RADAR

Research Archive and Digital Asset Repository

OXFORD BROOKES UNIVERSITY

When referring to this work, the full bibliographic details must be given as follows:

Hartman, H. (2013) *Genome-scale metabolic modelling of Salmonella and Lactobacillus species*. PhD Thesis. Oxford Brookes University

Genome-scale metabolic modelling of *Salmonella* and *Lactobacillus* species

Hassan Bruno Hartman

A thesis submitted in partial fulfilment of the requirements of
Oxford Brookes University
for the award of the degree Doctor of Philosophy

In collaboration with Animal Health and Veterinary Laboratories
Agency

November 2013

Abstract

Salmonella Typhimurium is a major cause of morbidity and mortality in humans. It is also a commonly used model organism for intracellular Gram negative pathogens, a group of bacteria that is becoming increasingly resistant to available antibiotics. Systemic *Salmonella* infection involves proliferation in the small intestine followed by infection of epithelial and later macrophage host cells. In order to advance the understanding of the rôle of metabolism in virulence, a genome-scale metabolic model of *S.* Typhimurium was constructed, based on genomic and biochemical data obtained from public databases.

A method for modelling metabolic interactions between cells was developed and applied to models of *S.* Typhimurium and the probiotic *Lactobacillus plantarum*, in order to simulate the intestinal stage of infection. The analysis indicated that interactions, involving the transfer of glycolate from *L. plantarum* to *S.* Typhimurium, that favour growth of *S.* Typhimurium, are possible, by unlikely to occur *in vivo*.

Data from Phenotype Microarray (PM), as well as DNA microarray data obtained during infection of cultured macrophage cells, was integrated with the *S.* Typhimurium model. The PM data was largely in agreement with model results for growth on carbon and nitrogen sources, and indicated moderate agreement for sulphur and phosphorus sources. A model-based method for analysis of nutrient availability during growth inside host cells, based on PM and DNA microarray data, was developed. This environment is poorly characterised and direct experimental methods for obtaining this information are not available. The analysis indicated a nutritionally complex host environment, dominated by glycerol 3-phosphate and certain nucleosides and amino acids.

Owing to the complexity of the host environment, a method for identification of a sub-network of the model, required for viability on all growth supporting carbon sources was developed. The impact of sequentially removing combinations of reactions in the sub-network from the genome-scale model was evaluated. This analysis suggested approximately 60 reactions that in various combinations could be of relevance for designing antimicrobial intervention strategies, including antimicrobial agents and live attenuated vaccines.

Acknowledgements

The past years in the Cell Systems Modelling Group have been extremely educational, rewarding, challenging and fun. I am deeply grateful to everyone who has made this possible. In particular, I would like to thank my supervisors Dr Mark Poolman and Professor David Fell for the inspiration and support they have given me through their knowledge and experience, as well as their passion for science. They have both been tremendously helpful and encouraging. I would also like thank my co-supervisor Dr William Newell at the Animal Health and Veterinary Laboratories Agency for encouragement and support during this project.

I would like to thank past and present members of the Cell Systems Modelling Group. As I started my project Dr Achuthanunni Chokkathukalam gave me crucial practical support. I have greatly enjoyed discussing, travelling and collaborating with Amar Ghaisas, Dr Patrick de Vries, Dr Matthew Levin, Aliah Hawari, Maurice Cheung, Kailash Adhikari, Dipali Singh and Cathy Derow.

I am very grateful for the support I have been given by collaborators. I would like to thank Professor John Elmerdahl Olsen for providing valuable experimental support. I am also very grateful for the discussions and support from Professors Martin Woodward and Roberto La Ragione. I sincerely thank Dr Arthur Thompson for giving me access to *Salmonella* DNA microarray data. I thank Professor Peter Ruhdal Jensen, Dr Mogens Kilstrup, and Dr Sebastien Lemire who first introduced me to the world of *Salmonella*. I wish to express my deep gratitude to Dr Sergio Rossell, who first introduced me to the wonderful world of metabolic modelling and who has been a good friend and a great support ever since.

I also thank the helpful staff at Oxford Brookes University. In particular Farida Ben Ghorab, Philip Voysey, Jill Organ, and Catherine Joyejob. I also acknowledge the financial support from Oxford Brookes University and Animal Health and Veterinary Laboratories Agency.

I am also deeply grateful for the support from my mother, Tesse, and my sister, Leila. Finally, I am grateful beyond words for the patience and support from my partner, Johanna. I would not have been where I am today without her. A special thanks to my daughter, Elvira, who makes sure I get up every morning and always greets me with enthusiasm as I returned home.

Contents

URLs

BioCyc	`http://biocyc.org/`
BRENDA	`http://www.brenda-enzymes.org/`
ETE	`http://ete.cgenomics.org/`
GLPK	`http://www.gnu.org/software/glpk/`
Gnuplot	`http://www.gnuplot.info/`
NumPy	`http://www.numpy.org/`
Ply	`http://www.dabeaz.com/ply/`
Python	`http://docs.python.org/`
SciPy	`http://www.scipy.org/`
ScrumPy	`http://mudshark.brookes.ac.uk/ScrumPy`
TECRDB	`http://xpdb.nist.gov/enzyme_thermodynamics/`

Abbreviations

Table 1: Summary of abbreviations and BioCyc identifiers of reactions and metabolites.

Abbreviation	Common name	BioCyc Identifier
Reactions		
2-KGDH	2-oxoglutarate dehydrogenase	2OXOGLUTARATEDEH-RXN
6PGDH-(NAD)	6-phosphogluconate dehydrogenase	RXN-3341
6PGDH-(NADP)	6-phosphogluconate dehydrogenase	RXN-9952
AcAldDH	acetaldehyde dehydrogenase	ACETALD-DEHYDROG-RXN
AcetKin	acetate kinase	ACETATEKIN-RXN
AconDehydr	aconitate hydratase	ACONITATEDEHYDR-RXN
AconHydr	aconitate dehydratase	ACONITATEHYDR-RXN
AlcDH	alcohol dehydrogenase	ALCOHOL-DEHYDROG-RXN
AlaTransAm	β-alanine aminotransferase	2.6.1.18-RXN
Aspartase	aspartate ammonia-lyase	ASPARTASE-RXN
AspDeCarb	aspartate 1-decarboxylase	ASPDECARBOX-RXN
AspKin	aspartate kinas	ASPARTATEKIN-RXN
AspSemAldDH	aspartate semialdehyde dehydrogenase	ASPARTATE-SEMIALDEHYDE-DEHYDROGENASE-RXN
AspTrans	aspartate aminotransferase	ASPAMINOTRANS-RXN
CitSynth	citrate synthase	CITSYN-RXN
CytOx	Cytochrome oxidase	-
D-LacDH	D-lactate dehydrogenase	DLACTDEHYDROGNAD-RXN
Enolase	phosphopyruvate hydratase	2PGADEHYDRAT-RXN
FBPAldolase	fructose-bisphosphate aldolase	F16ALDOLASE-RXN
FumHydr	fumarate hydratase	FUMHYDR-RXN
G6PDH	glucose-6-phosphate 1-dehydrogenase	GLU6PDEHYDROG-RXN
GapDH	glyceraldehyde 3-phosphate dehydrogenase	GAPOXNPHOSPHN-RXN
GCVMult	glycine cleavage system	GCVMULTI-RXN

GluDH	glutamate dehydrogenase	GLUTAMATE-DEHYDROGENASE-NADP+-RXN-(NADP)
GlyoCarLig	glyoxylate carboligase	GLYOCARBOLIG-RXN
HomoSerDH	homoserine dehydrogenase	HOMOSERDEHYDROG-RXN
HomoSerKin	homoserine kinase	HOMOSERKIN-RXN
IsoCitCleav	isocitrate lyase	ISOCIT-CLEAV-RXN
IsoCitDH	isocitrate dehydrogenase	ISOCITDEH-RXN
KDPAldolase	2-dehydro-3-deoxyphosphogluconate (KDP) aldolase	METHYLENETHFDEHYDROG-NADP-RXN
KDPGALDOL-RXN		
MalDH	malate dehydrogenase	MALATE-DEH-RXN
MalSyn	malate synthase	MALSYN-RXN
MetTHFDH	methylenetetrahydrofolate dehydrogenase	
MQ_oxidase	menaquinol oxidase	-
NADH_DH	NADH dehydrogenase	-
NADH_DH_MQ	NADH dehydrogenase (menaquinone)	-
PFK	6-phosphofructokinase	6PFRUCTPHOS-RXN
PGAM	phosphoglycerate mutase	3PGAREARR-RXN
PGIsomerase	glucose-6-phosphate isomerase	PGLUCISOM-RXN
PGKin	phosphoglycerate kinase	PHOSGLYPHOS-RXN
PGLactonase	6-phosphogluconolactonase	6PGLUCONOLACT-RXN
PGlucDehydr	phosphogluconate dehydratase	PGLUCONDEHYDRAT-RXN
PhoAcTrans	phosphate acetyltransferase	PHOSACETYLTRANS-RXN
PyrDH	pyruvate dehydrogenase	PYRUVDEH-RXN
PyrKin	pyruvate kinase	PEPDEPHOS-RXN
Rib5PEpi	ribulose-phosphate 3-epimerase	RIBULP3EPIM-RXN
RiboHyd	ribonucleoside hydrolase	INOSINE-NUCLEOSIDASE-RXN
RiboKin	RIBOKIN-RXN	ribokinase
Rub5PIso	ribose-5-phosphate isomerase	RIB5PISOM-RXN
SCoASynth	succinate-CoA ligase	SUCCCOASYN-RXN
SucDH	succinate-ubiquinone oxidoreductase	SUCCINATE-DEHYDROGENASE-UBIQUINONE-RXN

THX	transhydrogenase	1.6.1.2-RXN
Transald	transaldolase	TRANSALDOL-RXN
TransketI	transketolase	1TRANSKETO-RXN
TransketII	glycoaldehyde transferase	2TRANSKETO-RXN
ThreSyn	threonine synthase	THRESYN-RXN
TriPIsomerase	triosephosphate isomerase	TRIOSEPISOMERIZATION-RXN
UPDGluEpi	UDP-glucose 4-epimerase	UDPGLUCEPIM-RXN

Metabolites

αKG	α-ketoglutarate	2-KETOGLUTARATE
2-PGA	2-phosphoglycerate	2-PG
3-PGA	3-phosphoglycerate	G3P
6PG	6-phospho-D-gluconate	CPD-2961
ADP-GMH	ADP-L-glycero-D-emphmanno-heptose	ADP-L-GLYCERO-D-MANNO-HEPTOSE
ASP	aspartate	L-ASPARTATE
Ac-P	acetylphosphate	ACETYL-P
AcCoA	acetyl-coenzyme A	ACETYL-COA
Acet	acetate	ACET
AcAld	ACETALD	acetaldehyde
CMP-KDO	CMP-3-deoxy-D-*manno*-octulosonate	CMP-KDO
CO_2	carbon-dioxide	-
CisAcon	cis-aconitate	CIS-ACONITATE
Cit	citrate	CIT
DHAP	dihydroxyacetone phosphate	DIHYDROXY-ACETONE-PHOSPHATE
DPG	1,3-diphosphoglycerate	DPG
E4P	D-erythrose-4-phosphate	ERYTHROSE-4P
EtOH	ethanol	ETOH
F6P	fructose-6-phosphate	FRUCTOSE-6P
FBP	fructose-1,6-bisphosphate	FRUCTOSE-16-DIPHOSPHATE
Form	formate	FORMATE
Fum	fumarate	FUM
G6P	glucose-6-phosphate	GLC-6-P
GAP	glyceraldehyde-3-phosphate	GAP
GL6P	D-glucono-Δ-lactone-6-phosphate	D-6-P-GLUCONO-DELTA-LACTONE
GLT	glutamate	GLT
Gly-3P	glycerol 3-phosphate	GLYCEROL-3P

H_i^+	intracellular proton	-
H_p^+	periplasmic proton	-
Hypox	hypoxanthine	HYPOXANTHINE
IsoCit	iso-citrate	THREO-DS-ISO-CITRATE
KDP	2-keto-3-deoxy-6-phospho-gluconate	2-KETO-3-DEOXY-6-P-GLUCONATE
M-DAP	*meso*-diaminopimelate	MESO-DIAMINOPIMELATE
Mal	malate	MAL
Met-THF	5,10-methylenetetrahydrofolate	METHYLENE-THF
NH_3	ammonia	AMMONIA
O_2	molecular oxygen	-
OAA	oxaloacetate	OXALACETIC_ACID
PEP	phosphoenol pyruvate	PHOSPHO-ENOL-PYRUVATE
Pyr	pyruvate	PYRUVATE
R5P	D-ribose-5-phosphate	RIBOSE-5P
Rb5P	D-ribulose-5-phosphate	RIBULOSE-5P
Rib	ribose	RIBOSE
S7P	D-sedoheptulose-7-phosphate	D-SEDOHEPTULOSE-7-P
Suc	succinate	SUC
SucCoA	succinyl-coenzyme A	SUC-COA
UDP-GA	UDP-*N*-acetyl-D-glucosamine	UDP-N-ACETYL-D-GLUCOSAMINE
UDP-Gal	UDP-Galactose	UDP-GALACTOSE
UDP-Glc	UDP-glucose	UDP-GLUCOSE
UnDe-P	undecaprenyl phosphate	UNDECAPRENYL-P
X5P	D-xylulose-5-phosphate	XYLULOSE-5-PHOSPHATE

Chapter 1

Background and general introduction

1.1 Motivation

The introduction of antibiotics as therapeutic against pathogenic bacteria during the late 1930s and 1940s had a major positive impact on human health and welfare. But even before penicillins were clinically available, the first resistance-conferring penicillinase was identified [1, 2]. Introduction of streptomycin in mid-1940s to control *Mycobacterium tuberculosis* was followed by isolation of resistant strains of the pathogen in patients. This pattern was repeated for other antibiotics developed during the 20th century. It was not until the 1950s that the first report describing genetic transfer of antibiotic resistance between bacteria was published, and not until the late 1960s that this notion was widely recognised [3]. Until the end of the 20th century a high rate of development of new classes of antibiotics counteracted the emergence and spread of antibiotic resistance among pathogens [4]. Two worrying tendencies have coincided over the last couple of decades: the development of new antibiotics has decreased and multi-resistant bacteria have become more widespread [4, 5].

The development of new classes of antibiotics peaked in the 1940s to 1960s and no new class has been registered since 1987. Much of the early development of new antibiotics was based on empirical screening, i.e screening of growth inhibitory effects of natural products (primarily from antibiotic producing microorganisms) on pathogens. This method became less cost-effective after most of the interesting compounds produced by well characterised antibiotic producing microorganisms had been exploited [4]. Another reason why many large pharmaceutical companies have shifted focus away from development of antimi-

crobials is that this group of drugs are of limited commercial interest. Unlike diabetes or hypertension, which often require life long medication, treatment of bacterial infections with antimicrobials is usually short term, which means that the revenues from sales of antimicrobials are more limited. Many of the new antimicrobials are only approved for specific patient groups or situations, which further limits the return on investment. If, despite these obstacles, a company is successful in selling an antimicrobial, this only accelerates the emergence and spread of resistance [5]. Following the thalidomide disaster in the 1960s, many countries introduced stricter controls for approval of new drugs, which, apart from increased drug safety, also increased the cost of development of new drugs, including antimicrobials [2].

The origin and dynamics of antibiotic resistance in bacteria is a complex topic that will not be covered in detail here. In general, many different mechanisms for overcoming the effect of an antimicrobial are possible: gene duplications resulting in elevated concentration of cellular targets; broad spectrum efflux systems that export xenobiotics (including antimicrobials); specific enzymes that inactivate the antimicrobial, and mutations resulting in structural modifications of cellular targets. Many of the more specific mechanisms directed towards natural antibiotics are believed to have originated in the producing organisms [2]. Human activities have had a significant impact on the evolution and spread of antibiotic resistance. Large quantities of antibiotics have been produced industrially and eventually released to natural environments over the last 70 years, creating strong selective pressure for resistance in natural microbial populations [2].

Besides the direct release from industrial antibiotic production to natural environments, many uses of antimicrobials could favour resistant microbial populations. Example of these activities include: use for growth promotion in animal agriculture; use as prophylactic/therapeutic in humans; use as prophylactic/therapeutic in pets; use for pest control in agriculture; use as biocide in cleaning products, and use for genetic marker selection and sterile culture in biological research [2].

Outbreaks of multi-resistant pathogens, that are very difficult to control using existing drugs, are becoming less restricted by geographic boundaries owing to the increasing mobility of people and goods [6].

The clinical relevance of antibiotic resistance vary between different groups of bacteria. Gram positive pathogens, primarily staphylococci (including MRSA), streptococci, and enterococci, are becoming more manageable, owing to susceptibility to novel (Gram positive-specific) antimicrobials, including linezolid, quinupristin in combination with dalfopristin, and daptomycin, which are all derivatives of older classes of antimicrobials. Gram negative bacteria, especially

2

members of the *Enterobacteriaceae* family, are cause for more concern. The outer cell membrane of Gram negative bacteria is much less permeable to drug molecules than the cell envelope of Gram positive bacteria. This, in combination with prevalence of unspecific efflux systems, makes drug development against Gram negative bacteria highly challenging [2, 4]. Data from the UK show a rise in ESBL (extended spectrum β-lactamase)-producing *E. coli* and *Klebsiella* isolated from patients. Most of these isolates are resistant to one or more commonly used antibiotics. Resistance to carbapenems, which are often a last resort drug for multi-resistant Gram negative pathogens, is currently limited in the UK ($<2\%$ of isolated *Enterobacteriaceae*), but is becoming a more widespread problem in other parts of the world [7].

1.2   Aim and structure

It should be evident from the introduction above that the slow rate of development of new antimicrobials by pharmaceutical industry, in combination with widespread resistance, especially among Gram negative bacteria, to currently available ones, is highly problematic. The work presented here is intended to address some aspects of this problem.

One of the main aims of this project is to identify enzymatic sites of vulnerability in the metabolic network of *S.* Typhimurium that could aid development of antimicrobials or live attenuated vaccines. *S.* Typhimurium is a Gram negative pathogen of clinical importance (which is described in more detail in the next section), but the main motivation for focusing on this organism is that it is also widely used as a model organism for Gram negative intracellular pathogens. The number of antimicrobials that inhibit metabolic enzymes is very limited [8], thus metabolism is an underexploited source of targets for antimicrobial development. A reason for this, in the case of *S.* Typhimurium, is that the number of single enzymes in the network that are suitable drug targets is very limited, owing to a high degree of network redundancy [8], which calls for simultaneous inhibition of groups of two or more enzymes. This makes purely experimental procedures for target identification through selective gene deletion difficult, as it would involve construction of hundreds of thousands of mutant strain for a typical Gram negative pathogen. It will be argued in later chapters that the experimental efforts can be reduced significantly by simulating these deletions in a computer model of metabolism, and focusing the experimental effort to those enzymes that appear to be of interest based on the model analysis.

Metabolic enzymes have many properties that make them suitable drug targets: they perform a physiologically essential function; they are fairly well conserved among Gram negative pathogens; and their activity can be inhibited by

small molecules [9]. It is important however to note that the essentiality of a given metabolic enzyme could be limited to certain situations, e.g. when the organism is using a particular carbon source. This means that the suitability of a given enzyme as a drug target could be related to the nutrient availability. In the case of *S.* Typhimurium during infection this is not known in sufficient detail. Another aim of this project is thus to use metabolic modelling in order to gain insights into the metabolic behaviour of *S.* Typhimurium during infection. This is done using two separate model-based approaches: Potential metabolic interactions with the probiotic bacterium *L. plantarum* is explored; and system-wide experimental data obtained from *S.* Typhimurium during infection is analysed.

This thesis has the following organisation:

- The remainder of Chapter 1 gives a brief introduction to the biology *S.* Typhimurium.

- Chapter 2 describes the basis of much of the mathematical and computational methodology used in subsequent chapters.

- Chapter 3 describes the construction and analysis of a *S.* Typhimurium metabolic model.

- In Chapter 4 the analysis of a previously published model of *L. plantarum* is presented.

- Potential metabolic interactions between the models described in the previous two chapters is analysed in Chapter 5.

- Integration of experimental data relevant to infection of *S.* Typhimurium is described in Chapter 6.

- Chapter 7 is devoted to model-based identification of potentially vulnerable site in the metabolic network of *S.* Typhimurium.

- Chapter 8 is a general discussion of the results presented here.

1.3   Overview of the biology of *Salmonella*

1.3.1   Nomenclature and taxonomy

The *Salmonella* genus belongs to the gammaproteobacteria class and the *Enterobacteriacea* family. Salmonellae are phenotypically characterised as rod-shaped, Gram-negative, flagellated, and facultative anaerobes [10]. Classification below the genus level has long been subject to debate [11]. Currently, the genus is believed to include two species: *Salmonella enterica* and *Salmonella*

bongori, of which only *S. enterica* cause disease in humans. *S. enterica* is divided into six subspecies: *enterica*, *salamae*, *arizonae*, *diarizonae*, *houtenae*, and *indica*. Close to all (>99%) disease in humans is caused by the *enterica* subspecies [12]. Within *Salmonella enterica* subsp. *enterica* there are > 2100 serovars, phenotypically distinguished based on surface antigens, host specificity, and sensitivity to phages [11]. These serovars are usually named in relation to the symptoms they cause, their main host or the location of their discovery, e.g. *S.* Typhimurium causes symptoms in mice that resembles those caused by *S.* Typhi in humans, *S.* Colerasuis cause cholera-like symptoms in swine, and the first isolate of *S.* Dublin was made in that city. Despite the phenotypical diversity, the *Salmonella* serovars display a high degree of genetic similarity (96-99%) [13]. Host specificity of serovars is usually classified as host restricted, host adapted or unrestricted. Host restricted serovars are strongly associated with a particular host species and cause severe, systemic infection in that species, examples include Typhi, Paratyphi, and Sendai (human restricted), Gallinarum (fowl restricted), Abortusovis (sheep restricted), Typhisuis (swine restricted), Abortusequi (horse restricted). Host adapted serovars commonly cause systemic infection in a host species, but can occasionally cause less severe symptoms in other species, e.g. Cholerasuis, whose natural host is swine but which can cause disease in humans, and Dublin, whose natural host is cattle but which can also infect humans and sheep. The unrestricted serovars include Typhimurium and Enteritidis, which can cause systemic infection in some species, but cause self-limiting gastroenteritis in most mammals [14].

1.3.2   Epidemiology of *Salmonella*

The *Salmonella* serovars that cause disease in humans can be divided into typhoidal and non-typhoidal serovars, where the former category includes the main human restricted serovars Typhi and Paratyphi and the latter includes most serovars, but most frequently Typhimurium and Enteritidis. The typhoidal serovars cause typhoid fever, which is characterised by high fever, sepsis, and gastrointestinal bleeding [10]. Untreated typhoid fever has a fatality risk (case fatality) of 20% [15], this risk falls to 1% with treatment [16]. Non-typhoidal serovars primarily cause self-limiting diarrhoea, but cases where non-typhoidal serovars cause invasive, bloodstream infection (bacteremia) are an increasing problem [10]. Invasive non-typhoidal infection can involve a range of symptoms, typically (95% of cases) high fever and anaemia (40-50% of cases). The burden of different *Salmonella* serovars shows considerable regional variation. Invasive non-typhoidal salmonellae is a major concern in sub-Saharan Africa: Although epidemiological data for this region is uncertain, the annual incidence in chil-

dren aged 3-5 years has been estimated to 175-388 per 100000 and in adult HIV patients to 2000-7500 per 100000 [10]. The fatality risk for infection of invasive non-typhoidal salmonellae in this region is 22-47%, even when antimicrobials are applied [10]. Incidence of invasive non-typhoidal salmonellae is much less prevalent in Asia compared to sub-Saharan Africa: The highest rate, 7.2 cases per 100000 individuals, was reported in Pakistan for children aged 2-15 years [17]. Instead, typhoidal serovars are the dominating invasive salmonellae with incidence for 5-15 year-olds at ∼500 cases per 100000 in Pakistan (Karachi) and India (Kolkata) [18]. It has been suggested that the difference in epidemiology between these two regions is related to the high prevalence of malaria and HIV in sub-Saharan Africa [10]. In high income countries (Europe, Australia, New Zealand, and North America) the incidence of typhoidal infection is low, 10 cases per 100000. The incidence of non-typohoidal infection is considerably higher in these regions, 690 per 100000 in Europe [19]. Mortality from non-typhoidal infection in these regions is limited to 1-5% [10].

1.3.3   *Salmonella* infection

A common infection route for salmonellae is consumption of contaminated food or water, although pet animals, such as reptiles and amphibians, are also a source of infection [20]. Salmonellae are able to survive the low pH of the stomach and the hostile environment of the small intestine [21]. In the small intestine they can attach to epithelial cells and induce engulfment, thereby crossing the epithelial barrier [22, 23]. A preferred route is through microfold cells, which are epithelial cells that routinely sample the intestinal fluid for antigens by unspecific engulfment of extracellular material. Microfold cells then transport the vesicle to lymphoid cells situated under the epithelial cells [20]. Salmonellae can also be engulfed by a particular type of phagocyte (CD18 expressing) in the intestinal mucosa and transported across the epithelium [24, 25]. It has been suggested that the phagocyte route allows typhoidal serovars to infect the host without causing inflammation and diarrhoea [20]. After crossing the epithelium typhoidal serovars will induce engulfment by macrophage cells and subsequently spread to lymph nodes and the spleen, whereas non-typhoidal serovars will indirectly cause local inflammation by the action of white blood cells (polymorphonuclear leukocytes) [20]. The engulfment of typhoidal salmonellae by macrophage cells creates an intracellular vacuole, referred to as the Salmonella Containing Vacuole (SCV). Systemic infection can only proceed if the pathogen can replicate in the SCV [26]. The detailed nutrient availability in the SCV is unknown, but mutant strains that are defective in biosynthesis in certain nucleotides and amino acids are unable to replicate and cause infection, suggesting

that these nutrients are not provided by the host [27].

1.3.4 Human intervention strategies

The primary public health strategy for *Salmonella* control is to establish and maintain good general sanitation and hygiene (including safe practises for food and water handling), good nutrition, and vaccination programs [10, 19]. Medical treatment of typhoidal fever is becoming increasingly difficult, especially in developing countries, owing to the ubiquity of antibiotic resistance in salmonellae [19]. Drug resistance in typhoidal serovars varies between different locations, e.g. Ochiai et al. [18] reported that two-thirds of the typhoidal strains identified in Karachi, Pakistan were multi-resistant, which was the highest rate in the study. Chloramphenicol was the primary therapeutic for typhoid fever until the 1970s when resistance to this drug appeared. Fluoroquinolones, such as ciprofloxacin, largely replaced chloramphenicol, but fluoroquinolone resistance is an increasing problem, especially in south Asia [19, 28]. Non-typhoidal infections are usually self-limiting and left untreated, except for particularly vulnerable groups, such as noenatals, older adults, and immunosuppressed patients. Because of widespread drug resistance, non-typhoidal salmonellae are primarily treated with fluoroquinolones [19]. However, resistance to fluoroquinolones in non-typhoidal strains is reported to be on the rise [29]. More concerning is perhaps the isolation of carbapenem resistant *S.* Typhimurium strains, as this class is a last resort used clinically to treat *Salmonella* infection.

There are currently three vaccines against typhoid fever available internationally [19, 30]: (*i*) Heat killed whole cells of *S.* Typhi strain Ty2 administered by injection, with an efficacy of 50-94%, but also high rates of side effects. This vaccine has largely been replaced by alternative vaccines. (*ii*) Orally administered live attenuated *S.* Typhi strain Ty21a (derived from strain Ty2 by chemical mutagenesis [31]), has an efficacy of 50-80%, but very few side effect, and protection lasts for five years or more. Disadvantages of this vaccine include the need for multiple doses, dependency on cold chain, and possibly adverse effects in immunocompromised patients [19]; (*iii*) The Vi polysaccharide vaccine is administered by injection, has an efficacy of 60-80% and can by used in immunocompromised patients. Unlike the attenuated Ty21a vaccine, protection is limited to two years, and only towards strains that express the Vi antigen, which excludes *S.* Paratyphi and certain *S.* Typhi strains [19].

There are currently no vaccines for non-typhoidal infection approved for human use, but as multidrug resistance and invasive non-typhoidal infection are becoming more problematic research in this area is encouraged [10, 19].

It is for these reasons that this thesis investigates potential metabolic targets

that cause lethality or attenuation *S.* Typhimurium.

Chapter 2

General methodology

2.1 Introduction

As the first draft sequence of the human genome was announced in 2000, many individuals involved with the Human Genome Project had very high expectations as to the outcome of the project. Some of the more optimistic predictions stated that within a decade, i.e. by 2010, 'personalised medicine' would come of age. This would involve personalised genetic tests showing an individual's risk for common diseases, followed by personalised medical treatment. There was also an expectation that the sequencing project would reveal the 'key genes' for most of the major human diseases [32]. The promises of the early whole-genome sequencing projects have largely gone unmet. There cannot be any doubt that the availability of genome sequences has revolutionised biological sciences, but this data alone has so far failed to give a complete picture of the organism of interest. Moreover, the discrepancy between the expectations of the sequencing projects and the current state of biological sciences has motivated increased research efforts towards analysis of emergent systemic properties of biological components, rather than mere identification of components. Although biological reductionism has been hugely successful in identifying many of the components that make up cells and more complex biological systems, it does not offer a satisfactory explanation of how these components interact to form living cells, or how manipulation of these components could change cell behaviour in desirable directions [33]. Owing to this, a more holistic scientific paradigm, referred to as *systems biology*, has received increased interest over the last decade. Systems biology does not have a single coherent definition, but most people working in the field would consider a core feature to be construction and analysis of biological models, i.e. formal descriptions of biological systems of arbitrary complexity, which form a basis for integration of quantitative experimental data. This

methodology is highly interdisciplinary and borrows heavily from biology, chemistry, physics, computer science and mathematics, but with the ultimate aim of explaining biological phenomenon in terms of molecular interactions. One of the more mature sub-disciplines of systems biology is its application to metabolism. This stems primarily from the maturity of metabolic biochemistry: Many of the enzymes in central carbon metabolism (at least for some well known organisms) and the stoichiometry of the reactions they catalyse are known. Another factor is that metabolism is a network of material transformations, which means that it is subject to mass balance constraints. How this can be used in mathematical analysis of networks is described in more detail in the next section.

Whilst systems biology is a relatively new discipline (systems-wide data, such as genome sequences, having only been available since the late 1990s), the same underlying philosophy has been applied to smaller biological systems in the past. A prominent example of this is Metabolic Control Analysis (MCA). Since metabolic biochemistry emerged as a separate science it inherited from chemical kinetics the valid concept of *rate limiting steps*. In enzyme-catalysed reactions where more than two substrates are involved it is not uncommon that two of the substrates form an intermediate that reacts with the third substrate. For example, the net reaction

$$A \; + \; B \; + \; C \; \rightarrow \; P$$

can be decomposed to:

$$A \; + \; B \; \rightarrow \; X$$

$$X \; + \; C \; \rightarrow \; P,$$

where X is a reaction intermediate. If one of the intermediate steps is much slower than the other one, the overall rate of the net reaction will equal the rate of the slowest step, which is referred to the rate limiting step [34]. It is however not uncommon in biochemical literature to find references to enzymes as rate limiting for the overall flux through a multi-enzyme reactions systems, or pathway. The implication of certain enzymes being rate limiting is that increasing the concentration of these enzymes should increase the net flux through the pathway. Phosphofructokinase in glycolysis is an enzyme that has been considered rate limiting for the glycolytic flux [35, 36]. Construction of mutant strains with elevated phosphofructokinase concentration in baker's yeast did not, however, have any effect on glycolytic flux. There are other similar examples of how the concept of rate limiting enzymes in metabolism is problematic [37]. In order to overcome these discrepancies MCA was developed during the mid-1970s as a framework for analysis of how flux through metabolic networks is controlled.

In contrast to traditional biochemistry, MCA relies on formal descriptions of the system of interest. Most notable in this context is the introduction of the *Flux Control Coefficient (FCC)*, defined as the scaled sensitivity of an enzyme concentration on the flux through the network, i.e. how much the concentration of a given enzyme in a pathway affects the net flux through the pathway. For example, if the flux through a pathway changes linearly with the concentration of one single enzyme, and no other enzyme, in the pathway, then the enzyme has an FCC of 1, and all other enzymes an FCC of 0. As opposed to the traditional sorting of enzymes into rate or non-rate limiting, MCA explains how flux control is distributed over the enzymes in a pathway. Experimental and theoretical work has shown that in a majority of cases flux control is indeed distributed over the enzymes in a pathway, i.e. very few enzymes have FCCs close to 1, and that simple one-enzyme overexpressions are unlikely to affect the net flux through a pathway [37]. This brief description of MCA is in no way intended to be complete but only serves as an example of how formal descriptions of biological systems can be a great aid for understanding them.

2.2 Mathematical modelling of metabolism

Mathematical modelling of metabolic networks, as opposed to other cellular networks, is a relatively established field. Some of the reasons for this are outlined briefly:

- *Experimental:* The basic components of metabolic networks are biochemical reactions that transform metabolites, most of these reactions are catalysed by enzymes. Many metabolic enzymes are well characterised in terms of substrate specificity and, to a lesser extent, quantitative behaviour with varying concentrations of substrates, products and other metabolites. This accumulated data is a prerequisite for metabolic modelling.

- *Theoretical:* Quantitative experimental data relating the rate of an enzyme-catalysed reaction to concentrations of metabolites can be used to formulate a deterministic rate function. If the rate functions for a set of enzymes in a given network is known, as well as the concentrations of the metabolites assumed to affect the rate, this network can be modelled as a set of ordinary differential equations. Even if only the association between reactions and metabolites is known, but not the rate functions and concentrations of metabolites, the rate of change of metabolite concentrations are still constrained by mass conservation. This means that the change in metabolite concentration over time is the sum of the rates of all reactions involved with the metabolite. It follows that variables in

the system (reaction rates and metabolite concentrations) are more restricted than they would otherwise be and less information is needed to predict the behaviour of the system. Information transduction networks also involve transfer of material, e.g. signalling pathways often include phosphorylation and dephosphorylation steps, but the flow of information is not directly proportional to the flow of material.

- *Application:* There are strong incentives to develop predictable models of metabolic networks that are of technological (e.g. glutamate and lysine producing bacterium *Corynebacterium glutamicum* [38]) or medical (e.g. human [39, 40]) importance. These models can be used to devise rational intervention strategies for modifying the behaviour of the network.

2.2.1 Foundations and terminology

Information concerning the association between reactions and metabolites, specifically the *stoichiometry* or the molar proportions of substrates and product forms the basis of any metabolic model. A reaction that converts 2 moles of A into 1 mole of B can be expressed as:

$$2\,A \rightarrow\ B,$$
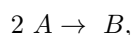
or, alternatively:

$$B\ -\ 2\,A\ =\ 0.$$

A convenient way of expressing the structure of a metabolic network is to compile the stoichiometries of all reactions in the network into a *stoichiometry matrix*, commonly abbreviated $\mathbf{N}$, where columns correspond to reactions, rows metabolites, and matrix elements are stoichiometric coefficients. This is illustrated in Fig. 2.1 for a small hypothetical network. In general all metabolites in the model are represented as rows in the matrix, but for many applications it is convenient to only include the *internal* metabolites, and exclude the *external* metabolites. Metabolites are considered external if their concentration is constant [41], if this is not the case they are internal. Examples of external metabolites include those whose cellular concentration is high enough to consider any change negligible (e.g. water), or metabolites that are maintained at a constant level outside the cell (and that can also be transported over the cell boundary, which is illustrated in Fig. 2.1). As will be illustrated in Section 2.3, biologically useful information can be extracted from a relatively information-poor stoichiometric model.

Since the internal metabolite concentrations are non-constant, they are time-dependent variables, determined by their initial values and the rates of consump-
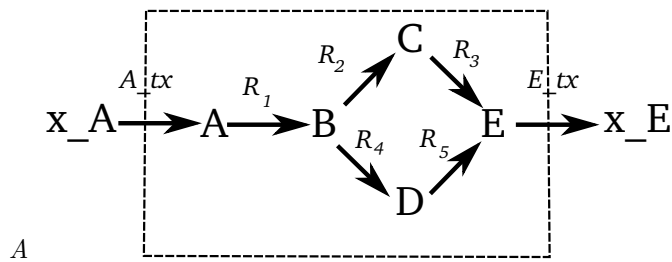
$$N = \begin{array}{c} \\ \\ \\ \\ \\ \\ \end{array} \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & -1 & 0 \end{bmatrix}$$

Figure 2.1: Metabolic network structure expressed with a stoichiometry matrix. The network in A contains the internal metabolites A - E and external metabolites x_A and x_E. The stoichiometry matrix (N) of the network is constructed as indicated in B. For each reaction (column in N) the stoichiometric coefficients appear as matrix elements at rows corresponding to metabolites involved with the reaction (elements of other metabolites are filled with zeroes). The external metabolites are not included in N.

tion and production by the reactions in the system (assuming abiotic processes that affect metabolite concentrations, such as convection or diffusion, are negligible). Mathematically, the rate of change of the internal metabolite pools can be represented as a set of balance equations, only determined by the stoichiometry of the network and the reaction rates. For the network shown in Fig. 2.1, the balance equations are:

$$
\begin{aligned}
\frac{dA}{dt} &= v_{A\_tx} - v_{R_1} \\
\frac{dB}{dt} &= v_{R_1} - v_{R_2} \\
\frac{dC}{dt} &= v_{R_2} - v_{R_3} \\
\frac{dD}{dt} &= v_{R_4} - v_{R_5} \\
\frac{dE}{dt} &= v_{R_3} + v_{R_5} - v_{E\_tx},
\end{aligned}
$$

which can be expressed using matrix notation as:

$$
\frac{d\mathbf{S}}{dt} = \mathbf{N} \cdot \mathbf{v}, \tag{2.1}
$$

where $\mathbf{S}$ is the vector of internal metabolite concentrations, $\mathbf{N}$ the stoichiometry matrix, and $\mathbf{v}$ the vector of reaction rates.

The values of vector $\mathbf{v}$ are primarily determined by enzyme kinetics (also referred to as rate laws or rate functions) and metabolite concentrations [34]. A numerically stable solution to Equation 2.1 is referred to as a *steady state solution*. This implies that the systems is in a state at which the variables, metabolite concentrations and reaction rates (which are called fluxes in a steady state solution), do not change with time. The time-invariance of the system at steady state is usually interpreted approximately, e.g. an oscillatory system with constant amplitude is also considered to be in steady state because there is no net accumulation or depletion of metabolites [42]. Fortunately, many biological phenomenon occur under stationary (or approximately stationary) conditions, well known examples from metabolic biochemistry include glycolysis and amino acid biosynthesis [43]. Given that a metabolic network of interest eventually reaches a steady state, Equation 2.1 simplifies to:

$$
\frac{d\mathbf{S}}{dt} = \mathbf{N} \cdot \mathbf{v} = \mathbf{0}, \tag{2.2}
$$

This is a very powerful constraint on the flux vector, $\mathbf{v}$. Although a unique solution is not obtained by the steady state constraint alone, Equation 2.2 implicitly defines all possible steady state solutions. Consequently, given the steady state assumption, $\mathbf{v}$ can be treated as a variable, and solved for independently of the kinetics. This is the basis of structural model analysis.

2.3 Structural modelling

Structural (or stoichiometric) models, where only the stoichiometry of the network, but not any kinetic parameters, are known, are more fundamental than kinetic ones, i.e. all kinetic models are also structural, but the converse is not true.

2.3.1 Null-space analysis

Much of structural analysis is centred around the *null-space* of the stoichiometry matrix. The flux vector $\mathbf{v}$ in Equation 2.2 is usually underdetermined and (infinitely) many values of $\mathbf{v}$ could satisfy the equation. This space of possible solutions is the null-space of the stoichiometry matrix. Although the null-space contains an infinite number of solutions (points in flux space), this space can be concisely defined as the linear combinations of columns of a *kernel matrix* ($\mathbf{K}$), which, unsurprisingly, is defined as:

$$\mathbf{N} \cdot \mathbf{K} = \mathbf{0}, \tag{2.3}$$

Since the null-space is defined by the columns of the kernel (the kernel matrix spans, or forms the basis of, the null-space) it follows that any steady state flux vector is a linear combination of kernel column vectors [42]:

$$\mathbf{v} = \mathbf{K} \cdot \lambda$$

where λ is a vector of scaling factors. For example, consider the network in Fig. 2.1, a kernel matrix for the network is given by:

$$\mathbf{K} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 1 & -1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}$$

where each row corresponds to a column in $\mathbf{N}$ (here, the row order is the same as the column order in $\mathbf{N}$).

If $\lambda = [\lambda_1 \ \lambda_2]^T$,

$$\mathbf{v} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 1 & -1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_1 - \lambda_2 \\ \lambda_1 - \lambda_2 \\ \lambda_2 \\ \lambda_2 \\ \lambda_1 \\ \lambda_1 \end{bmatrix} \tag{2.4}$$

From this representation of $\mathbf{K}$ it should be clear that if two linearly independent fluxes are observed all steady state fluxes in the system can be determined. For instance, if E_tx (or A_tx or R_1) and R_4 are measured and assigned fluxes J_1 and J_2, respectively, it can be concluded that $\lambda_1 = J_1$ and $\lambda_2 = J_2$, and further:

$$\mathbf{v} = \begin{bmatrix} J_1 \\ J_1 - J_2 \\ J_1 - J_2 \\ J_2 \\ J_2 \\ J_1 \\ J_1 \end{bmatrix}$$

It should be noted that the kernel of any given matrix is not unique, since different algorithms used for its computation cannot be guaranteed to return the exact same matrix, but as will be shown, invariant properties may still be obtained from the kernel. The simplest method for calculating the kernel of a matrix is Gauss-Jordan elimination [44].

Enzyme subsets

Since all possible steady state flux solutions are linear combination of kernel column vectors, kernel row vectors contain information about reactions that must apply for all possible solutions. For example, if a row vector contains only elements equal to zero the corresponding reaction cannot occur in any solution and is termed *dead* or strictly detailed balanced [45]. Further, sets of reactions that have identical or proportional row vectors must carry proportional (steady state) flux. These sets are called *enzyme subsets* (or reaction subsets) [46]. For the network in Fig. 2.1, inspection of the kernel reveals that R_1, A_{tx}, and E_{tx} all share the row vector $[1\ 0]$; R_2 and R_2 share $[1\ -1]$; and R_4 and R_5 share $[0\ 1]$. As indicated from network diagram this involves both sets of reactions that for linear pathways (R_2 and R_2, and R_4 and R_5), but also the less intuitive

16

set involving R_1, A_{tx}, and E_{tx}. In the algorithm suggested by Pfeiffer et al. [46], enzyme subsets are identified by (i) removing all dead reactions from $\mathbf{K}$; (ii) dividing all row vectors with the greatest common divisor; and (iii) comparing each row with each other. Two or more reactions belong to the same subset if their corresponding row vectors are identical and the sets do not conflict with reaction reversibilities. An example of a reversibility conflict could be if R_2 appeared with directionality $B \leftarrow C$ (i.e. the opposite direction to that indicated in Fig. 2.1), resulting in the subset: $B \leftarrow C \rightarrow D$. This would not change the kernel matrix and inspection of the row vectors of R_2 and R_3 would still suggest that they are in a subset. Sets like these are referred to as *inconsistent subsets*.

The procedure used within the framework of this thesis, which is implemented in the metabolic software used [47], differs slightly from that suggested by Pfeiffer et al. [46]. In step (ii) a *flux ratio*, $d_{i,j}$ for each row vector combination is defined as:

$$d_{i,j} = \begin{cases} \frac{\mathbf{K}_{i,a}}{\mathbf{K}_{j,a}} & \text{if } \frac{\mathbf{K}_{i,a}}{\mathbf{K}_{j,a}} = \frac{\mathbf{K}_{i,b}}{\mathbf{K}_{j,b}} \ \wedge \ (\mathbf{K}_{i,c} = \mathbf{K}_{j,c} = 0) \\ \\ 0 & \text{if } \frac{\mathbf{K}_{i,a}}{\mathbf{K}_{j,a}} \neq \frac{\mathbf{K}_{i,b}}{\mathbf{K}_{j,b}} \ \vee \ (\mathbf{K}_{i,c} \neq \mathbf{K}_{j,c}) \end{cases} \tag{2.5}$$

where $i, j \in \{1 \ldots m\}$, $\exists a \in \{1 \ldots r\}$, $\exists b \in \{1 \ldots r, \ b \neq a\}$, and $\forall c \in \{1 \ldots r, \ c \neq a, b\}$ (and $\mathbf{K}$ is the kernel matrix with m rows). In other words, each combination of row vectors from the kernel matrix are compared by component-wise division. If, for any two row vectors, all ratios are identical or zero, the two reactions are in the same subset and the flux ratio is equal to the constant ratio of the kernel elements. If this ratio is not constant or at least one of the elements in any of the vectors is zero and the corresponding element in the other vector is not, the reactions are not in the same subset.

Correlation analysis

Since the kernel matrix defines the solution space of Equation 2.2, the rows of the kernel can be seen as vectors in this space. If two reactions are associated with parallel row vectors in the kernel they are in the same enzyme subset. In fact, Equation 2.5 can be used for determining whether two vectors (of equal lenght) are parallel, which is true if the ratio $d_{i,j}$ is non-zero. A logical extension of the enzyme subset concept is then to asses the relation between any row vector in the kernel. This was explored by Poolman et al. [48], where the angle between row vectors was used for assessing the correlation between reactions in the null-space. Specifically, the cosine of the angle between a pair of row vectors in the *orthogonal* null-space was used, and referred as the *reaction correlation*

coefficient (ϕ_{ij} or RCC):

$$\phi_{ij} \;=\; \frac{\mathbf{K}_i\mathbf{K}_j^T}{\sqrt{\left(\mathbf{K}_i\mathbf{K}_i^T\right)}\sqrt{\left(\mathbf{K}_j\mathbf{K}_j^T\right)}} \;=\; \cos(\theta_{ij}^{\mathbf{K}}) \qquad (2.6)$$

where $\mathbf{K}$ is the orthogonal kernel matrix (i.e. all column vectors are perpendicular to each other) with m rows, $i, j \in \{1 \ldots m\}$, and $\theta_{ij}^{\mathbf{K}}$ is the angle between rows i and j in $\mathbf{K}$. It follows from the definition that the RCC must be in the range $[-1, 1]$, where a coefficient of ± 1 indicates that the reaction pair are completely correlated, i.e. in the same subset, and a coefficient of 0 indicates that they are completely disjoined, i.e. the reactions can never appear in the same flux solution. As mentioned above, the kernel matrix calculated by Gauss-Jordan elimination cannot be guaranteed to be unique. The angles between the row vectors of the kernel, however, can be shown to be unique, provided that the columns of the matrix are orthogonal [48]. This requirement can be ensured by either calculating the kernel using the Singular Value Decomposition method or by orthogonalising an existing kernel using the Gram-Schmidt method.

The reaction correlation coefficient is mathematically identical to Pearson's correlation coefficient (r) [48]. Although the form of Equation 2.6 is identical to r, it should be emphasised that none of the statistical interpretations of r are applicable to ϕ. Within the framework of this thesis r will mainly be applied to sets of flux solutions, rather than the entire null-space.

The reaction correlations of a network can be visualised using clustering of reactions into hierarchical trees, here referred to as *metabolic trees*. A commonly used method for generating hierarchical trees is the Weighted Pair Group Method using Arithmetic Averaging (WPGMA) algorithm [49, 50], which has been used for generating metabolic trees [48].

Conservation relationships

The definition of the null-space used so far is more specifically be referred to as the *right null-space* of $\mathbf{N}$. It is also possible to calculate the *left null-space*, defined as:

$$\mathbf{K}^T \cdot \mathbf{N} \;=\; \mathbf{0}^T$$

or

$$\mathbf{N}^T \cdot \mathbf{K} \;=\; \mathbf{0} \qquad (2.7)$$

where T in $\mathbf{N}^T$ denotes the transpose of matrix . Throughout this thesis, unless stated otherwise, null-space (kernel) refers to the right null-space (kernel). The left null-space can be used for identifying linear dependencies between
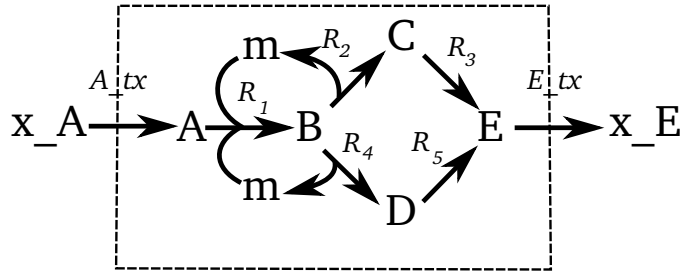
Figure 2.2: Hypothetical network where metabolites m and B are in a conservation relationship.

the rows of $\mathbf{N}$. Biochemically, this corresponds to conservation relationships between metabolite, i.e. metabolites that are associated such that the sum of their concentrations is constant [41].

An example of such a relationship is illustrated in Fig. 2.2, where the sum of the concentration of metabolites m and B must be constant. Metabolite m can be interpreted as a cofactor that participates in the catalysis of reactions R_1, R_2, and R_3, but is not consumed in the process. Since m does not leave the system all molecules of m must appear as free m or as a moiety of metabolite B, in other words $m + B = constant$. In this case m is a *conserved moiety* [41]. The stoichiometry matrix of the network is now:

$$\mathbf{N} = \begin{array}{c} m \\ A \\ B \\ C \\ D \\ E \end{array} \begin{bmatrix} -1 & 1 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & -1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 \end{bmatrix}$$

The left kernel of the new stoichiometry matrix will then be:

$$\mathbf{G} = \begin{array}{c} m \\ A \\ B \\ C \\ D \\ E \end{array} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

from which the conservation relationship between m and B can be deduced.

2.3.2   Elementary modes analysis

As demonstrated above, the null-space of the stoichiometry matrix is very useful for analysing relations between reactions that must be true for any flux solution realisable by the network. This does not mean that all solutions defined by the null-space are realisable by the network, since all reactions are treated as reversible in the kernel matrix. Moreover, the column vectors of the kernel are not unique. Hence, even if the column vectors forms a basis of the null-space, they are not invariant and individually they are not necessarily biologically meaningful. The *elementary modes* concept [51] overcomes many of the shortcomings of null-space based analysis. An elementary mode can be described as a component pathway, i.e. a combination of reactions that forms a route through the network. The complete set of elementary modes of a network gives a complete description of the network and any flux solution is a (non-negative) linear combination of elementary modes [52].

Definition

An elementary mode is a flux mode (i.e. a flux distribution with fixed proportions between reactions, but not fixed flux values [52]), such that it does not violate the steady state constraint, the reversibility constraints, and that the flux mode cannot be decomposed to smaller modes that fulfil these criteria. More formally, an elementary mode can be defined as a set of reactions, $\mathbf{M}$ [41]:

$$\mathbf{M} = \{\mathbf{v}|\ \mathbf{v} = \lambda\mathbf{v}^*,\ \lambda > 0\}.$$

Where three criteria must apply to the vector $\mathbf{v}^*$:

1. It must be a steady state solution.

2. Any irreversible reaction in $\mathbf{v}^*$ can only carry positive (or zero) flux.

3. There is no flux vector that fulfils the two first criteria and is a subset of $\mathbf{v}^*$ in the network.

For the network in Fig. 2.1 there are only two modes that fulfil these criteria, shown in Fig. 2.3. The mode in Fig. 2.3.C is only possible if the reactions involved are reversible.

An algorithm for the computation of all elementary modes of a network was suggested by Schuster and Hilgetag [51].

The original algorithm is efficient for relatively small networks. It does not, however, scale well for large and complex networks, since all possible routes must be evaluated [53]. Several alternative algorithms and improvements have
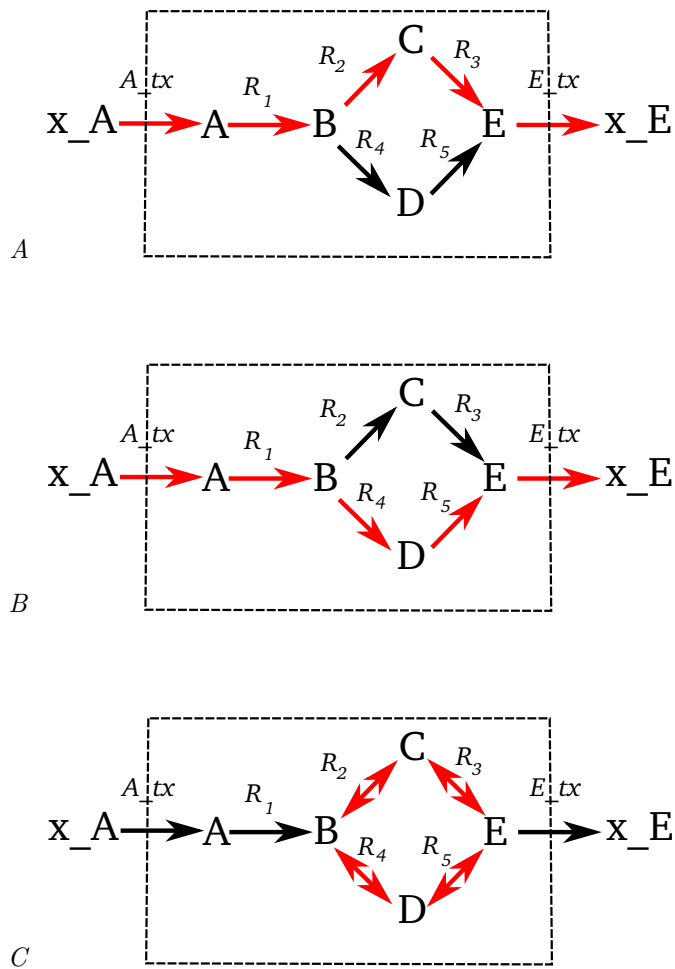
Figure 2.3: Elementary modes in a small hypothetical network. Reactions that participate in an elementary mode are indicated with red arrowheads. Modes in A and B have identical net stoichiometries ($x\_A \rightarrow x_E$), and are both irreversible. If reactions R_2 - R_5 are assumed to be reverible an additional elementary mode can be found (C). The mode in C is a metabolic cycle, thus it does not have a net stoichioemtry. If only one of the enzyme subsets in the cycle (R_2 and R_3, or R_4 and R_5) is reversible, the cycle will also be irreversibel.

been described, e.g. Wagner [54] suggested a null-space based approach and Gagneur and Klamt [55] suggested a binary approach.

Application

Elementary mode analysis has been successfully applied to several biological problems [56]. For example, Trinh et al. [57] identified a set of elementary modes in an *Escherichia coli* central carbon metabolism model that were not involved in desirable net reactions (the joint biosynthesis of biomass precursors and ethanol). By deleting the genes encoding the enzymes involved in the undesirable modes, very high yields of ethanol on glucose could be achieved. Klamt and Gilles [58] addressed the general problem of identifying minimal sets of reactions for deletion in order to abolish an undesirable reaction. This work, referred to as minimal cut set analysis and described in more detail in Section 2.6, has many potential application in metabolic engineering and rational drug discovery [59].

An alternative approach for enumerating transformation routes through a metabolic network, referred to as *extreme pathways*, was suggested by Schilling et al. [60]. Extreme pathways are similar to elementary modes in that they define flux modes that are non-decomposable and satisfy the steady state constraint. For a network where all reactions are irreversible the elementary modes and the extreme pathways are identical. If there are reversible reactions in the network the extreme pathways will be a subset of the elementary modes. Extreme pathways, unlike elementary modes, must by systemically independent [61]. This means that no extreme pathway is a non-negative linear combination of other extreme pathways. Although elementary modes are non-decomposable and linearly independent, some modes can be obtained by non-negative combinations with other. Consider the modes in Fig. 2.3. If the reversible mode in Fig. 2.3.C is split into two irreversible modes, the extreme pathways of the network would be the two cycles (with opposite direction) and one of the net transformations of x_A to x_E (i.e. Fig. 2.3.A or Fig. 2.3.B). This is because given e.g. the mode in Fig. 2.3.A, the mode in Fig. 2.3.B can be generated by adding the cycle where R_2 and R_3 are operating in reverse to the mode in Fig. 2.3.A. It should be noted that since the extreme pathways for some networks are a subset of the elementary modes, applications that rely on a complete enumeration of all routes through the network might fail if extreme pathways are used. Owing to the reduced complexity of extreme pathways, compared to elementary modes, calculating the extreme pathways of a network is less computationally expensive [62]. The similarities and differences between extreme pathways and elementary modes are discussed in [61–63].

2.3.3 Flux balance analysis

A modelling technique that has gained increasing popularity over the last decade is *Flux Balance Analysis (FBA)*. Unlike the methods mentioned above, FBA requires that some of the fluxes in the system are known, and produces a (possibly non-unique) flux solution. This makes the method useful for applications where integration of different types of experimental data is sought. A common type of experimental data used for FBA, especially when applied to whole cell models, are measurements of biomass components. These could include the concentrations of amino acids, nucleotides, membrane components, and cofactors.

FBA is the application of Linear Programming (LP) to underdetermined, stoichiometric models. Briefly, LP is a mathematical optimisation method for systems of linear equations, where given a set of constraints and an objective function, the method will find a solution that optimises the objective. The value of this solution, in terms of the objective function, is referred to as the objective value. Applied to metabolic models this involves minimisation or maximisation of some fluxes in the network, subject to some constraints. The constraints include the steady state and often bounds on fluxes (i.e. maximum or minimum, including negative, flux values). Reactions could also be assigned fixed fluxes. Commonly used objective functions include maximisation of the yield of biomass or ATP [64] and minimisation of total flux [65, 66]. Maximisation of biomass yield often involves defining an hypothetical *biomass reaction* where each substrate is a biomass component and the experimentally determined proportion of the component in biomass is used as stoichiometric coefficient, the product of this reaction is a unit of (external) biomass [67]. Yield maximisation is carried out by maximising the flux through this reaction, which is interpreted as the growth rate of the cell, while maintaining a fixed uptake of some carbon source. This optimisation is often referred to as maximisation of growth rate [68, 69]. It has been argued [70–72] that this type of optimisation is an optimisation of growth yield, and more generally, that any FBA problem is a yield optimisation problem. This is because the only way to optimise an output rate of a stoichiometric metabolic model is to select the most efficient route through the network, i.e. a route the minimises flux through output reactions that compete with the objective output reaction. This is indistinguishable from maximising the yield of the objective reaction. The growth rate of a cell is a function of kinetic properties of the metabolic network and cannot be modelled using currently available methods in structural modelling.

The first applications of LP to metabolic models were described by Fell and Small [73] and Watson [74]. It was later developed by Varma and Palsson [64, 75] under the name Flux Balance Analysis. FBA is currently applied to a great

variety of biological questions and metabolic networks of different complexity [76].

Apart from LP, a related optimisation technique, Mixed Integer Linear Programming (MILP), has also been applied to metabolic network modelling. In MILP problems each variable is associated with an integer value (0 or 1), as well as a continuous value. MILP can be used for problems where a discrete quantity is optimised, e.g. the number of active or inactive reactions. A noteworthy application is the iterative use of MILP to identify qualitatively different solutions to an FBA problem that have the same objective value [77]. MILP problems are, however, more complex and computationally expensive compared to LP problems.

Metabolic example

A simple example of how FBA can be applied to the network in Fig. 2.1 is given below. Assume the flux value of E_tx is known to be 10 and the solution with the minimal flux sum that can satisfy this constraint, as well as the steady state constraint, is sought. This problem can be stated as:

$$\text{Min} \quad : \sum_{j=1}^{7} v_j$$
$$\text{subject to} \quad \begin{cases} \mathbf{N}_{5,7} \cdot \mathbf{v} = \mathbf{0} \\ E\_tx = 10.0 \end{cases}$$

In this case the solution is the rather trivial elementary mode in Fig. 2.3.A with all active reactions normalised to 10.

2.4 Construction and analysis of genome-scale metabolic models

Structural modelling is popular owing to its mathematical simplicity and the availability of experimental data required for performing this type of modelling. These features also makes it highly scalable to large and complex networks, including whole cell models, or *genome-scale metabolic models (GSMs)*, and models of communities of different cells. Within the context of this thesis, this term refers to a structural model where all reactions are ultimately encoded by the genome of the organism of interest, and similarly, all known metabolic genes are accounted for in the model. There is of course no lack of interest in constructing genome scale kinetic models, but there are significant problems with this ambition, primarily related to the acquisition of experimental data and performing dynamic simulation, if a model could be constructed. These concerns are discussed in [78, 79].

2.4.1 Genome annotation

The availability of annotated genomes of high quality is a prerequisite for construction of GSMs. Since whole genome sequencing can be performed rapidly and with high accuracy and low cost, at least for bacteria [80], generation of sequence data is not a major obstacle for GSM construction. Genome annotation, i.e. assigning functions to the sequenced genes, is however potentially problematic. From the model construction viewpoint, the main concern is the annotation of genes encoding metabolic enzymes: the crucial information needed is the stoichiometry of the enzyme catalysed reaction and, if available, the reversibility of the reaction. It must be emphasised that the association between genes, proteins, enzymes, reactions, and metabolites is usually not a one-to-one, but a many-to-many, relationship. For example, a single gene could encode an enzyme that is capable of carrying out multiple metabolic functions. In that case the enzyme would be represented by multiple reactions in a GSM. Alternatively, an enzyme complex consisting of multiple proteins encoded by multiple genes could carry out a single reaction. This complexity must be taken into account when metabolic metabolic models are used for integration of gene expression data or when modelling results are used to guide gene deletion experiments. But, the association between different levels of biological information is essentially a problem relating to bioinformatics, not metabolic modelling. Hence, in this thesis, extraction and analysis of information relating to genome annotation is carried out using data obtained from publicly available databases (described in more detail below), and GSMs are constructed using this information. But the genome annotation does not form a part of the models.

Ultimately, all genome annotations are based on experimental data. For the central carbon metabolism of a set of well characterised model organisms, most metabolic genes can be assigned functions with high confidence. For more peripheral parts of metabolism, especially for poorly characterised organisms, the genome annotation is more indirect, and relies heavily on bioinformatic methods. Most automatic genome annotation pipelines identify potentially protein coding genes, translate the genes to protein sequences and perform some sort of similarity search where unknown predicted protein sequences are compared with a database of proteins with known sequence and function [81]. Most methods for performing similarity searches are based on the BLAST (Basic Local Alignment Search Tool) algorithm [82].

2.4.2 Biochemical databases

Apart from annotated genomes, biochemical databases are crucial aids for GSM construction. Given an annotated genome, where the association between genes

and metabolic enzyme reactions has been established, some of the remaining challenges include how to assign correct stoichiometries to the enzyme reactions and how to guarantee that names for reactions and metabolites are used consistently. For example, a genome annotation could suggest that a gene encodes an alcohol dehydrogenase (with EC number 1.1.1.1), but this only suggests a very generic stoichiometry of the type $alcohol + NAD \rightarrow aldehyde + NADH$. In order to incorporate this reaction into a metabolic model it would need to be replaced by a set of reactions with specific metabolites. Inconsistent use of metabolite identifiers, i.e. where a metabolite is referred to by different identifiers, introduces, potentially unintentional, constraints on the model since each identifier will be treated as a separate metabolite in the model. Inconsistent naming of reactions introduces unnecessary complexity since multiple reactions with identical stoichiometries will be included in the model.

These, and other, potential problems can be minimised by using well curated biochemical databases for model construction. Two well known databases suitable for model construction are KEGG [83, 84] and BioCyc [85, 86]. Both are metabolic databases that use their own identifiers for reactions, metabolites and pathways. The requirement that metabolite names should be used consistently has previously not been fully meet in KEGG (reviewed in [87]). A feature of the BioCyc database that makes it useful for model construction is the availability of organism specific databases, Pathway/Genome Databases (PGDBs), which are collections of the database entries (reactions, metabolites and pathways) that, based on genome annotation, are present in a given organism. These PGDBs are initially constructed using the software package Pathway Tools [88, 89]. Most databases undergo some manual curation, although this varies significantly depending on organism. A major concern with insufficiently curated PGDBs is the presence of "gap filling reactions". These are reactions that are introduced to PGDBs by the Pathway Tools software (specifically the PathoLogic component of the software) if some, but not all, of the reactions in a pathway are predicted to be present according to the genome annotation. The rationale for this procedure is the assumption that genome annotation software misannotates genes encoding multifunctional enzymes as unifunctional, thus creating holes in the metabolic network [88]. Since some of the gap filling reactions are likely to be false positives this could overestimate the metabolic capacity of the organism.

Other databases that are of relevance for model construction, and especially curation, include BRENDA and TECRDB. BRENDA (BRaunschweig Enzyme DAtabase) [90] is a database collecting information on metabolic enzymes. All information is manually collected from published data. The database contains much information relevant for kinetic modelling, e.g. kinetic parameters, reaction mechanisms, and known effectors. Information relevant for GSM curation

Table 2.1: Summary of databases used for model construction and curation within the framework of this thesis. URLs for the databases can be found in the list of URLs, under Nomenclature.

Database	Dates used	Version
BioCyc	1 November 2010 - 1 November 2012	15.0
BRENDA	1 November 2010 - 1 November 2012	-
TECRDB	1 November 2010 - 1 November 2012	-

includes reaction specificity and reversibility. The TECRDB [91] (Thermodynamics of Enzyme-Catalyzed Reactions DataBase) collects experimental data on thermodynamics of enzyme reactions. The is primarily useful for assigning reaction reversibility, since reactions that are strongly thermodynamically favoured in one direction are unlikely to be carry flux *in vivo* in the other direction. Entries in the database are sorted into classes depending on the quality of the experimental data. Databases used for model construction are summarised in Table 2.1.

2.4.3   Model curation

In initial metabolic model constructed from a public database is likely to contain numerous errors of varying severity [87]. Inconsistent naming of reactions and metabolites should be avoided since it results in unnecessary complexity (inconstant reaction names) and artificially disconnected networks (inconstant metabolites names). Missannotations, i.e. erroneous inclusion or exclusion of reactions, are a significant source of errors in GSMs. Unfortunately, these errors can only be dealt with by considerable manual effort. On the other hand, since these errors cause a deviation in network topology from the biological reality they are traceable. For example, consider the case where an FBA problem for a GSM is constrained such that consumption of a certain external carbon source is required, but no feasible solution can be obtained, despite experimental evidence that the organism can use the carbon source and feasibility of the FBA problem constrained for consumption on other carbons sources. In this case it can be assumed that the deviation is ultimately caused by missannotation, which can be resolved by considering biochemical literature, performing detailed similarity searches, and, ideally, experimentally confirming the activity of the missing reactions, followed by model correction.

A class of errors that are more difficult to detect, and thus more concerning, are related to material imbalances. These include simple imbalances that are relatively unproblematic to detect, such as `R186-RXN`, which in BioCyc version 11.1 was reported to have the stoichiometry:

$$\texttt{CPD-552} \ + \ \mathbf{2}\,\texttt{E-} \ + \ \mathbf{2}\,\texttt{HSO3} \leftrightarrow \texttt{S2O3} \ + \ \mathbf{2}\,\texttt{HSO3}$$
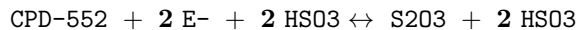
where CPD-552 is trithionate (with empirical formula S_3O_6). The stoichiometry implies that R186-RXN reduces CPD-552 and simultaneously eliminates one atom of sulphur and three atoms of oxygen, which is clearly incorrect. The inconsistency can be resolved by removing the two bisulphate (HSO3) molecules on the right hand side and one bisulphate molecule on the left hand side of the equation. In cases where the empirical formula is unknown imbalances like these are difficult to identify and resolve. A method for detecting material inconsistencies of this type was suggested by Gevorgyan et al. [92]. Reaction stoichiometries can be considered as statements about compounds. For instance, the reaction

$$A \ \rightarrow \ B$$
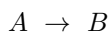
suggests that A and B have the same empiric formula. If the network also contains a reaction with the stoichiometry

$$A \ \rightarrow \ B \ + \ C$$

the two reactions can only be consistent if C has a mass of zero. These stoichiometric inconsistencies are always caused by material imbalances. The algorithm described in [92] identifies the maximum number of *unconserved metabolites* in the network, using a mixed integer linear programming approach. These metabolites can be identified by analysis of the left kernel of the stoichiometry matrix (Equation 2.7) as metabolites that are not involved in any nonnegative conservation relationships [93]. In the example above metabolites A and B are in such a relationship. An advantage of the algorithm suggested by Gevorgyan et al. [92] is that it can identify unconserved metabolites solely on the bases of reaction stoichiometry.

2.5 Analysis of genome-scale models

Since GSMs are large structural metabolic models all methods from structural analysis can be applied to GSMs. An advantage of null-space based analysis (Section 2.3.1) is that is is very general - conclusions drawn on the basis of the null-space of a stoichiometry matrix will apply to all particular steady state solutions achievable by the network, although all particular solutions defined by the null-space might not be achievable by the network owing to reversibility constraints. A consequence of the generality of null-space analysis for GSMs is that results tend to be difficult to relate to biochemically relevant situations.

For examples, as model complexity increases, size and number of enzyme subsets tend to decrease [48].

With FBA (Section 2.3.3) the solution space can be constrained to a relevant sub-space by incorporation of reaction reversibility and constraints on transporter reactions. On the other hand, FBA solutions are single points in solution space and thus on the opposite end of the generality spectrum, compared to kernel matrices. Poolman et al. [94] suggested an application of FBA where sets of solutions along one dimension of flux space (reaction) are collected and analysed. This is achieved by setting up an FBA problem and repeatedly solving it with the variable reaction fixed to different values. Reactions that appear in the collected solutions and change flux value in response to the changes in the constrained reaction can be considered to be coordinated with the this reaction. By selecting a physiologically relevant reaction, such as hydrolysis of ATP [94], the set of responding reactions will form a sub-network that balances the constrained reactions. This method is used for model analysis in Chapters 3, 7, and 4, and explained in more detail there.

2.6   Damage analysis

An important general application of metabolic modelling is analysis of how changes in network topology affect the solution space. A subset of this application involves identification of sets of reactions that when removed from the network abolish a certain function. When applied to metabolic models of pathogens this type of analysis can be used for identification of potential drug targets. A general feature of structural models is that the solution space is highly dependent on the metabolites assumed to be external. This can be seen by studying the algorithm for calculating elementary modes (Section 2.3.2), where at the start of the algorithm all metabolites are assumed to be external (and all reactions elementary modes), then as the algorithm progresses, metabolites are sequentially made internal and preliminary elementary modes are combined to form the actual elementary models of the model. Hence, if metabolites are incorrectly made external this could introduce unrealistic elementary modes. This is problematic when models of pathogen metabolism are used for drug target identification since the nutrient availability during infection is largely unknown [95]. This problem is addressed in Chapters 7 and 6.

Several methods for identification of drug targets in metabolic network have been suggested. The most fundamental method is perhaps *minimal cut set, (MCS)* analysis [58]. MCS involves the identifying the minimal sets of reactions that when removed from the network abolish a specific target reaction. The MCS of a network are commonly determined from the elementary modes, which

limits the analysis (as defined in [58]) to networks where elementary modes are computable. That MCSs can be determined from the elementary modes can be seen by considering the elementary modes in Fig. 2.3. MCS analysis is based on blocking each elementary mode that involves the target reaction by removing the minimal number of reactions from the network. If E_tx is the target reaction only elementary modes involving this reaction needs to be considered (i.e. Fig. 2.3.A and 2.3.B). From these two modes it can be concluded that reactions A_tx and R_1 appear in both modes and that they are each essential, hence each is an MCS. Next, each target mode involves two mode specific reactions, R_2 and R_3 for mode A and R_4 and R_5 for mode B. Hence, in order to block both modes, any combination of these two sets of reactions will also be MCSs. The complete MCSs of the network are: $\{A\_tx\}$, $\{R_1\}$, $\{R_2,\ R_4\}$, $\{R_2,\ R_5\}$, $\{R_3,\ R_4\}$, and $\{R_3,\ R_5\}$. Note that e.g. $\{R_1,\ R_2\}$ is also a cut set, but it is not minimal.

There are several methods for drug target identification that are based on the application of FBA to GSMs. One method is to identify essential genes in the network [96]. This is done by setting up a standard FBA problem and sequentially setting the flux of each reaction (or sets of reaction), associated with a gene, to zero and attempting to solve the FBA problem. If the problem is infeasible the constrained reaction, and consequently the gene, is essential. Gene essentiality analysis is often used to validate GSMs when experimental reference data is available. It is also possible to extend this analysis to combinations of two or more genes or reactions, although combinations of three or more is impractical owing to the combinatorial explosion for GSMs. In functional genomics a pair of genes are called synthetic lethal if the deletion of both genes results in a non-viable phenotype (but the single mutants are viable) [97]. This terminology is often used to describe FBA based essentiality analysis of GSMs, it should be emphasised that synthetic lethality only refers to the definition given above, and does not extend to reactions or sets of genes different from two.

Kim et al. [98] described an FBA based method, metabolite essentiality analysis, for identification of potential drugs, rather than drug targets. Computationally the method is similar to gene essentiality analysis, described above, but involves sequential removal of metabolites (by removing sets of reactions involved with a given metabolite) from the network.

2.7   Software for metabolic modelling

Owing to the size and complexity of metabolic models, biologically meaningful applications of metabolic modelling becomes impractical and error prone, even for fairly small models, without the use of computers.

2.7.1 Overview

There is a long history of applying computers to analysis and simulation of metabolic systems going back to the early 20th century [99–101]. Because of this there is a wide range of software available for metabolic modelling.

A common feature of most modelling software is that it simplifies the process of converting a file specifying properties of the model (e.g. stoichiometry and kinetics, if available) into mathematical objects, in addition to performing the mathematical analysis. Many of the early programs for metabolic modelling were developed for simulation of kinetic models, such as BIOSSIM [102], METASIM [103], FACSIMILE [104], and SCAMP [105, 106].

These early software packages were all script based and could only be run in batch mode, i.e. they consisted of files and libraries that could be called by program written by the user. A popular alternative to the script based approach is the Graphical User Interface (GUI), which is often considered easier to use for novice modellers since it does not require any knowledge in programming [101]. An early, and still widely used, application of GUIs in metabolic modelling software is Copasi (previously Gepasi) [107, 108], which is also primarily intended for kinetic modelling.

METATOOL [46] is one of the earlier software packages devoted to structural modelling. It includes implementations of algorithms for determining elementary modes and enzyme subsets, and uses a script based user interface. It has subsequently been succeeded by the GUI based YANA package [109]. Other software specialised on structural analysis include CellNetAnalyzer [110] (successor to FluxAnalyzer [111]), which includes algorithms for determination of minimal cut sets. CellNetAnalyzer (and FluxAnalyzer) run in the commercial computer environment MATLAB, and the user interface is the MATLAB Command Line Interface (CLI). The COBRA Toolbox [112] is another modelling package that runs on MATLAB, although an implementation using the free programming language Python (described in more detail in Section 2.7.2) was recently described [113]. It is mainly dedicated to FBA applied to GSMs. Similarly, SurreyFBA [114] is also focused on applications related to FBA analysis of GSMs. It is implemented in Jython (a combination of the programming languages Python and Java), and supports both GUI and CLI.

Two packages that are primarily written in Python are PySCeS and ScrumPy. PySCeS [115] is mainly designed for kinetic and metabolic control analysis, but offers some support structural modelling. ScrumPy [47] (described in more detail in Section 2.7.3) provides support for structural as well as kinetic modelling (including metabolic control analysis). All modelling work in this thesis was carried out using ScrumPy.

This diversity in metabolic modelling software is both an asset, in that users with different interests and preferences have a wide selection to choose from, and a liability, since it makes interactions between different groups and sub-disciplines difficult. Fortunately, a common model definition language, Systems Biology Markup Language (SMBL) [116], which aids exchange and comparison of models between modelling platforms, has been developed. The SMBL format supports many types of biological models and is not limited to metabolic models.

2.7.2   The Python programming language

Python is a multipurpose programming language, first released in 1991. It is a high-level language, which means that details concerning internal management of computer system, that may not be of immediate interest for the user, are automated. Python supports a number of programming paradigms [117]: *Object oriented* - data and functions can be encapsulated in objects, which are instances of generic classes, this paradigm promotes a highly modular and structured management of information; *imperative* - computation is carried out as a sequence of statements; *functional* - can be seen as the opposite of the imperative paradigm, functional programs only evaluate statements; *reflective* - programs are allowed to modify themselves under certain circumstances.

It was designed to combine many desirable traits [118]:

- *Readability:* Python syntax is designed to be clear and coherent, and thereby easy to maintain. This does not guarantee, but promotes, high quality and less error prone software.

- *Extensibility:* Many extensions of Python are available, this is especially important for scientific programming. Notable examples of this are SciPy and NumPy.

- *Productivity:* A given program can be written in Python using less code compared other languages, such as C, C++, or Java. This in combination with clear and readable syntax makes software development both fast and reliable.

- *Portability:* Python runs predictably on all major operating systems.

- *Libraries:* A large collection of special purpose software, referred to as the standard library, is distributed with Python.

- *Integration:* Python can be integrated with software written in other languages using a variety of integration mechanisms.

- *Flexibility:* The combination of extensibility, availability of libraries and interfaces to other languages makes Python highly flexible.

- *Transparency:* Python is free and open source.

2.7.3 ScrumPy: metabolic modelling with Python

The ScrumPy metabolic modelling package is a collection of programs written in Python for construction and analysis of metabolic models [87]. The currently available versions are adapted for Unix-like operating systems. The command line user interface of ScrumPy is built on the standard Python IDLE, which is an Integrated Development Environment (IDL) for Python development. The absence of GUIs, for some features of model analysis, in ScrumPy is a deliberate design feature motivated by the loss of flexibility and extensibility involved with overreliance on GUIs. ScrumPy can be used in both interactive and batch mode, which promotes two key features of Python: productivity - the interactive mode promotes a fast development cycle; reusability - reliable code for model analysis can be stored as programs and reused in batch mode. Many of the numerically demanding modelling tasks, such as elementary modes calculations, numerical integration of systems of rate equations, and the LP component of FBA, are carried out by integrated third party software, predominantly written in C.

The next chapter describes the construction and analysis of a genome-scale model of *Salmonella* Typhimurium, which involves much of the methodology covered here.

Chapter 3

Construction and analysis of a *Salmonella* Typhimurium genome-scale metabolic model

3.1 Introduction

This chapter describes the construction of a genome-scale model of the metabolic network of *Salmonella* Typhimurium LT2, and analysis of the resulting model. Much of the work discussed in later chapters is applied to, or otherwise involves, the model described in this chapter. The aim of the work described here is two-fold: (*i*) to construct a structural metabolic model based on the genome sequence of *S.* Typhimurium using publicly available data sources; (*ii*) to analyse the simulated response of the model to a physiologically relevant stimulus, and, based on this response, construct a functional sub-network of the genome-scale model. The methodologies and challenges associated with the first aim, construction and validation of GSMs, are reviewed in Chapter 2 (Section 2.4).

Three GSMs of the *S.* Typhimurium metabolic network have been published [119–121] previously. As reviewed in Chapter 2 (Section 2.4), stoichiometric inconsistencies in structural models can be identified using a MILP-based method [92]. In this chapter, the three previously published models, as well as the model presented here, are subjected to that test. One of the primary motivations for constructing a *Salmonella* GSM *de novo* is that this model will be used in later chapter for integration of experimental data. This work requires that the model

is open for modifications.

The second aim of this chapter, analysing the model's response to a simulated stimulus, involves computational approaches introduced in Chapter 2, such as FBA (Section 2.5) and null-space based methods (Section 2.2). In brief, the analysis is carried out by setting up and solving a set of FBA problems, where for each new solution, the flux constraint on a physiologically relevant reaction is increased by a constant value. The constraint that all biomass components must be produced in experimentally observed proportions is maintained in each solution. By collecting these solutions and identifying reactions that are correlated with the stimulus, a subset of the reactions in the model can be identified and analysed as a smaller, separate, model.

Null-space based methods can be used to group reactions in a model into correlated sets. Since these correlations apply for all possible states the sets of highly correlated reactions tend to be very small for GSMs. By defining a smaller model, with a smaller solution-space, based on the reactions that carry flux in response to an LP-imposed stimulus, a more focused and informative null-space analysis can be carried out. This is the motivation for the approach to model analysis used in this chapter. Here, the model response to changes in energy demand, modelled by the ATP hydrolysis reaction, is used to identify the set of reactions employed to meet this imposed increase in energy demand, thus in effect identifying the energy generating sub-network of the GSM. This method has been used previously in analysis of an *Arabidopsis thaliana* GSM [94]. Once this sub-network is extracted null-space based methods are used to characterise it.

3.2   Method

The model was constructed in a modular fashion using techniques described in Section 2.4, to generate five modules. These consisted of:

1. A 'top-level' module, the prime function of which is to import the other modules.

2. Automatically generated reactions.

3. Transporters.

4. Electron transport chain to generate proton-motive force.

5. Additional reactions.

and are described in the following sections. This defines a metabolic network that could be shown to be capable of generating all biomass components from a

minimal medium composed of glucose, NH_3, PO_4, SO_4 (and O_2). Biomass was assumed to be comprised of protein (as individual amino acids), lipopolysaccharides (LPS), peptidoglycan, membrane, DNA and RNA. For a complete table of species and relative abundance, see Appendix A.

Automatically generated reactions

The BioCyc flatfiles corresponding to gene, protein, enzyme, reaction and compound mappings for "*Salmonella enterica* enterica serovar Typhimurium str. LT2" were obtained from the BioCyc ftp site (registration required) [122, 123] and processed to extract relevant reaction information.

Reactions involved with non-metabolic species (e.g. "Damaged-DNA-Pyrimidine") or generic compounds (e.g. "Aromatic-Oxoacids") were removed (see Appendix C). The atomic balance, in terms of C, N, P and S, was verified for the remaining reactions.

Polymeric species, consisting of an undefined number of monomeric units, can give rise to mass inconsistancies if they appear in reactions with stoichiometric coefficient that are not proportional to the number of monomeric units involved in the reactions. For example, the BioCyc database reports a starch synthesis reaction with the stoichiometry:

$$\texttt{ADP-D-GLUCOSE} \rightarrow \texttt{ADP} + \texttt{Starch}$$
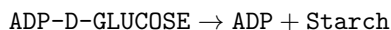
in which a single glucose moiety is added to starch, and an amylase reaction with the stoichiometry:

$$\texttt{Starch} \leftrightarrow \texttt{MALTOPENTAOSE}$$

in which five glucose moieties are removed from starch. This allows the two to operate in tandem, with a net stoichiometry:

$$\texttt{ADP-D-GLUCOSE} \rightarrow \texttt{ADP} + \texttt{MALTOPENTAOSE}$$

with an overall conversion of 1 glucose moiety into five glucose moieties. The solution is to re-write such stoichiometries with the stoichiometry of the polymeric species weighted by the number of monomeric moieties on the opposite side of the equation, so in the example above, glucose is taken as the monomer and the amylase reaction re-written as:

$$\textbf{5}\ \texttt{Starch} \leftrightarrow \texttt{MALTOPENTAOSE}$$

thus solving the problem.

Transport reactions

The transport module defines all reactions involving transfer between external and internal compartments, with the exception of protons (see following section), i.e. all biomass and media components plus CO_2.

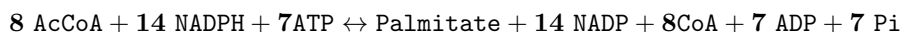All transport reactions were defined with the external species on the left hand side, so that all subsequent flux values, whether set as a constraint or as part of a solution, can be consistently interpreted as positive flux representing transport into the system and negative flux as loss from the system.

Electron transport chain

A realistic representation of the electron transport chain (ETC) depends upon a number of proton translocating reactions. In order to accommodate this, these reactions consists of a number of proton-translocations over the inner membrane, resulting in a proton concentration gradient which is used for driving phosphorylation of ADP. Since protons were declared external, translocation of protons cannot be coupled to ADP phosphorylation in a meaningful way. This problem was dealt with by treating the proton translocation process as a set of transformations of (external) intracellular protons to (internal) periplasmic protons under the consumption of redox energy. Due to the steady state constraint the periplasmic protons thus produced must be consumed in the ATP synthase reaction (which phosphorylates one ADP to ATP under the consumption of 4 perimplasmic protons). To this end, a separate ETC module was constructed. The ETC module had one single elementary mode, corresponding to a P/O ratio of 1.25, which can be compared with the literature value for *E. coli* of 1.33 [124]. This module also included a generic irreversible ATPase reaction (ATP hydrolysing).

Additional reactions

Since many of the automatically generated reactions in the *S.* Typhimurium PGDB involved in fatty acid synthesis included compounds without defined atomic composition, a generic fatty acid synthetase was defined. This generic reaction converts acetyl-CoA, NADPH, and ATP into palmitate, accordingly:

$$8\ \texttt{AcCoA} + 14\ \texttt{NADPH} + 7\texttt{ATP} \leftrightarrow \texttt{Palmitate} + 14\ \texttt{NADP} + 8\texttt{CoA} + 7\ \texttt{ADP} + 7\ \texttt{Pi}$$

Consequently, the synthesis of all lipid-containing biomass components were specifically defined with palmitate as a reactant. These biomass components included phosphatidate, phosphatidylserine, phosphatidylethanolamine, phos-

phatidylglycerol, cardiolipin, and 1,2-diacylglycerol. Similarly, the synthesis of the biomass components peptidoglycan and lipopolysaccharide (LPS) contained many atomically undefined intermediates and was replaced with a lumped reaction, based on pathway information obtained from BioCyc, of defined reactants:

$$\textbf{3 } \texttt{Ala} + \texttt{PEP} + \texttt{NADPH} + \texttt{Glt} + \textbf{5 } \texttt{ATP} \leftrightarrow$$

$$\texttt{Peptidoglycan} + \textbf{5 } \texttt{ADP} + \textbf{6 } \texttt{Pi} + \texttt{UMP} + \texttt{UDP} + \texttt{NADP}$$

$$\textbf{2 } \texttt{CMP-KDO} + \textbf{4} \texttt{ADP-GMH} + \textbf{3 } \texttt{UDP-Glc} + \textbf{2 } \texttt{UDP-GA} + \textbf{6 } \texttt{Palmitate} + \textbf{3 } \texttt{ATP} + \texttt{UDP-Gal} \leftrightarrow$$

$$\texttt{LPS} + \textbf{7 } \texttt{ADP} + \texttt{Acet} + \texttt{UMP} + \textbf{5 } \texttt{UDP} + \textbf{2 } \texttt{CMP}$$
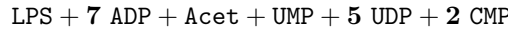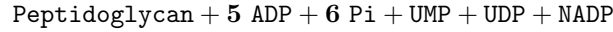
Nucleoside phosphatase reactions with generic substrates (nucleoside tri-, di- or mono-phosphate) were replaced with reactions including defined reactants. In a similar fashion, ribonucleoside reductase reactions with generic reactants were re-defined with specific reactants.

A number of additional reactions, that were required for synthesis of all biomass components individually, were added to the model after confirming the presence of genes encoding the corresponding enzymes.

3.2.1   Software and tools

The construction-process and subsequent analysis of the model was carried out using the ScrumPy software package [47]. All model analysis by LP was based on the Gnu Linear Programming Kit (GLPK) integrated with ScrumPy. Calculations of flux distributions were done with minimisation of the sum of fluxes as the objective function and the allowed solutions were constrained by fixing the flux of transport reactions specifying biomass precursor export. The LP was defined as stated below:

$$
\begin{aligned}
\text{minimize} \quad &: \quad \sum_{i=1}^{m} |v_i| \cdot c_i \\
\text{subject to} \quad &\begin{cases} \mathbf{N}_{n,m} \cdot \mathbf{v} = \mathbf{0} \\ v_j = t_j; \ k \leq j \leq m \\ v_{\text{ATPase}} = r_{\text{ATP}} \end{cases}
\end{aligned}
\tag{3.1}
$$

Where $\mathbf{v}$ is the vector of reaction rates, $\mathbf{N}$ is the stoichiometry matrix, with n rows (metabolites), and m rows (reactions), where each reaction is associated with an objective coefficient, c_i, which signifies the cost of a particular reaction. For the analysis conducted here, all objective coefficients were set to one. $v_{k...m}$ are the rates of the transporter reactions involved with biomass

components (collected as the kth to mth reactions of $\mathbf{N}$), fixed to the values or ranges specified in the vector $\mathbf{t}$. The transport vector $\mathbf{t}$ was calculated as the product of the vector of biomass component concentrations (Appendix A) and a fixed growth rate (μ_{max}). A generic ATPase reaction (v_{ATPase}) was used to represent the consumption of energy for maintenance and polymerisation of biomass compounds (r_{ATP}, determined using Equation 3.3). Thus, the LP is set to minimise the sum of fluxes, under the constraints that the steady state assumption is obeyed, that biomass precursors are produced in proportion to the biomass composition, and that specified nutrients and excretory products can be produced.

3.2.2 Model curation

As described in Section 3.2 all automatically generated reactions were assumed to be irreversible until shown otherwise. The rationale behind this strategy, as described elsewhere [94], is as follows: regardless of whether all automatically generated reactions are assumed to be reversible, and curated using thermodynamic data for all reactions, or assumed to be irreversible, and curated when required, the end result of both strategies is a set of reactions producing biomass precursors while obeying thermodynamic constrains. Two methods were used to identify reactions with possible inaccurate reversibility - identification of inconsistent enzyme subsets (described in Section 3.2.2) and analysis of biomass synthesis feasibility (described in Section 3.2.2). Reversible reactions are summarised in Appendix C.

Inconsistent subsets

Determination of inconsistent subsets, as defined in Chapter 2 Section 2.2, was used to identify reactions that needed to be redefined as reversible or defined in the opposite direction. A total of 32 inconsistent subsets were detected in the model, and reversibility evaluated based on information from on-line databases TECRDB and BRENDA.

Feasibility of biomass synthesis

In addition to the procedure described in Section 3.2.2, identification of candidate reactions for reversibility redefinition was done by constraining the LP (Section 3.2.1) to produce each biomass component in turn. When no feasible solution was found all reactions in the model were temporarily declared reversible, the LP re-solved, and all reactions carrying negative flux, that previously were defined as irreversible, were evaluated for re-definition as reversible or defined in the opposite direction.

Model-wide validation of energy and material consistency

Energy consistency was evaluated by setting glucose transport to zero and the generic ATPase to an arbitrary positive value and solving the LP. If a feasible solution existed, the reversibility and directionality of the reactions in the flux solution were checked against thermodynamic data as described in Section 3.2.2. A similar procedure was used for oxidation of the currency metabolites NADH and NADPH. The model was subjected to a check for stoichiometric inconsistency [92] and described in Chapter 2, Section 2.4.

Consistency validation of existing *S.* Typhimurium GSMs

The previously published *Salmonella* GSMs were converted to ScrumPy format and the biomass reactions of all three models were decomposed into several transporter reactions, one for each biomass component included in the original biomass reaction, and the models checked for stoichiometric inconsistency [92].

3.2.3   Model analysis

Simulated ATP demand variation

Variations of ATP demand were simulated by solving the LP:

$$
\begin{aligned}
\text{minimize} \quad &: \quad \sum_{i=1}^{m} |v_i| \cdot c_i \\
\text{subject to} \quad &\begin{cases}
\mathbf{N}_{n,m} \cdot \mathbf{v} = \mathbf{0} \\
v_j = t_j; \ k \leq j \leq m \\
v_{\text{ATPase}} = J_{\text{ATPase}}; \ J_{min} \leq J_{\text{ATPase}} \leq J_{max}
\end{cases}
\end{aligned} \tag{3.2}
$$

where v_{ATPase} (the rate of the ATPase reaction) is fixed to the set value J_{ATPase}. All other symbols have the same meaning as in Equation 3.1. The analysis was performed by solving the LP repeatedly, where for each solution all constraints remained fixed, except J_{ATPase} which was fixed to increasing values in a given range. In order to simulate changes of ATP demand in a physiologically relevant range, the lower limit of J_{ATPase} was set to an estimated rate of total ATP demand (Equation 3.3, r_{ATP})

$$
r_{\text{ATP}} = Y_{\text{ATP}} \cdot \mu + m_{\text{ATP}} \tag{3.3}
$$

which, using parameters derived from *E. coli* ($Y_{\text{ATP}} = 60$ mmol ATP (g DW)$^{-1}$, $m_{\text{ATP}} = 8.4$ mmol ATP (g DW)$^{-1}$ h^{-1}, and $\mu_{max} = 0.7$ h^{-1}, amounts to 51 mmol ATP (g DW)$^{-1}$ h^{-1} [125, 126]. The upper limit, r_{max}, was chosen as the point at which all responsive reactions responded with a constant increase to

the imposed ATPase flux (determined empirically as 350 mmol ATP $(\text{g DW})^{-1}$ h^{-1}). The set of flux solutions of the responsive reactions was used to construct a correlation tree, i.e. a dendrogram depicting similarities in flux response quantified as the magnitude of the Pearson's correlation coefficient between the set of flux values of all pairs of reactions [94].

Flux correlation analysis

The set of LP solutions obtained using Equation 3.2 in Section 3.2.3 were collected in a matrix, $\bar{\mathbf{S}}_{p,m}$, where each of the m columns corresponds to a reaction in the model, and each of the p rows corresponds to an LP solution, as such:

$$\bar{\mathbf{S}}_{p,m} = \begin{matrix} v_{1,J_{min}} & v_{2,J_{min}} & \cdots & v_{m,J_{min}} \\ \vdots & \vdots & \ddots & \vdots \\ v_{1,J_{max}} & v_{2,J_{max}} & \cdots & v_{m,J_{max}} \end{matrix} \tag{3.4}$$

$\bar{\mathbf{S}}$ was refined into matrix $\hat{\mathbf{S}}$ by removing reactions that did not respond to the increase in ATP demand, i.e. columns where all elements are identical were removed, or:

$$\hat{\mathbf{S}} = \{\bar{\mathbf{S}} \mid \bar{\mathbf{S}}_{1...\hat{n}...i,j} \neq \bar{\mathbf{S}}_{1...\hat{n}+1...i,j}\} \tag{3.5}$$

$$(\exists \hat{n} \in \{1,\ldots,p-1\}; \; i \in \{1,\ldots,p\}; \; j \in \{1,\ldots,m\}).$$

thus,

$$\hat{\mathbf{S}}_{p,l} = \begin{matrix} v_{x,J_{min}} & \cdots & v_{z,J_{min}} \\ \vdots & \ddots & \vdots \\ v_{x,J_{max}} & \cdots & v_{z,J_{max}} \end{matrix} \tag{3.6}$$

where $l < m$ and $1 \leq x, z \leq l$.

The flux correlation between the reactions in the $\hat{\mathbf{S}}$ matrix was quantified as the Pearson correlation coefficient between every combination of column in the matrix. The correlation coefficient was defined as:

$$r(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{3.7}$$

Equation 3.7 was applied to all combinations of column-vectors in matrix $\hat{\mathbf{S}}$, generating the square matrix $\boldsymbol{\Delta}$:

$$\mathbf{\Delta}_{l,l} = \begin{matrix} r\left(\hat{\mathbf{S}}_{1...p,x}, \hat{\mathbf{S}}_{1...p,x}\right) & \cdots & r\left(\hat{\mathbf{S}}_{1...p,x}, \hat{\mathbf{S}}_{1...p,z}\right) \\ \vdots & \ddots & \vdots \\ r\left(\hat{\mathbf{S}}_{1...p,z}, \hat{\mathbf{S}}_{1...p,x}\right) & \cdots & r\left(\hat{\mathbf{S}}_{1...p,z}, \hat{\mathbf{S}}_{1...p,z}\right) \end{matrix} \qquad (3.8)$$

As indicated in the definition, the diagonal of $\mathbf{\Delta}$ will only contain ones and the matrix is symmetric around the diagonal. The metabolic tree was constructed as described in Chapter 2, Section 2.2.

Core model extraction and analysis

All reactions that displayed a change in simulated flux as a response to the increase in ATP demand, i.e. the reactions represented by the columns of matrix $\hat{\mathbf{S}}$ defined in Equation 3.6, Section 3.2.3, were assembled into a new core model. The reactions of this core model included transporters that exhibited flux change, thus defining the external metabolites of the model as glucose, oxygen and carbon-dioxide. The data generated using Equation 3.2 was also used for calculating fixed rates of metabolite transfer between the core network and the global network. These values were determined by balancing the net rate of change of each metabolite in the core model for each level of fixed ATP demand, using the stoichiometry of the core network and the reaction rates determined in Equation 3.2. In other words, the matrix of metabolite balances, $\mathbf{E}$, in the core model, over the set of ATPase flux values investigated was obtained by multiplying the stoichiometry matrix of the core model, $\mathbf{N}'$, with an LP solution obtained with Equation 3.2, i.e. the transpose of any row-vector of matrix $\hat{\mathbf{S}}$, $\mathbf{v}$:

$$\mathbf{E} = \mathbf{N}'_{f,l} \cdot \mathbf{v^T} \qquad (3.9)$$

Where f is the number of metabolites in the core model, and l the number of reactions in the core model. If the net flux of a given metabolite amounted to zero, that metabolite was balanced within the sub-network, a negative net flux indicated export from the sub-network, and a positive net flux import to the sub-network.

Model condensation

In order to simplify the structure of the core model, it was condensed by replacing model reactions with the net-reactions of the corresponding enzyme subsets. The formal description in this section is based on existing methods available within the ScrumPy-evironment [47]. The procedure is based on expressing the enzyme subsets of the core model as a matrix, $\mathbf{S}_{m,d}$ (defined in Equation 3.11), where each of the d columns represents a subset (or a reaction

name if the reaction forms a singleton subset), each of the m rows represents a reaction, each element is the normalised flux ratio of the subsets (as described in Chapter 2, Section 2.2), and all structurally dead reactions and inconsistent subsets have been omitted. In order to construct $\mathbf{S}$, the intermediate square matrix $\mathbf{D}_{m,m}$ is defined as the matrix of all reaction flux ratios:

$$\mathbf{D}_{m,m} \quad = [d_{i,j}]_{m,m} \tag{3.10}$$

where $d_{i,j}$ is defined in Equation 2.5.

$\mathbf{S}$ can now be defined as:

$$\mathbf{S}_{m,d} = \{\mathbf{D}_{m,m} \,|\, \mathbf{D}_{1...m,a} \neq \mathbf{D}_{1...m,b}\} \tag{3.11}$$

$$(\forall a, b \in \{1, \ldots, m\})$$

i.e. $\mathbf{S}$ is the subset of unique column-vectors of $\mathbf{D}$. It follows from the symmetri of $\mathbf{D}$ that each row-vector in $\mathbf{S}$ has only one non-zero element, since a reaction can only be in one subset, but each column-vector can have one non-zero element (if the subset is singleton), or several (if the size of the subset is greater than one).

Condensation of the core model was carried out by recursively multiplying the stoichiometry matrix with the subset matrix ($\mathbf{S}$) and removing isostoichiometrical reactions until no further condensation could be achieved, as outlined in Algorithm 1. A simpel example of model condensation is shown in Fig. 3.1.

Algorithm 1 Minimal algorithm for condensing a given stoichiometry matrix $\mathbf{N}$ by replacing sets of reactions with enzyme subsets. The auxiliary function *delete_isoforms($\mathbf{N}$)* identifies all isostoichiometric reactions and for any given set of redundant reactions deletes all but one from $\mathbf{N}$, it returns a list of isostoichiometric reactions. The function *subsets_matrix($\mathbf{N}$)* returns the enzyme subsets of $\mathbf{N}$ expressed as a matrix (as described in Equations 3.10 and 3.11 *delete_isoforms()*).

---

   function CONDENSE($\mathbf{N}$)
      $iso := delete\_isoforms(\mathbf{N})$
      $\mathbf{S} := subsets\_matrix(\mathbf{N})$
      $\mathbf{N} := \mathbf{N} \cdot \mathbf{S}$
      $CanCondense := length(isos) > 0$ **or** $length(\mathbf{S}_{1...m,d}) > length(\mathbf{S}_{m,1...d})$
      $CanCondense := CanCondense$ **and** $length(\mathbf{N}_{1...n,m}) > 0$
      if $CanCondense$ **then**
         $Condense(\mathbf{N})$
      end if
      return $\mathbf{N}, isos$
   end function

$$\mathbf{K}=\begin{bmatrix} -1 & 0 \\ -1 & 1 \\ -1 & 1 \\ 0 & -1 \\ 0 & -1 \\ -1 & 0 \\ -1 & 0 \end{bmatrix} \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \\ E\_tx \\ A\_tx \end{matrix}$$

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \\ E\_tx \\ A\_tx \end{matrix}$$

$$\quad R_1 \; R_2 \; R_3 \; R_4 \; R_5 \; E\_tx \; A\_tx$$

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \\ E\_tx \\ A\_tx \end{matrix}$$
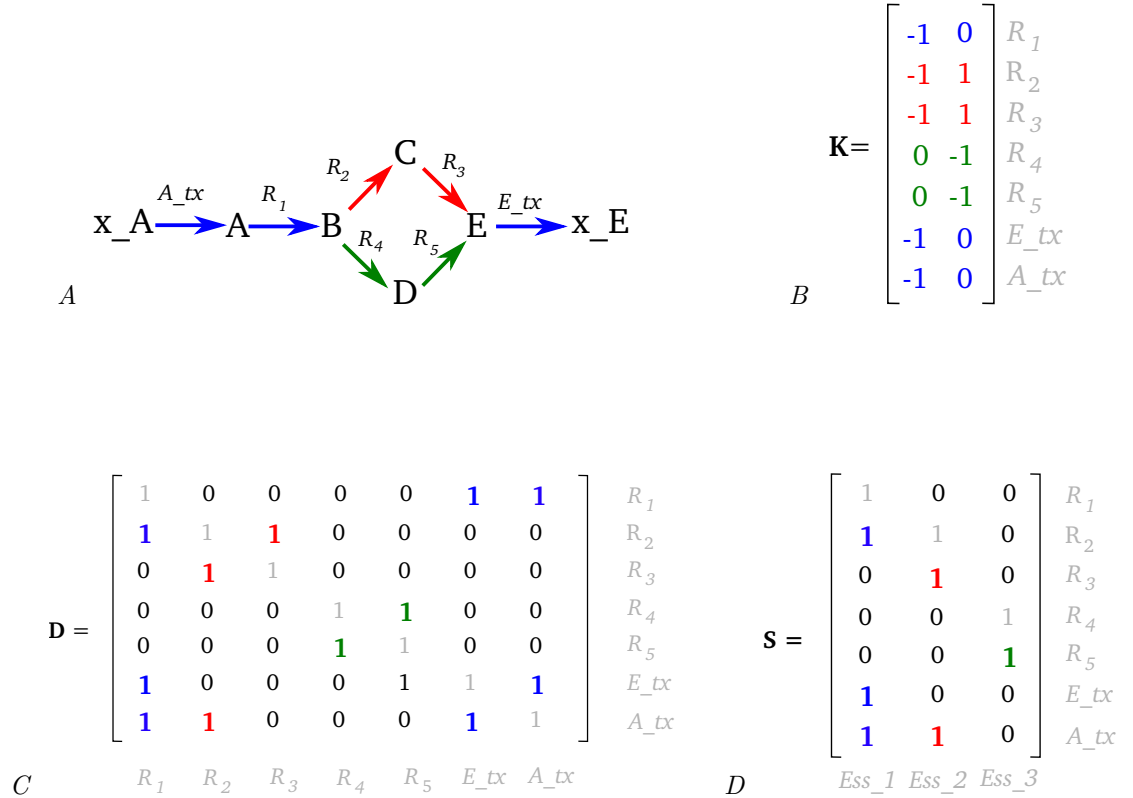
$$Ess\_1 \; Ess\_2 \; Ess\_3$$

Figure 3.1: Illustration of the model condensation procedure. Colours of reaction arrowheads and matrix elements indicate enzyme subset memebrship of a given reaction. **A**: Network diagram of model before condensation. **B**: Kernel matrix of model in A. The kernel indicate that steady state can be achieved by either carrying flux from external metabolite E to external metabolite A through R_2 and R_3 or by running a futile cycle involving R_2, R_3, R_4, and R_5 (by liear combination of these two solutions, a solution involving R_4, and R_5 instead of R_2 and R_3 can be obtained). Note that the kernel only give prototype solutions that are not necessarily thermodynamically feasible, i.e. in this case only the reverse of solution one is feasible. Enzyme subsets can be obtained from the kernel as the reactions with proportional row–vectors. **C**: Matrix **D** is obtained by applying Equation 3.10 to the kernel matrix. **D**: Matrix S is obtained as the set of unique column–vectors in matrix **D** (by applying Equation 3.11). By applying Algorithm 1 to the network, the net reaction is the conversion of external A to external E (with no internal metabolites).

Table 3.1: Model properties. Number of reactions and metabolites of the different modules constituting the model. Since metabolites can occur in more than one module the total number of metabolites is less than the sum of the number of metabolites in each module. Only genes whose gene-product are involved in at least one reaction are included in the enumeration. Note that the number of genes ultimately encoding reactions in the model is an underestimate, gene–associations are not established for all reactions. During the curation process 611 reactions and 374 metabolites defined in the database were excluded from the model.

Module	Reactions	Metabolites	Genes
S. Typhimurium PGDB	1623	1385	1,439
Automatically generated	1012	1011	-
Transporters	70	69	-
Input	*5*	*7*	-
Output	*65*	*64*	-
Electron transport chain	5	11	-
Additional reactions	74	147	-
Total	1161	1058	796

3.3 Results

3.3.1 General model properties

The final model had the characteristics summarised in Table 3.1. LP analysis showed that all biomass components could be produced in biologically relevant proportions from minimal media components, employing 318 reactions (including transporters) of the 1161 reactions, ultimately encoded by 268 genes. The capability to produce biomass precursors was retained under simulated variations in ATP demand. Approximately half of the reactions in the network, 510 out of 1161, were identified as "dead" (508 reactions were associated with zero-vector rows in the null-space matrix, in addition correlation coefficient analysis, described in [48], revealed two additional dead reactions), meaning that the reactions were unable to carry flux under the given media composition (glucose minimal media under aerobic conditions).

3.3.2 Analysis of model inconsistencies

Analysis of unconserved metabolites in the four *Salmonella* models considered here shows the two models published first (iRR1083 [119] and iMA945 [120]) did not have any unconserved metabolites, whereas for the consensus model [121] this number was 430 metabolites. For the model presented here, 10 unconserved metabolites were identified (which could be reduced to 7 by removing

Table 3.2: Metabolite import and export of core network. The table shows the constants fluxes of import and export of metabolites between the core and the main model during variation of ATPase flux. The number of involved reactions refers to reactions in the main model only. Metabolites that were transported with a flux below 1.00 were omitted from the table. Flux units are in mmol (g DW)$^{-1}$ h^{-1}.

Import			Export		
Metabolite	Flux	Reactions	Metabolite	Flux	Reactions
Conserved moieties					
ADP	27.00	33	ATP	27.00	48
NADP	10.38	26	NADPH	10.38	25
NADH	8.70	14	NAD	8.70	16
P_i	40.58	37	GLT	6.90	21
NH_3	11.39	12	PYR	8.57	10
αKG	7.70	10	PEP	3.95	6
CO_2	4.06	20	3-PGA	2.64	1
OAA	2.97	2	ASP	2.42	9
CoA	2.49	9	GAP	4.90	3
			AcCoA	1.76	7

compartment-specific metabolites). These compounds were limited to species that only consisted of hydrogen and oxygen, namely: proton, oxygen, water, superoxide, hydrogen peroxide, molecular hydrogen (and electron).

3.3.3 Model response to changing ATP demand

Of the 318 reactions required for synthesis of biomass, a total of 54 reactions responded to the imposed ATP demand variation. This set of reactions was reduced further to 34 by removing reactions where the total change in flux response was below a treshold of 0.01 mmol (g DW)$^{-1}$h^{-1}.

The remaining reactions formed a connected network, which to a large extent coincided with glycolysis, TCA, and the Entner-Doudoroff pathway (Fig. 3.2).

Since the ATP variation analysis involved increasing the ATP demand, whilst keeping the rate of biomass precursor synthesis constant, the flux through the core model could be divided into a variable and a constant component, as described in Section 3.2.3. The constant fluxes of metabolite transfer between the core network and the genome-scale network are summarized in Table 3.2.

Condensation of the core model, i.e. replacement of reactions with enzyme subsets as described in Section 3.2.3, resulted in a model consisting of six subset and five original reactions (Fig. 3.3). The stoichiometries of the enzyme subsets are summarized in Table 3.3.

The flux correlations between the reactions of the core obtained using Equation 3.8 from the LP data with the full model, were visualised using a metabolic
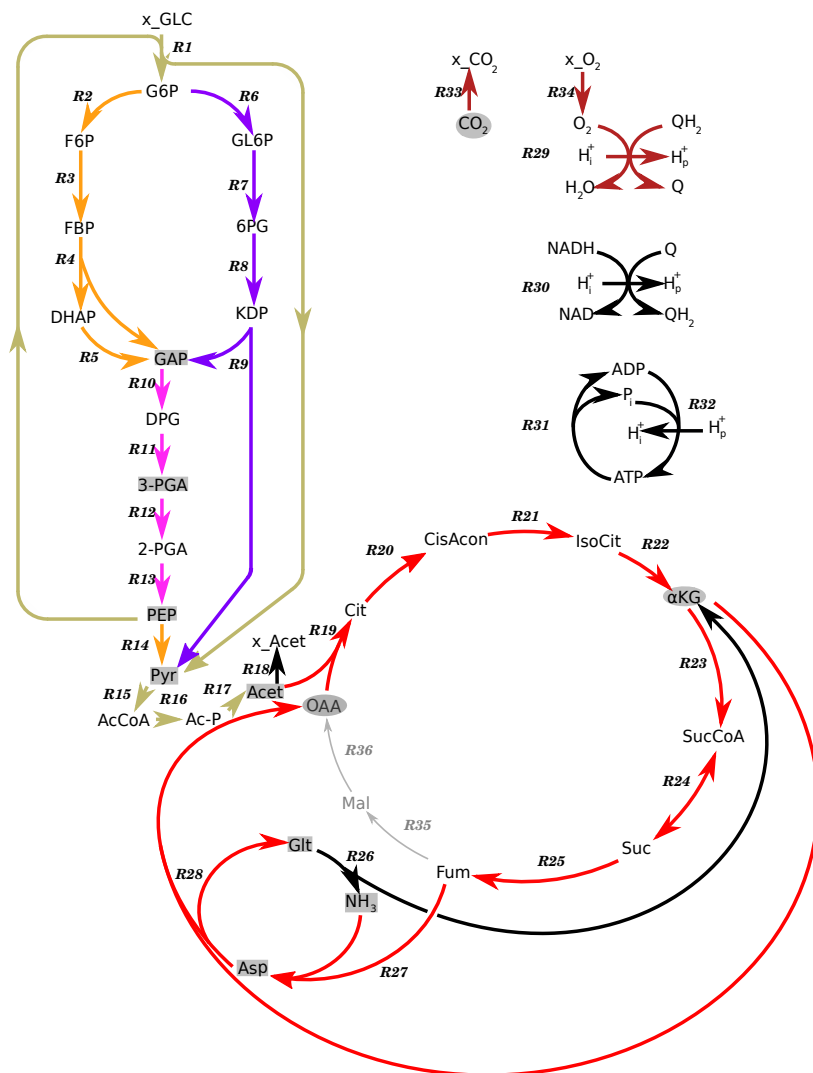
Figure 3.2: The core network. Colours represent different enzyme subsets. The enzyme subsets of the model are indicated by different colours. Grey boxes (circles) represent import (export) between the GSM and the core network, in accordance with the data in Table 3.2. The grayed reactions (R35 and R36) were not part of the core network, but have been included since they form part of the canonical TCA. The numbers refer to the abbreviated names of the reactions (Table 1), accordingly: R1 - glucose transporter, R2 - PGIsomerase, R3 - PFK, R4 - FBPAldolase, R5 - TriPIsomerase, R6 - G6PDH, R7 - PGLactonase, R8 - PGlucDehydr, R9 - KDPAldolase, R10 - GapDH, R11 - PGKin, R12 - PGAM, R13 - Enolase, R14 - PyrKin, R15 - PyrDH, R16 - PhoAcTrans, R17 - AcetKin, R18 - acetate transporter, R19 - CitLyase, R20 - AconDehydr, R21 - AconHydr, R22 - IsoCitDH, R23 - 2-KGDH , R24 - SCoASynth, R25 - SucDH, R26 - GluDH-(NAD(P)), R27 - Aspartase, R28 - AspTrans, R29 - CytOx, R30 - NADH_DH, R31 - ATPase, R32 - ATPSynth, R33 - CO_2 transporter, R34 - O_2 transporter, R35 - FumHydr, R36 - MalDH.
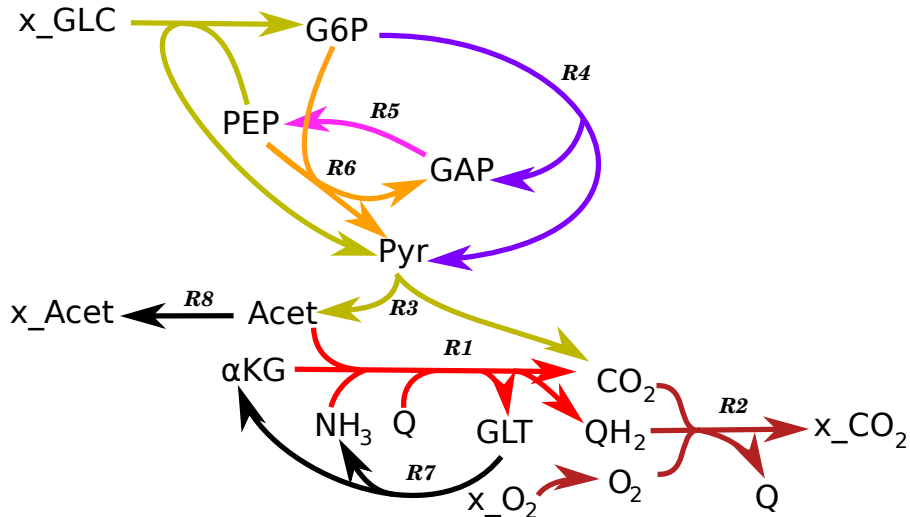
Figure 3.3: Structure of the condensed core network, based on the network in Fig. 3.2. Numbers refer to names of reactions and enzyme subset (see Table 1 and Table 3.3), accordingly: R1 - Ess_1, R2 - Ess_2, R3 - Ess_3, R4 - Ess_4, R5 - Ess_5, R6 - Ess_6, R7 - GluDH-(NAD(P)), R8 - acetate transporter.

Table 3.3: Stoichiometry of enzyme subsets generated from the core model. For each enzyme subset of the core network the condensed net stoichiometry is shown. External metabolites are indicated by the prefix "x_".

Enzyme subset	Stoichiometry		
	Reactants		Products
Ess_1	$NAD + ADP + Q + \alpha KG + NADP + NH_3 + Acet$	$\rightarrow$	$NADH + 2\ CO_2 + GLT + NADPH + ATP + QH_2$
Ess_2	$x\_O_2 + 2\ QH_2 + CO_2$	$\rightarrow$	$2\ Q + x\_CO_2$
Ess_3	$PEP + 2\ P_i + 2\ NAD + 2\ ADP + x\_Glc + Pyr$	$\rightarrow$	$2\ Acet + 2\ NADH + 2\ CO_2 + 2\ ATP + G6P$
Ess_4	$NADP + G6P$	$\rightarrow$	$Pyr + GAP + NADPH$
Ess_5	$P_i + ADP + NAD + GAP$	$\leftrightarrow$	$PEP + NADH + ATP$
Ess_6	$G6P + PEP$	$\rightarrow$	$Pyr + 2\ GAP$

48

tree (Fig. 3.4). Reactions that occur in the same enzyme subset in the core model are highly correlated in flux response, as expected. The agreement between null-space and and LP based correlation is not, however, absolute. This deviation is due to the omission of constant import and export transporters from the structural analysis of the core model. The flux responses of the core reactions to increasing ATP demand fell into two broad categories - (i) reactions that increased monotonically in magnitude with ATPase and (ii) reactions that eventually reached a plateau and maintained constant flux for ATP demands above a certain threshold. These two categories can be clearly distinguished as the two main clusters in the metabolic tree, where the reactions that continually increse include the reactions of enzyme subsets 2, 3, 5 and 6, the acetate transporter and the electron transport chain reactions. The reactions that reached a constant flux value with increasing ATP demand included the Entner-Doudoroff (subset 4) and TCA pathways (subset 1). The difference between the two categories of reactions is also pronounced in Fig. 3.5, where the reactions of the first category (i.e. those always increased with ATP demand) are found in Fig. 3.5.A – 3.5.C, whereas reactions from the second category are found in Fig. 3.5.B and 3.5.D. The elementary modes of the core model (Fig. 3.6) highlight the simplicity of the model: despite the seemingly complex structure of Fig. 3.2 there are only four elementary modes. Each mode involves import of glucose (and the remaining reactions in subset 3) and oxygen, hydrolysis and regeneration of ATP, export of CO_2, and lower glycolysis. Flux can pass through either upper glycolysis (subset 6) or Entner-Doudoroff (subset 4), which represents two separate modes. Likewise, after passing either subset 4 or 5, flux can either exit the system through the acetate transporter after passing subset 3, or enter TCA (subset 1). Thus, the redundancy between the acetate and the TCA route forms the additional two modes of the core model. The flux response of the core reactions can be seen as a shift away from primarily mode 1 towards mode 4, i.e. a shift from a subsets 1 and 4, which both generate NADPH, towards subset 6, which indirectly has a higher energy yield than subset 4, and the acetate transporter, which has a lower flux-cost than subset 1. Subset 6 has a higher energy yield because all the flux is directed towards subset 5 (lower glycolysis), which generates one unit of ATP and one NADH, unlike subset 4 which only directs half the flux to subset 5 and half goes directly to subset 3. In other words, the relative importance of the NADPH-generating subsets decreases as the energy demand increases.
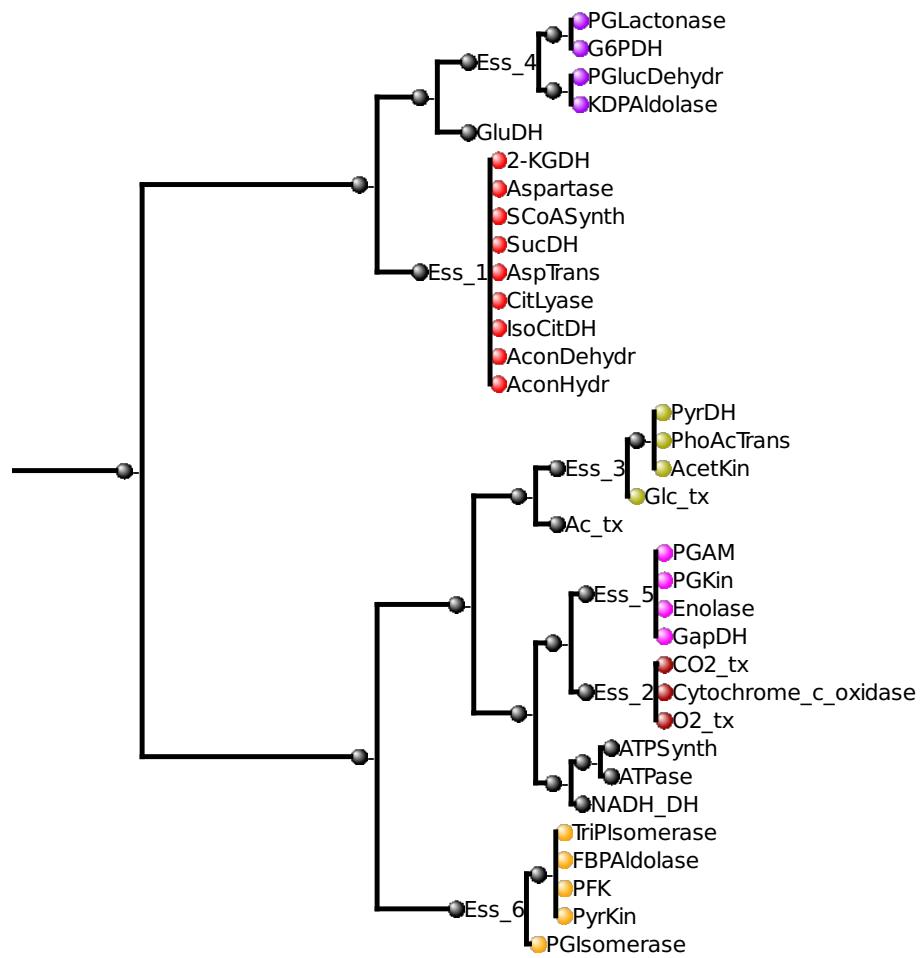
Figure 3.4: Flux correlation tree of the reactions of the core network. The tree display correlations in flux response to the imposed ATP demand between reactions in the core model. Enzyme subset identity of the reactions are indicated by colour codes, corresponding to that used in Fig. 3.2.
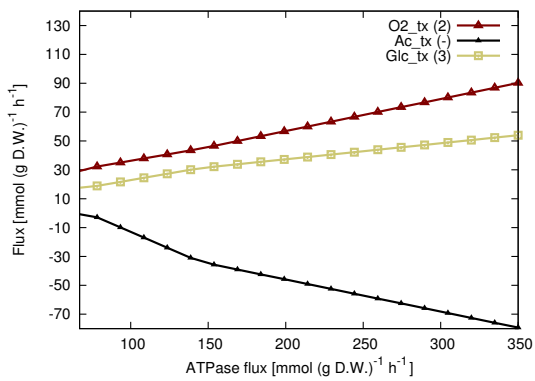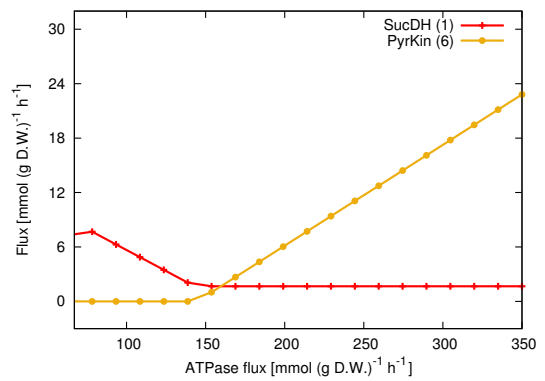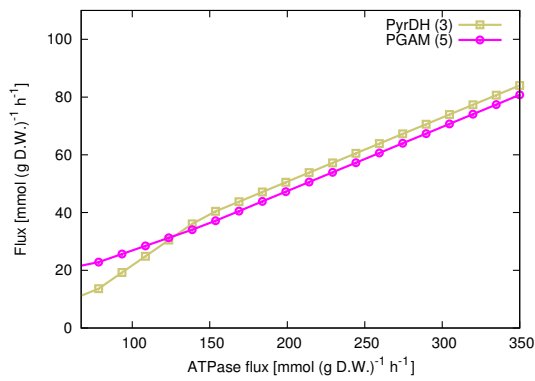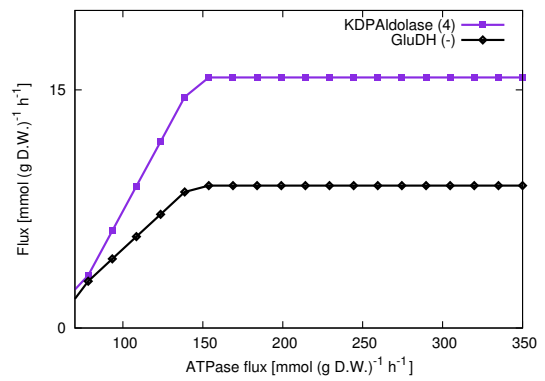
Figure 3.5: Responses of individual reactions (one per subset, except for subset 6, which is represented twice) to varying ATP demand. number in brackets refer to subset number. A: Response of glucose transporter (Ess_3), oxygen transporter and acetate transporter (export). B: Response of SucDH (Ess_1) and PyrKin (Ess_6). C: Response of PyrDH (Ess_3) and PGAM (Ess_5). D: Response of KDPAldolase (Ess_4) and GluDH.
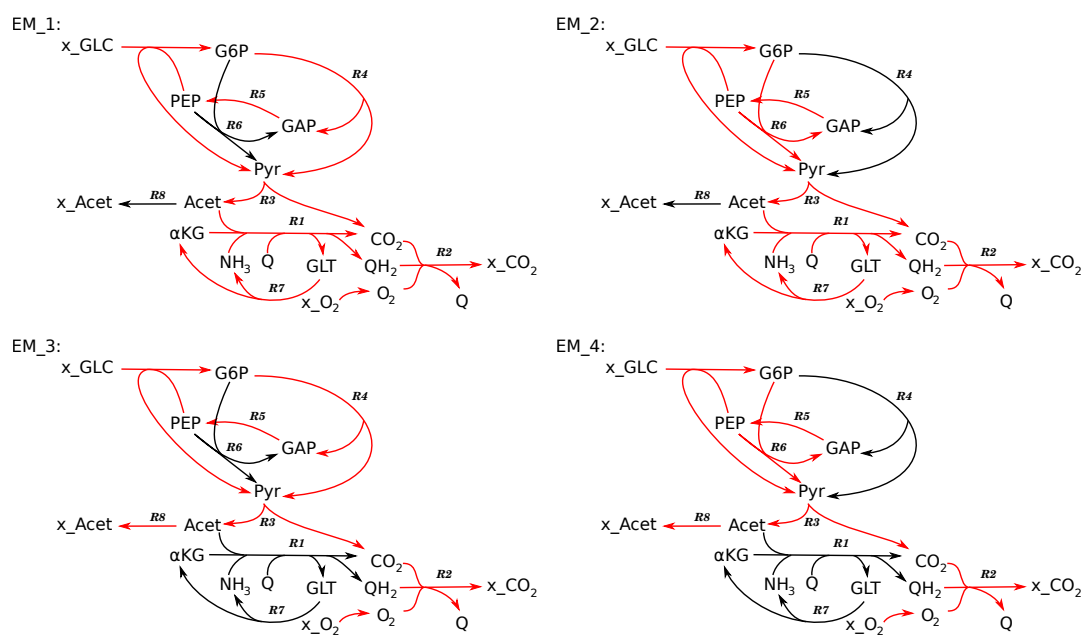
Figure 3.6: The elementary modes of the condensed core model. The elementary modes of the core model are displayed using the condensed network of the core model (Fig. 3.3). For each elementary mode the active reactions are in red (inactive are in black). Reaction identifiers have the same meaning as in Fig. 3.3.

3.4 Discussion

3.4.1 Model construction and characteristics

As indicated in Table 3.1, iterative refinement of the initial model based on the
S. Typhimurium PGDB into the GSM presented here involved removing 611
reactions and 374 metabolites from the final model. This curation was mainly
aimed at atomically unbalanced reactions and reactions involving atomically un-
defined metabolites. The former type of reaction is an obvious error, and should
be replaced with correct stoichiometry, whereas the latter type of reactions are
not necessarily erroneous; they are however unsuited for GSMs. The approach
used here is based on identifying undefined metabolites and where necessary, i.e.
when the metabolite in question represents a metabolically relevant compound
with undefined composition, replace the metabolite with defined instances of the
metabolite. An example of this is the replacement of the generic and undefined
metabolite "NTP" with the four nucleoside triphosphates (and thus replacing
one reaction with four reactions).

Apart from the risk of including atomically unbalanced reactions when equat-
ing PGDBs with genome-scale models, a major remaining obstacle is the correct
assignment of reaction reversibility. As described in Sections 3.2.2 and 3.2.2 the
method of identification of candidate reactions for re-definition of reversibility
used here is based on analysis of enzyme subsets and LP. The rational for this
approach is mainly practical. Most reactions in a metabolic network involved
with transforming substrate into biomass are assumed to primarily carry flux in
one direction, reactions that will inhibit this function due to their irreversibility
will be identified with the methods used here.

The size of the model can be compared to GSMs constructed using the same
methodology as used here, e.g. a *Streptococcus agalactiae* 2603V/R GSM with
631 reactions [127] based on a genome of 2.16 Mb encoding 2,175 genes [128] the
A. thaliana GSM in [94] had 1,406 reactions based on a 157 Mb genome [129]
with at least 27,000 genes [130]. The corresponding numbers for the *S.* Ty-
phimurium GSM presented here are 1161 reactions based on a 4.857 Mb genome
containing 4,489 protein-coding genes [122].

3.4.2 Comparison with other models

As mentioned, three models of *S.* Typhimurium metabolism have been con-
structed [119–121]. However the first two (iRR1083 [119] and iMA945 [120])
were constructed by modifying a pre-existing *E. coli* model (iAF1260) [126], i.e.
genetic differences, in terms of presence or absence of genes encoding metabolic
enzymes, between *E. coli* and *S.* Typhimurium were identified and the relevant

reactions subsequently removed from or added to the *S*. Typhimurium model. A potential problem with this approach is that it might underestimate phenotypical differences between *E. coli* and *S*. Typhimurium since genes are analysed in isolation. This is different from the approach used for construction of PGDBs, where the set of predicted enzymatic reactions of the organism, based on genome sequence alone, is compared to a library of "canonical" pathways, so that if a critical number of reactions in a given pathway is identified in the set of organism-specific reactions, this pathway (and all reactions) is included in the PGDB, even if all reactions in the pathway are not identified [123]. The benefit of this approach is that reactions that are spontaneous or where there is insufficient knowledge of the enzymatic basis of an observed reaction will not be omitted from the model. From a model construction point-of-view this is beneficial since the number of dead reactions will be minimised. The drawback is the risk of overestimating the metabolic capacity of the organism. Given that the reference model (iAF1260) is free from unconserved metabolites, any model constructed by appending or removing individual reactions (provided that the added reactions are consistent) will also be free from unconserved metabolites. Hence, the consistency of iRR1083 and iMA945 cannot be considered without taking their similarity to iAF1260 into account.

Unlike iRR1083 and iMA945, the *Salmonella* consensus model, which is based on iRR1083 and a previously unpublished *Salmonella* model (constructed using the same methodology as that used for iRR1083 [121]), contained 430 unconserved metabolites. So, although more detailed and larger than the previous models (2,201 reactions were included in the consensus model, 1,079 in iRR1083 and 1,964 in iMA945), this model suffers from the drawback that more than a third of the metabolites are inconsistent. As reviewed in Chapter 2 (Section 2.4), a consequence of these inconsistencies is the potential for flux distributions that violates the law of mass conservation, i.e. the appearance (or disappearance) of material from nothing. This is a serious modelling error that is hard (if not impossible) to detect, but can be prevented by using a consistent model. For the reasons presented in this section, the model presented here was constructed *de novo* from the *S*. Typhimurium genome and thoroughly curated to ensure stoichiometric consistency.

3.4.3   ATP demand analysis

The core model identified from the variation of ATP demand shared a considerable number of reactions with components of central carbon metabolism - e.g. glycolysis, TCA cycle and the electron transport chain. Three groups of reactions, however, are less expected: The reactions of subset 4, subset 3, and citrate

lyase, aspartate ammonia-lyase, and aspartate aminotransferase from subset 1.

The reactions of subset 4 (glucose-6-phosphate-1-dehydrogenase, 6-phosphogluconolactonase, phosphogluconate dehydratase and KDP aldolase) are considered components of the oxidative pentose phosphate pathway (glucose-6-phosphate-1-dehydrogenase and 6-phosphogluconolactonase) and the Entner-Doudoroff pathway (phosphogluconate dehydratase and KDP aldolase). The Entner-Doudoroff pathway is primarily associated with growth on sugar acids (e.g. gluconate, glucuronate and galacturonate) [131] which are metabolised to 6-P-gluconate. During growth on glucose, glucose-6-P is converted to 6-P-gluconate by the OPPP component of subset 4 (i.e. glucose-6-P-1-dehydrogenase and 6-phosphogluconolactonase). The response of the reactions in subset 4 to increasing ATP demand shows that the flux through the subset increased until reaching a constant value. This switch is coordinated with the activation of the reactions of subset 6 (upper glycolysis). As mentioned (Section 3.3.3), one interpretation of this observation is that it is part of an overall switch from elementary modes in the core network that generate NADPH to modes that generate NADH and ATP, i.e. a switch from a primarily biosynthetic flux distribution to one with a higher energy yield. This observation would suggest that an alternative function of the Entner-Doudoroff pathway, apart from its role in growth on sugar acids, is to offer an NADPH-generating alternative to upper glycolysis (subset 6), without involving all the reactions of the oxidative pentose pathway. This hypothesis is consistent with the observation that *E. coli* growing at maximum rate distributes a minor proportion (2%) of carbon flux through the Entner-Doudoroff pathway [132]. Similarly, Peng and Shimizu [133] showed, using 2-D protein electrophoresis applied to *E. coli* cultures, that all the enzymes of subset 4 were expressed under all conditions investigated. Actually, apart from expression profiles from aerobic gluconate medium (which resulted in high expression of Entner-Doudoroff enzymes and slightly lower OPPP expression) the highest expression of the enzymes in subset 6, especially the Entner-Doudoroff enzymes, was obtained from aerobic glucose medium.

A similar response pattern to that of the reactions in subset 4 was observed in an *A. thaliana* genome-scale model [94], where flux through the oxidative component of the oxidative pentose pathway decreased until being turned off in response to increasing ATP demand. As in the analysis presented here, a possible interpretation of this is that the redox energy is shifted from NADPH to NADH as the demand for ATP is increased.

Subset 3 involves the glucose transporter (via phosphotransferase system (PTS)), pyruvate dehydrogenase and the reactions of the acetate dissimilatory pathway (phosphate acetyltransferase and acetate kinase). The set is closely associated with, but does not include, the acetate exporter. The generation of

acetate is commonly associated with anaerobic metabolism in bacteria. However, *E. coli* has been shown to produce large amounts of acetate when grown aerobically on glucose. The function of this phenomenon is unclear. Traditionally, it has been attributed to wasteful "overflow" metabolism, however, mutation studies has shown that acetate dissimilation allows high growth rates and cell densities [134]. El-Masi [134] suggested that acetate formation serves the purpose of supplying ATP to the rest of metabolism and coenzyme-A to 2-oxoglutarate dehydrogenase in the TCA. The analysis presented here suggests that an optimal balance between ATP generation and flux minimisation (i.e. minimisation of protein investment of the cell), whilst synthesising biomass, can be achieved by diverting a significant proportion of the central carbon flux into acetate excretion. This is especially pronounced for high energy demand levels. Inspection of the elementary modes of the core model shows that the flux distribution that is obtained at high energy demand is not the one with the highest ATP yield, which is the mode that includes both TCA and glycolysis, but not Entner-Doudoroff and acetate excretion (mode 2 in Fig. 3.6). In contrast, the analysis suggests that the TCA primarily serves a biosynthetic rôle even at high energy demand, since the response curves reaches a constant value with increasing energy demand. This is consistent with the observation that high glucose concentrations repress expression of key TCA enzymes (succinate dehydrogenase, succinyl-CoA synthetase and 2-oxoglutarate dehydrogenase) under aerobic conditions [133, 135, 136]. This observation has been explained as an effect of oxygen limitation, i.e. the situation where the glucose availability exceeds the oxygen availability [136], the analysis presented here however shows that this mode of operation is theoretically optimal even if there are no limitations on glucose and oxygen uptake rates.

Another set of reactions not commonly associated with energy metabolism are citrate lyase, aspartate ammonia-lyase, and aspartate aminotransferase from subset 1 and the singleton subset glutamate dehydrogenase, which displays a similar flux response as subset 1. Citrate lyase is primarily considered to be an anaerobic enzyme that facilitates growth on citrate [137], although the failure to delete the genes encoding the enzyme would suggest a more essential function *in vivo* [138]. It has also been linked to survival of *S.* Typhimurium [139] and *Y. pestis* [140] in macrophages. It has been shown to be reversible [141] In the work presented here it appears as if the function of citrate lyase in relation to energy metabolism is closely associated with acetate dissimilation: when all acetyl-CoA is channelled towards acetate (thereby generating one ATP per acetate produced) no acetyl-CoA can enter TCA directly through citrate synthase. By synthesising citrate from acetate and oxaloacetate the pool of acetate that does not get excreted goes into TCA. This potential rôle of citrate lyase is not

supported by direct experimental evidence, but can not be ruled out. The set of reactions with the most complex topology involves aspartate ammonia-lyase, and aspartate aminotransferase and glutamate dehydrogenase. They are not part of a single biosynthetic pathway: in *E. coli* aspartate ammonia-lyase is primarily associated with glutamate degradation [142], aspartate transaminase is associated with both glutamate degradation and aspartate biosynthesis [143], and glutamate dehydrogenase is primarily associated with glutamate biosynthesis [144]. Although this set has the most complex topology of the subsets in the catalytic core, the net stoichiometry is simple:

$$\text{Fum} + \text{NADP} \leftrightarrow \text{OAA} + \text{NADPH}$$
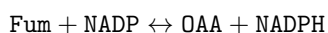
and therefore acts to bypass the fumarate and malate dehydrogenase reactions of the TCA cycle, again suggesting a possible rôle of the set of reactions that would be difficult to identify without the use of a metabolic model.

As a consequence of the condensation, some structural features of the core model are highlighted, primarily the association of pyruvate kinase to subset 6 (upper part of glycolysis, Fig. 3.2). This feature is a structural consequences of the inclusion of subset 4 and the acetate transporter, since in the absence of these subsets the network would form one single subset. The association of pyruvate kinase with the reactions of upper glycolysis is supported by experimental observations in *E. coli*: In the study by Siddique *et al.* [145] a pyruvate kinase isoenzyme (pyruvate kinase I, encoded by *pykF*) was deleted in *E. coli* and cultures grown on glucose minimal medium were subjected to transcriptional and metabolic flux analysis. The results showed that there was a correlation in flux decrease between pyruvate kinase (21% of wild type, measured as flux through pyruvate kinase II compared to wild type pyruvate kinase I) and some of the reactions of upper glycolysis, 6-phosphofructokinase (13.4% of wild type), triosephosphate isomerase (83% of wild type), and glucose-6-phosphate isomerase (13.4% of wild type). The flux through fructose-bisphosphate aldolase, which is the remaining member of ESS_6, was not explicitly reported. This decrease through upper glycolysis was concomitant with increased flux through the first and the third reaction of the pentose phosphate pathway (6-phosphogluconate dehydrogenase (273% of wild type) and glucose-6-phosphate dehydrogenase (273% of wild type)). Since the flux through the TCA cycle was only moderately decreased (97%), it could be assumed that flux from the pentose phosphate pathway to the TCA cycle involved GAP as an intermediate, indicating that lower glycolysis was not severely affected by the deletion of pyruvate kinase II. It was suggested that the cause of the flux decrease through upper glycolysis was due to inhibition of 6-phosphofructokinase by PEP, which

would be assumed to accumulate as an effect of the pyruvate kinase II deletion. These results thus lend support to the suggested association between pyruvate kinase and upper glycolysis.

Chapter 4

Analysis of a *Lactobacillus plantarum* genome-scale metabolic Model

4.1 Introduction

This chapter describes the conversion of an existing genome-scale model of the metabolic network of *L. plantarum* [70], into a version consistent with the modelling conventions of the ScrumPy-software (e.g. a modular model design and consistent definition of transport reaction external metabolites) [47] and compatible with the BioCyc database [123]. The *L. plantarum* model is then subjected to the same analysis as was the *S.* Typhimurium model in Chapter 3; response to ATP demand.

The conversion of the *L. plantarum* GSM to ScrumPy/BioCyc standards is motivated by the analysis in subsequent chapters, where it is used for analysis of metabolic interactions with *S.* Typhimurium, which requires that reaction and metabolite names in the two models are consistent. The original model was constructed using BiGG-nomenclature [146], which is not cross-referenced to BioCyc.

A limitation with FBA often encountered when applying the technique to metabolic models of microbial metabolism capable of both mixed acid and homolactic fermentation, is that the experimentally observed mode of metabolism, homolactic fermentation, has a lower energy yield than the alternative, and thus will not appear in an optimal LP solution. This phenomenon is not limited to lactic acid bacteria, FBA solutions for metabolic models of baker's yeast typically favour aerobic respiration over the experimentally observed combination

of respiration and fermentation [72].

The reason for this discrepancy is believed to be because *L. plantarum* favours growth rate over biomass yield [70]. As reviewed in Chapter 2, Section 2.5, nominal optimisation of growth rate in FBA problems are actually optimisations of growth yield. In order to obtain FBA solutions that are closer to the observed ones one practise has been to constrain the rate of key reactions to experimentally observed values [147]. Similarly, Simeonidis et al. [148] showed that more accurate FBA flux solutions could be obtained for a baker's yeast GSM by incorporating an energy cost proportional to the flux through respiratory reactions, representing the cost of synthesising and maintaining mitochondria.

Here a similar approach is used: a cost is added to output reactions that are not observed *in vivo*, thereby promoting, but not forcing, solutions that are experimentally observed.

4.2   Methods

The main issues covered in this section are: (*i*) reconstruction of the *L. plantarum* GSM [70] based on the BioCyc database; (*ii*) adaptation of FBA to deal with the discrepancy between optimal and experimentally observed flux distributions in *L. plantarum*; (*iii*) effect of FBA formulation (*iv*) on model response to energy demand variation. Since the technique of varying energy demand in order to identify reactions involved in catabolism was introduced in Chapter 3, Section 3.2.3, this will only be covered briefly here.

4.2.1   Modification of a *Lactobacillus plantarum* model

The ScrumPy-based *L. plantarum* was constructed with a similar modular structure to that of *S.* Typhimurium model described in Chapter 3, Section 3.2, where all reactions were ultimately derived from the original *L. plantarum* model. The modules included in the final model were: Automatically generated reactions; Transporters; Electron transport chain to generate proton-motive force; Additional reactions; Fatty-acid biosynthesis.

Computational methods

The *L. plantarum* model [70] was downloaded as a spreadsheet document. All reaction-information was extracted and used for programmatic sorting of the reactions into categories corresponding to the degree of manual curation needed to express the reaction using BioCyc-standards. The following categories were

used: *Matching reactions* - model-reactions whose annotation included EC-number and a common-name that could be uniquely associated with a BioCyc entry; *Transporter* - model-reactions that involved at least one external metabolite; *Non-unique reactions* - reactions that could only be partially matched to BioCyc entries (e.g. alcohol dehydrogenase which can use a multitude of substrates); and *Non-matching reactions* - all reactions that did not fall into any of the above categories (e.g. artificial reactions, such as).

The BioCyc equivalent of all reactions that were uniquely identified were incorporated into the automatically generated module. The transporter module was defined with the reactions identified as transporters (as described above). These reactions fell into two categories - importers of medium components (thus defining the growth medium available to *L. plantarum*) and exporters of metabolic by-products. Exporters of biomass components were manually added to the transporter module (see section below). The fatty-acid biosynthesis, additional reactions and electron transport chain were manually constructed, based on the reactions that did not match any BioCyc entries. The non-unique reactions were incorporated into the automatically generated module when unique, well-defined equivalents could be identified in BioCyc, otherwise they were manually defined in the additional reactions module.

Deviations from original model

The following two reactions with complex stoichiometries in the original model were replaced with sets of simpler reactions. A lumped reaction for biosynthesis of the complex lipid acyl-glycerol-3-phosphate (*L. plantarum* specific), "GAT1_LPL" in the original model, was expressed with BioCyc compound-identifiers and split into fatty-acid specific reactions:

```
0.12 "2-Hexadecenoyl-ACPs" + 0.32 "Octadec-2-enoyl-ACPs"
+ 0.25 c_propyl_octadecan_acp + "GLYCEROL-3P" + 0.26
"2-Hexadecenoyl-ACPs" + 0.02 "Stearoyl-ACPs" + 0.03
"Tetradec-2-enoyl-ACPs" -> "ACP" + acyl_gly_P_LP
```

was replaced with six acyltransferase reactions and one assembly reaction, in order to facilitate curation of the net reaction:

```
0.12 "2-Hexadecenoyl-ACPs"  -> 0.12 acyl1 + 0.12 "ACP"


0.32 "Octadec-2-enoyl-ACPs"  -> 0.32 acyl2 + 0.32 "ACP"


0.26 "2-Hexadecenoyl-ACPs" -> 0.26 acyl3 + 0.26 "ACP"
```

```
0.02 "Stearoyl-ACPs" -> 0.02 acyl4 + 0.02 "ACP"

0.03 "Tetradec-2-enoyl-ACPs" -> 0.03 acyl5 + 0.03 "ACP

0.25 c_propyl_octadecan_acp -> 0.25 acyl5 + 0.25 "ACP"

0.12 acyl1 + 0.32 acyl2 + 0.26 acyl3 + 0.02 acyl4 +
0.03 acyl5 + 0.25 acyl5 + "GLYCEROL-3P" -> acyl_gly_P_LP
```

In these reactions the metabolites `acylX` were used to represent the acyl moiety of an acyl-ACP molecule.

In the original model, biomass synthesis was modelled as a single reaction where each biomass component was expressed as a reactant with the stoichiometric coefficients corresponding to the concentration of the component. Export of certain polymeric biomass components (protein, DNA, and RNA) were expressed using the component metabolites, e.g. the transport of protein was split into 18 tRNA-dependent transporters with the stoichiometry:

```
x_AA + "AA-tRNAs" <> "Charged-AA-tRNAs"
```

The biosynthesis of glutamine- and glutamate-charged tRNAs are interdependent in *L. plantarum*, as can be seen from the stoichiometry of the last step in the pathway, "6.3.5.7-RXN":

```
1.0 GLN + 1.0 ATP + 1.0 GLT-tRNAs <> 1.0 Charged-GLN-tRNAs
+ 1.0 Pi + 1.0 GLT + 1.0 ADP
```

Thus the exporter of either of the amino acids must also export the other. This was modelled using a combined glutamine/glutamate exporter:

```
2.0 "GLT-tRNAs" + x_GLN + x_GLT <> "Charged-GLT-tRNAs" +
"Charged-GLN-tRNAs"
```

Model validation

The consistency of the model was tested using the same tests as described in Chapter 3, Section 3.2.2: Feasibility of synthesis of individual biomass components using FBA, energy and material consistency, identification of unconserved metabolites.

FBA problem formulation

Biomass synthesis was modelled using the same approach as applied to *S.* Typhimurium in Chapter 3 (Equation 3.1):

$$\begin{aligned}
\text{minimize} \quad &: \quad \sum_{i=1}^{m} |v_i| \cdot c_i \\
\text{subject to} \quad &\begin{cases} \mathbf{N}_{n,m} \cdot \mathbf{v} = \mathbf{0} \\ v_j = t_j; \ k \le j \le m \\ v_{\text{ATPase}} = r_{\text{ATP}} \end{cases}
\end{aligned} \qquad (4.1)$$

i.e. the objective was to minimise the sum of total flux through the network (with each reaction carrying a certain flux-cost, c), given the steady state constraint ($N_{n,m} \cdot \mathbf{v} = \mathbf{0}$), the constraint that biomass exporters (v_j) must carry flux fixed to a certain value (t_j), and that maintenance energy (modelled as flux through a generic ATPase), v_{ATPase}, was fixed to a value, r_{ATP}, determined using Equation 4.2:

$$r_{\text{ATP}} = Y_{\text{ATP}} \cdot \mu + m_{\text{ATP}} \qquad (4.2)$$

where the experimentally determined parameters were $Y_{\text{ATP}} = 27.4$ mmol ATP (g DW)$^{-1}$, $m_{\text{ATP}} = 0.36$ mmol ATP (g DW)$^{-1}$ h^{-1}, and $\mu = 0.4$ h^{-1} [70], which resulted in a flux of 11.3 mmol ATP (g DW)$^{-1}$ h^{-1}.

Each biomass exporter was fixed to the value corresponding to the product of the concentration of the respective component and the growth rate.

Equation 4.1 was used in two ways: (i) modelling of mixed acid fermentation by assigning an equal cost, c, of 1.0 for all reactions in the network and (ii) modelling of homolactic fermentation by assigning a 100-fold higher cost to exporters of metabolic by-products other than lactate (CO_2, ethanol, formate, acetate, succinate, and pyruvate).

4.2.2 Model analysis

Model response to ATP demand variation

The *L. plantarum* model was analysed using the same methods of variation of energy demand as described in Chapter 3, Section 3.2.3. Briefly, this involved re-solving Equation 4.1, for each iteration the value of the the fixed maintenance energy (i.e. the flux of the ATPase reaction) was increased by a constant value (i.e. v_{ATPase} was fixed to J_{ATPase} in the range $J_{min} \ldots J_{max}$), essentially as described in Equation 3.2.

The data obtained was analysed as described in Chapter 3, Section 3.2.3: Equations 3.4 – 3.8 were applied to the collected solutions in order to identify reactions that were responsive to the energy demand variation and calculate the Pearson's correlation coefficients of the flux-responses; Equation 3.9 was used for calculating the rates of constant metabolite transport between the catabolic core network and the global model; the responsive reactions were collected into

a separate sub-model, which was condensed into the enzyme subsets using Algorithm 1. The flux correlations were visualised in metabolic trees (as described in Chapter 2, Section 2.2).

Energy demand analysis was applied to mixed acid fermentation (with equal cost, c, for all reactions in the network) and homolactic fermentation (transporters of CO_2, ethanol, formate, acetate, succinate, and pyruvate assigned a 10-fold higher cost than other reactions). For both modes, the analysis was conducted assuming presence and absence of oxygen.

4.3   Results

4.3.1   Model properties

The reconstructed model had 675 reactions (and 568 metabolites), which compares to the 762 reactions (and 658 metabolites) of the original model. The reason for the larger number of reactions in the original model is the difference in how external and internal metabolites are treated in the two models: In the original model all metabolites that can be consumed or produced by the model are associated with an exchange reaction which converts an external metabolite into an internal metabolite, located in the extra-cytosolic compartment. The physically external (but mathematically internal) metabolite can then be imported to the cytosolic compartment by a transport reaction. In the modified model the exchange reactions were omitted and metabolites located in the external compartment treated as external metabolites, thus reducing the total number of reactions (and metabolites). All biomass components could be synthesised individually and in concert. The model was consistent in terms of energy and mass (validated as described in Chapter 3, Section 3.2.2). Analysis of stoichiometric inconsistencies identified five unconserved metabolites: O_2, H_2O, OH^-, H^+, and O_2^-. Since the unconserved metabolites were limited to species only consisting of hydrogen and oxygen the model was considered stoichiometrically consistent.

Of the 675 reactions in the model, 295 were needed for biomass synthesis under homofermentative conditions, 291 under mixed acid conditions.

4.3.2   ATP demand variation

For each of the four conditions for which ATP demand analysis was carried out (presence or absence of oxygen, mixed acid or homolactic fermentation), the total number of responsive reactions was reduced further by removing reactions whose net flux-change was below 0.01 mmol $(g\ DW)^{-1}h^{-1}$. The network-diagrams of the catabolic core models are shown in Figs. 4.1 – 4.3. Under mixed

acid conditions, some pairs of reactions functioned as net-transhydrogenases, i.e. these reactions involved a redox pair (NAD(P)/NAD(P)H) and were isostoichiometric apart from the redox pair. These reactions were represented as a single transhydrogenase reaction in the network diagrams. The size of the catabolic core models varied depending on the mode of fermentation - under homolactic conditions the number of responsive reactions was limited to 14 (including transporters of glucose and lactate, see Fig. 4.1), and the response was insensitive to oxygen availability; under mixed acid conditions, however, the size increased to 29 reactions (including a generic transhydrogenase) utilised under both aerobic (Fig. 4.2) and anaerobic (Fig. 4.3) conditions. The 14 reactions identified under homolactic conditions were limited to uptake and catabolism of glucose to the fermentation by-product lactate, through glycolysis and lactate dehydrogenase. Under mixed acid, anaerobic conditions 29 reactions responded to the imposed ATP demand. In isolation, however, additional reactions were required in order to obtain a steady state solution for the sub-model. Specifically, the regeneration of cofactors (NAD and CoA), which in the global model was carried out by reactions involved in biomass synthesis, was not available in the extracted core model. Steady state solution was obtained by including reactions for ethanol synthesis and export, from the global model. Under aerobic conditions a total of 32 reactions responded to the imposed ATP demand, these included oxygen uptake, menaquinol oxidase, and menaquinol-dependent NADH dehydrogenase. All catabolic core models involved fixed rates of metabolite exchange with the global model, shown in Table 4.1.

Compared to results obtained with the *S.* Typhimurium model (Chapter 3, Section 3.3.3), more metabolites that were in a conservation relationship were transported. Typically, for the conditions used here, two to four transported metabolites and two to five conserved moieties were obtained, compared to six to seven transported metabolites and three pairs of conserved moieties, obtained from the *S.* Typhimurium model. The exchange between core and global model also varied between the investigated conditions, the most pronounced difference occurring between mixed acid and homolactic fermentation, as would be expected based on the structural difference between the two categories.

Condensation of the catabolic core models greatly simplified the structure of the mixed acid networks (Figs. 4.2 and 4.3). The homofermentative model resulted in the trivial net-reaction:

```
x_GLC -> 2 x_D-Lact
```
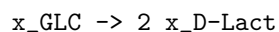
As can be seen in Figs. 4.2 and 4.3 some subsets (both singleton and multi-reaction) in the expanded model become integrated as components in other subsets during the condensation process. Under anaerobic conditions all the
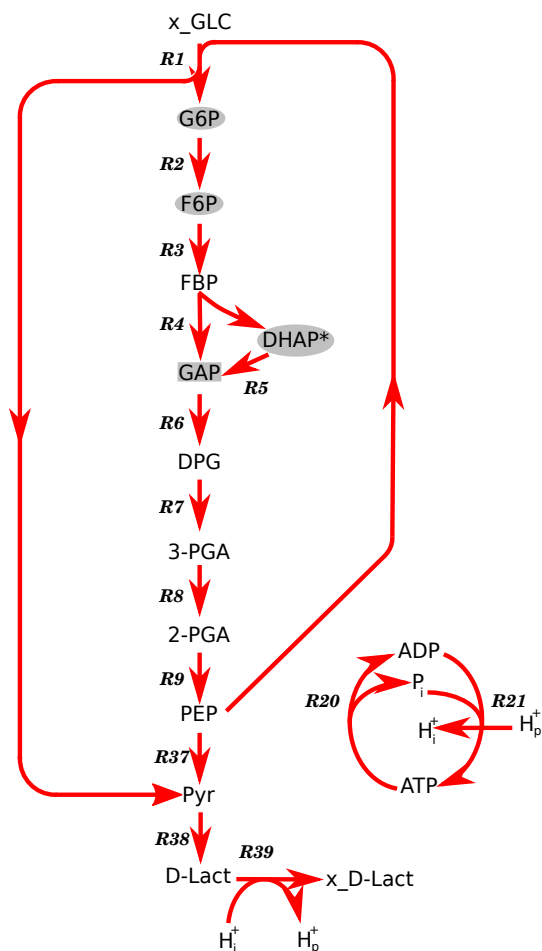
Figure 4.1: Catabolic core network under homolactic fermentation. The same network was obtained irrespective of oxygen availability. The responsive reactions involved glucose transport via the phosphotransferase system, glycolysis, conversion of pyruvate to lactate via lactate dehydrogenase, and export of lactate. Enzyme subset membership is indicated by reaction arrowhead colour; in this case all reactions belonged to the same subset. Metabolites that were imported at a fixed rate are indicated with gray boxes, exported metabolites are indicated with gray circles. * - DHAP was only exported under aerobic conditions. Reaction numbers refer to abbreviated reactions names (Table 1), accordingly: R1 - glucose transporter, R2 - PGIsomerase, R3 - PFK, R4 - FBP-Aldolase, R5 - TriPIsomerase, R6 - GapDH, R7 - PGKin, R8 - PGAM, R9 - Enolase, R20 - ATPase, R21 - ATPSynth, R37 - PyrKin, R38 - D-LacDH, R39 - lactate transporter.
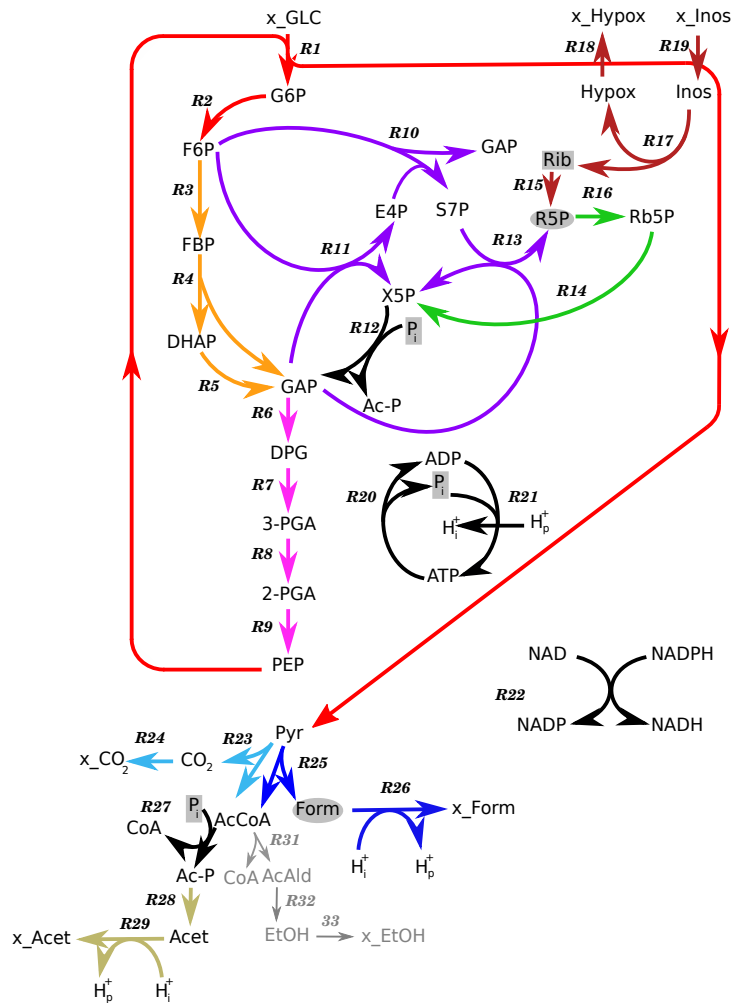
Figure 4.2: Catabolic core under mixed acid fermentative, anaerobic conditions. The enzymes subsets of the network are indicated by the colour of the reaction arrowheads. Metabolites that were imported at a fixed rate are indicated with grey boxes, exported metabolites are indicated with grey circles. Reactions R1 - R29 were common to the core model obtained in the absence and presence of oxygen. The greyed out reactions (R31 - R33, conversion of AcCoa to ethanol and subsequent export of ethanol from the network) were not part of the responsive reactions, but were necessary for obtaining a steady state solution for the core model. Reaction numbers refer to abbreviated reactions names (Table 1). Reactions R1 - R9 have the same meaning as in Fig. 4.1: R10 - Transald, R11 - TransketII, R12 - PhosKeto, R13 -TransketI, R14 - Rib5PEpi, R15 - RiboKin, R16 - Rib5PIso, R17 - RiboHyd, R18 - hypoxanthine transporter , R19 - inosine transporter, R20 - ATPase, R21 - ATPSynth, R22 - net-transhydrogenase, R23 - PyrDH-(NADP), R24 - CO_2transporter, R25 - PyrForm, R26 - formate transporter, R27 - PhoAcTrans, R28 - AcetKin, R29 - acetate transporter, R30 - AcAldDH, R31 - AlcDH, R32 - ethanol transporter.
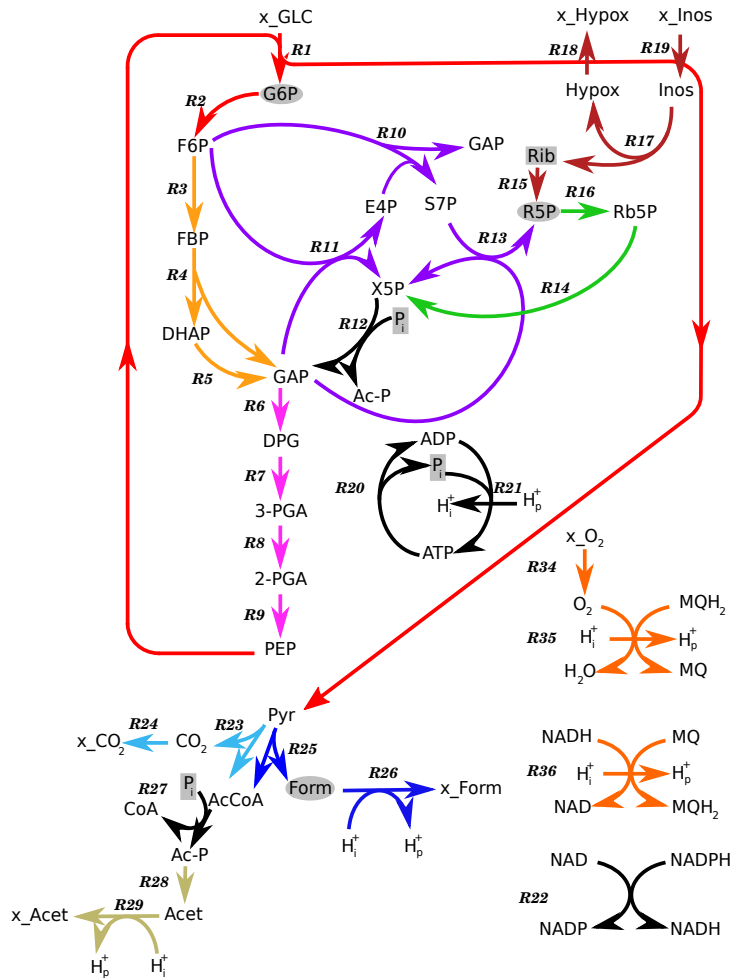
Figure 4.3: Catabolic core under mixed acid fermentative, aerobic conditions. The enzymes subsets of the network are indicated by the colour of the reaction arrowheads. Metabolites that were imported at a fixed rate are indicated with gray boxes, exported metabolites are indicated with gray circles. Reactions R1 - R29 were common to the core model obtained in the absence and presence of oxygen. Under these conditions the electron transport chain reactions (R34 - R36) were responsive to ATP demand variation. Reaction numbers refer to abbreviated reactions names (Table 1). Reactions R1 - R29 have the same meaning as in Fig. 4.2: R34 - oxygen transporter, R35 - MQ_oxidase, R36 - NADH_DH_MQ.

Table 4.1: Table of constant fluxes of import and export of metabolites between the core network and the global network. Data for all four conditions (mixed acid or homolactic fermentation, in the presence or absence of oxygen) are shown. For each conditions the data is split into conserved (upper part) and unconserved (lower part) metabolites. The numbers of reactions refer to reactions in the main model only. * - Core reactions involved with the conserved moieties responded with negligible, non-zero, flux-change to the change in ATP demand, they are subsequently not included in Figs. 4.1 – 4.3. ** - Gly-3P was involved in a net-transhydrogenase reaction.

Import			Export		
Metabolite	Flux	Reactions	Metabolite	Flux	Reactions
Mixed acid fermentation, anaerobic conditions					
ADP	116.11	33	ATP	116.11	62
NADP	45.60	12	NADPH	45.60	12
NADH	48.39	11	NAD	48.39	13
CoA	63.20	12	AcCoA	63.21	9
Met-THF*	1.80	2	5-Met-THF*	1.80	1
P_i	140.0	22	Gly-3P**	20.10	4
Rib	1.91	1	R5P	2.04	1
-	-	-	Form	1.66	3
Mixed acid fermentation, aerobic conditions					
ADP	99.62	29	ATP	99.62	58
NADP	45.50	13	NADPH	45.50	13
NADH	50.25	15	NAD	50.25	13
CoA	63.20	12	AsCoA	63.20	9
CDP*	16.45	2	CTP*	16.45	6
P_i	140.00	22	Gly-3P**	20.10	4
Rib	1.91	1	R5P	2.04	1
-	-	-	Form	1.66	3
-	-	-	G6P	1.25	2
Homolactic fermentation, anaerobic conditions					
ADP	118.40	35	ATP	118.40	64
NAD	48.14	15	NADH	48.14	13
NADPH	9.37	16	NADP	9.37	16
P_i	140.17	24	Pyr	58.65	5
GAP	2.90	4	Gly-3P**	20.10	4
-	-	-	F6P	3.30	3
-	-	-	G6P	1.25	2
Homolactic fermentation, aerobic conditions					
ADP	101.90	31	ATP	101.90	60
NAD	58.65	18	NADH	58.65	16
CDP*	16.45	2	CTP*	16.45	6
P_i	140.17	24	Pyr	58.65	5
GAP	2.90	4	DHAP	20.10	4
-	-	-	F6P	3.30	3
-	-	-	G6P	1.25	2

Table 4.2: Stoichiometry of enzymes subsets obtained by condensation of the catabolic core models. For each enzyme subset of the respective core network the condensed net stoichiometry is shown. Subsets 1 - 5 were obtained under mixed acid anaerobic conditions, subsets 5 - 8 under mixed acid aerobic conditions. External metabolites are indicated by the prefix "x_".

Enzyme subset	Stoichiometry		
	Reactants		Products
Ess_1	1 x_GLC + 2.5 P_i + 1 NAD + 1 ADP	→	1 NADH + 1 Pyr + 1 ATP + 1.5 Ac-P
Ess_2	1 Pyr + 1 CoA + 1 NAD	→	x_CO_2 + 1 AcCoA + 1 NADH
Ess_3	1 H_p^+ + 1 ATP + x_Acet	↔	1 Ac-P + 2 H^+ + 1 ADP
Ess_4	1 P_i + 2 Pyr + 2 H^+ + 1 CoA	↔	1 AcCoA + 1 Ac-P + 2 H_p^+ + 2 x_Form
Ess_5	1 AcCoA + 2 NADH + 1 H^+	↔	1 x_EtOH + 1 CoA + 2 NAD
Ess_6	2 x_GLC + 7 P_i + 2 AcCoA + 7 ADP + 2 NAD	→	2 CoA + 2 NAD + 2 Pyr + 5 H_p^+ + 7 ATP
Ess_7	x_O_2 + 1 NADH	→	2 H_p^+ + 1 NAD + H_2O
Ess_8	1 CoA + 1 Pyr	→	H_p^+ + 1 AcCoA + x_formate

reactions leading up to Pyr (i.e. in Fig. 4.2 reactions R1 - R19) form one subset, phosphate acetyltransferase (R27 in the figure) forms a subset with the formate producing subset in the expanded network (involving pyruvate formate-lyase and a formate transporter), and the net-transhydrogenase together with the CO_2 producing branch forms a subset in the condensed model. The acetate producing subset in the expanded model is conserved in the condensed model, also the added ethanol related reactions form a single net reaction in the condensed model. Under aerobic conditions reactions R1 - R19 (in Fig. 4.3) and R27 - R29 (i.e. phosphate acetyltransferase and the acetate producing branch) forms one single condensed reaction; formate is produced and exported by the subset involving pyruvate formate-lyase and formate transporter; and the reactions involving oxygen uptake and reduction to water, while generating proton-motive force for subsequent ATP synthesis, form a single subset (reactions R34 - R36 in Fig. 4.3, R7 in Fig. 4.4.B). The CO_2 producing subset is unmodified by the presence of oxygen.

Flux correlation analysis of the data generated under mixed acid conditions (irrespective of oxygen availability), show that reactions with perfectly correlated fluxes do not necessarily belong to the same subset in the respective core model (Figs. 4.6 and 4.5). Examples of this, assuming anaerobic conditions, include the correlation of the Fig. 4.2 reactions R6 - R9 (lower glycolysis) with reactions R26 - R27 (formate producing branch), and the association between R3 - R5 (upper glycolysis) and R15, R17 - R19 (net conversion of inosine to
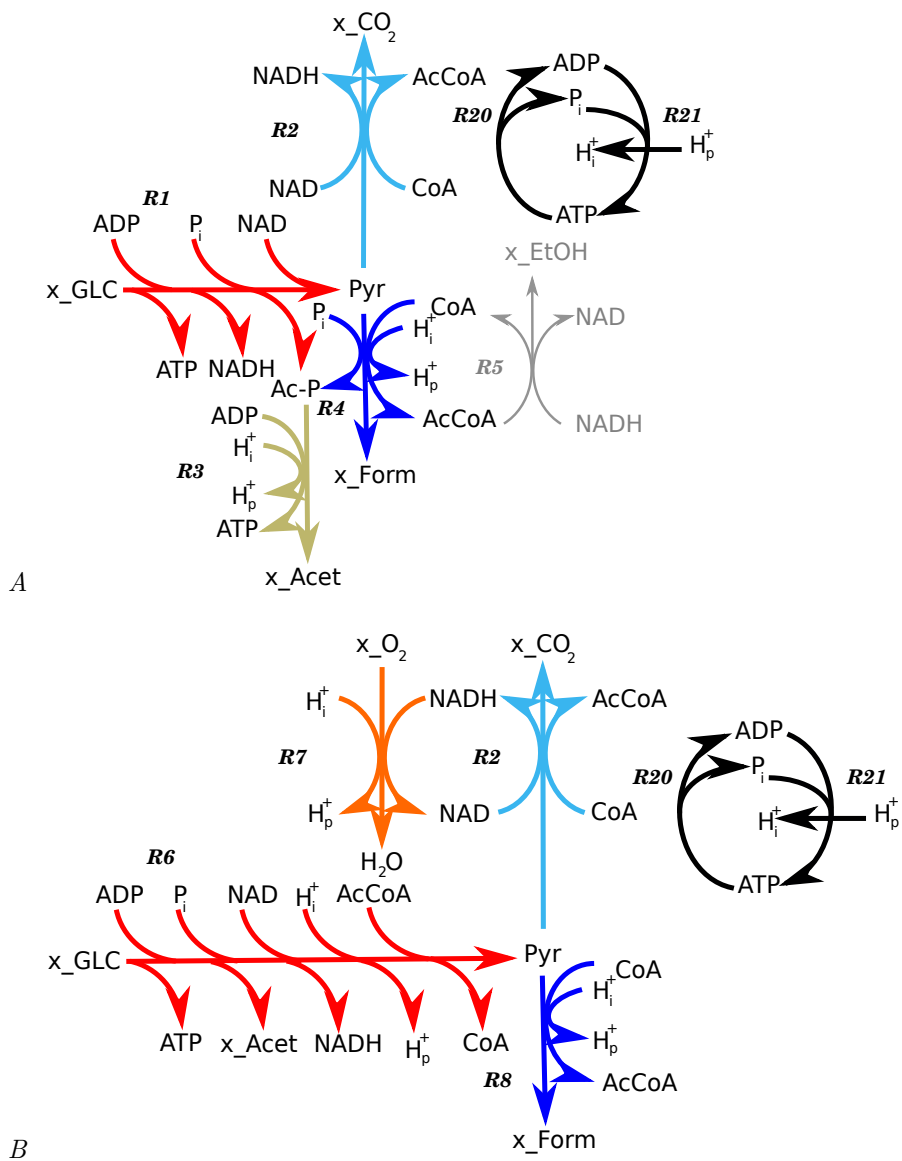
Figure 4.4: Condensed catabolic core network under mixed acid, anaerobic (A) and aerobic (B) conditions. The condensed networks are based on the networks in Figs. 4.2 and 4.3, colour codes are consistent with those used in the expanded networks. All metabolites involved with a particular reaction are indicated, but stoichiometric coefficients are not. For complete stoichiometries see Table 4.2. Reaction numbers refer to enzyme subsets names or reaction abbreviations for singleton subsets (unmodified from Figs. 4.2 and 4.3): R1 - Ess_1, R2 - Ess_2, R3 - Ess_3, R4 - Ess_4, R5 - Ess_5, R6 - Ess_6, R7 - Ess_7, R8 - Ess_8, R20 - ATPase, R21 - ATPSynth.
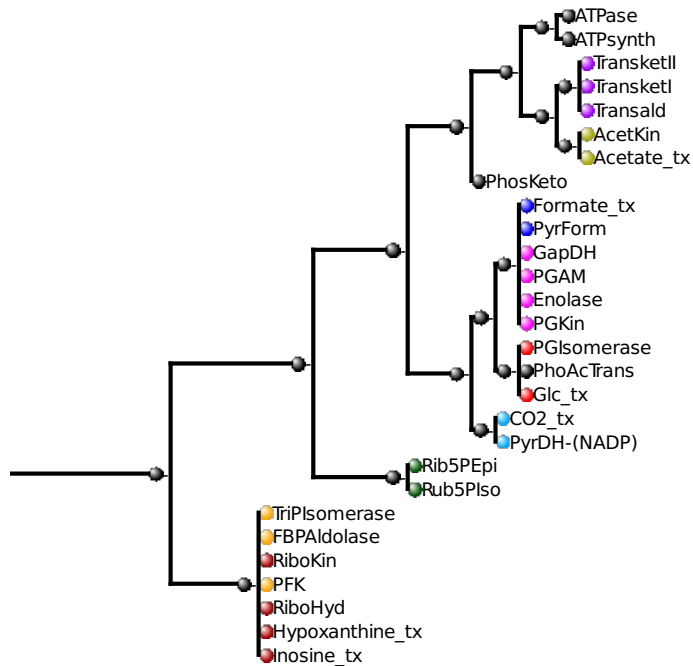
Figure 4.5: Flux correlation tree obtained under mixed acid, anaerobic conditions. Subsets in Fig. 4.2 are indicated with node colour.

ribose-5-phosphate). Similar observations are made under aerobic conditions. Analysis of the data generated under homolactic conditions show, trivially, that all responsive reactions are perfectly correlated (Fig. 4.7 and Fig. 4.8.E).

The response curves of the core networks (Fig. 4.8.A – 4.8.E) indicate relatively small differences between the reactions - unlike the responses obtained for *S.* Typhimurium (Chapter 3, Section 3.3.3), all core reactions obtained here were positively correlated with the ATPase flux.

4.4 Discussion

As shown by the different catabolic core models obtained by the ATP demand variation, the method used for simulating homolactic fermentation resulted in a less complex response, compared to the simulated mixed acid fermentation (Figs. 4.1 – 4.3). Since the FBA problem used for generating the flux data for
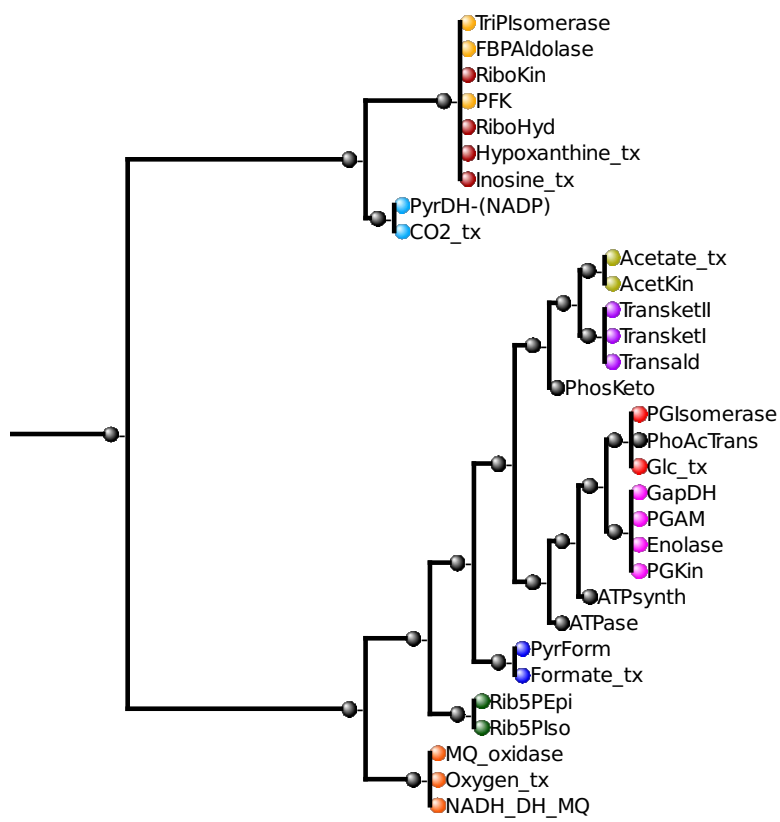
Figure 4.6: Flux correlation tree obtained under mixed acid, aerobic conditions. Subsets in Fig. 4.3 are indicated with node colour.
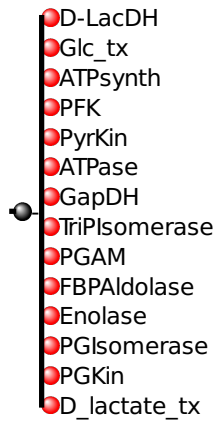
Figure 4.7: Flux correlation tree obtained under homofermentative conditions. As indicated all reactions in the core network (shown in Fig. 4.1) responded proportionally to the imposed ATP flux.

homolactic fermentation was more constrained than the mixed acid fermentation, this result is expected.

A noteworthy difference between the catabolic core models for the two modes of fermentation is the absence of the pyruvate kinase reaction from the mixed acid model. The reaction does however appear in all the FBA solutions used to extract the catabolic core models, but does not respond to changes in ATP demand under mixed acid conditions.

One surprising result was the inability of the mixed acid core model obtained under anaerobic conditions to sustain a steady state solution, unless the ethanol producing branch (subset 5 in Table 4.2) was included. The biochemical reason for this finding is obvious from the condensed network (Fig. 4.4.A): CoA and NAD cannot be recycled in the isolated network. Since these reactions (or any other functionally related reactions) were not identified in the analysis, it can be concluded that this function was carried out by the reactions carrying fixed flux.
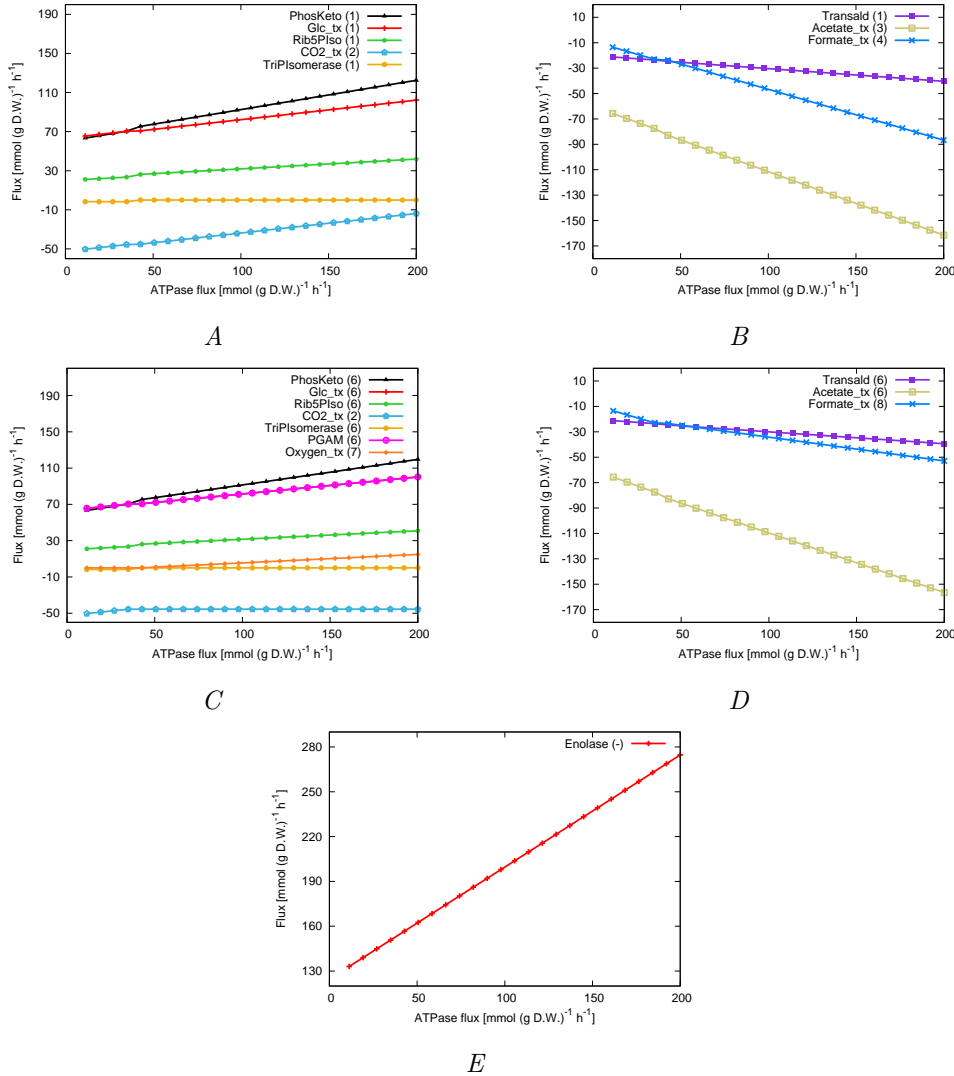
Figure 4.8: Responses of representative reactions from all subsets of the conditions analysed to changes in ATP demand. Subset number (as used in Fig. 4.4.A and 4.4.B, and Table 4.2) of the reactions are indicated in brackets. A: Response of PhosKeto, glucose transporter, Rib5PIso, CO_2 transporter, and TriPIsomerase under mixed acid, anaerobic conditions. B: Response of TransAld, and transporters of formate and acetate under mixed acid, anaerobic conditions. C: Response of PhosKeto, glucose transporter, Rib5PIso, CO_2 transporter, TriPIsomerase, PGAM, and oxygen transporter under mixed acid, aerobic conditions. D: Response of TransAld, and transporters of formate and acetate under mixed acid, aerobic conditions. E: Response of Enolase under both aerobic and anaerobic, homofermentative conditions.

Chapter 5

Modelling of metabolic interactions between *Salmonella* Typhimurium and *Lactobacillus plantarum*

5.1 Introduction

In this chapter a model combining those presented in the previous chapters (Chapters 3 and 4) is constructed and analysed. Application of FBA to multi-species systems has been described previously, mainly focusing on mutualistic interactions. Examples include modelling of interactions between sulphate-reducing *Desulfovibrio vulgaris* and the methanogen *Methanococcus maripaludis* [149]; and pairs of constructed auxotrophic strains of *E. coli* [150]. In these studies, interactions were modelled by treating the organism as separate compartments with a common exchange medium. In the case of the *D. vulgaris* - *M. maripaludis* network, the FBA was set to maximise the biomass yield of both organisms with various objective weightings - 10:1, 5:5, or 1:10 (*D. vulgaris*:*M. maripaludis*) [149]. Interactions between the *E. coli* strains were analysed using the Minimization Of Metabolic Adjustment (MOMA) algorithm [151], i.e. the objective for any given auxotrophic mutant was to minimise the deviation in flux solution from the wild type solution. This was done for all pairwise combinations of mutants where each pair was allowed to exchange exported metabolites.

Efforts to model antagonistic interactions in the form of competition has been based on identification of substrates that more than one of the organisms in the community can metabolise [152]. Zomorrodi and Maranas [153] developed a bi-level FBA framework for simultaneous optimisation of community-level objectives (e.g. maximisation of total biomass yield in the community) and species-specific objectives (maximisation of biomass yield of isolated organisms). The approach used here shares certain features with the methods described above: the organism-GSMs are incorporated as components, interconnected via a common medium. The integrated model is then subject to FBA where the biomass synthesis of the individual organisms is fixed and the objective is to minimise total flux of the integrated system. The FBA is repeated for different biomass ratios of the two organisms. This analysis will identify potentially mutually beneficial metabolic interactions between organisms in the community. Arguably, a limitation with this approach, as with those mentioned, is that only mutualistic interactions can be meaningfully modelled, provided only structural information of the network is available. As mentioned, some attempts to model competition have been based on identifying common substrates that can be utilised by more than one member of the community [152, 153], the outcome of this competition is assumed to be predictable using FBA, i.e. if the FBA solution assigns one organism a higher rate of biomass production for a given substrate uptake rate, compared to the competing species, the species with the higher growth rate out-competes the other species. There are at least two potential problems with this approach: (i) the equation of growth rate with biomass yield; (ii) the omission of the inhibitory effect of metabolic by-products. As reviewed in Section 4.1, based on [71], FBA optimises biomass yield, not growth rate. The assumption that a given species will out-grow a competing species because its biomass yield is potentially higher, may be true, but this cannot be concluded. The homolactic mode of fermentation in *L. plantarum*, presented in the previous chapter, is an example where this assumption is not true: despite the lower growth yield (and calculated growth rate based on FBA) of homolactic fermentation, this is the observed mode of fermentation, rather than mixed acid fermentation, precisely because homofermentative fermentation enables a higher growth rate. The second objection to this approach, which is also exemplified by the metabolism of *L. plantarum*, is that it ignores the inhibitory effect of excreted metabolic by-products. The lactate production of of various *Lactobacillus* species has been shown to have a strong inhibitory effect on the growth of competing, pathogenic bacteria [154]. This inhibition has been shown not to be a simple pH-effect, which makes the phenomenon difficult to model with a strictly structural approach. A solution to this problem will be not presented here, but these limitations of structural modelling of bacterial communities are

recognised.

5.2 Methods

In this section the methods used for integration of the GSMs of *L. plantarum* and *S.* Typhimurium, and an FBA-based approach for interrogation of metabolic interactions are described.

5.2.1 Model integration

The basic approach to model integration used here is shown in Fig. 5.1: Media components (consumable by any or both of the organisms) are identified and made available, as internal metabolites, to the common compartment; import reactions present in the original model are modified to allow uptake of the media components in the common compartment; metabolic by-products that are potentially producible by any of the organisms are made exportable to the common medium, available as media components to the other organism, and exportable from the system. The biomass components are the only compounds that can be made external directly by any of the organisms.

One prerequisite for the integration process is that metabolite identifiers are used consistently in the original models, since inconsistencies in naming would only result in creating structurally dead import reaction for the metabolic by-products. Here, this was guaranteed by using models that are constructed from a common database, BioCyc, as described in Chapters 3 and 4. In order to distinguish the different compartments of the integrated model, reaction and metabolite identifiers were modified by appending compartment-specific suffixes to the names: `_MetaLac` for *L. plantarum*; `_MetaSal` for *S.* Typhimurium; and `_InterMed` for the the common medium. Model integration was carried out using the following functions:

- `ModInt(...)` creates and saves an initial model containing the reactions of the two user-supplied structural models. Each reaction is distinguished by appending a suffix, equal to the file-name (omitting file extensions) of the respective model, to the reaction name and the metabolites it involves. These suffixes are returned as a list.

- `AppendModel(...)` is an auxiliary function to `ModInt(...)`. Provided a model to integrate, a stoichiometry matrix of integrated model (possibly empty), a list of suffixes (possibly empty), it integrates the model in the matrix of integrated models and returns the list of suffixes, with the name of the model appended.
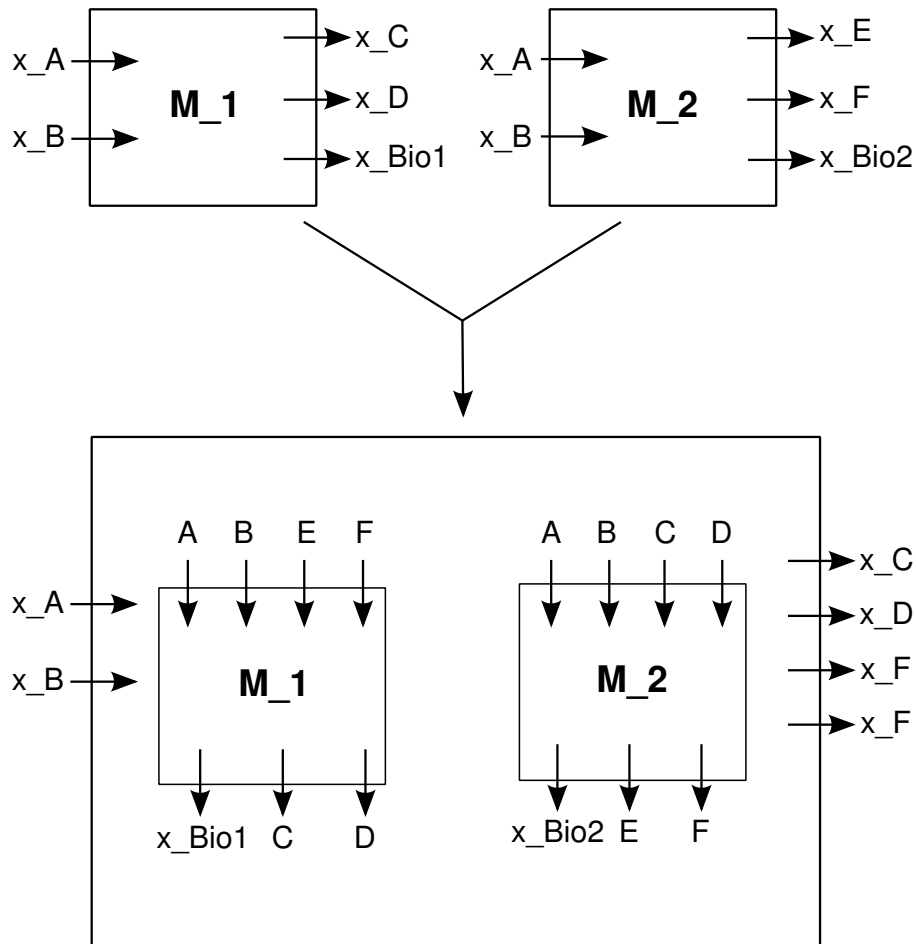
Figure 5.1: Overview of the model integration process. Models M_1 and M_2 can both take up the external metabolites A and B; M_1 can produce (external) biomass (Bio1), and metabolites C and D; M_2 can produce (external) biomass (Bio2), and metabolites E and F. The integration process includes identifying media components (A and B), exported compounds (C, D, E, and F), and biomass components (Bio1 and Bio2). In the integrated model external media components can be imported to a common medium and taken up by both species. All exported compounds can be released to the common medium and exported out of the integrated model or taken up by any other species than the producing species.

- `MakeInterMed(...)` replaces all external metabolites involved with transport reactions in the integrated model constructed by `ModInt(...)` with metabolites defined in the common medium. For each metabolite thus defined (e.g. `metA`), a transporter in the transporter module (named `fname_ext`) is defined, accordingly:

```
metA_tx_InterMed:
   x_metA -> metA_InterMed
```

The reversibility of the new transporter is identical to the reversibility of the original reaction, i.e. a medium component can only be imported, metabolic by-products can only be exported by the system. If any of the reaction-names generated automatically are already defined in the model a list of these is returned.

- `MakeUptakeReac(...)` enables a given organism in the integrated model to take up metabolites produced by any other organism from the common medium. The reactions are named by concatenating the identifier of the transported metabolite with the suffix of the receiving organism and the suffix `_UpTake`. Automatically generated reaction names that are already defined in the model are returned as a list.

- `SubHasTx(...)` is an auxiliary function to `MakeUptakeReac(...)` and is used to identify metabolites in the common medium that are already available to a given organism-specific compartment.

- `MakeNewTxDict(...)` is an auxiliary function to `MakeUptakeReac(...)`. Given an organism suffix (`org_suf`) and a common medium-metabolite (`met`), the function defines a reaction that links `met` with the equivalent metabolite in the organism-compartment indicated by `org_suf`, specified as a Python dictionary. The function also returns a reaction name indicating the name of metabolite being transporter, the recipient organism, and that the reaction is an uptake reaction (has suffix `_UpTake`). For example, if `met_A_InterMed`, which corresponds the originally internal metabolite `MET-A` and `_OrgX` are given, reaction:

```
met_A_OrgX_UpTake:
   met_A_InterMed -> MET-A_OrgX
```

is specified by returning:

```
met_A_OrgX_UpTake,{'met_A_InterMed':-1,'MET-A_OrgX':1}.
```

All uptake reactions are irreversible by default.

These functions were used to construct the full integrated model in terms of the following modules:

- *Integrated model:* Generated by `ModInt(...)` (and modified by `MakeInterMed(...)`). Contains all reactions from the original model, but modified as described above.

- *Common medium:* Generated by `MakeInterMed(...)`, defines the medium available to both sub-models.

- *Transport module:* Generated by `MakeUptakeReac(...)`, defines all input/output relations of the integrated model, except for export of biomass components, which is defined in the organism-specific compartments of the integrated model.

These modules were combined into a top-level model. The final model was validated in terms of stoichiometric and energetic constancy (as described in Chapter 3, Section 3.2.2).

5.2.2 Identification of mutualistic interactions

In order to identify possible metabolic interactions between *S.* Typhimurium and *L. plantarum* a set of FBA problems (Equation 5.1) for the integrated model were solved, where the organism-specific networks were constrained to produced biomass at experimentally observed rates. For each new solution the ratio between the two organisms was modified.

$$
\begin{aligned}
\text{minimize} \quad &: \quad \sum_{i=1}^{m} |v_i| \cdot c_i \\
\text{subject to} \quad &\begin{cases}
\mathbf{N}_{n,m} \cdot \mathbf{v} = \mathbf{0} \\
v_j = t_j \cdot r; \ e \leq j \leq f, \ \forall r \in \{0, \ \hat{n}, \ 2\hat{n}, \ \dots, 1\} \\
v_p = t_p \cdot q; \ g \leq p \leq h, \ q = 1 - r \\
v_{\text{ATPase}_{\text{MetaSal}}} = J_{\text{ATPase}_{\text{MetaSal}}} \\
v_{\text{ATPase}_{\text{MetaLac}}} = J_{\text{ATPase}_{\text{MetaLac}}}
\end{cases}
\end{aligned}
\tag{5.1}
$$

Where $\mathbf{v}$ is the vector of reaction rates, $\mathbf{N}$ is the $(n \times m)$ stoichiometry matrix, where each reaction is associated with an objective coefficient, c_i. As described in Chapter 4, homolactic mode of fermentation was modelled by assigning objective coefficients 10-fold higher to transporters of fermentation by-products associated with mixed acid fermentation, than the other reactions in the network, thus penalising mixed acid and (indirectly) promoting homolactic

fermentation. The impact of mode of fermentation on mutualism was investigated by setting the objective coefficients as described above for the equivalent transporters in the *L. plantarum* compartment. The rate of biomass production was fixed with the transport vector **t** consisting, as described in previous chapters, of the vector of biomass component concentrations of the respective organisms and a common growth rate (μ_{max}, set to 0.4 h^{-1}). The proportion of *S.* Typhimurium to *L. plantarum* biomass was fixed with the factor r (and q, which equals $1 - r$), which was multiplied with the organism-specific component of **t**. The biomass proportion factor r (and q) was sequentially set to all values in a range between 0 and 1, where the increment between two subsequent solutions, $\hat{n}$, was set to 0.02, corresponding to 50 solutions $\left(\hat{n} = \frac{1}{n_{solutions} - 1} \right)$. Maintenance energy, modelled as a generic ATPase (one for each organism), was fixed for the two sub-models as described previously (Section 3.2.3, for *S.* Typhimurium, and Section 4.2.1, for *L. plantarum*). Using the relationship:

$$ r_{\text{ATP}} = Y_{\text{ATP}} \cdot \mu + m_{\text{ATP}} $$

with μ = 0.4 h^{-1}, $Y_{\text{ATP,MetaLac}}$ = 27.6 mmol ATP (g DW)$^{-1}$, $m_{\text{ATP,MetaLac}}$ = 0.36 mmol ATP (g DW)$^{-1}$ h^{-1}, $Y_{\text{ATP,MetaSal}}$ = 60 mmol ATP (g DW)$^{-1}$, $m_{\text{ATP,MetaSal}}$ = 8.4 mmol ATP (g DW)$^{-1}$ h^{-1}, the following values were obtained:

$$ J_{\text{ATPase}_{\text{MetaLac}}} = 32.30 \text{ mmol (g DW)}^{-1} \text{h}^{-1}, $$
$$ J_{\text{ATPase}_{\text{MetaSal}}} = 11.4 \text{ mmol (g DW)}^{-1} \text{h}^{-1} $$

As in FBA problems presented in previous chapters, the objective was to minimise the sum of flux through the integrated network. In previous analysis this was motivated by the assumption that cells tend to minimise the protein investment for growth. Here, the purpose is to identify potential mutualistic interactions between two organisms that are assumed to minimise their protein investment. By using flux minimisation for the integrated network flux solutions that are mutually beneficial in terms of minimising protein investement are favoured.

The model integration was carried out so that reactions carrying flux between compartments could be identified easily based on identifier suffix, i.e. uptake of a metabolite, produced by one of the organism, by the other organism could only occur in two ways: either through an uptake reaction (constructed by `MakeUptakeReac(...)`, all having the suffix `_UpTake`); or through a fermentation by-product present in the original model operating in the importing direction.

Table 5.1: Summary of distributions of reactions and metabolites over the compartments of the integrated model. All metabolites exclusively defined in the common medium are involved with one transport reaction connecting it to the external medium and at least one sub-model transporter. The number of input and output reactions are equal by coincidence.

Module		Reactions		Metabolites	
S. Typhimurium GSM		1232		1072	
	Uptake		*36*		*36*
	Biomass		*50*		*50*
L. plantarum GSM		677		569	
	Uptake		*2*		*2*
	Biomass		*42*		*42*
Common medium		86		86	
	Input		*43*		*43*
	Output		*43*		*43*
Total		1995		1728	

5.3 Results and discussion

The number of reactions and metabolites of the integrated model are summarised in Table 5.1.

Like the two component models, the integrated model was stoichiometrically consistent for all carbon-containing metabolites.

The analysis of metabolic interactions between the two organisms identified one potential metabolite-exchange, namely transport of glycolate from *L. plantarum* to *S.* Typhimurium (Fig. 5.2).

As can be seen from Fig. 5.2 the uptake of glycolate by *S.* Typhimurium increases linearly with the relative concentration of the producer (*L. plantarum*), when the concentration of *L. plantarum* reaches 100% the uptake drops to 0%, since there is no organism to consume the glycolate. This shows that the demand for glycolate by *S.* Typhimurium is not saturated at any point where the *S.* Typhimurium biomass is non-zero. Glycolate excretion is required for biomass synthesis of the *L. plantarum* model with the biomass composition specified in [70]. Glycolaldehyde is a by-product of the biosynthesis of tetrahydrofoliate (THF), which can be converted to glycolate and subsequently excreted from the system (given the set of transporters reported in [70] glycolaldehyde cannot be exported). There are no reports of glycolate production as an essential by-product of biomass synthesis for *L. plantarum*, which could be explained by the low concentration of THF in biomass (10^{-5} mmol (g DW)$^{-1}$), resulting in a low rate of glycolate excretion (equal to the rate of tetrahydrofoliate export).

The glycolate consumed by *S.* Typhimurium is converted to glyoxylate and
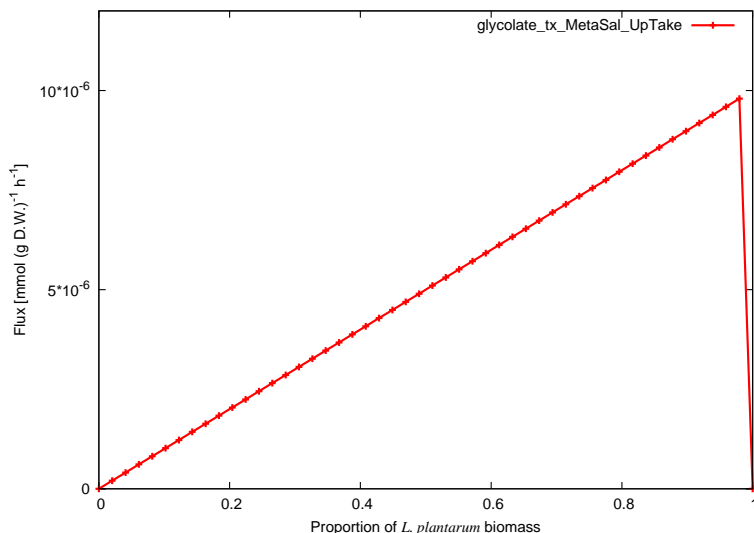
Figure 5.2: Transport of glycolate from *L. plantarum* to *S.* Typhimurium as a function of relative *L. plantarum* biomass concentration.

enters TCA via the glyoxylate bypass (Fig. 5.3). The flux through the reactions in *S.* Typhimurium involved with glycolate (aldehyde dehydrogenase and glycolate oxidase, reactions 10 and 12, respectively, in Fig. 5.3), is in the order of 10^{-2} mmol $(g \, DW)^{-1}h^{-1}$. The glycolate that is taken up by *S.* Typhimurium is consumed by glycolate oxidase, since the uptake flux is negligible in comparison with the flux of aldehyde dehydrogenase (which produces glycolate) the glycolate provided by *L. plantarum* is not sufficient to completely replace the flux from aldehyde dehydrogenase. This is a concequence of the relative crudness of the analysis; with the number of solutions ($n_{solutions}$) set to 50 (and $\hat{n} = 0.02$ or 2%) the highest possible *L. plantarum*:*S.* Typhimurium proportion is 98:2 (excluding the ratio 100:0). With this ratio the glycolate flux from *L. plantarum* to *S.* Typhimurium reaches its highest value, which approaches the rate of tetrahydrofoliate export assuming a ratio of 100:0 (10^{-5} mmol $(g \, DW)^{-1}$).

As $n_{solutions}$ is allowed to increase, and the *S.* Typhimurium proportion allowed to decrease, the glycolate uptake flux by *S.* Typhimurium approaches the flux of glycolate oxidase. This switch only occurs when the *S.* Typhimurium proportion is in the order of 10^{-6}.

There are no reports of mutualistic metabolic interactions between *S.* Typhimurium and *L. plantarum* *in vivo*. If this interaction did occur *in vivo* it would be difficult to show this experimentally owing to the very small fluxes

84
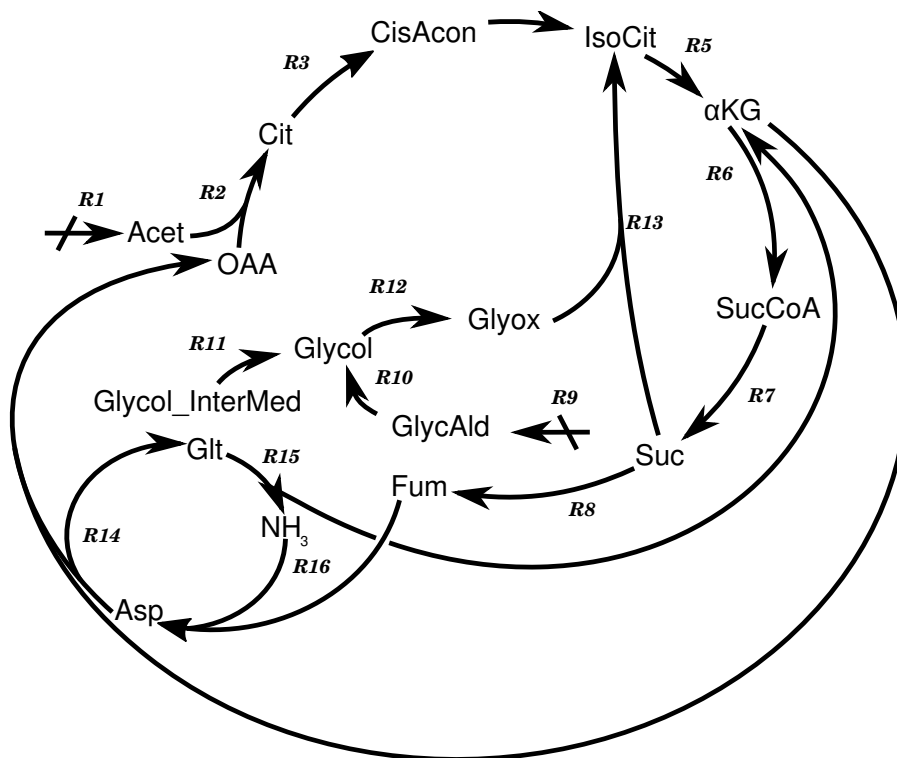
Figure 5.3: Partial network diagram of glycolate utilisation in *S.* Typhimurium. Unbalanced reactions are indicated with crossed-over reaction arrowheads. `Glycol_InterMed` indicates glycolate in the common medium. Key: R1 - AcetKin, R2 - CitLyase, R3 - AconDehydr, R4 - AconHydr, R5 - IsoCitDH, R6 -2-KGDH , R7 - SCoASynth, R8 - SucDH, R9 - DiHydrAld, R10 - AldDH, R11 - glycolate uptake reaction, R12 - GlycOx, R13 - IsoCitLyase, R14 - Asp-Trans, R15 - GluDH-(NAD(P)), R16 -Aspartase.

involved. Also, the effect might be masked by the inhibitory effect of other metabolic by-products from *L. plantarum*, not accounted for in this modelling framework. Although glycolate transfer from *L. plantarum* to *S.* Typhimurium has not been experimentally identified, there are examples of metabolic interactions involving glycolate. *Ralstonia solanacearum*, a Gram negative pathogen of the *Pseudomonas* genus, has been hypothesised to use glycolate, which is a by-product of photorespiration, as a carbon source during infection of tomatoes, based on *in vivo* expression studies [155].

The surprising finding here is that the gains, in terms of global flux minimisation, from metabolic interactions between *S.* Typhimurium and *L. plantarum* (which in this study is assumed to able to produce up to 36 compounds) are limited to a negligible transfer of glycolate.

Chapter 6

Integration of high-throughput data with genome-scale metabolic models

6.1 Introduction

The analysis of GSMs can be integrated with various types of experimental data. Since GSMs (ideally) incorporate all genes encoding metabolic enzymes they are well suited integration of genome wide data. In this chapter methods for integration of two types of high-throughput data with GSMs, Phenotype Microarray and DNA microarray data, are explored. Phenotype Microarray (PM) is an automated system for characterisation of growth phenotypes of cells on a standard set of media components [156]. The PM pipeline involves distributing cells (suspended in a buffer solution that maintains viability but not growth) into 96-well microplates (commonly 20 standard plates are used). Each well contains a specific (dried) compound, which potentially could serve as an alternative source of C, N, P, or S, as well as other compounds needed for metabolising the plate-specific compound (e.g. if a specific C source is provided in a certain well, standard (dried) sources of N, P, and S (and micronutrients) are also provided). After addition of cells to the microplates, they are maintained at a growth-permissive temperature (typically 35° C for human adapted enterobacteria, such as *E. coli* and *S.* Typhimurium) and the rate of respiration monitored for some lenght of time (usually between 24 to 48 h) [156]. The buffer

solution contains a redox-sensitive tetrazolium dye, which in its reduced form produces a purple colour. A respiring cell will generate an electron-flow through the electron transport chain, which will reduce the dye with a rate proportional to the rate of respiration [157]. The rate of colouration of the medium can be monitored by a plate reader [158].

Previously, integration of PM data with GSMs has mainly been focused on comparing FBA results with experimental data. For this type of analysis, the PM output is typically interpreted as a proxy for growth, and compared to the feasibility or infeasibility of the analogous set of FBA problems. AbuOun et al. [120] reported overall prediction accuracy of a *S.* Typhimurium model in the range 64% (for sulphur compounds) to 97% (for phosphorous compounds). Discrepancies between FBA and PM results can be used as a starting point for continued model curation [159].

DNA microarray analysis allows identification of differentially expressed genes. Briefly, the method centres around the following steps [160]:

- *Collecting mRNA from the organism of interest under different conditions.* This usually involves a reference condition, e.g. the wild type organism cultivated on standard medium, and a test condition, e.g. a mutant strain or wild type strain cultivated in the presence of sub-lethal concentrations of a toxin.

- *Conversion of mRNA to cDNA and labelling.* This is accomplished using reverse transcriptase, and incorporation of fluorescently labelled bases (one colour for each experimental condition).

- *Hybridisation of labelled cDNA to DNA probes.* Probes are DNA fragments that are fixed to a solid support (array). Each probe represents a complete or partial gene, and is identified based upon its position on the array. Microarray probes can be of a variety of types: single or double stranded, synthesised separately using PCR (polymerase chain reaction) [161] or *in situ* as oligonucleotides. cDNA from the two conditions are applied to the array simultaneously and allowed to hybridise to the probes.

- *Scanning of fluorescence.* The relative abundance of the two pools of cDNA is determined indirectly by quantifying the two fluorescence signals from the array.

- *Data analysis.* Statistical analysis of the raw data usually involves removing background noise, normalisation of experimental biases, and identification of genes whose expression is significantly different under the investigated conditions. Data is commonly transformed from expression ratios $\left(\text{i.e. } \frac{\text{signal}_{\text{test}}}{\text{signal}_{\text{reference}}}\right)$ to fold changes $\left(\text{i.e. } \log_2\left(\frac{\text{signal}_{\text{test}}}{\text{signal}_{\text{reference}}}\right)\right)$.

Several methods for integration of DNA microarray data with GSMs have been described previously. All approaches are based upon the assumption that expression rate of a given gene is correlated with the flux through the reaction ultimately encoded by the gene. The *GIMME* (Gene Inactivity Moderated by Metabolism and Expression) algorithm [162] is based upon solving a default FBA problem, then removing reactions in the network that are associated with under-expressed genes (given a user defined threshold). The FBA is then re-solved for the truncated model, if possible. If no feasible solution can be obtained reactions are re-introduced to the model so that the deviation from the expression data is minimised, until the FBA is feasible. *iMAT* (integrative Metabolic Analysis Tool) [163] involves sorting the reactions in the network into the categories, low, moderate, or high expression, based on the expression of the associated genes and user defined thresholds. A MILP is formulated, where the only constraints are that the solution must obey the steady state assumption and be thermodynamically feasible, the objective is to minimise the number of reactions associated with low expression in the solution and to maximise the number of highly expressed reactions. Similar to GIMME, iMAT allows a reaction associated with low expression to appear in the solution if no feasible solution can be obtained otherwise.

In this chapter a number of methods for model analysis relying on PM or DNA microarray data are presented. All analysis is applied to the *S.* Typhimurium model described in Chapter 3. The analysis presented in this chapter includes validation and subsequent curation of the model using comparisons between FBA solutions (mimicking the media conditions tested experimentally) and PM data. Solutions that are confirmed to support growth (more correctly, respiration) are subjected to a more detailed clustering analysis. A method for incorporating expression data in FBA is presented. Finally, an approach for integration of phenotype and expression data with FBA is presented.

6.2   Integration of PM data with metabolic models

A prerequisite for comparing sets of FBA solutions with PM data is that equivalence can be established between metabolite identifiers in the model used for generating the FBA solutions and the metabolite identifiers used in the experimental data set. Specifically, any source of C, N, S, or P for which growth data is available must have an unambiguous, if any, equivalent in the model. The compounds that fall into this category primarily involve complex, polymeric carbohydrates. These compounds are difficult to incorporate into metabolic models

since the monomers of which the polymers are composed are not uniquely defined. A more problematic situation is when a compound from the experimental data set is indeed defined in the model, but due to insufficient information this relationship cannot be established, and the compound is treated as undefined. Another potential problem is inconsistent use of compound names within either the model or the experimental data, which, if it cannot be guaranteed by the provider of the data, needs to be addressed in the integration process.

6.2.1 Software and computational methods

Software was developed to accomplished the following tasks: (i) harmonisation of metabolite identifiers used in the BioCyc database and the PM system; (ii) construction of transporter modules allowing uptake of metabolites investigated experimentally and expressed with BioCyc identifiers; (iii) comparing FBA solutions with corresponding PM data.

Database harmonisation

Harmonisation of the identifiers used in the PM system and the BioCyc database was primarily carried out using an auxiliary database, BiologKeggEntries, kindly provided by Dr William Newell (Animal Health and Veterinary Laboratories Agency). BiologKeggEntries collects information concerning PM compounds, each entry (compound) contains the Kegg database identifier, the position in the PM system (i.e. which well and plate it appears in), PM name, synonyms, mass, chemical formula, possibly compound identifiers in other databases, and reactions (Kegg identifiers) involved with the compound. Similar information is available for any compound defined in BioCyc. Since there are no direct links between BiologKeggEntries and BioCyc, these links were established indirectly by attempting to identify common links to at least one other database between all compounds defined in BioCyc and BiologKeggEntries.

Transporter-module construction

Each well on a PM plate contains one well-specific compound (source of C, N, P, or S) and three compounds (one for each element that is not being tested on the particular plate) that are common to all wells on the plate. This experimental set-up was modelled by constructing four transport modules that could replace the transport module described in Section 3.2. Each module represented one set of sources of a particular element, i.e. the carbon module contained transporters of all carbon sources tested and ammonia, sulphate, and phosphate; the nitrogen module contained all nitrogen sources and pyruvate (carbon source),

sulphate, and phosphate, etc. All modules allowed uptake of oxygen. All transport reactions involved only one internal and one external metabolite, except transporters that have been shown experimentally to take up metabolites (sugars) through the phosphotransferase system (PTS). These were defined as:

```
PEP + x_sugar -> Pyr + sugar-phosphate
```
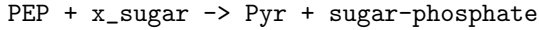
The following carbon sources were identified as PTS compounds: Glucose, mannitol, mannose, sorbitol, galactitol, glucosamine (and beta-D-glucosamine), tagatose, and fructose.

Comparison of FBA solutions with PM data

Growth on the nutrient sources defined in the model was modelled by solving the following equation, similar to Equation 3.1:

$$
\begin{aligned}
\text{minimize} \quad &: \quad \sum_{i=1}^{m} |v_i| \cdot c_i \\
\text{subject to} \quad &\begin{cases} \mathbf{N}_{n,m} \cdot \mathbf{v} = \mathbf{0} \\ v_j = t_j; \ k \leq j \leq m \\ v_e = 0; \ f \leq e \leq h, \ e \neq g \\ v_{\text{ATPase}} = r_{\text{ATP}} \end{cases}
\end{aligned} \tag{6.1}
$$

where, as previously, the problem is set to minimise the sum of flux, subject to the steady state constraint, biomass synthesis constraint, and ATP maintenance constraint. The additional constraint, that only one transporter (with index g in Equation 6.1) of alternative sources of a given element (transporters with indices in range f to h) can carry flux (while the other, transporters in that range cannot) was added to the problem. The problem was solved with g being set to each index in range f to h, sequentially. Each solution (or absence of solution) was compared with published PM data for S. Typhimurium (strain LT2) [120]. In AbuOun et al. [120] the phenotype data was reported as growth or no growth, which was compared to the feasibility or infeasibility of the corresponding FBA problem.

Cases where the FBA solution did not agree with the PM observation were subject to further scrutiny. Disagreement could be of two types: (i) The model predicted growth, but reactions required for catabolism could not be used *in vivo* due to regulation. (ii) Growth was observed experimentally, but not predicted by the model since reactions required for catabolism was missing from the database. The analysis was conducted using a total of 377 nutrients (190 C-sources, 94 N-sources, 59 P-sources, and 34 S-sources). The nutrients could be

divided into two broad categories - *known* nutrients, i.e. those that were defined in the model and *unknown* nutrients, that were not defined in the model and could thus not be tested explicitly, but were by default assumed not to support growth.

6.2.2 Compound clustering analysis

Cluster analysis was applied to the solutions generated by Equation 6.1. This analysis was applied in order to gain insight into the structure of the global catabolic network of *S.* Typhimurium, i.e. how complete flux solutions for biomass synthesis on different growth supporting compounds compare to each other. The clustering analysis was carried out in a similar fashion to the flux correlation analysis described in Section 3.2.3. The solutions generated by Equation 6.1 were collected into a matrix, $\mathbf{C}_{a,m}$:

$$\mathbf{C}_{a,m} = \begin{matrix} v_{1,tx_1} & \cdots & v_{m,tx_1} \\ \vdots & \ddots & \vdots \\ v_{1,tx_a} & \cdots & v_{m,tx_a} \end{matrix} \tag{6.2}$$

where a is the number of alternative nutrient sources available (using the symbols in Equation 6.1, $a = h - f$), m is the number of reactions in the model, and tx indicates a transporter in the model, as previously. Correlations between the flux solutions in $C_{a,m}$ were quantified by applying Pearson's correlation coefficient, $r(X, Y)$, defined in Equation 3.7, to all combinations of row-vectors (i.e. FBA solutions) in $C_{a,m}$ and collected into $\Delta_{a,a}$:

$$\boldsymbol{\Delta}_{a,a} = \begin{matrix} r\left(\mathbf{C}_{1,1\ldots m}, \mathbf{C}_{1,1\ldots m}\right) & \cdots & r\left(\mathbf{C}_{1,1\ldots m}, \mathbf{C}_{a,1\ldots m}\right) \\ \vdots & \ddots & \vdots \\ r\left(\mathbf{C}_{a,1\ldots m}, \mathbf{C}_{1,1\ldots m}\right) & \cdots & r\left(\mathbf{C}_{a,1\ldots m}, \mathbf{C}_{a,1\ldots m}\right) \end{matrix} \tag{6.3}$$

The correlations were visualised in a metabolic tree (Fig. 6.2), constructed as described in Section 2.2.

6.3 Integration of expression data with metabolic models

The least technical approach to metabolic interpretation of expression data is simple inspection of up and down regulated genes, whereas the method presented here is a systems based approach. As mentioned in Section 6.1, previously published methods for integration of expression data and FBA, rely, to varying degrees, on excluding reactions from the solution space based on the expression

of the corresponding gene(s). In contrast, the method presented here is based upon favouring reactions associated with highly expressed genes, and penalising those associated with genes expressed at a low rate, by assigning each reaction in the model an objective coefficient, or weighting factor, inversely proportional to the expression of their corresponding gene(s). If the FBA is formulated as a minimisation problem a high objective coefficient, proportional to a low expression rate, adds a high cost to the reaction assigned the coefficient. The benefits of this approach are that it is simple to implement, does not require user defined thresholds for high or low expression, nor exclude reactions from the solution space. A drawback is that it is limited to FBA formulations where all reactions in the model are included in the objective function, such as flux minimisation.

6.3.1 Methods

DNA microarray data for *S.* Typhimurium under several growth conditions and globally normalised using the BABAR package [164], was kindly provided by Dr Arthur Thompson (Institute for Food Research). Associations between reactions in the model and genes in the data set were obtained from the BioCyc database through the PyoCyc module in ScrumPy.

The original data was expressed as fold changes (FC), i.e.

$$\text{FC} \; = \; \log_2 \left(\frac{\text{signal}_{\text{test}}}{\text{signal}_{\text{reference}}} \right).$$

This formulation of expression data cannot be readily integrated into an FBA problem since objective coefficients must be non-negative. Further, the weighting factors needed for this must be inversely proportional to the gene expression rate. The data was transformed to inverse expression ratios in order to comply with these constraints:

$$\text{weight} = \frac{1}{\text{expression ratio}} = \frac{\text{signal}_{\text{reference}}}{\text{signal}_{\text{test}}} = \frac{1}{2^{\text{FC}}} \qquad (6.4)$$

This method was applied to ATP demand analysis (Section 3.2.3) by setting the objective coefficients in Equation 3.2, c_i, to values derived from Equation 6.4. In cases where a reaction was associated with more than one gene and no information on the relationship between genes (gene products constituting isoenzymes or components in a multimeric enzyme) was available, the highest signal was used. This analysis was carried out using microarray data extracted from *S.* Typhimurium during macrophage infection. In the analysis all reactions that were not associated with any genes were assigned an objective coefficient of 1. As previously, the analysis was conducted assuming only glucose minimal

medium.

6.4 Combining Microarray and PM data for model analysis

As mentioned in Section 1.3.3, the nutrients available to, and used by, *S.* Typhimurium during infection are largely unknown. In this section a method combining expression and PM data with metabolic modelling in order to identify compounds consumed by the organism from which the expression data was obtained is presented. In brief, the method uses FBA solutions generated with the method described in Section 6.2, identifies reactions unique for the catabolism of a set of (possibly singleton) compounds, and assigns a fold change to the set, based on the expression of the compound-associated reactions.

6.4.1 Method

Using the flux solutions collected in $C_{a,m}$ (Equation 6.2), internal reactions (columns in $C_{a,m}$) were associated with PM compounds (rows in $C_{a,m}$) on which they were active. The purpose of this procedure was to identify reactions that could be associated with as few compounds as possible, so that gene expression associated with a given reaction could be indicative of catabolism of a unique compound. Thus, the initial reaction-compound associations were modified by removing redundant information: If a set of reactions, $R = \{r_1, \ldots, r_x\}$, could be associated with a single compound, $A = \{a\}$, and another set of reactions, $\hat{R} = \{\hat{r}_1, \ldots, \hat{r}_y\}$, could be associated with a set of compounds, B, containing this single compound, $B \supset A$, then A was removed from B so that $\hat{R}$ was associated with $B' = B \Delta A$. This procedure could be generalised for any size of sets A and B, as long as $B \supset A$. Importantly, removal of redundant information from B based upon A could reveal redundant information in a set C, such that $C \supset B$, which would not have been apparent before simplification of B. In order to deal with this possibility the method was implemented as a recursive algorithm, i.e. the procedure was repeated until no more redundant information could be removed.

The method is illustrated in Fig. 6.1. This set of non-redundant reaction-compound associations will hereafter be referred to as *minimal reaction-compound associations*, and the process of determining this association will be referred to as *reaction-compound simplification*

The assignment of relative expression to compounds was done in three steps: (*i*) for all reactions in a minimal reaction-compound association the respective fold change (as in Section 6.3, reactions with multiple gene associations were
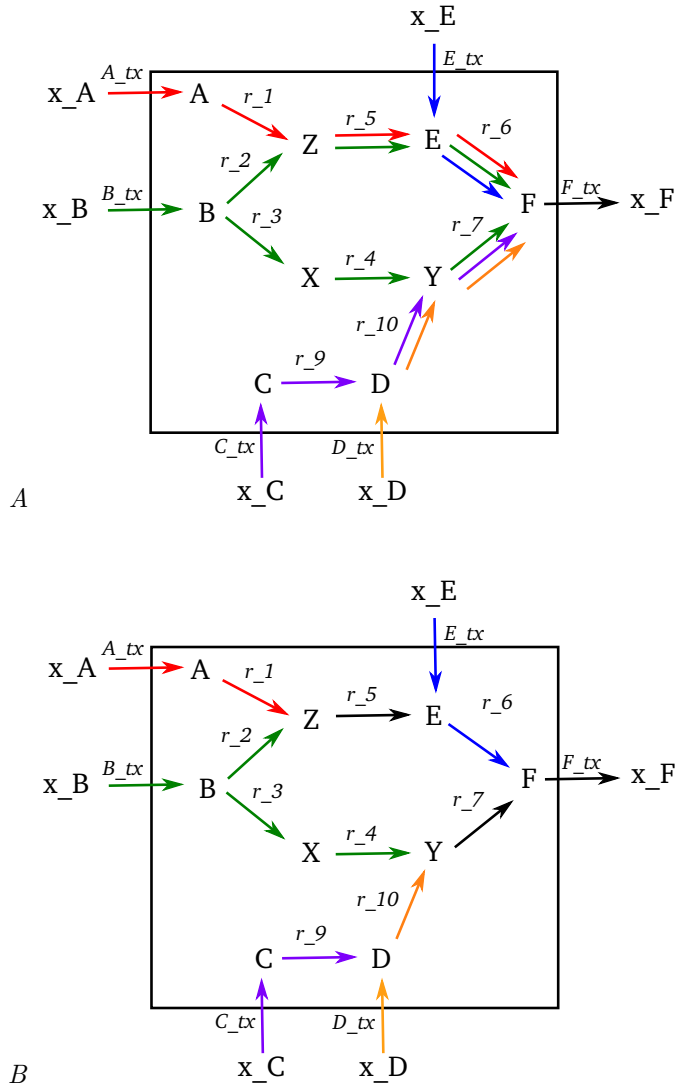
Figure 6.1: Illustration of reaction-compound simplification. **A**: Inspection of the FBA solutions obtained when using compounds A - E as single media component shows the associations indicated by the colours of the arrowheads. Specifically, $\{r_1\} \sim \{A\}$, $\{r_2, r_3, r_4\} \sim \{B\}$, $\{r_5\} \sim \{A, B\}$, $\{r_6\} \sim \{A, B, E\}$, $\{r_9\} \sim \{C\}$, $\{r_{10}\} \sim \{C, D\}$, and $\{r_7\} \sim \{B, C, D\}$ (where $\sim$ indicates association between sets of reactions and compounds). Any indication of activity of r_5 could be due to catabolism of either A or B. Reaction r_1 can only be active if A is available and reactions $r_2 - r_4$ can only be active when B is available. Thus, r_1, r_2, r_3 and r_4 are better indicators for A or B than r_5, and the association $\{r_5\} \sim \{A, B\}$ can be removed. This reasoning can be applied to all reactions with multiple compound associations, resulting in the figure in **B**.

95

Table 6.1: Results of comparison between FBA growth predictions and PM data. *Known* nutrients are defined in the model, whereas *unknown* nutrients are not and thus by default assumed to not support growth. n indicates the number of nutrients tested for each category.

	Known			Unknown		
Element	*Match(%)*	*Mismatch(%)*	*n*	*Match(%)*	*Mismatch(%)*	*n*
Carbon	90	10	97	71	29	93
Nitrogen	80	20	59	89	11	35
Sulphur	40	60	10	52	48	24
Phosphorus	83	17	29	40	60	5

assigned the maximum fold change) were collected; (*ii*) the average fold change for the reactions was calculated; (*iii*) the fold change calculated in (*ii*) was expressed as the relative fold change (in range [0,1]) in terms of the maximum and minimum fold change in the data set. The analysis was applied to the same expression data set, collected during intracellular growth of *S.* Typhimurium, as that used for integration of expression data and FBA (Section 6.3.1).

6.5   Results

6.5.1   Integration of PM data with metabolic models

The results of the comparisons between FBA solutions and PM data are shown in Table 6.1. As indicated, the agreement between experimental and FBA data was 80 - 90% for the carbon and nitrogen sources. For sulphur and phosphorus sources the agreement was poorer, indicating a lack of biochemical information on catabolism of S and P sources. These results were obtained following iterative curation based upon previous, poorer matches, until no more improvements could be made. The conditions where the agreement was below 50% (i.e. sulphur with known metabolites and phosphorus with unknown metabolites) was based upon few (10 and 5, respectively) samples. As mentioned, similar analysis was performed by AbuOun et al. [120] and yielded the following agreements between experimental and FBA data (number of sources in parenthesis): Carbon sources - 86% (92), nitrogen sources - 83% (66), phosphorous sources 97% (29), and sulphur sources - 64% (14).

Compound clustering analysis was carried out using the flux solutions of all growth supporting carbon compounds, as indicated by both experimental and FBA data. The result of the analysis is visualised as a metabolic tree (Fig. 6.2). In order to identify possible associations between chemical structure of compounds and compound clustering based upon FBA solutions, the compounds were sorted into 21 broad categories (listed in Fig. 6.2), based upon

BioCyc compound classes. As can be seen in the metabolic tree, there is some agreement between the structure of the compounds and the flux solutions for their conversion to biomass precursors. There are three main clusters, one is dominated by flux solutions for most amino acids and dipeptides. Most sugars (pentoses, hexoses, and their phosphorylated equivalents), appear in the same main cluster, closely correlated with the solutions for nucleosides. Solutions for most carboxylic acids cluster together, but this main cluster contains various other compound classes with fewer members, e.g. diols, cyclic alcohols, and sugar acids. There are also deviations from these patterns: Uridine (a pyrimidine) forms an isolated node; fucose and rhamnose (non-PTS hexoses) cluster together with the carboxylic acids; glucose 6- and 1-phosphate (hexose phosphates) and galactoses (hexoses) are closely correlated with the consumption of polysaccharides; some monocarboxylic acids (formate, glyoxylate, and glycolate) form isolated nodes.

As a result of the curation process 65 new reaction were included in the model. All results presented in this thesis have been obtained with this curated model. All elements-specific models are included in Appendix C

6.5.2 Expression data integration

The structure of the catabolic core model obtained using expression data from intracellular growth is shown in Fig. 6.3.

The catabolic core model is almost identical to the catabolic core model obtained for *L. plantarum* under homolactic fermentation (Fig. 4.1), the main difference between the networks is the production and excretion of lactate in *L. plantarum* and acetate and carbon dioxide in *S.* Typhimurium. Since similar computational methods were used in both cases, this seems to indicate that the method of identifying catabolic cores through variation of ATP demand is highly sensitive to the assignment of objective coefficients to reactions in the FBA problem.

As with the *L. plantarum* catabolic core model, the one obtained here also contained a single enzyme subset, and consequently a single elementary mode, which distinguishes it from the catabolic core of *S.* Typhimurium described in Chapter 3, where all reactions were assigned a uniform objective value. The original catabolic core is more complex, it involves 34 reactions and can be decomposed to four elementary modes. This reduction in complexity in the catabolic core model described here is presumably due to the use of non-uniform objective coefficients, i.e. the precise distribution of objective coefficients is less important than the fact that they are not identical. The composition of the resulting catabolic core is however likely to be determined by the distribution
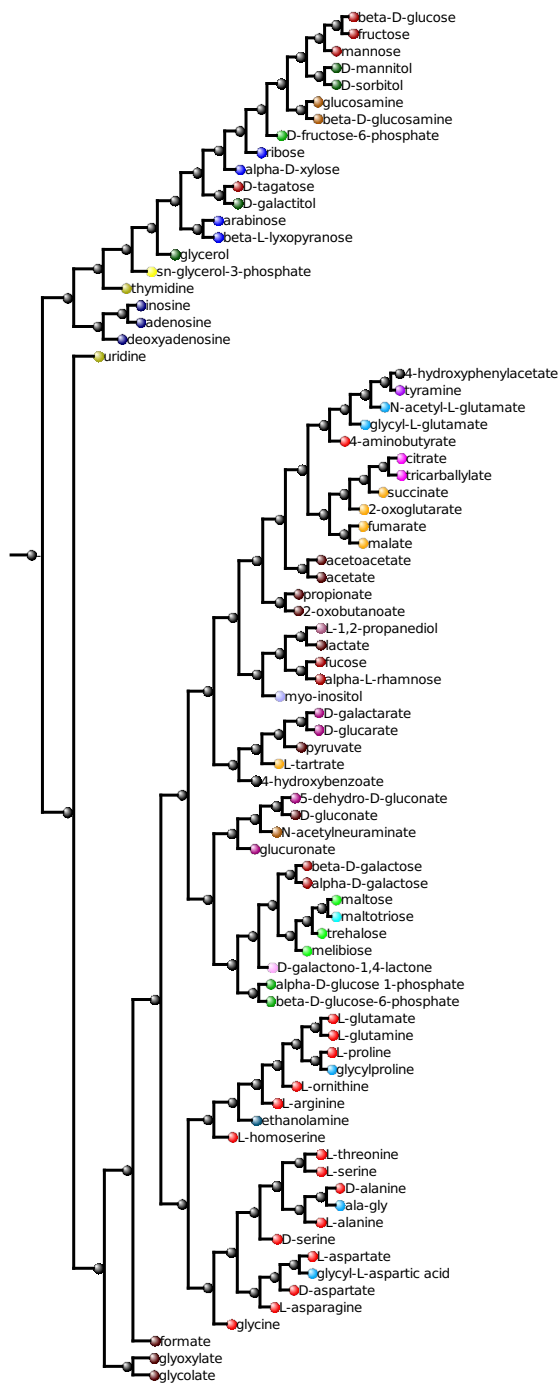
Figure 6.2: Metabolic tree based on compound clustering analysis. Compounds are sorted into the following categories (indicated by node colour): purines (•), amines (•), tricarboxylates (•), sugar acids (•), diols (•), pentoses (•), amino alcohols (•) dipeptides (•), cyclic alcohols (•), lactones (•), trisaccharides (•), disaccharides (•), pyrimidines (•), hexose phosphates (•), sugar alcohols (•), sugar alcohol phosphates (•), dicarboxylates (•), amino acids (•), hexoses (•), amino sugars (•), and monocarboxylates (•). Compounds which did not fit any category are indicated with a black node.
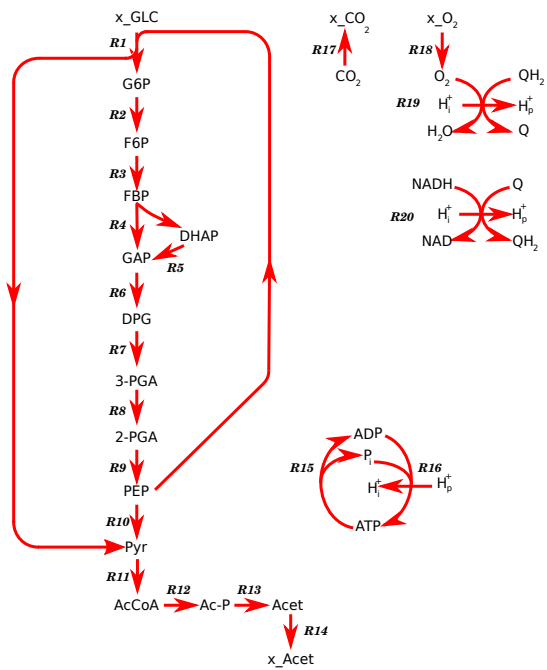
Figure 6.3: Catabolic core network for *S.* Typhimurium obtained using *in vivo* expression data as objective coefficients. Key: R1 - glucose transporter, R2 - PGIsomerase, R3 - PFK, R4 - FBPAldolase, R5 - TriPIsomerase, R6 - GapDH, R7 - PGKin, R8 - PGAM, R9 - Enolase, R10 - PyrKin, R11 - PyrDH, R12 - PhoAcTrans, R13 - AcetKin, R14 - acetate transporter, R15 - ATPase, R16 - ATPSynth, R17 - CO_2 transporter, R18 - O_2 transporter, R19 - CytOx, R20 - NADH_DH.

of objective coefficients.

An interesting observation is that the network obtained here corresponds exactly to one of the acetate producing elementary modes in the original catabolic core, namely EM_4 in Fig. 3.6. As shown in Chapter 3, this is also the mode that the original catabolic core tends towards as the ATP demand increases. Acetate excretion is commonly observed in fast growing *E. coli* in aerobic glucose-based medium, as reviewed in Section 3.4.3. Knowledge relating to the metabolism of *S.* Typhimurium during infection is scarce, which makes independent comparisons with the results obtained here difficult. The results of the integration of *in vivo* expression data and FBA presented here does however support the hypothesis that energy regeneration during macrophage growth occurs through the mode shown in Fig. 6.3.

6.5.3   Combining microarray and PM data for model analysis

The distribution of expression levels over minimal reaction-compound associations is shown in Fig. 6.4, and in more detail as a bar chart in Fig. 6.5. As can be seen in the figures, the compound associated with the highest expression is glycerol-3-phosphate (a sugar alcohol phosphate). Other compounds associated with high expression are 4-aminobutyrate (an amino acid), citrate (a tricarboxylate), and adenosine (a purine). In general, the distribution of compound associated expression does not seem to support consumption of particular groups of compounds, only particular compounds. For instance, it could be argued that the data supports consumption of nuceosides (thymidine, adenosine, inosine, deoxyadenosine, and uridine), since adenosine and thymidine are associated with high expression. On the other hand, inosine is associated with very low expression ($\sim$0) and deoxyadenosine with low to intermediate expression ($\sim$0.2). A group of compounds that deviates somewhat from this pattern are the amino acids. The majority of the amino acids in Figs. 6.2 and 6.4 appear in the same cluster in the lower part of the tree (together with dipeptides). In this cluster, the amino acid with the lowest associated expression is L-ornithine ($\sim$0.2) and the highest is L-aspargine ($\sim$0.6). Note that none of the compounds that are associated with very low expression in this cluster are amino acids according to the categories used in Fig. 6.2 - glycyl-L-aspartic acid and glycylproline are dipeptides and ethanolamine is an amino alcohol.
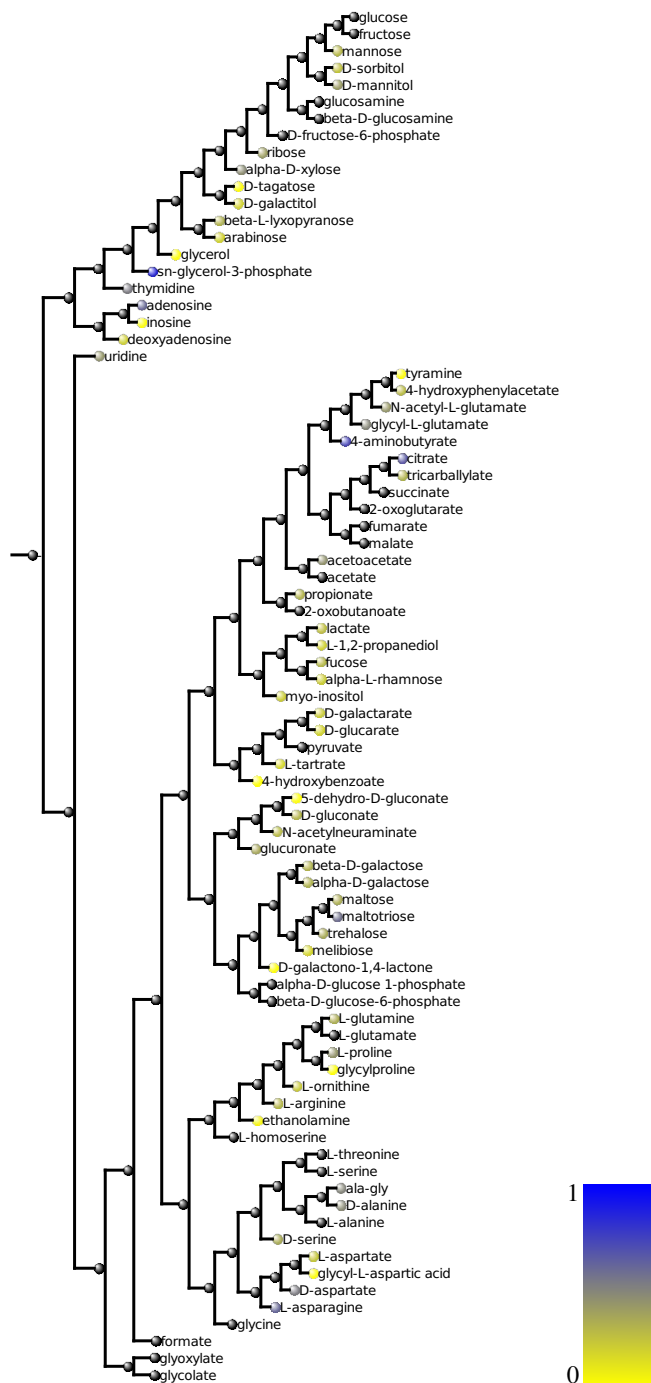
Figure 6.4: Visualisation of expression to compound associations. The metabolic tree shows the scaled relative expression of reactions associated with a given carbon compound, superimposed on the metabolic tree shown in Fig. 6.2. The node colours indicate the scaled relative expression associated with the compound, varying from yellow (indicating the minimum expression level identified in the data set) to blue (indicating the maximum expression level). Black nodes indicate compounds that could not be associated with an expression level.
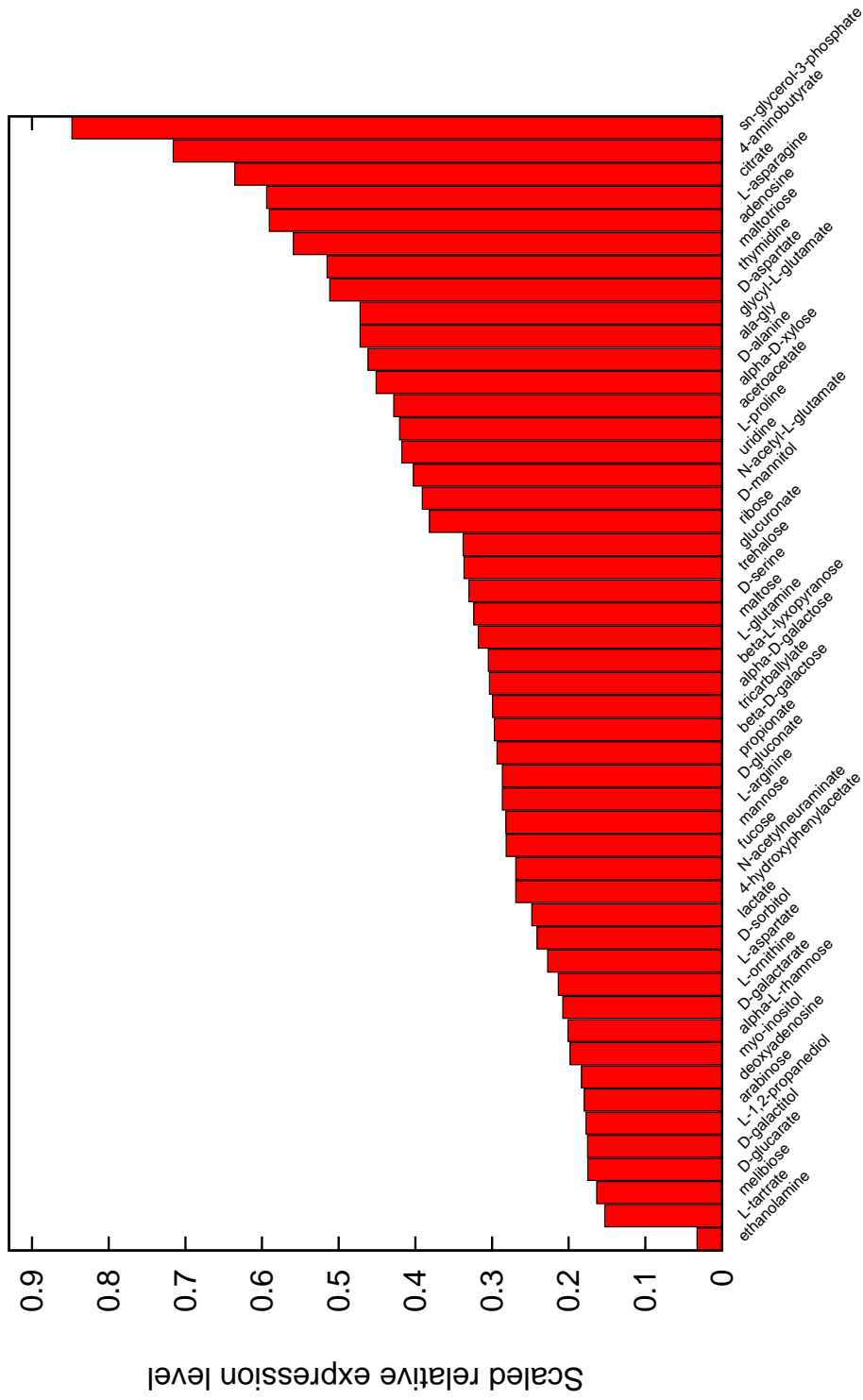
Figure 6.5: Bar chart relating reaction-compound associations (carbon sources) with relative expression levels. The chart is based on the same data as the metabolic tree in Fig. 6.4. Compounds associated with zero expression (white nodes in the tree) or where no change in expression level could be associated with the compound (black nodes in the tree) are not included in the chart.

6.6 Discussion

The motivation for developing the reaction-compound simplification method was to simplify the analysis of expression data in the context of metabolism, and more specifically to improve interpretations of carbon source utilisation from DNA microarray data. This reliance on expression data offers some important benefits, but also introduces some restrictions on the conclusions that can be drawn from the analysis. Expression data from DNA microarrays is genome-scale, the implication of this, in conjunction with PM data, for the analysis presented here is that all possible compounds that can be catabolised by an organism, are within the scope of the analysis. This is in contrast to methods based on carbon labelling, such as ^{13}C-isotopologue profiling (described in more detail below), where the scope of the analysis is limited to a selected set of metabolites. DNA microarray technology is just one method for expression analysis, a promising alternative method is RNA sequencing (RNA-seq). In brief, RNA-seq involves conversion of extracted RNA into a cDNA library, followed by deep sequencing (using some high-throughput sequencing technology) [165]. One of the benefits of RNA-seq is that it allows quantification of transcripts of low abundance, which implies that pathogen transcripts can be quantified even in situations where host transcripts are in much greater abundance [166]. This would in principle make the method applicable to expression data obtained from pathogens growing inside multicellular host organisms, in contrast to data obtained from cultured host cells. This type of analysis is possible with microarray based methods as well, but the quality of the low abundance pathogen expression data is very limited. There are currently no reported applications of methods based on carbon labelling which allows probing pathogen catabolism inside intact host organisms. The main drawback of the method presented here is inherent to all expression based analysis: Expression data relating to metabolic enzyme does not necessarily indicate enzyme abundance or carbon flux through a particular reaction. The implication of high expression rates of enzymes involved in catabolism of a particular compound could be that the compound is consumed. Alternatively, the implication could be that the intracellular environment is so variable in nutrient availability that the pathogen has evolved regulatory mechanisms to allow quick consumption of a broad set of compounds.

Direct experimental data related to the metabolism of *S.* Typhimurium during infection is difficult to obtain. One important contribution to this field is the study by Götz et al. [167], where a method for partial characterisation of the carbon catabolism during intracellular growth was applied to *S.* Typhimurium during infection of cultured human epithelial (Caco-2) cells. Although this

method was applied to epithelial, rather than macrophage, cells, microarray analysis indicate that expression of metabolic genes by *S*. Typhimurium grown in the two types of cell cultures do not differ much [168]. Briefly, the method, referred to as ^{13}C-*isotopologue profiling* [167, 169], is based upon exposing infected host cells to growth media containing (uniformly labelled) ^{13}C-glucose shortly after the onset of infection, followed by isolation of the bacterial cells and analysis of the labelling profile of bacterial biomass. Unlike Metabolic Flux Analysis (MFA), ^{13}C-isotopologue profiling does not result in detailed flux distributions, but based upon the distribution of labelled carbon in amino acids, some insights about the origin of the carbon can gained: E.g. if a given amino acid is enriched in labelled carbon this is an indication that the carbon atoms were derived from the labelled growth medium, rather than directly from the host cell. The results presented by Götz et al. [167] indicate that the main carbon sources used by *S*. Typhimurium during the epithelial stage of infection are glucose and most amino acids. Similarly, expression analyses of *S*. Typhimurium in both epithelial and macrophage cells have shown that transporters associated with glucose, branched-chain amino acids, and glycerol-3-phosphate are over-expressed [168, 170]. Previous studies have suggested, using auxotrophic *S*. Typhimurium strains, that purins, pyrimidines, and some amino acids (histidine and methionine) are synthesised *de novo* during systemic mouse infection [26, 171].

These findings could be considered to contradict some of the results presented here. However, it is important to emphasise the following drawbacks of the method: (*i*) Carbon sources that cannot be associated with unique catabolic reaction (or where these reactions cannot be identified indirectly through reaction-compound simplification) cannot be associated with an expression level. Hence, even if reactions associated with glucose catabolism are over-expressed they cannot be uniquely identified, as indicated in Fig. 6.4; (*ii*) The method suggested here is focused on finding experimental support for the *catabolism* of a given metabolite, i.e. if a biomass precursor, originating from the host cell, is incorporated in the biomass of the pathogen without any modifications, this metabolite would not be identified.

A previously published method, Differential Producibility Analysis (DPA) [172], bears some resemblance to the method suggested here. It addresses the related problem of associating expression data with production of a given metabolite. DPA is based upon identifying reactions (and in extension, genes) that are essential for biosynthesis of all metabolites defined in a GSM, using FBA and known gene-to-reaction associations of the organism of interest. All genes associated with a given metabolite are then separated into up- or down-regulated genes and the median signal for each metabolite (for both up- and

down-regulated genes) is calculated. Then, a statistical test, Rank Product Analysis [173], is used to rank the metabolites based upon the average microarray signal intensity over a set of experimental conditions. DPA was successfully applied to *Mycobacterium tuberculosis* expression data, isolated under various conditions (macrophage, human sputum, and a range of *in vitro* media). Bonde et al. [172] reported up-regulation of genes associated with biosynthesis of cell envelope components and aromatic amino acids, and down-regulation of central carbon metabolism, biosynthesis of (non-aromatic) amino acids and sugars. Although the metabolic behaviours of *S.* Typhimurium and *M. tuberculosis* during macrophage infection are not readily comparable, results presented by Bonde et al. [172], as well as those presented here and by Götz et al. [167], seems to indicate the availability of some amino acids in both the macrophage and epithelial environments.

A potentially useful continuation of the analysis presented here could be to complement it with DPA. DPA can be applied to identify metabolites that are being synthesised by an organism (that would be difficult to identify by other methods), whereas the method suggested here is intended for analysis of catabolism. A method that combines the two approaches would potentially give a more complete understanding of the media availability and the metabolic activity of an intracellular pathogen. Validation of the method presented here would be useful, but a test data set for this purpose has not been obtained to date.

Chapter 7

Damage analysis of *Salmonella* Typhimurium metabolism

7.1 Introduction

A major motivation for constructing and analysing metabolic models of pathogenic bacteria is to understand how their metabolism functions during infection, and in extension to prevent or terminate infection by disrupting their metabolism. The focus of this chapter is to identify sites of vulnerability in the metabolic network of *S.* Typhimurium that are relevant for its metabolism during infection. Specifically, the aim is to identify sets of reactions whose removal from the network will interfere with its performance under infection-relevant conditions. These reactions could be of interest from a pharmaceutical viewpoint, but since experimental verification of the importance of these reactions for infection has not been pursued owing to time constraints, this cannot be confirmed.

As reviewed in Section 2.6, several approaches for identifying potential drug targets using GSMs of pathogenic bacteria have been suggested previously. In the case of *S.* Typhimurium, the outcome in terms of clinically relevant drug targets or therapeutic strategies have so far been limited: Thiele et al. [121] identified 56 potential synthetic lethal gene pairs using FBA-based approach (double gene deletion analysis, described in Section 2.6) assuming LB growth medium (here assumed to consist of all amino acids, RNA and DNA monomers, and vitamins required for biomass synthesis). Previous work by Becker et al. [8] showed that approximately 150 metabolic enzymes are essential during *Salmonella* infection of mice, however most of these enzymes are already, or have been consider

as, targets for antimicrobials. It was suggested that a high degree of metabolic redundancy in *S.* Typhimurium resulted in a low proportion of single essential enzymes, and thus, that pairs (or larger sets) of enzymes that carried out essential metabolic functions could be used as drug targets [8]. Since identifying these pairs experimentally would require construction and characterisation of approximately 500 000 mutant strains, the model-based approach described in Thiele et al. [121] was used. There are a few potential problems with this approach: As discussed in Chapter 6 (Section 6.5.3), the nutrient availability during infection is poorly characterised, but since systemic infection in the mouse model requires that certain biomass components are synthesised *de novo* (e.g. purines, pyrimidines, histidine and methionine), LB medium is too rich to accurately model the nutrient availability of *S.* Typhimurium during infection. Solutions to FBA problems are highly dependent upon the media components assumed to be available to the organism being modelled. Consequently, if FBA based identification of synthetic lethal genes is carried out assuming a certain medium, the analysis will only be valid for that medium, or a subset of it. In the case of modelling a medium of unknown composition by assuming a composition which is definitely more complex, the analysis will underestimate the number of synthetic lethal genes. This is because the network redundancy will be obscured by the media redundancy. For example, consider an organism that can synthesise an essential biomass precursor from simpler metabolites by two different routes. Genes associated with reactions in these routes would constitute synthetic lethal pairs. If, however, the essential precursor was (falsely) assumed to be available in the medium, these pairs would not be detected with the FBA-based method.

Furthermore, by restricting the analysis to pairs of genes, rather than reactions, targets that would be relevant drug targets (i.e. involving two reactions) could be missed because more than two genes are associated with the reactions.

The approach described in this chapter to overcome some of these problems is to focus on the subset of metabolism that is involved with generation of energy and precursors for synthesis of biomass components. The rationale for this is that the ability to maintain these functions is essential for an organism, regardless of the detailed compositions of the medium and the biomass. As showed in Chapter 3 the catabolic core of a GSM can be identified by simulating a variation of energy demand (modelled by a generic ATPase reaction), whilst maintaining a fixed rate of biomass synthesis. A strength of this method is that the catabolic core is extracted from a realistic whole cell system and based upon its function within that system. By applying this method to all growth supporting carbons sources, as determined in Chapter 6, and constructing a consensus model from the catabolic cores, a global catabolic core model for *S.* Typhimurium is obtained. Owing to the relative simplicity of this model

(compared to a GSM) exhaustive damage analysis can be applied.

7.2   Methods

7.2.1   Catabolic core extraction

Equations 3.2 and 6.1 were combined in order to generate the flux data from which the global catabolic core was extracted:

$$
\begin{aligned}
\text{minimize} \quad &: \quad \sum_{i=1}^{m} |v_i| \cdot c_i \\
\text{subject to} \quad &\left\{
\begin{array}{l}
\mathbf{N}_{n,m} \cdot \mathbf{v} = \mathbf{0} \\
v_j = t_j; \ k \leq j \leq m \\
v_e = 0; \ f \leq e \leq h, \ e \neq g \\
v_{\text{ATPase}} = J_{\text{ATPase}}; \ J_{min} \leq J_{\text{ATPase}} \leq J_{max}
\end{array}
\right.
\end{aligned}
\tag{7.1}
$$

i.e. for each transporter of a growth-supporting carbon source (reactions f through h in $\mathbf{v}$) the constraint to carry zero flux was removed and the ATPase reactions was sequentially set to increasing values in the range $[J_{min}, \ J_{max}]$. As in Equations 3.2 and 6.1, the objective function was minimisation of total flux and the objective coefficients, c_i, were set to 1. Parameters J_{min} and J_{max} were used as previously (set to 63 and 350 mmol (g DW)$^{-1}$h^{-1}, respectively). For each carbon source, a matrix $\hat{\mathbf{S}}$ of reactions responsive to changes in energy demand (as defined in Equations 3.4 – 3.6) was created.

In order to remove unnecessary complexity from the global catabolic core, reactions that were only associated with a single carbon source (excluding transporters), as well as pairs of reactions that formed net-transhydrogenases, were removed. This was done by constructing an initial model of all responding reactions. Using the complete set of source specific responses (i.e. the set of $\hat{\mathbf{S}}$ matrices) all non-transport reactions unique for a given source were removed from the initial model. Null-space based methods (see Sections 3.2.3 and 2.2) were used to identify reactions that were rendered dead by the removal of the source specific ones, and by condensing the model to net-reactions, reactions that could only carry flux as a net-transhydrogenase were also identified and removed. These reactions were subsequently removed from the corresponding $\hat{\mathbf{S}}$ matrix. The procedure is outlined in Fig. 7.1.

This simplification was motivated by the limited contribution of reactions associated with few carbon sources to the global catabolic core, although their contribution to the source specific cores in which they appear is crucial. Similarly, net-transhydrogenases may provide a crucial function to a source specific catabolic core, but in the global core they will be functionally redundant.
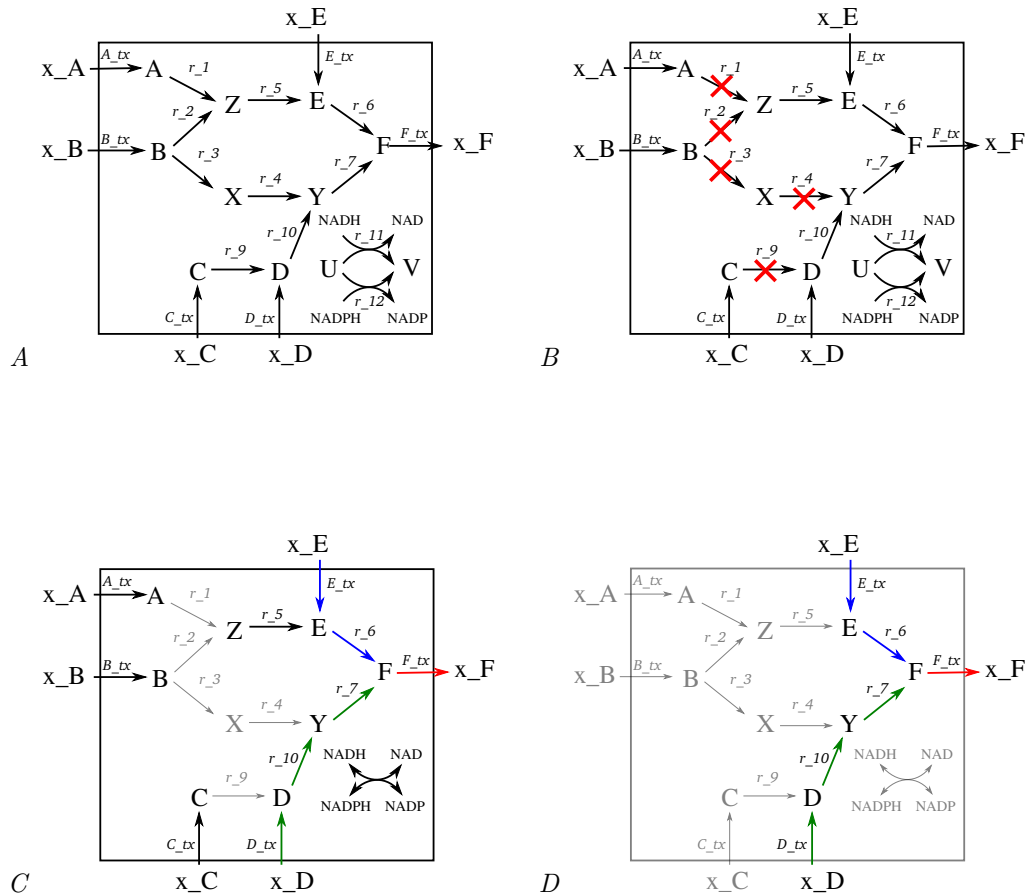
Figure 7.1: Illustration of catabolic core simplification. An initial model (A) is modified by removing all internal reactions asociated with single carbon sources (B). After removal of the reactions in B, the remaining model is condensed into enzyme subsets (C). The final step (D) involves removal of reactions that are dead (here transporters A_tx and B_tx, and internal reaction r_5) or condense into net–transhydrogenases (r_11 and r_12).

7.2.2 Flux correlation analysis

In order to investigate the flux correlations between the reactions in the global catabolic core, analysis similar to that described previously (Section 3.2.3) was used. The source specific flux response matrices, $\hat{\mathbf{S}}$, were combined in the matrix $\hat{\mathbf{Z}}$:

$$\hat{\mathbf{Z}}_{p,f} = \begin{matrix} v_{\hat{x},J_{min},tx_1} & \cdots & v_{\hat{z},J_{min},tx_1} \\ \vdots & \ddots & \vdots \\ v_{\hat{x},J_{max},tx_1} & \cdots & v_{\hat{z},J_{max},tx_1} \\ v_{\hat{x},J_{min},tx_2} & \cdots & v_{\hat{z},J_{min},tx_2} \\ \vdots & \ddots & \vdots \\ v_{\hat{x},J_{max},tx_l} & \cdots & v_{\hat{z},J_{max},tx_l} \end{matrix} \qquad (7.2)$$

where p is the product of the number of carbon sources used (l) and the number of flux solutions (rows) in each $\hat{\mathbf{S}}$ matrix, and f is the number of reactions in the global catabolic core. Each carbon source is associated with a transporter, tx_n. As previously, Pearson's correlation coefficient ($r(X,Y)$, Equation 3.7) was calculated for all combination of the columns in $\hat{\mathbf{Z}}$, as indicated in Equation 3.8, and a metabolic tree generated from the resulting correlation matrix.

7.2.3 Damage analysis

The procedure for identifying suitable candidate reactions involved the following steps: (i) All single reactions in the global catabolic core were sequentially removed from the complete GSM (with all growth supporting carbon sources available) and its ability to produce biomass was quantified; (ii) Reactions that did not impose damage on the GSM were combined in pairs and the procedure of quantifying the damage on the GSM repeated; (iii) The procedure was then repeated with the remaining reactions by combining them in sets of three. This was carried out by repeatedly solving the LP:

$$\begin{aligned} \text{minimize} \quad &: \quad \sum_{i=1}^{m} |v_i| \cdot c_i \\ \text{subject to} \quad &\begin{cases} \mathbf{N}_{n,m} \cdot \mathbf{v} = \mathbf{0} \\ v_j = t_j; \ k \le j \le m \\ v_{ATPase} = r_{ATP} \\ v_i, v_j, v_k = 0; \ i,j,k \in \{1 \dots l\} \end{cases} \end{aligned} \qquad (7.3)$$

where $\mathbf{N}_{n,m}$ is the stoichiometry matrix of the GSM. Assuming that the stoichiometry matrices of the GSM and the global catabolic core model have the same column order (i.e. the first l reactions of $\mathbf{N}$ are the reactions of the catabolic core), v_i, v_j, and v_k represent all combinations of these shared

reactions (including cases where $i = j = k$ and $i = j \neq k$). For each feasible flux solution the objective value, i.e. the sum of fluxes, was recorded. All reaction sets that rendered the LP unfeasible or caused an increase in objective value of more than 25% (compared to the solution with no reactions fixed to zero) were considered damaging and not included in larger reaction sets for further analysis.

In order to asses the contribution of individual reactions in the global catabolic core to the damage imposed on the GSM two metrics were defined: Damage coefficient (D_r) and relative frequency (F_r). The damage coefficient is the average, normalised damage associated with a given reaction, and the relative frequency is the ratio of the number of occurrences of a given reaction and maximum number of occurrences of any reaction in the set. The output of Equation 7.3 is a set of sets of reactions, $\{\{r_x, \ldots, r_z\}, \ldots, \{r_{\hat{x}}, \ldots, r_{\hat{z}}\}\}$, where reaction set can be associated with a damage, $\{r_x, \ldots, r_z\} \sim d_1, \ldots, \{r_{\hat{x}}, \ldots, r_{\hat{z}}\} \sim d_n$. In order to express a given reactions contribution to the damage of a particular set of reactions, this association can be expressed as the ratio of the damage and the cardinality (i.e. number of members, symbolised with $|\ldots|$) of the set, $r_x \sim \frac{d_1}{|\{r_x, \ldots, r_z\}|}$. For any reaction that appears in more than one reaction set this yields a set of ratios, collected in q, e.g.:

$$r_x \sim \left\{ \frac{d_1}{|\{r_x, \ldots, r_z\}|}, \ldots, \frac{d_m}{|\{r_x, \ldots, r_b\}|} \right\} = q_x$$

The damage coefficient of any reaction is then defined as:

$$D_{r_x} = \frac{\sum_{i \in q_x} \frac{i}{d_{max}}}{|q_x|} \tag{7.4}$$

where d_{max} is the maximum damage possible, and $|q_x|$ is the cardinality of the set q_x. Since the maximum damage is an unfeasible solution, d_{max} is set to an arbitrarily high value, 1000. The implication of this definition of D is that an essential reaction will have a coefficient of 1.0 and a reaction that always appear in essential sets of size two will have a coefficient of 0.5, etc.

Based upon the definition of q, F_r was defined as:

$$F_{r_x} = \frac{|q_x|}{max(|q_1|, \ldots, |q_n|)} \tag{7.5}$$

where $|q_x|$ represents the cardinality of q_x and $max(|q_1|, \ldots, |q_n|)$ is the maximum cardinality of any set in q. Unlike the standard definition of relative frequency, where absolute frequency of an item is compared with the total number of items, the frequency of a given reaction was compared to the most frequent reaction in the set, i.e. $max(|q_1|, \ldots, |q_n|)$ in Equation 7.5.

7.3 Results

7.3.1 Structure of the global catabolic core

Initially, the global catabolic core included 342 reactions. Using the simplification described in Section 7.2.1 the size was reduced to 173 reactions. Using condensation (see Section 3.2.3) the size could be reduced further to 128 reactions. This simplification also included reducing the number of carbon sources from 81 to 34. The metabolic tree based on correlations calculated using Equation 7.2 is shown in Fig. 7.2.

The structure of the correlation tree in Fig. 7.2 indicates that although most of the carbon source transporters are closely correlated with some internal reactions in the catabolic core, a significant minority do not correlate well with any internal reactions. Similarly, some internal reactions are not strongly correlated with any carbon sources. As shown in Fig. 7.2, the enzyme subsets of the glucose-based catabolic core (Chapter 3) are, in most cases, not maintained in the global catabolic core. Reactions that form subsets in the glucose core are closely, but not completely, correlated in the global core.

7.3.2 Damage analysis

Damage analysis identified 179 sets of reactions that caused an increase in objective value of 25% or more. Most of these sets included three reactions (164 sets), and only a small number (15 sets) were reaction pairs. No single reactions from the global catabolic core caused damage above the fixed threshold. These reaction sets that causes damage above the threshold of 25% involved a total of 65 reactions, 13 of which were transporters. The distribution of damage coefficients, as defined in Equation 7.4, over these reactions is shown in Fig. 7.4, and in more detail in Fig. 7.6. As evident in Fig. 7.6, most of the reactions with a damage coefficient of 0.33 or above, indicating participation in a lethal set of two or three reactions, were primarily involved in metabolism of amino acids, and thus not found in the glucose-based catabolic core. It should be emphasised that since synthesis of biomass components is used as a constraint in the damage analysis, reactions that, when removed from the network, interfere with this process will be assigned high damage coefficients. As mentioned in the introduction to this chapter, a motivation for focusing the damage analysis on the global catabolic core was to identify sets of reactions that were involved in generation of energy and precursors for biomass synthesis, without being too limited by the precise composition of the biomass, since this is not constant. Hence, reactions with high damage coefficients could carry a high damage *in vivo* if removed from the network, if the assumption concerning biomass com-
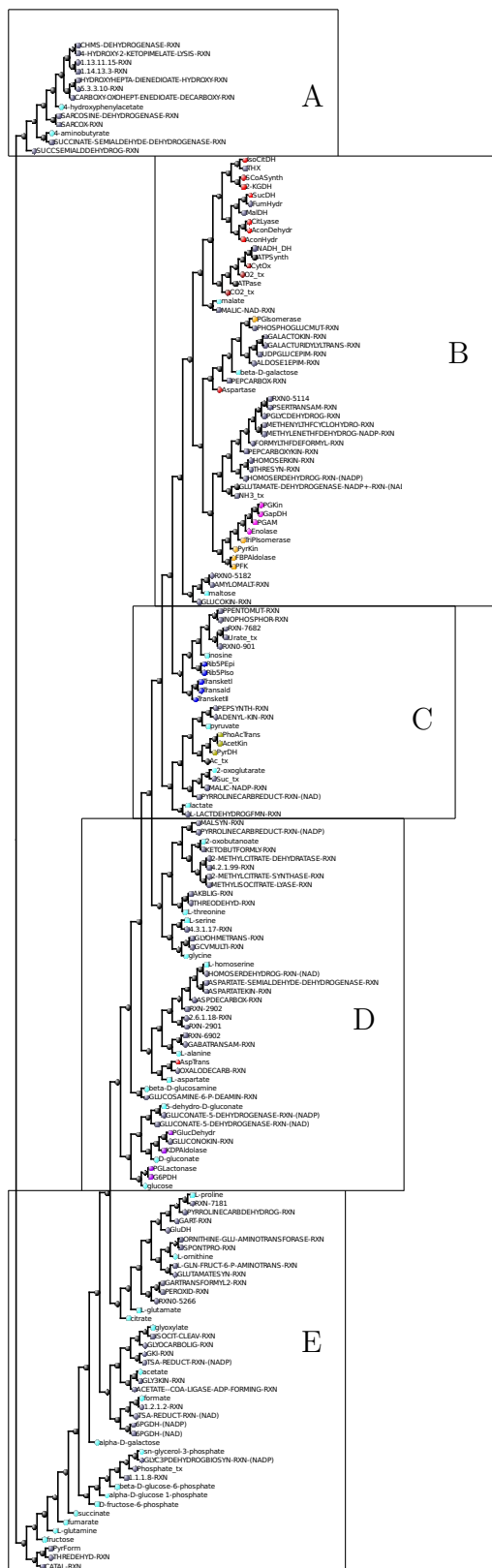
Figure 7.2: Flux correlations of the global catabolic core. Reactions that are also present in the glucose-based catabolic core are indicted by the node colour being consistent with the arrowhead colour in Fig. 3.2. Transporters of carbon sources are indicated by turquoise (●) nodes. Other reactions are indicated by gray (●) nodes. Also indicated in the tree are the reactions of the non-oxidative pentose phosphate pathway, with blue (●) nodes. Clusters A-E are shown in more detail in Fig. 7.3.

A

- CHMS-DEHYDROGENASE-RXN
- 4-HYDROXY-2-KETOPIMELATE-LYSIS-RXN
- 1.13.11.15-RXN
- 1.14.13.3-RXN
- HYDROXYHEPTA-DIENEDIOATE-HYDROXY-RXN
- 6.3.3.10-RXN
- CARBOXY-OXOHEPT-ENEDIOATE-DECARBOXY-RXN
- 4-hydroxyphenylacetate
- SARCOSINE-DEHYDROGENASE-RXN
- SARCOX-RXN
- 4-aminobutyrate
- SUCCINATE-SEMIALDEHYDE-DEHYDROGENASE-RXN
- SUCCSEMIALDDEHYDROG-RXN

B

- IsoCitDH
- THX
- 5CoASynth
- 2-KGDH
- SucDH
- FumHydr
- MalDH
- CitLyase
- AconDehydr
- AconHydr
- NADH_DH
- ATPSynth
- CytOx
- O2_tx
- ATPase
- CO2_tx
- malate
- MALIC-NAD-RXN
- PGIsomerase
- PHOSPHOGLUCMUT-RXN
- GALACTOKIN-RXN
- GALACTURIDYLYLTRANS-RXN
- UDPGLUCEPIM-RXN
- ALDOSE1EPIM-RXN
- beta-D-galactose
- PEPCARBOX-RXN
- Aspartase
- RXN0-5114
- PSERTRANSAM-RXN
- PGLYCDEHYDROG-RXN
- METHENYLTHFCYCLOHYDRO-RXN
- METHYLENETHFDEHYDROG-NADP-RXN
- FORMYLTHFDEFORMYL-RXN
- PEPCARBOXYKIN-RXN
- HOMOSERKIN-RXN
- THRESYN-RXN
- HOMOSERDEHYDROG-RXN-(NADP)
- GLUTAMATE-DEHYDROGENASE-NADP+-RXN-(NAD)
- NH3_tx
- PGKin
- GapDH
- PGAM
- Enolase
- TriPIsomerase
- PyrKin
- FBPAldolase
- PFK
- RXN0-5182
- AMYLOMALT-RXN
- maltose
- GLUCOKIN-RXN

C

- PPENTOMUT-RXN
- INOPHOSPHOR-RXN
- RXN-7682
- Urate_tx
- RXN0-901
- inosine
- Rib5PEpi
- Rib5PIso
- Transketl
- Transald
- Transketll
- PEPSYNTH-RXN
- ADENYL-KIN-RXN
- pyruvate
- PhoAcTrans
- AcetKin
- PyrDH
- Ac_tx
- 2-oxoglutarate
- Suc_tx
- MALIC-NADP-RXN
- PYRROLINECARBREDUCT-RXN-(NAD)
- lactate
- L-LACTDEHYDROGFMN-RXN

D

- MALSYN-RXN
- PYRROLINECARBREDUCT-RXN-(NADP)
- 2-oxobutanoate
- KETOBUTFORMLY-RXN
- 2-METHYLCITRATE-DEHYDRATASE-RXN
- 4.2.1.99-RXN
- 2-METHYLCITRATE-SYNTHASE-RXN
- METHYLISOCITRATE-LYASE-RXN
- AKBLIG-RXN
- THREODEHYD-RXN
- L-threonine
- L-serine
- 4.3.1.17-RXN
- GLYOHMETRANS-RXN
- GCVMULTI-RXN
- glycine
- L-homoserine
- HOMOSERDEHYDROG-RXN-(NAD)
- ASPARTATE-SEMIALDEHYDE-DEHYDROGENASE-RXN
- ASPARTATEKIN-RXN
- ASPDECARBOX-RXN
- RXN-2902
- 2.6.1.18-RXN
- RXN-2901
- RXN-6902
- GABATRANSAM-RXN
- L-alanine
- AspTrans
- OXALODECARB-RXN
- L-aspartate
- beta-D-glucosamine
- GLUCOSAMINE-6-P-DEAMIN-RXN
- 5-dehydro-D-gluconate
- GLUCONATE-5-DEHYDROGENASE-RXN-(NADP)
- GLUCONATE-5-DEHYDROGENASE-RXN-(NAD)
- PGlucDehydr
- GLUCONOKIN-RXN
- KDPAldolase
- D-gluconate
- PGLactonase
- G6PDH
- glucose

E

- L-proline
- RXN-7181
- PYRROLINECARBDEHYDROG-RXN
- GART-RXN
- GluDH
- ORNITHINE-GLU-AMINOTRANSFORASE-RXN
- SPONTPRO-RXN
- L-ornithine
- L-GLN-FRUCT-6-P-AMINOTRANS-RXN
- GLUTAMATESYN-RXN
- GARTRANSFORMYL2-RXN
- PEROXID-RXN
- RXN0-5266
- L-glutamate
- citrate
- glyoxylate
- ISOCIT-CLEAV-RXN
- GLYOCARBOLIG-RXN
- GKI-RXN
- TSA-REDUCT-RXN-(NADP)
- acetate
- GLY3KIN-RXN
- ACETATE--COA-LIGASE-ADP-FORMING-RXN
- formate
- 1.2.1.2-RXN
- TSA-REDUCT-RXN-(NAD)
- 6PGDH-(NADP)
- 6PGDH-(NAD)
- alpha-D-galactose
- sn-glycerol-3-phosphate
- GLYC3PDEHYDROGBIOSYN-RXN-(NADP)
- Phosphate_tx
- 1.1.1.8-RXN
- beta-D-glucose-6-phosphate
- alpha-D-glucose 1-phosphate
- D-fructose-6-phosphate
- succinate
- fumarate
- L-glutamine
- fructose
- PyrForm
- THREDEHYD-RXN
- CATAL-RXN

Figure 7.3: Enlarged clusters from Fig. 7.2.

114

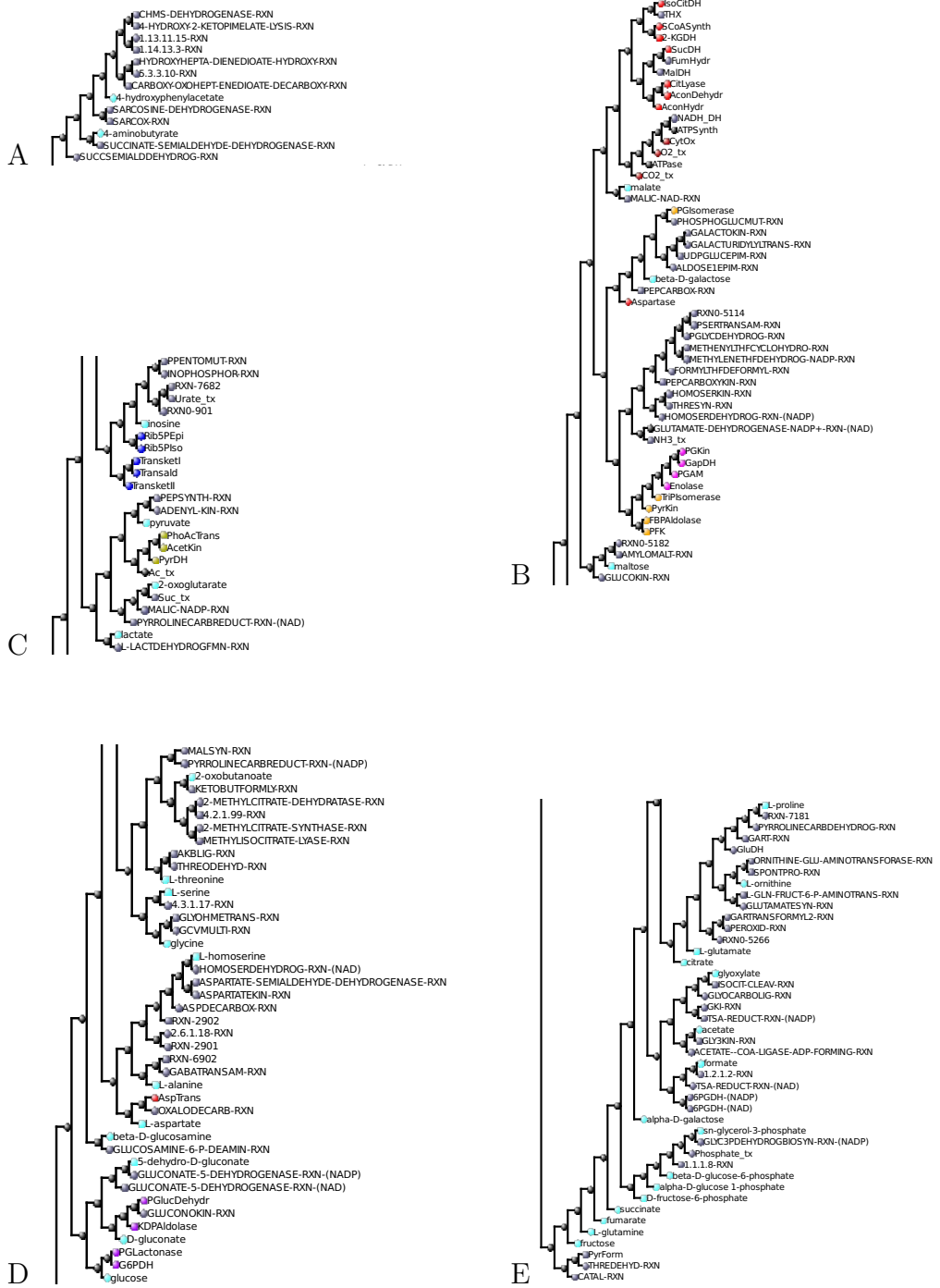position are accurate. Reactions with intermediate damage coefficients could be less damaging *in vivo*, but are also less biased by the assumptions regarding which biomass components are being synthesised.

The distribution of relative frequency is shown in Figs. 7.7 and 7.9. Unlike the distribution of damage coefficients, the reactions with high relative frequencies are dominated by reactions also found in the glucose-based catabolic core. Two reactions in particular have very high relative frequencies, ATP synthase (ATPSynth) and the carbon dioxide transporter (CO2_tx), indicating that both participate in a high proportion of the reaction sets, but as can be see in Fig. 7.6, both reactions have modest damage coefficients (below 0.1). This is a general pattern observed for the global catabolic core, and not an intrinsic property of the two metrics. Although in the extreme case of essential reactions, these will have a damage coefficient of 1 and only occur in one (singleton) reaction set, but for larger reactions sets only the network structure determines the relationship between the damage coefficient and relative frequency for a given reaction. For example, if an essential function in a metabolic network can be carried out by any of two equivalent elementary modes, then in order to abolish this function both modes must be removed, which can be done by removing any combination of one reaction from each mode. A reaction from any of these modes will have a damage coefficient of 0.5, but a relative frequency proportional to the size of the other mode. The complete set of damaging reactions are collected in Appendix B and C.

7.4 Discussion

The damage analysis highlights the essentiality of the pentose phosphate pathway. Out of the 179 reaction sets, 42 contained reactions from the non-oxidative pentose phosphate pathway (Rib5PIso, Rib5PEpi, TransketI, TransketII, and Transald). A majority of these sets were non-lethal (31), but 11 of the sets caused lethality in the GSM. All lethal sets involved at least two of the listed reactions, but owing to their involvement in non-lethal sets as well, their damage coefficients are relatively moderate: Rib5PIso - 0.1, Transald - 0.16, TransketI - 0.14, and TransketII - 0.16. The exception to this pattern is Rib5PEpi with a coefficient of 0.5. It is worth noting that owing to the high degree of redundancy of the genes encoding the enzymes catalysing these reactions (each reaction is associated with at least two genes: Rib5PIso - 2 genes; Transald - 3 genes; TransketI - 4 genes; TransketII - 4 genes; Rib5PEpi - 4 genes) this set would be difficult to identify with approaches focused on single and double gene knockouts.

The essentiality of some of the reactions identified in the analysis presented
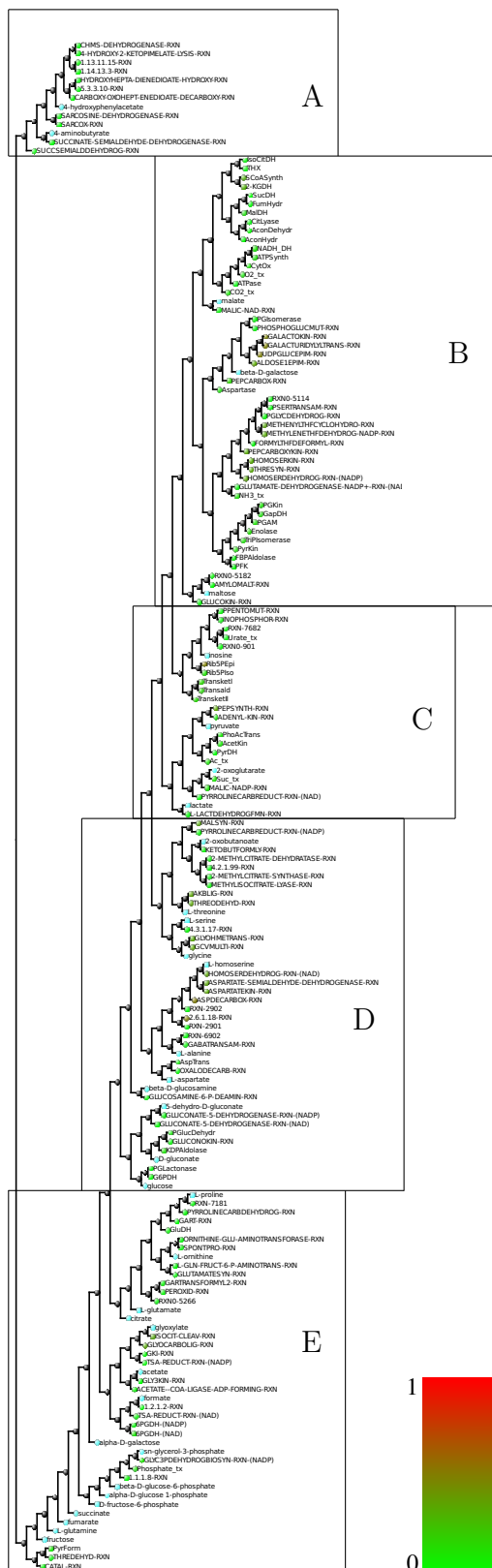
Figure 7.4: Distribution of damage over the global catabolic core. Damage coefficients (Equation 7.4) of reactions in the catabolic core are indicated by the gradient from red (indicating damage coefficient of 1) (●) to bright green (indicating a coefficient of 0) (●) (see scale). As in Fig. 7.2, carbon source transporters are indicated with turquoise nodes, and were not assigned damage coefficients. Clusters A-E are shown in more detail in Fig. 7.5. The damage coefficients are shown in more detail as a bar chart in Fig. 7.6.
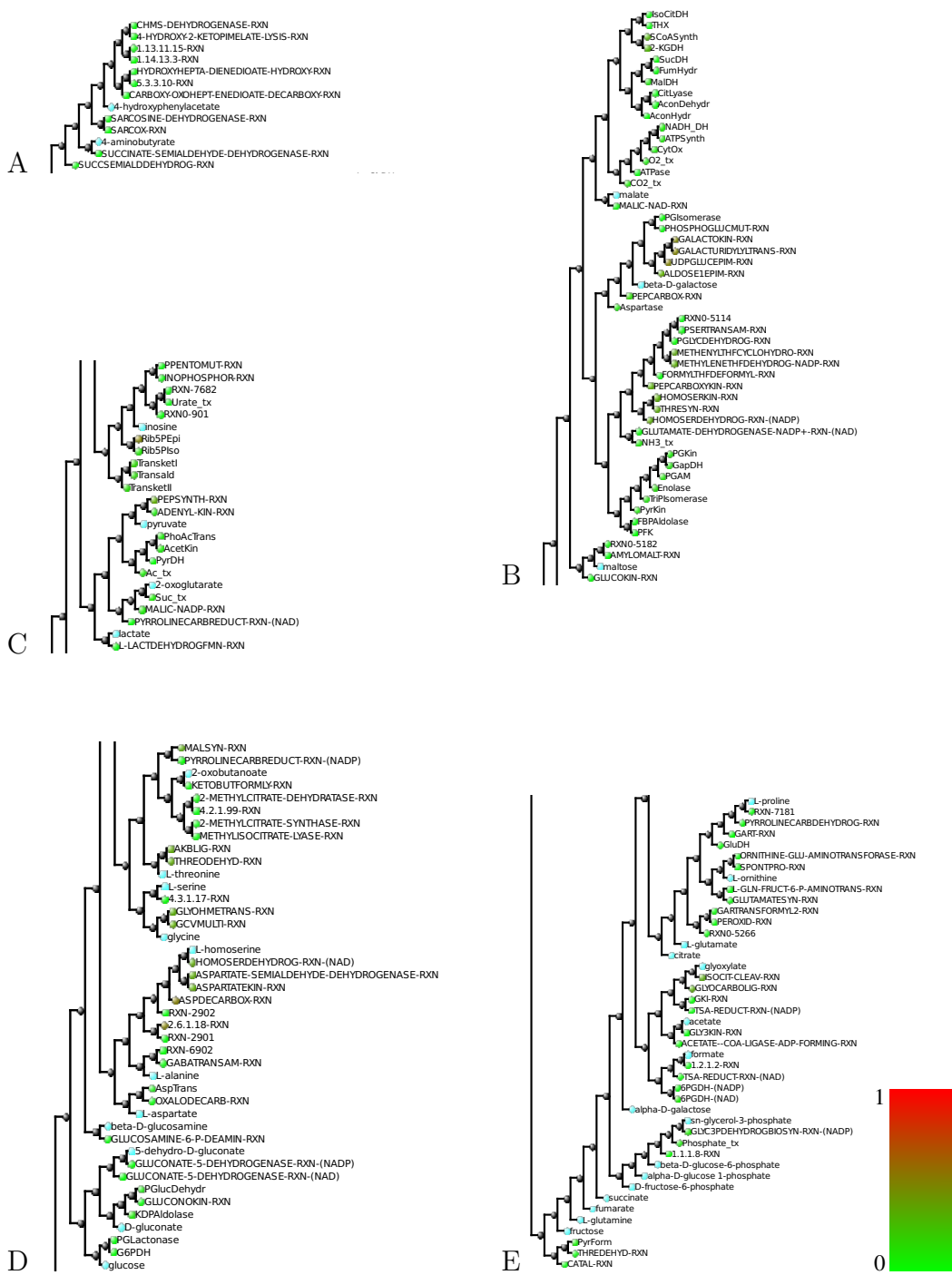
Figure 7.5: Enlarged clusters from Fig. 7.4.

117
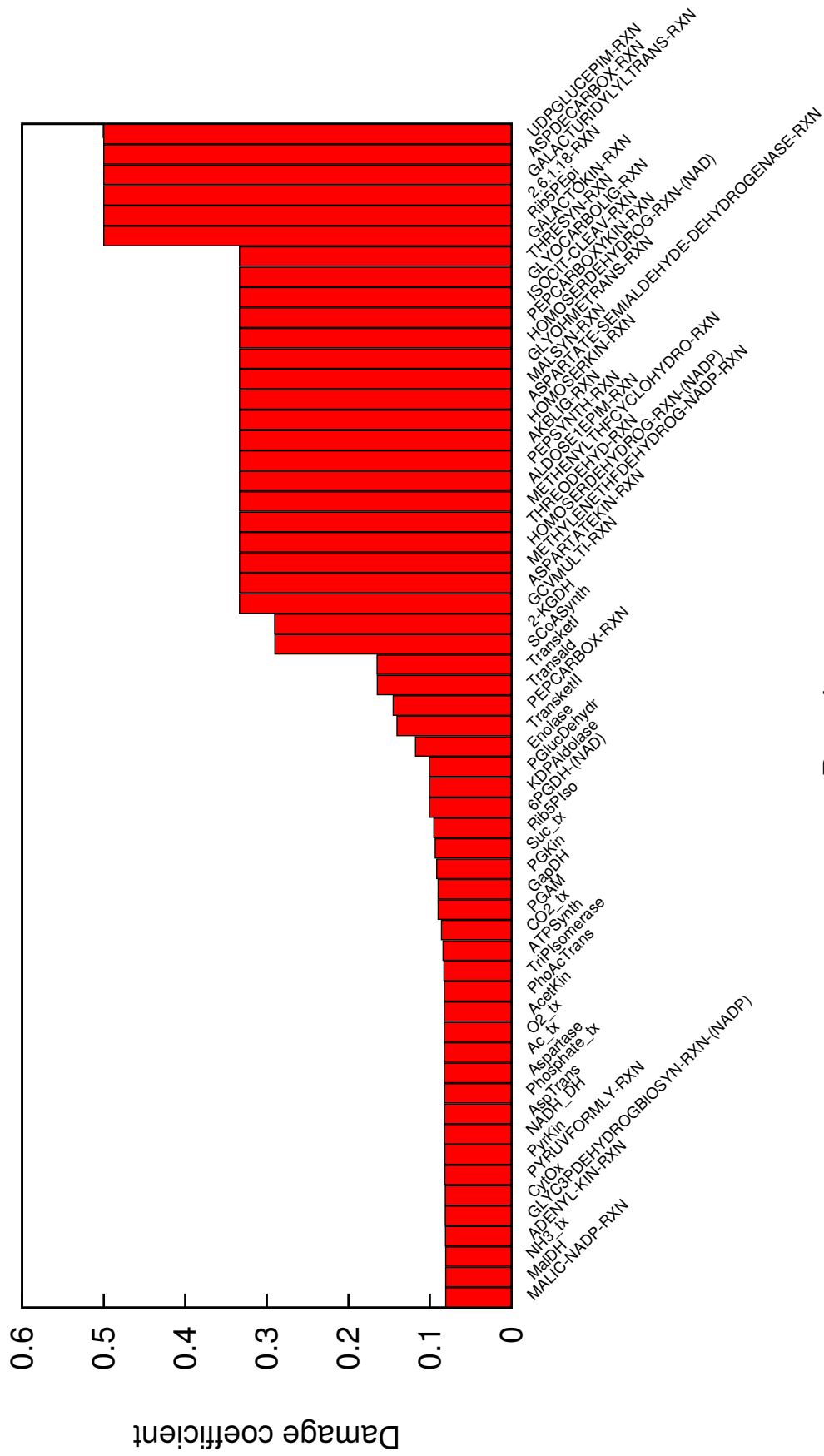
Figure 7.6: Bar chart of damage coefficients of the reactions in the global catabolic core.
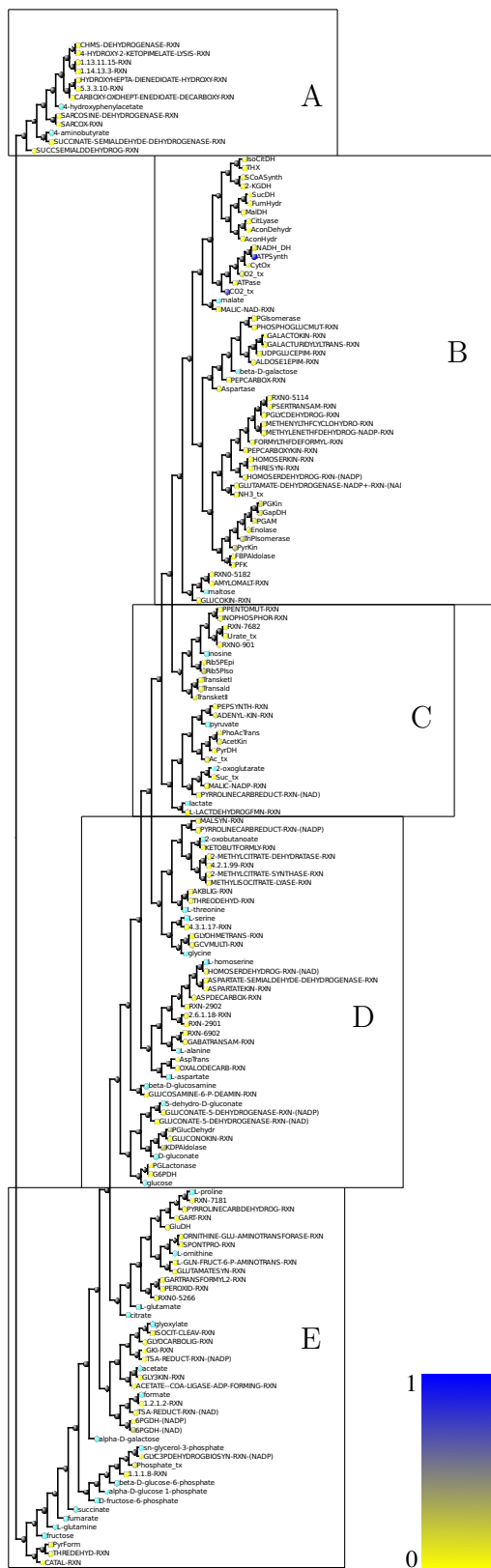
Figure 7.7: Relative frequency of reactions in the global catabolic core. Frequencies (as defined in Equation 7.5) are indicated by the node colour: blue nodes (●) indicate a relative frequency of 1, yellow nodes indicate a relative frequency of 0, gradients between blue and yellow (●) indicate frequencies between the extremes (see scale). As in Figs. 7.2 and 7.4, turquoise nodes are used for carbon source transporters. Clusters A-E are shown in more detail in Fig. 7.5. The relative frequencies are shown in more detail in Fig. 7.9.

119

Figure 7.8: Enlarged clusters from Fig. 7.4.

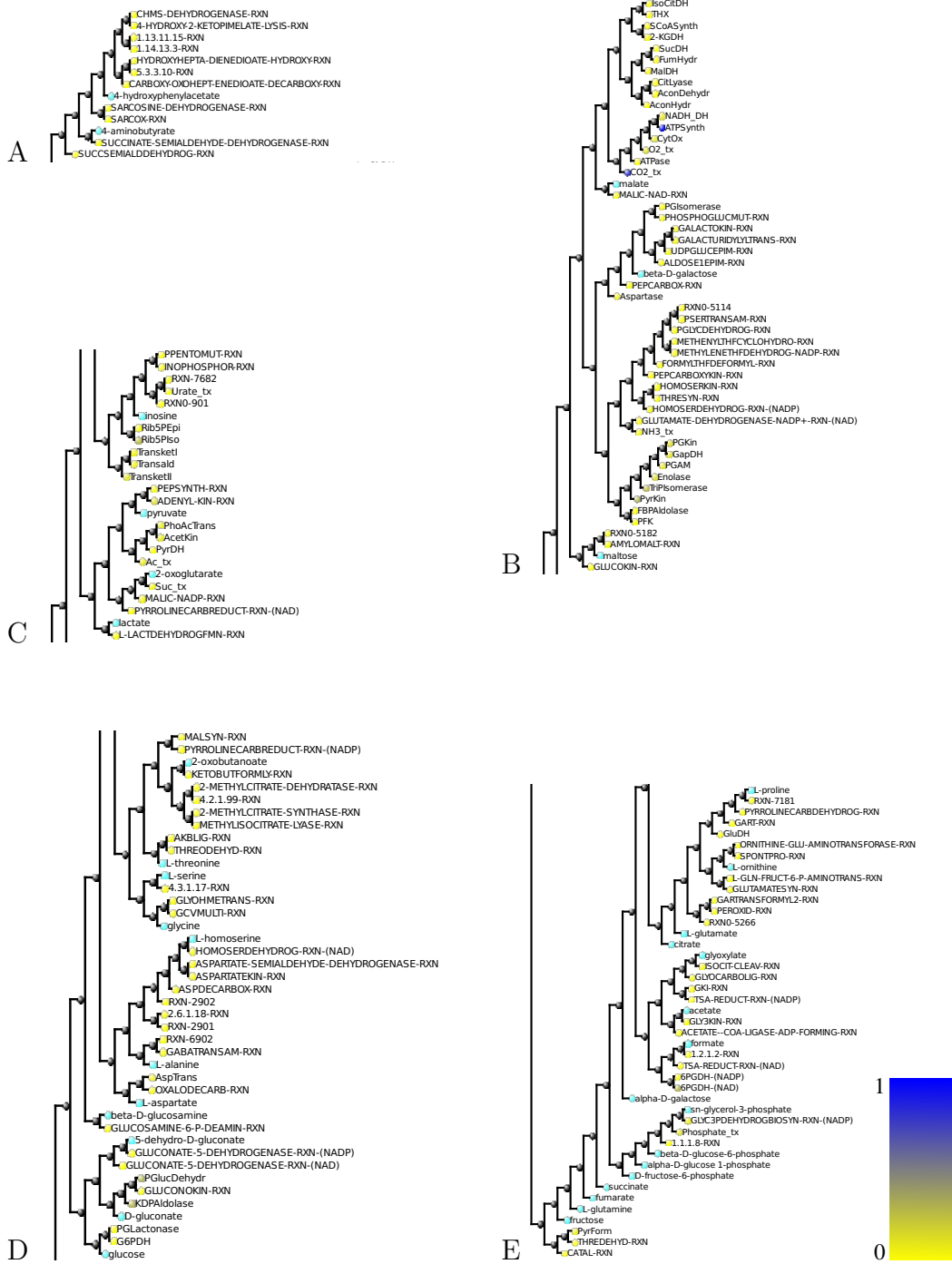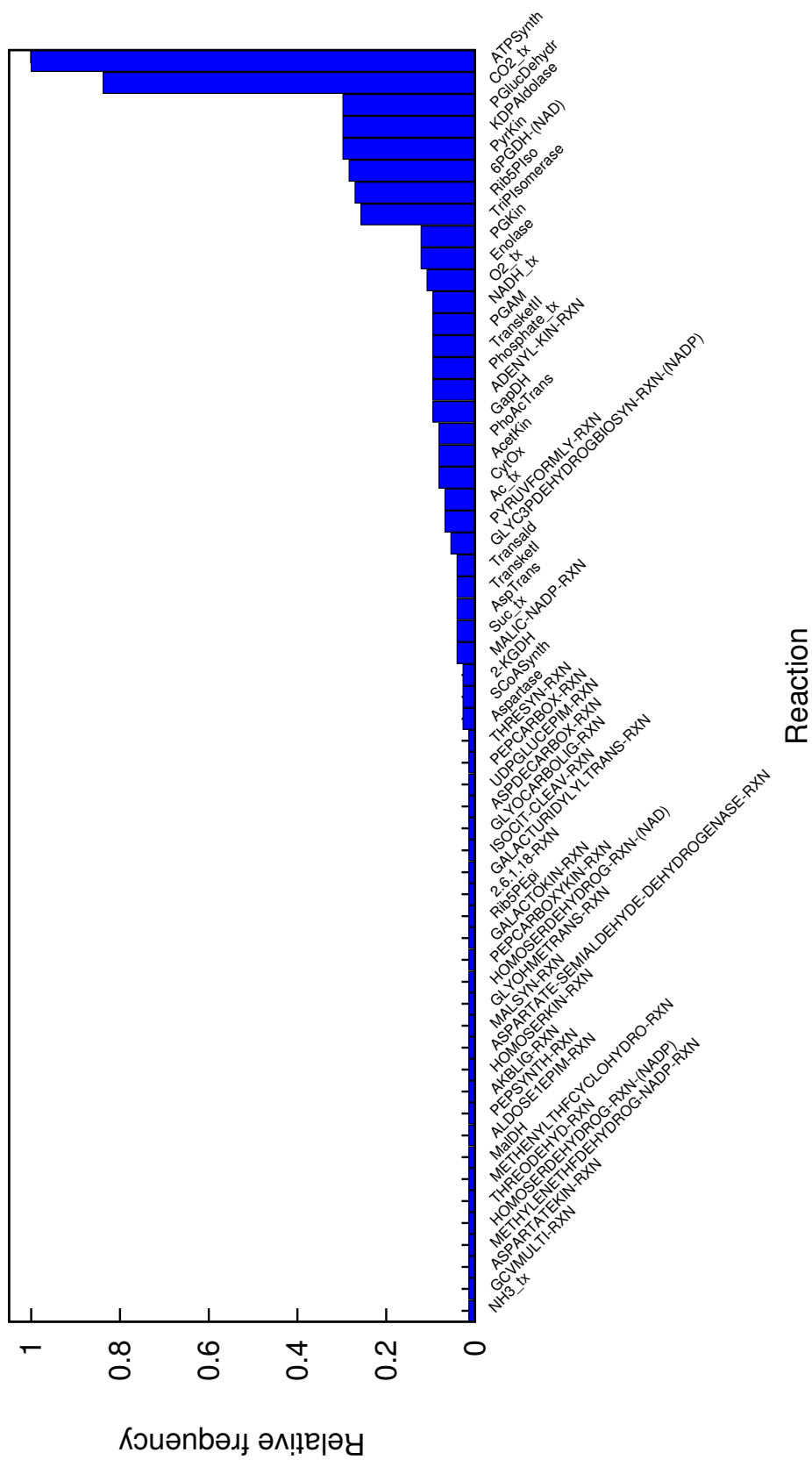Figure 7.9: Bar chart of relative frequencies (Equation 7.5) of reactions from the global catabolic core in damaging reaction sets.

here has been evaluated experimentally during infection. Becker et al. [8] used proteomics to identify enzymes that were expressed by *S.* Typhimurium during infection of mice. A subset of these enzymes were experimentally evaluated *in vivo* by constructing deletion mutants. The essentiality data relevant for the analysis presented here is summarised in Table 7.1. Becker et al. [8] identified a total of 155 proteins as potential drug targets, these coincide with 13 of reactions (and 24 protein-coding genes) identified in this analysis.

It should be emphasised that all reactions included in Table 7.1 occurred in sets of two or three in the data presented here, whereas the *in vivo* data was mainly based on single gene deletions (the only double deletion included in the table is $\Delta ackA\ \Delta pta$). It is also worth noting that none of the mutations listed in the table caused complete lethality, which is consistent with the modelling results. This observation also strengthens the motivation for including reaction deletions that are non-lethal, but cause an increase in objective value compared to the wild type, in the damage analysis: In this analysis the situation being modelled is growth in the macrophage where both the nutrient availability and the detailed biomass composition is unknown. Thus, it is reasonable to assume that a reaction deletion that cause a significant investment in enzyme (estimated as increase in flux sum) on a rich medium will have an even greater impact on growth in the nutritionally poorer and hostile macrophage environment. Furthermore, as indicated in the table, complete lethality is not required for avirulence, which suggests that reaction sets with a relatively low damage could be sufficient for avirulence *in vivo*.

An experimental strategy for implementing these reaction sets could be to focus on single reactions with high damage coefficients, and only implement the complete sets if these single deletions fail to cause attenuation. This way, reaction deletions that cause sufficient damage to prohibit infection, but not necessarily biomass synthesis, could be identified. The suggested experimental design would allow identification of high damage reactions sets that could potentially be useful as drug targets. It would also suggest low or intermediate damage sets, which is potentially useful for construction of live attenuated vaccine strains. If high damage reaction sets were to be used as drug target candidates, it would be desirable to avoid targeting any reactions which are shared between the host and the pathogen. A cursory comparison between the reactions included in the human PGDB from BioCyc and the reactions assigned a damage coefficient indicate that 15 reactions would be suitable targets (damage coefficient shown in parenthesis): AcetKin (0.08), PhoAcTrans (0.08), AspKin (0.33), KDPAl-dolase (0.1), HomoSerDH (0.33), AlaTransAm (0.5), ThreSyn (0.33), IsoCit-Cleav (0.33), AspDeCarb (0.5), AspSemAldDH (0.33), PyrForm (0.08), MalSyn (0.33), HomoSerDH (0.33), PGlucDehydr (0.1), GlyoCarLig (0.33).

Table 7.1: Summary of gene essentiality of reactions identified in damage analysis. D_r refers to the damage coeffcient of the reaction (as defiend in Equation 7.4) and F_r to the relative frequency (Equation 7.5). The summary is based on data and literature review presented in Becker et al. [8]. Note that abbreviated reaction names are used (see Table 1).

Reaction	D_r	F_r	Genes	Comment	Reference
AcetKin	0.08	0.08	ackA, pduW	$\Delta ackA$ Δpta (see PhoAc-Trans, below) has a wild type phenotype, but is suggested to be redundant.	[8, 174]
AspSemAldDH	0.33	0.01	usg, asd	Δasd is avirulant, presumably owing to inability to produce peptidoglycan.	[8]
AspKin	0.33	0.01	lysC, thrA, metL	$\Delta lysC$ is avirulant, presumably owing to inability to produce peptidoglycan.	[8]
ATPSynth	0.03	1	atpA– atpH	$\Delta atpB$ is avirulent, presumably owing to inability to generate ATP.	[8, 175]
Enolase	0.12	0.12	eno	Predicted essentiallity based on expression.	[8]
GapDH	0.09	0.09	gapA	Predicted essentiallity based on expression.	[8]
GCVMult	0.33	0.01	lpdA	$\Delta lpdA$ is avirulant.	[8]
IsoCitCleav	0.33	0.01	aceA	Predicted essentiallity based on expression.	[8, 176]
MetTHFDH	0.33	0.01	folD	Other reactions in the folate biosynthesis pathway (all expressed by the fol operon) are targets for sulfonamids.	[8]
PhoAcTrans	0.08	0.08	eutD, pta	$\Delta ackA$ Δpta (see AcetKin, above) has a wild type phenotype, but is suggested to be redundant.	[8]
PGKin	0.09	0.12	pgk	Predicted essentiallity based on expression.	[8]
6PGDH-(NAD)	0.1	0.28	gnd, rfbK, rfbM, rfbG, rfbF	Deletions of genes encoding reactions involved in biosynthesis of components of LPS are avirulant.	[8, 177]
UPDGluEpi	0.5	0.01	galE	$\Delta galE$ is avirulant.	[178, 179]

Chapter 8

General discussion

In the World Economic Forum's Global Risks report from 2013 [180] three risk cases of global concern are highlighted, one of those is the threat posed by antibiotic resistant bacteria in combination with the declining antimicrobial research and development pipeline. The report points out a few troubling aspects of this problem: Multi-resistant Gram negative pathogens, notably *Klebsiella pneumoniae*, that are resistant to all currently available antimicrobials (including those that are in the development pipeline) have been isolated. The success of treatment of infectious diseases over the last 70 years has depended on a steady input of new classes of antimicrobials, rather than the drugs being very sophisticated, since resistance would emerge eventually, especially for natural antibiotics. Thus, there is the possibility of a worst case, post-antibiotic scenario where all available antimicrobials have been rendered ineffective. Although for most healthy adults, at least in high income countries, this would not involve an immediate health threat, over the course of an individual's life there are many situations where efficient antibiotics are needed. Many currently available medical practises, e.g. preterm birth management, heart surgery, organ transplant, immunomodulating therapies to treat autoimmune diseases, as well as certain forms of cancers, would become associated with higher risk of complications in the absence of efficient antibiotics. This scenario would also, apart from high mortality rate, result in significant economic damage.

With the exception of multi-resistant Gram positive bacteria, the problem highlighted in the Global Risks report, and addressed in this thesis, has not improved since the work presented here was started. In the following section the results obtained here will be briefly reviewed, focusing on their relevance in addressing this problem.

8.1 Overview of the results obtained

Chapter 3 was focused on the construction of a genome-scale metabolic model of *S.* Typhimurium. The choice of modelling *S.* Typhimurium for addressing the concern of antibiotic resistant Gram negative pathogens was motivated by the fact that it is an established model organism for this group of bacteria, as well as a significant pathogen in its own right. The primary aim of the chapter was to construct a consistent model and analyse fundamental properties of the *S.* Typhimurium metabolic network. The starting point of the model construction was a previously constructed organism specific BioCyc database. As highlighted in the chapter, significant curation efforts were still required in order to construct a high quality metabolic model. The *S.* Typhimurium metabolic model was analysed by simulating changes in energy demand, using flux balance analysis. This method, which was also applied in Chapters 4 and 7, was used for identifying reactions that were correlated with energy demand, and therefor formed a catabolic sub-network within the genome-scale network.

The main focus of Chapter 4 was to adapt a previously published genome-scale model of *L. plantarum* to the ScrumPy and BioCyc standards in order to analyse metabolic interactions between this model and the *S.* Typhimurium model. Energy demand analysis showed that the catabolic core of *L. plantarum* shared some common features with the *S.* Typhimurium model and that the structure of the catabolic core was greatly influenced by which mode of fermentation (heterolactic or homolactic) the *L. plantarum* network was using.

The focus of Chapter 5 was analysis of metabolic interactions between *S.* Typhimurium and *L. plantarum*. Since the infection route of *S.* Typhimurium, and other enteric bacteria, involves ingestion and transmission through the epithelium of the small intestine, interactions with resident microflora (including *L. plantarum*) is an important part of the infection cycle. The initial ambition of this work was to identify different types of interactions, mutually beneficial, as well as antagonistic, ones. It was soon realised that the complex phenomenon of antagonism through metabolic interactions could not be accurately modelled using a structural approach alone, and that more experimental data would be required in order to pursue this aim. What could be modelled with structural models was potentially beneficial interactions. The results of Chapter 5 indicate that these are very limited. The only mutualistic interaction identified was transfer of glycolate from *L. plantarum* to *S.* Typhimurium.

Chapter 6 was dedicated to integration of experimental data with the *S.* Typhimurium genome-scale metabolic model. Specifically, Phenotype Microarray (PM) data and DNA microarray data obtained during intracellular growth was used. The integration of these two data types with the *S.* Typhimurium

model provided a basis for analysing nutrient availability and catabolism during infection. The results did not indicate any simple generalisations of the nutrient availability *in vivo*, but there were clear differences between carbon sources, e.g. glycerol-3-phosphate, some nucleosides, and some amino acids, appeared to be dominating. The distribution of metabolite-associated expressions could be indicative of a complex nutrient availability *in vivo* where many carbon sources are present at low concentration and co-metabolised. The results could also be indicative of a nutritionally variable environment to which *S.* Typhimurium has adapted by expressing metabolic genes related to catabolism of many unrelated compounds. The results obtained here and the general methodology for model-based analysis of expression could, pending experimental validation, be of great use for understanding the *in vivo* metabolism of *Salmonella* and other pathogenic bacteria. This information would not only be of general scientific interest, but be useful for designing intervention strategies against infective bacteria.

In the final chapter, results from Chapters 3 and 6 were combined in order to identify enzymes that could be of pharmaceutical interest. The method used in Chapter 7 was intended to identify sets of reactions that were important for viability of *S.* Typhimurium, regardless of the nutrient composition available to the cell. This approach was motivated by the results from Chapter 6, indicating an either variable or complex host environment. In order to achieve this, a consensus catabolic core network based on all growth supporting carbon sources was constructed and used for damage analysis. The focus on catabolism and energy generation was motivated by the importance of this metabolic function for overall viability. An alternative would have been to identify the reactions that were invariably associated with biomass synthesis. A problem with this approach is that with many carbon sources available the number of invariable biosynthetic reactions becomes quite limited. Catabolism and energy generation is arguably more fundamental that biomass synthesis for most pathogens. There are also experimental indications that growth of *Salmonella* during infection is limited. For example, after infection of human macrophages, *M. tuberculosis* can enter a state of dormancy where no or little growth occurs [181]. Despite the reduced metabolic activity and cellular ATP levels during dormancy, uncouplers of respiratory ATP generation have a strong antimicrobial effect on dormant cells *in vitro* [182]. Rapid growth of *S.* Typhimurium is usually observed when cultured macrophage cell lines are used for simulating *in vivo* macrophage infection. However, there is accumulating evidence for the occurrence of cells entering a metabolically dormant state during infection of cultured fibroblasts [183], similar to that of mycobacteria [184, 185]. Observations from animal tissue indicate an average total number of intracellular replications *in vivo* of two

or three [186].

8.2 Outlook

There is of course much additional work that would be both interesting and worthwhile pursuing. The most urgent priority for any continuation of the this work would be experimental investigation of the hypotheses generated here.

The method for data integration presented in Chapter 6 could be validated by applying it to expression data obtained during growth on media with known composition. The increasing accuracy of RNA-seq expression data [165, 166] makes it, in principle, very applicable to the method suggested in this work. This technology, in combination with metabolic modelling, also opens up the possibility of analysing *in vivo* metabolism and nutrient availability in environments that otherwise would be challenging to analyse experimentally, such as intracellular growth in intact living host tissue.

Perhaps the most interesting continuation of the results presented here would be construction and validation of the deletion mutations suggested in Chapter 7. A remaining question following validation would be how to use this information. An option would be to use the identified enzymes as targets for drug development. One of the benefits of antimicrobials targeting metabolic enzymes is that these targets are often conserved, which makes the drug suitable as a broad spectrum antimicrobial. A potential problem with this approach is of course the risk of inhibiting host enzymes. Although, drug candidates have been developed that specifically inhibit mycobacterial homologues of type-II NADH dehydrogenase (phenothiazines and analogous compounds) [187] and ATP synthase (diarylquinolines) [188], indicating the feasibility of targeting a conserved metabolic pathway.

The estimated damage imposed by the reaction sets in Chapter 7 falls within the extremes of no damage at all to complete lethality. Live attenuated vaccines for bacterial pathogens are strains that are viable, but genetically modified to be avirulant. In many cases this has been achieved by deletion of metabolic genes [189]. An alternative application of the less damaging reaction sets could then be construction of live attenuated vaccines for *S.* Typhimurium. Safe and efficient bacterial vaccines could dramatically reduce the need for routine use of antibiotics, thus decrease the selective pressure on multi-resistant pathogens, which in effect would preserve the potency of the last resort antibiotics [180, 190].

Bibliography

[1] E.P. Abraham and E. Chain. An Enzyme from Bacteria able to Destroy Penicillin. *Nature*, 146:837–837, 1940.

[2] J. Davies and D. Davies. Origins and Evolution of Antibiotic Resistance. *Microbiol. Mol. Biol. Rev.*, 74:417–433, 2010.

[3] J. Davies. Vicious circles: looking back on resistance plasmids. *Genetics*, 139:1465–1468, 1996.

[4] L. L. Silver. Challenges of antibacterial discovery. *Clin. Microbiol. Rev.*, 24:doi:10.1128/CMR.00030–10, 2011.

[5] S. R. Norrby, C. E. Nord, and R. Finch. Lack of development of new antimicrobial drugs: a potential serious threat to public health. *Lancet Infect. Dis.*, 5:115–119, 2005.

[6] A. J. Tatem, D. D. Rogers, and S. I. Hay. Global transport networks and infectious disease spread. *Adv Parasitol.*, 62:293–343, 2006.

[7] D. M. Livermore. Has the era of untreatable infections arrived? *J Antimicrob Chemother.*, 64:i29–i36, 2009.

[8] D. Becker, M. Selbach, C. Rollenhagen, M. Ballmaier, T. F. Meyer, M. Mann, and D. Bumann. Robust *Salmonella* metabolism limits possibilities for new antimicrobials. *Nature*, 440:303–308, 2006.

[9] L. E. Alksne and P.M. Dunman. Target-Based Antimicrobial Drug Discovery. *Methods in Molecular Biology*, 431, 271-283.

[10] N. A. Feasey, G. Dougan, R. A. Kingsley, R. S. Heyderman, and M. A. Gordon. Invasive non-typhoidal salmonella disease: an emerging and neglected tropical disease in Africa. *Lancet*, 379:2489–2499, 2012.

[11] L. Le Minor and M. Y. Popoff. Request for an Opinion Designation of *Salmonella enterica* sp. nov., norn. rev., as the Type and Only Species of the Genus *Salmonella. Int. J. Syst. Bacteriol.*, 37:465–468, 1987.

[12] J. R. McQuiston, S. Herrera-Leon, B. C. Wertheim, J. Doyle, P. I. Fields, R. V. Tauxe, and Jr J. M. Logsdon. Molecular Phylogeny of the Salmonellae: Relationships among *Salmonella* Species and Subspecies Determined from Four Housekeeping Genes and Evidence of Lateral Gene Transfer Events. *J Bacteriol.*, 190:7060–7067, 2008.

[13] R. A. Edwards, G. J. Olsen, and S. R. Maloy. Comparative genomics of closely related salmonellae. *TRENDS in Microbiology*, 10(2):94–99, 2002.

[14] S. Uzzau, D.J. Brown, T. Wallis, S. Rubino, G. Leori, S. Bernard, J. Casadesús, D.J. Platt, and J.E. Olsen. Host adapted serotypes of *salmonella enterica. Epidemiol Infect.*, 125:229–255, 2000.

[15] M.K. Bhan, R. Bahl, and S. Bhatnagar. Typhoid and paratyphoid fever. *Lancet*, 366:749–762, 2005.

[16] C.M. Parry, T.T. Hien, G. Dougan, N.J. White, and J.J. Farrar. Typhoid fever. *N Engl J Med.*, 347:1770–1782, 2002.

[17] M.I. Khan, R.L. Ochiai, L. von Seidlein, B. Dong, S.K.Bhattacharya, M.D. Agtini, Z.A. Bhutta, G.C. Do, M. Ali, D.R. Kim, M. Favorov, and J.D. Clemens. Non-typhoidal *Salmonella* rates in febrile children at sites in five Asian countries. *Trop Med Int Health.*, 15:doi: 10.1111/j.1365–3156.2010.02553.x., 2010.

[18] R. L. Ochiai, C. J. Acosta, M. C. Danovaro-Holliday, D. Baiqing, S. K. Bhattacharya, M. D. Agtini, Z. A. Bhutta, D. G. Canh, M. Ali, S. Shin, J. Wain, A.-L. Page, M. J. Albert, J. Farrar, R. Abu-Elyazeed, T. Pang, C. M. Galindo, L. von Seidlein, J. D. Clemens, and the Domi Typhoid Study Group. A study of typhoid fever in five Asian countries: disease burden and implications for controls. *Bull World Health Organ.*, 86:260–268, 2008.

[19] F.M. Sánchez-Vargas, M.A. Abu-El-Haija, and G.O. Gómez-Duarte. Salmonella infections: an update on epidemiology, management, and prevention. *Travel Med Infect Dis.*, 9:263–277, 2011.

[20] A. Haraga, M.B. Ohlson, and S.I. Miller. Salmonellae interplay with host cells. *Nat Rev Microbiol.*, 6:53–66, 2008.

[21] F. Garcia del Portillo, J.W. Foster, and B.B. Finlay. Role of acid tolerance response genes in *Salmonella typhimurium* virulence. *Infect Immun.*, 61: 4489–4492, 1993.

[22] A. Takeuchi. Electron microscope studies of experimental *Salmonella* infection. i. penetration into the intestinal epithelium by *Salmonella typhimurium*. *Am J Pathol.*, 50:109–136, 1967.

[23] A.J. Baumler, R.M. Tsolis, and F. Heffron. Contribution of fimbrial operons to attachment to and invasion of epithelial cell lines by *Salmonella typhimurium*. *Infect Immun.*, 64:1862–1865, 1996.

[24] A. Vazquez-Torres, J. Jones-Carson, A.J. Baumler, S. Falkow, R. Valdivia, W. Brown, M. Le, R. Berggren, W.T. Parks, and F.C. Fang. Extraintestinal dissemination of *Salmonella* by cd18-expressing phagocytes. *Nature*, 401:804–808, 1999.

[25] L.E. Smythies, M. Sellers, R.H. Clements, M. Mosteller-Barnum, G. Meng, W.H. Benjamin, J.M. Orenstein, and P.D. Smith. Human intestinal macrophages display profound inflammatory anergy despite avid phagocytic and bacteriocidal activity. *J Clin Invest*, 115:66–75, 2005.

[26] P. I. Fields, R. V. Swanson, C. G. Haidaris, and F. Heffron. Mutants of *Salmonella typhimurium* that cannot survive within the macrophage are avirulent. *PNAS*, 83:5189–5193, 1986.

[27] D. O'Callaghan, D. Maskell, F.Y. Liewa, C.S. Easmon, and G. Dougan. Characterization of aromatic- and purine-dependent *Salmonella typhimurium*: attention, persistence, and ability to induce protective immunity in balb/c mice. *Infect Immun.*, 56:419–423, 1988.

[28] E.E. Effa, Z.S. Lassi, J.A. Critchley, P. Garner, D. Sinclair, and Z.A. Bhutta. Fluoroquinolones for treating typhoid and paratyphoid fever (enteric fever). *Cochrane Database Syst Rev*, 10: doi:10.1002/14651858.CD004530.pub4, 2011.

[29] H.-M. Chen, Y. Wang, L.-H. Su, and C.-H. Chiu. Nontyphoid salmonella infection: Microbiology, clinical features, and antimicrobial therapy. *Pediatr Neonatol*, 54:147–152, 2013.

[30] A. Fraser, E. Goldberg, C.J. Acosta, M. Paul, and L. Leibovici. Vaccines for preventing typhoid fever. *Cochrane Database Syst Rev*, 3: doi:10.1002/14651858.CD001261.pub2, 2007.

[31] R. Germanier and E. Fuer. Isolation and characterization of Gal E mutant Ty 21a of *Salmonella typhi*: a candidate strain for a live, oral typhoid vaccine. *J Infect Dis.*, 131:553–558, 1975.

[32] S. S. Hall. Revolution postponed. *Scientific American*, 303:60–67, 2010.

[33] U. Sauer, M. Heinemann, and N. Zamboni. Getting Closer to the Whole Picture. *Science*, 316:550–551, 2007.

[34] Athel Cornish-Bowden. *Fundamentals of Enzyme Kinetics.* Portland Press, London, 1995.

[35] M. K. Campbell and S. O. Farrell, editors. *Biochemistry.* Brooks/Cole, 4 edition, 2000.

[36] D. L. Nelson and M.M. Cox, editors. *Lehninger Principles of Biochemistry.* Freeman, New York, 4 edition, 2004.

[37] D.A. Fell. *Understanding the Control of Metabolism.* Portland Press, London, 1997.

[38] Y. Shinfuku, N. Sorpitiporn, M. Sono, C. Furusawa T., and Hirasawa H. Shimizu. Development and experimental verification of a genome-scale metabolic model for *Corynebacterium glutamicum. Microb Cell Fact.*, 8: doi: 10.1186/1475–2859–8–43, 2009.

[39] N.C. Duarte, S. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas, and B.O. Palsson. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *PNAS*, 104(6):1777–1782, 2007.

[40] H. Ma, A. Sorokin, A. Mazein, A. Selkov, E Selkov, O. Demin, and I. Goryanin. The Edinburgh human metabolic network reconstruction and its functional analysis. *Molecular Systems Biology*, 3(135), 2007.

[41] Reinhardt Heinrich and Stefan Schuster. *The Regulation of Cellular Systems.* Chapman & Hall, London, England, 1996.

[42] S. Schuster and D. Fell. *Bioinformatics - From Genomes to Therapies*, volume 2, chapter 20, pages 755–805. Wiley-VCH, 2007.

[43] D. A. Fell and K. Snell. Control analysis of mammalian serine biosynthesis. *Biochem. J.*, 256:97–101, 1988.

[44] J. Heffron, editor. *Linear Algebra.* Department of Mathematics and Applied Mathematics, Virginia Commonwealth Univerity, Richmond, 2 edition, 2009.

[45] S. Schuster and R. Schuster. Detecting strictly detailed balanced subnetworks in open chemical-reaction networks. *J.Math.Chem.*, 6(1):17–40, 1991.

[46] T. Pfeiffer, I. Sanchez-Valdenebro, J.C. Nuno, F. Montero, and S. Schuster. METATOOL: for studying metabolic networks. *Bioinformatics*, 15 (3):251–257, 1999.

[47] M. G. Poolman. ScrumPy - metabolic modelling with Python. *IEE Proceedings Systems Biology*, 153(5):375–378, 2006.

[48] M. G. Poolman, C. Sebu, M. K. Pidcock, and D. A. Fell. Modular decomposition of metabolic systems via null-space analysis. *Journal of Theoretical Biology*, 249:691–705, 2007.

[49] G. N. Lance and W. T. Williams. A general theory of classificatory sorting strategies 1. Hierarchical systems. *Computer Journal*, 9(4):373–380, 1967.

[50] Byron J. T.; Morgan and Andrew P. G. Ray. Non-uniqueness and Inversions in Cluster Analysis. *Applied Statistics*, 44(1):117–34, 1995.

[51] S. Schuster and C. Hilgetag. On Elementary Flux Modes in Biochemical Systems at Steady State. *J.Biol.Syst.*, 2:165–182, 1994.

[52] S. Schuster, T. Dandekar, and D.A. Fell. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends. Biotech.*, 17(2):53–60, 1999.

[53] S. Klamt, S. Schuster, and E. D. Gilles. Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism of in purple nonsulphur metabolism. *Biotechnol. Bioeng.*, 77 (7):734–750, 2002.

[54] C. Wagner. Nullspace Approach to Determine the Elementary Modes of Chemical Reaction Systems. *J. Phys. Chem. B*, 108:2425–2431, 2004.

[55] J Gagneur and S Klamt. Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics*, 5(175), 2004.

[56] C.T. Trinh, A. Wlaschin, and F. Srienc. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl Environ Microbiol.*, 81:813–826, 2009.

[57] C.T. Trinh, P. Unrean, and F. Srienc. Minimal *escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. *Appl Environ Microbiol.*, 74:3634–3643, 2008.

[58] S. Klamt and E.D. Gilles. Minimal cut sets in biochemical reaction networks. *Bioinformatics*, 20(2):226–234, 2004.

[59] S. Klamt. Generalized concept of minimal cut sets in biochemical networks. *BioSystems*, 83:233–247, 2006.

[60] C.H. Schilling, D. Letscher, and B.O. Palsson. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway oriented perspective. *J.Theor.Biol.*, 203:229–248, 2000.

[61] Jason A. Papin, Nathin D. Price, Sharon J. Wiback, David A. Fell, and Bernhard O. Palsson. Metabolic pathways in the post-genomic era. *Trends Biochem. Sci.*, 28(5):250–258, 2003.

[62] S. Klamt and J. Stelling. Two approaches for metabolic pathway analysis? *Trends Biotechnol.*, 21:64–69, 2003.

[63] C. Wagner and R. Urbanczik. The Geometry of the Flux Cone of a Metabolic Network. *Biophys J.*, 89:3837–3845, 2005.

[64] A. Varma and B. O. Palsson. Metabolic Capabilities of *Escherichia coli*: II. Optimal Growth Patterns. *J. Theor. Biol.*, 165:503–522, 1993.

[65] H.-G. Holzhütter. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur. J. Biochem.*, 271:2905–2922, 2004.

[66] H.-G. Holzhutter. The generalized flux-minimization method and its application to metabolic networks affected by enzyme deficiencies. *Biosystems*, 83:98–107, 2006.

[67] I. Thiele and B. Ø. Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc.*, 5:93–121, 2010.

[68] J.S. Edwards, R.U. Ibarra, and B.O. Palsson. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.*, 19:125–130, 2001.

[69] S. S. Fong and B.O. Palsson. Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nat. Genet.*, 36:1056–1058, 2004.

[70] B. Teusink, A. Wiersma, D. Molenaar, C. Francke, W. M. de Vos, R. J. Siezen, and E. J. Smid. Analysis of Growth of *Lactobacillus plantarum* wcfs1 on a Complex Medium Using a Genome-scale Metabolic Model. *J Biol Chem*, 281:40041–40048, 2006.

[71] B. Teusink, A. Wiersma, L. Jacobs, R. A. Notebaart, and E. J. Smid. Understanding the Adaptive Growth Strategy of *Lactobacillus plantarum* by *In Silico* optimisation. *PLoS Comput Biol*, 5:e1000410, 2009.

[72] S. Schustera, T. Pfeiffer, and D. A. Fell. Is maximization of molar yield in metabolic networks favoured by evolution? *J. Theor. Biol.*, 252:497–504, 2008.

[73] D. A. Fell and J. R. Small. Fat synthesis in adipose tissue. *Biochem. J*, 238, 781-786.

[74] M. R. Watson. A discrete model of bacterial metabolism. *Comp. Appl. Biosci.*, 2(1):23–27, 1986.

[75] A. Varma and B. O. Palsson. Metabolic Capabilities of *Escherichia coli*: I. Synthesis of Biosynthetic Precursors and Cofactors. *J. Theor. Biol.*, 165:477–502, 1993.

[76] Ali R Zomorrodi, Patrick F Suthers, Sridhar Ranganathan, and Costas D Maranas. Mathematical optimization applications in metabolic networks. *Metab. Eng.*, 14(6):672–86, November 2012. doi: 10.1016/j.ymben.2012.09.005.

[77] S. Lee, C. Phalakornkule, M. M. Domach, and I. E. Grossmann. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Computers and Chemical Engineering*, 24:711–716, 2000.

[78] N. Jamshidi and B. Ø. Palsson. Formulating genome-scale kinetic models in the post-genome era. *Mol Syst Biol.*, 4:doi:10.1038/msb.2008.8, 2008.

[79] K. Smallbone, E. Simeonidis, N. Swainston, and P. Mendes. Towards a genome-scale kinetic model of cellular metabolism. *BMC Syst Biol.*, 4: doi: 10.1186/1752–0509–4–6, 2010.

[80] N. J. Loman, C. Constantinidou, J. Z. M. Chan, M. Halachev, M. Sergeant, C. W. Penn, E. R. Robinson, and M. J. Pallen. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat. Rev. Microbiol.*, 10:599–606, 2012.

[81] P. Stothard and D. S. Wishart. Automated bacterial genome analysis and annotation. *Curr. Opin. Microbiol.*, 9:505–510, 2006.

[82] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *J Mol Biol.*, 215:403–410, 1990.

[83] Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, and Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, 27:29–34, 1999.

[84] Minoru Kanehisa, Susumu Goto, Masahiro Hattori, Kiyoko F. Aoki-Kinoshita, Masumi Itoh, Shuichi Kawashima, Toshiaki Katayama, Michi-hiro Araki, and Mika Hirakawa. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research*, 34(Database Issue):D354–D357, 2006.

[85] Karp PD, Riley M, Saier M, Paulsen IT, Paley SM, and Pellegrini-Toole A. The EcoCyc and MetaCyc databases. *Nucleic Acids Res.*, 28(1):56–59, 2000.

[86] Peter D. Karp, Christos A. Ouzounis, Caroline Moore-Kochlacs, Leon Goldovsky, Pallavi Kaipa, Dag Ahrén, Sophia Tsoka, Nikos Darzentas, Victor Kunin, and Núria López-Bigas. Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Research*, 33(19):6083–6089, 2005.

[87] M. G. Poolman, B. K. Bonde, A. Gevorgyan, H. H. Patel, and D. A. Fell. Challenges to be faced in the reconstruction of metabolic networks from public databases. *IEE Proceedings Systems Biology*, 153(5):379–384, 2006.

[88] P. D. Karp, M. Latendresse, and R. Caspi. The Pathway Tools Pathway Prediction Algorithm. *Stand Genomic Sci*, 5:doi: 10.4056/sigs.1794338, 2011.

[89] P. D. Karp, S. M. Paley, M., Krummenacker, M. Latendresse, J. M. Dale, T. J. Lee, P. Kaipa, F. Gilham, A. Spaulding, L. Popescu, T. Altman, I. Paulsen, I. M. Keseler, and Ron Caspi. Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief Bioinform.*, 11:doi: 10.1093/bib/bbp043, 2011.

[90] I. Schomburg, A. Chang, C. Ebeling, M. Gremse, C. Heldt, G. Huhn, and D. Schomburg. BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Research*, 32:doi:10.1093/nar/gkh081, 2004.

[91] R.N. Goldberg, Y.B. Tewari, and T.N. Bhat. Thermodynamics of enzyme-catalyzed reactions - a database for quantitative biochemistry. *Bioinformatics*, 20:2874–2877, 2004.

[92] A. Gevorgyan, M. G. Poolman, and D. A. Fell. Detection of stoichiometric inconsistencies in biomolecular models. *Bioinformatics*, 24(19):2245–2251,

October 2008. ISSN 1460-2059. URL `http://view.ncbi.nlm.nih.gov/pubmed/18697772`.

[93] E.V. Nikolaev, A.P.Burgard, and C.D. Maranas. Elucidation and structural analysis of conserved pools for genome-scale metabolic reconstructions. *Biophys J*, 88:37–49, 2005.

[94] M. G. Poolman, L. Miguet, L. J. Sweetlove, and D. A. Fell. A genome-scale metabolic model of *Arabidopsis thaliana* and some of its properties. *Plant Physiol.*, 151(3):1570–1581, 2009.

[95] A. K. Chavali, K. M. DAuria E. L. Hewlett, R. D. Pearson, and J. A. Papin. A metabolic network approach for the identification and prioritization of antimicrobial drug targets. *Trends Microbiol.*, 20:doi: 10.1016/j.tim.2011.12.004, 2012.

[96] M.A. Oberhardt, B. Ø. Palsson, and J.A. Papin. Applications of genome-scale metabolic reconstructions. *Mol Syst Biol.*, 5:doi: 10.1038/msb.2009.77, 2009.

[97] B. Papp, R.A. Notebaart, and C. Pál. Systems-biology approaches for predicting genomic evolution. *Nat Rev Genet.*, 12:591–602, 2011.

[98] H.U. Kim, S.Y. Kim, H. Jeong, T.Y. Kim, J.J. Kim, H.E. Choy, K.Y. Yi, J.H. Rhee, and S.Y. Lee. Integrative genome-scale metabolic analysis of vibrio vulnificus for drug targeting and discovery. *Mol Syst Biol*, 7: doi:10.1038/msb.2010.115, 2011.

[99] B. Chance, J.J. Higgins, and D. Garfinkel. Analogue and digital computer representations of biochemical processes. *Fed Proc.*, 21:75–86, 1962.

[100] J. A. Burns. *Studies on Complex Enzyme Systems*. PhD thesis, University of Edinburgh, 1971.

[101] H.M. Sauro, M. Hucka, A. Finney, C. Wellock, H. Bolouri, J. Doyle, and H. Kitano. Next Generation Simulation Tools: The Systems Biology Workbench and BioSPICE Integration. *OMICS*, 7(4):355–372, 2003.

[102] D. Garfinkel. A Machine–Independent Language for the Simulation of Complex Chemical and Biochemical Systems. *Comput. Biomed. Res.*, 2: 31–44, 1968.

[103] D. J. M. Park and B. E Wright. Metasim, a General Purpose Metabolic Simulator for Studying Cellular Transformations. *Comput. Progm. Biomed.*, 3:10–26, 1973.

[104] E. M. Chance, A. R. Curtis, I. P. Jones, and C. R. Kirby. *FACSIMILE: a Computer Program for Flow and Chemistry Simulation, and General Initial Value Problems.*, 1977. AERE–R8775, HMSO London.

[105] H. M. Sauro. *Control analysis and simulation of metabolism.* PhD thesis, Oxford Polytechnic, 1986.

[106] H. M. Sauro. SCAMP: a general–purpose simulator and metabolic control analysis program. *Comput. Applic. Biosci.*, 9:441–450, 1993.

[107] P Mendes. Gepasi - a software package for modeling the dynamics, steady-states and control of biochemical and other systems. *Comp. Appl. Biosci.*, 9(5):563–571, 1993.

[108] 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*, 22:3067–3074, 2006.

[109] R. Schwarz, P. Musch, A. von Kamp, B. Engels, H. Schirmer, S. Schuster, and T. Dandekar. YANA a software tool for analyzing flux modes, gene-expression and enzyme activities. *BMC Bioinformatics*, 6: doi:10.1186/1471–2105–6–135, 2005.

[110] S. Klamt, J. Saez-Rodriguez, and E.D. Gilles. Structural and functional analysis of cellular networks with cellnetanalyze. *BMC Systems Biology*, 1, 2007.

[111] S. Klamt, J. Stelling, M. Ginkel, and E.D. Gilles. FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics*, 19(2):261–269, 2003.

[112] J. Schellenberger, R. Que, R. M. T. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, J. Kang, D. R. Hyduke, and B.Ø. Palsson. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc.*, 6:1290–1307, 2011.

[113] A. Ebrahim, J.A. Lermanand B.O. Palsson, and D.R. Hyduke. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst Biol.*, 7:doi: 10.1186/1752–0509–7–74, 2013.

[114] A. Gevorgyan, M. E. Bushell, C. Avignone-Rossa, and A. M. Kierzek. SurreyFBA: a command line tool and graphics user interface for constraint-based modeling of genome-scale metabolic reaction networks. *Bioinformatics*, 27:433–434, 2011.

[115] B. G. Oliver, J. M. Rohwer, and J.-H.S. Hofmeyr. Modelling cellular systems with pysces. *Bioinformatics*, 21:560–561, 2005.

[116] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics/computer Applications in The Biosciences*, 19:524–531, 2003.

[117] M. Lutz. *Programming Python*. O'Reilly & Associates, 2nd edition, 2001.

[118] M. Lutz and D. Ascher. *Learning Python*. O'Reilly & Associates, 1st edition, 1999.

[119] A. Raghunathan, J. Reed, S. Shin, B. O. Palsson, and S. Daefler. Constraint-based analysis of metabolic capacity of *Salmonella typhimurium* during host-pathogen interaction. *BMC Systems Biology*, 3 (38), 2009.

[120] M. AbuOun, P. F. Suthers, G. I. Jones, B. R. Carter, M. P. Sauders, C. D. Maranas, M. J. Woodward, and M. F. Anjum. Genome scale reconstruction of a *Salmonella* metabolic model: comparison of similarity and differences with a commensal *Escherichia coli* strain. *J.Biol.Chem.*, 234 (43):29480–29489, 2009.

[121] I. Thiele, D. R. Hyduke, B. Steeb, G. Fankam, D. K. Allen, S. Bazzani, P. Charusanti, F.-C. Chen, R. M. T. Fleming, C. A. Hsiung, S. C. J. De Keersmaecker, Y.-C. Liao, K. Marchal, M. L. Mo, E. Özdemir, A. Raghunathan, J. L. Reed, S.-I. Shin, S. Sigurbjörnsdóttir, J. Steinmann, S. Sudarsan, N. Swainston, I. M. Thijs, K. Zengler, B. O. Palsson, J. N. Adkins, and D. Bumann. A community effort towards a knowledge-base and mathematical model of the human pathogen *Salmonella* Typhimurium lt2. *BMC Systems Biology*, 5(8), 2011.

[122] M. McClelland, K. E. Sanderson, J. Spieth, S. W. Clifton, P. Latreille, L. Courtney, S. Porwollik, J. Ali, M. Dante, F. Du, S. Hou, D. Layman, S. Leonard, C. Nguyen, K. Scott, A. Holmes, N. Grewal, E. Mulvaney, E. Ryan, H. Sun, L. Florea, W. Miller, M. Nhan, R. Waterston, and

R. K. Wilson. Complete genome sequence of *Salmonella enterica* serovar typhimurium LT2. *Nature*, 413:852–856, 2001.

[123] Peter D. Karp, Suzanne Paley, and Pedro Romero. The pathway tools software. *Bioinformatics*, 18:S225–32, 2002.

[124] P. C. Maloney. *Coupling to an energized membrane: role of ion-motive gradients in the tranduction of metabolic energy. Escherichia coli* and *Salmonella typhimurium.* Cellular and Molecular Biology. American Society for Microbiology, 1987.

[125] T. Bauchop and S. R. Elsden. The Growth of Micro-organisms in Relation to their Energy Supply. *J. Gen. Microbiol.*, 23:457–469, 1960.

[126] A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, and B. O. Palsson. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, 3(121), 2007.

[127] A. Gevorgyan. *Analytic Methods for Genome-Scale Metabolic Networks Applied to* Streptococcus agalactie. PhD thesis, Oxford Brookes University, Oxford, U.K., 2009.

[128] H. Tettelin, V. Masignani, M. J. Cieslewicz, C. Donati, D. Medini, N. L. Ward, S. V. Angiuoli, J. Crabtree, A L. Jones, A. S. Durkin, R. T. DeBoy, T. M. Davidsen, M. Mora, M. Scarselli, I. Margarit y Ros, J. D. Peterson, C. R. Hauser, J. P. Sundaram, W. C. Nelson, R. Madupu, L., M. Brinkac, R. J. Dodson, M. J. Rosovitz, S. A. Sullivan, S. C. Daugherty, D. H. Haft, J. Selengut, M. L. Gwinn, L. Zhou, N. Zafar, H. Khouri, D. Radune, G. Dimitrov, K. Watkins, K. J. B. O'Connor, S Smith, T. R. Utterback, O. White, C. E. Rubens, G. Grandi, L. C. Madoff, D. L. Kasper, J. L. Telford, M. R. Wessels, R. Rappuoli, and C. M. Fraserabkm. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome". *PNAS*, 102(39), 2005.

[129] M. D. Bennet, I. J. Leitch, H. J. Price, and J. S. Johnston. Comparisons with *Caenorhabditis* (100 Mb) and *Drosophila* (175 mb) Using Flow Cytometry Show Genome Size in *Arabidopsis* to be 157 Mb and thus 25% Larger than the *Arabidopsis* Genome Initiative Estimate of 125 Mb. *Annals of Botany*, 91:547–557, 2005.

[130] M. Bevan and S. Walsh. The *Arabidopsis* genome: A foundation for plant research. *Genome Res.*, 15:1632–1642, 2005.

139

[131] N. Peekhaus and T. Conway. What's for Dinner?: Entner-Doudoroff Metabolism in *Escherichia coli*. *J. Bacteriol.*, 180(14):3495–3502, 1998.

[132] U. Sauer, F. Canonaco, S. Heri, A. Perrenoud, and E. Fischer. The Soluble and Membrane-bound Transhydrogenases UdhA and PntAB Have Divergent Functions in NADPH Metabolism of *Escherichia coli*. *J. Biol. Chem*, 279(8):6613–6619, 2004.

[133] L. Peng and K. Shimizu. Global metabolic regulation analysis for *Escherichia coli* K12 based on protein expression by 2-dimensional electrophoresis and enzyme activity measurement. *Applied Microbiology and Biotechnology*, 61:163 – 178, 2003. ISSN 0175-7598. doi: 10.1007/s00253-002-1202-6. URL `http://dx.doi.org/10.1007/s00253-002-1202-6`.

[134] M. El-Mansi. Flux to acetate and lactate excretions in industrial fermentations: physiological and biochemical implications. *J. Ind. Microbiol. Biotechnol.*, 31:295 – 300, 2003.

[135] S.-J. Park, G. Chao, and R. P. Gunsalus. Aerobic regulation of the sucabcd genes of escherichia coli, which encode α-ketoglutarate dehydrogenase and succinyl coenzyme a synthetase: Roles of arca, fnr, and the upstream sdhcdab promoter. *J. Bacteriol.*, 179:4138 – 4142, 1997.

[136] A. J. Wolfe. The acetate switch. *Microbiol. Mol. Biol. Rev.*, 69:12 – 50, 2005.

[137] S. Subramanian and C. Sivaraman. Bacterial citrate lyase. *J. Biosci.*, 6: 379 – 401, 1984.

[138] H. G. Kulla. Regulatory Citrate Lyase Mutants of *Salmonella typhimurium*. *J. Bacteriol*, 153:546 – 549, 1983.

[139] S. Eriksson, S. Lucchini, A. Thompson, M. Rhen, and J. C. D. Hinton. Unravelling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica*. *Mol Microbiol.*, 47:103 – 118, 2003.

[140] C. Pujol, J. P. Grabenstein, R. D. Perry, and J. B. Bliska. Replication of yersinia pestis in interferon γ-activated macrophages requires *ripA*, a gene encoded in the pigmentation locus. *PNAS*, 102:12909 – 12914, 2005.

[141] H. Inoue, F. Suzuki, H. Tanioka, and Y. Takeda. Studies on ATP Citrate Lyase of Rat Liver. III: The Reaction Mechanism. *J. Biochem.*, 63, 1968.

[142] L. J. Reitzer. *Sources of Nitrogen and Their Utilization*, volume 1, pages 380–391. ASM Press, 2nd edition, 1996.

[143] M. Marcus and Halpern Y. S. The metabolic pathway of glutamate in *Escherichia coli* K-12. *Biochem. Biophys. Acta.*, 177(2):314–320, 1969.

[144] L. J. Reitzer. *Ammonia Assimilation and the Biosynthesis of Glutamine, Glutamate, Aspartate, Aspargine, L-Alanine, and D-Alanine*, volume 1, pages 391–408. ASM Press, 2nd edition, 1996.

[145] K. A. Z. Siddiquee, M. J. Arauzo-Bravo, and K. Shimizu. Effect of a pyruvate kinase (*pykF*-gene) knockout mutation on the control of gene expression and metabolic fluxes in *Escherichia coli*. *FEMS Microbiol Lett*, 235:25–33, 2004.

[146] J. Schellenberger, J. O. Park, T. M Conrad, and Bernhard Ø Palsson. BiGG: a Biochemical Genetic and Genomic Database knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, 11, 2010.

[147] A. P. Oliveira, J. Nielsen, and J. Förster. Modeling *Lactococcus lactis* using a genome-scale flux model. *BMC Microbiology*, 5, 2005.

[148] E. Simeonidis, E. Murabito, K. Smallbone, and H. V. Westerhoff. Why does yeast ferment? a flux balance analysis study. *Biochem. Soc. Trans.*, 38:1225–1229, 2010.

[149] S. Stolyar, S. Van Dien, K. L. Hillesland, N. Pinel, T. J. Lie, J. A. Leigh, and D. A. Stahl. Metabolic modeling of a mutualistic microbial community. *Molecular Systems Biology*, 3:doi:10.1038/msb4100131, 2007.

[150] E. H. Wintermute and P. A. Silver. Emergent cooperation in microbial metabolism. *Molecular Systems Biology*, 6:doi:10.1038/msb.2010.66, 2010.

[151] D. Segrè, D. Vitkup, and G.M. Church. Analysis of optimality in natural and perturbed metabolic networks. *PNAS*, 99:15112–15117, 2002.

[152] S. Freilich, R. Zarecki, O. Eilam, E. Shtifman Segal, C. S. Henry, M. Kupiec, U. Gophna, R. Sharan, and E. Ruppin. Competitive and cooperative metabolic interactions in bacterial communities. *Nat Commun*, 2: doi:10.1038/ncomms1597, 2011.

[153] A. R. Zomorrodi and C. D. Maranas. OptCom: A Multi-Level Optimization Framework for the Metabolic Modeling and Analysis of Microbial Communities. *PLoS Comput. Biol.*, 8:e1002363, 2012.

[154] J. M. Neal-McKinney, X. Lu, T. Duong, C. L. Larson, D. R. Call, D. H. Shah, and M. E. Konkel1. Production of Organic Acids by Probiotic Lactobacilli Can Be Used to Reduce Pathogen Load in Poultry. *PLoS ONE*, 7:e43928, 2012.

[155] D. G. Brown and C. Allen. *Ralstonia solanacearum* genes induced during growth in tomato: an inside view of bacterial wilt. *Mol. Microbiol.*, 53: 1641–1660, 204.

[156] B. R. Bocher, P. Gadzinski, and E. Panomitros. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Research*, 11:1246–1255, 2001.

[157] B. R. Bochner and M. A. Savageau. Generalized Indicator Plate for Genetic, Metabolic, and Taxonomic Studies with Microorganisms. *Appl. Environ. Microbiol.*, 33:434 – 444, 1977.

[158] B. R. Bochner. New technologies to assess genotype-phenotype relationships. *Nat. Rev. Genet.*, 4:309 – 314, 2003.

[159] Y.-K. Oh, B. O. Palsson, S. M. Park, C. H. Schilling, and R. Mahadevan. Genome-scale Reconstruction of Metabolic Network in Bacillus subtilis Based on High-throughput Phenotyping and Gene Essentiality Data. *J. Biol. Chem.*, 282:28791–28799, 2007.

[160] A. Ehrenreich. DNA microarray technology for the microbiologist: an overview. *Appl Microbiol Biotechnol*, 73:255–273, 2006.

[161] D. J. Duggan, M. Bittner, Y. Chen, P. Meltzer, and J. M. Trent. Expression profiling using cdna microarrays. *Nat. Genet.*, 21:10 – 14, 1999.

[162] S. A. Becker and B. O. Palsson. Context-Specific Metabolic Networks Are Consistent with Experiments. *PLoS Comput. Biol.*, 4, 2008.

[163] T. Shlomi, M. N. Cabili, M. J. Herrgård, B. Ø. Palsson, and E. Ruppin. Network-based prediction of human tissue-specific metabolism. *Nat. Biotechnol.*, 4, 2008.

[164] M.J. Alston, J. Seers, J.C.D. Hinton, and S. Lucchini. BABAR: an R package to simplify the normalisation of common reference design microarray-based transcriptomic datasets. *BMC Bioinformatics*, 11, 2010.

[165] Z. Wang, M. Gerstein, and M. Snyder. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.*, 10:57–63, 2009.

[166] A.J. Westermann, S.A. Gorski, and J. Vogel. Dual RNA-seq of pathogen and host. *Nat Rev Microbiol.*, 10:618–630, 2012.

[167] A. Götz, E. Eylert, W. Eisenreich, and W. Goebel. Carbon Metabolism of Enterobacterial Human Pathogens Growing in Epithelial Colorectal Adenocarcinoma (Caco-2) cells. *PLoS ONE*, 5(5), 2010.

[168] I. Hautefort, A. Thompson, S. Eriksson-Ygberg, M. L. Parker, S. Lucchini, V. Danino, R. J. M. Bongaerts, N. Ahmad, M. Rhen, and J. C. D. Hinton. During infection of epithelial cells *Salmonella enterica* serovar typhimurium undergoes a time-dependent transcriptional adaptation that results in simultaneous expression of three type 3 secretion systems. *Cell Microbiol*, 10:958–984, 2008.

[169] E. Eylert, J. Schär, S. Mertins, Regina Stoll, A. Bacher, W. Goebel, and W. Eisenreich. Carbon metabolism of *Listeria monocytogenes* growing inside macrophages. *Mol. Microbiol.*, 69:1008–1017, 2008.

[170] T. Dandekar, A. Fieselmann, J. Popp, and Michael Hensel. *Salmonella enterica*: a surprisingly well-adapted intracellular lifestyle. *Front Microbiol*, 3:doi: 10.3389/fmicb.2012.00164, 2012.

[171] K. Y. Leung and B. B. Finlay. Intracellular replication is essential for the virulence of *Salmonella typhimurium. PNAS*, 88:11470–11474,, 1991.

[172] B. K. Bonde, D. J. V. Beste, E. Laing, A. M. Kierzek, and J. McFadden. Differential Producibility Analysis (DPA) of Transcriptomic Data with Metabolic Networks: Deconstructing the Metabolic Response of *M. tuberculosis. PLoS Comput. Biol.*, 7:e1002060, 2011.

[173] R. Breitling and P. Herzyk. Rank-based methods as a non-parametric alternative of the t-statistic for the analysis of biological microarray data. *J Bioinform Comput Biol*, 3:1171–1189, 2005.

[174] S. D. Lawhon, R. Maurer, M. Suyemoto, and C. Altier. Intestinal short-chain fatty acids alter *Salmonella typhimurium* invasion gene expression and virulence through BarA/SirA. *Mol Microbiol.*, 46:1451–1464, 2002.

[175] A. K. Turner, L. Z. Barber, P. Wigley, S. Muhammad, M. A. Jones, M. A. Lovell, S. Hulme, and P. A. Barrow. Contribution of proton-translocating proteins to the virulence of *Salmonella enterica* serovars Typhimurium, Gallinarum, and Dublin in chickens and mice. *Infect Immun.*, 71:3392–3401, 2003.

[176] F. C. Fang, S. J. Libby, M. E. Castor, and A. M. Fung. Isocitrate lyase (AceA) is required for salmonella persistence but not for acute lethal infection in mice. *Infect Immun.*, 73:2547–2549, 2005.

[177] L. V. Collins, S. Attridge, and J. Hackett. Mutations at *rfc* or *pmi* Attenuate *Salmonella typhimurium* Virulence for Mice. *Infect Immun.*, 59:1079–1085, 1991.

[178] D. Hone, R. Morona S. Attridge, and J. Hackett. Construction of defined *galE* mutants of *Salmonella* for use as vaccines. *J Infect Dis.*, 156:167–174, 1987.

[179] D. M. Hone, S. R. Attridge, B. Forrest, R. Morona, D. Daniels, J. T. LaBrooy, R. C. Bartholomeusz, D. J. Shearman, and J. Hackett. A *galE* via (Vi antigen-negative) mutant of *Salmonella typhi* Ty2 retains virulence in humans. *Infect Immun.*, 56:1326–1333, 1988.

[180] L. Howell, editor. *Global risks 2013, eighth edition: an initiative of the Risk Response Network.* 8 edition, 2013.

[181] L. Wayne and C. D. Sohaskey. Nonreplicating persistence of *Mycobacterium tuberculosis. Annu. Rev. Microbiol.*, 55:139–163, 2001.

[182] S. P. S. Rao, S. Alonso, L. Rand, T. Dick, and K. Pethe. The protonmotive force is required for maintaining ATP homeostasis and viability of hypoxic, nonreplicating *Mycobacterium tuberculosis. PNAS*, 105(33):11945–11950, 2008.

[183] D. A. Cano, M. G. Pucciarelli, M. Martìnez-Moya, J. Casadesus, and Francisco Garcìa del Portillo. Selection of Small-Colony Variants of *Salmonella enterica* Serovar Typhimurium in Nonphagocytic Eucaryotic Cells. *Infection and Immunity*, 71:3690–3698, 2003.

[184] D. J. Murphy and J. R. Brown. Identification of gene targets against dormant phase *Mycobacterium tuberculosis* infections. *BMC Infectious Disease*, 7(84), 2007.

[185] F. Garcia-del Portillo, C. Nunez-Hernandez, B. Eisman, and J. Ramos-Vivas. Growth control in the Salmonella-containing vacuole. *Current Opinion in Microbiology*, 11:46–52, 2008.

[186] A. Tierrez and F. Garcia-del Portillo. New concepts in *Salmonella* virulence: the importance of reducing the intracellular growth rate in the host. *Cellular Microbiology*, 7(7):901–909, 2005.

[187] E. A. Weinstein, T. Yano, L.-S. Li, D. Avarbock, A. Avarbock, D. Helm, A. A. McColm, K. Duncan, J. T. Lonsdale, and H. Rubin. Inhibitors of type II nadh:menaquinoneoxidoreductase represent a class of antitubercular drugs. *PNAS*, 102(12):4548–4553, 2005.

[188] K. Andries, P. Verhasselt, J. Guillemont, H. W. H. Gohlmann, J.-M. Neefs, H. Winkler, J. Van Gestel, P. Timmermann, M. Zhu, E. Lee, P. Williams, D. de Ghaffoy, E. Huitric, S. Hoffner, E. Cambau, C. Truffot-Pernot, N. Lounis, and V. Jarlier. A Diarylquinoline Drug Active on the ATP Synthase of *Mycobacterium tuberculosis*. *Science*, 307:223–227, 2005.

[189] A. Detmer and J. Glenting. Live bacterial vaccines a review and identification of potential hazards. *Microb Cell Fact.*, 23:doi:10.1186/1475–2859–5–23, 2006.

[190] B. Spellberg, J. G. Bartlett, and D. N. Gilbert. The Future of Antibiotics and Resistance. *N Engl J Med.*, 368:299–302, 2013.

[191] M.A. El-Khani and R.J. Stretton. Effect of growth medium on the lipid composition of log and stationary phase cultures of *Salmonella typhimurium*. *Microbos.*, 31:161–169, 1981.

[192] J. Gmainer and S. Schlecht. Molecular Organization of the Outer Membrane of *Salmonella typhimurium*. *Eur. J. Biochem.*, 93:609–620, 1969.

[193] A.A. Lindberg and C.G. Hellerqvist. Bacteriophage Attachment Sites, Serological Specificity, and Chemical Composition of the Lipopolysaccharides of Semirough and Rough Mutants of *Salmonella typhimurium*. *J.Bacteriol.*, 105(1):57–64, 1971.

Appendix A

Biomass

Table A.1: Concentration of biomass components of *Salmonella* Typhimurium used as constraints in FBA.

Component	Concentration (mmol(g DW)$^{-1}$))	Reference
Amino acids		
Isoleucine	0.087	[119]
Arginine	0.13	[119]
Alanine	0.75	[119]
Aspartate	0.37	[119]
Glutamate	0.26	[119]
Valine	0.28	[119]
Proline	0.26	[119]
Cysteine	0.017	[119]
Threonine	0.35	[119]
Phenylalanine	0.14	[119]
Glycine	1.5	[119]
Lysine	0.31	[119]
Tyrosine	0.04	[119]
Histidine	0.067	[119]
Serine	0.32	[119]
Tryptophan	0.088	[119]
Aspargine	0.27	[119]
Glutamine	0.20	[119]
Methionine	0.10	[119]
Leucine	0.21	[119]

Nucleotides

dATP	0.026	[119]
UTP	0.15	[119]
CTP	0.17	[119]
ATP	0.15	[119]
dCTP	0.028	[119]
dTTP	0.026	[119]
dGTP	0.029	[119]

Cell envelope

Phosphatidylethanolamine	$1.5 \cdot 10^{-3}$	[191]
Phosphatidylserin	$4.0 \cdot 10^{-6}$	[191]
Peptidoglycan	0.049	[192]
Phosphatidylglycerol	$4.2 \cdot 10^{-4}$	[191]
1,2-diacylglycerol	$1.88 \cdot 10^{-4}$	[191]
Lipopolysaccharide	$9.37 \cdot 10^{-4}$	[193]
Phosphatidate	$4.0 \cdot 10^{-6}$	[191]
cardiolipin	$1.4 \cdot 10^{-5}$	[191]
Starch	0.027	[191]

Other compounds

NADH	$5.0 \cdot 10^{-5}$	[126]
NAD	$2.15 \cdot 10^{-3}$	[126]
Methyl-tetrahydrofolate	0.05	[126]
Putrescine	0.035	[126]
Succinate-CoA	$3.0 \cdot 10^{-6}$	[126]
Acetyl-CoA	$5.0 \cdot 10^{-5}$	[126]
NADP	$1.3 \cdot 10^{-4}$	[126]
CoA	$6.0 \cdot 10^{-6}$	[126]
UDP-glucose	$3.0 \cdot 10^{-3}$	[126]
FAD	$1.0 \cdot 10^{-5}$	[126]
Spermidine	$7.0 \cdot 10^{-3}$	[126]
AMP	$1.0 \cdot 10^{-3}$	[126]

Appendix B

Reaction sets

Table B.1: Damage of reaction sets from Chapter 7, expressed as sum of total flux. The reference, wild type total flux is 192 mmol (g DW)$^{-1}$h^{-1}. This data also appears as a python file in Appendix C

Reaction sets	Damage (mmol (g DW)$^{-1}$h^{-1})
RXN-3341,Suc_tx,CO2_tx	300.4
UDPGLUCEPIM-RXN,GALACTURIDYLYLTRANS-RXN	lethal
ATPSynth,GLYCEROL-3P_tx,ACETATEKIN-RXN	253.1
ATPSynth,PHOSACETYLTRANS-RXN,TRIOSEPISOMERIZATION-RXN	253.9
HOMO-SER_tx,ATPSynth,GLYCEROL-3P_tx	241.2
RXN-3341,O2_tx,CO2_tx	249.6
PHOSGLYPHOS-RXN,CO2_tx,NADH_DH_ubi	246.1
O2_tx,CO2_tx,RIB5PISOM-RXN	252.4
RXN-3341,CO2_tx,Cytochrome_c_oxidase	243.5
TRANSALDOL-RXN,1TRANSKETO-RXN	lethal
1TRANSKETO-RXN,2TRANSKETO-RXN	lethal
NH3_tx,ATPSynth,PEPDEPHOS-RXN	242.0
ATPSynth,RIB5PISOM-RXN,TRIOSEPISOMERIZATION-RXN	247.1
ATPSynth,2OXOGLUTARATEDEH-RXN,PEPDEPHOS-RXN	242.6
3PGAREARR-RXN,O2_tx,CO2_tx	243.4
O2_tx,CO2_tx,KDPGALDOL-RXN	249.6
CO2_tx,2PGADEHYDRAT-RXN,Cytochrome_c_oxidase	243.5
Suc_tx,CO2_tx,KDPGALDOL-RXN	300.4

ATPSynth,PEPDEPHOS-RXN,RIB5PISOM-RXN	248.0
PGLUCONDEHYDRAT-RXN,CO2_tx,GLYCEROL-3P_tx	257.8
CO2_tx,ASPAMINOTRANS-RXN,KDPGALDOL-RXN	247.7
PGLUCONDEHYDRAT-RXN,RIB5PISOM-RXN,1TRANSKETO-RXN	lethal
RXN-3341,PHOSGLYPHOS-RXN,CO2_tx	290.4
ATPSynth,2-KETOGLUTARATE_tx,TRIOSEPISOMER-IZATION-RXN	241.1
ATPSynth,CO2_tx,KDPGALDOL-RXN	273.4
ASPDECARBOX-RXN,2.6.1.18-RXN	lethal
ATPSynth,PGLUCONDEHYDRAT-RXN,1TRANSKETO-RXN	241.0
HOMO-SER_tx,ATPSynth,RIB5PISOM-RXN	241.6
ATPSynth,2PGADEHYDRAT-RXN	244.8
ATPSynth,PEPDEPHOS-RXN,Ac_tx	250.2
ATPSynth,ADENYL-KIN-RXN,TRIOSEPISOMERIZATION-RXN	241.7
PGLUCONDEHYDRAT-RXN,O2_tx,CO2_tx	249.6
RXN-3341,CO2_tx,TRIOSEPISOMERIZATION-RXN	253.1
3PGAREARR-RXN,CO2_tx,NADH_DH_ubi	246.1
CO2_tx,GAPOXNPHOSPHN-RXN,2TRANSKETO-RXN	241.8
GCVMULTI-RXN,METHYLENETHFDEHYDROG-NADP-RXN,GLYOHMETRANS-RXN	lethal
CO2_tx,NADH_DH_ubi,RIB5PISOM-RXN	242.6
RXN-3341,CO2_tx,2PGADEHYDRAT-RXN	290.4
ATPSynth,PGLUCONDEHYDRAT-RXN,PEPDEPHOS-RXN	245.8
PGLUCONDEHYDRAT-RXN,CO2_tx,GAPOXNPHOSPHN-RXN	290.4
CO2_tx,GAPOXNPHOSPHN-RXN,RIB5PISOM-RXN	257.4
CO2_tx,2PGADEHYDRAT-RXN,RIB5PISOM-RXN	257.0
PGLUCONDEHYDRAT-RXN,CO2_tx,ASPARTASE-RXN	246.4
ATPSynth,RXN-3341,PEPDEPHOS-RXN	245.8
ATPSynth,KDPGALDOL-RXN,PEPDEPHOS-RXN	245.8
HOMOSERDEHYDROG-RXN-(NAD),HOMOSERDEHYDROG-RXN-(NADP), ASPARTATE-SEMIALDEHYDE-DEHYDROGENASE-RXN	lethal
GLYOCARBOLIG-RXN,MALSYN-RXN,ISOCIT-CLEAV-RXN	lethal
PGLUCONDEHYDRAT-RXN,CO2_tx,PEPCARBOX-RXN	434.9
3PGAREARR-RXN,CO2_tx,Cytochrome_c_oxidase	243.4
HOMOSERDEHYDROG-RXN-(NAD),ASPARTATEKIN-RXN,HOMOSERDEHYDROG-RXN-(NADP)	lethal

PGLUCONDEHYDRAT-RXN,CO2_tx,2PGADEHYDRAT-RXN	290.4
ATPSynth,RXN-3341,TRIOSEPISOMERIZATION-RXN	246.9
PHOSGLYPHOS-RXN,CO2_tx,KDPGALDOL-RXN	290.4
RXN-3341,Phosphate_tx,CO2_tx	250.8
PGLUCONDEHYDRAT-RXN,CO2_tx,ASPAMINOTRANS-RXN	247.7
ATPSynth,PHOSACETYLTRANS-RXN,RIB5PISOM-RXN	241.5
PGLUCONDEHYDRAT-RXN,CO2_tx,NADH_DH_ubi	245.9
RXN-3341,CO2_tx,GAPOXNPHOSPHN-RXN	290.4
PHOSGLYPHOS-RXN,CO2_tx,Cytochrome_c_oxidase	243.4
CO2_tx,2PGADEHYDRAT-RXN,NADH_DH_ubi	246.2
RIBULP3EPIM-RXN,1TRANSKETO-RXN	lethal
HOMO-SER_tx,ATPSynth,PEPDEPHOS-RXN	242.6
ATPSynth,Phosphate_tx,RIB5PISOM-RXN	243.1
RIBULP3EPIM-RXN,TRANSALDOL-RXN	lethal
ATPSynth,GLYCEROL-3P_tx,ADENYL-KIN-RXN	241.7
ATPSynth,GLYCEROL-3P_tx,Ac_tx	249.8
ATPSynth,GLYCEROL-3P_tx,2-KETOGLUTARATE_tx	241.2
ATPSynth,MALIC-NADP-RXN,TRIOSEPISOMERIZATION-RXN	241.3
PGLUCONDEHYDRAT-RXN,Phosphate_tx,CO2_tx	250.8
PGLUCONDEHYDRAT-RXN,3PGAREARR-RXN,CO2_tx	290.4
KDPGALDOL-RXN,RIB5PISOM-RXN,1TRANSKETO-RXN	lethal
ATPSynth,CO2_tx,RIB5PISOM-RXN	259.1
ATPSynth,GLYCEROL-3P_tx,PEPDEPHOS-RXN	247.3
CO2_tx,GLYCEROL-3P_tx,KDPGALDOL-RXN	257.8
CO2_tx,KDPGALDOL-RXN,GLYC3PDEHYDROGBIOSYN-RXN-(NADP)	243.5
RXN-3341,CO2_tx,ASPAMINOTRANS-RXN	247.7
PGLUCONDEHYDRAT-RXN,CO2_tx,GLYC3P-DEHYDROGBIOSYN-RXN-(NADP)	243.5
HOMO-SER_tx,ATPSynth,TRIOSEPISOMERIZATION-RXN	241.2
ATPSynth,KDPGALDOL-RXN,TRANSALDOL-RXN	240.9
RXN-3341,CO2_tx,PEPCARBOX-RXN	434.9
ATPSynth,3PGAREARR-RXN	244.8
ATPSynth,PGLUCONDEHYDRAT-RXN,TRIOSEPISOMERIZATION-RXN	251.9
ATPSynth,ASPAMINOTRANS-RXN,PEPDEPHOS-RXN	242.2
PHOSGLYPHOS-RXN,CO2_tx,2TRANSKETO-RXN	241.8
RXN-3341,RIB5PISOM-RXN,1TRANSKETO-RXN	lethal

THRESYN-RXN,THR_tx,THREODEHYD-RXN	lethal
ATPSynth,PHOSACETYLTRANS-RXN,PEPDEPHOS-RXN	250.2
ATPSynth,PYRUVFORMLY-RXN,TRIOSEPISOMERIZATION-RXN	245.6
GCVMULTI-RXN,METHENYLTHFCYCLOHYDRO-RXN,GLYOHMETRANS-RXN	lethal
HOMO-SER_tx,ASPARTATEKIN-RXN	lethal
ATPSynth,PEPDEPHOS-RXN,PYRUVFORMLY-RXN	248.9
ATPSynth,CO2_tx,GLYCEROL-3P_tx	241.1
2OXOGLUTARATEDEH-RXN,SUCCCOASYN-RXN	lethal
AKBLIG-RXN,THRESYN-RXN,THR_tx	lethal
Phosphate_tx,CO2_tx,KDPGALDOL-RXN	250.8
CO2_tx,KDPGALDOL-RXN,PEPCARBOX-RXN	434.9
PGLUCONDEHYDRAT-RXN,PHOSGLYPHOS-RXN,CO2_tx	290.4
ATPSynth,PEPDEPHOS-RXN,TRIOSEPISOMERIZATION-RXN	247.3
CO2_tx,GLYCEROL-3P_tx,RIB5PISOM-RXN	245.4
ATPSynth,Phosphate_tx,PHOSACETYLTRANS-RXN	241.8
O2_tx,CO2_tx,2PGADEHYDRAT-RXN	243.5
ATPSynth,Phosphate_tx,ADENYL-KIN-RXN	242.2
ATPSynth,PHOSACETYLTRANS-RXN,GLYCEROL-3P_tx	253.1
ATPSynth,CO2_tx,PEPDEPHOS-RXN	250.4
HOMO-SER_tx,ASPARTATE-SEMIALDEHYDE-DEHYDROGENASE-RXN	lethal
ATPSynth,MALATE-DEH-RXN,PEPDEPHOS-RXN	241.6
ATPSynth,PYRUVFORMLY-RXN,RIB5PISOM-RXN	242.5
CO2_tx,KDPGALDOL-RXN,2PGADEHYDRAT-RXN	290.4
O2_tx,CO2_tx,GAPOXNPHOSPHN-RXN	243.4
RXN-3341,3PGAREARR-RXN,CO2_tx	290.4
ATPSynth,GLYCEROL-3P_tx,2TRANSKETO-RXN	241.2
ATPSynth,ADENYL-KIN-RXN,ACETATEKIN-RXN	241.2
ATPSynth,PGLUCONDEHYDRAT-RXN,TRANSALDOL-RXN	240.9
PGLUCONDEHYDRAT-RXN,CO2_tx,TRIOSEPISOMER-IZATION-RXN	266.8
HOMOSERKIN-RXN,THR_tx,THREODEHYD-RXN	lethal
3PGAREARR-RXN,CO2_tx,RIB5PISOM-RXN	256.9
ATPSynth,RXN-3341,CO2_tx	273.4
ATPSynth,RXN-3341,TRANSALDOL-RXN	240.9
ATPSynth,ADENYL-KIN-RXN,PEPDEPHOS-RXN	246.8
ATPSynth,CO2_tx,TRIOSEPISOMERIZATION-RXN	241.1

ATPSynth,GLYCEROL-3P_tx,FRUCTOSE-6P_tx	241.1
ATPSynth,SUCCCOASYN-RXN,PEPDEPHOS-RXN	241.4
ATPSynth,Suc_tx,PEPDEPHOS-RXN	241.7
ATPSynth,ACETATEKIN-RXN,RIB5PISOM-RXN	241.5
ATPSynth,2TRANSKETO-RXN,TRIOSEPISOMERIZATION-RXN	241.2
CO2_tx,RIB5PISOM-RXN,TRIOSEPISOMERIZATION-RXN	245.3
ATPSynth,MALIC-NADP-RXN,GLYCEROL-3P_tx	241.3
ATPSynth,MALIC-NADP-RXN,PEPDEPHOS-RXN	241.9
ATPSynth,Phosphate_tx,CO2_tx	242.4
ATPSynth,PEPDEPHOS-RXN,2-KETOGLUTARATE_tx	243.2
ATPSynth,RIB5PISOM-RXN,Ac_tx	241.5
PHOSGLYPHOS-RXN,CO2_tx,RIB5PISOM-RXN	257.4
ATPSynth,GLYCEROL-3P_tx,RIB5PISOM-RXN	247.1
CO2_tx,KDPGALDOL-RXN,Cytochrome_c_oxidase	243.5
3PGAREARR-RXN,CO2_tx,2TRANSKETO-RXN	240.9
ATPSynth,GLYCEROL-3P_tx,KDPGALDOL-RXN	247.6
CO2_tx,KDPGALDOL-RXN,TRIOSEPISOMERIZATION-RXN	266.8
ATPSynth,RXN-3341,1TRANSKETO-RXN	241.0
CO2_tx,GAPOXNPHOSPHN-RXN,Cytochrome_c_oxidase	243.4
CO2_tx,GAPOXNPHOSPHN-RXN,KDPGALDOL-RXN	290.4
3PGAREARR-RXN,CO2_tx,KDPGALDOL-RXN	290.4
PGLUCONDEHYDRAT-RXN,TRANSALDOL-RXN,RIB5PISOM-RXN	lethal
ATPSynth,Ac_tx,TRIOSEPISOMERIZATION-RXN	250.5
PGLUCONDEHYDRAT-RXN,CO2_tx,Cytochrome_c_oxidase	243.5
ATPSynth,ADENYL-KIN-RXN,Ac_tx	241.2
ATPSynth,KDPGALDOL-RXN,TRIOSEPISOMERIZATION-RXN	251.9
ATPSynth,ACETATEKIN-RXN,TRIOSEPISOMERIZATION-RXN	253.9
KDPGALDOL-RXN,TRANSALDOL-RXN,RIB5PISOM-RXN	lethal
PGLUCONDEHYDRAT-RXN,Suc_tx,CO2_tx	300.4
CO2_tx,2PGADEHYDRAT-RXN,2TRANSKETO-RXN	241.0
ATPSynth,PGLUCONDEHYDRAT-RXN,CO2_tx	273.4
CO2_tx,KDPGALDOL-RXN,ASPARTASE-RXN	246.4
ATPSynth,GLYCEROL-3P_tx,PYRUVFORMLY-RXN	245.7
ATPSynth,PHOSGLYPHOS-RXN	244.8
PEPSYNTH-RXN,PEPCARBOXYKIN-RXN,2PGADEHYDRAT-RXN	lethal

ATPSynth,PEPDEPHOS-RXN,THR_tx	240.9
UDPGLUCEPIM-RXN,ALPHA-D-GALACTOSE_tx,ALDOSE1EPIM-RXN	lethal
ATPSynth,KDPGALDOL-RXN,1TRANSKETO-RXN	241.0
ATPSynth,ACETATEKIN-RXN,PEPDEPHOS-RXN	250.2
ATPSynth,RXN-3341,GLYCEROL-3P_tx	246.9
CO2_tx,KDPGALDOL-RXN,NADH_DH_ubi	245.9
ATPSynth,RIB5PISOM-RXN,GLYC3PDEHYDROGBIOSYN-RXN-(NADP)	243.8
ATPSynth,PGLUCONDEHYDRAT-RXN,GLYCEROL-3P_tx	247.6
ATPSynth,ADENYL-KIN-RXN,PYRUVFORMLY-RXN	242.2
GALACTOKIN-RXN,UDPGLUCEPIM-RXN	lethal
PHOSGLYPHOS-RXN,O2_tx,CO2_tx	243.4
RXN-3341,CO2_tx,NADH_DH_ubi	245.9
ATPSynth,PHOSACETYLTRANS-RXN,ADENYL-KIN-RXN	241.2
ATPSynth,PEPDEPHOS-RXN,PYRUVATE_tx	243.0
RXN-3341,CO2_tx,ASPARTASE-RXN	246.4
ATPSynth,FRUCTOSE-6P_tx,TRIOSEPISOMERIZATION-RXN	241.1
RXN-3341,CO2_tx,GLYC3PDEHYDROGBIOSYN-RXN-(NADP)	242.4
RXN-3341,CO2_tx,GLYCEROL-3P_tx	253.1
HOMOSERKIN-RXN,AKBLIG-RXN,THR_tx	lethal
ATPSynth,Phosphate_tx,ACETATEKIN-RXN	241.8
ATPSynth,GAPOXNPHOSPHN-RXN	244.8
TRANSALDOL-RXN,2TRANSKETO-RXN	lethal
RXN-3341,TRANSALDOL-RXN,RIB5PISOM-RXN	lethal
CO2_tx,GAPOXNPHOSPHN-RXN,NADH_DH_ubi	246.1

Appendix C

Additional material

Content of the attached CD.

C.0.1   Additional files

Deleted reactions

List of deleted reactions. Python file (deleted.py).

Reversible reactions

List of reversible reactions. Python file (reversible.py).

Model, glucose based, ScrumPy format

Salmonella Typhimurium GSM, ScrumPy format (MetaSal.spy). Includes modules Transporters.spy, AutoModel.spy, ETC.spy, and Extra.spy.

Model, glucose based, SBML format

Salmonella Typhimurium GSM, SBML format (MetaSal.sbml).

PM integrated model, C sources, ScrumPy format

Salmonella Typhimurium GSM, ScrumPy format (MetaBioSal_C.spy). Includes modules AutoModel.spy, ETC.spy, Extra.spy, Transporters_BioLog_C_test.spy, Tx_auto_C.spy, Transporters_out.spy, and Transporters_BioLog_sugars.spy

PM integrated model, C sources, SBML format

Salmonella Typhimurium GSM, integrated with PM C-sources, SBML format (MetaSal.sbml).

PM integrated model, C sources, ScrumPy format

Salmonella Typhimurium GSM, ScrumPy format (MetaBioSal_C.spy). Includes
modules AutoModel.spy, ETC.spy, Extra.spy, Transporters_BioLog_C_test.spy,
Tx_auto_C.spy, Transporters_out.spy, and Transporters_BioLog_sugars.spy

PM integrated model, C sources, SBML format

Salmonella Typhimurium GSM, integrated with PM C-sources, SBML format
(MetaSal_C.sbml).

PM integrated model, N sources, ScrumPy format

Salmonella Typhimurium GSM, ScrumPy format (MetaBioSal_N.spy). Includes
modules AutoModel.spy, ETC.spy, Extra.spy, Transporters_BioLog_N_test.spy,
Tx_auto_N.spy, and Transporters_out.spy.

PM integrated model, N sources, SBML format

Salmonella Typhimurium GSM, integrated with PM N-sources, SBML format
(MetaSal_N.sbml).

PM integrated model, P sources, ScrumPy format

Salmonella Typhimurium GSM, ScrumPy format (MetaBioSal_P.spy). Includes
modules AutoModel.spy, ETC.spy, Extra.spy, Transporters_BioLog_P_test.spy,
Tx_auto_P.spy, and Transporters_out.spy.

PM integrated model, P sources, SBML format

Salmonella Typhimurium GSM, integrated with PM P-sources, SBML format
(MetaSal_P.sbml).

PM integrated model, S sources, ScrumPy format

Salmonella Typhimurium GSM, ScrumPy format (MetaBioSal_S.spy). Includes
modules AutoModel.spy, ETC.spy, Extra.spy, Transporters_BioLog_S_test.spy,
Tx_auto_S.spy, and Transporters_out.spy.

PM integrated model, S sources, SBML format

Salmonella Typhimurium GSM, integrated with PM S-sources, SBML format
(MetaSal_S.sbml).

Damage analysis

Damaging reaction sets from the global catabolic core model, as defined in Chapter 7. Python file, contains a Python dictionary relating reaction sets (string of reaction names, separated by comma) to imposed damage (objective value (float) or the string "lethal" if set is lethal). (damage_reac_sets.py)