



The University of
Nottingham

UNITED KINGDOM • CHINA • MALAYSIA

GENOME-SCALE
METABOLIC MODELLING
of
Clostridium autoethanogenum

Rupert Oliver John Norman

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

September 2018

ABSTRACT

Recent international directives promoting the reduced consumption of fossil fuels have warranted methods for effective carbon recycling. Subsequently, *Clostridium autoethanogenum* has attracted academic and industrial interest due to its ability to convert syngas components (CO , CO_2 & H_2) into valuable platform chemicals, including ethanol and 2,3-butanediol. Developing the metabolic conversions catalysed by *C. autoethanogenum* into an efficient bioprocess requires the accurate prediction of optimal metabolic steady states, which in turn necessitates the construction of a genome-scale model (GSM).

In this work, a genome scale model of *C. autoethanogenum* has been constructed, experimentally parameterised and validated. Unlike previously published models, optimal flux distributions computed with this network reflect the native product profile of *C. autoethanogenum* without the requirement for additional constraints. The model has been used to investigate pH- and CO -supply- induced metabolic shifts, uncovering experimentally tested bioprocess conditions which promote ethanol and 2,3-butanediol production. The genes encoding Nfn were identified as knock-out targets for improved ethanol production, however the application of a new thermodynamic analysis reveals difficulties which were confirmed experimentally.

The work presented in this thesis ultimately provides a demonstrably useful tool for metabolic and bioprocess engineering of *C. autoethanogenum* for platform-chemical production.

To my parents – Heidi and David.

ACKNOWLEDGEMENTS

I owe an enormous debt of gratitude to my supervisors at Nottingham: Charlie Hodgman, Thomas Millat, Sarah Schatschneider and John King; and at Oxford Brookes: Mark Poolman, David Fell and Hassan Hartman. I will be forever grateful to Charlie for giving me this opportunity. I also owe thanks to Nigel Minton, Klaus Winzer, Alan Burbidge and Sarah English of the SBRC for all their support.

Thanks to the BBSRC and the School of Biosciences for funding my project.

A big thanks to my computational comrades in B18: Nicole Pearcy, Vanisha Patel, James Gilbert and Mohit Dalwadi; and at Brookes: Noah Mesfin, Kailash Adhikari and Dipali Singh.

I am grateful to Dov Stekel for encouraging my interest in computational biology when I was an undergraduate.

Finally, and most importantly, thanks to my family: Mum, Dad, Bas and Georgie.

NOMENCLATURE

ABBREVIATIONS

FBA	flux balance analysis
EMA	elementary modes analysis
2,3BD	2,3-butanediol
WLP	Wood-Ljungdahl pathway
GSM	genome-scale model
PGDB	pathway genome database
GPR	gene-protein-reaction relationship
KO	knock-out
PEP	phosphoenoylpyruvate
2PG	2-phosphoglycerate
G3P	glyceraldehyde-3-phosphate
3PHP	3-phospho-hydroxypyruvate
3-P-SER	3-phospho-serine
SER	serine
GLY	glycine
GLN	glutamine
GLT	glutamate
2KG	α -keto-glutarate
THF	tetrahydrofolate
H ₂ S	hydrogen sulfide
SLP	substrate level phosphorylation
LP	linear programming
ATP	adenosine triphosphate
ADP	adenosinde diphosphate
P _i	phosphate ion
PP _i	diphosphate ion
CoA	co-enzyme A

All chemical equations shown in this work are charge and mass balanced. For some compounds, the protonation state specified by their originating database (*i.e.* MetaCyc) is not shown.

Contents

1	GENERAL INTRODUCTION	12
1.1	MOTIVATION	12
1.2	<i>Clostridium autoethanogenum</i>	13
1.2.1	Energy conservation	16
1.3	MODELLING STEADY-STATE METABOLISM	21
1.4	OBJECTIVES AND THESIS STRUCTURE	23
2	DATA AND SOFTWARE	25
2.1	DATA	25
2.1.1	Genome sequences	25
2.1.2	Databases	25
2.1.3	Biomass composition	26
2.1.4	Gene protein reaction relationships	26
2.1.5	Fermentation data	26
2.2	SOFTWARE	27
2.2.1	ScrumPy	27
2.2.2	Inkscape	27
2.2.3	Fermentation data	28
2.2.4	EDGAR	28
3	GENERAL METHODS	29
3.1	MODEL CONSTRUCTION AND CURATION	29
3.1.1	Genome annotation	29
3.1.2	Biochemical databases	30

3.1.3	Curation	30
3.2	NULLSPACE OF THE STOICHIOMETRIC MATRIX	31
3.2.1	Enzyme subsets	32
3.2.2	Generating a condensed model	34
3.3	ELEMENTARY MODES ANALYSIS	37
3.3.1	Example: CO to formaldehyde	38
3.4	FLUX BALANCE ANALYSIS	39
3.4.1	Network flux minimization	40
3.4.2	Growth yield/ATP maximization	41
3.4.3	Energy conservation	42
3.4.4	FBA scanning	42
3.4.5	Unique reaction KO combinations	43
3.5	TESTING THERMODYNAMIC FEASIBILITY	43
3.5.1	Example: CO to formaldehyde	46
4	MODEL CONSTRUCTION	51
4.1	INTRODUCTION	51
4.2	METHODS	52
4.2.1	Draft metabolic network	52
4.2.2	Model curation	53
4.2.3	Database enhancement	57
4.3	RESULTS	58
4.3.1	Draft metabolic network	58
4.3.2	Model curation	58
4.3.3	Database enhancement	72
4.3.4	Comparison with published model	74
4.4	DISCUSSION	76
5	MODEL PARAMETERISATION AND VALIDATION	78
5.1	INTRODUCTION	78
5.2	METHODS	80
5.2.1	Biomass	80

5.2.2	ATP maintenance	81
5.2.3	Carbon sources	81
5.2.4	Product spectrum	82
5.2.5	Growth rate	82
5.2.6	Objective functions	82
5.3	RESULTS	83
5.3.1	Biomass	83
5.3.2	ATP maintenance	83
5.3.3	Carbon sources	84
5.3.4	Product spectrum	84
5.3.5	Growth rate	86
5.3.6	Objective functions	86
5.4	DISCUSSION	87
6	METABOLIC SHIFTS	90
6.1	INTRODUCTION	90
6.2	METHODS	93
6.2.1	pH shift	93
6.2.2	Gas shift	94
6.3	RESULTS	94
6.3.1	pH shift	94
6.3.2	Gas shift	95
6.4	DISCUSSION	97
7	METABOLIC ENGINEERING	104
7.1	INTRODUCTION	104
7.2	METHODS	105
7.3	RESULTS	106
7.3.1	Hydrogenase inactivation	106
7.3.2	Ethanol production subnetwork	106
7.4	Experimental data	112
7.5	DISCUSSION	117

8	GENERAL DISCUSSION	119
8.1	Future work	124
8.2	Conclusion	124
9	APPENDICES	144
A1	MINIMAL GROWTH MEDIUM FOR <i>C. autoethanogenum</i> . .	144
A2	AST AMINO ACID SEQUENCE	144
A3	BIOMASS COMPOSITION	144
A4	ATP MAINTENANCE	148
A5	$\Delta_r G'^{\circ}$ OF ATP SYNTHASE AND RNF	150
A5.1	ATP SYNTHASE	150
A5.2	RNF	153

Chapter 1

GENERAL INTRODUCTION

1.1 MOTIVATION

The adverse environmental and societal consequences of continued fossil-fuel dependence represent arguably *the* defining challenge for scientific research in the 21st century [Pachauri *et al.*, 2014; Head and Gray, 2016]. The need to meet this challenge is becoming increasingly urgent as the global demand for energy continues to rise [Sieminski *et al.*, 2014].

Biofuels, *i.e.* fuels derived directly from plants or microbial fermentation, offer a way to moderate this dependence [Antoni *et al.*, 2007]. Currently, most biofuels are produced using crops as feedstock for the microbial process from which they are derived [Sims *et al.*, 2008; Hemansi *et al.*, 2018]. Due to their crop requirement, these so-called ‘1st-generation’ biofuels raise a number of concerns [Pimentel *et al.*, 2013], including:

1. the use of agricultural land,
2. the conversion of natural habitats to farmland, and
3. an increase in fertiliser production.

1st-generation biofuels thus risk endangering food security and biodiversity while incurring high energy costs associated with increased fertiliser production [Sims *et al.*, 2008; Pimentel *et al.*, 2013].

The fermentation process for biofuel production pioneered by LanzaTech (www.lanzatech.com) avoids the above problems by using gaseous feedstock drawn directly from pre-existing industrial sites, offering a means of

carbon recycling [Simpson *et al.*, 2009; Daniell *et al.*, 2012]. This process uses the anaerobic microorganism *Clostridium autoethanogenum* to convert gas (mainly carbon monoxide) to ethanol [Köpke *et al.*, 2011a]. Furthermore, this process holds potential for the production of commodity chemicals other than ethanol, such as 2,3-butanediol [Köpke *et al.*, 2011b].

Optimisation and diversification of the LanzaTech process is desirable, but requires an understanding of the *C. autoethanogenum* metabolic network which governs the chemical conversions attained. This understanding is best achieved through computational modelling methods due to the scale and complexity of metabolism [Palsson, 2015]. Thus, the motivation for this work is the need for a computational model of *C. autoethanogenum* metabolism to help develop a promising bioprocess for the alleviation of fossil-fuel dependence.

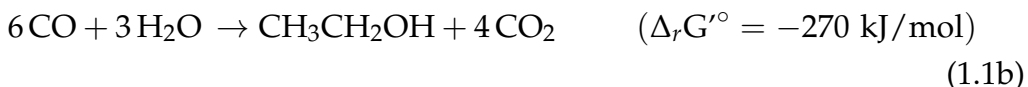
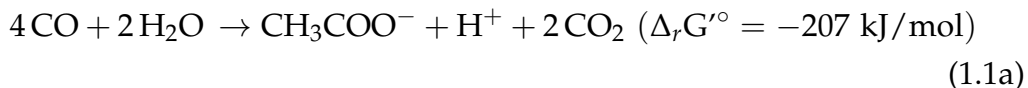
1.2 *Clostridium autoethanogenum*

The industrial biotechnology discussed in this work finds its origin in a seemingly unrelated place: animal nutrition. Abrini *et al.* inoculated bacterial growth medium with ‘hard fecal pellets’ from a rabbit to study the animal’s microflora [Abrini *et al.*, 1994]. Importantly, the enrichment cultures were incubated under an atmosphere of CO, CO₂ and N₂, and exhibited gas consumption with concomitant ethanol production. This metabolic conversion had been reported in the scientific literature only once before [Phillips *et al.*, 1993].

Microscopy and Gram staining of the successfully enriched cultures revealed a spore-forming, Gram-positive rod-shaped bacterium belonging to the genus *Clostridium*. The low genomic G+C content (25.9 %), unusual product spectrum and low pH optimum for growth (5.8–6.5) distinguished the isolate sufficiently for the authors to declare its novelty, giving the species name ‘*autoethanogenum*’ [Abrini *et al.*, 1994]. A later genus-wide phylogenetic study of *Clostridium* found *C. autoethanogenum* to be indistinguishable from *Clostridium ljungdahlii* [Stackebrandt *et al.*, 1999; Abrini *et al.*, 1994]. Thus, the early research on *C. autoethanogenum* yielded two important questions: 1) How is the organism able to grow on a gas mixture in which the only source of carbon is CO and CO₂? and 2) Is the classification of *C. autoethanogenum* as a separate species justified?

The first genome sequence for *C. autoethanogenum* gave a likely answer to the first of these questions [Bruno-Barcena *et al.*, 2013]. Genes encoding all of the enzymes required for the fixation of CO₂ to acetyl-CoA *via* the Wood-Ljungdahl Pathway (WLP) [Ljungdahl, 1986; Ljungdahl and Wood, 1965; Ljungdahl and Wood, 1969] are present in the automated annotation of the sequence (which is split into 100 sub-sequences, available from: www.ncbi.nlm.nih.gov/nuccore using accession numbers NZ_LASZX01000001.1–NZ_LASZX01000100.1).

This ancient biochemical pathway is best understood by categorising its constituent enzyme-catalysed reactions into two ‘branches’: the methyl branch and the carbonyl branch. The methyl branch proceeds by constructing a methyl group from CO₂ in a series of reducing reactions, while the carbonyl branch is simply the reduction of CO₂ to an enzyme-bound carbonyl group as catalysed by the CO dehydrogenase (CODH) component of the acetyl-CoA synthase/CODH complex (ACS/CODH). This same enzyme complex is responsible for the final step of the pathway, in which the methyl group and enzyme-bound carbonyl group are combined with co-enzyme A (CoA) to form one molecule of acetyl-CoA. Clearly, a source of reducing power is needed for these conversions to take place, a problem overcome by the encoding of two additional monofunctional CODH enzymes in the *C. autoethanogenum* genome which generate reduced ferredoxin and CO₂ from CO (and water). In this way, *C. autoethanogenum* is able to fix CO to acetyl-CoA *via* the WLP. Net reaction stoichiometries for the conversion of CO to ethanol and of CO to acetate are given below [Abrini *et al.*, 1994] (with standard Gibbs free energy of reaction values calculated using eQuilibrator [Flamholz *et al.*, 2012]):



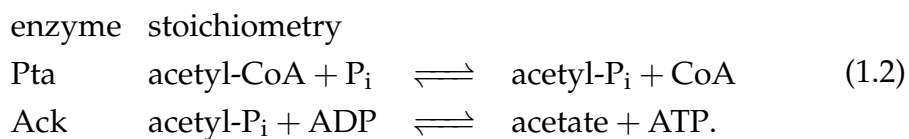
The first *C. autoethanogenum* genome sequence [Bruno-Barcena *et al.*, 2013] does not, however, answer the question of whether *C. autoethanogenum* is a separate species to *C. ljungdahlii*. The authors conclude that while *C. ljung-*

dahlia and *C. autoethanogenum* are indeed phylogenetically indistinguishable, they should be categorised as different strains due to reported phenotypic differences between the two organisms [Stackebrandt *et al.*, 1999; Bruno-Barcena *et al.*, 2013]. This uncertain answer relates back to an underlying weakness in the definition of the *Clostridium* genus, which lacks a distinct genetic basis and relies heavily on phenotype [Stackebrandt *et al.*, 1999].

Prior to the the publication of the Bruno-Barcena genome sequence [Bruno-Barcena *et al.*, 2013], studies involving *C. autoethanogenum* were conducted with an experimental bioprocess-engineering approach. These studies are largely concerned with reporting success or failure in increasing product yields after some alteration to the growth media, but do not attempt to understand the biochemical system involved [Guo *et al.*, 2010; Abubackar *et al.*, 2012; Cotter *et al.*, 2009a]. Nevertheless, important clues emerge: Abubakar *et al.* report the sensitivity of *C. autoethanogenum* fermentation products to pressure and pH [Abubackar *et al.*, 2012], while Cotter *et al.* report culture sensitivity to changes in the availability of nitrogen source [Cotter *et al.*, 2009a]. The former result, *i.e.* pH sensitivity of the acetogenic product profile during fermentation, was in fact reported in *C. ljungdahlii* before the discovery of *C. autoethanogenum* [Phillips *et al.*, 1993].

It was not until 2014 that a closed genome sequence of *C. autoethanogenum* was produced [Brown *et al.*, 2014]. Unlike the Bruno-Barcena study, this publication offers an answer to the question of whether *C. autoethanogenum* is a distinct species to *C. ljungdahlii*. Through inspection of the automated genome annotation generated in [Brown *et al.*, 2014], the authors report the presence of Clustered Regularly Interspaced Short Palindromic Repeats and Cas9-encoding genes (*i.e.* a CRISPR system) in *C. autoethanogenum* which are absent in *C. ljungdahlii* [Hsu *et al.*, 2014a]. This is a significant difference, especially in the context of industrial biotechnology, because native CRISPR systems are useful for the generation of mutants [Hsu *et al.*, 2014a].

The genome annotation presented in [Brown *et al.*, 2014] also includes enzymes of the classical acetogenic and ethanogenic pathways in *C. autoethanogenum*. Acetate is produced in a two-step pathway from acetyl-CoA proceeding *via* phosphotransacetylase (Pta) and acetate kinase (Ack):



The substrate-level phosphorylation (SLP) carried-out by Ack is an important feature of acetogenic metabolism, as discussed in the Section 1.2.1. Furthermore, acetaldehyde, the immediate precursor to ethanol, can be produced by reduction of acetyl-CoA or acetate as catalysed by acetaldehyde dehydrogenase (ADHE) or acetaldehyde:ferredoxin oxidoreductase (AFOR), respectively. The latter reaction is hypothesised to enable SLP to occur during ethanol formation in *C. autoethanogenum* [Wang *et al.*, 2013], see Fig. 1.1.

In 2011, the production of 2,3-butanediol (2,3BD) by wild-type strains of *C. autoethanogenum* was reported in [Köpke *et al.*, 2011c]. This was important because 2,3BD is a platform chemical with high global market value [Köpke *et al.*, 2011c], thus adding to the potential of *C. autoethanogenum* as the basis for an industrial bioprocess. As shown in Fig. 1.1, 2,3BD is produced from pyruvate *via* acetolactate synthase (ALS), acetolactate decarboxylase (ALDC) and butanediol dehydrogenase (BDH), and is hypothesised to function as an electron sink under certain redox conditions [Köpke *et al.*, 2011c]. Similarly, lactate production has been reported in *C. autoethanogenum* [Köpke *et al.*, 2011c]; likely catalysed by lactate dehydrogenase (LDH), providing an alternative electron sink as illustrated in Fig. 1.1.

The above text is designed to summarise relevant research on *C. autoethanogenum* conducted during the period 1994-2014 (*i.e.* between its discovery and the year this project began). A comprehensive review of research relating to *C. autoethanogenum* was published in 2018 and is included in the supplementary materials [Norman *et al.*, 2018].

1.2.1 Energy conservation

As shown in Eq. 1.2 and Fig. 1.1, acetogens can recover ATP by producing acetate *via* acetate kinase (Ack). However, the ATP consumed in the production of acetyl-phosphate *via* the WLP is equivalent to the ATP produced in its conversion to acetate, thus this site of substrate-level phosphorylation does not alone enable the cell to yield ATP. Acetogens growing autotrophically have another mechanism for generating ATP: transmembrane proton

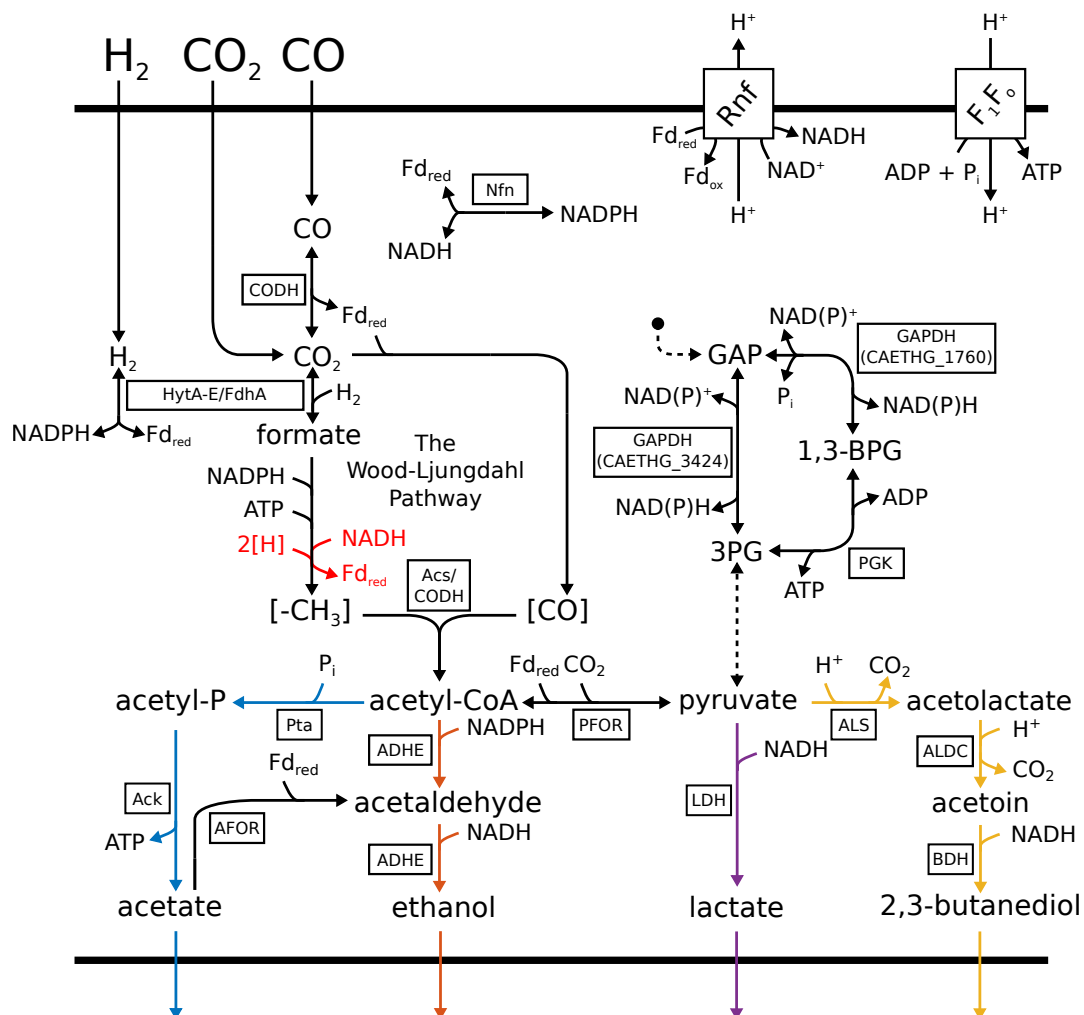


Figure 1.1: Autotrophic metabolic pathways in *C. autoethanogenum*. Arrows represent enzyme-catalysed chemical reactions, enzyme name abbreviations in boxes represent enzymes: CODH, carbon monoxide dehydrogenase; ACS/CODH, acetyl-CoA synthase/CO dehydrogenase complex; HytA-E/FdhA, NAD^+ -dependent hydrogenase/formate dehydrogenase complex; Pta, phosphotransacetylase; Ack, acetate kinase; AFOR, acetaldehyde:ferredoxin oxidoreductase; ADHE, bifunctional acetaldehyde/alcohol dehydrogenase; PFOR, pyruvate:ferredoxin oxidoreductase, LDH, lactate dehydrogenase; ALS, acetolactate synthase; ALDC, acetolactate decarboxylase; BDH, butanediol dehydrogenase. The membrane-bound F_1F_0 -type ATP synthase (F_1F_0) and Rnf complex are represented by boxes across the system boundary, with electron donors and acceptors shown. Elsewhere, only primary metabolites and electron donors are shown except for GAP metabolism, where all reactants are given. No stoichiometric weightings are included. Dotted reaction line between pyruvate represents reversible conversion of pyruvate to 3-phosphoglycerate (3PG) and dotted line between node and glyceraldehyde-3-phosphate (GAP) represents continuation of metabolic into glycolysis/gluconeogenesis-involved reactions. 1,3-bisphosphoglycerate abbreviated as 1,3-BPG. In the case of ACS/CODH, CO and the methyl-group (CH_3) are shown in square brackets to represent their association to proteins (a cobalt-iron sulphur protein and ACS/CODH, respectively). The methyl-branch of the Wood-Ljungdahl Pathway (from CO_2 to acetyl-CoA) is represented as a lumped reaction, where the red-coloured 2[H] represents an unknown electron donor, and the additional red reaction arrows represent the hypothetical transfer of electrons from $NADH$ to ferredoxin by electron bifurcation. Key production routes from acetyl-CoA are colour coded: blue = acetate, orange = ethanol, purple = lactate, yellow = 2,3-butanediol.

gradient-coupled ATP synthesis [Thauer *et al.*, 1977]. This process couples the transfer of electrons between electron carrier/donor pairs to the generation of a transmembrane cation potential, terminating with the production of ATP *via* a membrane-bound ATP synthase. For growth on CO, the scheme consists of 4 steps:

1. 2 electrons are transferred from CO to Fd by CODH,
2. 2 electrons are transferred from Fd^{2-} to NAD^+ by Rnf,
3. The exergonicity of Step 2 enables the translocation of two protons from the cytoplasm to the extracellular environment against the established proton gradient $\Delta\mu\text{H}^+$,
4. The ionic potential gained in Step 3 drives ATP synthesis by a membrane-bound F_1F_0 -type ATP synthase.

The exergonicity of the reaction described in Step 2 is the difference in redox potential between ferredoxin and NAD^+ redox couples [Berg *et al.*, 2002]:

$$\begin{aligned} E'_{\text{Fd}/\text{Fd}^{2-}} &\approx -500 \text{ mV} \\ E'_{\text{NAD}^+/\text{NADH}} &\approx -320 \text{ mV} \end{aligned} \quad (1.3)$$

The Rnf complex harnesses this potential, allowing the ‘pumping’ of protons from the cell to the extracellular environment against the established transmembrane proton gradient [Tremblay *et al.*, 2013]. The build-up of protons increases the transmembrane gradient, creating increased potential for the movement of protons back into the cytoplasm. This force drives the F_1F_0 -type ATP synthase rotor and subsequently produces ATP, as demonstrated experimentally in *A. woodii* and *C. ljungdahlii* [Tremblay *et al.*, 2013; Biegel *et al.*, 2009].

The efficiency of the ATP synthase reaction (*i.e.* number of translocated protons required for the production of 1 ATP) is hypothesised to be determined by the number of ‘c subunits’ (n_c) comprising the ATP synthase rotor (1 ATP per 120° rotation (see Eq. 1.4); $\frac{n_c}{3} \text{ H}^+$ per ATP generated) [Boyer, 1997]. Finally, it is important to note that the electron transport chain is also the primary source of reduced ferredoxin (Fd^{2-}) and NADH. Reactions

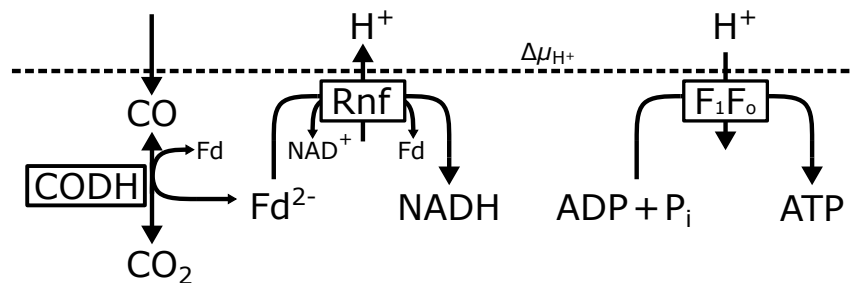
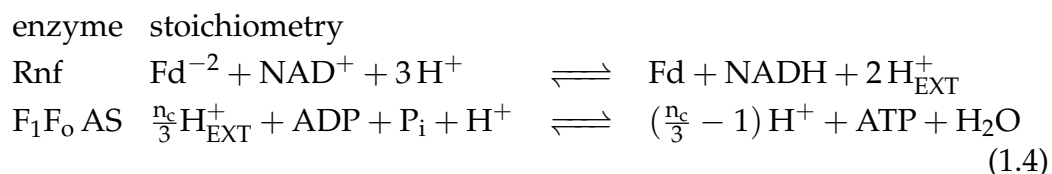


Figure 1.2: ATP synthesis coupled to electron transport and the transmembrane proton gradient ($\Delta\mu_{H^+}$). Solid arrow-headed lines and dashed lines respectively represent enzyme-catalysed reactions and the cell membrane. Labelled boxes represent enzymes (membrane-bound enzymes are drawn as boxes over the dashed line). Scheme includes electron-donors, translocated protons, ATP, ADP and P_i – other involved metabolites excluded. Stoichiometric weightings not shown.

for Rnf and F_1F_0 ATP synthase (AS) are given below. An illustration of the scheme is shown in Fig. 1.2.



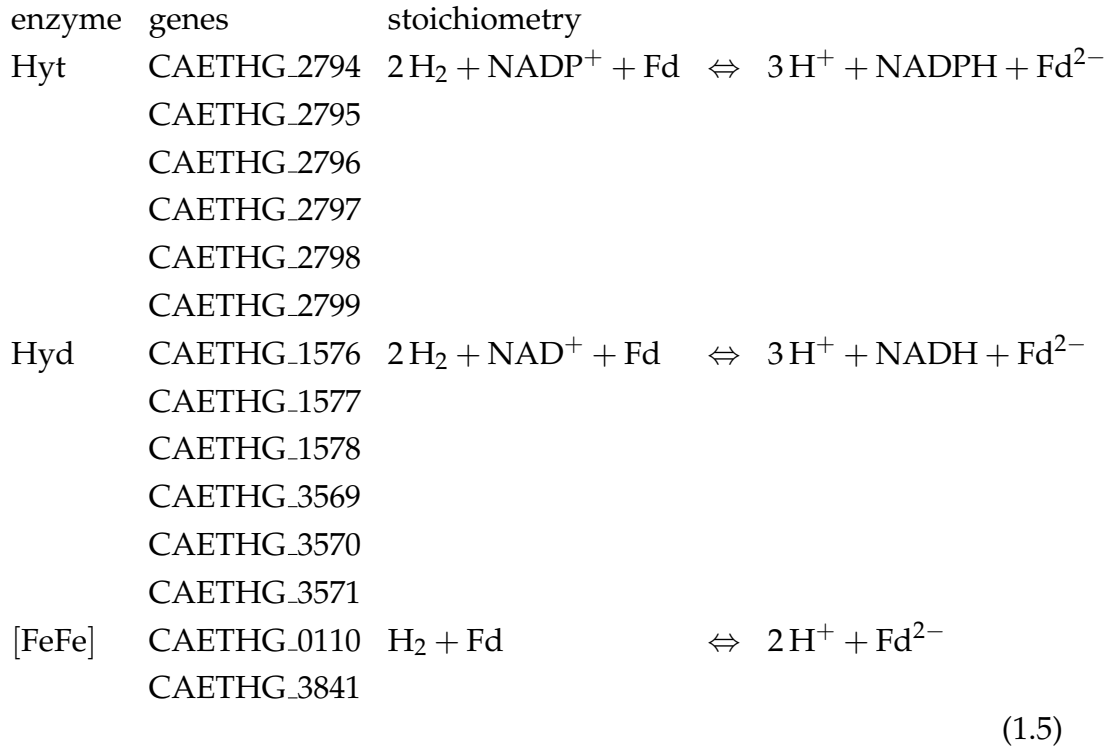
One of the first enzymological studies on *C. autoethanogenum* investigated the organism's growth on H_2 and CO_2 (where H_2 is the energy source). The outstanding discovery was that the multi-subunit hydrogenase, HytA-E, forms an enzyme-complex with formate dehydrogenase (FdhA). This complex exhibited reversible hydrogenase, formate dehydrogenase, and formate hydrogen lyase activities *in vitro*.

Another feature of HytA-E is that it facilitates electron bifurcation—a phenomenon allowing physiological electron carriers to donate their electrons to two different electron acceptors simultaneously. One of the acceptors involved will have a lower redox potential (*i.e.* E') than the donor, while the other will have a higher redox potential. The electron-bifurcating enzyme is able to harness energy released in the transfer of electrons to the higher-potential acceptor to facilitate the transfer of electrons to the lower-potential acceptor [Peters *et al.*, 2016; Herrmann *et al.*, 2008; Buckel and Barker, 1974; Li *et al.*, 2008].

In this case, HytA-E harnesses energy released in the transfer of electrons

from H_2 ($E'_{H^+/H_2} \approx -450$ mV) to $NADP^+$ ($E'_{NADP^+/NADPH} \approx -370$ mV), enabling the otherwise endergonic transfer of electrons from hydrogen to ferredoxin ($E'_{Fd/Fd^{2-}} \approx -500$ mV [Mock *et al.*, 2015]), see Eq. 1.5. This complex thus catalyses the first step in the methyl branch of the WLP (from CO_2 to formate), while also providing the means for the utilisation of hydrogen as an energy source.

In the case of growth on H_2 and CO_2 , electrons are taken directly from H_2 . With this electron source, the electron transport chain becomes more complex; according to available genome annotations [Brown *et al.*, 2014], there are 5 functional hydrogenases encoded in the *C. autoethanogenum* genome (3 different types) [Mock *et al.*, 2015], see Eq. 1.5 below.



Thus, when H_2 is the energy source there are 3 different possible reactions leading to reduced ferredoxin enabling the subsequent conversion to ATP *via* the Rnf/ATP synthase coupling.

The above information summarises what was known about energy conservation in *C. autoethanogenum* at the start of this project. The process of electron bifurcation adds a layer of complexity to energy conservation in autotrophic metabolism which sets acetogens apart from other bacteria. One point of interest throughout this work will be the relationship between the

activity of electron-bifurcating enzymes and the product spectrum of *C. autoethanogenum*.

1.3 MODELLING STEADY-STATE METABOLISM

As might be deduced from the literature so far discussed, devising a rationale for metabolic engineering through manual, logical analysis of known pathways is difficult. In fact, there are several metabolic engineering studies where enzyme/gene modifications proved ineffective or even deleterious to product yields [Juminaga *et al.*, 2011; Hauf *et al.*, 2000]. This is a well known consequence of system dynamics and network robustness [Alon, 2006; Palsson, 2015].

While mathematical techniques for modelling system dynamics are well established (*e.g.* ordinary differential equations [Jeffrey, 1989]), their application to large metabolic networks is problematic due to the large number of unknown enzyme kinetic parameters. In such cases, an approach accounting for the behaviour of the metabolic system as a whole is needed.

An alternative set of methods for computing the steady-states of large metabolic networks began development soon after the advent of genome sequencing [Watson, 1984a; Watson, 1984b; Papoutsakis, 1984; Fell and Small, 1986; Savinell and Palsson, 1992]. This set of methods, referred to as ‘structural modelling’ due to its dependence on the stoichiometric structure of a metabolic network [Fell *et al.*, 2010], is introduced in general terms below, and formally in Chapter 3. Structural modelling provides a readily achievable system-level approach to metabolic engineering. By assuming steady state, such methods are not restricted by the lack of enzymological information required to support kinetic modelling approaches. This is particularly useful in the case of industrial biotechnology since metabolic steady state coincides with the exponential growth phase during which the highest productivities are achieved.

The fundamental concept of structural modelling is the stoichiometric matrix. The stoichiometric matrix is a mathematical representation of metabolism; each row corresponds to a metabolite’s involvement in every chemical conversion possible in the subject metabolism, while each column corresponds to a reaction’s consumption and production of every metabolite present. The elements of the matrix are stoichiometric weightings, indicat-

ing how many moles of a given metabolite are consumed or produced in a given reaction.

A simple example is constructed below with a set of three reactions occurring between four hypothetical species 'A', 'B', 'C' and 'D':



For this trivial reaction system, the stoichiometric matrix is:

$$\begin{array}{c} \begin{array}{ccccc} & R1 & R2 & R3 & A_{in} & C_{out} & D_{out} \\ \begin{array}{l} A \\ B \\ C \\ D \end{array} & \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 \\ 1 & -1 & -2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \end{array} \end{array}, \tag{1.7}$$

where columns 'A_{in}', 'C_{out}' and 'D_{out}' represent transport processes enabling the uptake of A and excretion of C and D, respectively. The steady states of this system, *i.e.* the sets of reaction rates (or 'fluxes') that would prevent changes in the concentrations of all involved metabolites, are the vectors in the nullspace of this matrix (see Section 3.2) [Strang, 1976]. Optimal points in this solution space correspond to maximum or minimum values for a given set of reaction fluxes. Such points arise when limits on the range of allowable flux values are applied for a given set of reactions. Furthermore, optimal points can be found through linear programming analysis (LP) which uses an objective function to specify the set of flux variables requiring maximisation (or minimisation) [Strang, 1976]. This analysis is introduced further in Chapter 3.

Since the system shown in Eq. 1.7 has more reactions than metabolites (as is usually the case), it is 'underdetermined', meaning a set of basis vectors spans the space of possible solutions. Some of these solutions are optimal with respect to a given objective function and a set of constraints. For example, one objective function could be to maximize flux through R3, given the constraint that flux through R1 must not exceed an arbitrary value of 10. This type of problem can be solved with a computational technique called linear programming (LP), whereby the vertices of a linear solution space are

traversed programatically until the optimal solution is found [Strang, 1976]. The key point is that LP is just as easily applied to large networks as it is simple networks.

Flux Balance Analysis (FBA) is the name given to LP analysis of large metabolic networks. The first attempts to compute fluxes across a metabolic network were reported independently in [Watson, 1984b; Watson, 1984a] and [Papoutsakis, 1984]. An important development in the field was the incorporation of cellular biomass composition into the metabolic network undergoing FBA [Savinell and Palsson, 1992], enabling the prediction of growth rates and scaling fluxes towards experimentally measured values. The first metabolic model approaching genome scale (488 reactions, 343 metabolites) represented the metabolism of *Haemophilus influenzae* (which was also the first ‘free-living’ organism to have its genome sequenced) [Edwards and Palsson, 1999].

Another structural modelling method is elementary modes analysis. An elementary mode is an irreducible steady-state flux distribution over a metabolic network. ‘Irreducible’ here means that if one of the flux-carrying enzymes were removed from the flux distribution, no (non-zero) steady-state could be achieved with the remaining set of enzymes [Schuster *et al.*, 1999]. Popular platforms for structural modelling include ScrumPy [Poolman, 2006], SurreyFBA [Gevorgyan *et al.*, 2011] and COBRA [Heirendt *et al.*, 2017]. As detailed in Chapter 3, structural modelling provides a convenient and powerful way to work towards understanding the metabolism of *C. autoethanogenum* in order to meet the project’s objectives (set-out in the following section).

1.4 OBJECTIVES AND THESIS STRUCTURE

This project has three main objectives:

1. to build, parameterise and validate a genome-scale metabolic network (GSM) of *C. autoethanogenum* using available genomic and enzymological data,
2. to generate testable hypotheses regarding the tuning of bioprocess parameters for improved chemical production, and
3. to identify gene knock-out targets for improved chemical production.

Chapters 2 and 3 in this thesis respectively describe the data & software and general methods used throughout. Chapters 4 and 5 respectively describe the construction, parametrisation and validation of a GSM of *C. autoethanogenum*. Chapter 6 examines the effects of a change from carbon limitation to non-carbon limitation and variations in external pH on the subject metabolism, while Chapter 7 identifies gene knock-out targets that improve yields of commodity chemicals, particularly ethanol and 2,3-butanediol. In Chapters 6 and 7 model predictions are compared to experimental data. While each chapter contains its own discussion, a final, general discussion chapter evaluates the success of the project as a whole (in terms of meeting the metabolic engineering goals set-out above), and recapitulates any new insights into the metabolism of *C. autoethanogenum*. Finally, future directions leading on from this work are proposed.

Chapter 2

DATA AND SOFTWARE

2.1 DATA

This section details all sources of data, including genome sequences, cellular biomass composition, gene protein reaction relationships and fermentation data (*i.e.* high performance liquid chromatography (HPLC) and gas chromatography (GC)). Citations will be given explicitly where data sources are used throughout this work.

2.1.1 Genome sequences

Genome sequences for *C. autoethanogenum* referred to in this thesis are summarised in Table 2.1.

2.1.2 Databases

Two organism-specific MetaCyc Pathway Genome DataBases (PGDBs) for *C. autoethanogenum* were used in the construction of the GSM described in this thesis. Organism-specific PGDBs contain information about the genome,

Table 2.1: Published genome sequences for *Clostridia* spp. referred to throughout this work.

organism	reference	annotation	structure
<i>C. autoethanogenum</i>	[Bruno-Barcena <i>et al.</i> , 2013]	automated	100 contigs
	[Brown <i>et al.</i> , 2014]	automated	whole genome
	[Humphreys <i>et al.</i> , 2015]	manual	whole genome
<i>C. ljungdahlii</i>	[Köpke <i>et al.</i> , 2010]	automated	whole genome

Table 2.2: PGDBs for *C. autoethanogenum*.

name	genes	reactions	metabolites	reference
CAETHG	4107	1471	1044	[Brown <i>et al.</i> , 2014]
CLAU	4038	1429	1097	[Humphreys <i>et al.</i> , 2015]

proteome and metabolome of an organism, and are generated from an annotated genome sequence using the PathoLogic algorithm as implemented in Pathway Tools [Karp *et al.*, 2002]. Importantly, PGDBs associate reaction stoichiometries to enzymatic gene products, enabling the formation of a draft metabolic network. The library of MetaCyc PGDBs is split into the following three classes based on level of curation: Tier 1 (thorough manual curation), Tier 2 (preliminary manual curation) and Tier 3 (no manual curation). The PGDBs used in this work were both Tier 3. A summary of the relevant contents of the PGDBs used is given in Table 2.2.

2.1.3 Biomass composition

Details of the experiments conducted to measure the macromolecular biomass composition of *C. autoethanogenum* are given in Appendix A3.

2.1.4 Gene protein reaction relationships

46 GPR relationships were based on gene homology with *C. ljungdahlii*. This required both the genome sequence for *C. ljungdahlii* detailed in Table 2.1 and the GSM based on the same sequence [Nagarajan *et al.*, 2013]. Other gene associations were based on [Brown *et al.*, 2014] and [Humphreys *et al.*, 2015] annotations.

2.1.5 Fermentation data

All continuous fermentation data in this thesis were generated by Dr. Anne Henstra at the University of Nottingham. In Chapter 6, Fig. 6.5 shows steady-state efflux rates of liquid products in a continuous culture gas shift experiment. Fig. 6.1 also shows steady-state efflux rates of liquid products from a continuous culture pH shift experiment. In the latter experiment, data points and error values are estimates calculated from raw experimental data using matlab based software developed by Thomas Millat at the University of Nottingham.

2.2 SOFTWARE

2.2.1 ScrumPy

ScrumPy¹ [Poolman, 2006] is an open-source software package for metabolic modelling. It is written in Python and designed for use with Python's Integrated Development and Learning Environment (IDLE) and a Linux operating system, *e.g.* Ubuntu.

ScrumPy contains the modules 'PyoCyc' and 'BuildOrg' which generate a ScrumPy-formatted metabolic network from a given PGDB. BuildOrg includes automatic curation functions, including: *FixNads* – a function which maps any reaction involving the generic BioCyc metabolite classes 'NAD(P)' and 'NAD(P)H' to two reactions, one involving the NAD instances, and one involving the NADP instances; and *DelIsoforms* – a function which deletes duplicate reaction stoichiometries from the stoichiometric matrix (defined in Section 3.2), ensuring all reactions are unique. Furthermore, BuildOrg includes metabolite classes from the PGDB as metabolites (*e.g.* 'Reduced-2-4Fe-4S-Ferredoxins', 'Aldehydes', 'Purine-Bases' etc.) hence the increased number of metabolites in the raw draft network as compared to its originating database.)

2.2.2 Inkscape

Inkscape (<https://inkscape.org/>) is an open-source, general purpose graphics package used to draw pathway diagrams throughout this thesis. Its main advantage is that there are virtually no restrictions on how pathways can be represented.

A software package² for automatically colouring and annotating Inkscape pathway diagrams has been used to visualise flux distributions throughout this thesis [Pentjušs, unpublished].

¹Downloads and documentation available from mudshark.brookes.ac.uk/ScrumPy.

²For further details, including source code and descriptions, email mgpoolman@brookes.ac.uk (Mark Poolman).

2.2.3 Fermentation data

Bioreactor data were analysed using an automated python/matlab based software pipeline³ designed for the BioCommand software (BioCommand®, New Brunswick Scientific) [Gilbert & Millat, unpublished]. This software transfers bioreactor data to a web server, allowing real time updates before calibration and post processing using a suite of Matlab® scripts.

2.2.4 EDGAR

EDGAR (Efficient Database Framework for comparative Genome Analysis using BLAST score Ratios) [Blom *et al.*, 2009] was used to identify orthologous genes between different genome sequences of *C. autoethanogenum* [Humphreys *et al.*, 2015; Brown *et al.*, 2014], and between *C. autoethanogenum* and *C. ljungdahlii* [Köpke *et al.*, 2010].

³Further details of this toolset, including source code and descriptions, are available on request.

Chapter 3

GENERAL METHODS

This chapter contains general methods which are used in multiple chapters throughout this work.

3.1 MODEL CONSTRUCTION AND CURATION

3.1.1 Genome annotation

Any genome-scale model construction effort requires information about the enzymes encoded in the subject genome. This is typically provided in the form of a genome annotation. There are problems associated with automatic BLAST-based genome annotation procedures, including the ‘Chinese whispers’ effect (where mistaken annotations are transmitted from one genome to the next). Errors in annotation can affect the quality of the subsequent reconstructed network, necessitating its careful curation [Fell *et al.*, 2010; Thiele and Palsson, 2010].

For the purpose of metabolic network reconstruction, a genome annotation should provide information enabling the inference of enzyme-catalysed reaction stoichiometry. In the process of linking reaction stoichiometries to genes, one-to-many and many-to-many relationships can occur, *e.g.* an enzyme encoded by many genes and catalysing more than one chemical conversion (like HytA-E [Wang *et al.*, 2013] in *C. autoethanogenum*). This is a bioinformatics problem, requiring the assignment of multiple specific reaction stoichiometries to an enzyme with multiple possible activities. In this thesis, the PathoLogic algorithm [Karp *et al.*, 2002] is relied on to create a relational database of biochemical entities, capturing these one-to-many and

many-to-many relationships between genes and reactions. For full gene-protein-reactions relationships, which specify if genes encode subunit or iso-enzymes, manual curation is required since PathoLogic does not automatically infer this information.

3.1.2 Biochemical databases

The metabolic network reconstruction process described in this thesis is based on a pathway genome database (PGDB) generated by PathoLogic [Karp *et al.*, 2002], from a manually annotated genome sequence [Humphreys *et al.*, 2015]. PathoLogic refers to the MetaCyc database [Caspi *et al.*, 2008] during its PGDB-generation routine in order to populate the PGDB with metabolites, reactions, enzymes and genes. MetaCyc is a curated database which aims to provide a comprehensive coverage of known metabolism. Since MetaCyc is integrated with PathoLogic, it forms the primary reference for metabolic information throughout this thesis. MetaCyc also observes a standardised metabolite/reaction naming convention which is useful for software development and streamlining information transfer. Another important resource in this work is eQuilibrator [Flamholz *et al.*, 2012], which provides thermodynamic information on biochemical reactions.

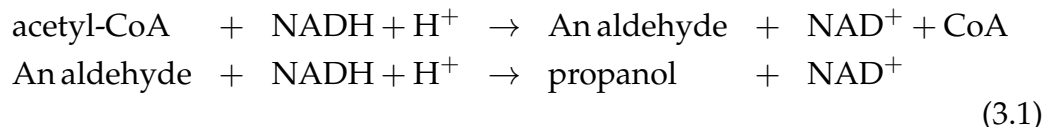
3.1.3 Curation

In the context of genome-scale models (GSMs), ‘curation’ refers to the process of identifying and rectifying errors in the definition of the model’s metabolic network, such as missing pathways and mass-imbalanced reaction stoichiometries. The aim of curation is thus to obtain a mass-balanced stoichiometric model which returns thermodynamically feasible flux distributions.

Throughout this work, the terms ‘transporter’ and ‘transport reaction’ are used interchangeably to refer to model reactions in which an external metabolite is converted to an internal metabolite (or *vice versa*).

An important part of the curation process for a draft metabolic model is the identification of generic metabolites with no defined chemical formula. This is important because problems can be caused by reactions involving both generic and specific metabolites. For example, while the reaction stoichiometries given in Eq. 3.1 are technically correct, these reactions would

allow the production of 3-carbon propanol directly from 2-carbon acetyl-CoA, clearly violating mass balance. By replacing the generic metabolite class with an appropriate metabolite instance, the inconsistency is avoided.



Conserved moieties do not require an explicit chemical formula definition. Thus generic metabolite electron/donor pairs like ‘A reduced ferredoxin’ and ‘An oxidized ferredoxin’ could remain in the network. In this way, information from the left nullspace of the stoichiometry matrix (see Section 3.2 below) is used to aid curation of the metabolic network.

One particularly important error that could arise from the automated network-generation process is the inclusion of oxygen-involving reactions. Such reactions cannot occur in the network because *C. autoethanogenum* is a strict anaerobe. There were four possibilities for the inclusion of oxygen-involving reactions: 1) the corresponding enzyme was erroneously added by pathologic to fill a gap, 2) the enzyme is encoded by *C. autoethanogenum* but is an anaerobic version of an enzyme usually studied in aerobic organisms, 3) the enzyme is a genuine oxygen-stress enzyme, or 4) the enzyme is capable of catalysing a reaction which might produce oxygen, and thus an additional oxygen stoichiometry was added by pathologic.

3.2 NULLSPACE OF THE STOICHIOMETRIC MATRIX

A stoichiometric matrix, here denoted $\mathbf{N}$, is an $m \times n$ matrix representing a set of m metabolites ($m_1 \dots m_m$) reacting according to a set of n reaction stoichiometries. These reactions take place at rates $v_1 \dots v_n$, which, along with the reaction stoichiometries, determine the rates of change of the metabolites ($\frac{dm_1}{dt} \dots \frac{dm_m}{dt}$). This system can be represented formally as the dot product of $\mathbf{N}$ with the vector of reaction rates $\mathbf{v}$ giving rise to a vector of metabolite concentration derivatives, $\mathbf{x}$:

$$\mathbf{x} = \mathbf{N}\mathbf{v} \quad (3.2)$$

$$\begin{bmatrix} \frac{dm_1}{dt} \\ \vdots \\ \frac{dm_m}{dt} \end{bmatrix} = \begin{bmatrix} n_{1,1} & \dots & n_{1,n} \\ \vdots & \ddots & \vdots \\ n_{m,1} & \dots & n_{m,n} \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

The nullspace is the space of all vectors, $\mathbf{v}$, which satisfy the special case where $\mathbf{N}\mathbf{v} = 0$, *i.e.* there is a steady state with respect to the metabolite concentrations. Non-trivial nullspace vectors (*i.e.* vectors other than the zero vector) will only exist if linear dependencies exist in the columns of the matrix, *i.e.* the rank of the matrix r is less than the number of columns ($r < n$). A basis for the nullspace of the stoichiometric matrix is a set of $(n - \text{rank}(\mathbf{N}))$ vectors which span the space of possible solutions to $\mathbf{N}\mathbf{v} = 0$. One such basis, the kernel matrix (denoted $\mathbf{K}$), can be computed with Gauss-Jordan elimination [Strang, 1976] (as implemented in ScrumPy [Poolman, 2006]). $\mathbf{K}$ contains information about steady-state behaviour, such as enzyme subsets [Gevorgyan *et al.*, 2008].

3.2.1 Enzyme subsets

The nullspace vectors of the stoichiometric matrix enable the identification of enzyme subsets (ES). These are sets of enzymes which carry flux at a fixed ratio at steady state. Enzyme subsets can reveal ‘amplifying’ or ‘dampening’ relationships between enzymes, *i.e.* the case where the flux carried by one enzyme in a subset is equal to a fraction of the flux carried by another. Thus, in the case where such a relationship applies, an increase in flux carried by one enzyme can be achieved with a relatively small increase in the flux of another (and *vice versa*), offering a potential rationale for identifying gene overexpression targets. Inconsistent enzyme subsets (IESs) are enzyme subsets in which the relative flux values for one or more of the involved reactions violates an irreversibility constraint [Hartman *et al.*, 2014; Gevorgyan *et al.*, 2008].

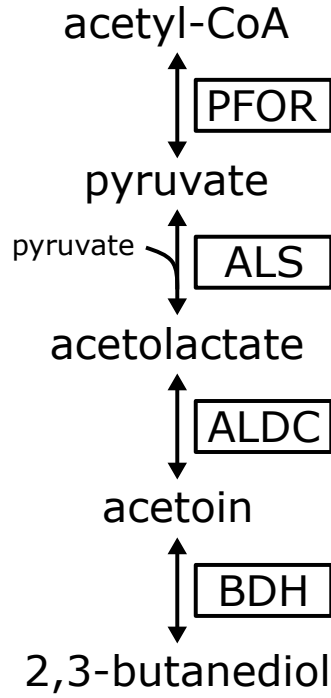


Figure 3.1: A diagram of the simplified linear pathway from acetyl-CoA to 2,3-butanediol (only primary metabolites are shown). Arrows represent chemical conversions between the given set of metabolites, with the corresponding enzyme names in adjacent boxes (PFOR = pyruvate:ferredoxin oxidoreductase, ALS = acetolactate synthase, ALDC = acetolactate decarboxylase, BDH = butanediol dehydrogenase).

As an example, the stoichiometric matrix, N , for the simple pathway given in Fig. 3.1 is

$$\begin{array}{c}
 \text{acetyl-CoA}_{\text{IN}} \quad \text{PFOR} \quad \text{ALS} \quad \text{ALDC} \quad \text{BDH} \quad 2,3\text{-BD}_{\text{OUT}} \\
 \begin{array}{l}
 \text{acetyl-CoA} \\
 \text{pyruvate} \\
 \text{acetolactate} \\
 \text{acetoin} \\
 2,3\text{-BD}
 \end{array}
 \left[\begin{array}{cccccc}
 1 & -1 & 0 & 0 & 0 & 0 \\
 0 & 1 & -2 & 0 & 0 & 0 \\
 0 & 0 & 1 & -1 & 0 & 0 \\
 0 & 0 & 0 & 1 & -1 & 0 \\
 0 & 0 & 0 & 0 & 1 & -1
 \end{array} \right], \quad (3.3)
 \end{array}$$

where $\text{acetyl-CoA}_{\text{IN}}$ and $2,3\text{-BD}_{\text{OUT}}$ represent a source of acetyl-CoA and a sink for 2,3-BD, respectively. The kernel matrix, K , for this network is given below:

$$\mathbf{K} = \begin{bmatrix} 2 \\ 2 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad (3.4)$$

where the n elements of $\mathbf{K}$ are relative flux values across the n reactions (columns) of $\mathbf{N}$. As there is just one column vector in $\mathbf{K}$ and all reactions carry flux, the entire pathway is an enzyme subset. Furthermore, the requirement of 2 moles of pyruvate for the formation of 1 mole of acetolactate gives rise to a flux dampening relationship. *I.e.*, the superior overexpression targets for increased 2,3-BD production are ALS, ALDC and BDH because the butanediol efflux is exactly equal to the flux carried by these enzymes at steady state, whereas it is equal to half the flux carried by PFOR. This assumes that the abundances of each enzyme in the subset are equally (and linearly) sensitive to the overexpression of their encoding genes.

3.2.2 Generating a condensed model

A model is condensed by representing all enzyme subsets (ES) as lumped reactions (and removing the individual reactions comprising the subsets from the model). Stoichiometry vectors representing lumped ES reactions are given by a linear combination of the column vectors of $\mathbf{N}$ which correspond to reactions in the ES.

$$\mathbf{n}_i^{condensed} = v_{n_1} \mathbf{n}_1^{ES_i} + \dots + v_{n_j} \mathbf{n}_j^{ES_i} \quad (3.5)$$

where $\mathbf{n}_i^{condensed}$ denotes a column-vector representing the i^{th} enzyme subset (ES_i , consisting of j reactions). $\mathbf{n}_k^{ES_i}$ is the column vector representing the k^{th} reaction in the enzyme subset, and v_{n_k} is the relative flux carried by the k^{th} reaction in the enzyme subset.

Example: CO to formaldehyde

Here is a simple example of model condensation based on a simple metabolic network comprising carbon monoxide dehydrogenase (CODH), a reversible NAD-dependent formate dehydrogenase catalysed reaction (Fdh_{NAD}),

an irreversible ferredoxin-dependent formate dehydrogenase catalysed reaction (Fdh_{NAD}) and formaldehyde dehydrogenase (FNOR). In this example, the metabolic pathway formed by these reactions includes transporters for CO , CO_2 , formate and formaldehyde (see Fig. 3.2).

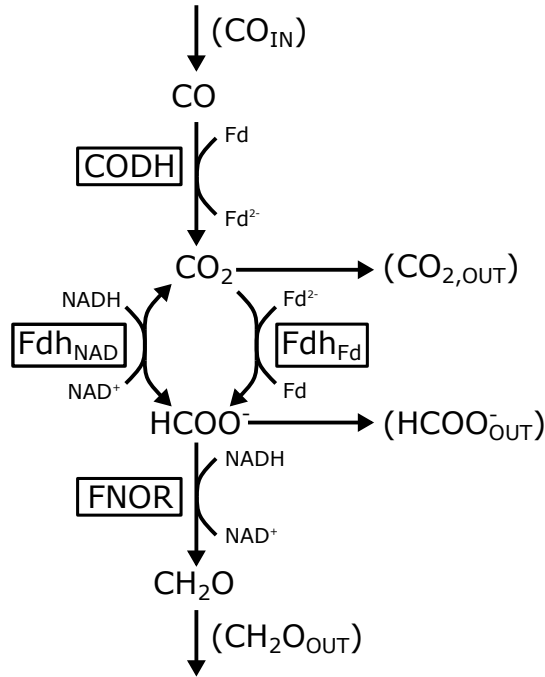


Figure 3.2: Hypothetical pathway from carbon monoxide to formaldehyde. Arrows represent enzyme-catalysed reactions between metabolites, of which only those containing carbon and/or accepting/donating electrons are shown. Enzyme names are given in boxes.

The stoichiometric matrix for this simple system is given below:

$$\mathbf{N} = \begin{array}{c} \begin{array}{c} \text{CO} \\ \text{CO}_2 \\ \text{HCOO}^- \\ \text{CH}_2\text{O} \\ \text{NADH} \\ \text{NAD}^+ \\ \text{Fd}^{2-} \\ \text{Fd} \end{array} \begin{array}{c} \text{CO}_{\text{IN}} \\ \text{CODH} \\ \text{Fdh}_{\text{NAD}} \\ \text{Fdh}_{\text{Fd}} \\ \text{FNOR} \\ \text{CH}_2\text{O}_{\text{OUT}} \\ \text{CO}_{2,\text{OUT}} \\ \text{HCOO}^-_{\text{OUT}} \end{array} \end{array} \quad (3.6)$$

The kernel matrix for this network, calculated with Gauss-Jordan elimination [Strang, 1976] as implemented in ScrumPy [Poolman, 2006]:

$$\mathbf{K} = \begin{bmatrix} 2 & 1 \\ 2 & 1 \\ -1 & 0 \\ 2 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (3.7)$$

Enzymes subsets are defined as sets of reactions carrying flux in a fixed ratio in all columns of $\mathbf{K}$. Thus, $\mathbf{K}$ shows that reactions corresponding to rows 1, 2 and 4 form an enzyme subset, while rows 3, 5, 6 and 7 form another. These reaction sets can each be lumped according to Eq. 3.5 and removed from $\mathbf{N}$ (the subsequent lumped reaction forming a new column 'ES'). The resulting condensed model's stoichiometric matrix is given below:

$$\mathbf{N}^{condensed} = \begin{array}{c} \text{CO} \\ \text{CO}_2 \\ \text{HCOO}^- \\ \text{C}_2\text{HO} \\ \text{NADH} \\ \text{NAD}^+ \\ \text{Fd}^{2-} \\ \text{Fd} \end{array} \begin{array}{c} \text{Ess 1} \\ \text{Ess 2} \\ \text{HCOO}^-_{\text{OUT}} \end{array} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 2 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (3.8)$$

This condensation can improve the computational performance of certain model analyses, particularly for very large networks. For example, time taken to compute the full set of elementary modes (see Section 3.3) increases with the size and complexity of the subject metabolic network.

3.3 ELEMENTARY MODES ANALYSIS

An elementary mode is an irreducible steady-state flux distribution over a metabolic network. ‘Irreducible’ here means that if one of the flux-carrying enzymes were removed from the flux distribution, no (non-zero) steady-state could be achieved with the remaining set of enzymes [Schuster *et al.*, 1999].

More formally, according to [Schuster and Hilgetag, 1994], a flux vector, $\mathbf{v}$, is an elementary mode if the condition of simplicity is met (in addition to the steady state and reversibility constraints set out in Section 3.2). The condition of simplicity is met if there are no two vectors $\mathbf{v}'$ and $\mathbf{v}''$ consistent with the following:

- A non-negative linear combination of $\mathbf{v}'$ and $\mathbf{v}''$ gives $\mathbf{v}$.
- $\mathbf{v}'$ and $\mathbf{v}''$ obey steady state and reversibility constraints.
- $\mathbf{v}'$ and $\mathbf{v}''$ both contain at least the same number of zero elements as $\mathbf{v}$, and one contains more zero elements than $\mathbf{v}$.
- for all indices i corresponding to boundary reactions, the components are not of opposite sign, $\text{sgn}(v_{1,i}) \neq \text{sgn}(v_{2,i})$.

The basic algorithm, applicable to an $n \times r$ stoichiometric matrix, $\mathbf{N}$, containing r exclusively irreversible reactions, works by constructing a series of n matrices ($\mathbf{T}^{(1)}, \mathbf{T}^{(2)}, \dots, \mathbf{T}^{(n)}$) called tableaux (where in this case n is the number of metabolites in the system). The starting tableau is constructed from the transpose of the stoichiometry matrix and an $r \times r$ identity matrix: $\mathbf{T}^{(0)} = (\mathbf{I}, \mathbf{N}^T)$. Following this, $\mathbf{T}^{(j+1)}$ is obtained from $\mathbf{T}^{(j)}$ by non-negative linear combination of its rows so that the $(r + j + 1)^{\text{th}}$ column is zero. Rows are either discarded or retained based on whether the criteria for simplicity set out above are met during comparison of indices of zero elements. The first r columns of $\mathbf{T}^{(n)}$ contain the elementary modes [Schuster and Schuster, 1993].

This basic algorithm is adapted to networks with some reversible reactions by partitioning the stoichiometric matrix forming $\mathbf{T}^{(0)}$ into reversible and irreversible parts [Schuster and Hilgetag, 1994]. A number of algorithms

have since been proposed for the computation of the elementary modes, including [Schuster and Schuster, 1993; Gagneur and Klamt, 2004; Wagner, 2004; Pérès *et al.*, 2014].

3.3.1 Example: CO to formaldehyde

Here is a simple example of EMA based on the system represented in Eq. 3.6 and illustrated in Fig. 3.2. The elementary modes of this network, illustrated in Fig. 3.3, are the columns of the matrix E, given below:

$$E = \begin{bmatrix} 2 & 1 \\ 2 & 1 \\ -1 & 0 \\ 2 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (3.9)$$

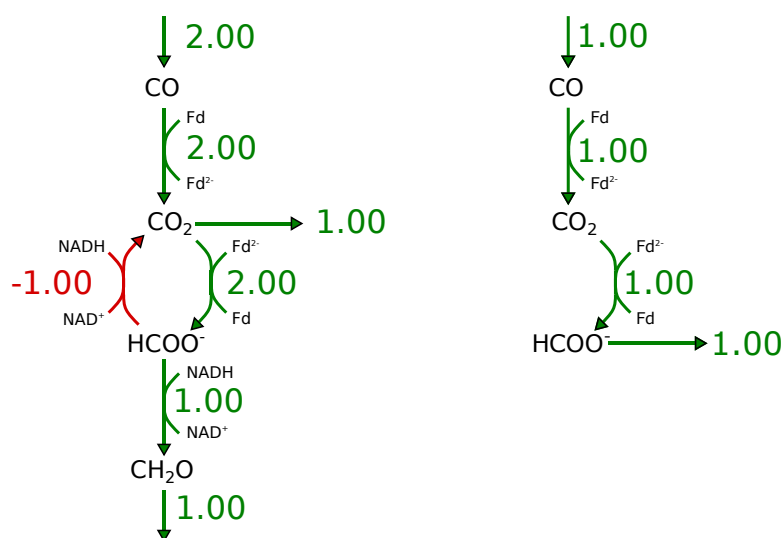


Figure 3.3: Elementary modes of the hypothetical pathway from carbon monoxide to formaldehyde. Green and red arrows represent relative flux carried in forward and reverse directions, respectively. (Reactions carrying no flux excluded.)

These elementary modes span the nullspace of this network’s stoichiometry matrix, meaning any steady state solution, $\mathbf{v}$, can be represented as a linear combination of rows $\mathbf{e}_1$ and $\mathbf{e}_2$ ($\mathbf{v} = \alpha\mathbf{e}_1 + \beta\mathbf{e}_2$, where α and β are positive scalars). Given this equation, statements can be made about the steady-state behaviour of the network, *e.g.* CO₂ production always accompanies formaldehyde production and CODH catalyses an essential reaction (since it is present in both elementary modes).

Min cut set analysis. An important question that can be answered with elementary modes is: Which reactions are associated with production of a particular compound? In the case of formate, this question can be answered by identifying reactions that carry flux in elementary modes involving formate production. Furthermore, formate production can be guaranteed to accompany steady-state solutions if sets of reactions are found which satisfy the following conditions:

1. members of the set carry no flux in at least one formate-producing mode, and
2. members of the set carry non-zero flux in all other modes.

In this example, both FNOR and Fdh_{NAD} fulfil this criteria, and are the minimum-cut set [Klamt and Gilles, 2004] with respect to formate production.

3.4 FLUX BALANCE ANALYSIS

Flux Balance Analysis (FBA) is the application of linear programming (LP) to large metabolic networks represented by a matrix of reaction stoichiometries. LP is a computational method which identifies optimal points in a linear solution space, *e.g.* the space of possible reaction rates associated with a set of reacting metabolites. The fundamental assumption behind FBA is the steady-state, *i.e.* that the concentration of metabolites in the network are unchanging, thus the associated vector of optimal reaction rates are found in the nullspace of the stoichiometric matrix. This fundamental subspace forms the foundation of almost all the analysis described in this work. A rigorous definition is given below.

Since FBA (and EMA) return theoretical extremes of the nullspace [Müller *et al.*, 2014], it is difficult to say to what extent these should match real-life metabolism. Rather, these techniques give an idea of the *capabilities* of the

metabolic network.

3.4.1 Network flux minimization

The preferred objective function in this study is minimization of absolute flux ($|v_i|$) through all reactions (3.10). For a mass-balanced metabolic network, total network flux minimization corresponds to a minimisation of protein investment across the metabolic network, representing enzymatic economy [Holzhütter, 2004]. A further advantage of flux minimisation is that it reduces the occurrence of thermodynamically infeasible energy-generating cycles [Holzhütter, 2004]:

$$\begin{aligned} & \text{minimize: } |\mathbf{v}_{\text{cat}}| \quad (\text{objective function}) \\ & \text{subject to } \begin{cases} \mathbf{N}\mathbf{v} = \mathbf{0} & (\text{steady state}) \\ v_{\text{ATPase}} = Y_{\text{xATP}} \times \mu + m_{\text{ATP}} \\ v_{k_i} = \mu \times b_i; i \in \{1, \dots, N_{\text{bio}}\} \\ v_{p_j} \leq 0; j \in \{1, \dots, N_{\text{prods}}\} \end{cases} \end{aligned} \quad (3.10)$$

where $\mathbf{N}$ is the stoichiometric matrix and $\mathbf{v}$ is a vector of all reaction flux values included in the metabolic network. $\mathbf{v}_{\text{cat}}$ is the set of network reactions catalyzed by enzymes *i.e.* the subset of $\mathbf{v}$ which excludes transport by diffusion (in this study: CO, CO₂ and H₂ and spontaneous reactions (which are automatically annotated as such by Pathway Tools). Y_{xATP} and m_{ATP} are equivalent to the growth and non-growth associated ATP maintenance costs (GAM and NGAM), respectively, and μ is the specific growth rate. v_{ATPase} is the flux through the network's ATPase reaction ($\text{ATP} + \text{H}_2\text{O} \longrightarrow \text{ADP} + \text{P}_i + \text{H}^+$). Reaction fluxes representing the production of N_{bio} biomass precursors are denoted v_{k_i} , where $\mathbf{k}$ is a vector containing the corresponding indices of these reactions in $\mathbf{v}$. The vector $\mathbf{b}$ contains relative abundance values for each biomass precursor which, when multiplied by μ , determine the value of each v_{k_i} . Reaction fluxes representing the transport of non-biomass-associated metabolites not required as growth substrates (*i.e.* potential products) are denoted v_{p_j} , where $\mathbf{p}$ is a vector of length N_{prods} containing the corresponding indices in $\mathbf{v}$.

3.4.2 Growth yield/ATP maximization

An alternative objective function is maximisation of biomass precursor production, an approximation of growth rate (μ) optimisation [Feist and Palsson, 2010]:

$$\begin{aligned} & \text{maximize: } \mu \\ & \text{subject to } \left\{ \begin{array}{l} \mathbf{N}\mathbf{v} = \mathbf{0} \\ v_{\text{ATPase}} = m_{\text{ATP}} \\ v_{k_i} = g_i; i \in \{1, \dots, N_{\text{growth}}\} \\ v_{p_j} \leq 0; j \in \{1, \dots, N_{\text{prods}}\} \end{array} \right. \end{aligned} \quad (3.11)$$

Here, maximisation of μ corresponds to the maximisation of a lumped chemical equation including all biomass precursors as reactants with their relative abundance given as stoichiometric coefficients. Y_{xATP} (GAM) is incorporated in the lumped equation as the stoichiometric coefficient on ATP. Reaction fluxes representing each of N_{growth} substrate-uptake reactions are denoted v_{k_i} , where $\mathbf{k}$ is a vector containing the corresponding indices in $\mathbf{v}$. The vector $\mathbf{g}$ contains fixed flux values for each of N_{growth} substrate-uptake reactions. In this study, the fixed substrate uptake rate is taken as the average experimentally observed uptake rate at steady state.

It is possible that organisms adapted for growth in low-nutrient environments (like *C. autoethanogenum*) regulate their metabolism according to this objective in order to gain an advantage over competitors. Counter arguments supporting the use of growth maximisation exist but are usually based on the assumption of a nutrient-rich environment [Feist and Palsson, 2010]. An important question is the extent to which simple microorganisms like *C. autoethanogenum* change their objectives.

An alternative form of this LP problem maximises ATP dissipation, *i.e.* flux carried by ATPase (v_{ATPase}) in the ATP hydrolysing direction. Solutions to this problem achieve optimal ATP yields. Furthermore, a complete hierarchy of products ordered in terms of Y_{ATP} can be generated by iteratively computing the Y_{ATP} -optimal solution while blocking (*i.e.* constraining to 0) flux across the transporter of the product formed in the previous iteration until ATP production is infeasible.

3.4.3 Energy conservation

The following analysis ensured thermodynamic consistency across the metabolic network: FBA was computed with fixed positive ATP-demand and without uptake of any carbon, nutrient or energy source in line with methods described in [Hartman *et al.*, 2014]:

$$\begin{aligned} & \text{minimize: } |\mathbf{v}| \\ & \text{subject to } \begin{cases} \mathbf{N}\mathbf{v} = \mathbf{0} \\ v_{\text{ATPase}} = 1 \\ v_{t_i} \leq 0; i \in \{1, \dots, N_{\text{trans}}\} \end{cases} \end{aligned} \quad (3.12)$$

where transport reactions are denoted v_{t_i} , and $\mathbf{t}$ is a vector of length N_{trans} containing all indices of transport reactions in $\mathbf{v}$. Any subsequent solution included a thermodynamic error, meaning definitions for one-or-more of the participant reactions required amendment or removal.

3.4.4 FBA scanning

An FBA scan is the repeated application of FBA with incremental changes in one or more constraints. The equation below (Eq. 3.13) defines an LP problem in which the variable v_{scan} is varied.

$$\begin{aligned} & \text{minimize: } |\mathbf{v}| \\ & \text{subject to } \begin{cases} \mathbf{N}\mathbf{v} = \mathbf{0} \\ v_{\text{ATPase}} = \text{NGAM} \\ v_{\text{biomass}} = \mu \\ v_{t_i} \leq 0; i \in \{1, \dots, N_{\text{trans}}\} \\ \min \leq v_{\text{scan}} \leq \max \end{cases} \end{aligned} \quad (3.13)$$

In the equation above, v_{t_i} , $\mathbf{t}$ and N_{trans} are defined as in Eq. 3.12. μ and NGAM are the growth rate and non-growth associated maintenance cost, respectively. The inclusion of these constraints ensure that the extracted network is valid in the context of biomass production and maintenance. Most important is the range of fixed fluxes, $\min \leq v_{\text{scan}} \leq \max$, for a given re-

action v_{scan} over which optimal solutions are calculated. A range is chosen such that at the maximum of the range, the absolute value of reactions are increasing monotonically with v_{scan} . Reactions whose flux-change is greater than a given cut-off value (Δv) are termed ‘changers’. For instance, if v_{scan} is chosen to be the transport reaction for a given compound of interest, the changers obtained will be a set of reactions associated with the compound’s production/consumption (depending of the range of fluxes chosen). In this way, changers can be used to define a subnetwork of reactions in the GSM associated with the behaviour associated with v_{scan} . A Python module for the extraction of changers from a ScrumPy FBA scan dataset (filename: ‘changers.py’) is included in the supplementary materials.

As outlined in Section 3.3, the elementary modes of a metabolic network give a complete description of its steady-state capabilities. Standard algorithms for the computation of the elementary modes are subject to a combinatorial explosion [Schuster and Schuster, 1993; Gagneur and Klamt, 2004; Wagner, 2004], making its application to smaller networks advantageous. The definition of a subnetwork of changer reactions thus provides a convenient way to obtain a smaller network amenable to the computation of elementary modes.

3.4.5 Unique reaction KO combinations

An FBA-based algorithm for the identification of gene KO targets for improved product yields is implemented as a function (‘KO_combinations’) in the ‘flux_balance_analysis’ Python module provided in the supplementary materials (see ‘flux_balance_analysis.py’). The function sets the flux of unique combinations of reactions to 0 for FBA while monitoring optimal solutions for increased product yields. In this study, unique reaction combinations of set length 2 and 3 were searched for.

An analogous approach could be applied to investigate overexpression but would require a rationale regarding the extent to which flux values of candidate reactions are increased.

3.5 TESTING THERMODYNAMIC FEASIBILITY

The thermodynamic favourability of a reaction depends on several factors: pH, ionic strength, temperature, pressure and the activities of the sub-

strates and products involved. One measure of the favourability of a chemical reaction is the Gibbs free energy of reaction, $\Delta_r G'$ [Gibbs, 1879]. This is defined according to the following equation,

$$\begin{aligned}\Delta_r G' &= \Delta_r G'^{\circ} + RT \ln Q_r \\ \Delta_r G'^{\circ} &= -RT \ln K_{eq}\end{aligned}\quad (3.14)$$

where $\Delta_r G'^{\circ}$ is the standard Gibbs free energy of reaction, R is the gas constant (8.314 J/mol), T is the temperature (298 °K), Q_r is the reaction quotient and K_{eq} is the equilibrium constant. Furthermore, for the purposes of this study, the pH and ionic strength under standard conditions are 6 and 0.1, respectively. This pH value is chosen because it lies within the optimum pH range for growth of *C. autoethanogenum* [Abrini *et al.*, 1994]. The equilibrium constant is the ratio of product activities over reactant activities in the special case where the overall flux of the subject chemical reaction is zero (*i.e.* the system is at equilibrium). The reaction quotient is the specified ratio of activities of products ($p_1, \dots, p_n$) over reactants ($s_1, \dots, s_m$). Thus, at equilibrium,

$$K_{eq} = Q_r = \frac{p_1^{x_1} \times \dots \times p_n^{x_n}}{s_1^{y_1} \times \dots \times s_m^{y_m}} \quad (3.15)$$

where x_i and y_j are the stoichiometric coefficients for the i^{th} product and j^{th} substrate, respectively. While the relationship between metabolite activities and $\Delta_r G'$ is non-linear, the relationship between $\Delta_r G'$ and the natural logarithm of the metabolite concentrations is linear (at constant pH, temperature, pressure and ionic strength). Thus, in the case where a given reaction is favourable (*i.e.* $\Delta_r G' < 0$), Eq. 3.14 can be re-written as the following linear inequality:

$$\sum_{i=1}^n x_i \ln p_i - \sum_{j=1}^m y_j \ln s_j < -\frac{\Delta_r G'^{\circ}}{RT} \quad (3.16)$$

While FBA and EMA do not explicitly incorporate the above relationships, they do use crude thermodynamic information through the definition of reaction reversibility constraints. These are usually based on the standard Gibbs energy of reaction $\Delta_r G'^{\circ}$ value of each reaction in question, where

the standard Gibbs energy of reaction is the difference in formation energies between the substrates and products (under standard conditions [Flamholz *et al.*, 2012]). The first problem with hard irreversibility constraints is that, in most cases, an arbitrary $\Delta_r G'^\circ$ cut-off value is used to decide whether a reaction is likely to proceed in one direction or the other. In reality, the direction of flux depends on the concentrations of the metabolites involved (according to Eq. 3.14). This leads to the danger that a given reaction may be constrained as irreversible in one direction when sets of physiologically feasible metabolite concentrations exist which would allow the reaction to carry flux in the opposite direction. Conversely, a distribution of computed fluxes consistent with steady-state and reversibility constraints will nonetheless be infeasible if no set of metabolite concentrations would enable all $\Delta_r G'$ values to agree with the computed flux directions.

Peres *et al.* (2018) propose the following linear program (LP) problem to check whether a set of metabolite concentrations exists which would enable negative $\Delta_r G'$ in the direction of flux for all reactions in a given flux distribution:

$$\begin{aligned} & \text{minimize: } \mathbf{c}^T \mathbf{0} \\ & \text{subject to } \begin{cases} \mathbf{N}_v^T \mathbf{c} < \mathbf{g} \\ c_{min} \leq c \leq c_{max} \end{cases} \end{aligned} \quad (3.17)$$

In the above equation, $\mathbf{N}_v^T$ is the stoichiometry matrix excluding columns representing reactions which carry no flux in a flux distribution, $\mathbf{v}$, derived from FBA or EMA of the full stoichiometry matrix ($\mathbf{N}$). $\mathbf{c}$ is a vector of log metabolite concentrations $[\ln c_i, \dots, \ln c_{m_v}]$ for all m_v metabolites involved in the subject flux distribution, $\mathbf{v}$. $\mathbf{g}$ is a vector of $\Delta_r G'^\circ$ values (divided by $-RT$) for each of the n_v internal reactions involved. Since the coefficients in the objective function (min: $\mathbf{c}^T \mathbf{0}$) are all zero, the LP will return the first solution encountered in the optimisation routine. Furthermore, since the matrix involved may be underdetermined, there may be an infinite number of solutions. To further constrain the model, a range of allowable metabolite concentrations may be imposed: $c_{min} \leq c \leq c_{max}$. The flux distribution, $\mathbf{v}$, is thermodynamically infeasible if no solution to this LP exists.

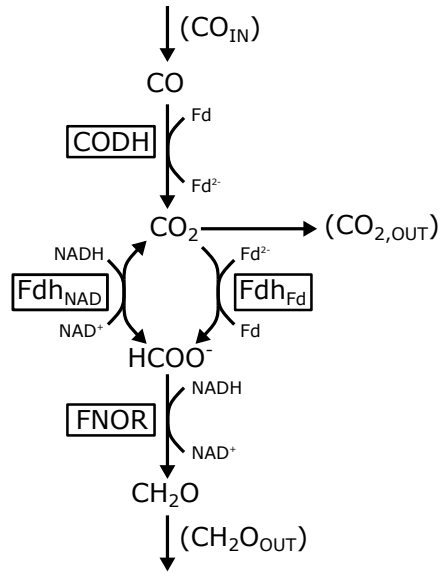


Figure 3.4: Hypothetical pathway from carbon monoxide to formaldehyde. Arrows represent enzyme catalysed reactions between metabolites. Enzyme names are given in boxes.

3.5.1 Example: CO to formaldehyde

Here is a simple example of the method set out above. Consider the metabolic pathway given in Fig. 3.4. The system represented by the stoichiometric matrix below has one elementary mode: $E = (2, 2, -1, 2, 1, 1, 1)$.

	CO_{IN}	CODH	Fdh_{NAD}	Fdh_{Fd}	FNOR	CH_2O_{OUT}	$CO_{2,OUT}$
CO	1	-1	0	0	0	0	0
CO_2	0	1	-1	-1	0	0	-1
$HCOO^-$	0	0	1	1	-1	0	0
CH_2O	0	0	0	0	1	-1	0
NADH	0	0	-1	0	-1	0	0
NAD^+	0	0	1	0	1	0	0
Fd^{2-}	0	1	0	-1	0	0	0
Fd	0	-1	0	1	0	0	0

(3.18)

The elementary mode for this system (displayed in Fig. 3.5) features two Fdh-catalysed reactions, one carrying flux in the forward direction which accepts electrons from ferredoxin, and one carrying flux in the reverse direction which donates electrons to NAD^+ . The idea of a chemical reaction

simultaneously carrying flux in both forward and reverse directions seems problematic from a thermodynamic perspective. But does the use of different electron acceptor/donor pairs, as in this case, make it feasible? Furthermore, formaldehyde dehydrogenase is largely favourable in the formaldehyde dehydrogenating direction under standard conditions (see Table 3.1). This means that the direction of flux carried in the displayed elementary mode (Fig. 3.5) would not normally be considered feasible. But is there a set of metabolite concentrations that would make it feasible? These are the questions that will be answered with the LP described in Section 3.5.

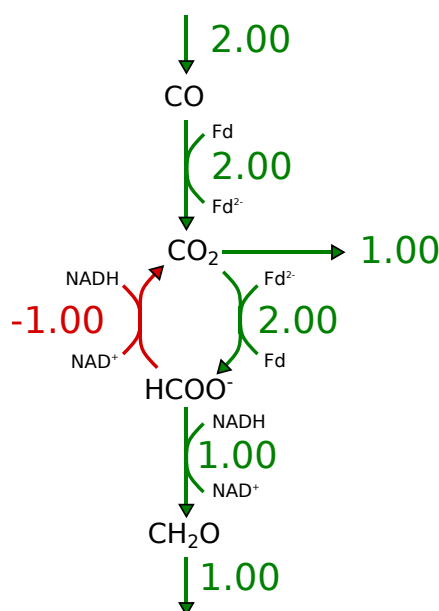


Figure 3.5: Elementary mode of the hypothetical pathway from carbon monoxide to formaldehyde.

The standard Gibbs energy of reaction ($\Delta_r G'^\circ$) for each of the internal reactions involved in the pathway are used to form the matrix of inequalities shown in Eq. 3.21. This matrix is the transpose of the stoichiometry matrix. Since error values for $\Delta_r G'^\circ$ are often significant (see Table 3.1), this uncertainty must be incorporated into the analysis.

Table 3.1: Gibbs energy of reaction for reactions in hypothetical formaldehyde pathway. Sourced from eQuilibrator [Flamholz *et al.*, 2012] (pH = 6, ionic strength = 0.1 M).

reaction	$\Delta_r G'^{\circ}$ (kJ/mol)	$\pm$
CODH	-16.3	14.3
Fdh _{NAD}	-8.2	13.3
Fdh _{Fd}	14	6.5
FNOR	31.6	3.9

For the elementary mode to be feasible, the following set of inequalities must be met:

$$\begin{aligned}
 \Delta_r G'_{\text{CODH}} + RT \ln \frac{[\text{CO}_2][\text{Fd}^{2-}]}{[\text{CO}][\text{Fd}]} &< 0 \\
 \Delta_r G'_{\text{Fdh}_{\text{NAD}}} + RT \ln \frac{[\text{NAD}^+][\text{HCOO}^-]}{[\text{CO}_2][\text{NADH}]} &< 0 \\
 \Delta_r G'_{\text{Fdh}_{\text{Fd}}} + RT \ln \frac{[\text{Fd}][\text{HCOO}^-]}{[\text{CO}_2][\text{Fd}^{2-}]} &< 0 \\
 \Delta_r G'_{\text{FNOR}} + RT \ln \frac{[\text{NAD}^+][\text{CH}_2\text{O}]}{[\text{NADH}][\text{HCOO}^-]} &< 0
 \end{aligned} \tag{3.19}$$

These inequalities can be linearised as follows,

$$\begin{aligned}
 \frac{\Delta_r G'_{\text{CODH}}}{RT} + \ln[\text{CO}_2] + \ln[\text{Fd}^{2-}] - \ln[\text{CO}] - \ln[\text{Fd}] &< 0 \\
 \frac{\Delta_r G'_{\text{Fdh}_{\text{NAD}}}}{RT} + \ln[\text{NAD}^+] + \ln[\text{HCOO}^-] - \ln[\text{CO}_2] - \ln[\text{NADH}] &< 0 \\
 \frac{\Delta_r G'_{\text{Fdh}_{\text{Fd}}}}{RT} + \ln[\text{Fd}] + \ln[\text{HCOO}^-] - \ln[\text{CO}_2] - \ln[\text{Fd}^{2-}] &< 0 \\
 \frac{\Delta_r G'_{\text{FNOR}}}{RT} + \ln[\text{NAD}^+] + \ln[\text{CH}_2\text{O}] - \ln[\text{NADH}] - \ln[\text{HCOO}^-] &< 0
 \end{aligned} \tag{3.20}$$

and represented in matrix form:

$$\begin{array}{l}
\text{CODH} \\
\text{Fdh}_{\text{NAD}} \\
\text{Fdh}_{\text{Fd}} \\
\text{FNOR}
\end{array}
\begin{bmatrix}
-1 & 1 & 0 & 0 & 0 & 0 & -1 & 1 \\
0 & -1 & 1 & 1 & -1 & 0 & 0 & 0 \\
0 & -1 & 1 & 0 & 0 & 0 & 1 & -1 \\
0 & 0 & -1 & 1 & 1 & -1 & 0 & 0
\end{bmatrix}
\cdot
\begin{bmatrix}
\ln[\text{CO}] \\
\ln[\text{CO}_2] \\
\ln[\text{HCOO}^-] \\
\ln[\text{CH}_2\text{O}] \\
\ln[\text{NAD}] \\
\ln[\text{NADH}] \\
\ln[\text{Fd}] \\
\ln[\text{Fd}^{2-}]
\end{bmatrix}
<
\begin{bmatrix}
\frac{-\Delta_r G'^{\circ}_{\text{CODH}}}{RT} \\
\frac{-\Delta_r G'^{\circ}_{\text{Fdh}_{\text{NAD}}}}{RT} \\
\frac{-\Delta_r G'^{\circ}_{\text{Fdh}_{\text{Fd}}}}{RT} \\
\frac{-\Delta_r G'^{\circ}_{\text{FNOR}}}{RT}
\end{bmatrix}.
\quad (3.21)$$

The equation above corresponds to the first constraint in Eq. 3.17. To include the error in $\Delta_r G'^{\circ}$ and upper/lower limits in metabolite concentrations, the following augmented matrix (Eq. 3.22) is constructed

$$\begin{bmatrix}
\mathbf{N}_{n \times m}^T & \mathbf{I}_{n \times n} \\
\mathbf{0}_{2n \times m} & \begin{bmatrix} \mathbf{I}_{n \times n} \\ -\mathbf{I}_{n \times n} \end{bmatrix} \\
\begin{bmatrix} \mathbf{I}_{m \times m} \\ -\mathbf{I}_{m \times m} \end{bmatrix} & \mathbf{0}_{2m \times n}
\end{bmatrix}
\cdot
\begin{bmatrix} \mathbf{c} \\ \mathbf{g} \end{bmatrix}
<
\begin{bmatrix} \mathbf{0} \\ \mathbf{u}_g \\ \mathbf{l}_g \\ \mathbf{u}_c \\ \mathbf{l}_c \end{bmatrix}
\quad (3.22)$$

where $\mathbf{c}$ is the vector of unknown log metabolite concentrations and $\mathbf{g}$ is a vector of unknown $\Delta_r G'^{\circ}$ values (divided by RT). The $\Delta_r G'^{\circ}$ values are unknown because their value is subject to error represented by the upper and lower bounds in $\mathbf{u}_g$ and $\mathbf{l}_g$, respectively. $\mathbf{u}_c$ and $\mathbf{l}_c$ are the upper and lower bound on log metabolite concentrations, chosen to ensure the solution is physiologically feasible. This equation (Eq. 3.22) defines the LP problem.

Since the 'less-than' inequality ($<$) cannot be implemented with standard LP software, Eq. 3.22 is modified so that the zeros in the first element of the right-hand-side vector are replaced by some arbitrary negative value (*e.g.* -1). This means $\leq$ can replace $<$ in Eq. 3.22 and avoids solutions supporting thermodynamic equilibrium (where no flux can occur).

In this example, a set of solutions are found which maximise and minimise the allowable concentrations of each metabolite. The range of allowable concentrations in this example is $0.001 \leq x \leq 10$ (mM), for metabolite x . The results for maximisation and minimisation given in Table 3.2 show both that the elementary mode is feasible and that there are certain restrictions on

metabolite concentrations necessary to achieve feasibility.

Table 3.2: Maximum and minimum allowable concentrations of metabolites supporting an elementary mode for formaldehyde production from CO.

metabolite	min (mM)	max (mM)
formaldehyde	0.001	0.5
formate	0.1	10
NAD	0.001	1.4
NADH	0.007	10
Fd	0.001	0.5
Fd ²⁻	0.02	10
CO ₂	0.001	10
CO	0.001	10

A ScrumPy-compatible implementation of this method is provided in the supplementary material ('metabolites.LP.py'). This implementation uses the LP solver provided in the SciPy package for Python [Jones *et al.*, 2014]. This implementation of the analysis, which is applied in Chapter 7, requires that constraints on metabolite concentrations be input as molar concentrations.

Chapter 4

MODEL CONSTRUCTION

4.1 INTRODUCTION

This chapter describes work carried out to construct a genome scale metabolic model of *C. autoethanogenum*. The model is needed to help improve an industrial bioprocess for carbon recycling, and must capture the fundamental process of acetogenic bacterial growth. This work was split into two broad sections: 1) creation of a draft model, and 2) model curation. In this work, ‘curation’ refers to the process of identifying and rectifying errors in the definition of the model’s metabolic network, such as missing pathways and mass-imbalanced reaction stoichiometries. The aim of curation is thus to obtain a mass-balanced stoichiometric model which returns thermodynamically feasible flux distributions.

The work described in this thesis is focused on one-carbon (C1) gas fermentation. This means particular attention was given to the enzymes of the Wood-Ljungdahl pathway [Ljungdahl and Wood, 1965; Ljungdahl and Wood, 1969], which enable autotrophy in many bacteria [Ragsdale and Pierce, 2008]. Similarly, special attention was given to identifying the enzymes responsible for transferring energy from C1 feedstocks to the ATP system, *i.e.* energy conservation – a process fundamental to life [Thauer *et al.*, 1977]. Knowledge of these enzymes is important because the exact mechanism of energy conservation varies greatly between anaerobic bacteria [Thauer *et al.*, 1977]. (A summary of published knowledge on the mechanisms for energy conservation in *C. autoethanogenum* is given in [Norman *et al.*, 2018].)

Another fundamental idea behind genome scale metabolic modelling is

that model-simulated fluxes represent, or are compatible with, cell viability and growth [Savinell and Palsson, 1992]. Thus, pathways towards all essential biomass precursors must be present in the network. This is an important aim of the work in this chapter.

Since at least some of the biomass precursors will require energy for their synthesis, some energy substrate will need to undergo exergonic conversion to byproducts to enable a positive ATP yield (Y_{ATP}) and meet the ATP requirement for growth. Biomass precursors for which this energy generation is necessary will be identified, as well as any cases where $Y_{\text{ATP}} \geq 0$. Biomass precursors whose production is accompanied by ATP generation may, in an industrial context, hold enhanced potential as target products. Indeed, many amino acids are produced for commercial purposes [Sahm *et al.*, 1995]. This is a good reason to investigate biomass-associated ATP yields.

The publication of a new genome sequence of *C. autoethanogenum* [Humphreys *et al.*, 2015] raised concerns regarding the older sequence described in [Brown *et al.*, 2014]. It was important to establish which of the reported discrepancies between different sequence versions might affect the metabolic network, and to make amendments where necessary. The possibility of errors in the Brown *et al.* sequence was not the only motivation for using the newly published genome annotation; the strain of *C. autoethanogenum* used to generate the new sequence was the exact strain available to the GASCHEM research group at the University of Nottingham (for which this work was conducted) [Humphreys *et al.*, 2015].

The process of model construction is described in detail in the methods section below. Following this, results from each stage of the curation process are given. Features of the curated model are discussed in the final part of this chapter, leading to an outline of the next steps.

4.2 METHODS

4.2.1 Draft metabolic network

The ScrumPy modules ‘PyoCyc’ and ‘BuildOrg’ [Poolman, 2006] were used to generate draft metabolic networks from the Pathway Genome Databases (PGDBs) described in Section 2.1.2. The initial draft networks built from the CAETHG and CLAU databases are referred to here as ‘AutoCauto’

and ‘AutoCLAU’, respectively.

4.2.2 Model curation

The first step in the curation of the draft metabolic network was to identify generic metabolites for which no chemical formula was defined *e.g.* ‘An alcohol’. Such metabolites were either replaced by specific instances with defined chemical formulae or removed from the network altogether (along with any reactions involving the subject metabolite).

Since *C. autoethanogenum* is a strict anaerobe [Abrini *et al.*, 1994], reactions involving oxygen were identified and inspected. Literature and sequence homology searches were used to decide which of the four reasons set out in Section 3.1 accounted for the reaction’s inclusion in the draft network (and whether the reaction should be retained).

ATP-generating futile cycles required identification and removal from the draft network since such flux distributions must necessarily contain a thermodynamic error [Fell *et al.*, 2010]. A futile cycle is a stoichiometrically-balanced flux distribution in which no metabolites are taken-up or excreted (*i.e.* where no net conversion takes place). Such flux distributions, forming cycles to fulfil the steady-state constraint, are impossible since cycles cannot maintain a negative energy gradient across every reaction involved. While some cycles will reside in the feasible region without interfering with optimal solutions, energy-generating cycles will appear in optimal solutions, thus undermining the actual feasibility of FBA results.

ATP-generating futile cycles were identified in the draft metabolic network using the LP problem described in Section 3.4.3. Since such cycles must necessarily contain a thermodynamic error, the directionality of one or more of the involved reactions was changed to disable the solution. In some cases, reactions involved in an ATP cycle carried flux contrary to the direction indicated by the standard Gibbs energy change. For example, reversible reactions carrying flux in the forward direction despite a high (positive) $\Delta_r G'^\circ$ value for the corresponding conversion may be constrained to carry flux only in the reverse direction on the basis that the necessary reaction quotient can only be achieved with physiologically unrealistic metabolite concentrations. The problem with this, as discussed in Chapter 3, is that a cut-off value for an acceptable $\Delta_r G'^\circ$ against which flux can be carried

Table 4.1: Gene associations for the WLP in *C. autoethanogenum*.

'branch'	enzyme	CLAU genes
carbonyl	CODH	CLAU_2924, CLAU_3807
	ACS/CODH	CLAU_1578-9
methyl	FdhA/HytA-E	CLAU_2712-21
	FtfL	CLAU_1576
	MTHFDH	CLAU_1574
	MetFV	CLAU_1572-3
	MT	CLAU_1567

must be chosen. For this reason, it is preferable to search for and remove reactions whose exclusion from the model can be justified by their lacking gene-reaction associations or containing unwanted metabolites.

In this case, such reactions would be constrained as irreversible to make ATP-cycling solutions infeasible. In other cases, reactions involved in ATP cycles were removed altogether.

A subnetwork of reactions forming the electron transport chain (ETC) was constructed manually based on the literature [Wang *et al.*, 2010; Wang *et al.*, 2013; Mock *et al.*, 2015]. Elementary Modes Analysis (EMA) was used to make sure that the ETC was capable of generating ATP *via* the membrane-bound Rnf and F_1F_0 ATP synthase complexes. The expected result was that at least one such mode would exist for each energy substrate included in the subnetwork.

Given that the WLP is understood to be the key mechanism of gas fermentation [Ragsdale and Pierce, 2008], it was deemed important to make sure that all genome-encoded enzymes of the WLP were represented in the metabolic network. After introducing the appropriate complement of WLP reactions to the network, FBA was conducted to ensure that C1 compounds (*i.e.* CO, CO₂) could be fixed to acetyl-CoA. A summary of genes encoding the enzymes of the WLP in *C. autoethanogenum* is provided in Table 4.1.

Biomass precursors

To ensure computed fluxes were compatible with cell viability, routes for the synthesis of biomass precursors had to be defined and curated. In GSMs, biomass is generally represented as a set of macromolecular biomass precursors – protein, DNA, RNA, lipid, cell wall material and other trace

metabolites. Each of these macromolecules are comprised of individual biomass precursor compounds, *e.g.* amino acids and nucleotides [Savinell and Palsson, 1992; Thiele and Palsson, 2010; Varma and Palsson, 1993]. An organism will typically synthesise these biomass precursors from sources of carbon, hydrogen, nitrogen, sulphur and phosphorous present in the environment. The liquid medium recipe for growth of *C. autoethanogenum* described in [Mihalcea *et al.*, 2008] provides a source of the elements essential for the production of biomass (including sulphur, nitrogen and phosphorous). Crucially, no amino acids are required in this minimal media recipe, meaning the cell must synthesise all 20 protein amino acids *de novo*. Recent research shows that pantothenate is an essential vitamin for growth of *C. autoethanogenum* along with thiamine and biotin, however the latter two are enzyme cofactors and do not participate in the metabolic network as metabolites in the conventional sense [Annan, 2018]. For this reason, uptake of biotin and thiamine is not considered in this work.

The biomass precursors that must be synthesised by the network are: all DNA and RNA nucleotides, all amino acids, the six lipid types identified in [Abrini *et al.*, 1994], NAD^+ , NADP^+ , FAD, FMN, THF, CoA, teichoic acid and peptidoglycan [Savinell and Palsson, 1992; Thiele and Palsson, 2010; Varma and Palsson, 1993]. Thus the biosynthesis of 42 precursors must be supported by the network.

To ensure biomass precursors can be made, an artificial transport reaction is added for each compound comprising these macromolecular classes to test if there is a steady state with production of the respective compound. The LP used to identify an optimal (flux minimal) steady state is described in Section 3.4.1. If no solution could be found, the literature was searched to try and identify valid bacterial biosynthetic pathways which could support production of the relevant precursor in *C. autoethanogenum*. Sequence homology methods (*i.e.* BLAST and PHYRE) can then be used to justify the addition of enzymes.

If the pathway towards the relevant biomass precursor is not resolved by a literature search, standard pathways as defined in MetaCyc were consulted to identify intermediate metabolite(s) participating in reaction(s) which produce the desired biomass component. Once identified, a transporter is added for the intermediate precursor(s) to the original biomass component before again attempting to compute a steady-state, this time adding a fixed-flux

constraint on the production of the intermediate. If again no solution is found, the process is repeated for the metabolite(s) participating one step further back from the desired biomass precursor, and so on until metabolites which the network can produce at steady-state are identified. Reactions between this metabolite (or set of metabolites) and the target biomass precursor are inspected for errors, including reversibility or the need for additional transporters enabling the excretion of by-products.

Anaerobes produce byproducts like acetate to yield energy for growth and cellular maintenance [Thauer *et al.*, 1977]. This is necessary when the conversion of growth substrate into biomass precursor is endergonic (*e.g.* asparagine biosynthesis, see Eq. 4.14). In other instances, efflux of a byproduct may be necessary simply to balance fluxes, the corresponding transporter acting as a sink, *e.g.* glycol production in THF biosynthesis, see Eq. 4.18. Glycol is a toxic compound which has been observed to cross the cellular membrane of *Clostridium glycolicum* [Gaston and Stadtman, 1963], with enzymes for its utilisation confined to protective microcompartments in *Acetobacter woodii* [Trifunović *et al.*, 2016].

In cases where efflux of a byproduct was necessary to produce a given biomass precursor, the corresponding solution was inspected to determine whether it was associated with energy production or was purely for flux-balancing purposes. Furthermore, for each biomass precursor, an attempt was made to partition the ATP generation involved in the solution into two categories:

1. ATP generated in the conversion of growth substrates into byproduct (*e.g.* CO to acetate), and
2. ATP generated in the pathway for precursor synthesis itself.

With this information, an ATP demand or yield could be associated with the net stoichiometry of each biomass precursor's synthesis. The ATP demand associated with the production of a particular biomass precursor was estimated in two ways:

1. an LP with transporters included for the uptake of ATP and water and the release of ADP, phosphate and protons is solved, minimising the uptake of energy source, and

2. the efflux rates of acetate and ethanol in the yield-maximal solution are multiplied by the maximum ATP yields respectively associated with acetate and ethanol. .

In order to check the conservation of mass across all network reactions. A balance-checking routine was written in Python (see python module ‘curation.py’ in the supplementary materials). This required the definition of manually defined chemical formulae for certain metabolites for which no chemical formula was given in the CLAU PGDB.

4.2.3 Database enhancement

This section describes the transition from a network with GPRs based on the older, Brown *et al.* genome sequence for *C. autoethanogenum* [Brown *et al.*, 2014] to the Humphreys *et al.* sequence [Humphreys *et al.*, 2015]. There were 243 discrepancies between the new and old sequence which appeared to relate to apocryphal stop-codons and frame-shift mutations in the older closed genome sequence. The erroneous stop codons frequently split single genes into two, so that when corrected, the Humphreys *et al.* sequence contained 73 fewer coding sequences [Humphreys *et al.*, 2015]. Furthermore, 142 errors were found in the annotation of protein coding sequences. These errors were identified by a thorough manual annotation procedure, which helped to avoid the transmission of errors typically associated with automated annotation approaches [Galperin and Koonin, 2010; Fell *et al.*, 2010; Thiele and Palsson, 2010]. It was important to establish where the reported discrepancies might affect the metabolic network and to make amendments where necessary.

The possibility of errors in the Brown sequence [Brown *et al.*, 2014] was not the only motivation for using the newly published genome annotation; the strain of *C. autoethanogenum* used to generate the new sequence was the exact strain available to the GASCHEM research group at the University of Nottingham (for which this work was conducted). Thus it was crucial that the metabolic network under construction was representative of this strain.

The PathoLogic algorithm (as implemented in Pathway Tools [Karp *et al.*, 2002]) was used to generate a PGDB from the *C. autoethanogenum* sequence [Humphreys *et al.*, 2015]. Network reactions unique to the CAETHG database were identified. Genes associated to these reactions (*i.e.* genes with CAE-

THG locus tags) were mapped to orthologs in the CLAU database (*i.e.* genes with CLAU locus tags) using a dictionary generated from bidirectional best blast hits analysis of the Bruno-Barcena *vs.* Humphreys (CAETHG *vs.* CLAU) genome annotations [Blom *et al.*, 2009]. The Python programs used to generate the locus-tag mapping are included in the supplementary materials (Python module: 'ClosDic').

4.3 RESULTS

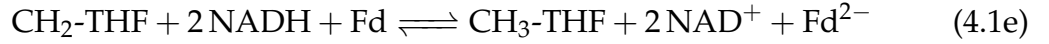
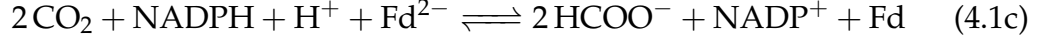
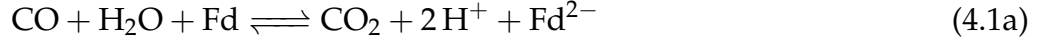
4.3.1 Draft metabolic network

AutoCauto consisted of 1451 reactions and 1584 metabolites, while AutoCLAU consisted of 1413 reactions and 1604 metabolites. The number of metabolites and reactions in the draft networks differ from their originating databases due to automatic curation functions executed by BuildOrg (see Section 2.2.1).

4.3.2 Model curation

A complete list of database reactions and metabolites excluded from the metabolic network is given in the file 'Unwanted.py' (see 'Model' directory in supplementary materials). Reactions requiring corrected stoichiometric weightings, direction change from irreversible to reversible or vice versa are respectively listed in supplementary files Corr_Stoich.spy, Corr_Revers.spy and Corr_Irrevs.spy.

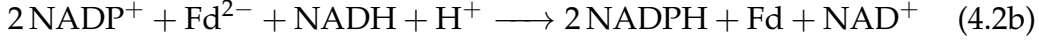
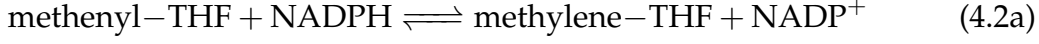
Initially, AutoCauto did not contain the complete WLP despite genomic and enzymological evidence for the presence of the complete pathway in *C. autoethanogenum* [Bruno-Barcena *et al.*, 2013; Brown *et al.*, 2014; Wang *et al.*, 2013; Mock *et al.*, 2015]. Carbon monoxide dehydrogenase (CODH, Eq. 4.1a), the acetyl-CoA synthase/CODH complex (ACS/CODH, Eq. 4.1b), formate dehydrogenase (FdhA, Eq. 4.1c), methyl transferase (MT, Eq. 4.1d) and methylene-THF reductase (MetFV, Eq. 4.1e) were all missing. The stoichiometries representing these enzyme-catalysed reactions, subsequently added to the network using the 'ExtraRxn.spy' file, are given in Eq. 4.1.



The set of reactions coupling ATP synthesis to the transport of electrons between electron carriers, *i.e.* the electron transport chain (ETC), was defined as a separate module (see supplementary file 'ETC.spy'). Of the five hydrogenases encoded in the genome of *C. autoethanogenum*, only HytA-E was retained in the ETC. This omission is justified by the high gene-expression levels of *hytA-E* relative to other hydrogenases during autotrophic growth of *C. autoethanogenum* reported in [Mock *et al.*, 2015]. The ETC accounts for the reduction of all three major electron carriers in the metabolism of *C. autoethanogenum*: Fd, NAD⁺ & NADP⁺. To achieve steady state, the metabolism must regenerate the oxidised form of these moieties. There are many potential sites for the regeneration of electron carriers, including oxidoreductase reactions leading to products such as ethanol, lactate or 2,3-butanediol. For example, one site for the oxidation of reduced ferredoxin is the pyruvate:ferredoxin oxidoreductase (PFOR) reaction enabling the production of pyruvate from acetyl-CoA [Mock *et al.*, 2015]. The reaction catalysed by this enzyme (encoded by CLAU_0896 or CLAU_2947) was introduced manually having been omitted by the PathoLogic algorithm.

Methylene-THF dehydrogenase was defined as NADPH-dependent, as reported in [Mock *et al.*, 2015] (see Eq. 4.2a). This presents a problem, however, since Rnf and CODH do not produce NADPH when CO is the carbon and energy source. The solution is that *C. autoethanogenum* encodes the enzyme Nfn [Wang *et al.*, 2010; Humphreys *et al.*, 2015]. Nfn couples the endergonic reduction of NADP⁺ with NADH to the exergonicity of the reduction

of NADP^+ with reduced ferredoxin [Wang *et al.*, 2010] (see Eq. 4.2b).



Another important consideration for the construction of the electron transport chain module is the efficiency of ATP synthesis *via* electron transport phosphorylation (ETP) [Thauer *et al.*, 1977]. ETP is a mechanism for ATP synthesis in which the transfer of electrons between electron carriers is coupled to the translocation of protons across the cellular membrane [Thauer *et al.*, 1977]. The subsequent proton motive force (PMF) can be exploited by the membrane-bound enzyme ATP synthase, which uses protons traveling back down the pH gradient into the cytoplasm to generate energy and phosphorylate ADP. The exact number of protons translocated per ATP synthesised ($\text{H}^+:\text{ATP}$) varies between organisms, and depends on the number of c-subunits in the ATP synthase rotor [Biegel *et al.*, 2011; Norman *et al.*, 2018]. The relationship between number of c-subunits and $\text{H}^+:\text{ATP}$ results from the hypothesis that every 120° turn of the ATP synthase rotor yields enough energy for the phosphorylation of 1 ADP. Thus, since each c-subunit is able to translocate only a single proton, $\text{H}^+:\text{ATP}$ is equal to the number of c-subunits divided by three. In *C. autoethanogenum*, the PMF is generated by a membrane-bound enzyme complex called Rnf which couples the reduction of NAD^+ by Fd^{2-} to the outwards translocation of two protons [Biegel *et al.*, 2009; Mock *et al.*, 2015]. The $\text{H}^+:\text{ATP}$ ratio is unknown for *C. autoethanogenum*, however a value of 3.66 has been determined for *Clostridium paradoxum* based on its ATP synthase c-rotor containing 11 subunits. This value was assumed for the study of *C. autoethanogenum* in previous work [Mock *et al.*, 2015] and is adopted here.

The full ETC network is shown below, with the corresponding stoichiometric matrix:

$$\mathbf{S}_{\text{ETC}} = [\mathbf{S}_{\text{ETC}}^{\text{enzymes}} \mathbf{S}_{\text{ETC}}^{\text{tx}}] \quad (4.3)$$

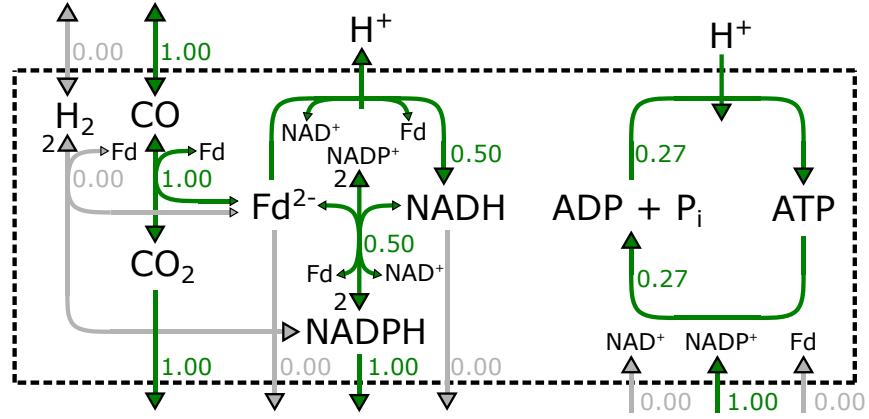


Figure 4.1: An elementary mode of the *C. autoethanogenum* electron transport chain sub-network. ATP is generated from CO but NADP^+ needs to be regenerated.

$$S_{\text{ETC}}^{\text{enzymes}} = \begin{array}{c} \text{CODH} \quad \text{Hyt} \quad \text{Nfn} \quad \text{Rnf} \quad \text{F}_1\text{F}_0 \quad \text{ATPase} \\ \begin{array}{l} \text{CO} \\ \text{H}_2 \\ \text{CO}_2 \\ \text{Fd}^{2-} \\ \text{Fd} \\ \text{NADH} \\ \text{NAD}^+ \\ \text{NADPH} \\ \text{NADP}^+ \\ \text{ATP} \\ \text{ADP} \\ \text{P}_i \\ \text{H}_2\text{O} \\ \text{H}^+ \\ \text{H}_{\text{EXT}}^+ \end{array} \end{array} \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 & 0 & 0 \\ -1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & -1 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ -1 & 0 & 0 & 0 & 1 & -1 \\ 2 & -3 & -1 & -3 & 2.67 & 1 \\ 0 & 0 & 0 & 2 & -3.67 & 0 \end{bmatrix} \quad (4.4)$$

The network is shown with the relative fluxes of one of its elementary modes represented by coloured arrows in Fig. 4.1. See supplementary materials for the ScrumPy-compatible model file ('ETC.spy'). As expected, this small network enables the production of ATP (and reduced electron carriers) from both CO and $[\text{CO}_2 + \text{H}_2]$, with the obvious requirement that the

Table 4.2: Summary of the *C. autoethanogenum* draft network with manually defined electron transport chain and curated WLP.

draft network	
Reactions	851
Transporters	16
Dead reactions	450
Metabolites	833
Orphan metabolites	320
Incons. enzyme subsets	21

oxidised form of the relevant electron carriers are regenerated elsewhere in metabolism. This sub-network will ensure that the model has at least one viable ATP-generating ‘engine’ during autotrophy. A summary of the *C. autoethanogenum* draft network features with manually defined electron transport chain and curated WLP is given in Table 4.2.

Among the oxygen-involving reactions in AutoCauto was an aspartate oxidase reaction (CLAU_0481, CLAU_0328, see Eq. 4.5a). The literature was searched (Google Scholar search string: ‘anaerobic aspartate oxidase’) for an anaerobic version of this enzyme to decide whether the reaction should be modified and retained or removed. The literature search revealed that under anaerobic conditions, the aspartate oxidase encoded in *Escherichia coli* is capable of using fumarate as an alternative electron acceptor under anoxic conditions [Gazzaniga *et al.*, 2009]. The best pBLAST hit of CLAU_0481 against the *E. coli* genome was aspartate oxidase (score: 830, e-val: 8.7e-106, alignment length 530, identities: 195), while the best hit of CLAU_0328 was fumarate reductase (score: 295, e-val: 1.0e-28, alignment length: 540, identities: 139), see supplementary materials for XML-formatted pBLAST result files ‘CLAU_0328_E.coli.xml’ and ‘CLAU_0481_E.coli.xml’. For both genes, all top-scoring PHYRE results for both CLAU_0481 and CLAU_0328 were succinate dehydrogenases or fumarate reductase, see supplementary materials ‘CLAU_0328.pdf’ and ‘CLAU_0481.pdf’. According to eQuilibrator, the $\Delta_r G'^{\circ}$ value for a fumarate-reducing aspartate oxidoreductase is close to equilibrium (≈ 0.6 at pH 6), and thus would be readily reversible (*i.e.* it could operate in the fumarate-reducing direction) with only small adjustments to metabolite concentrations [Flamholz *et al.*, 2012]. For these reasons, the oxygen-involving stoichiometry was replaced by a fumarate-reducing

stoichiometry, shown in Eq. 4.5b.



Biomass precursors

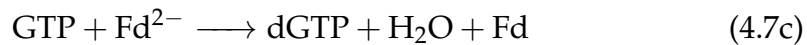
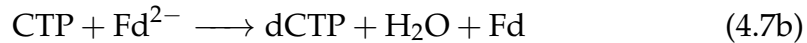
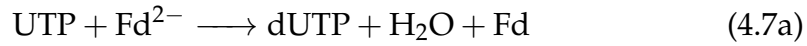
The draft network containing functioning ETC and full WLP was unable to synthesise 22 of 42 biomass precursors from either CO, [CO₂ + H₂] or fructose with unconstrained transporters for the following nutrients: hydrogen sulphide, ammonia and phosphate. All biomass precursors were present in the draft network except five of the six lipids detected in [Abrini *et al.*, 1994]: myristate, palmitoleate, petroselinate, stearate and methylenehexadecanoate. The text below is concerned with the curation of the network's biosynthesis pathways, focusing on biomass precursors for which manual curation was required. See Table 4.3 for a summary of network amendments. Complete flux-minimal solutions for all biomass precursors from all three main growth substrates (CO, [CO₂ + H₂] and fructose) are given in the supplementary materials.

DNA and RNA nucleotides. A change in the reversibility constraint on the draft network's diphosphatase reaction (encoded by CLAU_3055) was required to obtain feasible FBA solutions for the production of this class of biomass precursor. The diphosphatase in the draft network was irreversible in the PP_i-forming direction, and needed to be redefined as reversible. This is because PP_i is produced, but never consumed, in the pathway leading from ribose-5P to inosine monophosphate (IMP)—an intermediate common to each DNA and RNA nucleotide pathway. Thus, without the PP_i-consuming reaction of the diphosphatase, there is no way to balance PP_i production in the network. Furthermore, the PP_i-consuming reaction catalysed by diphosphatase has a $\Delta_r G'^{\circ}$ value of -15.8 [Flamholz *et al.*, 2012], supporting its favourability. The PP_i-yielding reaction, whose main function is to convert phosphoribose-diphosphate (PRPP) and glutamine to 5P-ribosylamine and glutamate, is catalysed by PRPP amidotransferase (encoded by CLAU_2869). The flux-minimal solution for all DNA and RNA nucleotides carries flux along the gluconeogenic route from pyruvate to fructose-1,6-diphosphate.

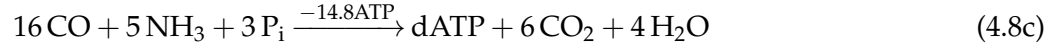
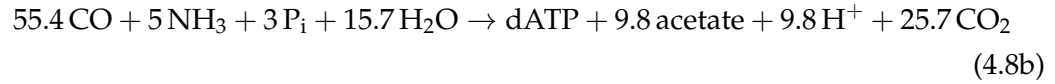
In addition to the diphosphatase, the transketolase reaction (Eq. 4.6) was redefined as reversible in accordance with [Sund *et al.*, 2015]. If left irreversible, the optimal solutions for DNA and RNA nucleotide synthesis were coupled with phenylalanine production, a behaviour for which no supporting literature evidence could be found.



In the case of DNA biosynthesis, electrons for the reduction of UTP to dUTP were provided by flavodoxin in the PGDB-defined reaction for ribonucleoside triphosphate reductase (CLAU.2231). However, since ferredoxin and flavodoxin are usually interchangeable in *Clostridia* (with the latter being used under conditions of iron limitation only [Buckel and Thauer, 2013]), this reaction was redefined as being ferredoxin-dependent (Eq. 4.7a). This correction enabled the reaction to carry flux, which subsequently allowed TTP production. The same correction was applied for the analogous cases of dCTP (Eq. 4.7b) and dGTP (Eq. 4.7c), for which these reduction reactions are the final step in their synthesis (unlike TTP).



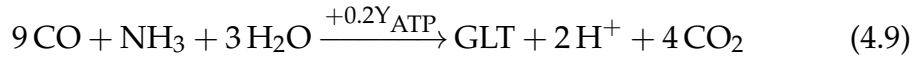
The ATP requirement associated with the dATP biosynthetic pathway was met with ethanol or acetate production (Eqs. 4.8a and 4.8b), depending on whether the $\min |\mathbf{v}_{\text{cat}}|$ or $\min v_{\text{CO}}$ objective function was used, respectively. (The solutions obtained with these objectives are respectively referred to as the ‘flux-minimal’ and ‘yield-maximal’ solutions). In fact, the flux distributions towards all DNA and RNA nucleotides required ATP generation and associated acetate/ethanol production.



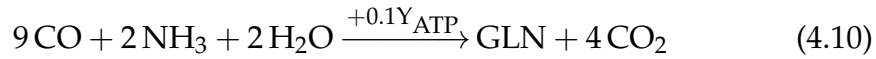
The draft network equipped with the WLP and ETC could not support production of six amino acids from either CO, CO₂ + H₂ or Fructose. After curation, synthesis routes for all amino acids were resolved for each of these major substrates. Details of amended precursor-specific pathways are given below.

Glutamate. A metabolic route for the uptake of glutamate (GLT, a known growth substrate for *C. autoethanogenum* [Abrini *et al.*, 1994]) was missing from the automatically generated network. More specifically, the omission of methylmalate dehydratase and citramalate lyase left gaps in the conventional pathway of glutamate degradation [Buckel and Barker, 1974]. A search of the manual genome annotation uncovered genes encoding citramalate lyase (CLAU_1014 or CLAU_2675) and isopropylmalate dehydratase (CLAU_2919 and CLAU_2920). A comparison of predicted protein structure performed with PHYRE [Kelley and Sternberg, 2009] revealed a high degree of homology (100% confidence, 90% coverage) between the 3-isopropylmalate large subunit (CLAU_2919) and methylmalate dehydratase in *Methanococcus jannaschii* [Bult *et al.*, 1996]. The subsequent addition of citramalate lyase and methylmalate dehydratase reactions completed this pathway for glutamate degradation, which is present in other *Clostridia* [Buckel and Barker, 1974].

This pathway for glutamate degradation [Buckel and Barker, 1974] was also utilised in reverse to enable glutamate production from CO, giving the net conversion (and positive ATP yield) shown in Eq. 4.9 as computed with FBA with the min $|\mathbf{v}_{\text{cat}}|$ objective function. This yield was possible because citramalate, an intermediate in the pathway between pyruvate and glutamate, was formed from pyruvate and acetate as catalysed by citramalate lyase (encoded by CLAU_2674). This enables substrate-level phosphorylation (SLP) in the formation of acetate from acetyl-phosphate as catalysed by Ack. Further ATP production contributing to the overall yield took place *via* Rnf/ATP synthase.



Glutamine. The flux-minimal solution for glutamine (GLN) production from CO with ammonium as nitrogen source follows the pathway for glutamate synthesis described above [Buckel and Barker, 1974] with glutamate being converted to glutamine by glutamine synthase (encoded by CLAU_1976). As with the production of glutamate, this solution enabled a positive ATP yield (see Eq. 4.10), however the yield is lower because glutamine synthase consumes ATP.

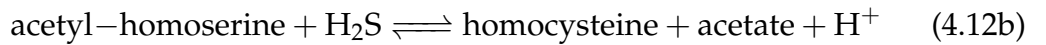


Aspartate. In the flux-minimal solution for aspartate (ASP) production, pyruvate was converted to oxalacetic acid by two routes simultaneously: 1) direct pyruvate carboxylation by pyruvate carboxylase (CLAU_1553) and 2) *via* phospho-enol-pyruvate (PEP) as catalysed by pyruvate kinase (PK, CLAU_2378) and PEP carboxykinase (CLAU_2645). Oxalacetate and glutamate are then converted to 2-ketoglutarate and aspartate by aspartate aminotransferase (CLAU_0207 or CLAU_0901 or CLAU_1310 or CLAU_2472 or CLAU_3333). In this solution, adenylate kinase catalyses the production of GTP and AMP from GDP and ADP. GTP is required in the reaction catalysed by PK, while AMP is rephosphorylated to ADP in a secondary reaction also catalysed by adenylate kinase [Berg and Joklik, 1954; Strominger *et al.*, 1959; Heppel *et al.*, 1959]. This intricate system for PEP production is avoided if PK is simply made ADP-dependent, however the inclusion of both ADP and GDP dependent versions of PK caused infeasible ATP cycles, thus the GDP version was kept and the ADP version removed. According to BRENDA [Schomburg *et al.*, 2002], there is only one known organism whose PK enzyme prefers GDP over ADP: *Toxoplasma gondii* [Saito *et al.*, 2008]. There are sixteen organisms listed whose PK shows activity with both GDP and ADP (and other nucleotides). This data does not provide a clear solution to the problem, thus the assumption of the GDP-dependent version is retained here.

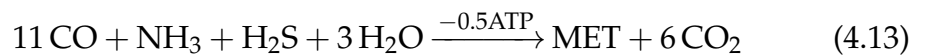


Methionine. A flux distribution for methionine biosynthesis from CO as energy source was feasible in the draft network with WLP and ETC, but proceeded *via* O-succinyl-homoserine-transferase (SST)—an enzyme associated with the production of methionine in *E. coli*. The absence of genes encoding SST indicated that this addition resulted from gap-filling steps applied in the Pathway Tools network construction algorithm.

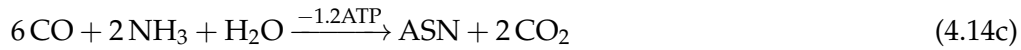
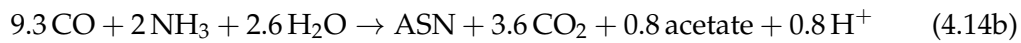
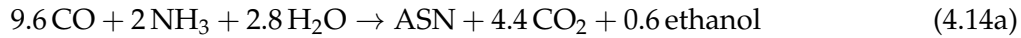
Ferla & Patrick *et al.* (2014) report an alternative route for methionine biosynthesis utilising O-acetyl-homoserine-transferase (AST), the amino acid sequence of which is distinguished from SST by the presence of a glutamate residue at position 111. This feature is shared in the corresponding CLAU_0472 sequence in the *C. autoethanogenum* genome (see Appendix A2), indicating that the first step in the pathway from homo-serine to methionine is in fact catalyzed by AST [Ferla and Patrick, 2014]. Methionine is synthesised directly in the reaction catalysed by acetyl homoserine aminocarboxypropyltransferase (AACPT) which takes 5-methyl-THF and homo-cysteine as substrate, the latter derived from homo-serine (as catalysed by AST) which is ultimately derived from glutamate *via* glutamate dehydrogenase. In summary, the additional reactions required for the production of methionine were AST (Eq. 4.12a, CLAU_0472) and AHS aminocarboxypropyltransferase (Eq. 4.12b, CLAU_2678).



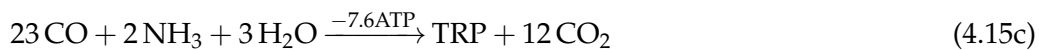
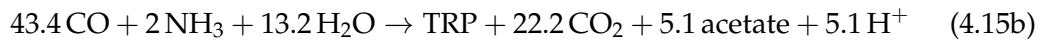
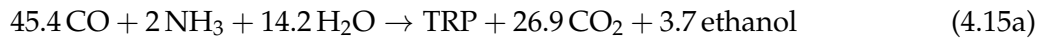
Methionine (MET) production was associated with acetate production in optimal solutions computed both with minimisation and max-yield objectives (Eq. 4.13). This acetate production results not from SLP in the Ack-catalysed reaction (which doesn't feature in the solution), but from the AACPT reaction (Eq. 4.12b) in which acetate is a byproduct which must be excreted to provide a flux-balanced solution. Furthermore, methionine production from CO, ammonia and H₂S yields ATP through the Rnf/ATP synthase couple (Eq. 4.13).



Asparagine. No feasible solution for asparagine (ASN) production was possible until diphosphatase was made reversible. After this amendment, the flux-minimal solution for asparagine production was associated with ethanol production (Eq. 4.14a). The maximal asparagine-yielding solution was associated with acetate production (Eq. 4.14b). The flux towards acetate/ethanol was required as a means of generating the 1.2 ATP required to produce 1 mole of asparagine from 6 moles of CO (Eq. 4.14c).

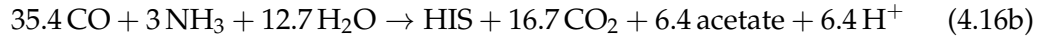
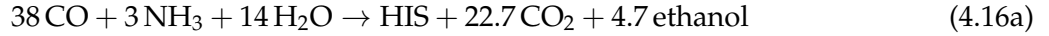


Tryptophan. The yield-maximal and flux-minimal FBA solutions for tryptophan (TRP) production were associated with acetate and ethanol production, respectively (Eqs. 4.15a and 4.15b) and required the reversible diphosphatase reaction. Both solutions use the tryptophan biosynthesis pathway described in [Crawford, 1975]. The minimal ATP cost of tryptophan production from CO is 7.6 ATP per tryptophan (Eq. 4.15c).



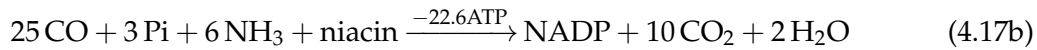
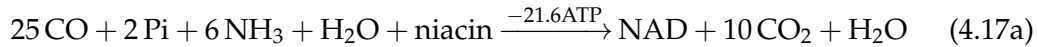
Histidine. Flux distributions producing histidine (HIS) from CO were feasible only with the reversible diphosphatase reaction. Once again, the formation of acetate or ethanol as byproduct was sensitive to the objective function (Eqs. 4.16a and 4.16b). ATP requirement of 9.5 mol per mol histidine was high relative to other amino acids (Eq. 4.16c). The enzymatic pathway catalysing the (net) conversion of PRPP to histidine is encoded in the block of genes spanning CLAU_3174 to CLAU_3189. This genetic structure

is similar in other bacteria [Martin *et al.*, 1971].

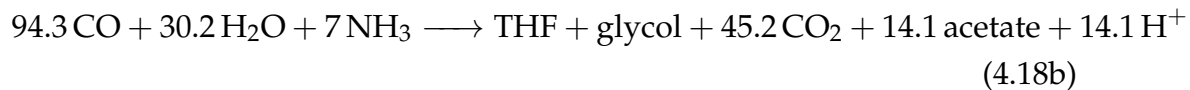
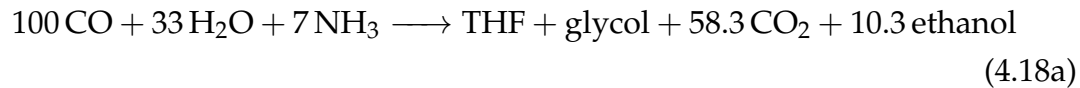


NAD^+ & NADP^+ . Flux distributions supporting NAD and NADP biosynthesis *de novo* (*i.e.* from CO and water) are only feasible when the diphosphatase reaction is defined as reversible. The pathways involved in the solutions are the anaerobic NAD and NADP biosynthesis pathways described in [Foster and Moat, 1980]. The fumarate-reducing aspartate oxidase described above (encoded by CLAU_0481 or CLAU_0328) carries flux in the NAD(P)^+ -synthesising solutions, consistent with its place in the pathways described in [Foster and Moat, 1980], and linked to the concomitant succinate production. Importantly, this result is inconsistent with the minimal-media formula described in Appendix A1, which includes niacin, a precursor required for NAD(P) biosynthesis in organisms which cannot produce it *de novo*. Furthermore, all previous studies which involve the growth of *C. autoethanogenum* include niacin or another NAD precursor (*e.g.* nicotinic acid) as a growth medium component [Abubackar *et al.*, 2016; Abubackar *et al.*, 2012; Abubackar *et al.*, 2015]. However, recent research into the essentiality of vitamins for *C. autoethanogenum* [Annan, 2018] demonstrates experimentally that niacin is non-essential for growth, thus supporting the network's validity.

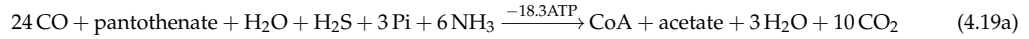
There is a different optimal solution for NAD(P)^+ biosynthesis when uptake of the vitamin niacin was enabled. Niacin was included in the minimal media recipe (Appendix A1) in experiments referenced later in this study. Thus, to allow comparison of model output and experimental data, the niacin transporter was retained in the model despite the feasibility of NAD(P)^+ *de novo* biosynthesis. The optimal solution for NAD(P)^+ production included both acetate and ethanol production with the min $|\mathbf{v}_{\text{cat}}|$ objective, consisting of 66 reactions. When the v_{CO} objective was used, acetate but no ethanol was produced in a solution consisting of 64 reactions. In both cases, the ethanol and acetate production was associated with the ATP generation required for NAD(P)^+ biosynthesis (Eq. 4.17).



Tetrahydrofolate. The optimal flux distribution for THF synthesis followed the pathway described in [Rossi *et al.*, 2011]. Biosynthesis of THF, which was defined as the mono-glutamate form (ch. formula: $\text{C}_{19}\text{H}_{21}\text{N}_7\text{O}_6$), required the introduction of a glycol transporter to allow feasible solutions in the GSM. The reversible diphosphatase reaction was also required. Glycol byproduct was required to allow flux through the reaction catalysed by an NADP dependent alcohol dehydrogenase, which converts glycolaldehyde and NADPH to glycol and NADP, of which there are several candidates encoded in the genome. This reaction thus serves as a sink for glycol, which is otherwise an orphan metabolite.

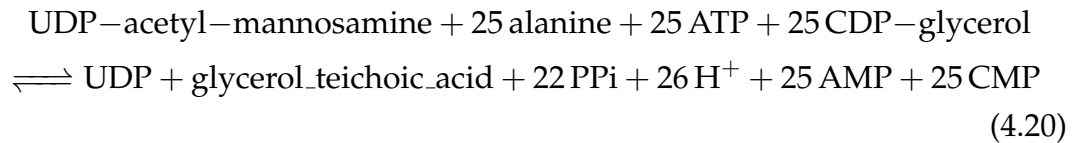


CoA. Like THF, co-enzyme A (CoA) is involved in the WLP pathway, so proper curation of the pathway responsible for its biosynthesis was important. With ch. formula $\text{C}_{21}\text{H}_{36}\text{N}_7\text{O}_{16}\text{P}_3\text{S}$, CoA incorporates atoms from each available (non-metallic) elemental source provided in the growth media. However, recent research [Annan, 2018] shows that *C. autoethanogenum* is auxotrophic with respect to pantothenate, an immediate precursor to CoA. More specifically, the genes *panBCD*, respectively encoding methyloxobutanate hydroxymethyltransferase, pantothenate synthase and aspartate decarboxylase, are not present in the *C. autoethanogenum* genome [Annan, 2018]. Thus to achieve CoA production, pantothenate was added to the list of media components (and is in any case listed as a component of the minimal media given in Appendix A1).



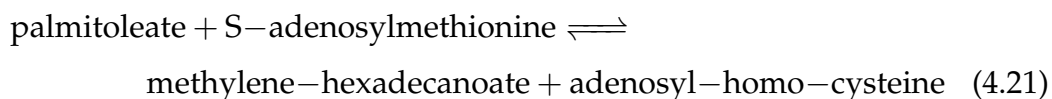
FAD & FMN. These compounds are important cofactors, essential for the function of many enzymes [Berg *et al.*, 2002]; their inclusion as biomass precursors represents this requirement. Neither could be synthesised without the reversible diphosphatase. In optimal solutions for both compounds, the PP_i was produced in PRPP amidotransferase, GTP cyclohydrolase and GMP synthase (as specified in the MetaCyc database [Caspi *et al.*, 2008]). Additional PP_i is produced in the conversion of FMN to FAD in the optimal solution for latter's biosynthesis from CO.

Peptidoglycan & teichoic acid. Again, PP_i hydrolysis necessary to achieve feasible solutions for the biosynthesis of peptidoglycan from CO. In the case of peptidoglycan, the MetaCyc-defined pathway for the polymerisation of isopentenyl diphosphate releases PP_i at every polymerisation step, necessitating recycling of PP_i . Teichoic acid production was included as a lumped-reaction based on the MetaCyc pathway for teichoic-acid production (MetaCyc ID: 'TEICHOICACID-PWY') [Caspi *et al.*, 2008]. This lumped reaction was given the ID 'glycerol_teichoic_acid_synth', producing a large teichoic-acid polymer with the ch. formula $\text{C}_{158}\text{H}_{313}\text{O}_{155}\text{P}_{25}\text{N}_{26}$ (model ID 'glycerol_teichoic_acid').



Lipids. Abrini *et al.* detected six different lipids in *C. autoethanogenum*: myristate, palmitate, palmitoleate, stearate, petroselinate and hexadecanoate. The remaining fraction was uncharacterised. For all lipids except hexadecanoate, lumped reactions were used to represent biosynthesis (see Appendix A3). Since the lumped reactions consume carbonate, the introduction of carbonic anhydrase (catalysing hydration of CO_2 to carbonate) was necessary for lipid biosynthesis. In *C. autoethanogenum*, carbonic anhydrase is encoded by CLAU_2699 or CLAU_3021 [Pander, 2018]. Hexadecanoate is a

cyclopropane side-chain fatty acid not synthesised in the same way as the fatty acids above. It is produced from palmitoleate and S-adenosyl methionine according to the following reaction [Kanemasa *et al.*, 1991; Bath and Tortorello, 2014]:



4.3.3 Database enhancement

After curation of the model was complete, a genome sequence and manual annotation of the GASCHEM strain of *C. autoethanogenum* became available [Humphreys *et al.*, 2015]. The annotation of this sequence uses the locus tag prefix 'CLAU_'. A pathway genome database (PGDB), referred to here as 'the CLAU database', was generated from this annotation using PathoLogic [Karp *et al.*, 2002].

In order to improve the model's representation of the GASCHEM *C. autoethanogenum* strain, reactions whose GPR referred only to CAETHG locus tags were linked to the GASCHEM annotation by mapping the CAETHG genes to CLAU genes using a dictionary of orthologs generated with best-BLAST hits as described in [Blom *et al.*, 2009]. The Python-readable dictionary is provided in the supplementary materials ('ClosDic.py'). 74 model reactions unique to the CAETHG PGDB were identified, 47 of which were successfully mapped to CLAU gene orthologs. 16 of the remaining model reactions unique to the CAETHG database were defined as spontaneous, and thus required no association to genes. Gene associations were defined manually for 10 of the remaining reactions, while 1 reaction had no gene-association defined in the CAETHG database. The composition of the model presented in this study before and after the mapping of CAETHG genes to CLAU genes is summarised in Fig. 4.2. This figure shows two key pieces of information: 1) before the necessary curation was carried-out, a significant proportion of model reactions were associated only to CAETHG locus tags (*i.e.* genes from the older genome annotation), and 2) after curation, all model reactions could be mapped to CLAU locus tags except transporters

Table 4.3: Summary of network modifications required to produce full complement of biomass precursors in the GSM of *C. autoethanogenum* for growth on CO and minimal media. ‘Gap’, ‘direction’ and ‘stoich.’ columns show reactions respectively requiring introduction into the network, change of direction, and change of stoichiometry.

precursor	additional transporters	gap	direction	stoich.
ATP			transketolase	
UTP			diphosphatase	
			transketolase	
GTP			diphosphatase	
			transketolase	
CTP			diphosphatase	
			transketolase	
dATP			diphosphatase	ribonucleoside-3P reduct.
			transketolase	
TTP			diphosphatase	
			transketolase	
dGTP			diphosphatase	ribonucleoside-3P reduct.
			transketolase	
dCTP			diphosphatase	ribonucleoside-3P reduct.
			transketolase	
glutamate		citramalate lyase,	diphosphatase	
		methylmalate dehydratase	diphosphatase	
methionine		homoacetyl transferase,	diphosphatase	
		AHS aminocarboxypropyltrans.	diphosphatase	
asparagine			diphosphatase	
tryptophan			diphosphatase	
histidine			diphosphatase	
NAD			diphosphatase	asp. oxidase
NADP			diphosphatase	asp. oxidase
THF			diphosphatase	
CoA	, glycol		diphosphatase	
FAD	, pantothenate		diphosphatase	
FMN			diphosphatase	
peptidoglycan			diphosphatase	
teichoic acid		teichoic acid synthase	diphosphatase	
lipids		carbonic anhydrase	diphosphatase	

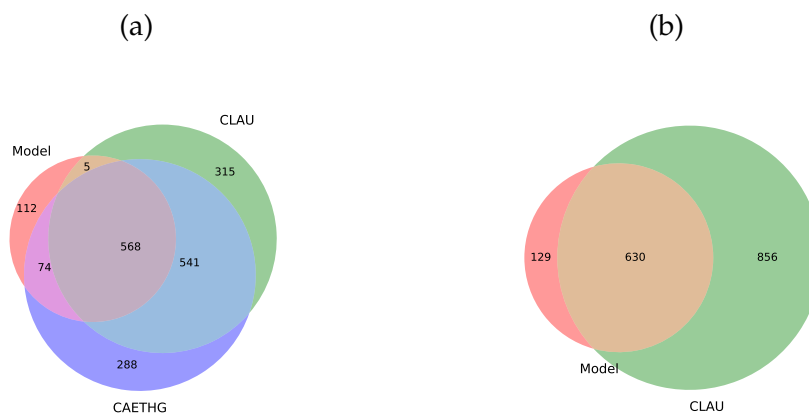


Figure 4.2: Model composition: CLAU *vs.* CAETHG databases. Venn diagrams show intersection of model reactions with database reaction sets before (a) and after (b) the mapping of CAETHG database genes to CLAU database orthologs. (a) Red circle area = model reactions, green circle area = CLAU database reactions, purple circle area = CAETHG database reactions. (b) Red circle area = model reactions, green circle area = CLAU database reactions, including 57 reactions whose gene associations were successfully mapped from CAETHG to CLAU genes (47 by identification of orthologs, 10 by inspection of manual genome annotation/literature searches). CLAU database reactions not included in the model were deemed relevant/helpful for metabolic modelling purposes.

whose inclusion is based on empirical evidence and hypothetical essential reactions. The final, curated model is called ‘MetaCLAU’.

The full set of GPRs are given in the supplementary materials as Python-readable document (‘manual.GPRs.py’) commented with justifications for the assignment of hypothetical AND/OR relationships where experimental evidence could not be found in the literature.

4.3.4 Comparison with published model

In 2015, a genome-scale metabolic model of *C. autoethanogenum* was published in [Marcellin *et al.*, 2016] and later updated in [Valgepea *et al.*, 2017b; Valgepea *et al.*, 2017a]. This section gives a general comparison of the model constructed in this chapter with the published model [Valgepea *et al.*, 2017b; Valgepea *et al.*, 2017a]. While both models satisfy the condition that no ATP is generated without the uptake and excretion of metabolites (see Section 3.4.3), differences in general statistics are given in Table 4.4.

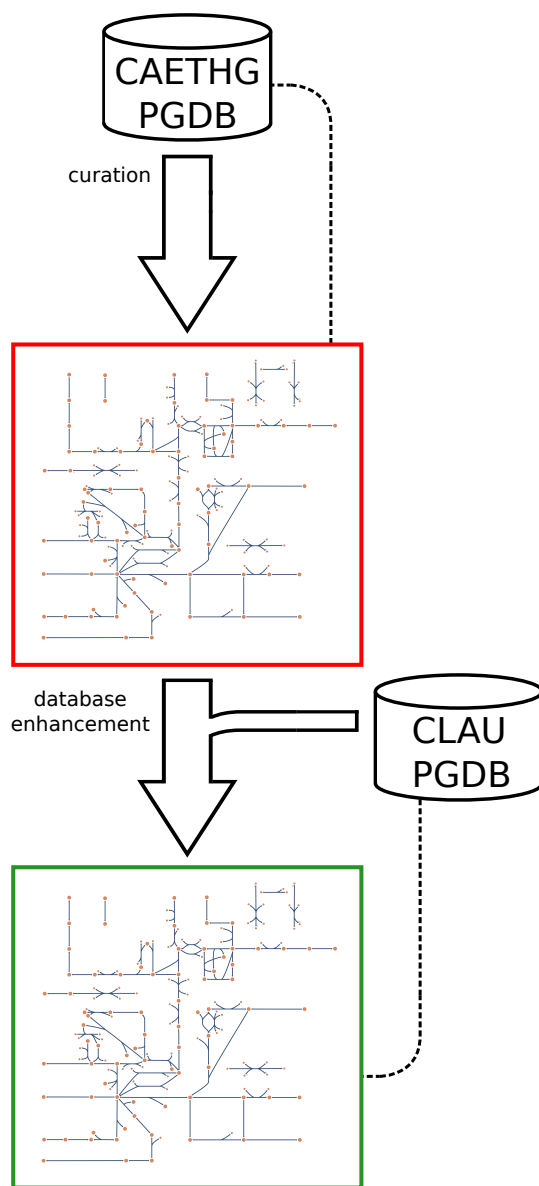


Figure 4.3: Model construction workflow. Arrows represent major construction steps (curation and database enhancement). Red- and green-framed network pictures respectively represent the model before and after database enhancement. Dashed lines show which PGDB is referred to by each model-version's GPRs.

Table 4.4: Comparison of *C. autoethanogenum* genome-scale metabolic models.

model	reactions	transporters	metabolites	genes	GPRs	IESs
MetaCLAU	758	79	772	699	600	2
iCLAU786	1001	30	1046	802	844	21

IESs = inconsistent enzyme subsets

While the MetaCLAU network is smaller in terms of number of reactions, metabolites and genes, it accounts for the transport of more metabolites than iCLAU786 and contains fewer inconsistent enzymes subsets. Furthermore, iCLAU786 does not contain the fumarate-reducing aspartate oxidase described in the results above (see Eq. 4.5b). The model instead defines an NAD-dependent aspartate dehydrogenase (with the reaction ID 'rxn05119_c0'), and thus does not capture the hypothetical relationship between succinate production and NAD(P)⁺ production included in MetaCLAU.

Model files

The directory of files required to load MetaCLAU as a ScrumPy model are provided in the supplementary materials (directory: 'Model'). A COBRA compatible JSON version is also given ('meta_clau'), along with a copy of the iCLAU786 SBML model ('iCLAU786.xml').

4.4 DISCUSSION

In this chapter, a GSM of *C. autoethanogenum* (MetaCLAU) has been constructed. This was done to provide support for the development of a fermentative bioprocess for carbon recycling. The model captures the fundamental concepts which govern bacterial growth: generation of ATP from an energy source and the production of biomass components. Certain aspects of the resulting metabolic network are discussed below.

One result which actually supports the validity of the model is that succinate is produced in solutions for the biosynthesis of NAD(P)⁺ without the inclusion of niacin in the media. This is consistent with, and may be an explanation for, the observation of succinate in cell supernatant in cultures of *C. autoethanogenum* [Breitkopf, 2018; Köpke *et al.*, 2013].

That the pyruvate kinase (PK) encoded in the *C. autoethanogenum* genome

(CLAU_2378) is GDP rather than ADP dependent in *C. autoethanogenum* is an unproven hypothesis. This stoichiometry was assigned automatically by the PathoLogic algorithm. The two PK enzymes in *Escherichia coli* appear to favour GDP as phosphate-accepting substrate but can also use ADP [Waygood and Sanwal, 1974; Waygood *et al.*, 1975], while in *Clostridium thermocellum*, several enzymes in the glycolytic pathway are GDP dependent, including glucokinase and PEP carboxykinase (though this organism does not encode PK) [Zhou *et al.*, 2013].

An important result is that the production of different biomass precursors is associated with different products. For example, methionine production is associated with acetate regardless of objective, and NAD(P)⁺ production is coupled to both acetate and ethanol production. Since flux distributions for individual biomass precursors were associated with different products, it follows that the ratios of precursors in *C. autoethanogenum* biomass could affect the product profile. The conclusion from this analysis is that it will be important to acquire an experimentally measured biomass composition of *C. autoethanogenum*. This is followed-up in the next chapter.

Chapter 5

MODEL PARAMETERISATION AND VALIDATION

5.1 INTRODUCTION

In order to scale computed fluxes towards biologically relevant values, species-specific parameters were derived experimentally by GASCHEM researchers at the University of Nottingham. This was intended to improve the accuracy of model predictions [Toro *et al.*, 2018]. In the case of genome scale modelling of bacterial cells, two key parameters are required: biomass composition and ATP maintenance costs [Fell *et al.*, 2010].

In the previous chapter, pathways leading to essential biomass precursors were established in the genome scale model (GSM). What remains unknown is the distribution of flux towards each of these precursors in a viable *C. autoethanogenum* cell. Incorporation of this knowledge into model analyses will ensure that results are given against a realistic background of biomass production [Feist and Palsson, 2010; Pramanik and Keasling, 1997]. Despite the large number of biotechnologically interesting *Clostridia*, detailed physiological information suitable for deriving cellular biomass composition is available for a small range of species only [Lee *et al.*, 2008; Senger and Papoutsakis, 2008a; Dash *et al.*, 2016]. Thus the first aim of the work described in this chapter is to parameterise the GSM with an experimentally derived, *C. autoethanogenum*-specific macromolecular biomass composition.

The second important set of parameters for genome scale modelling is the ATP required for growth and non-growth associated maintenance pro-

cesses, GAM and NGAM, respectively. These parameters vary greatly from organism to organism [Stephanopoulos *et al.*, 1998], and affect the demand on energy substrate(s). In light of this variability and potential impact to model behaviour, it was important not to source GAM and NGAM values from other GSM studies. Instead, the aim of the work here described was to evaluate the importance of experimentally derived values.

This chapter also describes efforts to validate the genome scale metabolic model (GSM) of *C. autoethanogenum* (MetaCLAU). Model validation is the set of tests and analyses used to ensure that a mathematical model can predict known behaviour before it is applied for the prediction of unknown behaviour [Torres and Eberhard, 2002]. Without this initial test of validity, there are no grounds for claiming that a model is capable of describing real-world phenomena. Thus, in this chapter, published data has been used to evaluate the validity of the capabilities of MetaCLAU.

A good first test of the validity of a GSM is the simulation of growth on a range of potential substrates [Nagarajan *et al.*, 2013; Toro *et al.*, 2018]. Published research on *C. autoethanogenum* has involved testing the organism's growth on a variety of compounds experimentally, thus there is a substantial base of knowledge against which to test the predictive capability of MetaCLAU in terms of substrate utilisation. Furthermore, an understanding of the relative metabolic advantages of various growth substrates can help to direct the composition of feedstocks for industrial bioprocess design.

The model's ability to capture changes in product profile in response to environmental and genetic perturbation is of the utmost importance to this project. As discussed in the general introduction, the native aqueous product profile of *C. autoethanogenum* (when grown at its pH optimum for growth) consists mainly of acetate and ethanol with trace amounts of lactate and 2,3-butanediol (2,3BD) [Abrini *et al.*, 1994; Köpke *et al.*, 2011c]. The model will be valid only if it is able to produce these compounds while supporting biomass production, a requirement tested in this chapter.

Following this, a comparison of the growth rate of continuous cultures grown on CO was made to model-predicted growth yields. Culture data were generated by Anne Henstra at the University of Nottingham.

Finally, an aspect of validation which pertains specifically to FBA is the choice of objective function. Finding an objective function which is both

biologically relevant and convenient to use is challenging – as evidenced by the large amount of literature on the subject [Feist and Palsson, 2010; Schuetz *et al.*, 2007; Sendin *et al.*, 2010]. For this reason, an evaluation of the validity of various objectives was carried out.

5.2 METHODS

5.2.1 Biomass

The macromolecular biomass composition was split into the following quantifiable categories: protein, DNA, RNA, lipid, polysaccharide and trace metabolites, based on genome scale modelling standards [Thiele and Palsson, 2010] and known composition of gram positive cells [Archibald *et al.*, 1968a]. Experiments for the determination of macromolecular biomass composition were conducted by Florence Annan, Ronja Brietkopf, Bart Pander, Pawel Piatek and Laudina Safo at the University of Nottingham. Details of methods are given in Appendix A3.

The relative abundance of major lipids in *C. autoethanogenum* are taken from [Abrini *et al.*, 1994]. Abrini *et al.* quantified the relative proportion of six different lipids: myristate (14:0) = 5.5%, palmitate (16:0) = 50.6 %, palmitoleate (16:1) = 3.4 %, stearate (18:0) = 3.3 %, petroselinate (18:1) = 1.1 % and hexadecanoate (17 cycl.) = 5.3 %.

DNA nucleotide ratios were estimated from the full genome sequence [Humphreys *et al.*, 2015], whereas RNA nucleotide and amino acid ratios were estimated from transcribed rRNA and translated ribonucleoprotein sequences, respectively (see supplementary materials, ‘ribonucleoprotein_genes.csv’). The use of ribosomal RNA and ribonucleoprotein sequences to characterise the prevalence of RNA and amino acids is based on the observation that the largest allocation of protein synthesis capacity is in the production of ribosomes [Li *et al.*, 2014]. Genes encoding ribonucleoproteins were identified from the manual genome annotation [Humphreys *et al.*, 2015] by searching for the 21 proteins (S1–S21) and 34 proteins (L1–L34) comprising the 30S and 50S ribosomal subunits, respectively [Berg *et al.*, 2002]. The final biomass composition, represented as a ScrumPy-format lumped reaction (with growth-associated ATP maintenance costs included), is given below:

BIOMASS:

111 WATER + 0.024 CYS + 0.097 L-ASPARTATE + 0.16 ILE + 0.13 SER + 0.045 GLN + 0.22 LYS + 0.071 PRO + 0.1 THR + 0.043 PHE + 0.19 VAL + 0.26 GLY + 0.026 HIS + 0.15 LEU + 0.088 ARG + 0.0064 TRP + 0.2 L-ALPHA-ALANINE + 0.13 GLT + 0.099 ASN + 0.047 TYR + 0.042 MET + 0.1 TTP + 0.05 DGTP + 0.1 DATP + 0.05 DCTP + 0.1 GTP + 0.07 UTP + 0.08 CTP + 111.2 ATP + 0.06 METHYLENE HEXADECANOATE + 0.07 MYRISTATE + 0.6 PALMITATE + 0.01 PETROSELINATE + 0.04 STEARATE + 0.04 PALMITOLEATE + 0.05 NAD + 0.04 THF-GLU-N + 0.001 FMN + 0.01 CO-A + 0.05 NADP + 0.00086 FAD + 1e-06 glycerol.teichoic.acid + 3.5e-05 CPD-12231 $\longrightarrow$ 111 ADP + 111 Pi + 111 PROTON + x.BIOMASS $\sim$

5.2.2 ATP maintenance

Experiments for the determination of ATP maintenance costs were conducted by Anne Henstra and Louise Sewell at the University of Nottingham. Details of methods are given in Appendix A4.

Growth-associated and non-growth-associated rates of CO consumption were respectively converted to GAM and NGAM by multiplication with the Y_{ATP}^{max} value estimated from FBA as described in Section 3.4.2 and [Toro *et al.*, 2018]. This estimation is based on the organism's ability to use CO as sole source of energy. It follows from this that if the yield of ATP from CO is known (or, as in this case, assumed to be Y_{ATP}^{max}), the CO uptake rate can be converted into ATP requirement.

A proportion of GAM is accounted for by biomass precursor production in the metabolic network. This proportion must be subtracted from the total GAM before incorporation as a flux constraint. (Units for growth associated and non-growth associated maintenance are mmol gDCW⁻¹ and mmol gDCW⁻¹ h⁻¹, respectively.)

5.2.3 Carbon sources

To test growth on a range of substrates, a transporter for the substrate in question was added to the model and the model was constrained so that uptake of the substrate in question was the only source of energy/carbon. Transporters added to test the known inability for utilisation of a substrate (*e.g.* glucose) were not retained (unless the substrate was also a known prod-

uct). FBA was then used (as described in Section 3.4.1) to test if production of biomass and ATP for maintenance was feasible. Each successfully predicted growth substrate supports the model’s validity [Toro *et al.*, 2018].

5.2.4 Product spectrum

A complete hierarchy of products ordered in terms of Y_{ATP} was generated by iteratively computing the Y_{ATP} -optimal solution while blocking (*i.e.* constraining to 0) flux across the transporter of the product formed in the previous iteration until ATP production was infeasible. The FBA applied at each iteration is described in Section 3.4.2. This analysis can be run using the `ipython-notebook` and `COBRApy`-compatible JSON model file given in the supplementary materials (`‘model_analysis.ipynb’` and `‘metaclau.json’`).

After computing the range of products associated with ATP generation from CO, the linear program proposed in [Peres *et al.*, 2018] and given in Section 3.5 will be applied to 1) ensure flux distributions are thermodynamically feasible, and 2) assess the maximum and minimum allowable concentrations for each individual metabolite.

5.2.5 Growth rate

The maximal yield of biomass precursors from carbon monoxide was computed using FBA as described in Section 3.4.2. The growth rate calculated by maximising biomass yield cannot be exceeded experimentally without an increase in substrate uptake; it is the rate that would be achieved if the maximal biomass-yielding flux distribution is adopted by the microbe at some defined substrate uptake rate. When interpreting fluxes and growth rates computed through the maximisation of biomass yield, it is important to note that a high biomass-yielding microbe could grow slower than a low biomass-yielding microbe if the substrate uptake rate is sufficiently high in the latter case.

5.2.6 Objective functions

FBA was applied with the following three objective functions: minimisation of total network flux, minimisation of total network *enzyme catalysed* flux, and growth yield maximisation. For the minimisation objectives, product spectrum, substrate uptake, and change in objective value (z) were

Table 5.1: Overall biomass composition of *C. autoethanogenum* DSM 10061

component	%	$\pm$
protein	26.3	2.3
DNA	14.6	7.5
RNA	17.9	4.2
lipid	21.0	1.7
polysaccharide	10.2	7.8
teichoic acid	0.6	0.1
other	9.4	-

recorded. For growth maximisation, the product profile and objective value (*i.e.* the growth yield (μ)) were recorded.

5.3 RESULTS

5.3.1 Biomass

The relative proportions of protein, DNA, RNA, lipid, teichoic acid, polysaccharide and other (*i.e.* essential trace metabolite pools) were measured for *C. autoethanogenum* (Table 5.1).

Tables showing the estimated relative proportions of individual nucleotides and amino acids are given in Appendix A3 along with metabolite concentrations comprising the pool of trace metabolites.

5.3.2 ATP maintenance

A GAM value of 155.7 ± 34.6 and an NGAM value of 2.2 ± 0.5 resulted from multiplication of growth and non-growth associated CO uptake rates (415 ± 92 and 5.7 ± 1.4 , respectively) by the computed $Y_{\text{ATP}}^{\text{max}}$ value of 0.375.

After subtracting the proportion of growth associated ATP maintenance incorporated in the model's metabolic network ($44.5 \text{ mmol gDCW}^{-1}$) from the experimentally derived GAM parameter, the actual GAM value used as a constraint in FBA simulations was $111.2 \text{ mmol gDCW}^{-1} \text{ h}^{-1}$. Total ATP maintenance (r_{ATP}) resulting from GAM/NGAM parameters sourced from published GSMs and an assumed growth rate of 0.028 are shown in Table 5.2.

Table 5.2: Maintenance costs (r_{ATP}) calculated at $\mu = 0.028 \text{ h}^{-1}$ with GAM and NGAM values from a range of GSMs/organisms. Units for growth associated and non-growth associated maintenance are mmol gDCW^{-1} and $\text{mmol gDCW}^{-1} \text{ h}^{-1}$, respectively.

Organism	GAM	NGAM	r_{ATP}	Reference
<i>C. autoethanogneum</i>	111.2	2.2	5.3	–
<i>C. autoethanogneum</i>	41.3	–	1.2	[Marcellin <i>et al.</i> , 2016]
<i>Clostridium ljungdahlii</i>	46.7	0.45	1.8	[Nagarajan <i>et al.</i> , 2013]
<i>Clostridium beijerinckii</i>	40.0	8.5	9.6	[Milne <i>et al.</i> , 2011]
<i>Acetobacter woodii</i>	150.0	0.8	5.0	[Bainotti and Nishio, 2000]
<i>Bacillus subtilis</i>	105.0	9.0	11.9	[Oh <i>et al.</i> , 2007]
<i>Escherichia coli</i>	97.0	18.9	21.6	[Hempfling and Mainzer, 1975]
<i>Saccharomyces cerevisiae</i>	71–91	< 1	<3.3	[Verduyn <i>et al.</i> , 1990]

5.3.3 Carbon sources

The ability of *C. autoethanogneum* to use different carbon/energy sources for the production of biomass (and maintenance) was predicted with FBA. Each compound was used as sole carbon source in the *in silico* minimal medium by allowing influx across the corresponding transporter while applying the $|\mathbf{v}|$ objective. The simulation results were compared to experimental data (Table 5.3), showing good agreement with known carbon/energy sources.

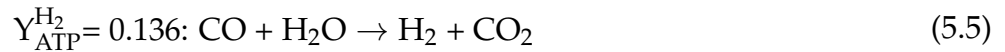
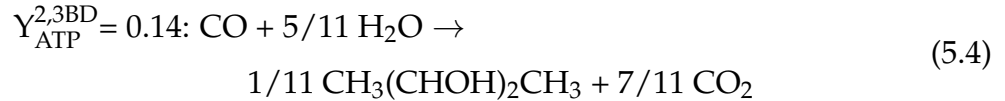
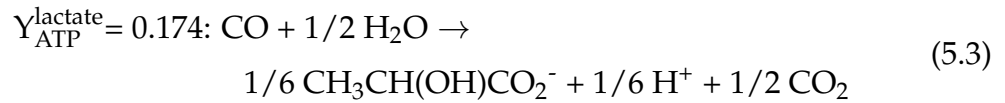
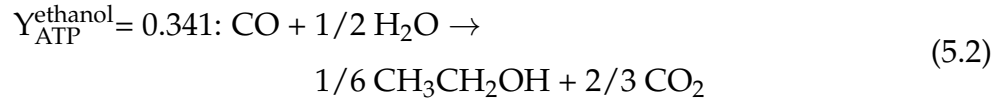
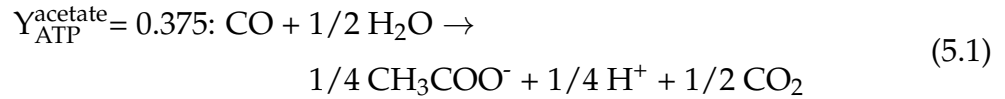
5.3.4 Product spectrum

The optimal ATP yield (Y_{ATP}) per mole substrate uptake with CO as the sole carbon and energy source was 0.375 with acetate as the sole product (net stoichiometry shown in Eq. 5.1). A sub-optimal Y_{ATP} value (0.342) is associated with ethanol production (Eq. 5.2). A hierarchy of products based on their associated ATP-yield is shown below, demonstrating that the model can return flux distributions associated with the expected range of aqueous/gaseous products:

Table 5.3: Carbon-source validation results.

carbon source	Model predicts growth?	Growth observed experimentally?	reference
CO	YES	YES	a,b,c
CO ₂ (+ H ₂)	YES	YES	a,c
fructose	YES	YES	a,c
xylose	YES	YES	a,d
arabinose	YES	YES	a
rhamnose	YES	YES	a
pyruvate	YES	YES	a
glutamate	YES	YES	a
formate	YES	NO	a
fumarate	YES	YES**	a,e
glucose	NO*	NO	a
lactate	NO	NO	a
acetate	NO	–	–
ethanol	NO	–	–
2,3-butanediol	YES	–	–

references: a = [Abrini *et al.*, 1994], b = [Marcellin *et al.*, 2016], c = [Mock *et al.*, 2015], d = [Cotter *et al.*, 2009a], e = [Breitkopf, 2018]. * *C. autoethanogenum* does not encode a glucose transporter. ** References a and e disagree, see discussion.



Molecular hydrogen was the product associated with the lowest ATP yield from carbon monoxide. This was the only product (not counting carbon dioxide) which, under standard culture conditions for *C. autoethanogenum*, would be in a gaseous state. The flux distribution leading to hydrogen production uses only four enzymes (not including the artificial ATPase reaction): CODH, Rnf, ATP synthase, Nfn, and HytA-E. In this solution, the

protons reduced to hydrogen in the HytA-E reaction are released from the water consumed in the CODH reaction.

5.3.5 Growth rate

An average measured CO uptake rate of 16.52 ± 0.02 mmol gDCW⁻¹ h⁻¹ was applied as a fixed-flux constraint on CO transport, which, with the objective function set to maximise biomass production, lead to a predicted optimal specific growth yield of 0.027 h⁻¹. The measured specific growth rate was 0.028 ± 0.001 h⁻¹.

5.3.6 Objective functions

The sensitivity of the model's product fluxes to different objective functions was tested. The results are shown in Table 5.4. The results are compared to iCLAU786, for which the min $|v_{\text{cat}}|$ objective could not be trialled since information on spontaneous reactions is not available for this model [Marcellin *et al.*, 2016].

Table 5.4: Sensitivity of product fluxes to choice of FBA objective. Includes results computed with iCLAU786 [Marcellin *et al.*, 2016].

model	objective	v_{acetate}	v_{ethanol}	v_{CO_2}	v_{CO}
MetaCLAU	min $ v_{\text{cat}} $	0.06	2.7	12	18.4
	min $ v $	0.07	2.7	12	18.4
	max μ	2.3	0	5	10
iCLAU786	min $ v $	0	2.4	10.9	16.7
	max μ	1.9	0	5.1	10

5.4 DISCUSSION

A genome scale model of *C. autoethanogenum* has been constructed and parameterised with experimentally derived values representing ATP maintenance costs and cellular biomass composition. [Thauer *et al.*, 1977] states that the largest cause of NGAM cost is maintaining cell turgor pressure. If this is true, the high NGAM reported here may suggest that turgor pressure is a major expenditure for *C. autoethanogenum*. This could have ramifications for industrial scale-up where cells can be subject to intense pressure stress. The next step is to test the validity of the parametrised model.

Table 5.2 shows that the GAM for *C. autoethanogenum* was higher than

other well-studied organisms. This may relate to the maintenance of cellular turgor pressure, which has been cited as an important function in growing autotrophic bacteria [Thauer *et al.*, 1977]. Conversely, the NGAM was considerably lower than other organisms, especially *Escherichia coli*. This may relate to relatively high numbers of flagella and pili produced and maintained by *E. coli* [Pratt and Kolter, 1998]. Notably, the r_{ATP} value estimated in this study is over four times higher than the value applied in the previously published model of *C. autoethanogenum* [Marcellin *et al.*, 2016]. Since r_{ATP} is applied as a constraint in model analysis, this result raises doubts about the parameterisation of the [Marcellin *et al.*, 2016] model and its ability to accurately simulate *C. autoethanogenum* growth and viability.

The GSM model built in Chapter 4 has undergone analyses to test its validity.

The model is able to support biomass and maintenance on all known growth substrates. This demonstrates good predictive capability. Formate is the only disagreement. One possible explanation is that the 100 mM formate concentrations used in [Abrini *et al.*, 1994] were toxic; 100 mM formate was found to be toxic to *Escherichia coli* [Zaldivar and Ingram, 1999], while 30 mM has been shown to inhibit growth of *Clostridium ljungdahlii* [Ramíó-Pujol *et al.*, 2014]. On the other hand, this disagreement may arise from some toxic or thermodynamic constraint which the structural modelling approach does not account for.

Analysis of ATP-yielding metabolism revealed that the model was capable of producing the full experimentally observed product profile of *C. autoethanogenum*. Hydrogen was also an ATP-yielding product. The hierarchy of products in terms of ATP yield is in itself instructive: acetate and ethanol are at the top, with yields around three times higher than lactate and butanediol. This fits with the established behaviour of *C. autoethanogenum* as a producer of acetate and ethanol, primarily [Abubackar *et al.*, 2016; Abubackar *et al.*, 2015; Abubackar *et al.*, 2012]. Since acetate is the superior product in terms of ATP yield, it is not immediately clear why the organism makes ethanol at all. This question will be investigated in the following chapter.

The growth rate computed with the model matches well to the experimentally measured specific growth rate. Interpretation of this result must take into account that the growth rate is estimated from the growth yield.

This estimation requires knowledge of the substrate (CO) uptake rate and biomass composition. Agreement between the subsequently calculated growth rate and the experimentally measured rate indicates that the organism has adopted a flux distribution which maximises its biomass yield.

The product spectrum is sensitive to the choice of objective function. Optimal solutions computed with both minimising total network flux ($|\mathbf{v}|$) and minimising total network enzyme catalysed flux ($|\mathbf{v}_{\text{cat}}|$) as objective functions include ethanol as the dominant product. In contrast, optimal solutions computed with maximisation of growth as objective function produce acetate only. One possible interpretation of this is that under conditions in which *C. autoethanogenum* pursues enzyme economy, flux distributions leading to ethanol production are superior, while under conditions in which it is advantageous for *C. autoethanogenum* to maximise growth, acetate will be the preferred product.

The model has shown a good degree of validity, especially with respect to the feedstock which this thesis focuses on (CO). In the following chapters, the model will be used to investigate ways of increasing the production of ethanol and 2,3BD.

Chapter 6

METABOLIC SHIFTS

6.1 INTRODUCTION

As discussed in the general introduction to this thesis, published attempts to engineer *C. autoethanogenum* have focused on the tuning of bioprocess parameters (pH, feedstock, media composition *etc.*) [Abubackar *et al.*, 2016; Abubackar *et al.*, 2012; Abubackar *et al.*, 2015]. These studies have not produced a clear strategy for improved ethanol or 2,3BD production from CO. This is important because CO is the main component of the industrial waste-gas used as feedstock in the bioprocess which this thesis aims to support [Mock *et al.*, 2015; Köpke *et al.*, 2011c]. Furthermore, studies have demonstrated the influence of bioprocess parameters on the metabolic activity of many other industrially relevant *Clostridia* [Skonieczny and Yargeau, 2009; Zhu and Yang, 2004; Dabrock *et al.*, 1992; Bothun *et al.*, 2004], most notably the pH shift in *Clostridium acetobutylicum* [Gottwald and Gottschalk, 1985; Bahl *et al.*, 1982]. For these reasons, it was decided that initial analyses of the model would focus on the network’s response to two important bioprocess parameters: pH and CO-supply.

An important bioprocess parameter is the pH of the growth medium [Harms *et al.*, 2002]. This is of particular importance to *C. autoethanogenum* which relies on a transmembrane proton gradient ($\Delta\mu$) to yield ATP *via* ATP synthase during autotrophic growth [Biegel *et al.*, 2011; Mock *et al.*, 2015]. In *C. autoethanogenum*, $\Delta\mu$ is thought to have a value of ≈ 1 , with the organism’s optimum pH for growth somewhere between 5.5 and 6. Up to now, pH-shift experiments in *C. autoethanogenum*—which involve dropping

the growth media pH below 5—have given contradictory results. Some experiments enhanced ethanol production [Abubackar *et al.*, 2016; Abubackar *et al.*, 2012; Abubackar *et al.*, 2015; Guo *et al.*, 2010] while others diminished ethanol production [Martin *et al.*, 2016]. Here the genome-scale model (GSM) established in the previous chapters will be used to investigate the effect of pH adjustments of product profile. The challenge was to find an appropriate representation of pH in the GSM, which does not incorporate metabolite concentrations – another instance of the inherent limitations of standard FBA.

Some authors have developed a method which computes ‘proton flux’ as a function of the rate of dissociation of every acidic metabolite involved in a given GSM [Senger and Papoutsakis, 2008a; Senger and Papoutsakis, 2008b]. A simpler method is possible, however, since the range of pH values applied may drop below the pK_a of acetate (≈ 4.8). This means that acetate efflux may become infeasible due to diffusion of the protonated acetic acid back across the cellular membrane. Thus, in this chapter, restrictions on acetate efflux are used to investigate the effect of pH on the metabolism of *C. autoethanogenum*.

Due to the scale of gas emissions from steel mills (> 1 million tonnes of CO per year per furnace), the hypothetical upper limit of gas supply to a bioreactor is very high [Worrell *et al.*, 2001; Siitonen *et al.*, 2010]. Furthermore, since the bioprocess under development should prevent the release of CO into the environment, it follows that the more CO incorporated into biomass or commercial products, the better. Obviously, one way to achieve greater consumption of CO is to scale-up the entire process under a carbon-limited growth regime (see [Marcellin *et al.*, 2016]). This would mean that, while the CO uptake per unit of bacterial culture (*e.g.* gDCW/L) remains constant, the total amount of CO uptake increases. While scaling-up the bioprocess is generally desirable, the optimisation of process efficiency, *i.e.* increasing the CO uptake per unit of bacterial culture and its conversion to products, is an equally important approach. In fact, given the cost associated with maintaining and powering a large facility, the smaller, high-efficiency version may be superior.

A factor which may affect bioprocess efficiency is carbon-limitation. This is the scenario in which rate of growth of a bacterial population is limited by the supply of its carbon source. Experimental studies aimed at manipulating

the product profile of *C. autoethanogenum* have so far focused on carbon-limited growth regimes [Abubackar *et al.*, 2016; Abubackar *et al.*, 2012; Abubackar *et al.*, 2015; Cotter *et al.*, 2009b; Marcellin *et al.*, 2016; Valgepea *et al.*, 2017a] while neglecting non-carbon limitation – the case where some other nutrient becomes limiting and the carbon source is in excess. Importantly, the carbon source studied in this work (CO) is also the energy source. This means that an excess supply of CO brought about by the inability to increase biomass represents an excess supply of energy. We argue, therefore, that non-carbon-limited growth regimes are worth investigating because an excess supply of energy may result in the diversion of flux towards other, more reduced compounds to maintain redox balance [Johnson *et al.*, 1931]. While it is also possible that excess energy could be released as heat, it seems unlikely that an organism adapted to survive at the ‘thermodynamic limit of life’ (*i.e.* adapted so that its metabolism operates close to thermodynamic equilibrium) could easily adopt this strategy [Schuchmann and Müller, 2014; Thauer *et al.*, 1977].

It seems probable that hydrogen production might form part of the cell’s metabolic strategy to sink excess reducing power. This follows from the fact that just 5 enzymes are required to carry electrons from CO to H₂ (CODH, Rnf/ATP synthase, Nfn, HytA-E) at steady state while releasing carbon as CO₂ (see Section 5.3.4). However, it has been observed that hydrogen production becomes thermodynamically unfavourable if its partial pressure in the external environment exceeds a critical value (3×10^4 Pa for H₂ production from ferredoxin under standard conditions) [Angenent *et al.*, 2004].

One product of high importance is the platform chemical 2,3-butanediol (2,3-BD) – a trace component of the native *C. autoethanogenum* product profile, the downstream products of which have an estimated global market value of \$43 billion in sales [Köpke *et al.*, 2011b]. While the potential for 2,3-BD synthesis from industrial waste gas is well established [Köpke *et al.*, 2011b], a clear strategy for its increased production has not been published despite the GSM analyses presented in [Marcellin *et al.*, 2016]. The work below uses the novel GSM established in Chapters 4 and 5 to find such a strategy. Specifically, we test the hypotheses that 2,3-BD is an efficient sink for excess reducing power supplied under non-carbon-limited conditions before and after a thermodynamic limit on hydrogen production is imposed.

In order to assess the qualitative predictive capability of the model, exper-

imental data from pH-shift and gas-shift experiments conducted by Anne Henstra at the University of Nottingham are displayed in Figs. 6.1 and 6.5, respectively. HPLC and micro GC analyses were conducted on the supernatant and headspace of the subsequent culture, respectively [Henstra *et al.*, unpublished]. The pH shift was a continuous-culture experiment in which the external environment of a steady-state bacterial culture was reduced from pH 5 to pH 4. The gas shift experiment involved enforcing a non-carbon-limiting growth regime in which increasing CO inflow rates, achieved with continuous sparging, could not support increased growth due to insufficient levels of other nutrients. Non-carbon-limitation in this case corresponds to an excess energy supply (since CO is both the carbon and energy source). In both experiments, the steady-state growth rate was 0.028 h^{-1} (as determined by the dilution rate of the bioreactor).

6.2 METHODS

In this chapter, the FBA scanning technique (see Section 3.4.4) is used to investigate the model's response to changes in culture conditions, including pH, carbon *vs.* non-carbon limitation (*i.e.* a 'gas shift') and constraints on hydrogen production.

6.2.1 pH shift

To simulate the metabolic consequences of a pH shift, FBA scanning analysis was conducted in which flux values in the range 0 to 10 ($\text{mmol gDCW}^{-1} \text{ h}^{-1}$) were set as upper-bound constraints on the acetate transporter. Optimal solutions were computed with minimisation of CO uptake (v_{CO}) as the objective function and $\mu = 0.028 \text{ h}^{-1}$, the latter constraint representing the growth-rate set in the pH-shift experiment. Network flux minimisation objectives were avoided since the flux-minimal solutions produce little acetate (see Section 5.3.6). The subsequent changes in product profile and changes in intracellular fluxes could thus be monitored. As stated in the introduction to this chapter, this approach was based on the idea that rates of acetate efflux become infeasible due to a pH-induced shift in the acetic acid dissociation equilibrium.

6.2.2 Gas shift

To simulate the gas-shift under non-carbon limitation, an FBA scan was applied in which CO uptake spanned the range 15-45 mmol gDCW⁻¹ h⁻¹ (which covers the experimental range, see Fig. 6.5). The lower bound of this range was the uptake rate required to support a fixed growth rate of 0.028 h⁻¹, thus all CO uptake rates above this value represent non-carbon limitation. To test the effect of a restriction on hydrogen efflux [Angenent *et al.*, 2004], the hydrogen transporter was closed (*i.e.* fixed to zero), and the scan repeated.

6.3 RESULTS

6.3.1 pH shift

Restrictions on the upper bound of acetate efflux were applied to investigate the metabolism of *C. autoethanogenum* under a pH shift. With the minimisation of v_{CO} objective function, ethanol production diminished linearly with increasing acetate efflux, reaching zero at an acetate efflux rate of ≈ 3.75 mmol gDCW⁻¹ h⁻¹. In the region involving ethanol production, acetaldehyde:ferredoxin oxidoreductase (AFOR) and alcohol dehydrogenase (Adh) carry flux in the acetate- and aldehyde-reducing directions, respectively. These activities converting intracellular acetate to ethanol while enabling ATP recovery at the site of Ack. Fig 6.1 shows a plot of the FBA scan results and experimental data from the pH shift. Simulated ethanol-efflux values form a line with gradient = -0.73 and intercept = 2.7 .

The model-calculated trend seems generally to involve higher acetate and ethanol efflux rates than the trend suggested by the data points. This could be explained by an inaccuracy in the measured biomass composition or estimated ATP maintenance costs, which could cause an over-demand on ATP generation *via* acetate/ethanol production. Despite this inaccuracy, the trend suggested by the measured data points appears consistent with the model-calculated trend.

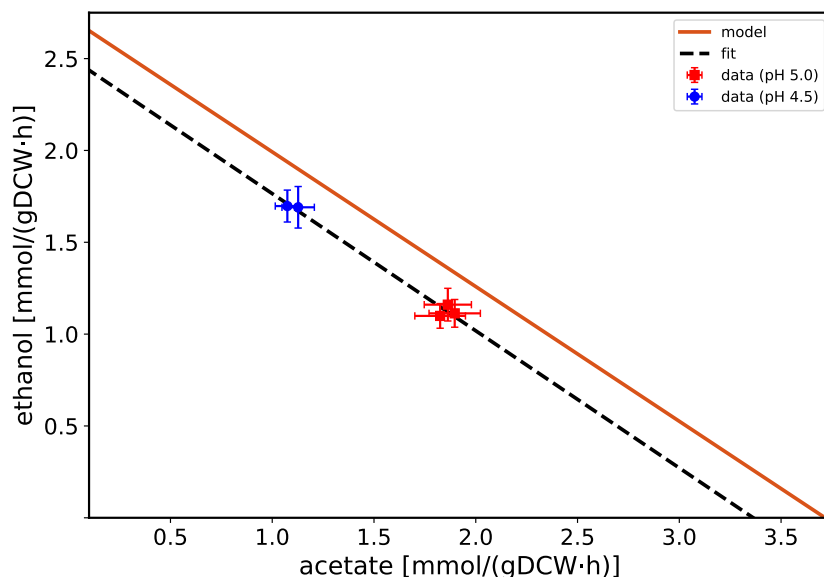


Figure 6.1: Simulated pH shift with experimental data. Orange line shows inverse linear relationship between acetate and ethanol efflux computed with the *C. autoethanogenum* GSM (gradient = -0.73 , intercept = 2.7). Orange markers and error-bars respectively show experimentally observed efflux rates and confidence intervals from curves fitted to raw experimental data (see Section 2.1.5). Dashed line shows linear fit to data computed with orthogonal distance regression (as implemented in [Jones *et al.*, 2014]). Experiments were carried out by Anne Henstra at the University of Nottingham.

6.3.2 Gas shift

A range of CO uptake rates exceeding the minimum required to support growth was applied as fixed flux constraints to investigate metabolic changes under non-carbon limitation. The FBA scan over CO uptake rates resulted in a shift from majority acetate production to ethanol production between uptake values of 17 - 18.4 mmol gDCW $^{-1}$ h $^{-1}$ (Fig. 6.2a), and subsequent hydrogen production with CO uptake values > 18.4 mmol gDCW $^{-1}$ h $^{-1}$ (Fig. 6.2b). When the model was constrained to allow no production (or uptake) of hydrogen, 2,3-butanediol and lactate efflux appear in optimal solutions after the initial switch from acetate to ethanol (where 2,3-butanediol efflux increases with CO uptake but lactate efflux remains constant, see Fig. 6.3). The constraint on hydrogen efflux was applied to mimic increasing hydrogen partial pressure in the external environment [Angenent *et al.*, 2004].

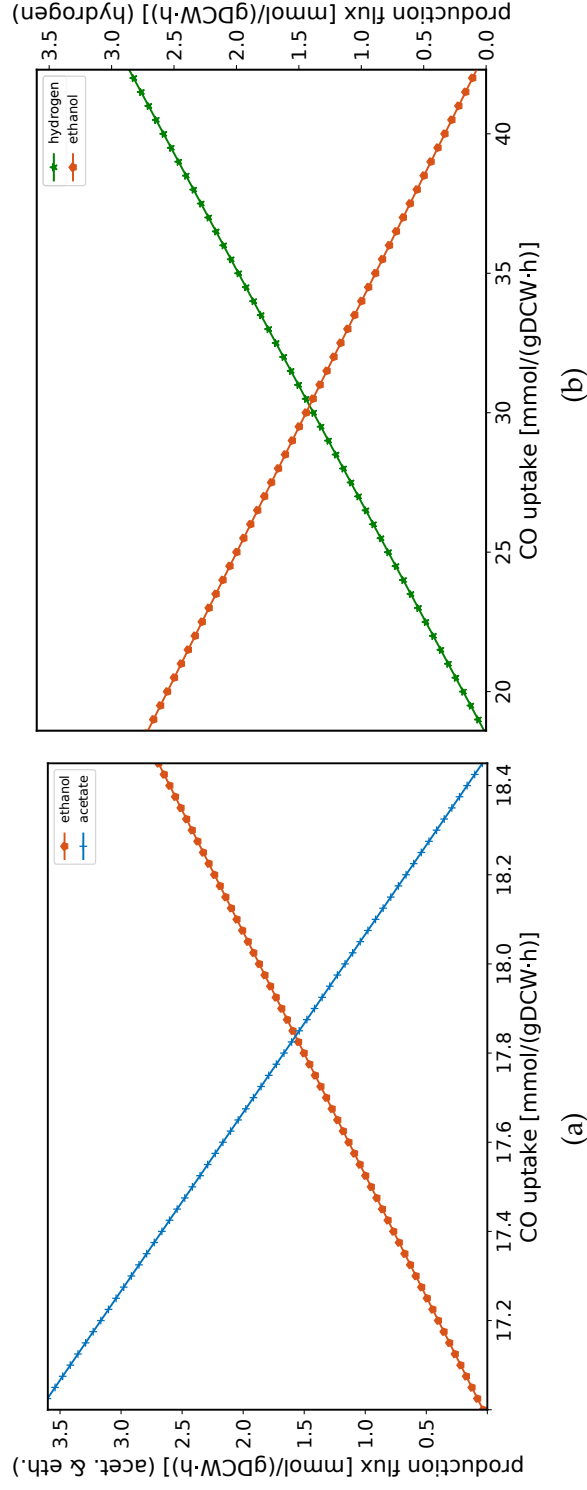


Figure 6.2: Production fluxes in optimal solutions with v_{CO} values in the range (a) $17 < v_{\text{CO}} < 18.4 \text{ mmol gDCW}^{-1} \text{ h}^{-1}$, and (b) $18.4 < v_{\text{CO}} < 45$. (All fluxes in $\text{mmol gDCW}^{-1} \text{ h}^{-1}$.)

Between CO uptake values of 17-18.4 mmol gDCW⁻¹ h⁻¹, there were 18 changers (see Section 3.4.4), 13 of which were internal reactions (*i.e.* not transporters). Of these 13 internal reactions, 3 increase their (absolute) flux value over this range (CODH, AFOR, Adh), while the remaining 10 decrease (Pta, Ack, Nfn and the WLP), see Fig. 6.4b.

In solutions with CO uptake in the range $18.4 < v_{CO} < 42.5$, the previously inactive HytA-E increased its flux in strict proportion to the rate of efflux of hydrogen ($v_{H_2} = 2v_{HytA-E}$). The NADPH requirement for HytA-E was met with increasing Nfn flux. Thus, the onset of increasing hydrogen production requires Nfn to move from the set of decreasing reactions to the set of increasing reactions. When hydrogen efflux was constrained to zero, the pathway for 2,3-butanediol production reported in [Köpke *et al.*, 2011c] (*i.e.* ALS, ALDC and BDH) was activated and increased its flux monotonically with CO uptake. In addition, LDH was activated but did not increase flux with increasing CO uptake.

6.4 DISCUSSION

The modelling results reported above suggest that a restriction on acetate efflux, as hypothetically caused by a drop in external pH, will lead to increased ethanol efflux. As minimisation of v_{CO} was used as the objective function, this result is expected since acetate production is associated with the maximum Y_{ATP} from CO while ethanol is associated with the second highest Y_{ATP} (see Section 5.3.4). Furthermore, the utilisation of AFOR for the reduction of acetate in these solutions is consistent with published data [Liew *et al.*, 2017]. The experimentally observed inverse linear relationship between acetate and ethanol is predicted by the model; the gradient of the model-generated line is within the confidence value of a linear fit to the experimental data. This result shows that a restriction on acetate efflux—as is hypothetically caused by a reduction in external pH—corresponds to an increase in ethanol efflux.

To summarise, results in this chapter suggest an inverse linear trend between a drop in pH and ethanol production (assuming that a constraint on acetate efflux is the governing phenomenon behind this process) which, on a quantitative level, agrees well with experimental data. The difference in the intercept parameter between the GSM-generated and data-fitted lines may

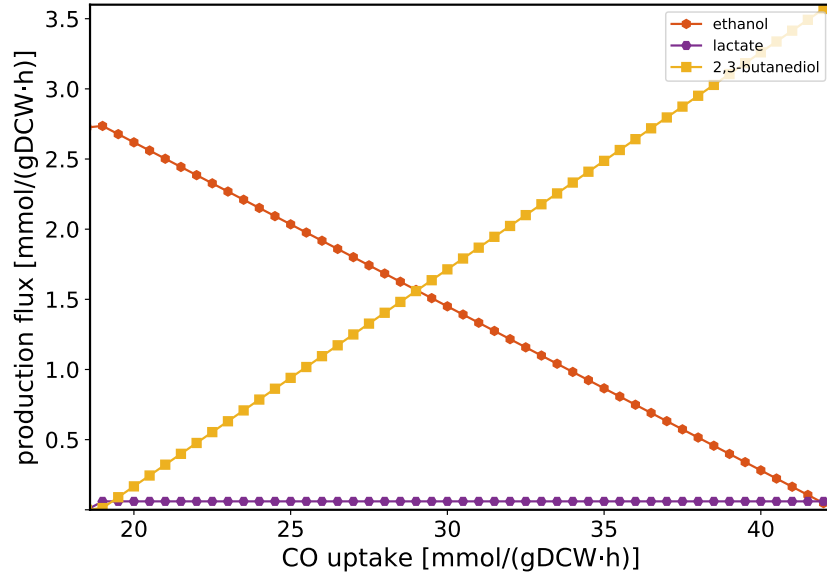


Figure 6.3: FBA-simulated gas-shift in *C. autoethanogenum*: hydrogen production blocked. Orange, yellow and purple lines respectively denote ethanol, 2,3-BD and lactate efflux.

relate to an inadequacy in the model's biomass composition.

An important step forward will be to incorporate a quantitative relationship between acetate efflux and external pH into the model analysis. This would enable a scan over pH directly, rather than the use of acetate efflux as a proxy for pH. This may be achieved by incorporating additional aspects of the bioprocess with, for example, dynamic modelling techniques. Indeed, a recent study involving kinetic modelling of *C. autoethanogenum* metabolism derived, through somewhat more in-depth modelling approach, a linear trend between 'proton flux' (*i.e.* pH) and ethanol production [Chen *et al.*, 2018], supporting the findings reported above.

One potential development of the results obtained in this chapter is the degree to which protons from acetate can drive ATP synthase. An interesting avenue for future analysis of the model would be to enable protons coupled to acetate efflux to drive the ATP synthase. However, the simultaneous repression of the Rnf complex (by decreasing pH) would be best accounted for using dynamic modelling approaches. Thus this idea remains a suggested future direction.

Oversupply of CO to the metabolic network leads to excess reducing

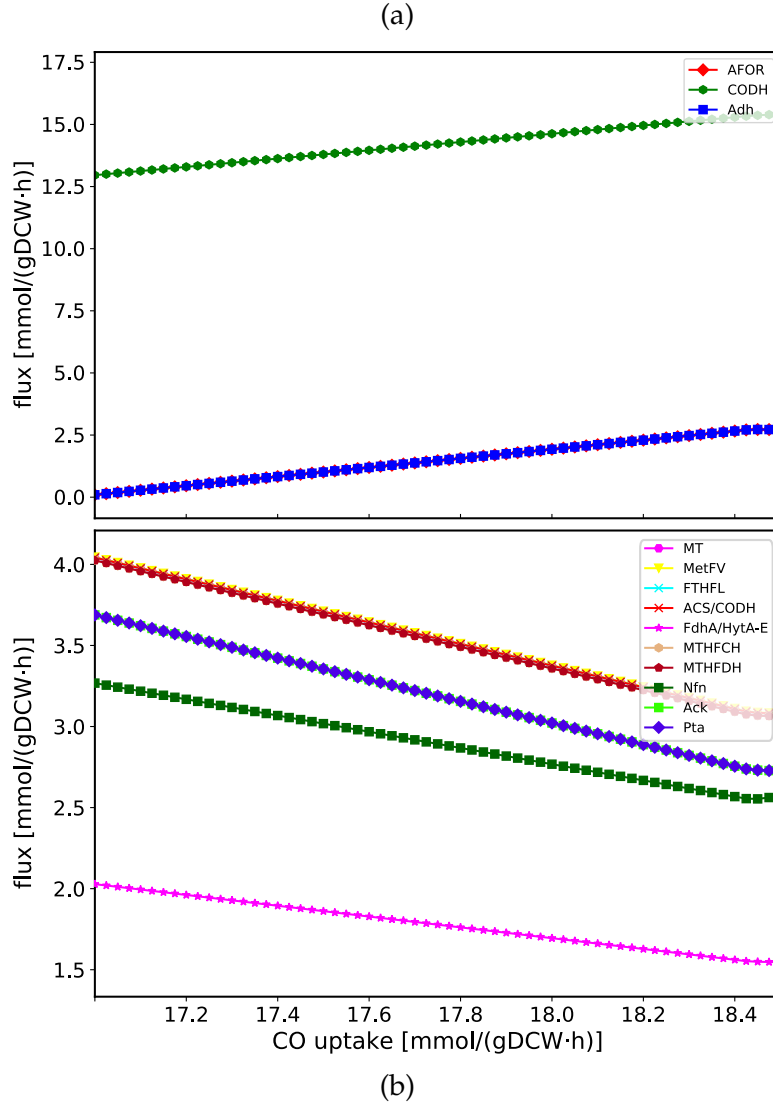


Figure 6.4: Internal fluxes in optimal solutions with v_{CO} values in the range $17 < v_{CO} < 18.4$ mmol gDCW⁻¹ h⁻¹. (a) Reactions with increasing flux: red, green and blue line-markers respectively represent flux carried by AFOR, CODH and Adh. (b) Reactions with decreasing flux: methyl transferase (MT), methyl-THF synthase (MetFV), formate-THF ligase (FthL), acetyl-CoA/CO dehydrogenase (ACS/CODH), formate dehydrogenase (Fdh/HytA-E), methenyl-THF cyclohydro-lase (MTHFCH), methylene-THF dehydrogenase (MTHFDH), Nfn, acetate kinase (Ack) and phosphotransacetylase (Pta). (All fluxes in mmol gDCW⁻¹ h⁻¹.)

equivalents, forcing flux through pathways allowing the removal of superfluous reducing power (electrons) from the system [Köpke *et al.*, 2011b]. This has been tested using FBA by increasing the ratio between energy-substrate supply (in this case, CO) and growth rate over a fixed range (Fig. 6.2a) representing non-carbon limited conditions. Results show that increasing the

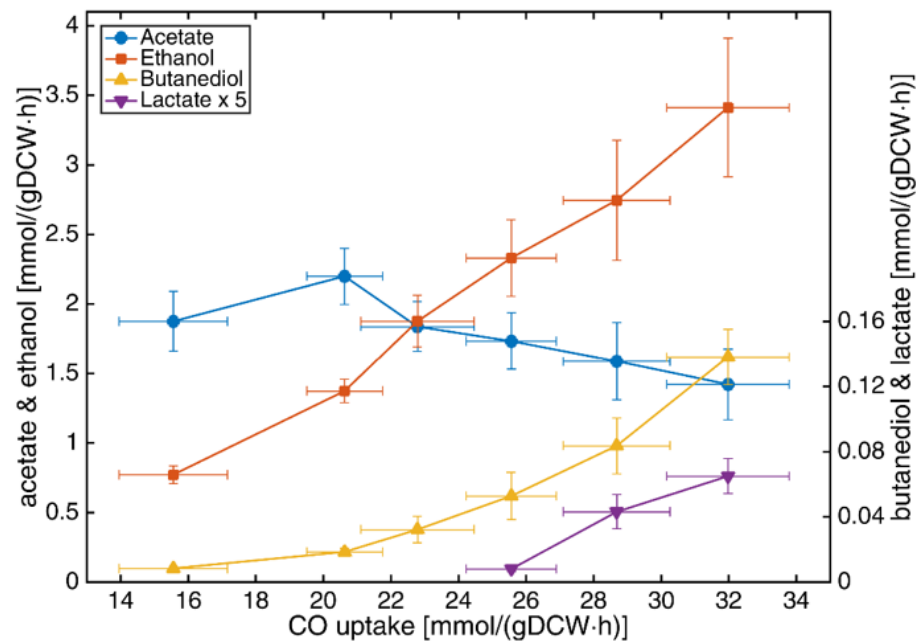


Figure 6.5: Gas shift data. Liquid product efflux rates *vs.* steady-state CO uptake rates. Data presentation courtesy of Thomas Millat at the University of Nottingham. Orange, blue, yellow and purple points respectively represent ethanol, acetate, 2,3-butanediol and lactate efflux rates (lines connect data points, *i.e.* they are not fitted or model-generated). Error bars show standard error.

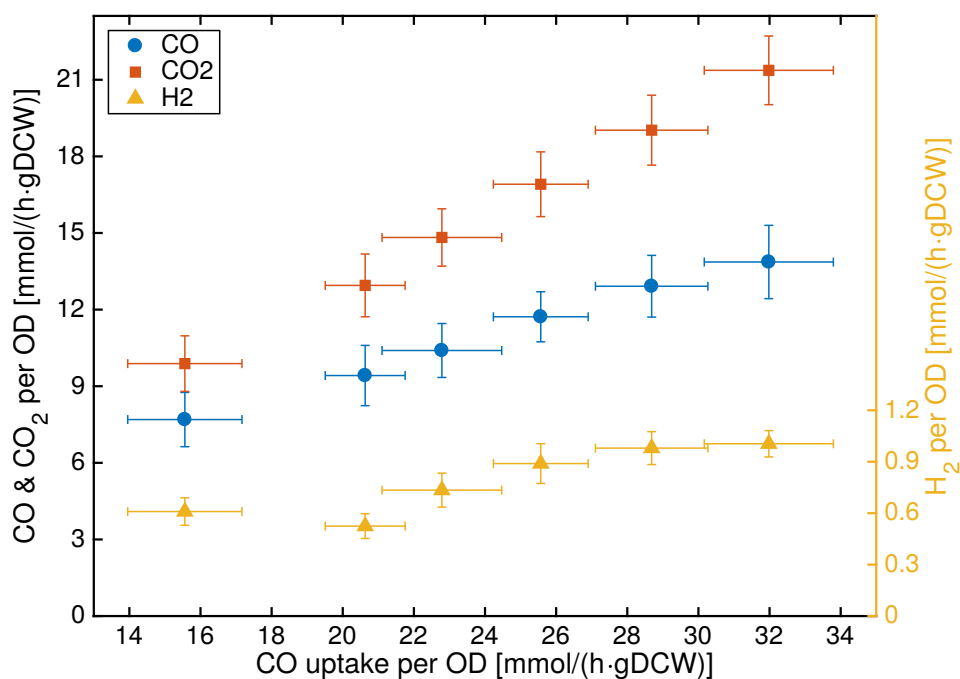


Figure 6.6: Gas shift data. Gas efflux rates *vs.* steady-state CO uptake rates. Data presentation courtesy of Thomas Millat at the University of Nottingham. Orange, blue, and yellow points respectively represent CO, CO₂ and H₂. Error bars show standard error.

influx of CO beyond the rate necessary to achieve a fixed growth rate causes a switch from acetate to ethanol production over the uptake range 17-18.5 mmol gDCW⁻¹ h⁻¹. This provides a new strategy for increasing ethanol production, supported by the experimental data shown in Fig. 6.5, which shows a shift from acetate to ethanol production after the onset of a non-carbon limited growth regime starting at a CO uptake rate $\approx$ 20 mmol gDCW⁻¹ h⁻¹.

The inflection points in Figures 6.2 to 6.4 correspond to alternative pathways being ‘switched’ on or off as the bounds of the solution space change. This allows new optimal points to be found by FBA, which, depending on the geometry of the bounded solution space, may involve the use of otherwise inactive reactions. A hypothetical metabolic cause for pathways being switched on or off might be the thermodynamic infeasibility of a reaction caused by the accumulation of internal metabolites. In the case of the onset of hydrogen production seen in Figure 6.2, one possible biological cause could be the upregulation of genes encoding (or related to the control of) hydrogenase by the Rex regulator, which detects cellular redox poise [Brekasis and Paget, 2003].

It has also been hypothesised that bacterial hydrogen production will be restricted by external constraints during continuous culture. More specifically, hydrogen production may become thermodynamically unfavourable if its partial pressure in the external environment approaches some critical value. This will cause metabolism to favour the production of other electron sinks (such as lactate) [Claassen *et al.*, 1999; Levin *et al.*, 2004; Angenent *et al.*, 2004]. To test this, the CO uptake scan was repeated with hydrogen transport restricted to zero. Optimal flux distributions resulting from this simulation showed 2,3-butanediol production increasing monotonically with CO uptake after the acetate/ethanol switch (Fig. 6.3). Furthermore, lactate production appeared in optimal flux distributions, but was unchanged with increasing CO uptake, suggesting an association with biomass production. Thus the analysis has uncovered a testable set of bioprocess conditions for which 2,3BD is included in optimal metabolic flux distributions, increasing with the gas inflow.

On a qualitative level, the relationship between hydrogen production and 2,3BD production predicted with the GSM appears to hold in the experimental data shown in Figures 6.5 and 6.6, in which the rate of 2,3BD production increases as H₂ production saturates in the range of CO uptake

starting at $\approx 20 \text{ mmol gDCW}^{-1} \text{ h}^{-1}$.

An important discussion point arises from the use of the flux minimisation objective to simulate the gas shift: if the flux minimisation objective represents a minimal protein investment, is this object appropriate for simulating non-CO limitation, where cells are exposed to excess energy? One answer relates to the organism's evolutionary background: autotrophic acetogens like *C. autoethanogenum* have evolved over vast time periods to survive environments with a low availability of nutrients [Schuchmann and Müller, 2014], thus it is possible that their affinity for a strategy of energetic efficiency is not easily overcome.

A more obvious question and direction for future research which arises from this work is the following: could 2,3-butanediol production be induced by knocking-out hydrogenase activity. This would be a novel strategy for 2,3-butanediol production which could be executed using recently-developed CRISPR technology [Hsu *et al.*, 2014b].

Chapter 7

METABOLIC ENGINEERING

7.1 INTRODUCTION

This chapter reports efforts to identify gene knock-out (KO) targets for the improved production of ethanol and 2,3-butanediol in *C. autoethanogenum*. Devising knock-out strategies through the manual, logical consideration of biochemical literature is difficult [Hauf *et al.*, 2000]. In contrast, the analysis of a genome scale model (GSM) provides a view of the entire metabolic system, offering a superior approach for predicting metabolic fluxes before and after the inactivation of a set of genes [Oberhardt *et al.*, 2009].

Results from the previous chapter suggest that a restriction on hydrogen efflux causes the inclusion of 2,3-butanediol production in optimal flux distributions computed with the flux-minimisation objective. An obvious question arising from this result is the following: Could this behaviour be enforced by inactivating hydrogen-evolving hydrogenase activity? This question is investigated here.

No gene knock-outs strategy for the improved production of ethanol emerged from the previous chapter. Therefore, an important goal for this chapter was to identify gene KO targets for increased ethanol yield using techniques which consider the entire steady-state solution space of an extracted sub-network [Klamt and Gilles, 2004]. Since such methods are designed to reveal gene KO targets whose relationship to the target compound is not obvious, it was important to answer two questions: 1) *Why* do the knock-out targets cause an increase in product yield? and 2) How does the metabolic network support biomass and maintenance in the absence of the

reactions associated to the gene targets?

Since the optimal flux distributions induced by the inactivation of reactions may represent entirely novel modes of metabolism, a new method for the assessment of the thermodynamic feasibility in terms of metabolite concentrations proposed in [Peres *et al.*, 2018] was implemented and applied.

7.2 METHODS

A simple algorithm which set the flux of unique combinations of reactions to zero for FBA while monitoring optimal solutions for increased product yields was applied (see supplementary materials, ‘flux.balance.analysis.py’). Unique reaction combinations of set length 2 and 3 were searched for.

FBA with $v_{\text{HytA-E}} = 0$ and $v_{\text{CO}} = 25$ was used to test the hypothesis that 2,3BD production could be induced under non-carbon limitation by knocking-out genes responsible for hydrogen production. This analysis was performed in a ipython-notebook available in the supplementary materials (see ‘model analyses’ directory).

The subnetwork-extraction technique described in Section 3.4.4 was used to extract a subnetwork of reactions associated with ethanol production. The range of fixed ethanol efflux values applied was $0 \leq v_{\text{ethanol}} \leq 100$. The scan included a background of biomass and ATP maintenance so that the resulting subnetwork was compatible with cell growth. The enzyme subsets of the extracted subnetwork were computed and all consistent enzyme subsets were converted into lumped reactions enabling model reduction.

Elementary modes of the extracted, reduced subnetwork were computed to help identify knock-out targets by min-cut set analysis [Klamt and Gilles, 2004]. The aim was to identify reactions which fit two criteria: reactions should 1) carry flux in all non-ethanol producing modes, and 2) carry no flux in at least one ethanol-producing mode.

Any identified knock-out targets were tested as such using FBA of the GSM. If optimal solutions had an improved ethanol yield, a second FBA solution was computed with ethanol minimisation as the objective to check whether ethanol production had been made essential by the inactivation of the KO target. (If $v_{\text{ethanol}} > 0$ in the ethanol-minimised solution computed with biomass and ATP maintenance, ethanol production would be essential

for cell viability.)

Finally, any optimal solution resulting in improved product yields was tested using the LP problem proposed by Peres *et al.* detailed in 3.5. This analysis was used to return maximum and minimum metabolite concentrations allowing negative Gibbs energy of reaction values in the appropriate directions for the subject flux distribution. Two ranges of allowable metabolite concentrations were used: $0.1 \text{ nM} \leq m \leq 100 \text{ mM}$ and $0.1 \text{ nM} \leq m \leq 10 \text{ mM}$ (based on the known range of metabolite concentrations in *E. coli* [Flamholz *et al.*, 2012]). Standard Gibbs energy values were taken from equilibrator [Flamholz *et al.*, 2012] (pH = 6 and ionic strength = 0.1), except for ATP synthase and Rnf, for which $\Delta_r G'$ values were estimated using published data [Tremblay *et al.*, 2013; Henry *et al.*, 2006; Henry *et al.*, 2007] (see Appendix A5 for workings).

Mutant strains were generated and characterised by Chris Humphreys at the university of Nottingham.

7.3 RESULTS

7.3.1 Hydrogenase inactivation

FBA with $v_{\text{HytA-E}} = 0$ and $v_{\text{CO}} = 25$ was used to test the hypothesis that 2,3BD production could be induced under non-carbon limitation by knocking-out genes responsible for hydrogen production. When the allowable flux carried by HytA-E (the H_2 -producing reaction in the network) was constrained to 0, 2,3BD production was included in optimal solutions computed with the flux-minimisation objective. An ipython notebook demonstrates these results (see supplementary materials, 'model_analyses.ipynb').

7.3.2 Ethanol production subnetwork

The subnetwork-extraction technique described in Section 3.4.4 was used to extract a subnetwork of reactions associated with ethanol production. A plot of reaction fluxes changing with increasing enforced ethanol efflux is shown in Fig. 7.1. These 'changers' formed an ethanol-production associated sub-network consisting of 85 reactions and 86 metabolites (see 'ethanol_subnet.spy' in supplementary materials). GPRs were defined for all enzymatic-reactions.

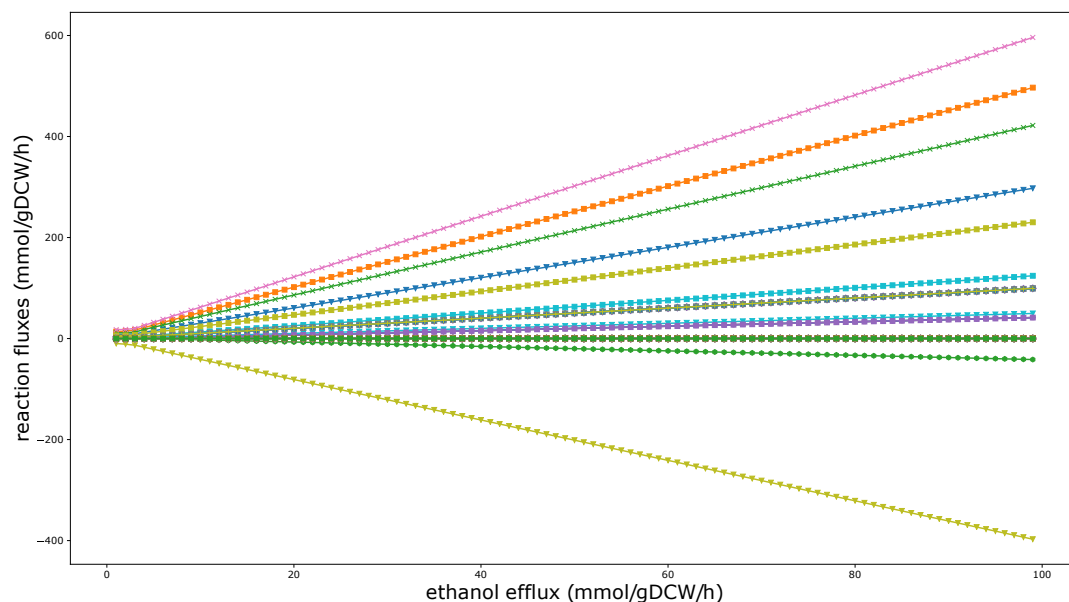


Figure 7.1: Steady-state reaction fluxes changing with increasing ethanol efflux. Each coloured line represents a different reaction. Markers along a given line represent flux values for the corresponding reaction in optimal solutions at a given ethanol efflux constraint. These reactions form a sub-network associated with ethanol production.

Next, the enzyme subsets of the subnetwork were computed to further reduce the model and identify any flux-amplification relationships (as outlined in Section 3.2.1). There were 39 enzyme subsets in total, 20 of which were consistent (see Table 7.1. Each set of reactions belonging to a consistent enzyme subset was converted into a lumped reaction. The inclusion of these lumped reactions into the subnetwork and the removal of the individual reactions comprising them further reduced the model to 60 reactions and 58 metabolites (see ‘reduced_model.spy’ in supplementary materials).

The elementary modes of this reduced subnetwork were computed to help identify knock-out targets. There were 105 elementary modes, 38 of which produced ethanol (‘ethanol modes’) and 67 involving either ethanol consumption or no ethanol transporter activity (‘non-ethanol modes’). The aim was to identify reactions which fit two criteria: reactions should 1) carry flux in all non-ethanol modes, and 2) carry no flux in at least one ethanol mode. These reactions, if inactivated, would disable non-ethanol producing steady-states while preserving at least one ethanol-producing steady state. Thus, their associated genes would be knock-out candidates.

The set of reactions carrying flux in all 67 non-ethanol modes was identi-

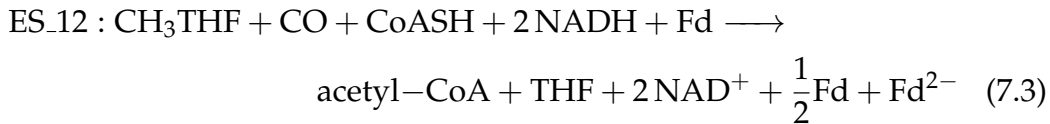
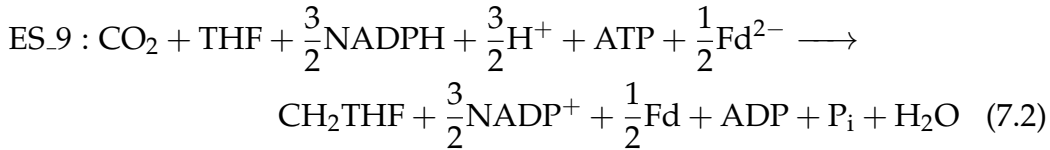
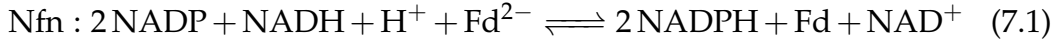
Table 7.1: Consistent enzyme subsets in a set of reactions associated with ethanol production in *C. autoethanogenum*.

	BioCyc reaction IDs
Ess_3	CITRAMALATE-LYASE-RXN METHYLASPARTATE-AMMONIA-LYASE-RXN ISOCITRATE-DEHYDROGENASE-NAD+-RXN S-2-METHYLMALATE-DEHYDRATASE-RXN METHYLASPARTATE-MUTASE-RXN RXN-14047 CITLY-RXN
Ess_4	6.3.4.16-RXN UREA-CARBOXYLASE-RXN ARGINASE-RXN ORNCARBAMTRANSFER-RXN ARGININE-DEAMINASE-RXN ALLOPHANATE-HYDROLASE-RXN
Ess_7	atp_synthase rnf
Ess_9	FORMATETHFLIG-RXN 1.2.1.2-RXN METHENYLTHFCYCLOHYDRO-RXN METHYLENETHFDEHYDROG-NADP-RXN
Ess_10	RXN-7737 N-ACETLDIAMINOPIMELATE-DEACETYLASE-RXN 2.3.1.89-RXN RXN-4822
Ess_12	methyl_transferase acetcoa_synth methylthf_synth
Ess_13	RXN-14117 RXN-14074
Ess_16	PHOSACETYLTRANS-RXN ACETATEKIN-RXN
Ess_18	carbonic_anhydrase carbonic_acid_diss PYRUVATE-CARBOXYLASE-RXN
Ess_19	ALCOHOL-DEHYDROG-RXN ethanol_tx

Table 7.2: Ethanol and acetate production fluxes in optimal flux distributions for KO mutants. (z is the corresponding solution's objective value when $|\mathbf{v}_{\text{cat}}|$ is the objective. Y_{ethanol} is the molar yield of ethanol, $\frac{v_{\text{ethanol}}}{v_{\text{CO}}}$.)

enzyme	objective	v_{acetate}	v_{ethanol}	μ	v_{CO}	Y_{ethanol}	no. of reactions (z)
WT	min $ \mathbf{v}_{\text{cat}} $	0.06	2.7	0.028	16.4	0.165	237 (113.61)
	max μ	4.36	0	0.036	20	0	242
Δnfn	min $ \mathbf{v}_{\text{cat}} $	0	5.28	0.028	33.75	0.156	236 (269.63)
	max μ	0	3.19	0.01	20	0.160	237
ΔES_9	min $ \mathbf{v}_{\text{cat}} $	0	0	0.028	43.34	0	235 (272)
	max μ	4.42	0	0.031	20	0	237
ΔES_{12}	min $ \mathbf{v}_{\text{cat}} $	-	-	-	-	-	-
	max μ	-	-	-	-	-	-

fied. This set contained 5 reactions: nfn , Ess_9 , Ess_{12} , acetate_tx and proton_tx . 55 reactions carried zero flux in at least one ethanol-producing mode. This set contained all 5 reactions carrying flux in every non-ethanol producing mode. Details of the enzyme-catalysed reactions in this intersecting set are given below:



The hypothesis is simple: Any member of this set of reactions will force steady-state ethanol production if inactivated. Each was tested as a knock-out for increased ethanol production in the GSM with both net flux minimisation and growth maximisation as objective. Results are summarised in Table 7.2.

No ethanol-growth coupling was induced by any of the knock-outs (*i.e.* ethanol efflux was 0 in all ethanol-minimised solutions). Despite this, it is clear from Table 7.2 that the superior knock-out target is nfn. When the growth maximisation objective is used with $v_{\text{Nfn}} = 0$, ethanol replaces acetate as the sole product, while with the min. $|\mathbf{v}_{\text{cat}}|$ objective, acetate production was abolished but the ethanol yield decreased by 5% compared to the WT solution (where $\text{yield} = \frac{v_{\text{ethanol}}}{v_{\text{CO}}}$).

To understand this result, the elementary modes of the reduced model after removal of Nfn were computed and inspected. Surprisingly, the conversion of carbon monoxide to ethanol induced in the genome scale model was not represented in the elementary modes of the reduced model. The only available net conversion was: $4\text{CO} + 2\text{acetate} + 2\text{H}^+ + 2\text{H}_2\text{O} \longrightarrow 2\text{ethanol} + 4\text{CO}_2$ ($\Delta_r G^\circ = -127\text{KJ/mol}$). Furthermore, only one of these acetate-consuming modes gave a positive ATP yield (Y_{ATP}) (see Fig. 7.2).

The flux distribution returned from FBA did not include acetate uptake, thus elementary modes supporting the above conversion were not responsible for the improved ethanol production. To investigate how the improved ethanol production was achieved, the optimal (flux minimal) solution was inspected. The solution consisted of 236 reactions, of which 28 appeared responsible for the generation of ATP for maintenance and growth based on their carrying large flux (*i.e.* $v > 0.01$) relative to other fluxes in the solution. This was confirmed by computing the max ATP yield FBA solution, which contained exactly the same set of 28 reactions.

Since in Chapter 4 the reaction catalysed by Nfn is shown to be the only NADPH source with CO-fed growth, NADPH-producing reactions in the $v_{\text{nfn}} = 0$ solution were identified. There was just one: An NADP⁺-dependent glutamate dehydrogenase. This enzyme catalyses the reaction shown in Eq. 7.4, taking over from Nfn as the source of NADPH for growth.



The maximum ATP yield ($Y_{\text{ATP}}^{\text{max}}$) with $v_{\text{nfn}} = 0$ was 0.197 on CO, 53% of the WT $Y_{\text{ATP}}^{\text{max}}$ (0.375). The flux distribution carried by the reactions of the Y_{ATP} solution is shown in Fig. 7.3. In this flux distribution, CO is

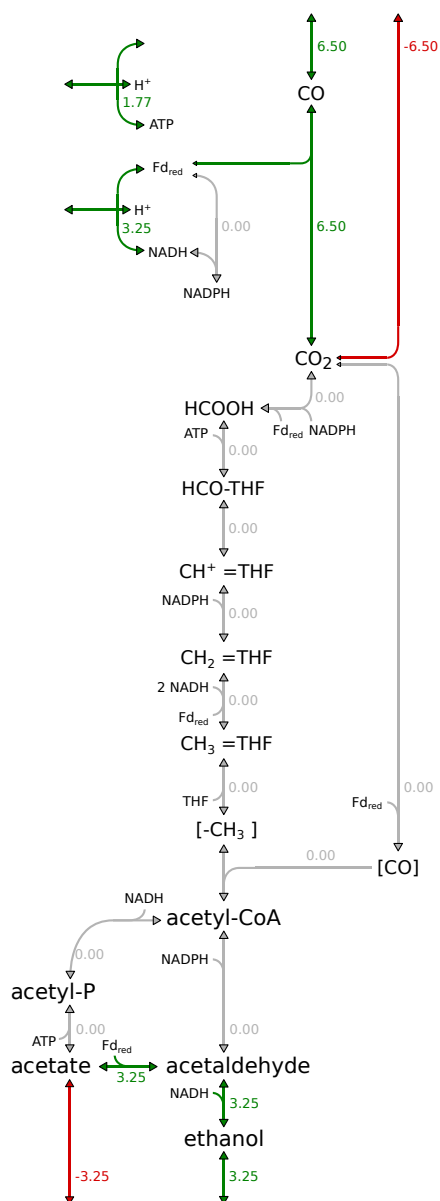


Figure 7.2: An ethanol-producing elementary mode of the subnetwork with $v_{\text{nfn}} = 0$. $Y_{\text{ATP}} = 0.27$. Arrows represent enzyme-catalysed chemical reactions or transport processes. Reaction arrows coloured green and red respectively represent forward and reverse reaction directions (where the forward direction runs from the top of the figure to the bottom and/or left to right). Grey arrows represent unused reactions. The flux value for a given reaction is written next to the corresponding reaction arrow.

fixed to acetyl-CoA without the conventional WLP. Formate dehydrogenase, formyl-THF ligase, formyl-THF cyclohydrolase and methylene-THF dehydrogenase are all inactive. Instead of this conventional linear route, half of the acetyl-CoA produced is recycled *via* pyruvate, phosphoenolpyruvate (PEP), 2-phosphoglycerate (2PG), glyceraldehyde-3-phosphate (G3P), 3-phospho-hydroxypyruvate (3PHP), 3-phospho-serine (3-P-SER), serine (SER) and glycine (GLY). Glutamine (GLN), glutamate (GLT) and α -keto-glutarate (2KG) are also involved in a small motif which balances the production of ammonium in the conversion of GLY to methylene-THF as catalysed by the glycine cleavage system (MetaCyc ID: GCVMULTI-RXN).

The flux distribution shown in Fig.7.3 is more complicated than the standard WLP pathway route in that it involves more reactions and utilises a novel cyclic flux pattern through central metabolism. Despite this radical re-configuration of metabolism, the set of essential genes did not change with $v_{nfn} = 0$. Gene associations for each of the involved reactions are given in Table 7.3.

A further necessary condition for feasibility is that a set of metabolite concentrations must exist which enable all changes in Gibbs energy of reaction to be negative. The LP problem proposed in [Peres *et al.*, 2018] and detailed in 3.5 was used to return maximum and minimum metabolite concentrations allowing negative Gibbs energy of reaction values in the appropriate directions for this flux distribution.

For the *nfn* knock-out flux distribution, no solutions enabling thermodynamic feasibility could be found with metabolite concentrations (m) in the range $1 \text{ nM} \leq m \leq 100 \text{ mM}$ (based on the known range of metabolite concentrations in *E. coli* [Flamholz *et al.*, 2012]). For the WT network, the maximum Y_{ATP} solution was thermodynamically feasible with metabolite concentrations in the physiological range. Metabolites whose maximum and minimum feasible concentrations were respectively lower and higher than the maximum and minimum limits of the physiological range are given in Table 7.4.

7.4 Experimental data

Experimental testing of the growth and product profile of wild-type (WT) and *nfn* deletion mutant (Δnfn) strains of *C. autoethanogenum* was conducted

Table 7.3: Fluxes and gene-protein-reaction relationships (GPRs) for internal reactions involved in Y_{ATP} -maximal solution computed with $v_{\text{nm}} = 0$. The ATPase referred to here is an artificial reaction whose flux value represents the generation of ATP for work.

reaction ID	flux	GPR
RXN-14074	0.167	CLAU_1878
GCVMULTI-RXN	0.167	CLAU_0457
GLUTAMATE-SYNTHASE-NADH-RXN	0.167	CLAU_3759
ACETATEKIN-RXN	0.167	CLAU_3275
RXN0-5114	0.167	CLAU_0799
codh	0.667	(CLAU_1561-1579) OR (CLAU_2922-2924)
PHOSACETYLTRANS-RXN	0.167	CLAU_3274
PYRUFERRREDUCT-RXN	0.167	CLAU_0896 OR CLAU_2947
methy1thf_synth	0.333	(CLAU_1572 AND CLAU_1573)
atp_synthase	0.364	(CLAU_2285-2294)
2PGADEHYDRAT-RXN	0.167	CLAU_1710
ALCOHOL-DEHYDROG-RXN	0.167	CLAU_0030 OR CLAU_1040 OR CLAU_1460 OR CLAU_1764 OR CLAU_1766 OR CLAU_1794 OR CLAU_2382 OR CLAU_3196 OR CLAU_3514 OR CLAU_3655 OR CLAU_3656 OR CLAU_3861
RXN-14117	0.167	CLAU_2378
acetoac_synth	0.333	(CLAU_1561 AND CLAU_1564 AND CLAU_1566 AND CLAU_1568-1570 AND CLAU_1577-1579) OR (CLAU_2922-2924)
ADENNYL-KIN-RXN	0.167	CLAU_1878
PSEKTRANSAM-RXN	0.167	CLAU_1189
acetaldehyde-oxidoreductase	0.167	(CLAU_0089) OR (CLAU_0099)
mf	0.667	(CLAU_3144-3149)
ATPASE-RXN	0.197	-
methy1_transferase	0.333	CLAU_1567
GLYOHEMETRANS-RXN	0.167	CLAU_0663, CLAU_3158
PGLYCDEHYDROG-RXN	0.167	CLAU_1027 OR CLAU_1138 OR CLAU_1205 OR CLAU_1332 OR CLAU_1990 OR CLAU_2158 OR CLAU_3166 OR CLAU_3170
3PGAREARR-RXN	0.167	CLAU_0686, CLAU_1711
GLUTAMINESYN-RXN	0.167	CLAU_1976

Table 7.4: Metabolite concentrations supporting fluxes computed from ATP-yield-maximising FBA of the ‘wild-type’ GSM.

metabolite ID	min concentration (mM)	max concentration (mM)
FORMATE	0.256	100
ATP	1×10^{-6}	3.15×10^{-5}
5-10-METHENYL-THF-GLU-N	1×10^{-6}	0.000391
FORMYL-THF-GLU-N	4.07×10^{-6}	0.00159
Pi	3.18	100
ext_PROTON	39.0	100
ADP	3.176	100
THF-GLU-N	0.256	100

by Dr. Chris Humphreys at the University of Nottingham in 2017. The Δnfn mutant strain was generated using CRISPR gene editing technology and grown in batch cultures with CO as the sole source of carbon and energy. HPLC and GC analysis were conducted on the headspace of the subsequent culture. Figs. 7.4 & 7.5 show results.

For the wild-type batch culture experiment, the lag growth-phase commences immediately and reaches a final relative OD (OD/OD_0) of 0.84 at $t = 96$ hours. Acetate remains the dominant product throughout, though ethanol is also produced in lower amounts. The presence of both products, and the dominance of acetate is consistent with the genome-scale model predictions shown in Table 7.2 to the extent that the growth-maximal solution gives acetate as the dominant (sole) product and both products are included in the flux-minimal solution ($|v_{cat}|$). Conversely, there must be some inadequacy in the FBA performed since the experimentally observed phenotype is not reproduced by either of the individual optimal solutions in Table 7.2. This disagreement may relate to batch culture conditions, which differ from the continuous culture conditions from which the models parameter values were derived. Furthermore, the cultures may have been subject to some chemical or thermodynamic constraint not captured during standard FBA of the GSM.

For the Δnfn mutant, the time point which most closely corresponds to the start of the classic microbial growth pattern is $t = 72$ hours. From the growth-curve fit, the time point at which the maximal exponential growth rate was achieved was calculated as $t_{\mu_{max}} = 146$ hours. Since metabolic steady-state occurs during the exponential growth phase, the HPLC data points taken nearest to $t_{\mu_{max}}$ are the most suitable for comparison with model sim-

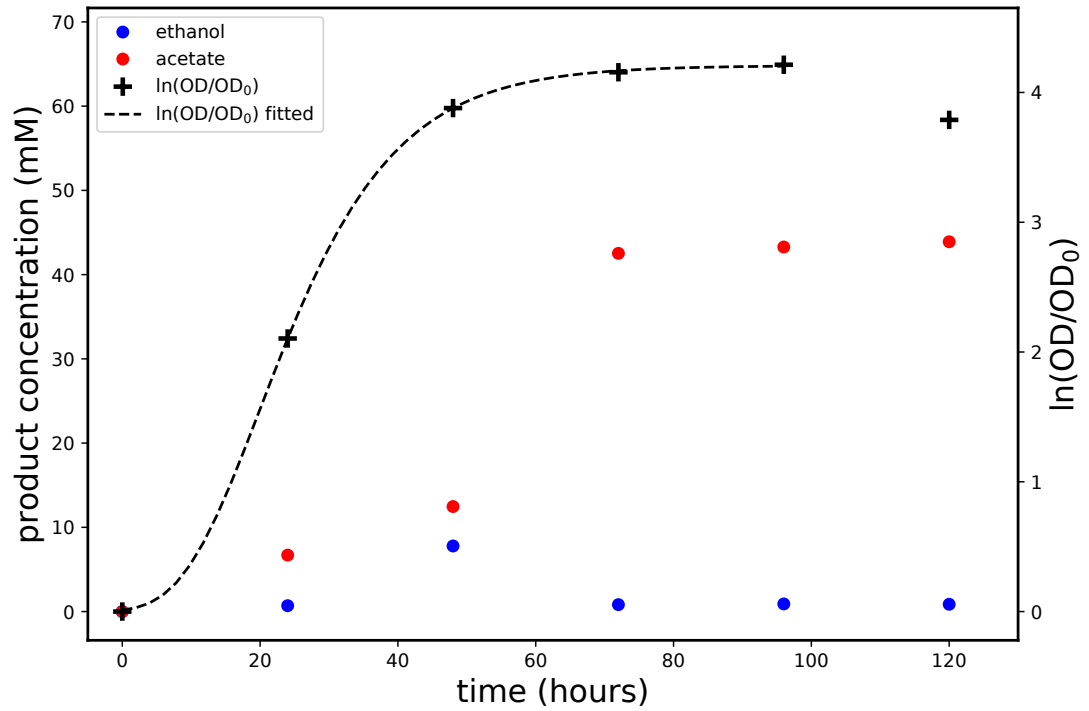


Figure 7.4: Growth curve and product profile of the WT strain. Data generated by Dr. Chris Humphreys at the University of Nottingham, 2017.

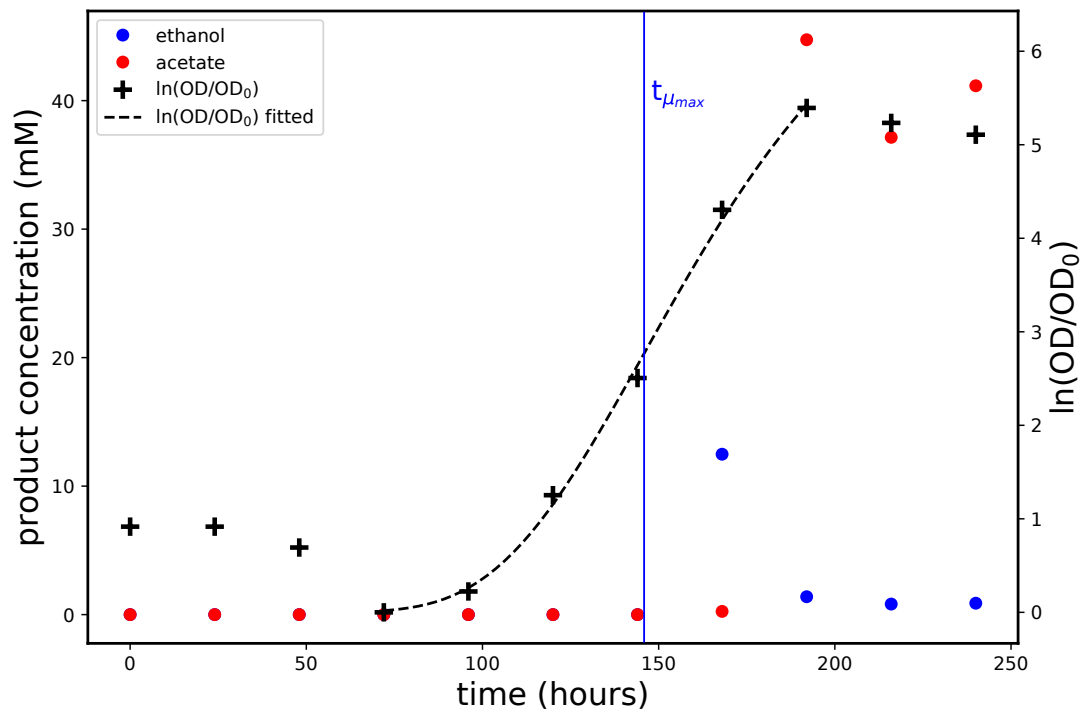


Figure 7.5: Growth curve and product profile of the Δnfn KO mutant strain. Mutant and data generated by Dr. Chris Humphreys at the University of Nottingham, 2017.

ulations. While neither acetate nor ethanol is detected prior to $t_{\mu_{max}}$, it is notable that ethanol is the sole detected product at the next sampling point. This is consistent with the predicted optimal behaviour of the Δnfn mutant.

In both experiments, the cultures reach similar maximum relative (OD/OD_0), however the fitted maximum growth rate parameter is much lower for the Δnfn mutant (0.07) than the wild-type (0.14). Unfortunately, since the uptake rate of carbon monoxide cannot be calculated from the available data, no comparison between the growth-rates calculated using the genome scale model and the experimental data can be performed.

The Δnfn mutant culture reduces its optical density before commencing exponential growth. The significant initial OD values are indicative of a carry-over effect from the fructose-fed starter cultures (personal communication, Chris Humphreys), *i.e.* the cells may have been able to maintain themselves with residual amounts of fructose present in the media.

7.5 DISCUSSION

The main aim of this chapter was to identify candidate gene KO targets for increased ethanol production and assess the thermodynamic feasibility of subsequent flux distributions. These primary aims have been met, since a testable KO candidate for the improved yield of ethanol was identified (*nfn*) and a new methodology for testing the thermodynamic feasibility [Peres *et al.*, 2018] of a flux distribution has been developed and implemented.

The inclusion of Nfn in the set of candidate KOs is surprising because its activity is thought to be the only source of NADPH during CO-fed growth. This would at first appear to be crucial since NADPH is required by methylene-THF dehydrogenase in the WLP of *C. autoethanogenum*. However, flux distributions in the flux-minimal FBA solution computed using the genome scale model show that the reaction catalyzed by methylene-THF dehydrogenase can be bypassed by recycling half of the acetyl-CoA generated in the reaction catalysed by ACS/CODH (see Fig. 7.3). Even so, NADPH is required for biomass production, so some alternative means of NADPH production must be taking place in the Δnfn solution. The only reaction producing NADPH in the aforementioned solution is the glutamate dehydrogenase (glutamate + $NADP^+ + H_2O \longrightarrow \alpha\text{-ketoglutarate} + NADPH + H^+ + NH_4^+$) participating in the small cycling pathway seen in Fig. 7.3.

In the Δnfn solution depicted in Fig. 7.3, ethanol production *via* acetaldehyde:ferredoxin oxidoreductase and alcohol dehydrogenase replaces the oxidation of reduced ferredoxin and NADH performed in WT solutions (by Nfn). In summary, the removal of Nfn forces the metabolism to oxidise reduced electron carriers elsewhere; the enzymes leading to ethanol being the flux-minimal solution. Furthermore, NADPH production for biomass is taken-over by glutamate dehydrogenase, and the NADPH demand represented by methylene-THF dehydrogenase is circumvented all together.

While all reactions involved in the novel flux distribution described above had valid (*i.e.* non-hypothetical) gene associations (see Table 7.3), it remained unclear if a set of physiological metabolite concentrations existed that would support thermodynamic feasibility of the fluxes. This was checked using an implementation of a recently proposed method for testing thermodynamic feasibility which returns maximum and minimum metabolite concentrations supporting a flux distribution. Results showed that no metabolite concentrations could be found for the flux distribution given an allowable concentration range of $1 \text{ nM} \leq x \leq 100 \text{ mM}$, suggesting that the novel mode of metabolism returned from FBA is infeasible, unlike the ATP-maximal solution computed with the wild-type network (see Table 7.4). This result conflicts with experiments showing batch growth of the Δnfn mutant in which ethanol is the dominant product at samples taken during exponential growth (Fig. 7.5).

Recent Tradis data generated at the University of Nottingham (Craig Woods *et al.*, unpublished) suggest strongly that Nfn is in fact essential. This result confirms the findings of the thermodynamic-based metabolite concentration analysis, but conflicts with the experimental data shown in Section 7.4. One possible explanation could be that ‘fructose carry-over’ is responsible for the growth of the Δnfn mutant seen here (Humphreys, personal communications).

While [Rydzak *et al.*, 2015] propose a similar serine-involving route for growth on glucose without formate production in a Δpfl *Clostridium thermocellum* mutant, the Δnfn metabolism described above remains a novel hypothesis. Transcriptomic and proteomic data of a Δnfn mutant are needed to attempt its refutation.

Chapter 8

GENERAL DISCUSSION

As stated in the General Introduction, the motivation for this work was the need for a computational model of *C. autoethanogenum* metabolism to help engineer a bioprocess for carbon recycling. If achieved, this biotechnological advancement will provide a strategy to alleviate fossil-fuel dependence. Recently, the company LanzaTech has built and initialised the world's first commercial-scale plant for the conversion of waste CO to ethanol ¹ using laboratory-evolved strains of *C. autoethanogenum*; the scope for implementing rational metabolic engineering strategies for the improvement and diversification of this bioprocess is enormous. Despite this, a model of *C. autoethanogenum* published in 2015 has not yielded a strategy for increased 2,3BD production, nor has it been used to design mutants for improved ethanol production. This fact enhances the importance of the work reported below.

The objectives for this thesis, which forms part of the GASCHEM group project, were: 1) to build, parametrise and validate a genome-scale metabolic network (GSM) of *C. autoethanogenum*, 2) to use the model to investigate how bioprocess parameters can be optimised, and 3) to identify gene knock-out targets for the creation of enhanced production strains. This final chapter provides a critical discussion of the progress made towards each of these targets in Chapters 4 to 7 and considers the wider consequences of the results obtained.

In Chapter 4, a GSM of *C. autoethanogenum* was constructed. A GSM of this organism was desired because its metabolism is the basis for a commercial bioprocess for the utilisation of waste CO; such technology could be

¹see www.lanzatech.com

improved with a model capable of simulating the organism's steady-state metabolism. The model was constructed from a publicly available PGDB for *C. autoethanogenum*, requiring manual curation of energy metabolism and biosynthetic pathways. The curated model was enhanced with a PGDB based on the GASCHEM strain by linking model reactions to locus tags in the GASCHEM genome annotation. This was important to the overall aim of the project because it meant that the model could be linked to the organism actually used in industry. The curated model is able to simulate flux distributions in which the energy substrate is utilised for the generation of ATP, and subsequently to support the production of all identified biomass precursors. These two results are fundamental to the simulation of bacterial growth, *i.e.* the phenomenon upon which the subject bioprocess relies. One outstanding result from the curation of biosynthetic pathways was that *de novo* biosynthesis of NAD(P)⁺ is coupled to succinate production. This hypothesis, not captured by a previously published model of *C. autoethanogenum*, must be investigated further since succinate is a valuable commodity chemical. Furthermore, it was observed that optimal flux distributions for the synthesis of individual biomass precursors were often coupled to different byproducts, meaning that the biomass composition could affect the model's ability to predict the product profile of the industrial strain. Obtaining an accurate biomass composition thus formed one focus of the next chapter.

In Chapter 5 the GSM was parameterised. Parameterisation of the GSM is important to improve the accuracy of model-predictions. This in turn helps to develop the bioprocess because the model can be used to test and generate realistic strategies for its improvement. The first key parameter is biomass composition, the importance of which was emphasised by results in the previous chapter—*i.e.* the variety of byproducts appearing in solutions for the synthesis of individual biomass precursors. This would impact the the model's prediction of the product profile of *C. autoethanogenum*, which is its main purpose. The biomass composition was measured experimentally by colleagues using the GASCHEM strain, meaning the model does not rely on parameter values originating from an organism of unknown relevance (as is the case with the previously published GSM). Of note was the high lipid content and low protein content, for which a definite explanation has not emerged in this work. While macromolecular biomass composition

is known to vary dramatically with different growth rates [Taymaz-Nikerel *et al.*, 2010], it is unclear what influence this could have on the biomass measurements made for this study; samples were taken from continuous culture with a fixed growth rate. It is also possible that an increase in lipid production could form part of a stress response to high levels of solvent production [Isken and de Bont, 1998].

The second key parameter is the ATP maintenance cost of cellular maintenance as represented by the growth associated- and non-growth associated-maintenance costs (GAM and NGAM, respectively). These values, again experimentally measured by GASCHEM colleagues, significantly changed the yields of biomass achieved by the model. Biomass yields are important to the project aims because they partially determine the growth rate of the bacterium and the productivity of the bioprocess. Notably, the experimentally derived maintenance costs implement here differ to those in the previously published *C. autoethanogenum* model [Marcellin *et al.*, 2016] (for which no experimental derivation is described).

Chapter 5 also describes the final step in establishing the model: validation. Since the model in this study has been parameterised differently to the previously published model, *i.e.* experimentally, it was important to test the performance of the model against experimental data. Successful validation shows that the construction and parameterisation steps have worked; it shows that the model can confidently be used to simulate *C. autoethanogenum* metabolism and help to develop the desired conversion of waste carbon to valuable products. The model is able to support growth on all 10 experimentally tested growth substrates. This is a good result because it suggests the model's metabolic network contains degradation pathways which capture the real behaviour of *C. autoethanogenum*. Furthermore, it shows that the model is equipped to explore possible avenues for the diversification of the bioprocess: the use of different feedstocks. The next form of validation was to test if the model could support flux towards known products. Given the project aims, this was the most important behaviour that the model should get right. The model could support flux towards all experimentally observed products (acetate, ethanol, hydrogen, 2,3BD and lactate), associating each with an ATP yield. Finally, the model-simulated growth yield agrees well with the measured growth rate on CO. Given these results, and the project's focus on CO-fed growth, it was concluded that the model was suitably valid

to continue.

In Chapter 6, metabolic shifts were investigated to uncover bioprocess conditions conducive to the production of ethanol and 2,3BD. This approach was taken since the tuning of bioprocess conditions like gas-supply and external pH can be achieved relatively immediately, *i.e.* it does not require the generation of mutants. The pH shift was approximated by assuming that a restriction on acetate efflux occurs if the external medium pH drops below its pK_a value ($pH \approx 4.8$). The optimal solutions computed with the model showed an inverse linear relationship between acetate and ethanol where low acetate efflux corresponded to increased ethanol efflux. Subsequently, a continuous culture pH shift experiment was conducted by collaborators in which the external pH was shifted from 5 to 4.5. The experimentally observed trend between acetate and ethanol closely matched model predictions. This exciting result requires testing at industrial scale. Other future work will be to better define the relationship between external pH and acetate efflux so that pH can be incorporated directly as a model parameter. (Some progress on this question has already been published [Chen *et al.*, 2018].)

The gas shift was simulated by increasing the CO uptake flux to levels exceeding the required rate for growth. This is especially relevant to the industrial application of the subject biotechnology since the potential gas supply rates are very large. While again flux distributions computed with the model show a shift in ethanol *vs.* acetate, the key finding was that 2,3-butanediol production is included in solutions after a restriction of hydrogen efflux is applied. This was supported by experimental data from gas-shift experiments conducted by the GASCHEM group which appear to show 2,3-butanediol production increasing as hydrogen levels saturate. This result offers a new strategy for 2,3BD production from CO: enforce a restriction H_2 production. The model predicts that inactivating the hydrogenase responsible for H_2 production could achieve this phenotype. This hypothesis is currently being tested experimentally. A thermodynamic restriction on hydrogen production could also be tested by artificially filling the bioreactor headspace with H_2 , as shown in [Valgepea *et al.*, 2018]. In summary, the model analysis reported in this chapter has provided novel strategies for the optimisation of the production of commodity chemicals from CO. This will improve the commercial viability of the subject biotechnology, bringing the

alleviation of fossil-fuel dependence closer to reality.

In Chapter 7, KO targets for improved ethanol production were investigated. Ethanol is the main target product from *C. autoethanogenum* fermentation, however progress has been achieved through strain evolution rather than model-directed engineering—an approach which could offer further improvements. The electron bifurcating enzyme Nfn was identified as a candidate KO target using minimum-cut set analysis of an extracted sub-network associated with ethanol production. Experimental data as to the effectiveness of the *nfn* KO mutant were conflicting: initial batch experiments suggested ethanol production was increased in the mutant during exponential growth, however recently-generated TraDIS data (Woods, unpublished) suggest that *nfn* is an essential gene. The application of a new technique suggested by [Peres *et al.*, 2018] has helped to provide an explanation for the likely difficulties with achieving a viable Δnfn mutant; flux distributions responsible for supporting biomass and ethanol production in the mutant network were not thermodynamically feasible unless an extreme range of metabolite concentrations can be maintained in the cells.

The essentiality of Nfn for CO growth indicates that it is the sole source of NADPH and must carry flux for this reason. Since Nfn couples three redox reactions through electron bifurcation, its essentiality offers a possible site for coupling a synthetic redox pathway to the energy metabolism of *C. autoethanogenum*. In terms of future research, the field of incorporating thermodynamic information into the analysis of genome scale metabolic models is gathering momentum, offering two important advancements: 1) a route for the incorporation of large metabolomics datasets as reaction reversibility constraints in GSMs, and 2) a way to avoid physiologically infeasible flux distributions in GSM analysis, this is a weakness associated with the hard reversible/irreversible constraints used in conventional GSM analysis like FBA. The results in this chapter have helped to investigate the autotrophic metabolism of *C. autoethanogenum*, providing an explanation for the essentiality of Nfn. Furthermore, a novel implementation of thermodynamic feasibility analysis has been demonstrated, offering a way to improve the interpretation of model output and produce better strategies for the development of a sustainable bioprocess.

8.1 Future work

A general problem that has arisen from this study is the organism's multiple electron sinks (*i.e.* biomass, hydrogen, ethanol, lactate and 2,3-butanediol) and difficulty of directing flux towards one in particular. As mentioned in the discussion of the metabolic shifts chapter (6), the regeneration of oxidised electron carriers can occur at multiple sites, and while the various reasons for choosing one site over the other have been elucidated above, an easy way to direct flux towards one site of regeneration over the other is lacking. A possible solution to this problem arose from Chapter 7 in relation to the essentiality of Nfn: introduce foreign or artificial electron carrier/donor agents which can only donate/accept electrons from enzymes which have been engineered to do. It is possible to imagine a metabolic system in which the oxidation of such an agent is coupled to the flux carried by Nfn, while its reduction is coupled to the production of a target product. To achieve steady state, the system would have to regenerate the oxidised form of the foreign donor/acceptor at the specified site and would be in 'check-mate', with no choice but to produce the target compound at rates proportional to flux carried in the energy metabolism. This idea, of using foreign/artificial electron donor/acceptor agents for metabolic engineering should form a future research direction.

8.2 Conclusion

In conclusion, a genome-scale metabolic network of *C. autoethanogenum* has been successfully constructed, parameterised and validated. Analysis of the model uncovered strategies for enhanced 2,3BD and ethanol production through modification of bioprocess parameters and gene KOs. Furthermore, a new method for testing the thermodynamic feasibility of model simulations has been implemented and used to gain insight into the autotrophic metabolism surrounding the enzyme Nfn. The results in this work will help to secure the futurity of a promising bioprocess for the reduction of global fossil-fuel dependence.

Bibliography

- [Abrini *et al.*, 1994] Abrini, J., Naveau, H., and Nyns, E.-J. (1994). *Clostridium autoethanogenum*, sp. nov., an anaerobic bacterium that produces ethanol from carbon monoxide. *Archives of Microbiology*, 161(4):345–351.
- [Abubackar *et al.*, 2016] Abubackar, H. N., Fernández-Naveira, Á., Veiga, M. C., and Kennes, C. (2016). Impact of cyclic pH shifts on carbon monoxide fermentation to ethanol by *Clostridium autoethanogenum*. *Fuel*, 178(Supplement C):56 – 62.
- [Abubackar *et al.*, 2012] Abubackar, H. N., Veiga, M. C., and Kennes, C. (2012). Biological conversion of carbon monoxide to ethanol: Effect of pH, gas pressure, reducing agent and yeast extract. *Bioresource Technology*, 114(Supplement C):518 – 522.
- [Abubackar *et al.*, 2015] Abubackar, H. N., Veiga, M. C., and Kennes, C. (2015). Carbon monoxide fermentation to ethanol by *Clostridium autoethanogenum* in a bioreactor with no accumulation of acetic acid. *Bioresource Technology*, 186(Supplement C):122–127.
- [Alon, 2006] Alon, U. (2006). *An introduction to systems biology: design principles of biological circuits*. Chapman and Hall/CRC.
- [Angenent *et al.*, 2004] Angenent, L. T., Karim, K., Al-Dahhan, M. H., Wrenn, B. A., and Domínguez-Espinosa, R. (2004). Production of bioenergy and biochemicals from industrial and agricultural wastewater. *TRENDS in Biotechnology*, 22(9):477–485.
- [Annan, 2018] Annan, F. J. (2018). *B-vitamin requirements of Clostridium autoethanogenum*. PhD thesis, University of Nottingham.

- [Antoni *et al.*, 2007] Antoni, D., Zverlov, V. V., and Schwarz, W. H. (2007). Biofuels from microbes. *Applied microbiology and biotechnology*, 77(1):23–35.
- [Archibald *et al.*, 1968a] Archibald, A., Baddiley, J., and Blumsom, N. (1968a). The teichoic acids. *Advances in Enzymology and Related Areas of Molecular Biology, Volume 30*, pages 223–253.
- [Archibald *et al.*, 1968b] Archibald, A., Baddiley, J., and Button, D. (1968b). The glycerol teichoic acid of walls of *Staphylococcus lactis* i3. *Biochemical Journal*, 110(3):543–557.
- [Axelsson and Gentili, 2014] Axelsson, M. and Gentili, F. (2014). A single-step method for rapid extraction of total lipids from green microalgae. *PloS one*, 9(2):e89643.
- [Bahl *et al.*, 1982] Bahl, H., Andersch, W., Braun, K., and Gottschalk, G. (1982). Effect of pH and butyrate concentration on the production of acetone and butanol by *Clostridium acetobutylicum* grown in continuous culture. *European journal of applied microbiology and biotechnology*, 14(1):17–20.
- [Bainotti and Nishio, 2000] Bainotti, A. and Nishio, N. (2000). Growth kinetics of *Acetobacterium* sp. on methanol-formate in continuous culture. *Journal of Applied Microbiology*, 88(2):191–201.
- [Bath and Tortorello, 2014] Bath, C. A. and Tortorello, M.-L., editors (2014). *Encyclopedia of Food Microbiology*, volume 1. Academic press, 2nd edition.
- [Berg *et al.*, 2002] Berg, J. M., Tymoczko, J. L., and Stryer, L. (2002). *Biochemistry*. W. H. Freeman & Co., 5th edition.
- [Berg and Joklik, 1954] Berg, P. and Joklik, W. (1954). Enzymatic phosphorylation of nucleoside diphosphates. *Journal of Biological Chemistry*, 210(2):657–672.
- [Biegel *et al.*, 2011] Biegel, E., Schmidt, S., González, J. M., and Müller, V. (2011). Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes. *Cellular and Molecular Life Sciences*, 68(4):613–634.

- [Biegel *et al.*, 2009] Biegel, E., Schmidt, S., and Müller, V. (2009). Genetic, immunological and biochemical evidence for a rnf complex in the acetogen *Acetobacterium woodii*. *Environmental Microbiology*, 11(6):1438–1443.
- [Blom *et al.*, 2009] Blom, J., Albaum, S. P., Doppmeier, D., Pühler, A., Vorhölter, F.-J., Zakrzewski, M., and Goesmann, A. (2009). EDGAR: a software framework for the comparative analysis of prokaryotic genomes. *BMC bioinformatics*, 10(1):154.
- [Bothun *et al.*, 2004] Bothun, G., Knutson, B., Berberich, J. A., Strobel, H., and Nokes, S. (2004). Metabolic selectivity and growth of *Clostridium thermocellum* in continuous culture under elevated hydrostatic pressure. *Applied Microbiology and Biotechnology*, 65(2):149–157.
- [Boyer, 1997] Boyer, P. D. (1997). The ATP synthase-a splendid molecular machine. *Annual Review of Biochemistry*, 66(1):717–749.
- [Bradford, 1976] Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry*, 72(1-2):248–254.
- [Breitkopf, 2018] Breitkopf, R. (2018). *Understanding the C₄ dicarboxylic acid metabolism in Clostridium autoethanogenum*. PhD thesis, University of Nottingham.
- [Brekasis and Paget, 2003] Brekasis, D. and Paget, M. S. (2003). A novel sensor of NADH/NAD⁺ redox poise in *Streptomyces coelicolor* A3(2). *The EMBO Journal*, 22(18):4856–4865.
- [Brown *et al.*, 2014] Brown, S. D., Nagaraju, S., Utturkar, S., De Tissera, S., Segovia, S., Mitchell, W., Land, M. L., Dassanayake, A., and Köpke, M. (2014). Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of *Clostridium autoethanogenum* and analysis of CRISPR systems in industrial relevant Clostridia. *Biotechnology for Biofuels*, 7(1):40.
- [Bruno-Barcena *et al.*, 2013] Bruno-Barcena, J. M., Chinn, M. S., and Grunden, A. M. (2013). Genome sequence of the autotrophic acetogen *Clostridium autoethanogenum* JA1-1 strain DSM 10061, a producer of ethanol from carbon monoxide. *Genome Announcements*, 1(4).

- [Buckel and Barker, 1974] Buckel, W. and Barker, H. (1974). Two pathways of glutamate fermentation by anaerobic bacteria. *Journal of bacteriology*, 117(3):1248–1260.
- [Buckel and Thauer, 2013] Buckel, W. and Thauer, R. K. (2013). Energy conservation via electron bifurcating ferredoxin reduction and proton/na⁺ translocating ferredoxin oxidation. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1827(2):94 – 113. The evolutionary aspects of bioenergetic systems.
- [Bult *et al.*, 1996] Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., *et al.* (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science*, 273(5278):1058–1073.
- [Caspi *et al.*, 2008] Caspi, R., Foerster, H., Fulcher, C. A., Kaipa, P., Krumnacker, M., Latendresse, M., Paley, S., Rhee, S. Y., Shearer, A. G., Tissier, C., *et al.* (2008). The metacyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research*, 36(suppl 1):D623–D631.
- [Chen *et al.*, 2018] Chen, J., Daniell, J., Grin, D., Li, X., and Henson, M. A. (2018). Experimental testing of a spatiotemporal metabolic model for carbon monoxide fermentation with *Clostridium autoethanogenum*. *Biochemical Engineering Journal*, 129:64 – 73.
- [Claassen *et al.*, 1999] Claassen, P., Van Lier, J., Contreras, A. L., Van Niel, E., Sijtsma, L., Stams, A., De Vries, S., and Weusthuis, R. (1999). Utilisation of biomass for the supply of energy carriers. *Applied microbiology and biotechnology*, 52(6):741–755.
- [Cotter *et al.*, 2009a] Cotter, J. L., Chinn, M. S., and Grunden, A. M. (2009a). Ethanol and acetate production by *Clostridium ljungdahlii* and *Clostridium autoethanogenum* using resting cells. *Bioprocess and Biosystems Engineering*, 32(3):369–380.
- [Cotter *et al.*, 2009b] Cotter, J. L., Chinn, M. S., and Grunden, A. M. (2009b). Influence of process parameters on growth of *clostridium ljungdahlii* and *clostridium autoethanogenum* on synthesis gas. *Enzyme and Microbial Technology*, 44(5):281–288.

- [Crawford, 1975] Crawford, I. P. (1975). Gene rearrangements in the evolution of the tryptophan synthetic pathway. *Bacteriological reviews*, 39(2):87.
- [Dabrock *et al.*, 1992] Dabrock, B., Bahl, H., and Gottschalk, G. (1992). Parameters affecting solvent production by *Clostridium pasteurianum*. *Applied and environmental microbiology*, 58(4):1233–1239.
- [Daniell *et al.*, 2012] Daniell, J., Köpke, M., and Simpson, S. D. (2012). Commercial biomass syngas fermentation. *Energies*, 5(12):5372–5417.
- [Dash *et al.*, 2016] Dash, S., Ng, C. Y., and Maranas, C. D. (2016). Metabolic modeling of Clostridia: current developments and applications. *FEMS Microbiology Letters*, 363(4):fnw004.
- [Edwards and Palsson, 1999] Edwards, J. S. and Palsson, B. Ø. (1999). Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *Journal of Biological Chemistry*, 274(25):17410–17416.
- [Feist and Palsson, 2010] Feist, A. M. and Palsson, B. Ø. (2010). The biomass objective function. *Current Opinion in Microbiology*, 13(3):344–349.
- [Fell *et al.*, 2010] Fell, D. A., Poolman, M. G., and Gevorgyan, A. (2010). Building and analysing genome-scale metabolic models. *Biochemical Society Transactions*, 38(5):1197–1201.
- [Fell and Small, 1986] Fell, D. A. and Small, J. R. (1986). Fat synthesis in adipose tissue. *Biochemical Journal*.
- [Ferla and Patrick, 2014] Ferla, M. P. and Patrick, W. M. (2014). Bacterial methionine biosynthesis. *Microbiology*, 160(8):1571–1584.
- [Flamholz *et al.*, 2012] Flamholz, A., Noor, E., Bar-Even, A., and Milo, R. (2012). equilibrato—the biochemical thermodynamics calculator. *Nucleic Acids Research*, 40:D770–D775.
- [Foster and Moat, 1980] Foster, J. W. and Moat, A. G. (1980). Nicotinamide adenine dinucleotide biosynthesis and pyridine nucleotide cycle metabolism in microbial systems. *Microbiological reviews*, 44(1):83.
- [Gagneur and Klamt, 2004] Gagneur, J. and Klamt, S. (2004). Computation of elementary modes: a unifying framework and the new binary approach. *BMC bioinformatics*, 5(1):175.

- [Galperin and Koonin, 2010] Galperin, M. Y. and Koonin, E. V. (2010). From complete genome sequence to complete understanding? *Trends in biotechnology*, 28(8):398–406.
- [Gaston and Stadtman, 1963] Gaston, L. W. and Stadtman, E. (1963). Fermentation of ethylene glycol by *Clostridium glycolicum*, sp. n. *Journal of Bacteriology*, 85(2):356–362.
- [Gazzaniga *et al.*, 2009] Gazzaniga, F., Stebbins, R., Chang, S. Z., McPeck, M. A., and Brenner, C. (2009). Microbial nad metabolism: lessons from comparative genomics. *Microbiology and Molecular Biology Reviews*, 73(3):529–541.
- [Gevorgyan *et al.*, 2011] Gevorgyan, A., Bushell, M. E., Avignone-Rossa, C., and Kierzek, A. M. (2011). Surreyfba: a command line tool and graphics user interface for constraint-based modeling of genome-scale metabolic reaction networks. *Bioinformatics*, 27(3):433–434.
- [Gevorgyan *et al.*, 2008] Gevorgyan, A., Poolman, M. G., and Fell, D. A. (2008). Detection of stoichiometric inconsistencies in biomolecular models. *Bioinformatics*, 24(19):2245–2251.
- [Gibbs, 1879] Gibbs, J. W. (1879). On the equilibrium of heterogeneous substances. *Transactions of the Connecticut Academy of Arts and Sciences*.
- [Gottwald and Gottschalk, 1985] Gottwald, M. and Gottschalk, G. (1985). The internal pH of *Clostridium acetobutylicum* and its effect on the shift from acid to solvent formation. *Archives of microbiology*, 143(1):42–46.
- [Guo *et al.*, 2010] Guo, Y., Xu, J., Zhang, Y., Xu, H., Yuan, Z., and Li, D. (2010). Medium optimization for ethanol production with *Clostridium autoethanogenum* with carbon monoxide as sole carbon source. *Biore-source technology*, 101(22):8784–8789.
- [Harms *et al.*, 2002] Harms, P., Kostov, Y., and Rao, G. (2002). Bioprocess monitoring. *Current Opinion in Biotechnology*, 13(2):124 – 127.
- [Hartman *et al.*, 2014] Hartman, H. B., Fell, D. A., Rossell, S., Jensen, P. R., Woodward, M. J., Thorndahl, L., Jelsbak, L., Olsen, J. E., Raghunathan, A., Daefler, S., and Poolman, M. G. (2014). Identification of potential drug

targets in salmonella enterica sv. typhimurium using metabolic modelling and experimental validation. *Microbiology*, 160(6):1252–1266.

[Hauf *et al.*, 2000] Hauf, J., Zimmermann, F. K., and Müller, S. (2000). Simultaneous genomic overexpression of seven glycolytic enzymes in the yeast *Saccharomyces cerevisiae*. *Enzyme and Microbial Technology*, 26(9):688 – 698.

[Head and Gray, 2016] Head, I. M. and Gray, N. D. (2016). Microbial biotechnology 2020; microbiology of fossil fuel resources. *Microbial Biotechnology*, 9(5):626–634.

[Heirendt *et al.*, 2017] Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., Haraldsdottir, H. S., Keating, S. M., Vlasov, V., Wachowiak, J., *et al.* (2017). Creation and analysis of biochemical constraint-based models: the cobra toolbox v3. 0. *arXiv preprint arXiv:1710.04038*.

[Hemansi *et al.*, 2018] Hemansi, R. G., Yadav, G., Kumar, G., Yadav, A., Saini, J. K., and Kuhad, R. C. (2018). Second generation bioethanol production: The state of art. *Sustainable Approaches for Biofuels Production Technologies: From Current Status to Practical Implementation*, 7:121.

[Hempfling and Mainzer, 1975] Hempfling, W. P. and Mainzer, S. E. (1975). Effects of varying the carbon source limiting growth on yield and maintenance characteristics of *Escherichia coli* in continuous culture. *Journal of Bacteriology*, 123(3):1076–1087.

[Henry *et al.*, 2007] Henry, C. S., Broadbelt, L. J., and Hatzimanikatis, V. (2007). Thermodynamics-based metabolic flux analysis. *Biophysical journal*, 92(5):1792–1805.

[Henry *et al.*, 2006] Henry, C. S., Jankowski, M. D., Broadbelt, L. J., and Hatzimanikatis, V. (2006). Genome-scale thermodynamic analysis of *escherichia coli* metabolism. *Biophysical journal*, 90(4):1453–1461.

[Heppel *et al.*, 1959] Heppel, L. A., Strominger, J. L., and Maxwell, E. S. (1959). Nucleoside monophosphate kinases: Ii. transphosphorylation between adenosine monophosphate and nucleoside triphosphates. *Biochimica et Biophysica Acta*, 32:422–430.

- [Herrmann *et al.*, 2008] Herrmann, G., Jayamani, E., Mai, G., and Buckel, W. (2008). Energy conservation via electron-transferring flavoprotein in anaerobic bacteria. *Journal of Bacteriology*, 190(3):784–791.
- [Holzhütter, 2004] Holzhütter, H.-G. (2004). The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *The FEBS Journal*, 271(14):2905–2922.
- [Hsu *et al.*, 2014a] Hsu, P. D., Lander, E. S., and Zhang, F. (2014a). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157(6):1262 – 1278.
- [Hsu *et al.*, 2014b] Hsu, P. D., Lander, E. S., and Zhang, F. (2014b). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157(6):1262–1278.
- [Humphreys *et al.*, 2015] Humphreys, C. M., McLean, S., Schatschneider, S., Millat, T., Henstra, A. M., Annan, F. J., Breitskopf, R., Pander, B., Piatek, P., Rowe, P., *et al.* (2015). Whole genome sequence and manual annotation of *Clostridium autoethanogenum*, an industrially relevant bacterium. *BMC Genomics*, 16(1):1.
- [Isken and de Bont, 1998] Isken, S. and de Bont, J. A. (1998). Bacteria tolerant to organic solvents. *Extremophiles*, 2(3):229–238.
- [Jeffrey, 1989] Jeffrey, A. (1989). *Mathematics for Engineers and Scientists*. Chapman and Hall.
- [Johnson *et al.*, 1931] Johnson, M., Peterson, W., and Fred, E. (1931). Oxidation and reduction relations between substrate and products in the acetone-butyl alcohol fermentation. *Journal of Biological Chemistry*, 91(2):596–591.
- [Jones *et al.*, 2014] Jones, E., Oliphant, T., and Peterson, P. (2014). {SciPy}: open source scientific tools for {Python}.
- [Juminaga *et al.*, 2011] Juminaga, D., Baidoo, E. E., Redding-Johanson, A. M., Bath, T. S., Burd, H., Mukhopadhyay, A., Petzold, C. J., and Keasling, J. D. (2011). Modular engineering of L-tyrosine production in *Escherichia coli*. *Applied and environmental microbiology*, pages AEM-06017.

- [Kanemasa *et al.*, 1991] Kanemasa, Y., Nagamachi, E., and Shibuya, S. (1991). The presence of cis-9,10-methylene hexadecanoic acid in *Yersinia enterocolitica*. *Microbiology and Immunology*, 35:77–81.
- [Karp *et al.*, 2002] Karp, P. D., Paley, S., and Romero, P. (2002). The pathway tools software. *Bioinformatics*, 18(suppl 1):S225–S232.
- [Kelley and Sternberg, 2009] Kelley, L. A. and Sternberg, M. J. (2009). Protein structure prediction on the web: a case study using the Phyre server. *Nature protocols*, 4(3):363.
- [Klamt and Gilles, 2004] Klamt, S. and Gilles, E. D. (2004). Minimal cut sets in biochemical reaction networks. *Bioinformatics*, 20(2):226–234.
- [Köpke *et al.*, 2010] Köpke, M., Held, C., Hujer, S., Liesegang, H., Wiezer, A., Wollherr, A., Ehrenreich, A., Liebl, W., Gottschalk, G., and Dürre, P. (2010). *Clostridium ljungdahlii* represents a microbial production platform based on syngas. *Proceedings of the National Academy of Sciences*, 107(29):13087–13092.
- [Köpke *et al.*, 2011a] Köpke, M., Mihalcea, C., Bromley, J. C., and Simpson, S. D. (2011a). Fermentative production of ethanol from carbon monoxide. *Current Opinion in Biotechnology*, 22(3):320–325.
- [Köpke *et al.*, 2011b] Köpke, M., Mihalcea, C., Liew, F., Tizard, J. H., Ali, M. S., Conolly, J. J., Al-Sinawi, B., and Simpson, S. D. (2011b). 2, 3-butanediol production by acetogenic bacteria, an alternative route to chemical synthesis, using industrial waste gas. *Applied and Environmental Microbiology*, 77(15):5467–5475.
- [Köpke *et al.*, 2011c] Köpke, M., Mihalcea, C., Liew, F., Tizard, J. H., Ali, M. S., Conolly, J. J., Al-Sinawi, B., and Simpson, S. D. (2011c). 2,3-butanediol production by acetogenic bacteria, an alternative route to chemical synthesis, using industrial waste gas. *Applied and Environmental Microbiology*, 77(15):5467–5475.
- [Köpke *et al.*, 2013] Köpke, M., Nagaraju, S., and Chen, W. (2013). Recombinant microorganisms and methods of use thereof, patent number wo 2013/115659 a2. US Patent.

- [Lee *et al.*, 2008] Lee, J., Yun, H., Feist, A. M., Bernhard, P. Ø., and Lee, S. Y. (2008). Genome-scale reconstruction and *in silico* analysis of the *Clostridium acetobutylicum* ATCC 824 metabolic network. *Applied Microbiology and Biotechnology*.
- [Levin *et al.*, 2004] Levin, D. B., Pitt, L., and Love, M. (2004). Biohydrogen production: prospects and limitations to practical application. *International journal of hydrogen energy*, 29(2):173–185.
- [Lewis *et al.*, 2000] Lewis, T., Nichols, P. D., and McMeekin, T. A. (2000). Evaluation of extraction methods for recovery of fatty acids from lipid-producing microheterotrophs. *Journal of Microbiological Methods*, 43(2):107–116.
- [Li *et al.*, 2008] Li, F., Hinderberger, J., Seedorf, H., Zhang, J., Buckel, W., and Thauer, R. K. (2008). Coupled ferredoxin and crotonyl coenzyme a (coa) reduction with nadh catalyzed by the butyryl-coa dehydrogenase/etf complex from *clostridium kluyveri*. *Journal of Bacteriology*, 190(3):843–850.
- [Li *et al.*, 2014] Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J. S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157(3):624–635.
- [Liew *et al.*, 2017] Liew, F., Henstra, A. M., Köpke, M., Winzer, K., Simpson, S. D., and Minton, N. P. (2017). Metabolic engineering of *Clostridium autoethanogenum* for selective alcohol production. *Metabolic Engineering*, 40(Supplement C):104 – 114.
- [Ljungdahl, 1986] Ljungdahl, L. G. (1986). The autotrophic pathway of acetate synthesis in acetogenic bacteria. *Annual Review of Microbiology*, 40:415–450.
- [Ljungdahl and Wood, 1965] Ljungdahl, L. G. and Wood, H. G. (1965). Incorporation of c^{14} from carbon dioxide into sugar phosphates, carboxylic acids, and amino acids by *Clostridium thermoaceticum*. *Journal of Bacteriology*, 89(4):1055–1064.
- [Ljungdahl and Wood, 1969] Ljungdahl, L. G. and Wood, H. G. (1969). Total synthesis of acetate from CO_2 by heterotrophic bacteria. *Annual Reviews in Microbiology*, 23(1):515–538.

- [Marcellin *et al.*, 2016] Marcellin, E., Behrendorff, J. B., Nagaraju, S., DeTissera, S., Segovia, S., Palfreyman, R. W., Daniell, J., Licon-Cassani, C., Quek, L.-e., Speight, R., Hodson, M. P., Simpson, S. D., Mitchell, W. P., Köpke, M., and Nielsen, L. K. (2016). Low carbon fuels and commodity chemicals from waste gases – systematic approach to understand energy metabolism in a model acetogen. *Green Chemistry*, 18:3020–3028.
- [Martin *et al.*, 2016] Martin, M. E., Richter, H., Saha, S., and Angenent, L. T. (2016). Traits of selected *Clostridium* strains for syngas fermentation to ethanol. *Biotechnology and Bioengineering*, 113(3):531–539.
- [Martin *et al.*, 1971] Martin, R. G., Berberich, M. A., Ames, B. N., Davis, W. W., Goldberger, R. F., and Yournon, J. D. (1971). [147] enzymes and intermediates of histidine biosynthesis in *Salmonella typhimurium*. In *Methods in Enzymology*, volume 17, pages 3–44. Elsevier.
- [Masuko *et al.*, 2005] Masuko, T., Minami, A., Iwasaki, N., Majima, T., Nishimura, S.-I., and Lee, Y. C. (2005). Carbohydrate analysis by a phenol–sulfuric acid method in microplate format. *Analytical biochemistry*, 339(1):69–72.
- [Mihalcea *et al.*, 2008] Mihalcea, C. D., Fung, J. M. Y., Al-Sinawi, B., and Loan Phuong, T. (2008). Optimised fermentation media.
- [Milne *et al.*, 2011] Milne, C. B., Eddy, J. A., Raju, R., Ardekani, S., Kim, P.-J., Senger, R. S., Jin, Y.-S., Blaschek, H. P., and Price, N. D. (2011). Metabolic network reconstruction and genome-scale model of butanol-producing strain *Clostridium beijerinckii* NCIMB 8052. *BMC systems biology*, 5(1):130.
- [Mock *et al.*, 2015] Mock, J., Zheng, Y., Mueller, A. P., Ly, S., Tran, L., Segovia, S., Nagaraju, S., Köpke, M., Dürre, P., and Thauer, R. K. (2015). Energy conservation associated with ethanol formation from H₂ and CO₂ in *Clostridium autoethanogenum* involving electron bifurcation. *Journal of Bacteriology*, 197(18):2965–2980.
- [Müller *et al.*, 2014] Müller, S., Regensburger, G., and Steuer, R. (2014). Enzyme allocation problems in kinetic metabolic networks: optimal solutions are elementary flux modes. *Journal of theoretical biology*, 347:182–190.

- [Nagarajan *et al.*, 2013] Nagarajan, H., Sahin, M., Nogales, J., Latif, H., Lovley, D. R., Ebrahim, A., and Zengler, K. (2013). Characterizing acetogenic metabolism using a genome-scale metabolic reconstruction of *Clostridium ljungdahlii*. *Microbial Cell Factories*, 12(1):1.
- [Norman *et al.*, 2018] Norman, R. O., Millat, T., Winzer, K., Minton, N. P., and Hodgman, C. (2018). Progress towards platform chemical production using *Clostridium autoethanogenum*. *Biochemical Society Transactions*.
- [Oberhardt *et al.*, 2009] Oberhardt, M. A., Palsson, B. Ø., and Papin, J. A. (2009). Applications of genome-scale metabolic reconstructions. *Molecular Systems Biology*, 5(1).
- [Oh *et al.*, 2007] Oh, Y.-K., Palsson, B. Ø., Park, S. M., Schilling, C. H., and Mahadevan, R. (2007). Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *Journal of Biological Chemistry*, 282(39):28791–28799.
- [Pachauri *et al.*, 2014] Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., Church, J. A., Clarke, L., Dahe, Q., Dasgupta, P., *et al.* (2014). *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. IPCC.
- [Palsson, 2015] Palsson, B. (2015). *Systems biology*. Cambridge University Press.
- [Pander, 2018] Pander, B. (2018). *Nutrient limitation in Clostridium autoethanogenum and characterisation of its carbonic anhydrase*. PhD thesis, University of Nottingham.
- [Papoutsakis, 1984] Papoutsakis, E. T. (1984). Equations and calculations for fermentations of butyric acid bacteria. *Biotechnology and Bioengineering*, 26(2):174–187.
- [Pérès *et al.*, 2014] Pérès, S., Morterol, M., and Simon, L. (2014). Sat-based metabolics pathways analysis without compilation. In *International Conference on Computational Methods in Systems Biology*, pages 20–31. Springer.

- [Peres *et al.*, 2018] Peres, S., Schuster, S., and Dague, P. (2018). Thermodynamic constraints for identifying elementary flux modes. *Biochemical Society Transactions*.
- [Peters *et al.*, 2016] Peters, J. W., Miller, A.-F., Jones, A. K., King, P. W., and Adams, M. W. W. (2016). Electron bifurcation. *Curr. Opin. Chem. Biol.*, 31:146–152.
- [Phillips *et al.*, 1993] Phillips, J. R., Klasson, K. T., Clausen, E. C., and Gaddy, J. L. (1993). Biological production of ethanol from coal synthesis gas. *Applied Biochemistry and Biotechnology*.
- [Pimentel *et al.*, 2013] Pimentel, D., Hess, R., Diaz-Chavez, R., Ravindranath, R., and Cortez, L. B. (2013). *Food versus fuel: an informed introduction to biofuels*. Zed Books Ltd.
- [Poolman, 2006] Poolman, M. (2006). Scrumpy: metabolic modelling with python. *IEE Proceedings-Systems Biology*, 153(5):375–378.
- [Pramanik and Keasling, 1997] Pramanik, J. and Keasling, J. (1997). Stoichiometric model of *Escherichia coli* metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnology and bioengineering*, 56(4):398–421.
- [Pratt and Kolter, 1998] Pratt, L. A. and Kolter, R. (1998). Genetic analysis of *Escherichia coli* biofilm formation: roles of flagella, motility, chemotaxis and type I pili. *Molecular microbiology*, 30(2):285–293.
- [Ragsdale and Pierce, 2008] Ragsdale, S. W. and Pierce, E. (2008). Acetogenesis and the Wood-Ljungdahl pathway of CO₂ fixation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1784(12):1873 – 1898.
- [Ramió-Pujol *et al.*, 2014] Ramió-Pujol, S., Ganigué, R., Bañeras, L., and Colprim, J. (2014). Impact of formate on the growth and productivity of *Clostridium ljungdahlii* PETC and *Clostridium carboxidivorans* P7 grown on syngas. *International Microbiology*, 17:195–204.
- [Rossi *et al.*, 2011] Rossi, M., Amaretti, A., and Raimondi, S. (2011). Folate production by probiotic bacteria. *Nutrients*, 3(1):118–134.

- [Rydzak *et al.*, 2015] Rydzak, T., Lynd, L. R., and Guss, A. M. (2015). Elimination of formate production in *Clostridium thermocellum*. *Journal of industrial microbiology & biotechnology*, 42(9):1263–1272.
- [Sahm *et al.*, 1995] Sahm, H., Eggeling, L., Eikmanns, B., and Krämer, R. (1995). Metabolic design ain amino acid producing bacterium *Corynebacterium glutamicum*. *FEMS Microbiology Reviews*.
- [Saito *et al.*, 2008] Saito, T., Nishi, M., Lim, M. I., Wu, B., Maeda, T., Hashimoto, H., Takeuchi, T., Roos, D. S., and Asai, T. (2008). A novel GDP-dependent pyruvate kinase isozyme from *Toxoplasma gondii* localizes to both the apicoplast and the mitochondrion. *Journal of Biological Chemistry*.
- [Savinell and Palsson, 1992] Savinell, J. M. and Palsson, B. O. (1992). Network analysis of intermediary metabolism using linear optimization. i. development of mathematical formalism. *Journal of Theoretical Biology*, 154(4):421 – 454.
- [Schomburg *et al.*, 2002] Schomburg, I., Chang, A., and Schomburg, D. (2002). BRENDA, enzyme data and metabolic information. *Nucleic Acids Research*, 30(1):47–49.
- [Schuchmann and Müller, 2014] Schuchmann, K. and Müller, V. (2014). Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nature Reviews Microbiology*, 12(12):809–821.
- [Schuetz *et al.*, 2007] Schuetz, R., Kuepfer, L., and Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology*.
- [Schuster and Schuster, 1993] Schuster, R. and Schuster, S. (1993). Refined algorithm and computer program for calculating all nonnegative fluxes admissible in steady states of biochemical reaction systems with or without some flux rates fixed. *Bioinformatics*, 9(1):79–85.
- [Schuster *et al.*, 1999] Schuster, S., Dandekar, T., and Fell, D. A. (1999). Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends in biotechnology*, 17(2):53–60.

- [Schuster and Hilgetag, 1994] Schuster, S. and Hilgetag, C. (1994). On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems*, 2(02):165–182.
- [Sendin *et al.*, 2010] Sendin, J. O. H., Exler, O., and Banga, J. (2010). Multi-objective mixed integer strategy for the optimisation of biological networks. *IET Systems Biology*.
- [Senger and Papoutsakis, 2008a] Senger, R. S. and Papoutsakis, E. T. (2008a). Genome-scale model for *Clostridium acetobutylicum*: Part I. metabolic network resolution and analysis. *Biotechnology and Bioengineering*.
- [Senger and Papoutsakis, 2008b] Senger, R. S. and Papoutsakis, E. T. (2008b). Genome-scale model for *Clostridium acetobutylicum*: Part ii. development of specific proton flux states and numerically determined subsystems. *Biotechnology and bioengineering*, 101(5):1053–1071.
- [Sieminski *et al.*, 2014] Sieminski, A. *et al.* (2014). International energy outlook. *Energy Information Administration (EIA)*, 18.
- [Siitonen *et al.*, 2010] Siitonen, S., Tuomaala, M., and Ahtila, P. (2010). Variables affecting energy efficiency and CO₂ emissions in the steel industry. *Energy Policy*, 38(5):2477–2485.
- [Simpson *et al.*, 2009] Simpson, S. D., Forster, R. L., and Rowe, M. (2009). Microbial fermentation of gaseous substrates to produce alcohols. US Patent App. 12/296,300.
- [Sims *et al.*, 2008] Sims, R., Taylor, M., Saddler, J., and Mabee, W. (2008). From 1st-to 2nd-generation biofuel technologies. *Paris: International Energy Agency (IEA) and Organisation for Economic Co-Operation and Development*.
- [Skonieczny and Yargeau, 2009] Skonieczny, M. T. and Yargeau, V. (2009). Biohydrogen production from wastewater by *Clostridium beijerinckii*: effect of pH and substrate concentration. *International journal of hydrogen energy*, 34(8):3288–3294.
- [Stackebrandt *et al.*, 1999] Stackebrandt, E., Kramer, I., Swiderski, J., and Hippe, H. (1999). Phylogenetic basis for a taxonomic dissection of the

- genus *Clostridium*. *FEMS Immunology & Medical Microbiology*, 24(3):253–258.
- [Stephanopoulos *et al.*, 1998] Stephanopoulos, G., Aristidou, A. A., and Nielsen, J. (1998). *Metabolic engineering: principles and methodologies*. Academic press.
- [Strang, 1976] Strang, G. (1976). *Linear algebra and its applications* academic. New York, 19802.
- [Strominger *et al.*, 1959] Strominger, J. L., Heppel, L. A., and Maxwell, E. S. (1959). Nucleoside monophosphate kinases: I. transphosphorylation between adenosine triphosphate and nucleoside monophosphates. *Biochimica et biophysica acta*, 32:412–421.
- [Sund *et al.*, 2015] Sund, C. J., Liu, S., Germane, K. L., Servinsky, M. D., Gerlach, E. S., and Hurley, M. M. (2015). Phosphoketolase flux in *Clostridium acetobutylicum* during growth on L-arabinose. *Microbiology*, 161(2):430–440.
- [Taymaz-Nikerel *et al.*, 2010] Taymaz-Nikerel, H., Borujeni, A. E., Verheijen, P. J., Heijnen, J. J., and van Gulik, W. M. (2010). Genome-derived minimal metabolic models for *Escherichia coli* MG1655 with estimated in vivo respiratory ATP stoichiometry. *Biotechnology and Bioengineering*, 107(2):369–381.
- [Thauer *et al.*, 1977] Thauer, R. K., Jungermann, K., and Decker, K. (1977). Energy conservation in chemotrophic anaerobic bacteria. *Bacteriological reviews*, 41(1):100.
- [Thiele and Palsson, 2010] Thiele, I. and Palsson, B. O. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*.
- [Toro *et al.*, 2018] Toro, L., Pinilla, L., Avignone-Rossa, C., and Ríos-Estapa, R. (2018). An enhanced genome-scale metabolic reconstruction of *Streptomyces clavuligerus* identifies novel strain improvement strategies. *Bioprocess and biosystems engineering*, 41(5):657–669.

- [Torres and Eberhard, 2002] Torres, Néstor, V. and Eberhard, Voit, O. (2002). *Pathway Analysis and Optimization in Metabolic Engineering*. Cambridge University Press.
- [Tremblay *et al.*, 2013] Tremblay, P.-L., Zhang, T., Dar, S. A., Leang, C., and Lovley, D. R. (2013). The Rnf complex of *Clostridium ljungdahlii* is a proton-translocating ferredoxin:nad⁺ oxidoreductase essential for autotrophic growth. *mBio*, 4(1).
- [Trifunović *et al.*, 2016] Trifunović, D., Schuchmann, K., and Müller, V. (2016). Ethylene glycol metabolism in the acetogen *Acetobacterium woodii*. *Journal of Bacteriology*, pages JB-00942.
- [Valgepea *et al.*, 2017a] Valgepea, K., de SP Lemgruber, R., Meaghan, K., Palfreyman, R. W., Abdalla, T., Heijstra, B. D., Behrendorff, J. B., Tappel, R., Köpke, M., Simpson, S. D., Nielsen, L. K., and Marcellin, E. (2017a). Maintenance of ATP homeostasis triggers metabolic shifts in gas-fermenting Acetogens. *Cell Systems*, 4(5):505 – 515.e5.
- [Valgepea *et al.*, 2018] Valgepea, K., Lemgruber, R. S. P., Abdalla, T., Binos, S., Takemori, N., Takemori, A., Tanaka, Y., Tappel, R., Köpke, M., Simpson, S. D., *et al.* (2018). H₂ drives metabolic rearrangements in gas-fermenting *Clostridium autoethanogenum*. *Biotechnology for biofuels*, 11(1):55.
- [Valgepea *et al.*, 2017b] Valgepea, K., Loi, K. Q., Behrendorff, J. B., de SP Lemgruber, R., Plan, M., Hodson, M. P., Köpke, M., Nielsen, L. K., and Marcellin, E. (2017b). Arginine deiminase pathway provides ATP and boosts growth of the gas-fermenting acetogen *Clostridium autoethanogenum*. *Metabolic Engineering*, 41:202 – 211.
- [Varma and Palsson, 1993] Varma, A. and Palsson, B. O. (1993). Metabolic capabilities of escherichia coli: I. synthesis of biosynthetic precursors and cofactors. *Journal of theoretical biology*, 165(4):477–502.
- [Verduyn *et al.*, 1990] Verduyn, C., Postma, E., Scheffers, W. A., and Van Dijken, J. P. (1990). Energetics of *Saccharomyces cerevisiae* in anaerobic glucose-limited chemostat cultures. *Microbiology*, 136(3):405–412.
- [Wagner, 2004] Wagner, C. (2004). Nullspace approach to determine the elementary modes of chemical reaction systems. *The Journal of Physical Chemistry B*, 108(7):2425–2431.

- [Wang *et al.*, 2013] Wang, S., Huang, H., Kahnt, J., Mueller, A. P., Köpke, M., and Thauer, R. K. (2013). NADP-specific electron-bifurcating [FeFe]-Hydrogenase in a functional complex with Formate Dehydrogenase in *Clostridium autoethanogenum* grown on CO. *Journal of Bacteriology*, 195(19):4373–4386.
- [Wang *et al.*, 2010] Wang, S., Huang, H., Moll, J., and Thauer, R. K. (2010). NADP⁺ reduction with reduced ferredoxin and NADP⁺ reduction with NADH are coupled *via* an electron-bifurcating enzyme complex in *Clostridium kluyveri*. *Journal of Bacteriology*, 192(19):5115–5123.
- [Watson, 1984a] Watson, M. R. (1984a). Metabolic maps for the apple ii. *Biochemical Society Transactions*.
- [Watson, 1984b] Watson, M. R. (1984b). *Microcomputers in biochemical education*, chapter Use of the Apple II in teaching metabolism, pages 143e–158. Taylor & Francis. editor: Wood, Edward J.
- [Waygood *et al.*, 1975] Waygood, E. B., Rayman, M. K., and Sanwal, B. (1975). The control of pyruvate kinases of *Escherichia coli*. II. effectors and regulatory properties of the enzyme activated by ribose 5-phosphate. *Canadian Journal of Biochemistry*, 53(4):444–454.
- [Waygood and Sanwal, 1974] Waygood, E. B. and Sanwal, B. (1974). The control of pyruvate kinases of *Escherichia coli* I. physicochemical and regulatory properties of the enzyme activated by fructose 1, 6-diphosphate. *Journal of Biological Chemistry*, 249(1):265–274.
- [Worrell *et al.*, 2001] Worrell, E., Price, L., and Martin, N. (2001). Energy efficiency and carbon dioxide emissions reduction opportunities in the US iron and steel sector. *Energy*, 26(5):513–536.
- [Zaldivar and Ingram, 1999] Zaldivar, J. and Ingram, L. O. (1999). Effect of organic acids on the growth and fermentation of ethanologenic *Escherichia coli* ly01. *Biotechnology and Bioengineering*, 66(4):203–210.
- [Zhou *et al.*, 2013] Zhou, J., Olson, D. G., Argyros, D. A., Deng, Y., van Gulik, W. M., van Dijken, J. P., and Lynd, L. R. (2013). An atypical glycolysis in *Clostridium thermocellum*. *Applied and Environmental Microbiology*, pages AEM-04037.

[Zhu and Yang, 2004] Zhu, Y. and Yang, S.-T. (2004). Effect of pH on metabolic pathway shift in fermentation of xylose by *Clostridium tyrobutyricum*. *Journal of Biotechnology*, 110(2):143–157.

Chapter 9

APPENDICES

A1 MINIMAL GROWTH MEDIUM FOR *C. autoethanogenum*

See supplementary materials: 'Fermenter medium recipe for GASCHEM C auto.docx'.

A2 AST AMINO ACID SEQUENCE

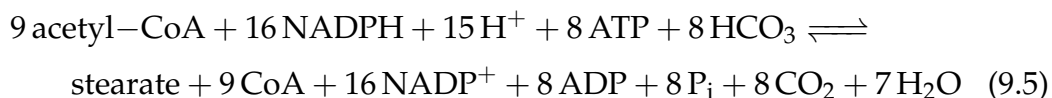
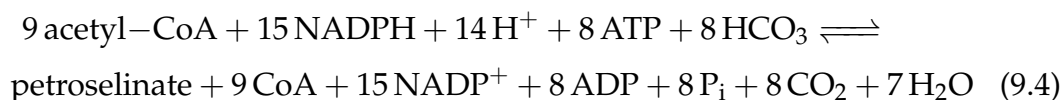
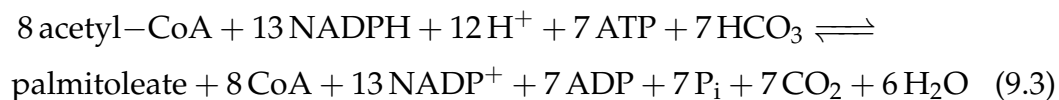
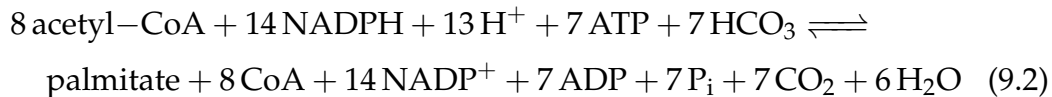
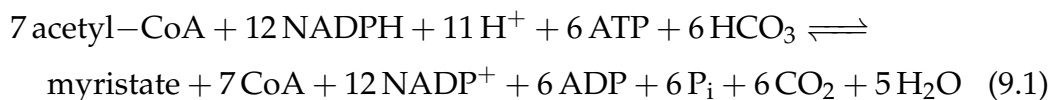
Amino acid sequence of homoserine acetyl transferase (AST) encoded by CLAU_0472 [Humphreys *et al.*, 2015]. The glutamate residue at position 111 (which indicates AST) is shown in bold font [Ferla and Patrick, 2014].

>CLAU_0472

MPINIPNDLPASKTLKAENIFVMNEKRALKQDIRPLNIVILNLMPVKTVT
EVQLLRLLSNSPIQVDITLMYPKSHRSKNTSEQYLLKYYETFDEIKGRKFD
GMIITGAPVEKCLKFEDVDYWDELTEIMDWSKKNVYSTLHICWGSQAGL
YHHFGIPKYILDKKMFGVFAHIINEENTQLLRGFDDIFYAPHSRHTCVKK
EDIEKVPPEILSESKDAGVYIVSTKGGRQIFVTGHPEYDTLTLKSEYDRD
VAANLPIEIPKNYFPHNDTAKQPRATWKSHANLLFLNWLNYVYQETP
YNLDSLQ

A3 BIOMASS COMPOSITION

Lipid biosynthesis reactions. Lumped reactions were used to represent the biosynthesis of the lipids reported in [Abrini *et al.*, 1994] (Eqs. 9.1 to 9.5).



Total lipid measurement. In order to obtain the total cellular lipid content, a biphasic fractional separation reaction was used. 20 ml samples of fermenter culture were extracted and centrifuged at $15000 \times g$ for 15 minutes before washing in 100 ml of 1% NaCl. The supernatant was discarded and the sample freeze dried for 20 hours. Solvents were added to the sample in order of increasing polarity to give a chloroform:methanol:water ratio of 1:2:0.8. The subsequent cell suspension was washed into a separatory funnel and left overnight. *Phase separation:* The desired solvent mixture occurred when the solvent ratio was changed to 1:1:0.9 (chloroform:methanol:water). The lower lipid-containing phase was separated off and the chloroform evap-

orated using a rotary evaporator. The weight of the remaining lipids was measured using an analytical balance [Axelsson and Gentili, 2014; Lewis *et al.*, 2000]. Proportions of lipids constituting the total lipid mass were taken from [Abrini *et al.*, 1994].

Total gDNA measurement. Phenol chloroform extraction was performed to measure the total gDNA of *C. autoethanogenum*. Samples of *C. autoethanogenum* were cultured in 5 ml of YTF (Yeast extract, Tryptone, Fructose) media for 3 days before centrifugation at $6000 \times g$ for 2 min. The cell pellet was resuspended in 180 μ l of lysozyme digestion buffer (25 mM Tris-HCl, pH 8.0, 2.5 mM EDTA, 1% Triton X-100), containing 20 mg/ml of lysozyme VI (MP Biomedicals), for 30 min at 37 °C. Following incubation, 25 μ l Proteinase K (Sigma Aldrich), 85 μ l of dH₂O and 110 μ l of 10% SDS (Sigma Aldrich) were added to the lysed samples and mixed by inversion, and incubated at 65 °C for 30 min. After incubation, 400 μ l of phenol chloroform (or equal volumes to sample volume), were pipetted into 2 ml Manual Phase Lock Gel™ Heavy tubes (5 Prime, UK), the tube contents were inverted for approx. 15 sec before centrifugation for 3 min at maximum speed. The clear top organic layer of the sample was transferred into a fresh 1.5 mL Eppendorf tube and 5 μ l of RNase A (Sigma Aldrich) was added and the sample incubated at room temperature for 10 min. The sample was placed into a fresh Phase Lock tube containing 400 μ l of phenol chloroform and mixed by inversion, then centrifuged at maximum speed for 3 min. This was repeated twice and after the final chloroform wash, the top layer was placed into a 1.5 ml tube containing pre-cooled (-80 °C) sodium acetate and ethanol, and mixed by inversion. The sample was placed back into the -80 °C freezer for 30 min. For ethanol precipitation, samples were thawed to room temperature, centrifuged at maximum speed for 15 min, and supernatant removed leaving a visible pellet. 1 ml of 70 % ethanol was added to the sample and mixed by inversion, then centrifuged at maximum speed for 3 min. The samples were air dried at room temperature for 45 min, and resuspended in 50 μ l of sterile, nuclease-free dH₂O.

Total RNA measurement. RNA was isolated using the following kits following manufacturers recommendations; FastRNA PRO Blue (MP Biomedicals) for cell disruption in a FastPrep 24 homogenizer (MP Biomedicals); DirectZol RNA miniprep (Zymo Research Products) for RNA isolation; Ambion Turbo DNase (Thermo-Fisher Scientific) for DNA removal; RNA clean

& concentrator (Zymo Research Products) for RNA clean up. An on column DNA digestion at room temperature (RT) and a post column DNA digestion at 37 °C for 30 minutes were performed. Only small batches of up to 4 cell pellets were processed through the homogenisation steps at a time. Presence of remaining DNA was tested by PCR with primers F and R. Quality and quantity of the obtained RNA was assessed by Nanodrop (Thermo Scientific), Qubit (Thermo-Fisher Scientific) and RNA-nano gel-electrophoresis (2100 Bioanalyser, Agilent).

Total Protein measurement. Cells pellets were lysed using Bugbuster(EMDmillipore) with lysozyme. Protein content was determined by using a Bradford assay (Sigma, Bradford ref) [Bradford, 1976].

Total Polysaccharide measurement. Polysaccharides were measured according to (masuko). Fructose, Xylose, Glucose and Rhamnose were used as references and whole cell suspension used for measurements [Masuko *et al.*, 2005]. The carbohydrate content was determined using a quick method that is not optimised for total carbohydrate content of cells and is somewhat sensitive to carbohydrate composition which was not determined.

Total Teichoic acid measurement. For analysis of the teichoic acid content of *C. autoethanogenum*, the method from Archibald [Archibald *et al.*, 1968b] was used. For this, 250 ml of *C. autoethanogenum* culture were centrifuged at 4000 rpm for 15 min at 4 °C. The gained cells were then lysed using 5 ml of 1 x BugBuster solution and 1 mg lysozyme. The mix was left shaking at RT for 20 min at 25 rpm. Afterwards, the lysed cells were centrifuged at 4 °C, 7000 rpm for 20 min, the gained pellet was washed with 100 mM sodium phosphate buffer (pH 7) and freeze dried over night. This resulted in 186 mg of cell wall material. The material was dissolved in 20 ml of 10 % (w/v) trichloroacetic acid and stirred at 4 °C for 24 h. The mix was afterwards centrifuged at 7000 rpm, 4 °C for 15 min and the supernatant immediately mixed with 20 ml ice-cold ethanol. The resulting mixture was kept at 4 °C until later use. Further 20 ml of 10 % trichloroacetic acid were used to dissolve the pellet and stirred at 4 °C for 48 h. Then, the mix was centrifuged again at 7000 rpm and 4 °C for 20 min and the resulting supernatant added to 20 ml ice cold ethanol. Both supernatants mixes were joined together and their precipitants dissolved in 5 ml of 10 % trichloroacetic acid. The insoluble material was removed *via* centrifugation and 20 ml ice cold ethanol added to the supernatant. This mix was then centrifuged again for a last time for

Table 9.1: Estimated amino acid proportions.

amino acid	mmol gDCW ⁻¹
Alanine	0.203
Arginine	0.088
Asparagine	0.099
Aspartate	0.097
Cysteine	0.024
Glutamate	0.131
Glutamine	0.045
Glycine	0.260
Histidine	0.026
Isoleucine	0.159
Leucine	0.150
Lysine	0.216
Methionine	0.042
Phenylalanine	0.043
Proline	0.071
Serine	0.126
Threonine	0.102
Tryptophan	0.006
Tyrosine	0.047
Valine	0.185

15 min at 4 °C and 7000 rpm. The teichoic acid precipitated as a fine powder which was weighted. Out of 186 mg of cell wall material, 1.4 mg of teichoic acid could be isolated.

A4 ATP MAINTENANCE

The following experimental procedure was conducted to measure growth and non-growth ATP maintenance costs (GAM and NAGM, respectively) in CO-fed continuous culture of *C. autoethanogenum*: The set dilution rate was increased incrementally from 0.01 to 0.028 (hour⁻¹), covering the following range of values: 0.01, 0.012, 0.0125, 0.015, 0.0183, 0.0188, 0.027, 0.028. Steady state was reached after each change, as determined by stable culture optical density. For each steady-state period, the average CO uptake rate was calculated and plotted against the corresponding specific growth rate, which was taken as equivalent to the set dilution rate according to Eq. 9.6 (where [OD] denotes the optical density of the culture). The slope of a linear fit to this data is the growth associated CO uptake rate, the y-intercept is the

Table 9.2: Estimated DNA and RNA nucleotide proportions.

nucleotide	mmol gDCW ⁻¹
DATP	0.103
TTP	0.105
DGTP	0.045
DCTP	0.049
ATP	0.104
UTP	0.074
GTP	0.104
CTP	0.078

non-growth associated CO uptake rate (see Fig. 9.1).

$$\frac{d[\text{OD}]}{dt} = [\text{OD}](\mu - D) = 0 \quad (9.6)$$

(D = dilution rate, μ = specific growth rate.) ATP costs were calculated by multiplying growth and non-growth associated CO uptake rates by the model-predicted maximal ATP yield (Y_{ATP}) from CO. A proportion of the growth associated cost (GAM) is accounted for by biomass precursor production in the metabolic network. This proportion must be subtracted from the total GAM before incorporation as a flux constraint.

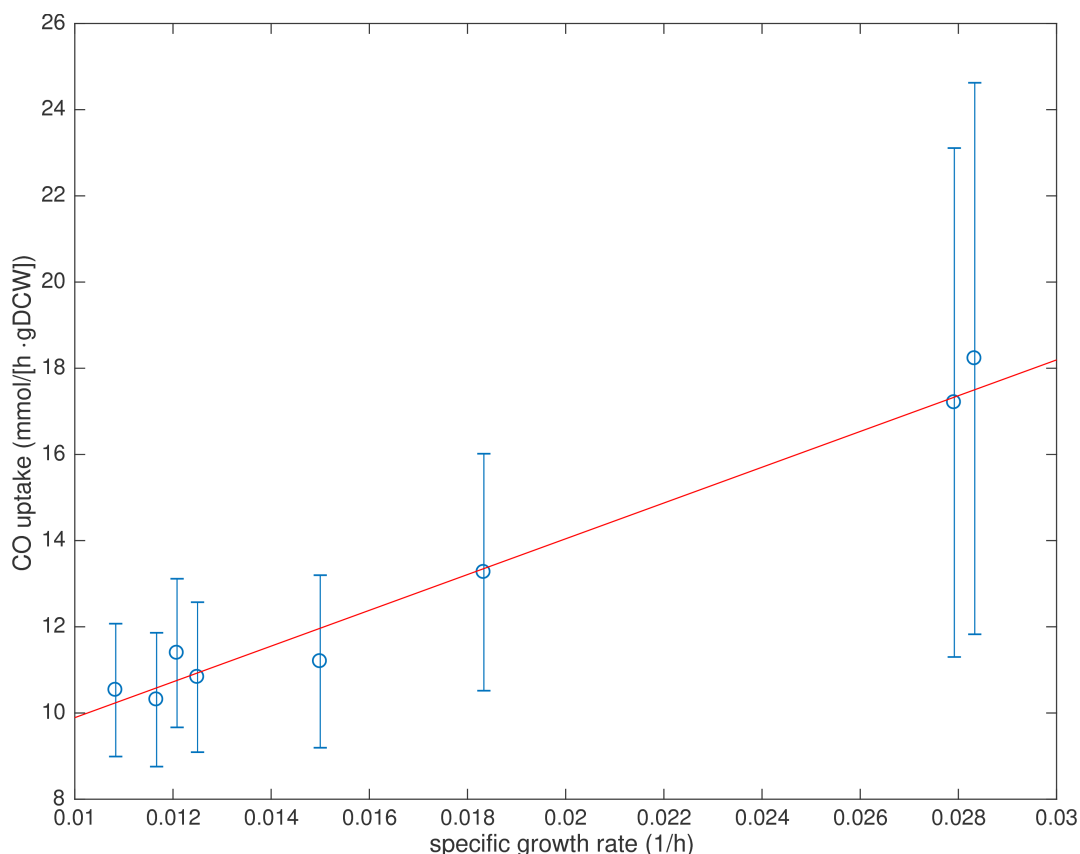


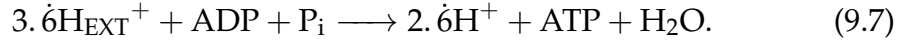
Figure 9.1: Specific growth rate of *C. autoethanogenum* continuous culture *vs.* CO uptake rate. Gradient and intercept of trend line represent growth associated and non-growth associated CO uptake parameters, respectively. Gradient = 415 ± 92 mmol gDCW⁻¹ ($p = 3.3 \times 10^{-5}$), intercept = 5.7 ± 1.4 mmol gDCW⁻¹ h⁻¹ ($p = 5.5 \times 10^{-5}$), $R^2 = 0.957$, error bars show the relative standard deviation (RSD).

A5 $\Delta_r G'^{\circ}$ OF ATP SYNTHASE AND RNF

eQuilibrator does not provide standard Gibbs energy of reaction values ($\Delta_r G'^{\circ}$) for reactions coupled to the transport of compounds between biological compartments. The standard Gibbs energy of reaction for F₁F₀-ATP synthase ($\Delta_r G'_{\text{ATPsyn}}^{\circ}$) and Rnf ($\Delta_r G'_{\text{ATPsyn}}^{\circ}$) were estimated using the method described in [Henry *et al.*, 2006] and published experimental data for the closely related *C. ljungdahlii* (grown autotrophically) [Tremblay *et al.*, 2013].

A5.1 ATP SYNTHASE

As stated in Chapter 4, the stoichiometry of the reaction catalysed by F₁F₀-ATP synthase is



This reaction can be split into two processes: 1) the cytoplasmic phosphorylation of ADP, and 2) the translocation of protons from the extracellular environment into the cytoplasm. *I.e.*,



These two processes each have an associated Gibbs energy change ($\Delta_r G'^{\circ}_{\text{cyt.}}$ and $\Delta_r G'^{\circ}_{\text{trans.}}$, respectively). As stated in [Henry *et al.*, 2006], it follows from this that

$$\Delta_r G'^{\circ}_{\text{ATPsyn}} = \Delta_r G'^{\circ}_{\text{cyt.}} + \Delta_r G'^{\circ}_{\text{trans.}}, \quad (9.10)$$

where

$$\Delta_r G'^{\circ}_{\text{trans.}} = \Delta_{\Delta\Psi} G'^{\circ} + \Delta_{\Delta\text{pH}} G'^{\circ}. \quad (9.11)$$

$\Delta_{\Delta\Psi} G'^{\circ}$ and $\Delta_{\Delta\text{pH}} G'^{\circ}$ respectively represent the Gibbs energy change associated with the ionic potential ($\Delta\Psi$ [mV]) and pH potential (ΔpH) across the cellular membrane. Equations for calculating $\Delta_{\Delta\Psi} G'^{\circ}$ and $\Delta_{\Delta\text{pH}} G'^{\circ}$ from $\Delta\Psi$ and ΔpH are given below:

$$\Delta_{\Delta\Psi} G'^{\circ} = nF\Delta\Psi \quad (9.12)$$

$$\Delta_{\Delta\text{pH}} G'^{\circ} = -2.3hRT\Delta\text{pH}, \quad (9.13)$$

where n is the net charge transported across the membrane (+3.66), F is the Faraday constant (0.0965 kJ/mV mol), h is the number of protons being transported across the membrane (3.66), R is the universal gas constant (0.008314 kJ/mol) and T is the temperature (298 K) [Henry *et al.*, 2006]. Average experimentally measured values for $\Delta\Psi$ and ΔpH were taken from a study of the closely related organism *C. ljungdahlii* [Tremblay *et al.*, 2013]:

$$\Delta\Psi = ((-72.7) + (-53.6) + (-58.2))/3 = -62 \pm 8 \text{ (mV)} \quad (9.14)$$

$$\Delta\text{pH} = (0.29 + 0.29 + 0.46)/3 = 0.3 \pm 0.1, \quad (9.15)$$

(Error value is standard deviation from the mean.) The resulting value for $\Delta_r G'_{\text{ATPsyn}}$ is $14 \pm 6 \text{ kJ/mol}$. Workings below:

$$\Delta_{\Delta\Psi} G'^{\circ} = nF\Delta\Psi = 3.6 \times 0.0965 \times -62 = -21.767... \quad (9.16)$$

$$\Delta_{\Delta\text{pH}} G'^{\circ} = -2.3hRT\Delta\text{pH} = -2.3 \times 3.6 \times 0.008314 \times 298 \times 0.3 = -6.128..., \quad (9.17)$$

$$\Delta_r G'_{\text{trans.}} = \Delta_{\Delta\Psi} G'^{\circ} + \Delta_{\Delta\text{pH}} G'^{\circ} = -21.767... + (-6.128...) = -27.886... \quad (9.18)$$

$$\Delta_r G'_{\text{ATPsyn}} = \Delta_r G'_{\text{cyt.}} + \Delta_r G'_{\text{trans.}} = 41.14 + (-27.886...) = 13.513... \quad (9.19)$$

To account for error in the experimental data, Eq. 9.10 is re-written with error terms (ξ):

$$\Delta_r G'_{\text{ATPsyn}} \pm \xi_{\Delta_r G'_{\text{ATPsyn}}} = \Delta_r G'_{\text{trans.}} \pm \xi_{\Delta_r G'_{\text{trans.}}} + \Delta_r G'_{\text{cyt.}} \pm \xi_{\Delta_r G'_{\text{cyt.}}}. \quad (9.20)$$

Thus,

$$\xi_{\Delta_r G'_{\text{ATPsyn}}} = \xi_{\Delta_r G'_{\text{trans.}}} + \xi_{\Delta_r G'_{\text{cyt.}}}. \quad (9.21)$$

$\xi_{\Delta_r G'_{\text{cyt.}}}$ has a value of 0.6 kJ/mol, which taken from eQuilibrator as the error in $\Delta_r G'^{\circ}$ for ADP phosphorylation at pH 6. $\xi_{\Delta_r G'_{\text{trans.}}}$ is calculated by re-writing Eq. 9.11 with error terms:

$$\Delta_r G'_{\text{trans.}} \pm \xi_{\Delta_r G'_{\text{trans.}}} = \Delta_{\Delta\Psi} G'^{\circ} \pm \xi_{\Delta_{\Delta\Psi} G'^{\circ}} + \Delta_{\Delta\text{pH}} G'^{\circ} \pm \xi_{\Delta_{\Delta\text{pH}} G'^{\circ}}. \quad (9.22)$$

Thus,

$$\zeta_{\Delta_r G'_{\text{trans.}}} = \zeta_{\Delta_{\Delta\Psi} G'^{\circ}} + \zeta_{\Delta_{\Delta\text{pH}} G'^{\circ}}, \quad (9.23)$$

and since n, h, F, R and T are theoretical constants and have no error value,

$$\Delta_{\Delta\Psi} G'^{\circ} \pm \zeta_{\Delta_{\Delta\Psi} G'^{\circ}} = nF \times \Delta\Psi \pm \zeta_{\Delta\Psi} \quad (9.24)$$

$$\Delta_{\Delta\text{pH}} G'^{\circ} \pm \zeta_{\Delta_{\Delta\text{pH}} G'^{\circ}} = -2.3hRT \times \Delta\text{pH} \pm \zeta_{\Delta\text{pH}}, \quad (9.25)$$

and

$$\frac{\zeta_{\Delta_{\Delta\Psi} G'^{\circ}}}{\Delta_{\Delta\Psi} G'^{\circ}} = \frac{\zeta_{\Delta\Psi}}{\Delta\Psi} \quad (9.26)$$

$$\frac{\zeta_{\Delta_{\Delta\text{pH}} G'^{\circ}}}{\Delta_{\Delta\text{pH}} G'^{\circ}} = \frac{\zeta_{\Delta\text{pH}}}{\Delta\text{pH}}. \quad (9.27)$$

Finally,

$$\zeta_{\Delta_r G'_{\text{trans.}}} = \frac{\zeta_{\Delta\text{pH}}}{\Delta\text{pH}} \Delta_{\Delta\text{pH}} G'^{\circ} + \frac{\zeta_{\Delta\Psi}}{\Delta\Psi} \Delta_{\Delta\Psi} G'^{\circ} \quad (9.28)$$

$$= \left| \frac{0.1}{0.3} \times -6.128... \right| + \left| \frac{8.13}{-61.5} \times -21.767... \right| = 4.920... \text{ kJ/mol}, \quad (9.29)$$

so that

$$\zeta_{\Delta_r G'_{\text{ATPsyn}}} = 0.6 + 4.920... = 5.520... \quad (9.30)$$

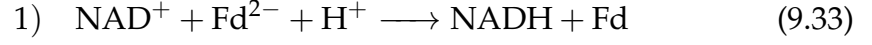
$$\Delta_r G'_{\text{ATPsyn}} \pm \zeta_{\Delta_r G'_{\text{ATPsyn}}} = \underline{14 \pm 6 \text{ kJ/mol.}} \quad (9.31)$$

A5.2 RNF

As stated in Chapter 4, the stoichiometry of the reaction catalysed by Rnf is



This reaction can be split into two processes: 1) the cytoplasmic reduction of NAD^+ with reduced ferredoxin, and 2) the translocation of two protons from the cytoplasm to the extracellular environment. *I.e.*,



The process of calculating $\Delta_r G'_{\text{Rnf}}^{\circ}$ is analogous to ATP synthase.

$$\Delta_r G'_{\text{Rnf}}^{\circ} \pm \xi_{\Delta_r G'_{\text{Rnf}}^{\circ}} = \Delta_r G'_{\text{trans.}}^{\circ} \pm \xi_{\Delta_r G'_{\text{trans.}}^{\circ}} + \Delta_r G'_{\text{cyt.}}^{\circ} \pm \xi_{\Delta_r G'_{\text{cyt.}}^{\circ}}. \quad (9.35)$$

However, in this case $\Delta_r G'_{\text{cyt.}}^{\circ} \pm \xi_{\Delta_r G'_{\text{cyt.}}^{\circ}} = 22.3 \pm 11.7\text{mV}$ as sourced from eQuilibrator [Flamholz *et al.*, 2012]. The remaining calculation of $\Delta_r G'_{\text{trans.}}^{\circ} \pm \xi_{\Delta_r G'_{\text{trans.}}^{\circ}}$ is exactly as with ATP synthase, except the parameters n and h are now 2.

$$\xi_{\Delta_r G'_{\text{trans.}}^{\circ}} = \frac{\xi_{\Delta\text{pH}}}{\Delta\text{pH}} \Delta_{\Delta\text{pH}} G'^{\circ} + \frac{\xi_{\Delta\Psi}}{\Delta\Psi} \Delta_{\Delta\Psi} G'^{\circ} \quad (9.36)$$

$$= \left| \frac{0.1}{0.3} \times -3.343... \right| + \left| \frac{8.13}{-61.5} \times -11.867... \right| = 2.683... \text{ kJ/mol}, \quad (9.37)$$

The resulting value for $\Delta_r G'_{\text{trans.}}^{\circ} \pm \xi_{\Delta_r G'_{\text{trans.}}^{\circ}} = -15.210... \pm 2.683... \text{ kJ/mol}$, and thus $\Delta_r G'_{\text{Rnf}}^{\circ} \pm \xi_{\Delta_r G'_{\text{Rnf}}^{\circ}} = -26 \pm 15 \text{ kJ/mol}$. This value is close to the value estimated in [Schuchmann and Müller, 2014] (-25 kJ/mol).