x10^{-2}$

(e)

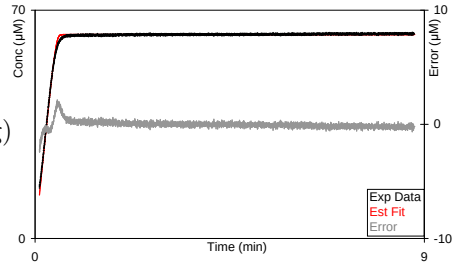


(f)

Fit Stats	RMSE	SD
	2.45	2.45

Parameter	EV	RMSE
k_{cat} (s^{-1})	63.1	0.163
[U2P](0) (μM)	81	$4.34x10^{-2}$

(g)



(h)

Fit Stats	RMSE	SD
	0.336	0.337

Parameter	EV	RMSE
k_{cat} (s^{-1})	23.9	$1.04x10^{-2}$
[U3P](0) (μM)	62.5	$3.398x10^{-3}$

Figure 7.4: Results of numerical fitting to time courses for the amino acid ligases. Graphs (a), (c), (e) and (g) plot the experimental data (Exp Data), the fitted curve (Est Fit) and the difference between these curves (Error) for MurC, MurD, MurE and MurF respectively. Tables (b), (d), (f) and (h) contain the overall fit statistics: root mean square error (RMSE) and estimated standard deviation (SD) for each fit, the estimated parameter values (EV), and the RMSE associated with these estimates.

Enzyme	k_{cat} (Chapter 5) (s^{-1})	k_{cat} (Direct PE) (s^{-1})
MurA	8.07	28.8
MurB	3.17	4.12
MurC	7.58	26.6
MurD	36.5	21.5
MurE	73.5	63.1
MurF	29.2	23.9

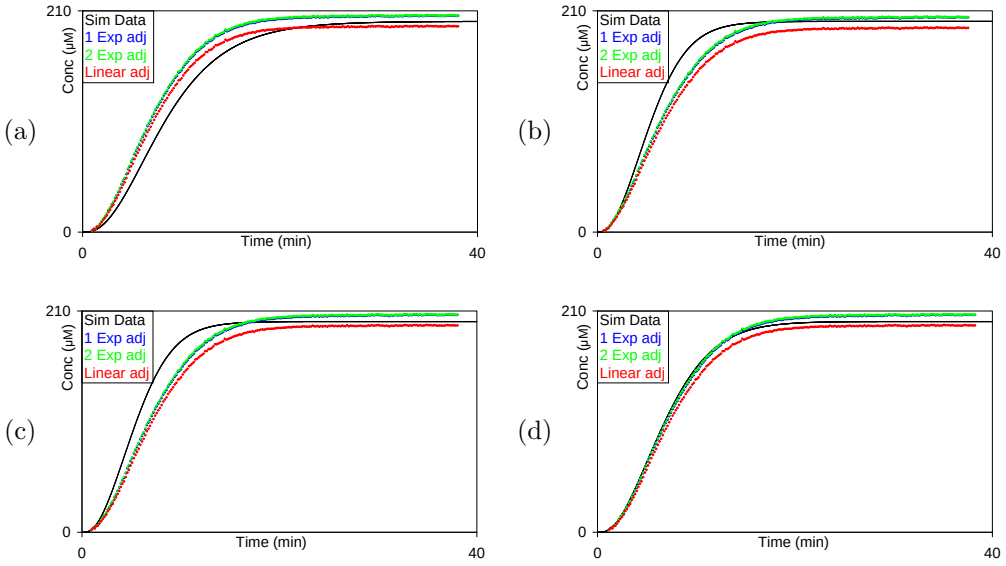
Table 7.3: Table of k_{cat}s obtained either using initial rate data, as described in Chapter 5, or by numerical parameter estimation from time course data (Direct PE).

discrepancy between experimental and simulated pathway dynamics.

7.2.3 Comparison of data simulated using new k_{cat}s to experimental results

A series of simulations were undertaken using COPASI using differential equations (7.1) and a combination of parameters from Tables 5.1, 7.1, and 7.3. Three cases were considered where k_{cat}s of the following subsets of enzymes: MurA and MurB; MurA to MurC; and MurA to MurF, were changed to the values found by direct parameter estimation, Models 7.2.3.1, 7.2.3.2, and 7.2.3.3 respectively. The k_m values used were unchanged from those used in the previous section. A fourth case, Model 7.2.3.4, was created from Model 7.2.3.3 by using the substrate inhibited parameters for MurB from Table 5.1. No change to the form of the equations given was necessary since the levels of substrate present were insufficient to create any significant inhibitory effect. The parameters used are presented in Tables C.2-C.5 in Appendix C. A comparison between simulated and experimental time courses for the full pathway, MurA to MurF, is presented below in Figure 7.5.

Simulated NAD^+ production in Model 7.2.3.1 remains slower than the experimental time course (Figure 7.5(a)). This model is however much closer than Model 7.2.1.1 as seen by the significantly reduced root mean square error and Figure 7.3. Once the k_{cat} of MurC is changed (Model 7.2.3.2) the rate of simulated NAD^+ production exceeds that of the experimental time



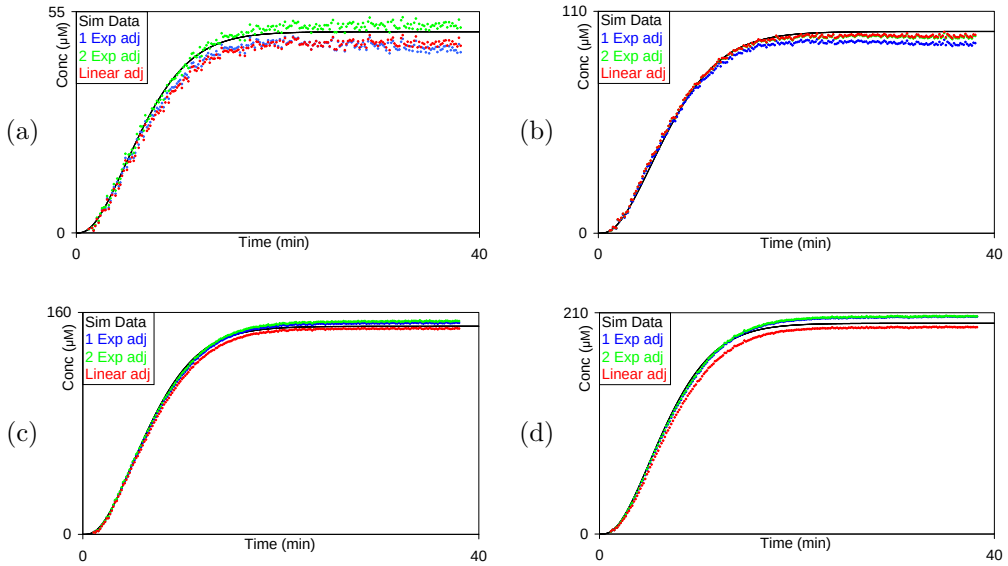
(e)

Model	Graph	Best RMSE (μM)
7.2.1.1	Figure 7.3(d)	39.4
7.2.3.1	(a)	9.20
7.2.3.2	(b)	11.1
7.2.3.3	(c)	13.3
7.2.3.4	(d)	4.63

Figure 7.5: Comparison of experimental time courses to those simulated using Models 7.2.3.1-4. Graphs (a)-(d) plot concentration of NAD^+ produced (Conc) against time in minutes for the pathway A to F. The adjusted experimental time courses are plotted in colour according to the keys provided. Graphs (a)-(d) correspond to Models 7.2.3.1-4 respectively. Table (e) presents the best root mean square errors for each of the graphs.

course (Figure 7.5(b)). While one might expect lower k_{cat}s for MurD to MurF (Model 7.2.3.3) to correct for this, they in fact make little difference (Figure 7.5(c)). The overall NAD^+ production is slightly reduced but the difference between these two cases is minimal, see Appendix B. The difference in RMS error between these two cases highlights the problem with manual elimination of the observed absorbance jump (Figure 7.5(e)). The simulated curve for Model 7.2.3.3 should lie slightly closer to the data, however it seems to produce a slightly higher RMS error. It would be preferable to minimise the RMS error computationally when comparing experimental and simulated data. The final case, where MurB is modelled using the substrate

inhibited k_ms and k_{cat} (Model 7.2.3.4) however compares very well to the experimental data (Figure 7.5(d)). Thus these are the parameters used for the remainder of this analysis (Table C.5). A more detailed comparison of the various cases considered can be found in Appendix B.



(e)

	RMS Error (μM)			
Adjust type	A-C	A-D	A-E	A-F
Single Exp	2.94	4.33	2.17	4.63
Double Exp	1.44	1.96	2.75	4.89
Linear	2.91	1.69	3.26	6.48

Figure 7.6: Comparison of experimental time courses to those simulated using Model 7.2.3.4. Graphs (a)-(d) and Table (e) are as those found in Figure 7.3; the confidence interval is omitted.

Comparisons of the simulation time courses for Model 7.2.3.4 to experimental time courses are presented in Figure 7.6. The RMS error values are consistently low (Figure 7.6(e)) and the simulated data are a close visual match to the experimental data. The adjustment of the experimental data for the A-D pathway require some modification, towards the end of the time course the concentration of NAD^+ drops suggesting that the predicted background NAD^+ production is too high. A similar issue exists for the single exponential and linear adjustments for the A-C pathway. The remaining adjustments seem to produce reasonable results. The double exponential adjustment seems to give the most consistent fit, given that it either has

the lowest or is close to the lowest RMS error.

Although ADP is only released by the last four reactions in the pathway the analysis presented here suggests that the dynamics of release are strongly influenced by the earlier MurA, MurB and MurC reactions. This is indicated by the relatively small effect observed by adjustment of the k_{cat}s of MurD to MurF compared to that observed when the k_{cat}s of MurA to MurC were altered. This difference is however partially explained by the relative size of these changes of k_{cat}.

Further note the relatively large change in NAD^+ production produced by changing the parameters used for MurB. While the k_{cat} used increased, the increases of k_m significantly slowed the turnover of that reaction, slowing the pathway as a whole. Both substrate concentrations for this reaction were low allowing the k_ms to have a much larger effect on pathway dynamics than might be expected.

Correctly accounting for background production of NAD^+ is a significant problem and requires further study. Ideally the use of additional assay techniques to measure release of alternative species within the reaction mixture should also be developed. Similar measurements were attempted using the phosphate release assay, however phosphate contamination of the isocitrate dehydrogenase used precluded accurate measurement.

7.2.4 *Comparison of* E. coli *and* S. pneumoniae *pathway dynamics*

The dynamics of *E. coli* and *S. pneumoniae* pathways are now compared based on time courses simulated in COPASI. Model 7.2.3.4, Table C.5, being the closest to experimental data under the conditions considered, was used in the following analysis. The *E. coli* pathway has the same form as *S. pneumoniae* pathway and so was described by differential equations (7.1). However the dynamics of each enzyme is different thus an alternative set of parameters, Table 7.4, was derived from the available literature. These parameters were not all obtained under the same

conditions. As such the results of this analysis must be treated with some caution.

Enzyme	Parameter	Value	Enzyme	Parameter	Value
MurA	k_{cat} (s^{-1})	4.75	MurB	k_{cat} (s^{-1})	21.7
[81, 91]	k_m (PEP) (μM)	0.400	[85]	k_m (NADPH) (μM)	19.9
	k_m (UGP) (μM)	15.0		k_m (UDPPEE) (μM)	12.5
MurC	k_{cat} (s^{-1})	16.3	MurD	k_{cat} (s^{-1})	15.5
[98, 99]	k_m (ATP) (μM)	130	[100]	k_m (ATP) (μM)	138
	k_m (L-Ala) (μM)	48.0		k_m (D-Glu) (μM)	55.0
	k_m (UMN) (μM)	44.0		k_m (U1P) (μM)	19.4
MurE	k_{cat} (s^{-1})	11.7	MurF	k_{cat} (s^{-1})	19.4
[101]	k_m (ATP) (μM)	620	[102]	k_m (ATP) (μM)	164
	k_m (mDAP) (μM)	36		k_m (D-Ala-D-Ala) (μM)	208
	k_m (U2P) (μM)	76.0		k_m (U3P) (μM)	78

Table 7.4: Table of kinetic parameters describing the reactions in the *E. coli* pathway. References below the enzyme abbreviations indicate the source of these data.

Both pathways were simulated in COPASI under the initial conditions described in Section 7.2.1. The concentrations of UGP and U5P and their rates of change are presented in Figure 7.7. Based on the parameters obtained the *E. coli* pathway completes turnover of UGP to U5P more quickly under the conditions used than the *S. pneumoniae* pathway. Initial rates of production of U5P are higher for the *S. pneumoniae* pathway, due to its generally faster amino acid ligases which rapidly turnover any UDP-intermediates which reach them. However the faster early stages of the *E. coli* pathway allow it to attain a higher maximum rate of U5P production, resulting in faster production overall.

The concentrations of UDPPEE, UMN and U2P and their rates of change are presented in Figure 7.8. The *S. pneumoniae* pathway accumulates a very high concentration of UDPPEE, almost half initial concentration of UGP, which is then slowly turned over to UMN (Figure 7.8(a)-(b)). This is due to the high ratio between the MurA and MurB k_{cat}s. This ratio is reversed in the *E. coli* pathway resulting in a very low accumulation of UDPPEE which is

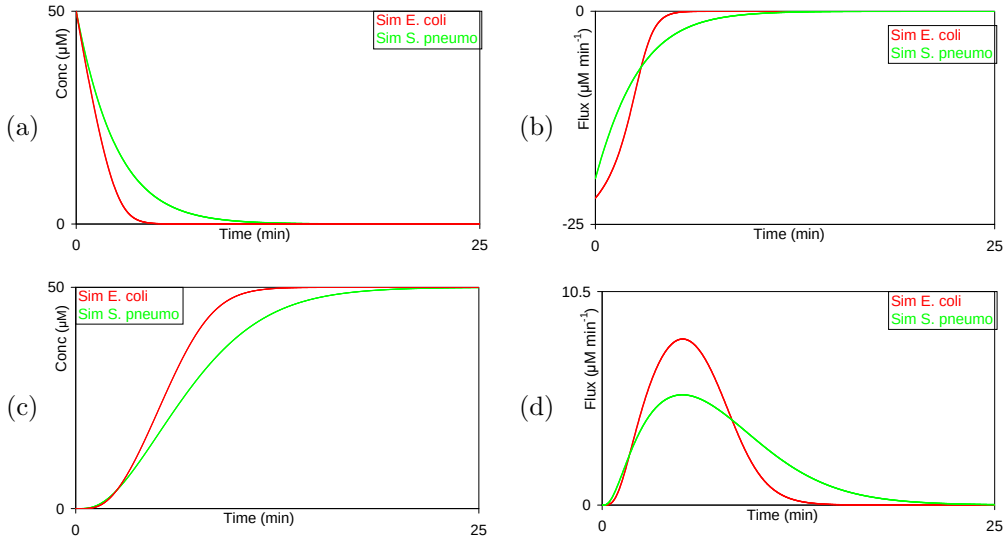


Figure 7.7: Comparison of simulated time courses for *E. coli* and *S. pneumoniae* pathways. Graphs (a) and (c) plot the concentrations of UGP and U5P respectively against time. Graphs (b) and (d) plot the rates of change of these species against time.

quickly turned over to UMN.

The concentrations of the remaining intermediates in the *E. coli* pathway follow a consistent pattern (Figure 7.8(c)-(f)). The concentration of intermediate peaks rapidly and then equally quickly is turned over. Comparing Figure 7.8(c) and (e), the peak concentration of U2P is reached only when UMN has been nearly completely depleted. This is reflected in the rates of change of these species which have quite large maxima and minima (Figure 7.8(d) and (f)). The reactions come close to operating in stages, the following reaction only reaching high effective rates when the previous reaction has nearly stopped. The *S. pneumoniae* pathway forms a clear contrast. Here the later UDP intermediates, U1P, U2P and U3P, do not accumulate in high concentrations; instead the high k_{cat}s of the enzymes using these substrates ensure that they are rapidly turned over.

The shapes of concentration curves of UMN, U1P, U2P and U3P tend to mirror the shape of the UDPPEE curve; further emphasizing the influence of the MurB reaction on the pathway dynamics. This reaction seems to act as a bottleneck and as such seems critical to the control of

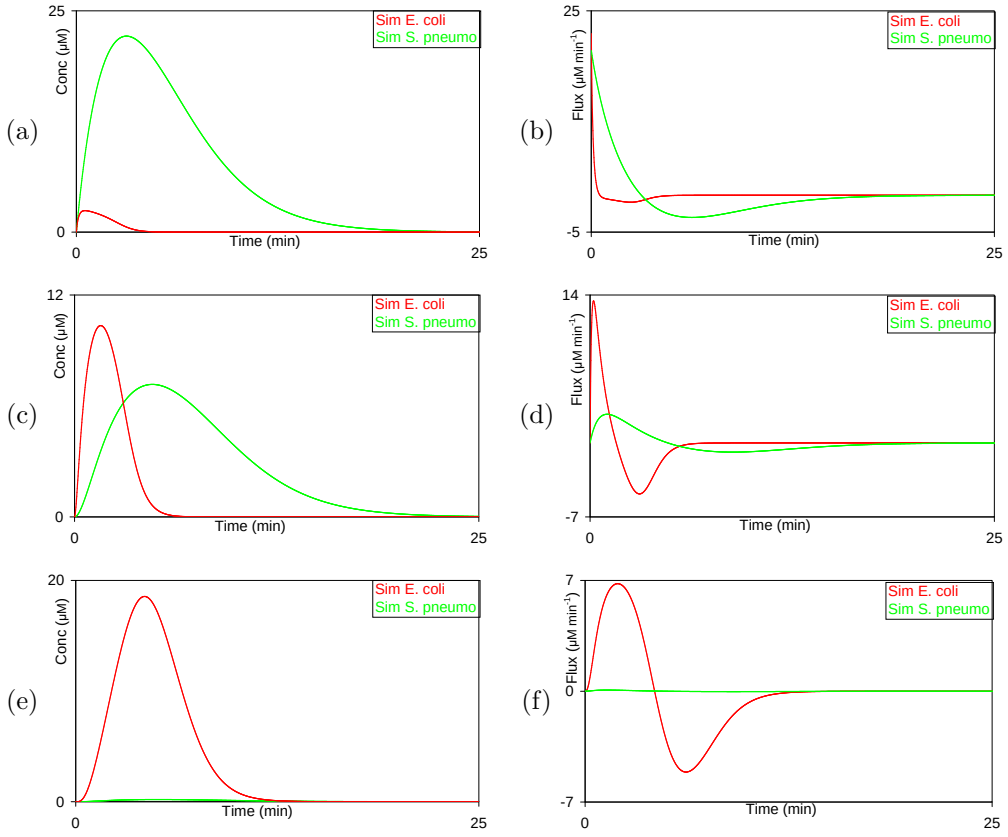


Figure 7.8: Comparison of simulated time courses for *E. coli* and *S. pneumoniae* pathways. Graphs (a), (c), and (e) plot the concentrations of UDPPEE, UMN and U2P respectively against time. Graphs (b), (d), and (f) plot the rates of change of these species against time.

the pathway. By contrast no single reaction in the *E. coli* pathway has such a strong influence on the pathway dynamics. The latter stage intermediates accumulate at high concentrations, like UDPPEE, but then are rapidly depleted.

Both types of behaviour suggest possible antibiotic strategies. The *S. pneumoniae* pathway may be susceptible to inhibition of the MurB and MurC catalysed reactions, especially since UDPPEE and UMN act as inhibitors of MurA. This possibility will be investigated further later in the chapter. The quick accumulation of intermediate concentrations in the *E. coli* pathway would overwhelm inhibitors which compete with the UDP intermediates. As such an uncompetitive inhibitor would be preferable as suggested in [67]; given the speed of the pathway such an inhibitor might rapidly drive the affected UDP-intermediate to toxic concentrations.

Note however that the pathway simulations undertaken here are unlikely to correspond to conditions *in vivo*. For example it is unlikely that all enzyme concentrations are static and equal, or that a small finite initial concentration of UGP is provided. Issues of compartmentalisation of enzymes and the effects of enzyme crowding on kinetics have not even been considered. More work is required to model the *in vivo* pathways, first simply to determine *in vivo* concentrations of enzymes and then possibly to consider the more complex problems of compartmentalisation and crowding which are hard to duplicate *in vivo*.

7.3 Characterisation of two inhibitors of the peptidoglycan biosynthesis pathway

The analysis presented in the previous section assumes an unchallenged pathway. It can be extended by incorporating the known internal inhibitors of the pathway, UDPPEE and UMN, and by incorporating inhibitors of other enzymes. First however any inhibitors used must be characterised. In this section two inhibitors of enzymes in the peptidoglycan biosynthesis pathway are characterised. The first, UDP-MurNAc a product of the reaction catalysed by MurB and a substrate of MurC, is an inhibitor of MurA[186]. As such it is not appropriate for use as an antibiotic; but may enhance the effects of inhibitors challenging reactions further along the pathway. The second, C-1, is an inhibitor of MurC, and is a potential antibiotic (Figure 7.9) [63]. Kinetic parameters describing the inhibitory effect of these compounds on *S. pneumoniae* enzymes were determined experimentally. They are used in pathway simulations presented in Section 7.4.

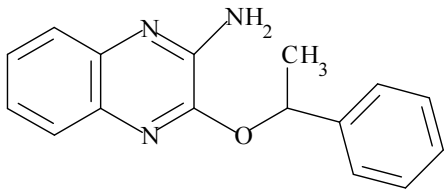
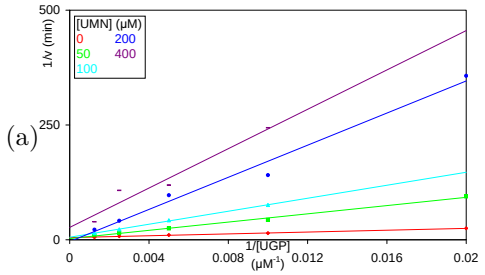
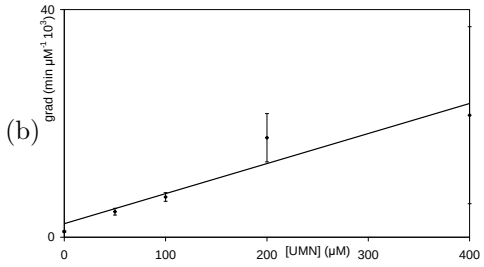


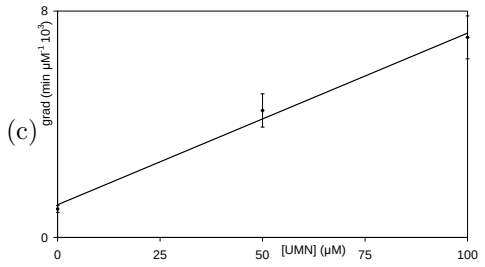
Figure 7.9: Chemical structure of C-1 [63].

7.3.1 Inhibition of MurA

It is known that UDP-MurNAc (UMN) had an inhibitory effect on the reaction catalysed by MurA [186]. The k_m of MurA with respect to PEP is very low relative to that of UDP-GlcNAc (UGP), Section 5.1. As such the inhibitory effect is likely to be most pronounced with respect to UGP. It was characterised using the phosphate release assay as previously described in Sections 2.4.3 and 5.3. Reaction mixtures contained the components listed in Section 2.4.3 with the addition of PEP to a final concentration of 480μM. For each of five concentrations of UMN, 0, 50, 100, 200 and 400μM, five initial rates were determined dependent on the following concentrations of UGP, 50, 100, 200, 400 and 800μM. Reactions were initiated by the addition of MurA to a final concentration of 28.5nM.







(d)

		m $\left(\frac{\text{min}}{\mu\text{M}^2}\right)$		y_0 $\left(\frac{\text{min}}{\mu\text{M}}\right)$	
Plot	RMSE	EV	RMSE	EV	RMSE
(b)	74.0	52.8	9.55	2380	1970
(c)	6.70	60.6	5.19	1160	335

Figure 7.10: Primary and secondary plots used to characterise the inhibitory effect of UMN on MurA with respect to UGP. Graph (a) plots inverted initial rates (1/v) against inverted UGP concentrations (1/[UGP]) and lines of best fit for the five UMN concentrations used. Graphs (b) and (c) plot the gradients of these lines (grad) against concentration of UMN ([UMN]) and a line of best fit. Error bars indicate a 90% confidence interval for each gradient. Graph (c) omits the points corresponding to 200 and 400μM UMN. Table (d) contains the RMSE of the fits, the estimated parameter values, EV, and associated RMSE, for gradient (m) and y-intercept (y_0) for Graphs (b) and (c).

The data were linearised and plotted using the Lineweaver-Burke plot (Figure 7.10(a)) [144]. Two data points were omitted from this plot due to the very low rates observed. Low rates are

poorly determined due to the low signal measured. Furthermore since the linearisation used requires these values to be inverted any errors in these readings are magnified disproportionately to those arising at higher rates. For low concentrations of UMN, 0-100μM, the lines plotted appear to intersect on the 1/v axis; suggesting competitive inhibition with respect to UGP [144]. For higher concentrations of UMN this convergence does not appear to be preserved. This could suggest that an alternative mode of inhibition is more appropriate. However given the quality of these data no definitive statement can be made. The analysis proceeds on the assumption of competitive inhibition with respect to UGP.

The gradients of the lines plotted in Figure 7.10(a) were plotted against concentration of UMN in Figure 7.10(b). Error bars were estimated as described in Section 3.4 for a Student's t-distribution to constitute a 90% confidence interval. For concentrations of UMN exceeding 100μM the confidence intervals are so large as to render the corresponding estimates meaningless. As such the remaining three data points were replotted separately as Figure 7.10(c).

Lines of best fit were estimated in each case; the results are summarised in Figure 7.10(d). The best fit is obtained for the second plot, omitting the two concentrations of UMN exceeding 100μM, however in both cases the intercept with the y axis is poorly determined. The k_i with respect to UGP is estimated by taking $\frac{y_0}{m}$, where y_0 is the y-intercept and m the gradient, for this graph [144]. The estimated values are 45.0 and 19.1μM for Graphs (b) and (c) respectively. However given the high error associated with the estimated values of b these values cannot be regarded with any great confidence. Extending the series of assays to encompass a greater range of initial concentrations should allow this parameter to be determined more accurately. Furthermore better results will be achieved if the initial rates measured are higher; since the linearisation used is poorly behaved for low rates. It may also be necessary to eliminate any confounding effect caused by product inhibition due to the presence of UDPPEE, Section 5.3.

The second value of 19.1μM is used for k_i for the remainder of this chapter. MurA is product inhibited by UDPPEE. It will be assumed that the inhibitory effects of UDPPEE and UMN

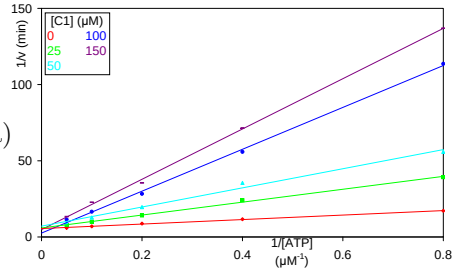
on MurA are the same unless otherwise stated. This assumption is justifiable given the high degree of similarity between the two species; however it would be preferable to determine the real value of this parameter.

7.3.2 Inhibition of MurC

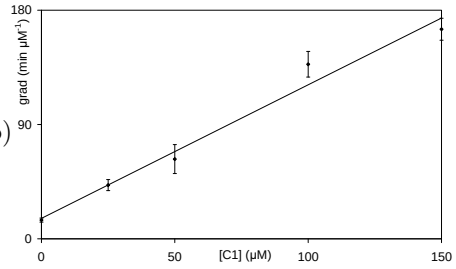
The compound C-1 has been shown to inhibit the reaction catalysed by *E. coli* MurC [63]. This inhibition is thought to be competitive with respect to ATP. It was shown, using phosphate release assays, to also have an inhibitory effect on *S. pneumoniae* MurC. The effect with respect to ATP was characterised using ADP release assays, Section 2.4.2; C-1 having been shown not to affect this assay system. Reaction mixtures contained the components detailed in Section 2.4.2 with the addition of UMN and MurC to final concentrations of 550μM and 35nM respectively. For each of five concentrations of C-1: 0, 25, 50, 100, 150μM, five initial rates were determined dependent on the following respective concentrations of ATP: 1.25, 2.5, 5, 10, 20μM. At the required concentration for a stock solution C-1 was found to be insoluble in water, but dissolved in DMSO. All reaction mixtures were made up to 3% DMSO to eliminate any effect of DMSO concentration on the characterisation. Reactions were initiated by the addition of L-Ala to a final concentration of 10mM.

The data were linearised and plotted using the Lineweaver-Burke plot (Figure 7.11(a)). The lines plotted do not intersect perfectly on the 1/v axis, however they are close to doing so. As such, informed by the results for *E. coli* MurC [63], it is reasonable to assume that competitive inhibition with respect to ATP. The gradients of the lines plotted in Figure 7.11(a) were plotted against concentration of C-1 (Figure 7.11(b)). Error bars were estimated as described in Section 7.3.1; all gradients estimated were determined to a reasonable degree of accuracy. A line of best fit was estimated for these data; the results are summarised in Figure 7.11(c). A good fit was obtained although the estimated value of the y-intercept is poorly determined. Note however that this value is within the 90% confidence interval of the corresponding point plotted. As such

(a)



(b)



(c)

		m $\left(\frac{\text{min}}{\mu\text{M}^2}\right)$		y_0 $\left(\frac{\text{min}}{\mu\text{M}}\right)$	
Plot	RMSE	EV	RMSE	EV	RMSE
(b)	8.68	1.05	0.0930	16.1	7.85

Figure 7.11: Primary and secondary plots used to characterise the inhibitory effect of C-1 on MurC with respect to ATP. Graph (a) plots inverted initial rates (1/v) against inverted ATP concentrations (1/[ATP]) and lines of best fit for the five C-1 concentrations used. Graph (b) plots the gradients of these lines (grad) against concentration of C-1 ([C1]) and a line of best fit. Error bars indicate a 90% confidence interval for each gradient. Table (c) contains the RMSE of the fit, the estimated parameter values (EV) and associated RMSEs for gradient (m) and y-intercept (y_0) for Graph (b).

this estimate can be treated more confidently than initial statistics suggest. The k_i estimated, using the formula in Section 7.3.1, was 15.3±6.00μM. This value is used for the remainder of the chapter.

7.4 Modelling of the inhibited peptidoglycan biosynthesis pathway

Having obtained characterisations of pathway inhibitors in the previous section it is now possible to model and simulate the effects of these inhibitors on the pathway. In addition it is possible to measure time courses of pathways challenged by these inhibitors experimentally. In the next section experimental and simulated data are compared. Initially the models used consider only the inhibitors internal to the pathway, UDPPEE and UMN. The models are compared to the experimental data introduced in Section 7.2.1. The models are then expanded to consider inhibition by C-1. Experiments were conducted for the subpathway starting with MurA and terminating at MurC, the results were compared to simulated data. In the subsequent section the effects of inhibition are examined in more detail using simulated data. The interaction

between inhibition by C-1 and the internal pathway inhibitors is examined in the context of antibiotic development.

7.4.1 Comparison of simulated inhibited pathway dynamics to experimental data

Two sources of internal pathway inhibition are considered, Figure 7.12; inhibition of MurA by UDPPEE or by UMN. Three models were used: MurA was inhibited either competitively or uncompetitively by UMN (Models 7.4.1.1 and 7.4.1.2) or competitively by both UDPPEE and UMN (Model 7.4.1.3). The quasi-steady state equations to describe these models of inhibition are as follows:

$$f_6(S_1, S_2, I) = \frac{V_{max} S_1 S_2}{S_1 S_2 + k_{m,1} S_2 (1 + \frac{I}{k_i}) + k_{m,2} S_1}, \tag{7.2}$$

$$f_7(S_1, S_2, I) = \frac{V_{max} S_1 S_2}{(1 + \frac{I}{k_i})(S_1 S_2 + k_{m,1} S_2 + k_{m,2} S_1)}, \tag{7.3}$$

$$f_8(S_1, S_2, I_1, I_2) = \frac{V_{max} S_1 S_2}{S_1 S_2 + k_{m,1} S_2 (1 + \frac{I_1}{k_{i,1}})(1 + \frac{I_2}{k_{i,2}}) + k_{m,2} S_1} \tag{7.4}$$

where f_6, f_7 and f_8 replace f_1 in Equations (4.5) for competitive, uncompetitive or two competitive inhibitors respectively. These expressions are well-known [144]. The inhibitory terms, $(1 + \frac{I}{k_i})$, decrease the rate of reaction by increasing the denominator as the concentration of I increases. For competitive inhibition inhibition directly increases the effective k_m of a particular substrate. Uncompetitive inhibition by contrast increases the entire denominator.

To model these changes $f_1^A([\text{UGP}], [\text{PEP}])$ was replaced in Equations (7.1) by $f_6^A([\text{UGP}], [\text{PEP}], [\text{UMN}])$, $f_7^A([\text{UGP}], [\text{PEP}], [\text{UMN}])$ or $f_8^A([\text{UGP}], [\text{PEP}], [\text{UDPPEE}], [\text{UMN}])$. No other changes were made to the structure of the model. The parameter values were initially unchanged from the best case obtained in Section 7.5; those for the MurB reaction were returned to the uninhibited values for the fourth simulation below (Model 7.4.1.4). Tables of the parameters used can be found in Appendix C, Tables C.6-C.7. The results were compared to the experimental time

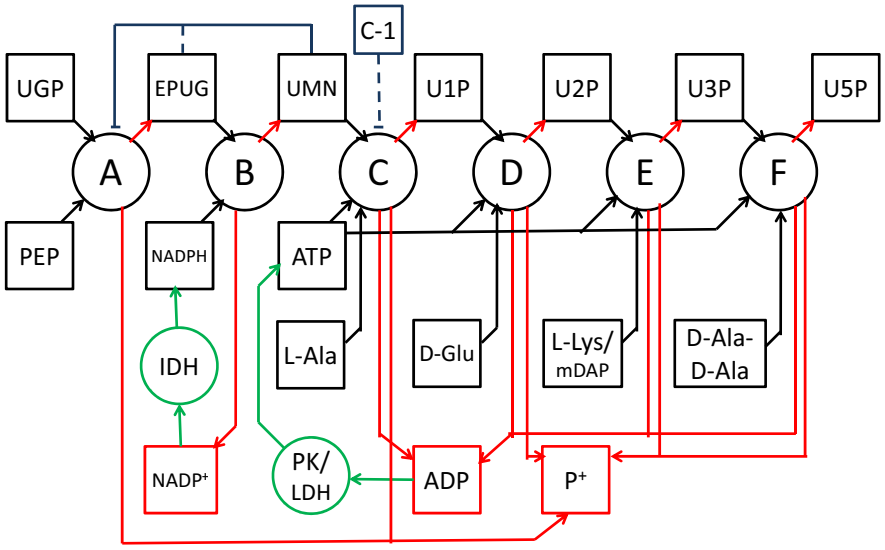
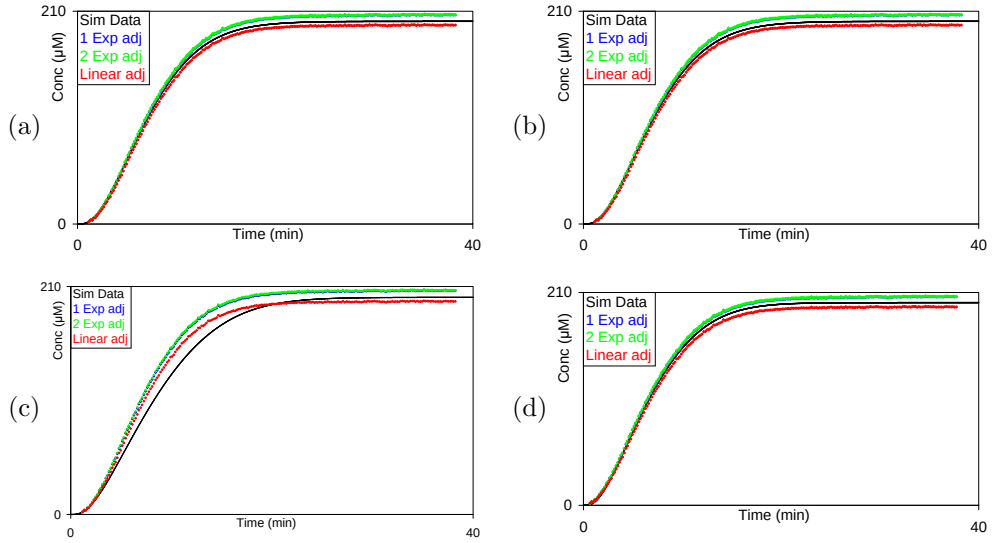


Figure 7.12: Schematic diagram of cytoplasmic phase of peptidoglycan biosynthesis with inhibition steps. Colours and shapes are as previously described, Figure 7.1. Blue lines connect inhibitors to the process inhibited. Dashed lines indicate an inhibitory effect which is not included in all models.

courses described in Section 7.2.1 (Figure 7.13).

The two modes of inhibition of MurA by UMN, competitive or uncompetitive, are practically indistinguishable for the simulations undertaken, the RMS difference between the two pathway models being 0.122μM. They also fit the experimental data, slightly better than the previous best model (Figure 7.13(e)). Where inhibition from both UDPPEE and UMN was simulated NAD^+ production is clearly slower after the first five minutes (Figure 7.13(c)). However when MurB was modelled using the uninhibited parameters the simulated curve corresponds well to the experimental data (Figure 7.13(d)).

In order to investigate the effects of C-1 on the pathway further experimental time courses were recorded for the subpathway containing the processes catalysed by MurA, MurB and MurC. Two sets of assays were undertaken, in the first NAD^+ release was measured as previously described and in the second, NADPH oxidation was monitored using the MurB assay, Section 2.4.4. Reaction mixtures were unchanged from those stated in Section 7.2.1 with the following exceptions: UGP concentration was 200μM; MurD to F and the corresponding amino acids



(e)

Model	Graph	Best RMSE (μM)
7.2.3.4	Figure 7.6(d)	4.63
7.4.1.1	(a)	4.08
7.4.1.2	(b)	4.00
7.4.1.3	(c)	7.39
7.4.1.4	(d)	4.03

Figure 7.13: Comparison of experimental time courses to pathway simulations taking into account inhibition by UDPPEE and UMN. Graphs (a)-(c) plot concentration of NAD^+ (μM) produced against time in minutes. The adjusted experimental curves are as described in Section 7.2.1. Graphs (a)-(d) plot simulated pathway production for Models 7.4.1.1-.4 respectively. Table (e) presents the lowest root mean square error for each simulated time course.

were omitted; and C-1 or DMSO was added to each reaction mixture to a final concentration of 200μM or 3% respectively. When using the MurB assay IDH and NADH were omitted and NADPH concentration was increased to 300μM. This eliminates the absorbance change caused by oxidation of NADH and allows the oxidation of NADPH to be monitored. Reactions were initiated by the addition of MurA and allowed to continue for about 40 minutes; two time courses were measured in each case and were found to agree well. Experimental time courses were adjusted using a single exponential model, derived from the average of the initial background rates of NAD^+ production and the final rate observed in the uninhibited pathways.

These experiments were modelled as follows. MurA was modelled as being inhibited competitively by either UMN or UMN and UDPPEE, i.e. f_1^A was replaced by f_6^A or f_8^A respectively in

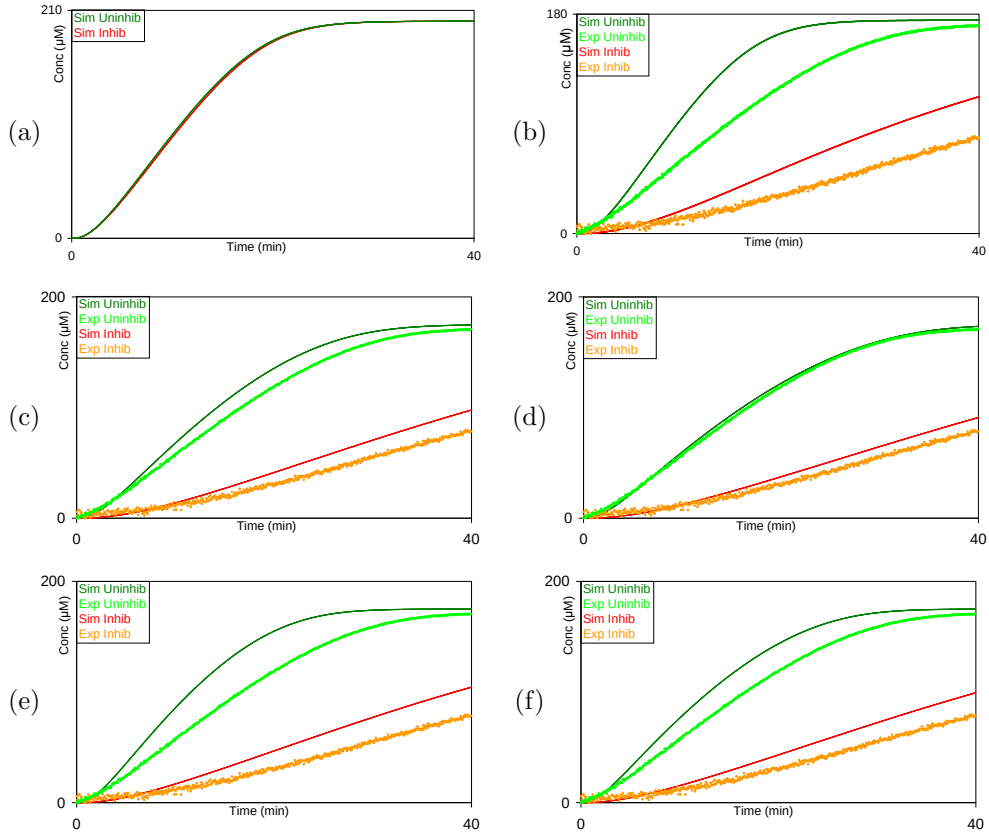


Figure 7.14: Comparison of experimental time courses to pathway simulations for NAD^+ production in the MurA to MurC pathway. Graph (a) plots simulated NAD^+ concentration for Model 7.4.1.5. Graphs (b)-(f) plot simulated and experimental concentration of NAD^+ (μM) produced against time in minutes for Models 7.4.1.6-10 respectively.

Equations (7.1). When C-1 was modelled as a competitive inhibitor (Model 7.4.1.5) using the following function in place of $f_3^C([\text{UMN}], [\text{ATP}], [\text{L-Ala}])$ in Equations (7.1) [144]:

$$f_9(S_1, S_2, S_3, I) = \frac{V_{max} S_1 S_2 S_3}{S_1 S_2 S_3 + k_{m,1} S_2 S_3 + k_{m,2} S_1 S_3 (1 + \frac{I}{k_i}) + k_{m,3}}, \tag{7.5}$$

for $S_1 = [\text{UMN}]$, $S_2 = [\text{ATP}]$, $S_3 = [\text{L-Ala}]$ and $I = [\text{C-1}]$ and appropriate kinetic parameters, no significant pathway inhibition is seen in the simulated data (Figure 7.14(a)). This does not correspond well to experimental data which show a significant degree of inhibition under the conditions used. The k_m with respect to ATP does not seem to significantly regulate the rate of the MurC catalysed reaction in simulations; the reason for this remains unclear. It does however

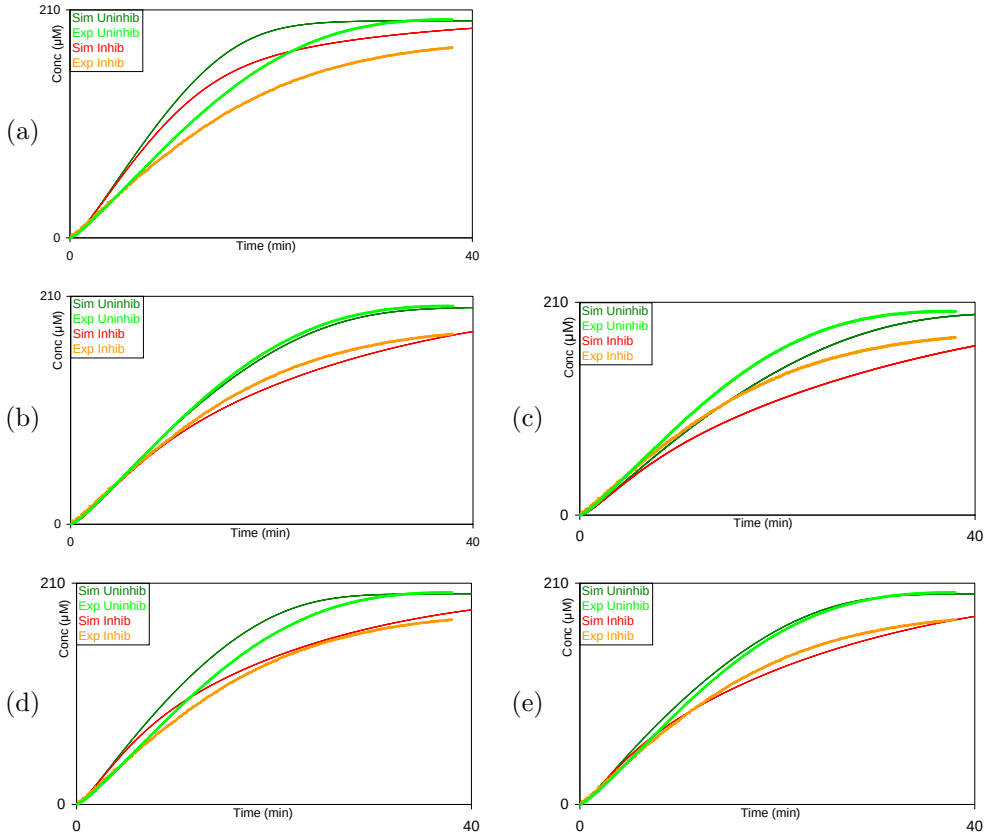


Figure 7.15: Comparison of experimental time courses to pathway simulations for $NADP^+$ production in the MurA to MurC pathway. Graphs (a)-(e) plot simulated and experimental concentration of $NADP^+$ (μM) produced against time in minutes. Graph (a)-(e) plot the simulated results for Models 7.4.1.6-10 respectively.

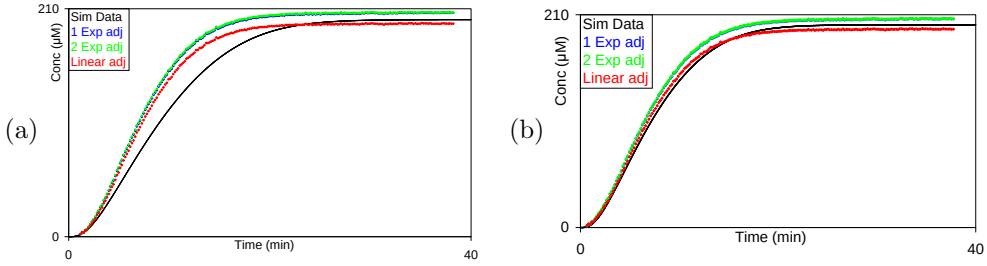


Figure 7.16: Comparison of experimental time courses to pathway simulations taking into account inhibition by UDPPEE and UMN. Graphs (a) and (b) plot concentration of NAD^+ (μM) produced against time in minutes. The adjusted experimental curves are as described in Section 7.2.1. Simulations for Graphs (a) and (b) used Models 7.4.1.8 and 7.4.1.10.

explain the apparent lack of sensitivity to competitive inhibition in this case. Since the ATP concentration is constant, due to the recycling by pyruvate kinase and lactate dehydrogenase in the models used and experiments undertaken it is reasonable to model C-1 as an uncompetitive inhibitor. Thus the following function was used in place of f_9 [144]:

$$f_{10}(S_1, S_2, S_3, I) = \frac{V_{max} S_1 S_2 S_3}{S_1 S_2 S_3 + k_{m,1} S_2 S_3 + k_{m,2} S_1 S_3 (1 + \frac{I}{k_i}) + k_{m,3}}, \tag{7.6}$$

with the same mapping of S_is and I to substrate or inhibitor concentrations as above. This did result in pathway inhibition (Figures 7.14(b)-(f) and 7.15). Parameters for these models are found in Appendix C, Table C.8. Initial conditions were chosen to correspond to those used in the experiments described above.

When MurA was inhibited by only UMN (Model 7.4.1.6) simulations did not describe the experimental data well (Figures 7.14(b) and 7.15(a)). The experimental NAD^+ and $NADP^+$ production whether inhibited or uninhibited is significantly slower than simulations predict. Given that it is known that MurA is inhibited by both UDPPEE and UMN this was to be expected.

Four simulations were undertaken where MurA was inhibited by both UMN and UDPPEE. The first two used the substrate inhibited MurB parameters (Models 7.4.1.7-8) the second two the uninhibited parameters (Models 7.4.1.9-10). For each pair two values of the k_i with respect to UDPPEE were used, 19.1 and 10μM. Parameters are presented in Appendix C, Tables C.9 and C.10. Predicted NAD^+ production, using Model 7.4.1.8, is a good match for experimental NAD^+ production (Figure 7.14(d)). However simulated $NADP^+$ production is slower than expected (Figure 7.15(c)). Similarly where simulated $NADP^+$ production is close to experimental data (Figure 7.15(b) and (e)), corresponding to Models 7.4.1.7 and 7.4.1.10, the simulated NAD^+ production is faster than experimental observations (Figure 7.14(c) and (f)). This is indicative that the ratio between the rates of the reactions catalysed by MurB

and MurC is not being accurately predicted by these models. Thus either the kinetic constants used are incorrect or an additional, unidentified, interaction between pathway intermediates and enzymes exists. The combination of uninhibited MurB and a k_i of 19.1μM with respect to UDPPEE (Model 7.4.1.9) is a poor match for both measurements (Figures 7.14(e) and 7.15(d)).

No combination of parameters logically arising from the data available fully describes the observed dynamics of the pathway. As such further work is needed to determine the cause of these discrepancies. The use of ping-pong models for all enzymes and the use of an uncompetitive model of C-1 inhibition may be contributory factors. Substrate inhibition of MurB is unlikely to be a factor given the relatively low concentrations of NADPH and UGP used.

Use of the lower k_i results in a poor match to the uninhibited pathway data unless the uninhibited MurB parameters are used (Figure 7.16). As such the best compromise is to use Model 7.4.1.7, corresponding to Figures 7.13(c), 7.14(c) and 7.15(b). This is the model used in the following section.

7.4.2 Simulated effects of inhibition on the pathway

In Section 7.2.4 the dynamics of simulated *E. coli* and *S. pneumoniae* pathways were simulated and compared. In this section the same approach is used to investigate the effect of inhibition on concentrations of the UDP-intermediates in the pathway using Model 7.4.1.7. In the presence of 200μ C-1 the overall production of U5P is reduced about 25% over 40 minutes (Figure 7.17(e)). Note, however, that for the majority of the time course the concentration of the U5P in the inhibited pathway lags significantly behind that of the uninhibited pathway; it catches up due to exhaustion of UGP. In the presence of C-1 rates of U5P production are significantly reduced resulting in a lower rate of production being maintained for a longer period (Figure 7.17(f)). Exhaustion of UGP is slightly slower and UDPPEE accumulation is slightly lower in the inhibited time courses however the differences are fairly minimal (Figure 7.17(a)-(d)).

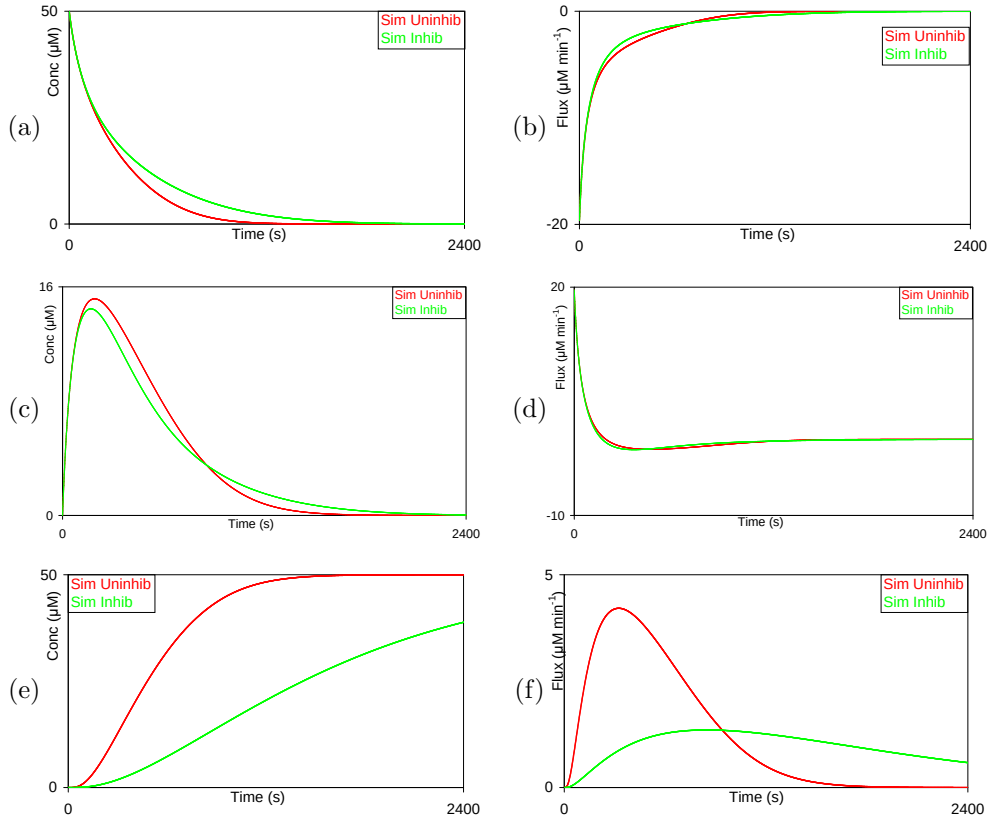


Figure 7.17: Comparison of simulated C-1 inhibited and uninhibited time courses. Graphs (a), (c), and (e) plot concentration of UGP, UDPPEE, and U5P (μM) respectively against time in minutes. Graphs (b), (d), and (f) plot the corresponding rates of the change of concentration of these species. Inhibited and uninhibited curves are colour coded.

Inhibition has a greater effect on the concentration of UMN (Figure 7.18(a)-(b)). The inhibition caused by C-1 causes a large accumulation of UMN; which in turn causes the small changes observed in the UGP and UDPPEE time courses. Exhaustion of this accumulation is not achieved within the time course simulated. The UDP-intermediates downstream of MurC have similar dynamics (Figure 7.18(c)-(d)). Under inhibition production and accumulation of these intermediates is reduced forming flattened versions of the UMN curve. Thus it would appear that, under inhibition, control of the reaction shifts from the MurB to the MurC reaction.

The potential synergy between C-1 and internal pathway inhibition is an area of some interest as it may influence whether C-1 proves an effective antibiotic. However for limited concentrations of UGP this interaction has little impact. The build up of UDPPEE and UMN is limited by

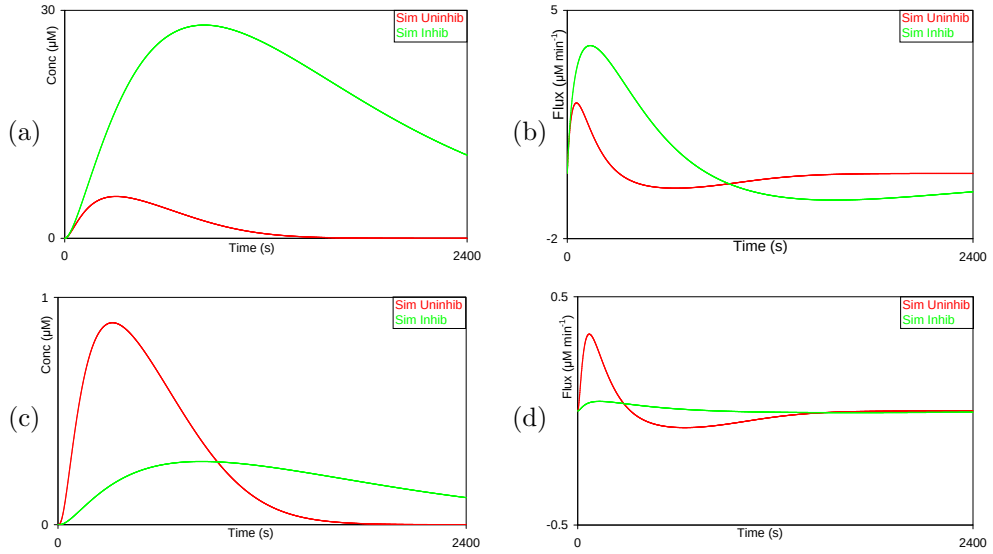
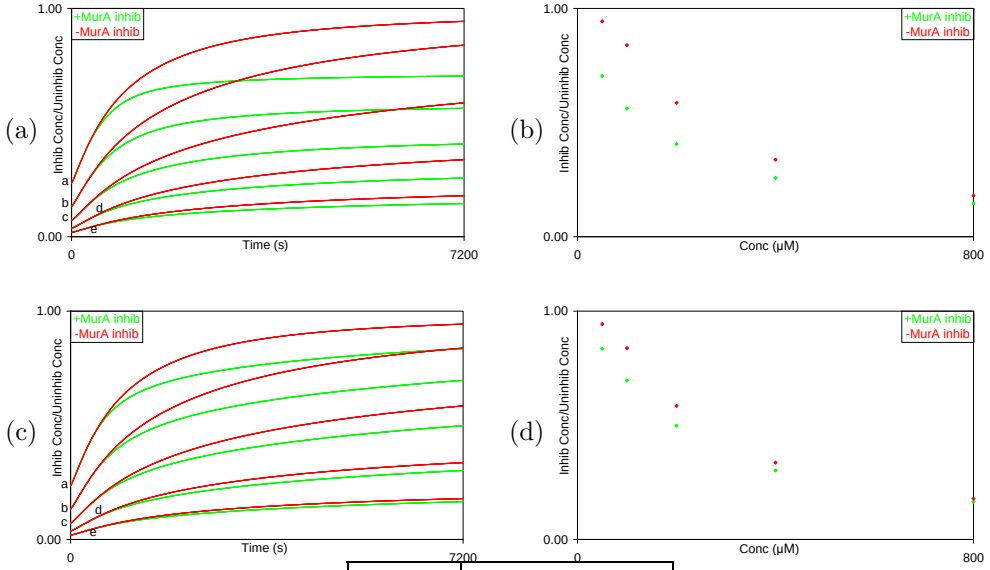


Figure 7.18: Comparison of simulated C-1 inhibited and uninhibited time courses. Graphs (a) and (c) plot concentration of UDPPEE, UMN and U1P (μM) respectively against time in minutes. Graphs (b) and (d) plot the corresponding rates of the change of concentration of these species. Inhibited and uninhibited curves are colour coded.

the availability of UGP. Where a low concentration of UGP is used inhibitory concentrations of these species only form when UGP is mostly exhausted. However a high concentration of UGP can overwhelm inhibition by these species. As such the more significant effects of this interaction are likely to be less pronounced under typical experimental conditions than they would be *in vivo* where UGP can be continuously provided to the pathway. An analysis of the case where UGP is constantly provided to the pathway is presented below.

DIfferential equations (7.1) were modified as previously described to include inhibition of MurC by C-1. Replacement of UGP was modelled in two ways. In the first case the MurA reaction produced a molecule of UGP for every molecule used, i.e. $\frac{d[\mathrm{UGP}]}{dt} = 0$ in Equations 7.1. This represents rapid replacement by a carefully controlled upstream pathway and was used for Models 7.4.2.1 and 7.4.2.2. Alternatively a constant flow of UGP into the system was allowed, i.e. $\frac{d[\mathrm{UGP}]}{dt} = v_{\mathrm{UGPrep}} - f^A_{1,8}$, (Models 7.4.2.3 and 7.4.2.4); this flow was chosen so that in uninhibited conditions UGP concentration remained at approximately the initial value of 100μM. In odd numbered models, MurA was modelled as inhibited by UDPPEE and UMN using the

modifications to Equations (7.1) previously described, in even numbered models this inhibition was not included. The k_{cat} of MurA was reduced in the even numbered models so that, in the absence of C-1, overall production by both pathways was approximately equivalent. Six concentrations of C-1 were used 0, 50, 100, 200, 400, and 800μM and simulations of a two hour time course were undertaken. The results are presented below in Figures 7.19 and 7.20. Model 7.4.1.7 was modified to form four new models, Models 7.4.2.1-4, parameters can be found in Appendix C, Table C.11.



(e)

Curves	Conc C-1 (μM)
a	50
b	100
c	200
d	400
e	800

Figure 7.19: Gross effect of interaction between MurC inhibition by C-1 and MurA inhibition by UDPPEE and UMN. Graphs (a) and (c) plot the ratio of U5P produced in inhibited and uninhibited pathways against time in seconds. Curves are colour coded according to whether MurA was inhibited. Pairs of curves are associated with a letter which indicates the concentration of C-1 used, Table (e). Graphs (b) and (d) plot the ratio of inhibited to uninhibited U5P production after 7200 seconds against the concentration of C-1 used (μM). Graphs (a) and (b) correspond to Models 7.4.2.1 and 7.4.2.2, Graphs (c) and (d) to Models 7.4.2.3 and 7.4.2.4.

The graphs in Figure 7.19 plot the production of U5P at the non-zero concentrations of C-1

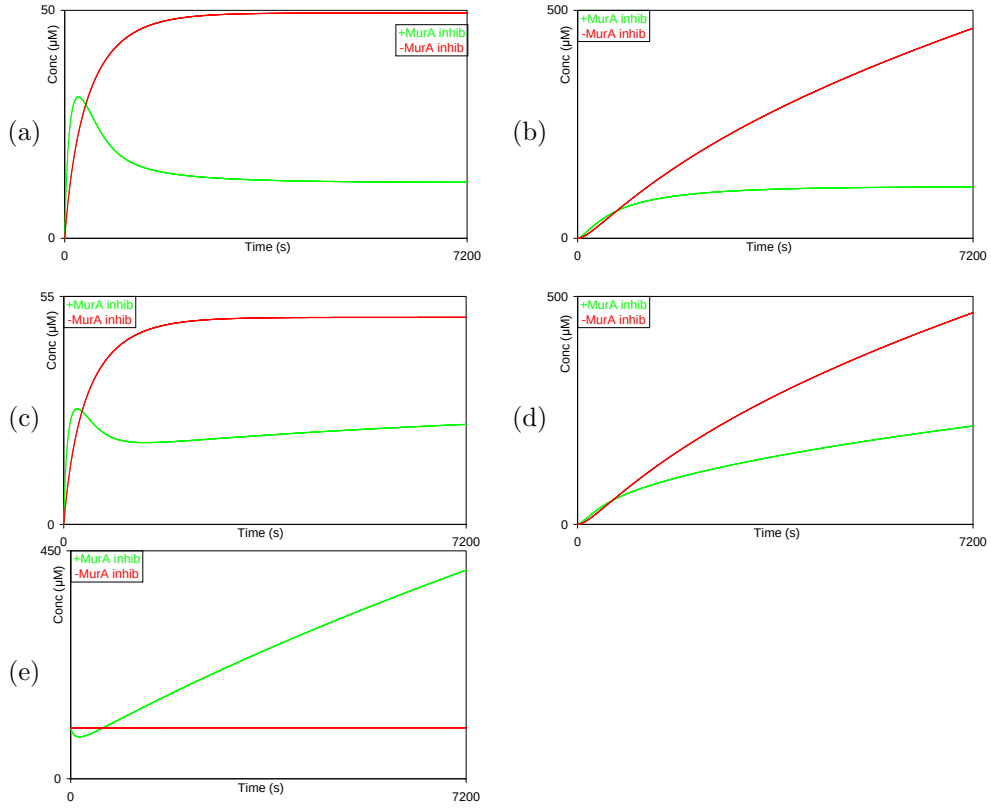


Figure 7.20: Effect of the interaction between MurC inhibition by C-1 and MurA inhibition by UDPPEE and UMN on pathway dynamics. Graphs (a) and (b) plot concentrations of UDPPEE and UMN respectively for Models 7.4.2.1-2 against time. Graphs (c), (d) and (e) concentrations of UDPPEE, UMN, and UGP respectively for Models 7.4.2.3-4 against time. Curves are colour coded according to whether MurA was inhibited. Concentration of C-1 was 50μM.

relative to corresponding production when C-1 is zero. Over the initial fifteen to twenty minutes production of U5P is reduced due to the inhibition of MurC by C-1 (Figure 7.19(a) and (c)). However the red and green curves are very similar indicating that inhibition of MurA is not having a significant effect. Indeed at high concentrations of C-1, 800μM, this remains true for the entire time course (see the curves marked "e" in Figure 7.19(a) and (c)). However for lower concentrations of C-1, 50-200μM, inhibition of MurA starts to have a significant effect after the first twenty minutes. The green curves, corresponding to models where MurA is inhibited by UDPPEE and UMN, lag behind the red curves, where MurA is not inhibited.

The differences are most pronounced where UGP is replaced proportionate to use (Figure

7.19(a) and (b)). Figure 7.19(b) plots the ratio of U5P production at a given concentration of C-1 to that of an uninhibited pathway against concentration of C-1. Note the difference between the curves; where MurA is uninhibited twice as much C-1 is required to achieve an equivalent reduction in U5P production for concentrations of C-1 less than 400μM. The time courses suggest that this effect may become more pronounced over longer time courses (Figure 7.19(a)). Compare the green curve for 50μM C-1 to the red curve for 200μM C-1; relative rates of increase suggest that the red curve would intersect the green curve if the time course was extended another two hours. Thus for a four hour time course it may take four times as much C-1 to achieve equivalent reductions in U5P production if MurA is uninhibited.

Where UGP is provided to the system at a constant rate inhibition of MurA has a reduced effect, green curves generally lie closer to red curves (Figure 7.19(c) and (d)). This can be explained by comparing the concentrations of UGP, UDPPEE, and UMN within the pathways for a given C-1 concentration, 50μM in this case (Figure 7.20(a)-(e)). If UGP is replaced rather than being provided at a constant rate then inhibition of MurA by UDPPEE and UMN reduces the rate of influx of UGP into the system. The concentrations of these three species then reach constant levels (Figure 7.20(a) and (b)). If MurA is not inhibited by these substrates the rate of influx of UGP into the system is constant allowing higher concentrations of UDPPEE and UMN to be attained, thus allowing the pathway as a whole to produce more U5P. When UGP is provided at a constant rate inhibition of MurA causes the accumulation of UGP (Figure 7.4.2(e)) eventually overcoming this competitive inhibition (Figure 7.4.2(c) and (d)). In these graphs the green curves do not settle to steady states indicating that the rate of the MurA reaction is increasing as the concentration of UGP increases. Thus the MurA-C subpathway can attain higher rates of reaction when UGP is provided at a constant rate, increasing the rate of production of U5P.

Note that accumulation of UMN does not overcome C-1 inhibition in the same way. The model for C-1 inhibition used was uncompetitive which is not strictly correct since C-1 is competitive

with ATP. However it is believed to be uncompetitive with UMN so this should not unduly affect the results above. In addition ATP is unlikely to accumulate *in vivo* given its use in many cellular processes, thus this inhibitor is unlikely to be overcome *in vivo*.

Further information is required to truly assess the effectiveness of C-1 in inhibiting the peptidoglycan biosynthesis pathway. As observed in previous sections, the enzyme concentrations are not representative of *in vivo* conditions. In addition the *in vivo* fluxes into and out of this pathway have yet to be determined and it is unknown what reductions in output can be tolerated. However the analysis does suggest that the interaction between C-1 and internal pathway inhibition would reduce the amount of C-1 required to attain any specific reduction in overall output. Thus lower doses of C-1 would be required potentially increasing its efficacy as an antibiotic. The magnitude of this reduction in dose is dependent in part on the way UGP is provided to the pathway. Given that UGP is used in a number of bacterial processes [80], it is unlikely that accumulations of the type observed in the second pair of models could occur. As such the more optimistic predictions presented may hold more weight.

7.5 Summary

A model has been developed that accurately predicts the production of ADP by the pathway when unchallenged by inhibitors for the experimental conditions used. Further validation of this model would be beneficial, both for alternative initial conditions and for measurement of other species especially since for the conditions used the later stages of the pathway seem to influence NAD^+ production relatively little. As yet it has not been possible to use the phosphate release assay described in Section 2.4.3 to measure phosphate production of the pathway, due to phosphate contamination of the recycling enzyme used to prevent exhaustion of NADPH. A UDP-pentapeptide assay also exists [187]. However concentrations of the available coupling enzymes proved insufficient for kinetic assays. Further work would allow either of these assays

to be used to measure pathway products.

Use of higher initial concentrations of UGP proved difficult, since in the full pathway every molecule of UGP produces four molecules of ADP. Providing sufficient NADH at high concentrations of UGP increases the background absorbance to levels at which the time courses measured are too noisy to be useful. The use of alternative assays may allow this problem to be circumvented.

The model was extended to include the effects of three inhibitors. The predicted time courses match less well to experimental data as such the model needs to be refined to better account for the discrepancies observed.

The *E. coli* and *S.pneumoniae* pathways were compared using simulations. The kinetic parameters used for the *E. coli* pathway were obtained from a number of different sources and were not recorded under consistent conditions; an area of potential improvement. The analysis highlights some differences between the two pathways which may provide insight into the differences between Gram-positive and Gram-negative peptidoglycan biosynthesis. However given the current lack of data concerning *in vivo* concentrations of the enzymes no strong conclusions can be drawn at this time. The differences suggest distinct antibiotic strategies for the two species.

The *S. pneumoniae* strategy, informed by the observed accumulations of UDPPEE and UMN in the pathway, was targeting MurB and MurC. This strategy could be partially tested given the availability of C-1, a MurC inhibitor. Simulations showed that inhibition of MurC would cause further accumulation of UMN in the pathway. In addition the inhibition of MurA by UMN and UDPPEE enhanced the effectiveness of this inhibitor.

8. WIDER DISCUSSION AND CONCLUSIONS

8.1 Introduction

The preceding results describe significant developments in two key areas: the direct estimation of kinetic parameters from experimental time courses and the modelling of the cytoplasmic phase of the peptidoglycan biosynthesis pathway. The tools developed for direct parameter estimation are applicable to any kinetic characterisation and are discussed in the following section. The limitations of the models constructed and their application in the development of antibacterials is then discussed. Brief conclusions are presented in the final section.

8.2 Parameter identifiability and estimation

The normal approach to kinetic characterisations of enzymes, via the determination of initial rates of reaction for a range of initial substrate concentrations, is time consuming due to the large number of experiments required. If the enzymes characterised use multiple substrates and a mechanism more complex than a ping-pong mechanism this problem is exacerbated. Two and three substrate reactions are relatively common; for example around a fifth of *E. coli* K-12 reactions use ATP and some other substrate [188]. As such modelling the majority of metabolic pathways will require kinetic characterisations of such reactions. The estimation of kinetic parameters from experimental time courses significantly reduces the experimental cost of these characterisations.

While this approach has been proposed in the past [146, 147] it is underdeveloped. The impor-

tance of ensuring that the parameters obtained are uniquely determined from the data was not noted. As such the results obtained were unreliable it was possible that alternative parameters would produce equally good experimental fits. It is essential to conduct a structural identifiability analysis of the models used to determine whether parameter vectors can be uniquely determined given the type of experiment conducted. In addition where a number of candidate reaction mechanisms exist indistinguishability analysis is necessary to determine which experiments are need to determine which is appropriate.

Furthermore direct parameter estimation has yet to enter general usage as a tool for kinetic characterisation. Where possible it has been used in the kinetic characterisations undertaken in this work to demonstrate its potential. Thus while the individual characterisations may only be of interest for those studying the peptidoglycan biosynthesis pathway, the modelling, analysis and techniques presented here (Chapters 4-6), are more broadly applicable. In the following sections the results obtained concerning this subject are briefly summarised and areas of further work are suggested. Theoretical analysis is considered in the first three sections; the following sections are concerned with analysis of experimental results.

8.2.1 Steady state identifiability and indistinguishability analysis

An analysis of the identifiability and indistinguishability of models of enzyme reactions had not previously been undertaken for any significant subset of the potential two or three substrate reaction models. Quasi-steady state forms of such models for two and three substrate ping-pong reaction mechanisms were found to be structurally globally identifiable (SGI) for the measurement of a single reaction species (Section 4.3). Equivalent analyses of quasi-steady state simple ordered mechanisms demonstrated that these models are always unidentifiable; although their V_{max}s are always SGI. The most significant results obtained from indistinguishability analysis were the indistinguishability of the simple ordered mechanisms from ping-pong mechanisms with the same number of substrates. Practically distinguishing between reactions with differing

numbers of substrates is relatively easy (Section 4.5).

The expression describing the dynamics of the two substrate simple ordered mechanism (Equation (4.14)) is more generally applicable to all two substrate mechanisms where product release is irreversible [144]. The ping-pong mechanism is the only special case; one of the parameters, $k_{s,12}$ being zero. As such it is the only SGI mechanism; all other mechanisms being unidentifiable and indistinguishable from the ping-pong mechanism over the quasi-steady state time scale. Consequently the two substrate reaction mechanisms that do not involve some inhibitory steps can be considered fully analysed.

The general expression for three substrate mechanisms under the same constraints is [144]:

$$f_{11} = \frac{V_{max} S_1 S_2 S_3}{S_1 S_2 S_3 + \sum_{i=1}^{3} (k_{m,i} \prod_{j \neq i} S_j) + k_{s,12} S_3 + k_{s,13} S_2 + k_{s,23} S_1 + k_{s,123}}. \tag{8.1}$$

As such the analysis presented does not encompass all possibilities. Note however that the input-output relationships derived for the ping-pong and simple ordered mechanisms are constructed from the same monomials (Sections 4.3.4 and 4.3.5); only the coefficients are affected by the choice of mechanism. It is reasonable to suppose that this structure is inherent to the input-output relation and thus that a maximum of seven unknown parameters can be SGI. Note further that three of these coefficients seem to determine identifiability of the initial substrate concentrations; demonstrated by the local identifiability of these parameters for the unidentifiable simple ordered model. Combining these observations it seems likely that only the ping-pong mechanism is SGI, and that all other mechanisms are unidentifiable and indistinguishable from it on the quasi-steady state time scale. This supposition, however, requires further analysis before it can be confirmed.

A number of possible reaction mechanisms are yet to be analysed. In the context of pathway modelling and development of antibacterials, mechanisms which describe inhibition, either by reaction substrates and products or by species which are not active within the reaction pathway,

are particularly important. The effects of reversible product release steps are also of interest. This constitutes a significant area for further work.

8.2.2 Transient identifiability analysis

Identifiability analysis of the transient models of two substrate reactions proved significantly more complex. Measurement of each individual reaction species was considered in conjunction with a number of possible product release mechanisms (Section 6.2). A number of SGI outputs were identified which are experimentally reasonable. Typically showing that a model corresponding to measurement of the first substrate to bind is SGI was relatively simple. The other significant experimental output, product concentration, proved more intractable, typically requiring the input-output relationship approach or atypical initial conditions be used.

An indistinguishability analysis of these models is needed to determine whether they can be distinguished experimentally. There are also several other possible mechanisms describing two substrate reactions which have yet to be analysed. Furthermore there are numerous inhibited or three substrate reaction mechanisms which have yet to even be considered. The structural identifiability and indistinguishability questions associated with transient reaction mechanisms remain relatively unexplored.

Given that typically metabolic pathways would be studied over hours, it could be argued that studying transient, pre-steady state, time scales is unnecessary. Note however the results presented in Section 5.7.1. Substrate inhibition was found to have an observable effect, on the steady state time scale, which was not predicted by the quasi-steady state equations, but was predicted by the transient model. The effect of accumulating inhibitors on pathway dynamics may also be best described by transient models. As such studying transient models may not be strictly necessary, but it does seem to provide additional understanding, even on relatively long time scales.

8.2.3 The input-output relationship approach

Concepts in differential algebra have been applied to the structural identifiability problem in the past [189]. Indeed one of these techniques uses this relation in the identifiability analysis [190]. However the problem of linear independence of the monomials has not previously been considered. As such the conditions, presented in Section 3.2.3.1, constitute a significant step in the development of this approach.

The input-output relationship approach proved highly effective in the analysis of the quasi-steady state models. However in most cases it proved impossible to compute input-output relationships if the measured species was involved in a reversible reaction; somewhat limiting its utility. It is currently unclear whether this is a fundamental limitation of the approach or simply of the algorithm used. However, when the input-output relationship could be derived, this approach allowed the analysis of models that were computationally intractable using the Taylor series approach. As such further development of this approach would seem worthwhile.

The application of the input-output relationship approach to indistinguishability analysis has not been previously documented. If the combined relations of the two models produce a polynomial, linear independence is easy to establish. However it is unclear how to do so if the result is a differential polynomial. Each input-output relation is associated with a distinct prime ideal; thus if the intersection of these ideals is non-trivial then it should also be prime. As such if the combined input-output relation is a generator of the intersection of these ideals, then the arguments advanced previously apply. However that this expression is necessarily such a generator is not immediately apparent. This then remains an area for further work.

8.2.4 Identification of substrate inhibition

Atypically shaped progress curves were observed in the kinetic characterisation of MurF, in Section 5.7.1. Such abnormalities are typically thought to arise from insufficient coupling enzymes

[144]. However in this case they appear to be associated with substrate inhibition a connection which appears not to have been noted previously. Curves simulated using transient, but not quasi-steady state, models of such a reaction were similarly shaped. The transient model also predicted that the change of shape would occur at a substrate concentration comparable to that observed experimentally. This may prove a useful test for substrate inhibition, a phenomenon which can be easily confused with assays containing insufficient coupling enzymes. Demonstration of this behaviour in other substrate inhibited enzymes would validate the explanation proposed. Further modelling may allow a link between the substrate concentration at which the curves change shape to the kinetic parameters of the reaction to be established. The discrepancy between the predictions of transient and quasi-steady state models emphasises the importance of transient models in developing a full understanding of pathway dynamics even over quasi-steady state time scales.

8.2.5 Estimation of quasi-steady state parameters

The results of direct parameter estimation from time courses were mixed. On the quasi-steady state time scale the majority of the enzymes considered catalyse unidentifiable reactions. As such, using a single progress curve, it was possible to determine only a single parameter, k_{cat} for these enzymes. The reaction catalysed by MurB was SGI and estimates of all parameters were obtained using direct parameter estimation. The parameters obtained corresponded reasonably well to those obtained using initial rate data; although there is some question as to whether a substrate inhibited model should be used for this reaction. It had been previously noted that initial conditions play a significant role in determining whether k_m values could be estimated [146]. This observation was substantiated by sensitivity analysis.

The two approaches also yield similar k_{cat} estimates for MurD, MurE and MurF (Table 7.3) the discrepancies observed being consistent with loss of enzyme activity due to time in storage. However when these parameters were estimated for MurA and MurC the resulting values were

significantly greater than obtained using the alternative approach. For MurA this discrepancy would appear to arise from product inhibition by UDPPEE; in any case it prevents the proposed deconvolution of MurA kinetic parameters (Section 4.3.3) from being tested. For MurC the cause of this discrepancy is unclear.

The results for MurB demonstrate the potential effectiveness of this approach for reactions which are believed to proceed by a ping-pong mechanism. Two time courses, rather than one, were required in order to determine all kinetic parameters due to the sensitivities of each of the k_ms. Estimates are improved by averaging several experimental time courses. In comparison typically ten to twenty initial rates are required to determine the k_{cat} and a single k_m; and repeat measurements are also typically taken. As such direct parameter estimation should consistently require significantly fewer experiments than methods using initial rates.

The simple ordered models are unidentifiable, however the results of the indistinguishability analysis (Section 4.4.1) suggest a way to use direct parameter estimation regardless. A series of time courses must be recorded using the same concentration of the first substrate to bind and different concentrations of the remaining substrate. If a ping-pong model is fitted to each of these time courses differing estimates for $k_{m,2}$ should be obtained. From Equations (4.75) plotting $k_{m,2}S$ against $S = (S_{20} - S_{10})$ should yield a line with y-intercept $\widehat{k_{s,12}}$ and gradient $\widehat{k_{m,2}}$; finally $\widehat{k_{m,1}} = k_{m,1} + k_{m,2} - \widehat{k_{m,2}}$. More data are required than would be for a ping-pong mechanism, $x + 1$ time courses, where x is the number of points desired in the line. However an equivalent characterisation using initial rates would require x^2 time courses, since a square matrix of initial substrate concentrations would be used [144], see for example the characterisation of inhibitors in Section 7.3. Thus direct parameter estimation would again require significantly fewer experiments. A similar approach may be possible for the three substrate simple ordered model, it is however not immediately obtainable by inspection of Equations (4.90). If initial rates are used however, x^3 time courses will be required [144], typically a minimum of 125, as such it is certainly worth developing such an approach.

8.2.6 Estimation of transient rate constants

Using a stopped flow spectrometer it was possible to record pre-steady state time courses for two enzymes, MurB and LDH (Section 6.3.1). LDH has been extensively studied in the past. It is not the intention of this work to further that analysis, but rather to use LDH in the development of this direct parameter estimation approach. The MurB reaction was described using a ping-pong mechanism [85] while LDH was described using a simple ordered model [180].

Studies with MurB found that four of the model parameters could be determined to some degree of confidence using experimental data. Quasi-steady state parameters were calculated from the transient parameters obtained; two of these parameters corresponded well to those determined from quasi-steady state kinetics. For simulated time courses of this reaction all six parameters could be estimated accurately for appropriate choices of initial conditions.

Studies using LDH demonstrated that this approach is highly sensitive to obtaining the correct model for the experiment undertaken. Furthermore they also suggest that for complex models the initial conditions may have a strong influence over which of the parameters may be successfully determined. This further substantiates the observations of Bates et al. [146] and emphasises the importance of sensitivity analysis after parameter estimation is completed.

Enzyme availability is a significant limiting factor in the analysis of pre-steady state kinetics using traditional techniques due to the need to have similar substrate and enzyme concentrations and sufficient substrate to observe a significant spectroscopic change. The results obtained suggest that, for a mechanistic fitting technique, the ratio of enzyme and substrate concentrations is less critical if the speed of the reaction is sufficiently low [145]. As such this technique has the potential to make pre-steady state kinetics easier to investigate experimentally.

The majority of enzymes considered do not produce a directly observable spectroscopic change. As such in order to study them using spectroscopic techniques a coupled assay is required. Using such systems when collecting pre-steady state time courses complicates parameter estimation,

since the coupling reactions may affect the time courses measured. A model could be developed which would include all reaction steps, however structural identifiability analysis of such a model is likely to prove computationally intractable. However each of the coupling reactions used is enzyme catalysed and uses two substrates, as such individual structural identifiability analyses should prove possible. Thus it may be possible to characterise these systems as follows. First the reaction directly responsible for the measured absorbance change would be characterised in the absence of the other reactions. Then this reaction would be used to allow time courses for the immediately preceding reaction. It should be possible to eliminate the effect of the coupling reaction using the rate constants obtained; allowing the next reaction to be characterised. In this way it may be possible to characterise each reaction.

However the time courses obtained for the reaction of interest may be insufficiently sensitive to the rate constants of this reaction to allow parameter estimation. As such an alternative approach may be required. Modification of the enzyme, to incorporate a fluorescing amino acid within the active site, might allow the binding of enzyme and substrate to be monitored directly [191]. This method suffers from uncertainty as to which reaction species are measured especially where there are multiple enzyme substrate complexes within the reaction mechanism. The use of a fluorescent probe of phosphate concentration avoids this problem and has been used for pre-steady state kinetic characterisations [192, 193].

It should also be noted that techniques like isothermal calorimetry [182], nuclear magnetic resonance [194], and even HPLC (in conjunction with quenched flow techniques) [84] have also been successfully applied to studying this sort of kinetic problem. Thus it may not be necessary to use spectroscopic techniques at all. The focus on spectroscopic techniques in this work arises principally from the particular suitability of the data produced by such techniques to direct parameter estimation.

8.3 Pathway modelling

An understanding of the dynamics of the fluxes through the peptidoglycan biosynthesis pathway may provide insight valuable in the development of new antibacterials targeting this pathway. Of particular interest are the interactions of inhibitors on this pathway since if an antibacterial influences multiple reactions, natural metabolic resistance may be reduced. In addition in such cases development of drug resistance by alteration of the target is more difficult, due to the increased number of mutations necessary. Species specific differences between the pathway dynamics may provide a similarly specific exploitable target. This may allow the development of drugs which target pathogenic bacteria while leaving the majority of the bacterial population unaffected. Such a drug would not create the opportunities for rapid proliferation of a newly resistant bacteria which are a factor the development and spread of antibacterial resistance [7].

Experimental data are ultimately limited to describing the system under a specific set of conditions, those used in the experiment. A more general understanding of the system requires a mathematical model which describes the results of the individual experiments. Models of the cytoplasmic phase of peptidoglycan biosynthesis have been developed for *E. coli* and *S. pneumoniae*. The results that can be obtained from these models and their limitations are discussed below. The first two sections discuss ways to improve model validation. The results are applied to the problem of antibacterial development in the third section. Ways to extend and improve these models are discussed in the final section.

8.3.1 Model validation

Mathematical models of quasi-steady state enzyme kinetics are well developed [144]. An *in silico* model of the cytoplasmic phase of the peptidoglycan biosynthesis pathway was constructed by coupling steady state representations of the reactions catalysed by MurA-F (Chapter 7). This model can be adjusted to represent the pathways arising in particular species by appropriate

choice of parameters. The pathway was also reconstructed *in vitro* using purified enzymes from *S. pneumoniae*; and monitored using the ADP release assay. There is nothing inherently novel in either of these two techniques. However this is the first time they have been applied to the study of this pathway. It should be further noted that while many of these reactions have been separately characterised, Table 7.4, they have never been characterised under a consistent set of experimental conditions. This is a vital requirement for parameters to be used in pathway modelling, since it makes little sense to use parameters corresponding to different experimental conditions in the same model.

The *in silico* model accurately predicted the ADP production of this *in vitro* pathway under the conditions considered. The model was expanded to incorporate inhibition of MurA by two pathway intermediates and MurC by an external inhibitor. The predictions of this model corresponded less well to experimental data; although they remain similar. The results obtained are encouraging; however it is clear that further validation is required. It would clearly be beneficial to consider other outputs of the pathway, that is the concentrations of other species. This would require the use of other assays and is discussed in detail in the following section. In the remainder of this section a number of other issues in the validation process are considered.

The model developed considers only processes directly related to peptidoglycan biosynthesis. The possibility of background processes, such as background ATPase activity or degradation of substrates, is not considered. However data from the *in vitro* reconstruction show significantly more NADH oxidation than can be accounted for by the available substrate. It is necessary to either incorporate this background oxidation into the model or eliminate it from experimental data. The latter option was used here; as this process seemed most likely to be an experimental artifact. The model chosen was however rather crude based simply on exponential or linear decay of the available NADH. A superior model could be developed if the cause of this oxidation were to be isolated and then characterised. A similar analysis would be needed for each of the alternative measurements described below.

The production of adenosine tetraphosphate by MurD and MurE is unlikely to influence these results, given the presence of an excess of the relevant amino-acids [109, 110]; it might however have a confounding effect under alternative initial conditions. It would be advisable to determine whether the amino-acid ligases and MurB have (other) background ATPase or NADPH oxidase activity in the absence of other substrates. Similarly the possibility of enzyme degradation, with resulting loss of activity, over time courses measured has not been considered. UDP-intermediates can be produced using appropriately truncated *in vitro* pathways, which are typically run overnight. As such it seems likely that no significant loss of activity would occur over the time courses considered and that the reactions will proceed to completion. Nonetheless it would be worth checking that these inferences are accurate.

The model constructed currently corresponds to an *in vitro* situation. Given that the *in vivo* concentrations of the enzymes catalysing this pathway are as yet unknown it is not yet possible to construct a model of the *in vivo* pathway. However for *E. coli* the *in vivo* steady state concentrations of several UDP-intermediates are known [116, 117, 118]. The ratio between any two pathway adjacent intermediates is determined by the rates of turnover of these intermediates. As such, given the relevant enzymes have been kinetically characterised, it should be possible to infer ratios of the enzyme concentrations from these data.

8.3.2 Alternative measurements

Phosphate release can be measured for a reconstructed pathway using the assay described in Section 2.4.3. This would provide additional data with which to validate the *in silico* model. Phosphate and ADP are produced by the four amino-acid ligases in equal amounts; however phosphate is also produced by MurA. Thus the difference between these two time courses should be entirely due to the contribution of MurA. Thus the activity of MurA within the pathway could be monitored directly if all background production of phosphate and ADP can be eliminated. Unfortunately the enzyme used to recycle NADPH for MurB, isocitrate dehydrogenase,

was found to be heavily contaminated with phosphate. This exhausted the available MESG, preventing this assay from being used to record time courses for the reconstructed pathway. The contaminating phosphate could be eliminated by FPLC using a size exclusion column (Section 2.3.5).

Direct monitoring of MurB is also possible if NADPH is not recycled (Section 2.4.4); this would result in depletion of NADPH which would affect the pathway dynamics. Increasing the concentration of NADPH so that this depletion has no significant effect may cause substrate inhibition of MurB to become a significant factor. If MurB has a significant NADPH oxidase activity it would have to be accounted for in the analysis of this assay.

An assay has been developed which uses a penicillin binding protein, DacB, to release D-Ala from UDP-MurNAc-pentapeptide. The oxidative deamination of D-Ala is used to produce hydrogen peroxide, which in turn is reduced using Amplex Red. Oxidation of Amplex Red produces an absorbance peak at 563nm [187]. This assay would allow the activity of MurF within the pathway to be monitored directly. However quite high concentrations of the relatively unstable DacB are required to allow rapid coupling, such concentrations are hard to achieve. In addition DTT, which is universally present in all assays undertaken in this work, will reduce Amplex Red, interfering with measurement; as such DTT should not be used in conjunction with this assay.

The Cary® 100 spectrometer can be used to record absorbance spectra over a time course (Varian Medical Systems Inc, Palo Alto, USA). This approach reduces the number of data points recorded at any given absorbance proportionally to the range of absorbances which must be measured. However it also permits several reaction species to be monitored in a single time course. The absorbances used by the phosphate, ADP and NADPH assays are very similar; which could confuse readings made using this approach. However NADH analogues, which absorb at a higher wavelength, could be substituted in the ADP assay, to allow the phosphate, ADP and UDP-MurNAc-pentapeptide assays to be combined in this way (Sigma-Aldrich, St

Louis, USA).

The techniques described above make use of real-time continuous assays. Such assays produce very dense time courses for a relatively small number of reaction species. It is possible to produce time courses with complementary characteristics, relatively sparse but for a majority of the reaction species, using stopped assays. Ongoing reactions are quenched at different time points by rapid denaturation of the enzymes or introduction of potent inhibitors. The reaction mixture can then be analysed using FPLC or mass spectrometry to determine the concentrations of most, if not all, of the reaction species.

These techniques can also be applied to analysis of *in vivo* concentrations of reaction species. Samples are taken from a culture of bacteria, lysed and the reactions quenched. The cell contents are then analysed as before. The dynamics of *in vivo* pathways could be investigated by challenging the culture with a potent antibacterial targeting the beginning of the pathway, fosfomycin for example [26]. Samples could then be taken at suitable intervals and analysed as described above; allowing the depletion of substrates within the pathway to be followed.

8.3.3 Comparison of pathway dynamics as a tool in antibacterial development

In silico models were used to compare the dynamics of the *E. coli* and *S. pneumoniae* pathways (Section 7.2.4). Since the enzyme concentrations used are not representative of *in vivo* concentrations, these results only allow the relative efficiencies of the enzymes to be compared. Nonetheless they highlight some differences between the two pathways which could be exploited in the design of antibacterials, notably that the early stage of the *S. pneumoniae* pathway may be particularly vulnerable to inhibitors.

Competitive inhibitors increase the apparent k_m with respect to the substrate with which the inhibitor competes [67]. Thus if this substrate accumulates to a sufficient degree normal rates of reaction will be restored. This provides a form of metabolic resistance to an inhibitor. As

such the effectiveness of a competitive inhibitor is maximised if the substrate with which it competes cannot accumulate; either because it is used by other reactions or due to feedback control within the pathway. A substrate like ATP, which is used by many reactions within the cell is thus a good target for a competitive inhibitor. Similarly the inhibition of MurA by UDPPEE and UDP-MurNAc provides an example of an exploitable feedback loop. A recently developed inhibitor of MurC, C-1 [63], was investigated in this context (Section 7.4.2). The interaction was shown to potentially reduce the concentration of C-1 needed to achieve a lethal reduction in overall production of UDP-MurNAc-pentapeptide.

8.3.4 Possible extensions to the models developed

The models developed in Section 7 describe an *in vitro* reconstruction of the cytoplasmic phase of the biosynthesis pathway. The *E. coli* model could be significantly improved by conducting kinetic characterisations of the enzymes under a consistent set of initial conditions. As alluded to throughout the previous sections both models could be modified to describe *in vivo* dynamics by incorporating *in vivo* enzyme concentrations. Clearly this requires the determination of these concentrations. Model predictions could then be compared to *in vivo* intermediate concentrations.

The model could then be extended in two directions. Given that enzyme concentrations have been incorporated it is perhaps natural to consider the mechanisms by which those enzyme concentrations are achieved. It is thought that the genes encoding the enzymes associated with the cytoplasmic phase are expressed constitutively, that is continuously, and that protein concentrations are insensitive to growth conditions [195, 186]. However these conclusions were drawn from studies which only considered *E. coli*. Gram-positive bacteria express two distinct MurA proteins [82]. This suggests potential for additional regulation of protein expression in this pathway. Furthermore in resistant *S. aureus* regulation of cell wall synthesis steps has been identified [196]. As such there may be, as yet unidentified, regulatory networks affecting

these genes which are not connected to growth conditions. Thus a model of the cytoplasmic phase in which enzyme concentrations are constant is not unrepresentative of *in vivo* conditions. However investigation of regulation on the transcriptional level seems apposite, especially for Gram-positive bacteria.

The alternative to modelling protein expression is to extend the pathway to later and earlier steps. The majority of the reaction steps in both directions are known [80, 119]; so such an extension may be relatively simple. Certainly the side reactions, providing the amino acids, and earlier stages of the hexosamine biosynthesis pathway could be incorporated with ease. The membrane bound phases may prove more challenging. These reactions can be kinetically characterised [125, 126, 127, 128, 124]; however the models used do not take into account the motion restrictions imposed by the cell membrane. Given that enzyme and at least one substrate are constrained to move only within two dimensions it is necessary to modify even the most general assumptions of mass action kinetics [130]. As such extension of this model to the membrane bound phases is a non-trivial problem. Finally it is worth noting that, as in the majority of such studies, the issues of crowding and compartmentalisation have been ignored. The theory necessary to handle these problems is under development [197, 198, 199]. Practical investigation of such problems is however experimentally challenging. Some effects of crowding can be reproduced using inert polymers [200]. However the more general problem, the increased efficiency inherent in a pathway composed of closely associated enzymes, is less tractable *in vitro*. Furthermore intracellular arrangement, and hence the degree of compartmentalisation, of the peptidoglycan biosynthesis enzymes is as yet uninvestigated. These issues constitute a significant area for further work assuming appropriate tools can be developed with which to investigate them.

8.4 Conclusions

A model of the cytoplasmic phase of the *S. pneumoniae* peptidoglycan biosynthesis pathway has been developed and validated. A similar unvalidated model has been constructed for the equivalent *E. coli* pathway using kinetic parameters taken from the available literature. Comparison of these models suggests different antibacterial strategies for each species. An inhibitory feedback loop was identified and modelled in the *S. pneumoniae* pathway. Experimental and simulated data were used to show that this process could be exploited to enhance the effectiveness of an inhibitor of the pathway.

The models constructed provide a basis from which more detailed models of these pathways can be developed. The results obtained illustrate the potential of pathway models to inform the development of antibacterials.

Direct parameter estimation from progress curves has been shown to be a viable alternative to typical approaches to kinetic characterisation of enzymes. The essential *a priori* analysis necessary before use of such techniques has been undertaken for a significant subset of biologically relevant reaction mechanisms. The input-output relationship approach to identifiability and indistinguishability proved highly effective in the study of quasi-steady state models, but less effective using the more complex, and reversible, transient models, where the Taylor series approach was generally more effective.

Nonetheless a structurally identifiable model is not guaranteed to yield high quality parameter estimates. The sensitivity of parameters to the time courses considered must also be assessed. It was found that estimation of the parameters for quasi-steady state ping-pong models required at least two time courses for optimal parameter sensitivity. The investigation of these issues, through parameter estimation to simulated time courses and local sensitivity analysis of each parameter, is just as important as the structural identifiability in the parameter estimation process.

BIBLIOGRAPHY

[1] AJ Alanis. Resistance to antibiotics: are we in the post-antibiotic era? *Archives of Medical Research*, 36(6):697–705, 2005.

[2] F Von Nussbaum, M Brands, B Hinzen, S Weigand, and D Habich. Antibacterial Natural Products in Medicinal Chemistry-Exodus or Revival? *Angew Chem Int Ed Engl*, 45:5072–5129, 2006.

[3] G H Cassell. Emergent antibiotic resistance: health risks and economic impact. *FEMS Immunol Med Microbiol*, 18(4):271–4, 1997.

[4] SE Cosgrove and Yehuda Carmeli. The impact of antimicrobial resistance on health and economic outcomes. *Clin Infect Dis*, 36(11):1433–7, 2003.

[5] IDSA (Infectious Diseases Society of America). Bad bugs, no drugs. As antibiotic discovery stagnates . . . a public health crisis brews., July 2004.

[6] SB Levy. The challenge of antibiotic resistance. *Scientific American*, 278(3):32–39, 1998.

[7] C Walsh. Molecular mechanisms that confer antibacterial drug resistance. *Nature*, 406(6797):775–81, 2000.

[8] E A Eady and J H Cove. Staphylococcal resistance revisited: community-acquired methicillin resistant *Staphyloccocus aureus* - an emerging problem for the management of skin and soft tissue infections. *Curr Opin Infect Dis*, 16:10324, 2003.

[9] MA Espinal. The global situation of MDR-TB. *Tuberculosis*, 83(1-3):44–51, 2003.

[10] D Morens, G Folkers, and A Fauci. The challenge of emerging and re-emerging infectious diseases. *Nature*, 2004.

[11] EM Zager and R McNerney. Multidrug-resistant tuberculosis. *BMC Infectious Diseases*, 8(1):10, 2008.

[12] J Clardy, MA Fischbach, and CT Walsh. New antibiotics from bacterial natural products. *Nature biotechnology*, 24(12):1541–1550, 2006.

[13] SJ Projan. Why is Big Pharma getting out of antibacterial drug discovery? *Curr Opin Microbiol*, 6(5):427–430, 2003.

[14] D Smith, K Daniel, Z Wang, WC Guida, TH Chan, and QP Dou. Docking studies and model development of tea polyphenol proteasome inhibitors: Applications to Rational Drug Design. *Proteins: Structure, Function and Bioinformatics*, 54:58–70, 2004.

[15] D B Kell. Systems biology, metabolic modelling and metabolomics in drug discovery and development. *Drug discovery today*, 11(23-24):1085–92, 2006.

[16] C R Cho, M Labow, M Reinhardt, J van Oostrum, and M C Peitsch. The application of systems biology to drug discovery. *Curr Opin Chem Biol*, 10(4):294–302, 2006.

[17] H Barreteau, Andreja Kovac, Audrey Boniface, Matej Sova, Stanislav Gobec, and Didier Blanot. Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS Microbiol Rev*, 32(2):168–207, 2008.

[18] M R Hicks, A Damianoglou, A Rodger, and T R Dafforn. Folding and membrane insertion of the pore-forming peptide gramicidin occur as a concerted process. *J Mol Biol*, 383(2):358–66, 2008.

[19] J J Champoux. DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem*, 70:369–413, 2001.

[20] I Chopra. Research and development of antibacterial agents. *Curr Opin Microbiol*, 1(5):495–501, 1998.

[21] S Admiraal, C Khosla, and C Walsh. A Switch for the Transfer of Substrate between Nonribosomal Peptide and Polyketide Modules of the Rifamycin Synthetase Assembly Line. *J Am Chem Soc*, 125(45):13664–13665, 2003.

[22] D I Edwards. Nitroimidazole drugs–action and resistance mechanisms. I. Mechanisms of action. *J Antimicrob Chemother*, 31(1):9–20, 1993.

[23] D Leitsch, D Kolarich, I B H Wilson, F Altmann, and M Duchêne. Nitroimidazole action in Entamoeba histolytica: a central role for thioredoxin reductase. *PLoS Biol*, 5(8):e211, 2007.

[24] P Nissen, J Hansen, N Ban, PB Moore, and TA Steitz. The structural basis of ribosome activity in peptide bond synthesis. *Science*, 289(5481):920, 2000.

[25] H J Rogers. *Microbial cell walls and membranes*, chapter Structure of peptidoglycan, pages 190–214. Chapman & Hall, 1980.

[26] F M Kahan, J S Kahan, P J Cassidy, and H Kropp. The mechanism of action of fosfomycin (phosphonomycin). *Ann N Y Acad Sci*, 235(0):364–86, 1974.

[27] M P Lambert and F C Neuhaus. Mechanism of D-cycloserine action: alanine racemase from Escherichia coli W. *J Bacteriol*, 110(3):978–87, 1972.

[28] L Silver. Does the cell wall of bacteria remain a viable source of targets for novel antibiotics? *Biochemical pharmacology*, 71:996–1005, 2006.

[29] X Fang, K Tiyanont, Y Zhang, J Wanner, D Boger, and S Walker. The mechanism of action of ramoplanin and enduracidin. *Mol Biosyst*, 2(1):69–76, 2006.

[30] D R Storm. Mechanism of bacitracin action: a specific lipid-peptide interaction. *Ann N Y Acad Sci*, 235(0):387–98, 1974.

[31] RR Yocum, JR Rasmussen, and JL Strominger. The mechanism of action of penicillin. Penicillin acylates the active site of Bacillus stearothermophilus D-alanine carboxypeptidase. *J Biol Chem*, 255(9):3977, 1980.

[32] J B Rake, R Gerber, R J Mehta, D J Newman, Y K Oh, C Phelen, M C Shearer, R D Sitrin, and L J Nisbet. Glycopeptide antibiotics: a mechanism-based screen employing a bacterial cell wall receptor mimetic. *J Antibiot*, 39(1):58–67, 1986.

[33] F Baquero and J Blázquez. Evolution of antibiotic resistance. *Trends in Ecology & Evolution*, 12(12):482–487, 1997.

[34] S P Denyer and J-Y Maillard. Cellular impermeability and uptake of biocides and antibiotics in Gram-negative bacteria. *J Appl Microbiol*, 92 Suppl:35S–45S, 2002.

[35] M M Ochs, M P McCusker, M Bains, and R E Hancock. Negative regulation of the Pseudomonas aeruginosa outer membrane porin OprD selective for imipenem and basic amino acids. *Antimicrob Agents Chemother*, 43(5):1085–90, 1999.

[36] E Dé, A Baslé, M Jaquinod, N Saint, M Malléa, G Molle, and J M Pagès. A new mechanism of antibiotic resistance in Enterobacteriaceae induced by a structural modification of the major porin. *Mol Microbiol*, 41(1):189–98, 2001.

[37] P A Lambert. Cellular impermeability and uptake of biocides and antibiotics in Gram-positive bacteria and mycobacteria. *Journal of applied microbiology*, 92 Suppl:46S–54S, 2002.

[38] L Cui, H Murakami, K Kuwahara-Arai, H Hanaki, and K Hiramatsu. Contribution of a thickened cell wall and its glutamine nonamidated component to the vancomycin resistance expressed by Staphylococcus aureus Mu50. *Antimicrob Agents Chemother*, 44(9):2276–85, 2000.

[39] IT Paulsen, MH Brown, and RA Skurray. Proton-dependent multidrug efflux systems. *Microbiology and Molecular Biology Reviews*, 60(4):575, 1996.

[40] J-M Pagès, M Masi, and J Barbe. Inhibitors of efflux pumps in Gram-negative bacteria. *Trends in Molecular Medicine*, 11(8):382–9, 2005.

[41] S B Levy. Active efflux mechanisms for antimicrobial resistance. *Antimicrob Agents Chemother*, 36(4):695–703, 1992.

[42] D Hogan and R Kolter. Why are bacteria refractory to antimicrobials? *Curr Opin Microbiol*, 5(5):472–7, 2002.

[43] M Stavri, L J V Piddock, and S Gibbons. Bacterial efflux pump inhibitors from natural sources. *J Antimicrob Chemother*, 59(6):1247–60, 2007.

[44] A A Medeiros. Evolution and dissemination of beta-lactamases accelerated by generations of beta-lactam antibiotics. *Clin Infect Dis*, 24 Suppl 1:S19–45, 1997.

[45] H C Neu. The crisis in antibiotic resistance. *Science*, 257(5073):1064–73, 1992.

[46] D M Livermore and N Woodford. The beta-lactamase threat in Enterobacteriaceae, Pseudomonas and Acinetobacter. *Trends in Microbiology*, 14(9):413–20, 2006.

[47] J.-E Hugonnet, L. W Tremblay, H. I Boshoff, C. E Barry, and J. S Blanchard. Meropenem-Clavulanate Is Effective Against Extensively Drug-Resistant Mycobacterium tuberculosis. *Science*, 323(5918):1215–1218, 2009.

[48] K J Shaw, P N Rather, R S Hare, and G H Miller. Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes. *Microbiol Rev*, 57(1):138–63, 1993.

[49] WC Hon, GA McKay, PR Thompson, RM Sweet, DSC Yang, GD Wright, and AM Berghuis. Structure of an enzyme required for aminoglycoside antibiotic resistance reveals homology to eukaryotic protein kinases. *Cell*, 89(6):887–895, 1997.

[50] G Dantas, MOA Sommer, RD Oluwasegun, and GM Church. Bacteria subsisting on antibiotics. *Science*, 320(5872):100, 2008.

[51] M D Song, M Wachi, M Doi, F Ishino, and M Matsuhashi. Evolution of an inducible penicillin-target protein in methicillin-resistant Staphylococcus aureus by gene fusion. *FEBS Lett*, 221(1):167–71, 1987.

[52] C T Walsh, S L Fisher, I S Park, M Prahalad, and Z Wu. Bacterial resistance to vancomycin: five genes and one missing hydrogen bond tell the story. *Chem Biol*, 3(1):21–8, 1996.

[53] Y Carmeli, N Troillet, A W Karchmer, and M H Samore. Health and economic outcomes of antibiotic resistance in Pseudomonas aeruginosa. *Arch Intern Med*, 159(10):1127–32, 1999.

[54] S E Cosgrove. The relationship between antimicrobial resistance and patient outcomes: mortality, length of hospital stay, and health care costs. *Clin Infect Dis*, 42 Suppl 2:S82–9, 2006.

[55] M L Cohen. Epidemiology of drug resistance: implications for a post-antimicrobial era. *Science*, 257(5073):1050–5, 1992.

[56] C E Phelps. Bug/drug resistance. Sometimes less is more. *Medical Care*, 27(2):194–203, 1989.

[57] A El Zoeiby, F Sanschagrin, and R C Levesque. Structure and function of the Mur enzymes: development of novel inhibitors. *Mol Microbiol*, 47(1):1–12, 2003.

[58] Miha Kotnik, Jan Humljan, Carlos Contreras-Martel, Marko Oblak, Katja Kristan, Mireille Hervé, Didier Blanot, Uros Urleb, Stanislav Gobec, Andréa Dessen, and Tom Solmajer. Structural and functional characterization of enantiomeric glutamic acid deriva-

tives as potential transition state analogue inhibitors of MurD ligase. *J Mol Biol*, 370(1):107–15, 2007.

[59] T Bratkovic, M Lunder, U Urleb, and B Strukelj. Peptide inhibitors of MurD and MurE, essential enzymes of bacterial cell wall biosynthesis. *J Basic Microbiol*, 48(3):202–6, 2008.

[60] C Paradis-Bleau, M, L Boudreault, A Lloyd, F Sanschagrin, T D H Bugg, and R C Levesque. Selection of peptide inhibitors against the Pseudomonas aeruginosa MurD cell wall enzyme. *Peptides*, 27(7):1693–700, 2006.

[61] C Paradis-Bleau, A Lloyd, F Sanschagrin, T Clarke, A Blewett, T D H Bugg, and R C Levesque. Phage display-derived inhibitor of the essential cell wall biosynthesis enzyme MurF. *BMC Biochem*, 9:33, 2008.

[62] R Sink, A Kovac, T Tomasić, V Rupnik, A Boniface, J Bostock, I Chopra, D Blanot, L P Masic, S Gobec, and A Zega. Synthesis and biological evaluation of N-acylhydrazones as inhibitors of MurC and MurD ligases. *Chem Med Chem*, 3(9):1362–70, 2008.

[63] L E Zawadzke, M Noreia, C R Desbonnet, H Wang, K Freeman-Cook, and T J Dougherty. Identification of an Inhibitor of the MurC Enzyme, Which Catalyzes and Essential Step in the Peptidoglycan Precursor Synthesis Pathway. *Assay Drug Dev*, 6(1):93–103, 2008.

[64] M E Tanner, S Vaganay, J van Heijenoort, and D Blanot. Phosphinate Inhibitors of the D-Glutamic Acid-Adding Enzyme of Peptidoglycan Biosynthesis. *J Org Chem*, 61(5):1756–1760, 1996.

[65] R I Christopherson and R G Duggleby. Metabolic resistance: the protection of enzymes against drugs which are tight-binding inhibitors by the accumulation of substrate. *Eur J Biochem*, 134(2):331–5, 1983.

[66] R G Duggleby and R I Christopherson. Metabolic resistance to tight-binding inhibitors

of enzymes involved in the de novo pyrimidine pathway. Simulation of time-dependent effects. *Eur J Biochem*, 143(1):221–6, 1984.

[67] R Eisenthal and A Cornish-Bowden. Prospects for Antiparasitic Drugs: The case of trypanosoma brucei, the causative agent of African Sleeping Sickness . *J Biol Chem*, 273(10):5500–5, 1998.

[68] C Gram. Uberdle isolirte Farbung der Schizomyceten in Schnltt-und Trockenpraparaten. *Fortschr Med*, 2:185–9, 1884.

[69] A Popescu and R J Doyle. The Gram stain after more than a century. *Biotechnic and histochemistry*, 71(3):145–51, 1996.

[70] T D Bugg and C T Walsh. Intracellular steps of bacterial cell wall peptidoglycan biosynthesis: enzymology, antibiotics, and antibiotic resistance. *Natural product reports*, 9(3):199–215, 1992.

[71] W Vollmer, D Blanot, and M A De Pedro. Peptidoglycan structure and architecture. *FEMS Microbiol Rev*, 32(2):149–167, 2008.

[72] W Vollmer. Structural variation in the glycan strands of bacterial peptidoglycan. *FEMS Microbiol Rev*, 32(2):287–306, 2008.

[73] D J Tipper, J L Strominger, and J C Ensign. Structure of the cell wall of Staphylococcus aureus, strain Copenhagen. VII. Mode of action of the bacteriolytic peptidase from Myxobacter and the isolation of intact cell wall polysaccharides. *Biochemistry*, 6(3):906–20, 1967.

[74] W Vollmer and U Bertsche. Murein (peptidoglycan) structure, architecture and biosynthesis in Escherichia coli. *Biochim Biophys Acta*, 1778(9):1714–34, 2008.

[75] J B Ward. The chain length of the glycans in bacterial cell walls. *Biochem J*, 133(2):395–8, 1973.

[76] K H Schleifer and O Kandler. Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol Rev*, 36(4):407–77, 1972.

[77] T D Bugg, S Dutka-Malen, M Arthur, P Courvalin, and C T Walsh. Identification of vancomycin resistance protein VanA as a D-alanine:D-alanine ligase of altered substrate specificity. *Biochemistry*, 30(8):2017–21, 1991.

[78] M Arthur, C Molinas, T D Bugg, G D Wright, C T Walsh, and P Courvalin. Evidence for in vivo incorporation of D-lactate into peptidoglycan precursors of vancomycin-resistant enterococci. *Antimicrob Agents Chemother*, 36(4):867–9, 1992.

[79] DJ Waxman and JL Strominger. Penicillin-Binding Proteins and the Mechanism of Action of Beta-Lactam Antibiotics1. *Annu Rev Biochem*, 52(1):825–869, 1983.

[80] S Milewski. Glucosamine-6-phosphate synthase–the multi-facets enzyme. *Biochim Biophys Acta*, 1597(2):173–92, 2002.

[81] J L Marquardt, D A Siegele, R Kolter, and C T Walsh. Cloning and sequencing of Escherichia coli murZ and purification of its product, a UDP-N-acetylglucosamine enolpyruvyl transferase. *J Bacteriol*, 174(17):5748–52, 1992.

[82] W Du, J R Brown, D R Sylvester, J Huang, A F Chalker, C Y So, D J Holmes, D J Payne, and N G Wallis. Two active forms of UDP-N-acetylglucosamine enolpyruvyl transferase in gram-positive bacteria. *J Bacteriol*, 182(15):4146–52, 2000.

[83] JL Marquardt, ED Brown, CT Walsh, and KS Anderson. Isolation and structural elucidation of a tetrahedral intermediate in the UDP-N-acetylglucosamine enolpyruvoyl transferase enzymic pathway. *J Am Chem Soc*, 115(22):10398–10399, 1993.

[84] E D Brown, J L Marquardt, J P Lee, C T Walsh, and K S Anderson. Detection and characterization of a phospholactoyl-enzyme adduct in the reaction catalyzed by UDP-N-acetylglucosamine enolpyruvoyl transferase, MurZ. *Biochemistry*, 33(35):10638–45, 1994.

[85] T E Benson, J L Marquardt, A C Marquardt, F A Etzkorn, and C T Walsh. Overexpression, purification, and mechanistic study of UDP-N-acetylenolpyruvylglucosamine reductase. *Biochemistry*, 32(8):2024–30, 1993.

[86] D R Sylvester, E Alvarez, A Patel, K Ratnam, H Kallender, and N G Wallis. Identification and characterization of UDP-N-acetylenolpyruvylglucosamine reductase (MurB) from the Gram-positive pathogen Streptococcus pneumoniae. *Biochem J*, 355(Pt 2):431–5, 2001.

[87] T Skarzynski, A Mistry, A Wonacott, S E Hutchinson, V A Kelly, and K Duncan. Structure of UDP-N-acetylglucosamine enolpyuvyl transferase, an enzyme essential for the synthesis of bacterial peptidoglycan, complexed with substrate UDP-N-acetylglucosamine and the drug fosfomycin. *Structure*, 4:1465–74, 1996.

[88] S Eschenburg and E Schonbrunn. Comparative X-ray analysis of the un-liganded fosfomycin-target MurA. *Proteins*, 40:290–8, 2000.

[89] J L Marquardt, E D Brown, W S Lane, T M Haley, Y Ichikawa, C H Wong, and C T Walsh. Kinetics, stoichiometry, and identification of the reactive thiolate in the inactivation of UDP-GlcNAc enolpyruvoyl transferase by the antibiotic fosfomycin. *Biochemistry*, 33(35):10646–51, 1994.

[90] C Wanke and N Amrhein. Evidence that the reaction of the UDP-N-acetylglucosamine 1-carboxyvinyltransferase proceeds through the O-phosphothioketal of pyruvic acid bound to Cys115 of the enzyme. *Eur J Biochem*, 218(3):861–70, 1993.

[91] D H Kim, W J Lees, K E Kempsell, W S Lane, K Duncan, and C T Walsh. Characterization of a Cys115 to Asp substitution in the Escherichia coli cell wall biosynthetic enzyme UDP-GlcNAc enolpyruvyl transferase (MurA) that confers resistance to inactivation by the antibiotic fosfomycin. *Biochemistry*, 35(15):4923–8, 1996.

[92] K A De Smet, K E Kempsell, A Gallagher, K Duncan, and D B Young. Alteration

of a single amino acid residue reverses fosfomycin resistance of recombinant MurA from Mycobacterium tuberculosis. *Microbiology*, 145 (Pt 11):3177–84, 1999.

[93] C Ramilo, R J Appleyard, C Wanke, F Krekel, F Amrhein, and J N Evans. Detection of the covalent intermediate of UDP-N-acetyl glucosamine enolpyruvyl transferase by solution-state and time resolved solid-state NMR spectroscopy. *Biochemistry*, 33:15071–9, 1994.

[94] T E Benson, C T Walsh, and J M Hogle. The structure of the substrate-free form of MurB, an essential enzyme for the synthesis of bacterial cell walls. *Structure*, 4(1):47–54, 1996.

[95] A M Dhalla, J Yanchunas, H T Ho, P J Falk, J J Villafranca, and J G Robertson. Steady-state kinetic mechanism of Escherichia coli UDP-N-acetylenolpyruvylglucosamine reductase. *Biochemistry*, 34(16):5390–402, 1995.

[96] K L Constantine, L Mueller, V Goldfarb, M Wittekind, W J Metzler, and J Yanchunas Jr. Characterisation of NADP+ binding to perdeuterated MurB: backbone atom NMR assignments and chemical shift changes. *J Mol Biol*, 267:1223–46, 1997.

[97] A Bouhss, D Mengin-Lecreulx, D Blanot, J van Heijenoort, and C Parquet. Invariant amino acids in the Mur peptide synthetases of bacterial peptidoglycan synthesis and their modification by site-directed mutagenesis in the UDP-MurNAc:L-alanine ligase from Escherichia coli. *Biochemistry*, 36(39):11556–63, 1997.

[98] D Liger, A Masson, D Blanot, J Heijenoort, and et al. Over-production, Purification and Properties of the Uridine-diphosphate-N-Acetylmuramate: l-alanine ligase from Escherichia coli. *Eur J Biochem*, 230:80–7, 1995.

[99] F Nosal, A Masson, R Legrand, D Blanot, B Schoot, J van Heijenoort, and C Parquet. Site-directed mutagenesis and chemical modification of the two cysteine residues of the

UDP-N-acetylmuramoyl:L-alanine ligase of Escherichia coli. *FEBS Lett*, 426(3):309–13, 1998.

[100] F Pratviel-Sosa, D Mengin-Lecreulx, and J van Heijenoort. Over-production, purification and properties of the uridine diphosphate N-acetylmuramoyl-L-alanine:D-glutamate ligase from Escherichia coli. *Eur J Biochem*, 202(3):1169–76, 1991.

[101] C Michaud, D Mengin-Lecreulx, J van Heijenoort, and D Blanot. Over-production, purification and properties of the uridine-diphosphate-N-acetylmuramoyl-L-alanyl-D-glutamate: meso-2,6-diaminopimelate ligase from Escherichia coli. *Eur J Biochem*, 194(3):853–61, 1990.

[102] M S Anderson, S S Eveland, H R Onishi, and D L Pompliano. Kinetic mechanism of the Escherichia coli UDPMurNAc-tripeptide D-alanyl-D-alanine-adding enzyme: use of a glutathione S-transferase fusion. *Biochemistry*, 35(50):16264–9, 1996.

[103] J Emanuele Jr, H Jin, B Jacobson, C Chang, and et al. Kinetic and crystallographic studies of Escherichia coli UDP-N-acetylmuramate: L-alanine ligase. *Protein Sci*, 2:2566–74, 1996.

[104] C A Smith. Structure, Function and Dynamics in the mur Family of Bacterial Cell Wall Ligases. *J Mol Biol*, 362(4):640 – 55, 2006.

[105] A El Zoeiby, F Sanschagrin, J Lamoureux, A Darveau, and R C Levesque. Cloning, over-expression and purification of Pseudomonas aeruginosa murC encoding uridine diphosphate N-acetylmuramate: L-alanine ligase. *FEMS Microbiol Lett*, 183(2):281–8, 2000.

[106] A Bouhss, S Dementin, C Parquet, D Mengin-Lecreulx, J A Bertrand, D Le Beller, O Dideberg, J van Heijenoort, and D Blanot. Role of the ortholog and paralog amino acid invariants in the active site of the UDP-MurNAc-L-alanine:D-glutamate ligase (MurD). *Biochemistry*, 38(38):12240–7, 1999.

[107] S Dementin, A Bouhss, G Auger, C Parquet, D Mengin-Lecreulx, O Dideberg, J van Heijenoort, and D Blanot. Evidence of a functional requirement for a carbamoylated lysine residue in MurD, MurE and MurF synthetases as established by chemical rescue experiments. *Eur J Biochem*, 268(22):5800–7, 2001.

[108] JA Bertrand, E Fanchon, L Martin, L Chantalat, G Auger, D Blanot, J van Heijenoort, and O Dideberg. "Open" structures of MurD: domain movements and structural similarities with folylpolyglutamate synthetase. *J Mol Biol*, 301(5):1257–1266, 2000.

[109] A Bouhss, S Dementin, J van Heijenoort, C Parquet, and D Blanot. Formation of adenosine 5'-tetraphosphate from the acyl phosphate intermediate: a difference between the MurC and MurD synthetases of Escherichia coli. *FEBS Lett*, 453(1-2):15–9, 1999.

[110] A Blewett. *The substrate specificity of peptidoglycan biosynthesis enzymes from* Streptococcus pneumoniae. PhD thesis, University of Warwick, 2005.

[111] G Rosso, K Takashima, and E Adams. Coenzyme content of purified alanine racemase from Pseudomonas. *Biochem Biophys Res Commun*, 34(1):134–40, 1969.

[112] P Doublet, J van Heijenoort, J P Bohin, and D Mengin-Lecreulx. The murI gene of Escherichia coli is an essential gene that encodes a glutamate racemase activity. *J Bacteriol*, 175(10):2970–9, 1993.

[113] M A Taal, S E Sedelnikova, S N Ruzheinikov, P J Baker, and D W Rice. Expression, purification and preliminary X-ray analysis of crystals of Bacillus subtilis glutamate racemase. *Acta Crystallogr D Biol Crystallogr*, 60(Pt 11):2031–4, 2004.

[114] T D Bugg, G D Wright, S Dutka-Malen, M Arthur, P Courvalin, and C T Walsh. Molecular basis for vancomycin resistance in Enterococcus faecium BM4147: biosynthesis of a depsipeptide peptidoglycan precursor by vancomycin resistance proteins VanH and VanA. *Biochemistry*, 30(43):10408–15, 1991.

[115] A de Dios, L Prieto, J Martin, and A Rubio. 4-Substituted D-Glutamic Acid Analogues: The First Potent Inhibitors of Glutamate Racemase(MurI) Enzye with Antibacterial Activity. *Journal of Medicinal Chemistry*, 45:4559–70, 2002.

[116] D Mengin-Lecreulx, B Flouret, and J van Heijenoort. Cytoplasmic steps of peptidoglycan synthesis in Escherichia coli. *J Bacteriol*, 151(3):1109–17, 1982.

[117] D Mengin-Lecreulx, L Texier, M Rousseau, and J van Heijenoort. The murG gene of Escherichia coli codes for the UDP-N-acetylglucosamine: N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase involved in the membrane steps of peptidoglycan synthesis. *J Bacteriol*, 173(15):4625–36, 1991.

[118] M H Buckstein, J He, and H Rubin. Characterization of nucleotide pools as a function of physiological state in Escherichia coli. *J Bacteriol*, 190(2):718–26, 2008.

[119] A Bouhss, A E Trunkfield, T D H Bugg, and D Mengin-Lecreulx. The biosynthesis of peptidoglycan lipid-linked intermediates. *FEMS Microbiol Rev*, 32(2):208–233, 2008.

[120] N Ruiz. Bioinformatics identification of MurJ (MviN) as the peptidoglycan lipid II flippase in Escherichia coli. *Proc Natl Acad Sci USA*, 105(40):15553–7, 2008.

[121] A Inoue, Y Murata, H Takahashi, N Tsuji, S Fujisaki, and J Kato. Involvement of an essential gene, mviN, in murein synthesis in Escherichia coli. *J Bacteriol*, 190(21):7298–301, 2008.

[122] A Bouhss, D Mengin-Lecreulx, D Le Beller, and J van Heijenoort. Topological analysis of the MraY protein catalysing the first membrane step of peptidoglycan synthesis. *Mol Microbiol*, 34(3):576–85, 1999.

[123] S Ha, D Walker, Y Shi, and S Walker. The 1.9 A crystal structure of Escherichia coli MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Sci*, 9(6):1045–52, 2000.

[124] E Sauvage, F Kerff, M Terrak, J A Ayala, and P Charlier. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS Microbiol Rev*, 32(2):234–58, 2008.

[125] A Bouhss, M Crouvoisier, D Blanot, and D Mengin-Lecreulx. Purification and characterization of the bacterial MraY translocase catalyzing the first membrane step of peptidoglycan biosynthesis. *J Biol Chem*, 279(29):29974–80, 2004.

[126] S Ha, E Chang, MC Lo, H Men, P Park, M Ge, and S Walker. The kinetic characterization of Escherichia coli MurG using synthetic substrate analogues. *J Am Chem Soc*, 3:21–28, 1996.

[127] L Chen, H Men, S Ha, XY Ye, L Brunner, Y Hu, and S Walker. Intrinsic Lipid Preferences and Kinetic Mechanism of Escherichia coli MurG. *Biochemistry*, 41(21):6824–6833, 2002.

[128] G Auger, J van Heijenoort, D Mengin-Lecreulx, and D Blanot. A MurG assay which utilises a synthetic analogue of lipid I. *FEMS Microbiology Letters*, 219(1):115–119, 2003.

[129] U Kohlrausch, F B Wientjes, and J V Höltje. Determination of murein precursors during the cell cycle of Escherichia coli. *J Gen Microbiol*, 135(6):1499–506, 1989.

[130] R Kopelman. Fractal Reaction Kinetics. *Science*, 241(4873):1620–1626, 1988.

[131] JV Holtje. Growth of the stress-bearing and shape-maintaining murein sacculus of Escherichia coli. *Microbiology and Molecular Biology Reviews*, 62(1):181, 1998.

[132] W Vollmer, B Joris, P Charlier, and S Foster. Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiol Rev*, 32(2):259–286, 2008.

[133] A Antignac, K Sieradzki, and A Tomasz. Perturbation of cell wall synthesis suppresses autolysis in Staphylococcus aureus: evidence for coregulation of cell wall synthetic and hydrolytic enzymes. *J Bacteriol*, 189(21):7573–80, 2007.

[134] J T Park. Why does Escherichia coli recycle its cell wall peptides? *Mol Microbiol*, 17(3):421–6, 1995.

[135] E W Goodell and U Schwarz. Release of cell wall peptides into culture medium by exponentially growing Escherichia coli. *J Bacteriol*, 162(1):391–7, 1985.

[136] D Mengin-Lecreulx, J van Heijenoort, and J T Park. Identification of the mpl gene encoding UDP-N-acetylmuramate: L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase in Escherichia coli and its role in recycling of cell wall peptidoglycan. *J Bacteriol*, 178(18):5347–52, 1996.

[137] W Vötsch and M F Templin. Characterization of a beta -N-acetylglucosaminidase of Escherichia coli and elucidation of its role in muropeptide recycling and beta -lactamase induction. *J Biol Chem*, 275(50):39032–8, 2000.

[138] A Asgarali, K A Stubbs, A Oliver, D J Vocadlo, and B L Mark. Inactivation of the glycoside hydrolase NagZ attenuates antipseudomonal beta-lactam resistance in Pseudomonas aeruginosa. *Antimicrob Agents Chemother*, 53(6):2274–82, 2009.

[139] F Llaneras and J Picó. Stoichiometric modelling of cell metabolism. *J Biosci and Bioeng*, 105(1):1 – 11, 2008.

[140] A K Gombert and J Nielsen. Mathematical modelling of metabolism. *Current Opinion in Biotechnology*, 11(2):180 – 6, 2000.

[141] JD Varner. Large-scale prediction of phenotype: concept. *Biotechnology and bioengineering*, 69(6):664–678, 2000.

[142] BN Kholodenko. Cell signalling dynamics in time and space. *Nature reviews. Molecular cell biology*, 7(3):165, 2006.

[143] C J Morton-Firth and D Bray. Predicting temporal fluctuations in an intracellular signalling pathway. *J Theor Biol*, 192(1):117–28, 1998.

[144] M Dixon and E C Webb. *Enzymes*. Longman Group Ltd, 3rd edition, 1979.

[145] I M Henderson and T D Bugg. Pre-steady-state kinetic analysis of 2-hydroxy-6-keto-nona-2,4-diene-1,9-dioic acid 5,6-hydrolase: kinetic evidence for enol/keto tautomerization. *Biochemistry*, 36(40):12252–8, 1997.

[146] D Bates and C Frieden. Treatment of Enzyme Kinetic Data III. The use of the full time course of a reaction, as examined By computer simulation, in defining enzyme mechanisms. *J Biol Chem*, 248(22):7878–84, 1973.

[147] C Zimmerle and C Frieden. Analysis of progress curves by simulations generated by numerical integration. *Biochem J*, 258:381–387, 1989.

[148] R Bellman and K J Astrom. On Structural Identifiability. *Math Biosci*, 7:329–39, 1970.

[149] R Bellman. Topics in pharmacokinetics–IV: Approximation in Process Space and Fitting by Sums of Exponentials. *Math Biosci*, 14:45–47, 1972.

[150] KR Godfrey and MJ Chapman. The problem of model indistinguishability in pharmacokinetics. *Journal of Pharmacokinetics and Pharmacodynamics*, 17(2):229–267, 1989.

[151] S G Grant, J Jessee, F R Bloom, and D Hanahan. Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proceedings of the National Academy of Sciences of the United States of America*, 87(12):4645–4649, 1990.

[152] PJ Lopez, I Marchand, SA Joyce, and M Dreyfus. The C-terminal half of RNase E, which organizes the Escherichia coli degradosome, participates in mRNA degradation but not rRNA processing in vivo. *Mol Microbiol*, 33(1):188–99, 1999.

[153] D Hanahan. Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol*, 166(4):557–80, 1983.

[154] FW Studier. Protein Production by auto-induction in high density shaking cultures. *Protein Expr Purif*, 41(1):207–34, 2005.

[155] U K Laemmli. Cleavage of Structural proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, 227(5259):680–5, 1970.

[156] D E Wampler and E W Westhead. Two aspartokinases from Escherichia coli. Nature of the inhibition and molecular changes accompanying reversible inactivation. *Biochemistry*, 7(5):1661–71, 1968.

[157] M Webb. A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems. *Proceedings of the National Academy of Sciences*, 1992.

[158] M B Monagan, K O Geddes, K M Heal, G Labahn, S M Vorkoetter, J McCarron, and P DeMarco. *Maple 10 Programming Guide*. Maplesoft, Waterloo ON, Canada, 2005.

[159] S Hoops, S Sahle, R Gauges, C Lee, J Pahle, N Simus, M Singhal, L Xu, P Mendes, and U Kummer. COPASI - a COmplex PAthway SImulator. *Bioinformatics*, 22:3067–74, 2006.

[160] M Chappell, N Evans, K Godfrey, and M Chapman. Structural identifiability of controlled state space systems: aquasiautomated methodology for generating identifiable reparameterisation of unidentifiable systems. *Symbolic Computation for Control (Ref. No. 1999/088)*, 1999.

[161] N Evans, L White, M Chapman, and K Godfrey. The structural identifiability of the susceptible infected recovered model with seasonal forcing. *Math Biosci*, 194:175–97, 2005.

[162] N Evans, M Chapman, M Chappell, and K Godfrey. Identifiability of uncontrolled nonlinear rational systems. *Automatica*, 38:1799–805, 2002.

[163] H Pohjanpalo. System Identifiability Based on the Power Series Expansion of the Solution. *Math Biosci*, 41:21–33, 1978.

[164] S Vajda. Structural identifiability of linear, bilinear, polynomial and rational systems. In L Ljung and A Titli, editors, *Proceedings of the 9th IFAC World Congress*, page 107, 1984.

[165] G Margaria, E Riccomagno, MJ Chappell, and HP Wynn. Differential algebra methods for the study of the structural identifiability of rational function state-space models in the biosciences. *Math Biosci*, 174(1):1–26, 2001.

[166] H Pohjanpalo. Identifiability of deterministic differential models in state space. Research Report 56, Technical Research Centre of Finland, 1982.

[167] JF Ritt. *Differential Algebra*. AMS Colloquium Publications, 1st edition, 1950.

[168] F Boulier, D Lazard, F Ollivier, and M Petitot. Representation for the radical of a finitely generated differential ideal. In *ISSAC'95: Proceedings of the 1995 International Symposium on Symbolic and Algebraic Computation*, pages 158–166, 1995.

[169] E Hubert. *Étude algébrique et algorithmique des singularités des équations différntielles implicites*. PhD thesis, Institut National Polytechnique de Grenoble, 1997.

[170] F Boulier, D Lazard, F Ollivier, and M Petitot. Computing representations for radicals of finitely generated differential ideals. *Applicable Algebra in Engineering*, 2009.

[171] M Krusemeyer. Why Does the Wronskian Work. *Am Math Monthly*, 95(1):46–49, 1988.

[172] S Schnell, MJ Chappell, ND Evans, and MR Roussel. The mechanism distinguishability problem in biochemical kinetics: The single-enzyme, single-substrate reaction as a case study. *Comptes rendus-Biologies*, 329(1):51–61, 2006.

[173] *Facsimile (Version 4.0) User Guide*. AEA Technology, Harwell Laboratory, Didcot, Oxfordshire, UK, 1995.

[174] DW Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J Soc Ind App Math*, 11(2):431–41, 1963.

[175] D B Fogel, L J Fogel, and J W Atmar. Meta-evolutionary programming. In *25th Asiloma Conference on Signals, Systems and Computers*, pages 540–5, 1992.

[176] L Petzold. Automatic Selection of Methods for SOlving Stiff and Nonstiff Systems of Ordinary Differential Equations. *SIAM Journal of Scientific and Statistical Computing*, 4(1):136–48, 1983.

[177] K Radhakrishnan and A C Hindmarsh. Description and Use of LSODE, the Livermore Solver for Ordinary Differential Equations. Reference Publication 1327, NASA, 1993.

[178] S Eschenburg, W Kabsch, ML Healy, and E Schonbrunn. A New View of the Mechanisms of UDP-N-Acetylglucosamine Enolpyruvyl Transferase(MurA) and 5-Enolpyruvylshikimate-3-phosphate Synthase(AroA) Derived from X-ray Structures of Their Tetrahedral Reaction Intermediate States. *J Biol Chem*, 278(49):49215–49222, 2003.

[179] B Chance. Rapid and Sensitive Spectrophotometry. I. The Accelerated and Stopped-Flow Methods for the Measurement of the Reaction Kinetics and Spectra of Unstable Compounds in the Visible Region of the Spectrum. *Review of Scientific Instruments*, 22:619–627, 1951.

[180] J Südi. Macroscopic rate constants involved in the formation and interconversion of the two central enzyme-substrate complexes of the lactate dyhrogenase turnover. *Biochem J*, 139:261 – 71, 1974.

[181] B V Plapp. On calculation of rate and dissociation constants from kinetic constants for the ordered bi bi mechanism of liver alcohol dehydrogenase. *Arch Biochem Biophys*, 156:112–4, 1973.

[182] H d'A Heck. Porcine heart lactate dehydrogenase: optical rotatory dispersion, thermodynamics, and kinetics of binding reactions. *J Biol Chem*, 244(16):4875 – 81, 1969.

[183] N G Bennett and H Gutfreund. The kinetics of the interconversion of intermediates of the reaction of pig muscle lactate dehydrogenase with oxidised nictoninamide-adenine dinucleotide and lactate. *Biochem J*, 135:81–5, 1973.

[184] J J Holbrook and H Gutfreund. Approaches to the study of enzyme mechanisms lactate dehydrogenase. *FEBS Lett*, 31(2):157–69, 1973.

[185] R A Stinson. Equilibrium binding of nicotinamide nucleotides to lactate dehydrogenase. *Biochem J*, 131:719–28, 1973.

[186] S Mizyed, A Oddone, B Byczynski, D W Hughes, and P J Berti. UDP-N-acetylmuramic acid (UDP-MurNAc) is a potent inhibitor of MurA (enolpyruvyl-UDP-GlcNAc synthase). *Biochemistry*, 44(10):4011–7, 2005.

[187] T B Clarke, F Kawai, S-Y Park, J R H Tame, C G Dowson, and D I Roper. Mutational analysis of the substrate specificity of Escherichia coli penicillin binding protein 4. *Biochemistry*, 48(12):2675–83, 2009.

[188] I M Keseler, C Bonavides-Martínez, J Collado-Vides, S Gama-Castro, R P Gunsalus, D A Johnson, M Krummenacker, L M Nolan, S Paley, I T Paulsen, M Peralta-Gil, A Santos-Zavaleta, A G Shearer, and P D Karp. EcoCyc: a comprehensive view of Escherichia coli biology. *Nucleic Acids Research*, 37(Database issue):D464–70, 2009.

[189] L Ljung and T Glad. On global identifiability for arbitrary model parametrizations. *Automatica*, 30(2):265–276, 1994.

[190] M Saccomani and G Bellu. DAISY: An efficient tool to test global identifiability. Some case studies. *Control and Automation*, 2008.

[191] C L Baird, M S Gordon, D M Andrenyak, J F Marecek, and J E Lindsley. The ATPase reaction cycle of yeast DNA topoisomerase II. Slow rates of ATP resynthesis and P(i) release. *J Biol Chem*, 276(30):27893–8, 2001.

[192] M Brune, J L Hunter, J E Corrie, and M R Webb. Direct, real-time measurement of rapid inorganic phosphate release using a novel fluorescent probe and its application to actomyosin subfragment 1 ATPase. *Biochemistry*, 33(27):8262–71, 1994.

[193] M L Moyer, S P Gilbert, and K A Johnson. Pathway of ATP hydrolysis by monomeric and dimeric kinesin. *Biochemistry*, 37(3):800–13, 1998.

[194] T E Benson, C T Walsh, and V Massey. Kinetic characterization of wild-type and S229A mutant MurB: evidence for the role of Ser 229 as a general acid. *Biochemistry*, 36(4):796–805, 1997.

[195] D Mengin-Lecreulx and J van Heijenoort. Effect of growth conditions on peptidoglycan content and cytoplasmic steps of its biosynthesis in Escherichia coli. *J Bacteriol*, 163(1):208–12, 1985.

[196] A M Beltramini, C D Mukhopadhyay, and V Pancholi. Modulation of cell wall structure and antimicrobial susceptibility by a Staphylococcus aureus eukaryote-like serine/threonine kinase and phosphatase. *Infect Immun*, 77(4):1406–16, 2009.

[197] T C Laurent. Enzyme reactions in polymer media. *Eur J Biochem*, 21(4):498–506, 1971.

[198] F Horn and R Jackson. General mass action kinetics. *Archive for Rational Mechanics and Analysis*, 47(2):81–116, 1972.

[199] S Schnell and T E Turner. Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws. *Progress in biophysics and molecular biology*, 85(2-3):235–60, 2004.

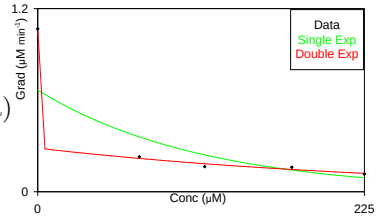
[200] I Pozdnyakova and P Wittung-Stafshede. Non-linear effects of macromolecular crowding on enzymatic activity of multi-copper oxidase. *BBA - Proteins and Proteomics*, 1804(4):740–744, 2010.

APPENDIX

A. DETERMINATION OF BACKGROUND NAD^+ PRODUCTION

The raw data presented in Section 7.2.1 were affected by background production of NAD^+. Measurements were made for approximately 2 minutes prior to the addition of MurA in which base rates were observed for each of the four reaction mixtures. (Given the sampling rates described in Section 2.4.2 this corresponds to between twelve and thirty measurements dependent on the number of time courses being measured.) In addition as the reactions reached completion the rate of NAD^+ production did not converge to zero but rather to a positive non-zero rate. These start and end background rates differed by factors of between five and ten and as such it was insufficient merely to estimate NAD^+ production from either rate. Furthermore the end rates differed in a manner which appeared dependent on the amount of NAD^+ produced (Figure A.1(a)). This suggests that the background rate may arise from oxidation of NADH by LDH in the absence of pyruvate in a concentration dependent manner. As such a single exponential model, $f([\mathrm{NAD}^+]) = Ae^{B[\mathrm{NAD}^+]}$, relating concentration of NAD^+ ($[\mathrm{NAD}^+]$) was fitted to the data (Figure A.1(a)). This model did not adequately fit the data obtained so a double exponential model, $f([\mathrm{NAD}^+]) = A_1e^{B_1[\mathrm{NAD}^+]} + A_2e^{B_2[\mathrm{NAD}^+]}$, was also fitted, with better results (Figure A.1(a)). Finally a third model referred to as a linear model was constructed, to create a smooth transition between the two rates, $f(t) = r_0\frac{t_1-t}{t_1} + r_1\frac{t}{t_1}$, where r_0 and r_1 are the start and end rates, t is time and t_1 the end time. The resulting predictions of background rate and NAD^+ production are presented below (Figure A.2(a)-(f)).

(a)



(b)

Model	A_1	A_2	B_1	B_2
1 Exp	1.07	-	-1.16x10^{-2}	-
2 Exp	0.782	0.286	-1.32	-3.84x10^{-3}

Figure A.1: Construction of exponential models of background NAD$^+$ production. Graph (a) plots observed background rate against concentration of NAD$^+$ produced, single exponential and double exponential fits, curves are colour coded according to the key provided. Table (b) contains the constants used to construct the single and double exponential plots.

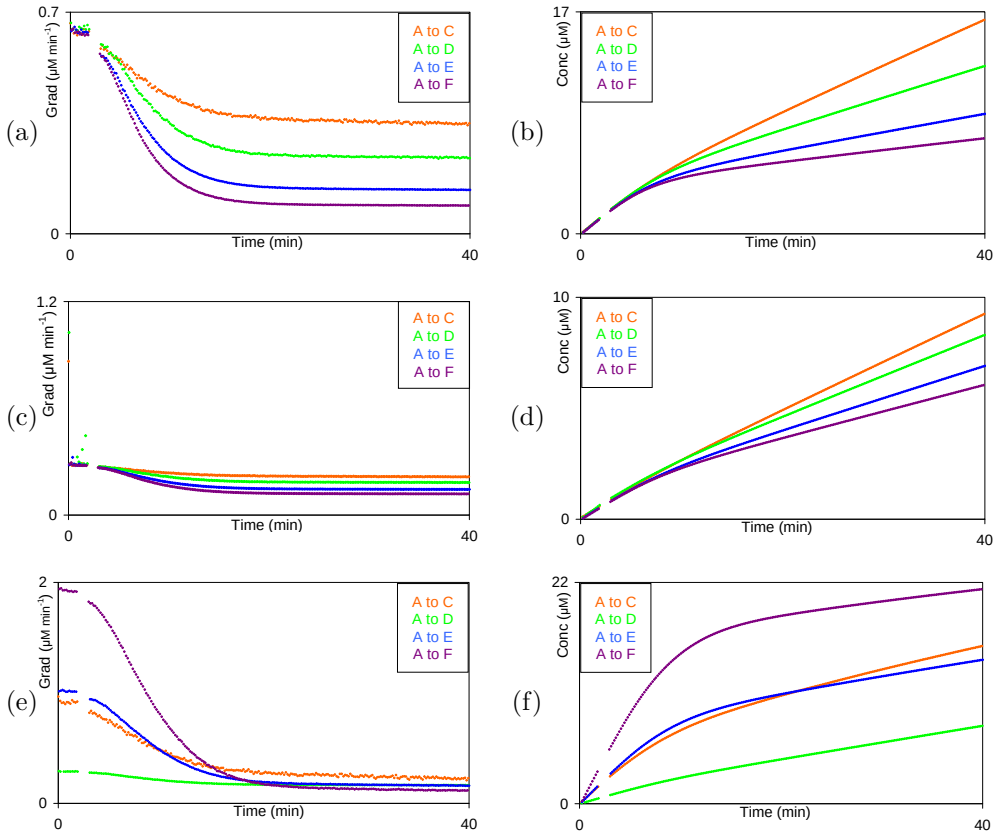


Figure A.2: Predicted background rate and NAD$^+$ production dependent on model used. Graphs (a), (c), and (e) plot the predicted background rate against time for the single exponential, double exponential and linear models respectively. Graphs (b), (d), and (f) plot the predicted NAD$^+$ production for the same models.

B. EXTENDED COMPARISON OF SIMULATED TO EXPERIMENTAL DATA

A total of sixteen different pathways were simulated using four different sets of parameters as described in Section 7.2.3. The simulations arising from three of these sets of parameters did not result in close matches between experimental and simulated data and as such were not covered in detail. The resulting graphs are shown below (Figures B.1, B.2, B.3). The graphs show a significant difference between predicted and experimental NAD^+ production. Figure B.4(a) plots the best RMS error values between predicted and experimental curves, showing the fourth case, used in the subsequent analysis was the best predictor of the experimental data. The NAD^+ production under cases 2 and 3 are compared in Figure B.4(b). The two curves are almost indistinguishable; NAD^+ production for case 3 is slightly slower than that for case 2.

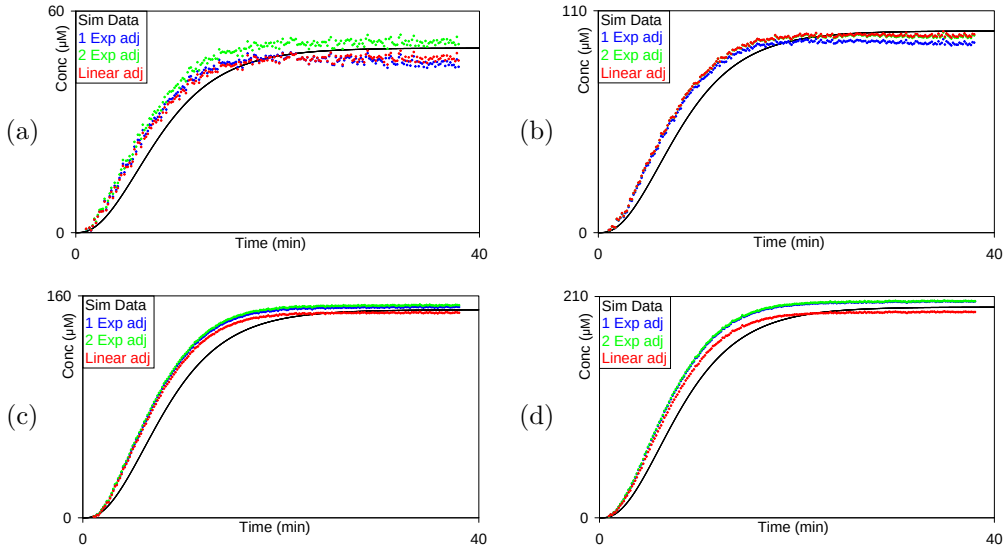


Figure B.1: Comparison of experimental time courses to those simulated using Model 7.2.3.1. Graphs (a)-(d) plot concentration of NAD^+ produced (Conc) against time in minutes. Graphs (a), (b), (c) and (d) correspond to the pathways A to C, A to D, A to E and A to F respectively. The adjusted experimental time courses are plotted in colour according to the keys provided. The simulated time course is plotted in black.

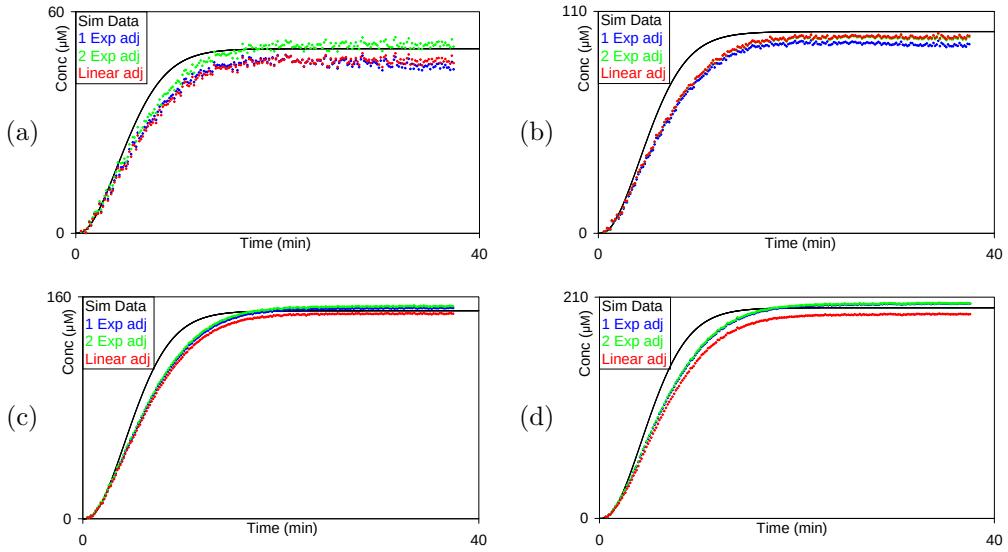


Figure B.2: Comparison of experimental time courses to those simulated using 7.2.3.2. Graphs (a)-(d) are as in Figure B.1.

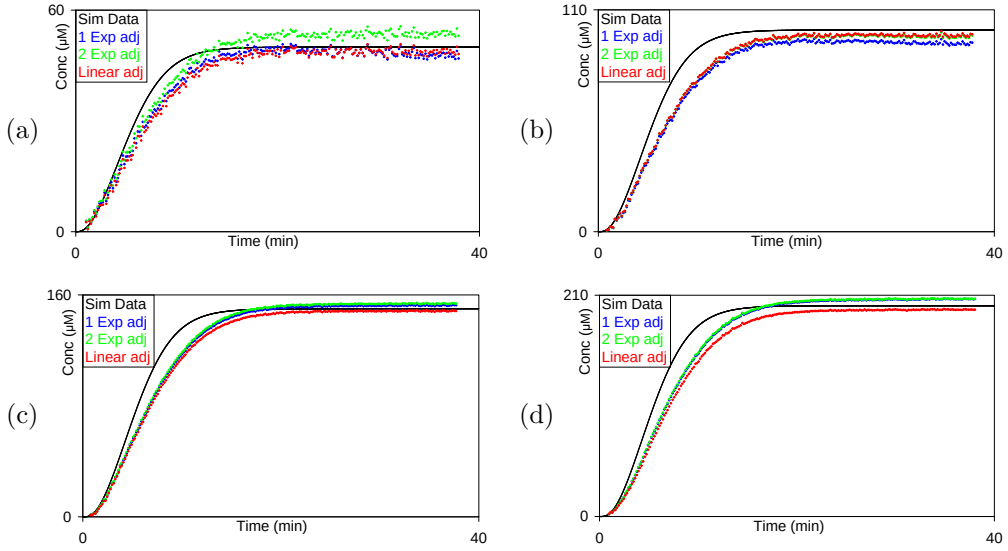


Figure B.3: Comparison of experimental time courses to those simulated using 7.2.3.3. Graphs (a)-(d) are as in Figure B.1.

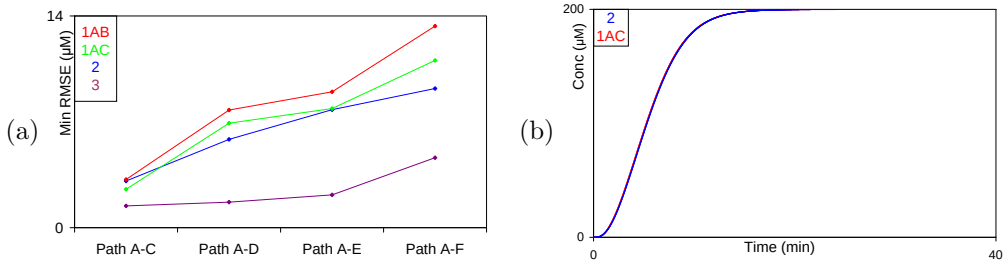


Figure B.4: Comparison of RMS error for the models for NAD^+ production and subpathways used. Graph (a) plots the lowest RMSE for each model and pathway used. Models 7.2.3.1-4 are referred to as 1AB, 1AC, 2 and 3 respectively. Graph (b) plots the simulated concentration of NAD^+ produced by the full pathway for Models 7.2.3.2 and 7.2.3.3 referred to as 1AC and 2 respectively.

C. TABLES OF PARAMETERS FOR THE MODELS USED IN CHAPTER 7

The parameters of each of the models used in Chapter 7 are tabulated below. Models were implemented in *COPASI* as described in Section 3.5. Certain reactions, corresponding to necessary recycling and monitoring reactions, are modelled only by a rate. It is assumed that these reactions are sufficiently fast to ensure immediate turnover of their substrates.

Enzyme	Parameter	Value	Enzyme	Parameter	Value	Reactions	Rate (s^{-1})
MurA	k_{cat} (s^{-1})	8.07	MurB	k_{cat} (s^{-1})	3.17	v_{IDH}	10000
	k_m (PEP) (μM)	4.81		k_m (NADPH) (μM)	20.9	$v_{PK/LDH}$	10000
	k_m (UGP) (μM)	390		k_m (UDPPEE) (μM)	32.8	Species	Init Conc (μM)
MurC	k_{cat} (s^{-1})	7.58	MurD	k_{cat} (s^{-1})	36.5	UGP	50
	k_m (ATP) (μM)	22.4		k_m (ATP) (μM)	38.2	UDPPEE	0
	k_m (L-Ala) (μM)	716		k_m (D-Glu) (μM)	293	UMN	0
	k_m (UMN) (μM)	202		k_m (U1P) (μM)	26.2	U1P	0
MurE	k_{cat} (s^{-1})	73.5	MurF	k_{cat} (s^{-1})	29.2	U2P	0
	k_m (ATP) (μM)	191.2		k_m (ATP) (μM)	12.3	U3P	0
	k_m (L-Lys) (μM)	498.6		k_m (D-Ala-D-Ala) (μM)	31.7	PEP	1230
	k_m (U2P) (μM)	13.4		k_m (U3P) (μM)	3.40	ATP	200
Species	Init Conc (μM)	Species	Init Conc (μM)	Species	Init Conc (μM)	NADPH	100
L-Ala	10000	D-Glu	10000	L-Lys	10000	NADH	1000
D-Ala-D-Ala	10000	MurA-F	0.1				

Table C.1: Table of parameters for Model 7.2.1.1. Kinetic parameters for error bars found in Table 5.1. Subpathways were modelled by eliminating unnecessary enzymes and substrates.

Enzyme	Parameter	Value	Enzyme	Parameter	Value
MurA	k_{cat} (s^{-1})	28.8	MurB	k_{cat} (s^{-1})	4.12
	k_m (PEP) (μM)	4.81		k_m (NADPH) (μM)	20.9
	k_m (UGP) (μM)	390		k_m (UDPPEE) (μM)	32.8
MurC	k_{cat} (s^{-1})	7.58	MurD	k_{cat} (s^{-1})	36.5
	k_m (ATP) (μM)	22.4		k_m (ATP) (μM)	38.2
	k_m (L-Ala) (μM)	716		k_m (D-Glu) (μM)	293
	k_m (UMN) (μM)	202		k_m (U1P) (μM)	26.2
MurE	k_{cat} (s^{-1})	73.5	MurF	k_{cat} (s^{-1})	29.2
	k_m (ATP) (μM)	191.2		k_m (ATP) (μM)	12.3
	k_m (L-Lys) (μM)	498.6		k_m (D-Ala-D-Ala) (μM)	31.7
	k_m (U2P) (μM)	13.4		k_m (U3P) (μM)	3.40

Table C.2: Table of parameters for Model 7.2.3.1. Initial conditions and reactions were as found in Table C.1. Subpathways were modelled by eliminating unnecessary enzymes and substrates.

Enzyme	Parameter	Value	Enzyme	Parameter	Value
MurA	k_{cat} (s^{-1})	28.8	MurB	k_{cat} (s^{-1})	4.12
	k_m (PEP) (μM)	4.81		k_m (NADPH) (μM)	20.9
	k_m (UGP) (μM)	390		k_m (UDPPEE) (μM)	32.8
MurC	k_{cat} (s^{-1})	26.6	MurD	k_{cat} (s^{-1})	36.5
	k_m (ATP) (μM)	22.4		k_m (ATP) (μM)	38.2
	k_m (L-Ala) (μM)	716		k_m (D-Glu) (μM)	293
	k_m (UMN) (μM)	202		k_m (U1P) (μM)	26.2
MurE	k_{cat} (s^{-1})	73.5	MurF	k_{cat} (s^{-1})	29.2
	k_m (ATP) (μM)	191.2		k_m (ATP) (μM)	12.3
	k_m (L-Lys) (μM)	498.6		k_m (D-Ala-D-Ala) (μM)	31.7
	k_m (U2P) (μM)	13.4		k_m (U3P) (μM)	3.40

Table C.3: Table of parameters for Model 7.2.3.2. Initial conditions and reactions were as found in Table C.1. Subpathways were modelled by eliminating unnecessary enzymes and substrates.

Enzyme	Parameter	Value	Enzyme	Parameter	Value
MurA	k_{cat} (s^{-1})	28.8	MurB	k_{cat} (s^{-1})	4.12
	k_m (PEP) (μM)	4.81		k_m (NADPH) (μM)	20.9
	k_m (UGP) (μM)	390		k_m (UDPPEE) (μM)	32.8
MurC	k_{cat} (s^{-1})	26.6	MurD	k_{cat} (s^{-1})	21.5
	k_m (ATP) (μM)	22.4		k_m (ATP) (μM)	38.2
	k_m (L-Ala) (μM)	716		k_m (D-Glu) (μM)	293
	k_m (UMN) (μM)	202		k_m (U1P) (μM)	26.2
MurE	k_{cat} (s^{-1})	63.1	MurF	k_{cat} (s^{-1})	23.9
	k_m (ATP) (μM)	191.2		k_m (ATP) (μM)	12.3
	k_m (L-Lys) (μM)	498.6		k_m (D-Ala-D-Ala) (μM)	31.7
	k_m (U2P) (μM)	13.4		k_m (U3P) (μM)	3.40

Table C.4: Table of parameters for Model 7.2.3.3. Initial conditions and reactions were as found in Table C.1. Subpathways were modelled by eliminating unnecessary enzymes and substrates.

Enzyme	Parameter	Value	Enzyme	Parameter	Value
MurA	k_{cat} (s^{-1})	28.8	MurB	k_{cat} (s^{-1})	4.46
	k_m (PEP) (μM)	4.81		k_m (NADPH) (μM)	41.2
	k_m (UGP) (μM)	390		k_m (UDPPEE) (μM)	68.7

Table C.5: Table of parameters for Model 7.2.3.4. Initial conditions and reactions were as found in Table C.1. Parameters for MurC, MurD, MurE and MurF are unchanged from those found in Table C.4. Subpathways were modelled by eliminating unnecessary enzymes and substrates.

Enzyme	Parameter	Value	Enzyme	Parameter	Value
MurA	k_{cat} (s^{-1})	28.8	MurB	k_{cat} (s^{-1})	4.46
	k_m (PEP) (μM)	4.81		k_m (NADPH) (μM)	41.2
	k_m (UGP) (μM)	390		k_m (UDPPEE) (μM)	68.7
	k_i (UMN) (μM)	19.1			

Table C.6: Table of parameters for Model 7.4.1.1 and 7.4.1.2. Initial conditions and reactions were as found in Table C.1. Parameters for MurC, MurD, MurE and MurF are unchanged from those found in Table C.4. Subpathways were modelled by eliminating unnecessary enzymes and substrates.

Enzyme	Parameter	Value	Enzyme	Parameter	Val 3	Val 4
MurA	k_{cat} (s^{-1})	28.8	MurB	k_{cat} (s^{-1})	4.46	4.12
	k_m (PEP) (μM)	4.81		k_m (NADPH) (μM)	41.2	20.9
	k_m (UGP) (μM)	390		k_m (UDPPEE) (μM)	68.7	32.8
	k_i (UMN) (μM)	19.1				
	k_i (UDPPEE) (μM)	19.1				

Table C.7: Table of parameters for Model 7.4.1.3-4. Initial conditions and reactions were as found in Table C.1. Parameters for MurC, MurD, MurE and MurF are unchanged from those found in Table C.4. Model 7.4.1.3 uses parameters for MurB given in column Val 3. Model 7.4.1.4 uses parameters for MurB given in column Val 4. Subpathways were modelled by eliminating unnecessary enzymes and substrates.

Enzyme	Parameter	Value	Enzyme	Parameter	Value
MurA	k_{cat} (s^{-1})	28.8	MurB	k_{cat} (s^{-1})	4.46
	k_m (PEP) (μM)	4.81		k_m (NADPH) (μM)	41.2
	k_m (UGP) (μM)	390		k_m (UDPPEE) (μM)	62.7
	k_i (UMN) (μM)	19.1			
MurC	k_{cat} (s^{-1})	26.6	MurD	k_{cat} (s^{-1})	21.5
	k_m (ATP) (μM)	22.4		k_m (ATP) (μM)	38.2
	k_m (L-Ala) (μM)	716		k_m (D-Glu) (μM)	293
	k_m (UMN) (μM)	202		k_m (U1P) (μM)	26.2
	k_i (C-1) (μM)	15.3			
MurE	k_{cat} (s^{-1})	63.1	MurF	k_{cat} (s^{-1})	23.9
	k_m (ATP) (μM)	191.2		k_m (ATP) (μM)	12.3
	k_m (L-Lys) (μM)	498.6		k_m (D-Ala-D-Ala) (μM)	31.7
	k_m (U2P) (μM)	13.4		k_m (U3P) (μM)	3.40

Table C.8: Table of parameters for Models 7.4.1.5 and 7.4.1.6. Initial conditions and reactions were as found in Table C.1, C-1 concentration was either 0 or 200 μM. Subpathways were modelled by eliminating unnecessary enzymes and substrates. The IDH reaction was eliminated when simulating measurement of NADPH. Initial conditions were changed as experimental conditions dictated.

Enzyme	Parameter	Value	Enzyme	Parameter	Value
MurA	k_{cat} (s^{-1})	28.8	MurB	k_{cat} (s^{-1})	4.46
	k_m (PEP) (μM)	4.81		k_m (NADPH) (μM)	41.2
	k_m (UGP) (μM)	390		k_m (UDPPEE) (μM)	62.7
	k_i (UMN) (μM)	19.1			
	k_i* (UDPPEE) (μM)	var			

Table C.9: Table of parameters for Models 7.4.1.7 (k_i (UDPPEE) 19.1) and 7.4.1.8 (k_i (UDPPEE) 10). Initial conditions and reactions were as found in Table C.1; concentrations of C-1 were either 0 or 200μM. Parameters for MurC, MurD, MurE and MurF are unchanged from those found in Table C.8. Subpathways were modelled by eliminating unnecessary enzymes and substrates. The IDH reaction was eliminated when simulating measurement of NADPH. Initial conditions were changed as experimental conditions dictated.

Enzyme	Parameter	Value	Enzyme	Parameter	Value
MurA	k_{cat} (s^{-1})	28.8	MurB	k_{cat} (s^{-1})	4.12
	k_m (PEP) (μM)	4.81		k_m (NADPH) (μM)	20.9
	k_m (UGP) (μM)	390		k_m (UDPPEE) (μM)	32.8
	k_i (UMN) (μM)	19.1			
	k_i* (UMN) (μM)	var			

Table C.10: Table of parameters for Models 7.4.1.9 (k_i (UDPPEE) 19.1) and 7.4.1.10 (k_i (UDPPEE) 10). Initial conditions and reactions were as found in Table C.1; concentrations of C-1 were either 0 or 200μM. Parameters for MurC, MurD, MurE and MurF are unchanged from those found in Table C.8. Subpathways were modelled by eliminating unnecessary enzymes and substrates. The IDH reaction was eliminated when simulating measurement of NADPH. Initial conditions were changed as experimental conditions dictated.

Enzyme	Parameter	Value	Enzyme	Parameter	Value	Reactions	Rate (s^{-1})
MurA	k_{cat}* (s^{-1})	28.8 (7.8)	MurB	k_{cat} (s^{-1})	4.46	v_{IDH}	10000
	k_m (PEP) (μM)	4.81		k_m (NADPH) (μM)	41.2	$v_{\mathrm{PK/LDH}}$	10000
	k_m (UGP) (μM)	390		k_m (UDPPEE) (μM)	62.7	$v^{\ddagger}_{\mathrm{UGPrep}}$	0.145 (0.16) (μM)
	k_i† (UMN) (μM)	19.1				Species	Init Conc (μM)
	k_i† (UMN) (μM)	19.1				UGP	100
MurC	k_{cat} (s^{-1})	26.6	MurD	k_{cat} (s^{-1})	21.5	UDPPEE	0
	k_m (ATP) (μM)	22.4		k_m (ATP) (μM)	38.2	UMN	0
	k_m (L-Ala) (μM)	716		k_m (D-Glu) (μM)	293	U1P	0
	k_m (UMN) (μM)	202		k_m (U1P) (μM)	26.2	U2P	0
	k_i (C-1) (μM)	15.3				U3P	0
MurE	k_{cat} (s^{-1})	63.1	MurF	k_{cat} (s^{-1})	23.9	PEP	10000
	k_m (ATP) (μM)	191.2		k_m (ATP) (μM)	12.3	ATP	200
	k_m (L-Lys) (μM)	498.6		k_m (D-Ala-D-Ala) (μM)	31.7	NADPH	100
	k_m (U2P) (μM)	13.4		k_m (U3P) (μM)	3.40	NADH	1000
Species	Init Conc (μM)	Species	Init Conc (μM)	Species	Init Conc (μM)	MurA-F	0.1
L-Ala	10000	D-Glu	10000	L-Lys	10000	D-Ala-D-Ala	10000

Table C.11: Table of parameters for Models 7.4.2.1-4. A range of initial concentration of C-1 were used. For even numbered models value in brackets was used for the k_{cat} of MurA and the k_is marked with a † were eliminated. For Models 7.4.2.3 and 7.4.2.4 the reaction UGPrep was included. It represents simple flux of UGP into the system at the given rate, the value in brackets being used for Model 7.4.2.4.