
Structural Metabolic Modelling of the Acetogen *Acetobacterium woodii*

Noah Adam Mesfin

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

This project was funded by a BBSRC iCASE studentship
January 2020

Acknowledgements

My time here at Oxford Brookes in the Cell Systems Modelling Group has been incredibly gratifying and enriching. In particular, my supervisors Prof. David Fell and Dr. Mark Poolman have been unfaltering sources of an expansive knowledge, encouragement and rigorous critical thinking throughout the project, for which I will always be thankful.

I would also like to thank the diligent people at ZuvaSyntha and the Technical University of Denmark, where I went on secondments, for making those fruitful periods of collaborative research and being instructive in the laboratory for practices in anaerobiosis. I especially thank my co-supervisor Dr. Sheila Ingemann Jensen and past co-supervisors Dr. Annette Alcabas and Dr. Michelle Gradley.

I am thankful to the past and present members of our group. To Dr. Hassan Hartman, Dr. Dipali Singh, Dr Kailash Adhikari, Elzbieta Al Saidi and Yanica Said thank you for being great company in the lab and on our travels. To the past members I've met in passing or by proxy, I would also like to thank you because much of the work carried out here is built on your foundations. Thank you also to the visiting members to our group, especially Dr. Rupert Norman, Dr. Nicole Pearcy and Maite Diaz-Calvo.

I would like to thank the staff in the Faculty of Life Sciences for their administrative support, in particular Catherine Joyejob, Jill Organ, Philip Voysey, Prof. Susan Brooks, Dr. David Evans and Dr Helen Packer.

The completion of this PhD has been, for me, the fulfilment of a plan that had its inception in Turku, Finland during an Erasmus placement in 2010, so I would like to acknowledge formative people throughout this period. I thank Prof. Natalia Battchikova, Prof. Michael Bushell, Prof. Nicolas Locker, Dr. Andrez Kierzek, Dr. Jose Jimenez, Dr. Andrea Rocco and Dr. Emma Laing.

A special thanks also goes to the researchers in the CCNet (formerly C1Net) community, especially Prof. Volker Müller and Prof. Klaus Winzer for interesting discussions and data supplied before publication that helped shape some of the work contained here.

I would be remiss to not mention that all the work carried out throughout this project including the preparation of this manuscript was done using free and open source software, namely $\text{\LaTeX}$, Gnuplot, GNU and Linux - it has been a joy working this way.

To my friends and family who have sustained me in the most important ways, I cannot

thank you all enough. To my brothers Alan and Meron, my cousin Luam, and to Haruka Sophia and Oliver, thanks for being there.

Finally, my deepest gratitude goes to my parents, Bisrat and Dawit, who were steadfast in their support. Thank you for your never ending encouragement and for helping provide a space for me to think and to grow.

Abstract

Acetogens are microbes that produce acetate as a fermentation by-product from a diverse range of substrates including gaseous substrates. They achieve this by employing the Wood-Ljungdahl Pathway (WLP), which enables the fixation of carbon dioxide into cellular material for growth and to generate energy via a non-photosynthetic route. Electrons for this process are derived from diverse substrates that include molecular hydrogen and carbon monoxide. The ability of acetogens to utilise synthesis gas (H_2 , CO, CO_2) make them an attractive target for industrial biotechnology since these are potentially renewable feedstocks. A model acetogen is *Acetobacterium woodii*, but as of yet no computational analysis has been carried out on it.

Here we construct metabolic models, including a Genome-scale Metabolic Model (GSM) of *Acetobacterium woodii* using the sequenced and annotated genome of strain DSM1030. The GSM consists of 867 metabolites, 841 reactions and 98 transporters and is predictive of growth from the diverse substrates utilised by *A. woodii*. It is validated against experimental data produced here and that reported in the scientific literature. To analyse the impact of producing heterologous compounds with *A. woodii*, a core model was constructed from the GSM and then used in conjunction with heterologous synthesis pathways for elementary modes analysis. The GSM was also used to analyse how growth would be affected by heterologous compound production. For this effort, a combination of linear programs were written that consistently calculated a $Y_{\text{ATP}}^{\text{max}}$. It was found that no single LP could consistently calculate a $Y_{\text{ATP}}^{\text{max}}$. ATP calculations were also used to compare *A. woodii* maintenance costs across several conditions.

Elementary modes analysis was carried out on a smaller WLP-only model, the results of which were used to interpret seemingly inconsistent reports on CO-dependent growth of *A. woodii*. Additionally it is highlighted that an as of yet unknown regulatory framework is likely biasing acetate production in conditions that could otherwise produce other compounds. Novel biosynthesis routes for commodity chemicals are also reported here and analysed for their impact on *A. woodii* metabolism along with other synthetic routes reported in the scientific literature and patents. Based on modelling results, suggestions are made for metabolically engineering *A. woodii* for applications in industrial biotechnology.

Contents

1	Introduction	2
1.1	Project Background	2
1.2	Bacterial Metabolism	3
1.2.1	Anabolism	4
1.2.2	Catabolism	4
1.3	Acetogenic Bacteria	7
1.3.1	<i>Acetobacterium woodii</i> as a Model Acetogen	7
1.4	Theoretical Underpinnings	8
1.4.1	Mathematical Description of Metabolic Networks	8
1.4.2	Metabolism at Steady State and the Nullspace	11
1.4.3	Elementary Modes Analysis	13
1.4.4	Linear Programming	15
1.4.5	Genome-scale Metabolic Models	18
1.5	ScrumPy	18
1.6	Aims and Thesis Outline	22
2	Computational Methods	24
2.1	Building and Analysing Models	24
2.1.1	Databases	24
2.1.2	Model System Definition	26
2.1.3	Model Review	29
2.1.4	Model Validation	31
2.2	Model Analysis	32
3	A Metabolic Model of the Wood-ljungdahl Pathway	34
3.1	Introduction	34
3.2	Model Construction and Curation	35
3.2.1	Construction	35
3.2.2	Review and Validation	36
3.2.3	Model Properties	36

3.3	Elementary Modes Analysis	38
3.3.1	Elementary modes consuming CO ₂ -H ₂	39
3.3.2	Elementary modes consuming CO	41
3.3.3	Elementary modes consuming CO ₂ -CO-H ₂	43
3.3.4	Elementary modes consuming CO-H ₂	43
3.3.5	Formate, fructose and methanol metabolism (with CO)	47
3.4	Discussion	49
3.4.1	Introduction	49
3.4.2	CO-inhibited reactions in modes of the extended WLP model	49
3.4.3	CO production and re-utilisation	56
3.4.4	Biphasic growth patterns on CO ₂ -CO-H ₂	58
3.4.5	Potential novel metabolism	62
3.4.6	Conclusions	63
4	Genome-Scale Metabolic Model Construction and Overview	64
4.1	Introduction	64
4.2	Curation	65
4.2.1	Lysine Biosynthesis	65
4.2.2	Vitamin B12 - Cobalamin Biosynthesis	66
4.2.3	Methanol Catabolism	71
4.2.4	Central Metabolic Reactions	74
4.3	Model Overview	79
4.3.1	General Metrics	80
4.4	Discussion	82
4.4.1	Overview	82
5	Bioenergetics of <i>A.woodii</i> Metabolism	86
5.1	Introduction	86
5.2	Y_{ATP}^{MAX} Calculations	86
5.2.1	Y_{ATP}^{max} using four LPs and substrate combinations	88
5.2.2	ATP yields table	92
5.3	Calculating maintenance costs	93
5.3.1	Maintenance costs with five growth substrates	94
5.4	Bioenergetic Core	95
5.4.1	Methodolgy	95
5.4.2	Bioenergetic core metrics and properties	98
5.4.3	Reaction essentiality	100
5.5	Using alternative electron acceptors changes ATP yields	102
5.6	Discussion	104

6	Experimental Growth of <i>A.woodii</i>	107
6.1	Chapter Introduction	107
6.2	Aims	107
6.3	Methods	107
6.3.1	Laboratory data analysis	107
6.3.2	Modelling data generation	111
6.3.3	Carbon mass balance	112
6.4	Laboratory Results	112
6.4.1	Fructose batch growth	112
6.4.2	Using nitrite as nitrogen source	113
6.5	Comparing <i>in vivo</i> observations to modelling predictions	115
6.5.1	Comparing the batch fructose fermentation to model predictions	115
6.5.2	Comparing model predictions with different maintenance costs	117
6.5.3	Model analysis of ethanol-nitrite based growth	119
6.5.4	Comparing other reported data to modelling predictions	120
6.6	Chapter Discussion	120
6.6.1	Fructose metabolism in experiments and modelling	120
6.6.2	Alternative nitrogen sources in acetogens	122
7	Producing Value-Added Compounds	124
7.1	Introduction	124
7.2	Method	124
7.2.1	Construction of heterologous pathways	124
7.2.2	Construction and analysis of <i>in silico</i> engineered <i>A.woodii</i> models	125
7.2.3	Growth analysis of <i>in silico</i> engineered <i>A.woodii</i>	126
7.2.4	Yield calculations	127
7.2.5	Metabolite turnover rate calculations	127
7.2.6	ATP scan	127
7.3	Results - producing higher-value added compounds	128
7.3.1	3-Hydroxypropionate production from CO ₂ -H ₂	128
7.3.2	A survey of higher-value added chemical production with single carbon substrates	132
7.3.3	Growth analysis of <i>A.woodii</i> using the genome-scale metabolic model	138
7.4	Discussion	141
8	Conclusions and Outlook	145
8.1	Overview	145
8.2	Summary of Results	145
8.3	Outlook	147

Bibliography	161
Appendix A	162
Appendix B	166
Appendix C	174

List of Figures

1.1	Architecture of metabolic networks	4
1.2	Fates of glucose during fermentation	6
1.3	Phase-contrast microscopic image of <i>Acetobacterim woodii</i>	8
1.4	Mathematical representation of a metabolic matrix as a stoichiometric matrix	10
1.5	Reaction subsets and elementary modes of a simple metabolic model	14
1.6	ScrumPy model snippet	20
2.1	Model construction work cycle	25
2.2	Fragment of <i>A.woodii</i> cell wall	29
2.3	ScrumPy formatted <i>A.woodii</i> top level module	30
3.1	A structural metabolic model of the Wood-Ljungdahl pathway	37
3.2	Elementary modes with CO ₂ -H ₂ as inputs	40
3.3	Elementary modes with CO as input	42
3.4	Elementary modes with CO ₂ -H ₂ -CO as inputs	44
3.5	Elementary modes with CO-H ₂ as inputs (part 1)	45
3.6	Elementary modes with CO-H ₂ as input (part 2).	46
3.7	Schematic of the Wood-Lungdahl pathway model extended to include fructose, formate and methanol metabolism.	48
3.8	Findings from a report on CO production and re-utilisation	55
3.9	Findings from a report on biphasic CO-based growth	59
4.1	The curated lysine biosynthesis pathway	67
4.2	The curated adenosylcobalamin (vitamin B12) biosynthesis pathway	69
4.3	Venn diagram of methanol related enzymes	72
4.4	TCA short caption	77
4.5	An overview of the <i>A.woodii</i> Genome-scale Model	81
5.1	Venn diagram comparing four LPs that calculate Y_{ATP}^{max}	89
5.2	Maintenance costs in amounts of acetate	96
5.3	Extracting a bioenergetic core	97

5.4	The bioenergetic core	99
5.5	The bioenergetic effect of a hypothetical NADH sink	103
6.1	Ammonia calibration curve	109
6.2	HPLC chromatograms of fructose experiments	110
6.3	Cell dry-weight calibration curve	111
6.4	Phase-contrast microscopy of <i>A.woodii</i>	113
6.5	Fructose growth experiments	114
6.6	Comparing model rates to observed rates	116
6.7	Comparing different maintenance costs	118
6.8	Comparing model and observed acetate yields across several conditions	121
7.1	Product survey workflow	126
7.2	Routes to 3-hydroxypropionate	128
7.3	Elementary modes of 3-hydroxypropionate synthesis	129
7.4	Analysis of 3-hydroxypropionate elementary modes	131
7.5	Route to ethanol	133
7.6	Routes to 1,3-butanediol	133
7.7	Routes to 3-hydroxybutyrate	134
7.8	Routes to acetone and isopropanol	135
7.9	Routes to butyrate	136
7.10	Routes to butanol	137
7.11	Product survey using C1 substrate feedstocks	140
7.12	ATP wastage increases 1,3-butanediol yields	143

List of Tables

3.1	Elementary modes of gaseous exchange in Wood-Ljungdahl pathway	38
3.2	Elementary modes of fructose and formate in the extended Wood-Ljungdahl pathway	50
3.3	Elementary modes of methanol in the extended Wood-Ljungdahl pathway model	51
4.1	Reactions in central metabolism that were either removed or added to the model	76
4.2	TCA short caption	78
4.3	Vitamin B12 yields in <i>A. woodii</i> when using different carbon and energy sources for growth.	83
5.1	Comparing EtOH-CO ₂ ATP yields	90
5.2	The Y_{ATP}^{max} of various substrates	91
5.3	Comparing maintenance costs in ATP amounts	94
5.4	Using alternative electron acceptors	104
6.1	Comparing model rates to observed rates	119
6.2	Validated elemental breakdown of <i>A. woodii</i>	120
7.1	Summary of 3-hydroxypropionate modes	130
7.2	Product survey using C1 substrate feedstocks	139
8.1	Every possible paired substrate combination: ATP Yields and Growth Phenotypes	166
8.2	Macromolecular composition adopted from [26]	174
8.3	Breakdowns of macromolecular biomass composition	174

Chapter 1

Introduction

1.1 Project Background

The Earth's human population is projected to grow to between 8.5 - 8.6 billion by 2030 and between 9.4-10.1 billion by 2050 [140], making the need to modernise industries towards sustainability in a world of limited resources ever pressing. The European Union and the United Kingdom have set targets to be world leaders in pioneering sustainable biobased economies [23, 48]. The UK for example has funded Networks in Industrial Biotechnology and Bioenergy (NIBB) which brings together academics, industry partners and policy makers to facilitate knowledge transfer and translatable research outcomes. A second phase of funding was provided for these NIBBs in 2018 of which CCNet (formerly C1Net) is a part of and with which this project is associated. The aim of CCNet is to develop technologies that use readily available feedstocks such as single carbon gases and utilise waste streams to synthesise valuable commodity chemicals. This is a BBSRC iCASE funded project to investigate the potential of using single-carbon substrate (C1) consuming bacteria for bio-industrial purposes. An iCASE studentship requires that the academic aspect of the work is partnered to a biotechnology company to attempt to more directly translate research into industrial application.

It is with this in mind that this project aims to apply metabolic modelling techniques to a chassis microbe to contribute to its engineering for a more circular economy. Acetogenic bacteria are able to grow on C1 gases such as those found in synthesis gas (a combination of CO_2 - CO - H_2). Throughout the work carried out here the focus is on using metabolic modelling to gain insight into *A. woodii*, a model acetogen, and then to explore engineering strategies for an industrially relevant strain. We have previously collaborated with ZuvaSyntha to target 1,3-butadiene production from C1 substrates, but that partnership has been interrupted due to business insolvency. We have since collaborated with Novo Nordisk Foundation Center for Biosustainability, based at Denmark Technical University to engineer *A.*

woodii a 3-hydroxypropionic acid producing strain from C1 substrates.

1.2 Bacterial Metabolism

A bacterial cell like all living organisms reproduces itself as its primary function. To build the biological components needed for a new cell, matter from its surroundings is consumed and transformed in a series of biochemical reactions which as a whole is called metabolism. These biochemical reactions can be classified by their function [84]:

- **Assembly reactions** make modifications to macromolecules, transport them to pre-specified locations within the cell, and allows for the association of macromolecules into larger cellular structures such as cell membranes, and walls.
- **Polymerisation reactions** form branched or unbranched polymeric chains of various molecules and are otherwise known as macromolecules. This is achieved by sequential linkage of smaller precursor molecules and compounds.
- **Biosynthetic reactions** build all of the necessary biomass precursors from a central pool of central metabolites for polymerisation reactions. They also build cofactors and other metabolic factors needed for the synthesis of these. Many of these precursor metabolites are essential to all living organisms such as amino acids and nucleic acids.
- **Fuelling reactions** generate energy in the form of adenosine triphosphate (ATP) and other energetic molecular currency necessary for cellular functions. Additionally, these reactions include those that break down compounds into a central pool of metabolites from which biomass precursors are built.

Given that the physical state and chemical composition of the surroundings of a cell is dynamic, the metabolism of a cell is suitably dynamic to adapt to environmental changes. The classification of the reactions above can be further organised into two larger groups called anabolic and catabolic reactions, where catabolic reactions comprise the fuelling reactions and anabolic reactions build the molecules needed for reproduction with assembly, polymerisation and biosynthetic reactions. The overall architecture of metabolic networks resembles that of a “bow tie” and is depicted in Figure 1.1. Catabolic reactions funnel or fan in the products of degraded and broken down nutrients into a central pool of metabolites and anabolic reactions fan out from these central metabolites into various biomass building blocks and growth by-products. The by-products are excreted into the extracellular environment either for functions such as intercellular signalling or as a consequence of a metabolic dead-end. Reaction sequences are commonly organised into functional units called metabolic pathways. Some of the readily recognisable metabolic pathways in central metabolism for example are glycolysis, the tricarboxylic acid cycle (TCA) and the pentose phosphate pathway (PPP).

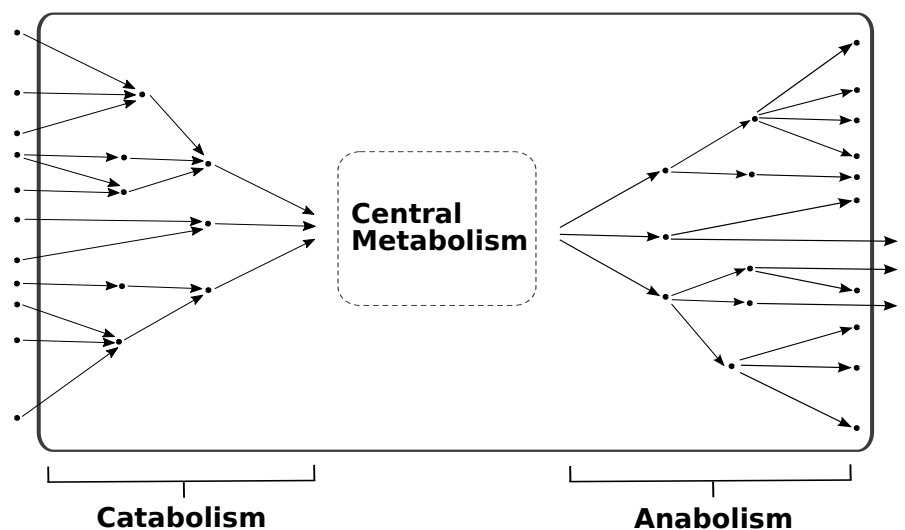


Figure 1.1: **The architecture of metabolic networks.** Catabolic reactions fan-in nutrients found in the extracellular environment to a central pool of metabolites and generate chemical energy. Anabolic reactions fan-out from the central pool of metabolites building the biological components of cells. Central metabolic reactions are the interface between catabolic and anabolic reactions. The outer solid box is the boundary between extracellular and intracellular environments.

1.2.1 Anabolism

Anabolic processes as a whole are endothermic and so require chemical energy that is produced from catabolic processes. Of the cellular constituents built in anabolism 92% of their dry weight contains the elements carbon, hydrogen, oxygen, nitrogen, phosphorus and sulphur. The main groups of biomass precursors and the polymers they are used to build are amino acids for building proteins and other cellular structures, nucleic acids for building DNA, RNA and several metabolic cofactors, fatty acids and sterol esters for cellular membranes, peptidoglycan monomers and teichoic acids for cell walls and other cofactors such as vitamins which are used in essential metabolic processes. Central metabolites used to synthesise nucleotides are ribose-5-phosphate and 3-phosphoglycerate for purine nucleotides and oxaloacetate for pyrimidine nucleotides. Amino acids synthesised from several central metabolites such as phosphoenol pyruvate, pyruvate, oxaloacetate, erythrose-4-phosphate and glucose-6-phosphate [126].

1.2.2 Catabolism

Catabolic processes are ultimately exothermic, yielding energy in the form of ATP and reduced cofactors (often as NADH or NADPH), and also provide the carbon skeletons that

central metabolites are comprised of. Compounds that enter catabolic pathways are often referred to as substrates. Substrates that supply carbon are called carbon sources and substrates that supply energy are called energy sources. Many substrates such as glucose serve to provide both carbon and energy and so are called carbon and energy sources. As glucose is such a widely distributed and available substrate in nature, glycolysis - catabolic pathways by which glucose is broken down to pyruvate - is found almost universally in all living organisms [24].

Further distinctions can be made between organisms that metabolise substrates aerobically (with O_2), anaerobically (without O_2), or metabolise substrates as a facultative anaerobe (both with and without O_2). The fate of pyruvate from glycolysis for example is dependant in part on whether aerobic or anaerobic catabolic pathways act upon it.

Aerobic

The role of oxygen in aerobic catabolism of substrates is to accept terminal electrons after passage through an electron transport chain (ETC) to form H_2O . Passage of electrons through the ETC is powered by reduced cofactors. Thus the recycling of cofactors is only limited by the amount of O_2 in the system. This type of metabolism is also referred to as respiration. As compared to anaerobic catabolism, using oxygen yields a higher amount of new cells per substrate. The end products of glucose if catabolised this way is the fully oxidised CO_2 and H_2O . This is achieved via the action of three catabolic pathways, glycolysis, TCA cycle and the ETC.

Anaerobic

In anaerobic pathways the terminal electron acceptor varies, and so the product of catabolism are generally more varied also. Anaerobic metabolism can also utilise an ETC but will do so with electron acceptors other than O_2 . Less oxidised substrates such as nitrate, nitrite, sulphate, fumarate or sulphur act as electron acceptors instead and this is called anaerobic respiration. Anaerobic metabolism without use of an ETC is called fermentation. In Figure 1.2 fates of glucose during fermentation from various bacteria have been compiled and are shown [61, 101].

Maintenance energy

The rate of energy production is dependent on the energy needed for growth and non-growth maintenance. Microbial growth is distinct from reproduction in that growth refers to the increase in the number of cells and is exponential in nutrient-rich and favourable conditions. During growth the sum of energy produced from catabolism can be categorised into growth associated maintenance (GAM), and non-growth associated maintenance (NGAM). Anabolism has already been discussed and involves transport, polymerisation and biosyn-

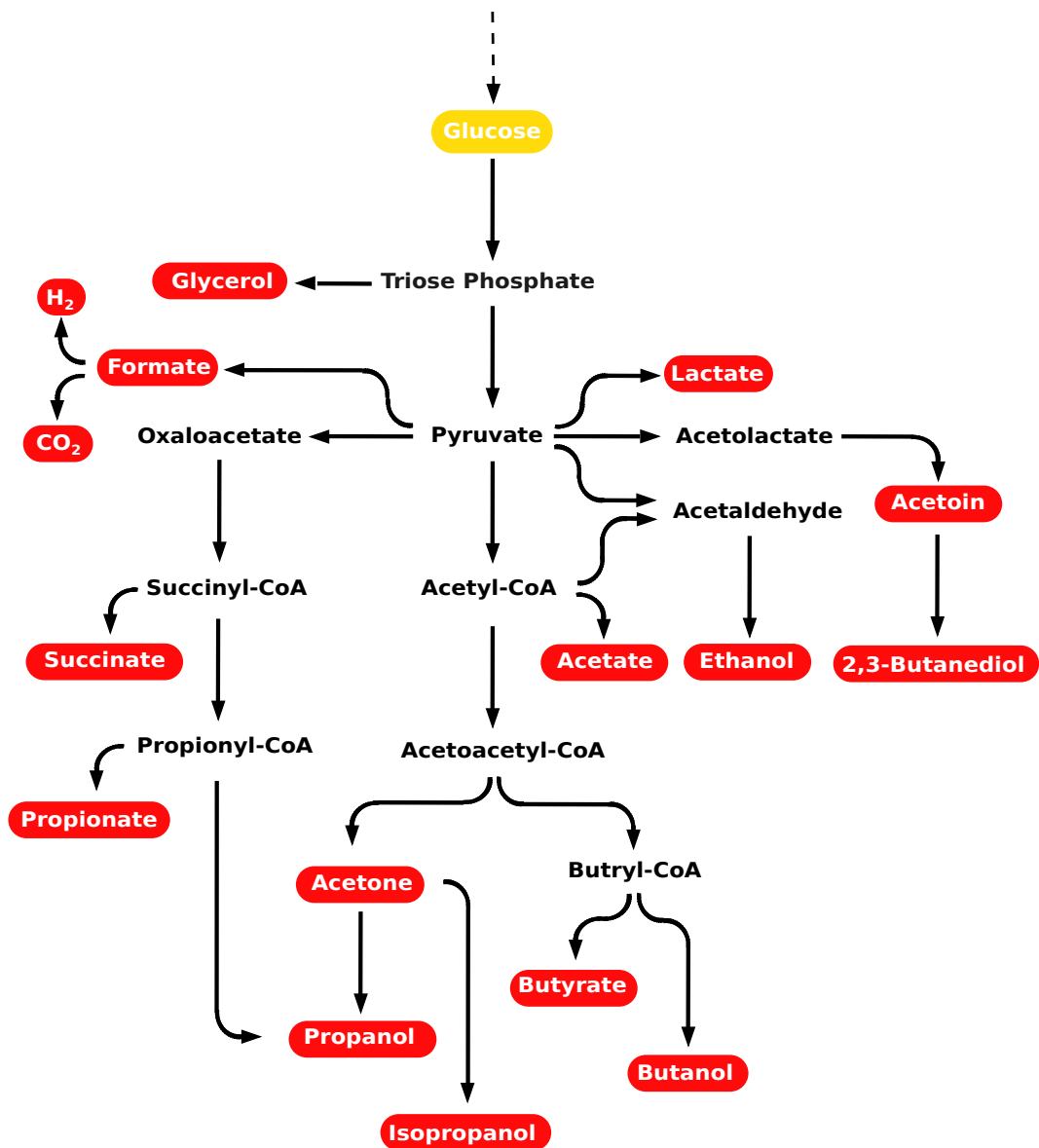


Figure 1.2: **Fates of glucose during fermentation from various bacteria.** Arrows represent single reactions or multiple lumped reactions, compounds highlighted in red are fermentation end products and excreted into extracellular environments. These pathways have been simplified for demonstration purposes and so exact stoichiometries are not shown.

thetic processes which are all processes needed for growth. NGAM encapsulates the energetic cost of processes such as macromolecular turnover, motility and the maintenance of concentration gradients and electric potential gradients [90]. Through precise measurements of substrate uptake rates and stoichiometric growth calculations at different growth rates, it is possible to experimentally determine GAM and NGAM.

1.3 Acetogenic Bacteria

Acetogenic bacteria are a diverse group of C1 fermenting microbes, thriving in rich CO₂, CO and H₂ environments. They utilise a metabolic pathway called the acetyl-CoA pathway, otherwise known as the Woods-Ljungdahl Pathway (WLP) after Harland G. Wood and Lars G. Ljungdahl who made major contributions to the elucidation of it [100]. WLP allows for the fixation CO₂ into cell carbon via acetyl-CoA, generating acetate as a by-product [33]. This is achieved by reducing one CO₂ into formate which is ligated with a tetrahydrofolate (THF) to form formyl-THF at the expense of ATP. Formyl-THF then undergoes several methyl transformations to form methyl-THF. A second CO₂ is reduced to carbon monoxide via a bifunctional carbon monoxide dehydrogenase acetyl CoA synthase (CODH-AcS). The methyl group from methyl-THF is then combined with carbon monoxide and CoA into acetyl-CoA. Acetyl-CoA is then either utilised for biomass or energy generation by producing and excreting acetate. It is thought to be one of the ancient metabolic pathways and receives interest as candidate for perhaps the oldest of the modern CO₂ fixation pathways [125]. Although all acetogens produce acetate as a by-product, some acetogens do not produce it as an end-product [1].

1.3.1 *Acetobacterium woodii* as a Model Acetogen

A. woodii was first isolated as a type strain for the novel *Acetobacterium* genus on a summer microbial ecology course in Woods Hole, Massachusetts, US in sediments of a small creek estuary [6, 32]. It is a gram-positive bacteria growing homoacetogenically in rod-shaped pairs, and on a wide range of substrates [6]. It is roughly 1 μ m wide and 2 μ m in length and has a singular subterminal flagellum. Of the acetogens, it is currently the best biochemically characterised since the elucidation of the WLP enzyme methylene-THF dehydrogenase [9], and has all reactions in WLP stoichiometrically defined. Uniquely among acetogens, it employs a Na⁺ translocating Rhodobacter nitrogen fixation complex (Rnf complex) in its cell membrane which couples ferredoxin and NADH transhydrogenation with Na⁺ translocation to build an electrochemical gradient [56]. Apart from the Rnf no other membrane-embedded cytochromes have been found. Based on the growth substrate being used and the redox balance of the cell the Rnf can operate in a fully reversible manner. The Na⁺ gradient is utilised by a F/V-hybrid rotor ring ATP synthase that couples ATP synthesis with Na⁺ import and ATP hydrolysis with Na⁺ export [79]. Two further notable and unique enzymes

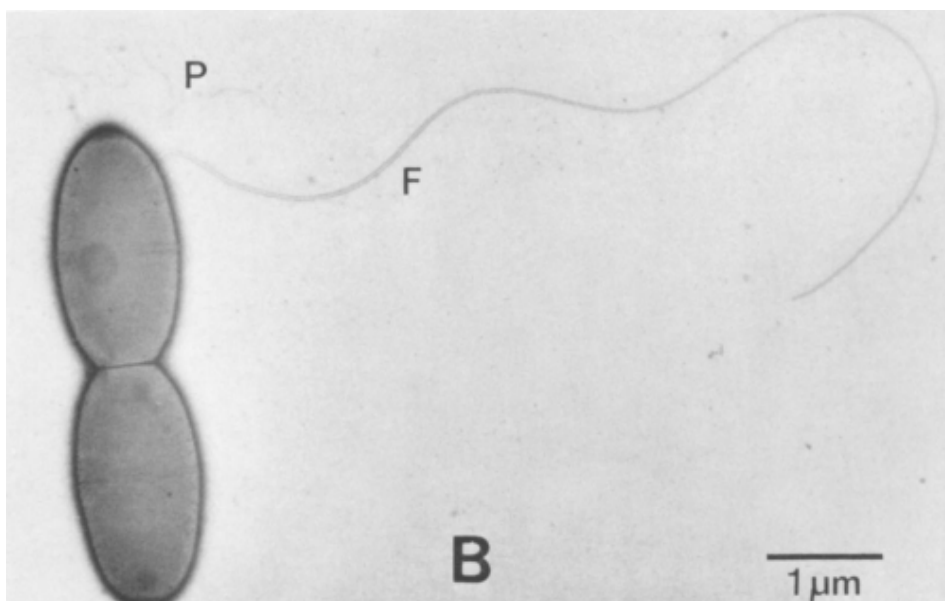


Figure 1.3: **Phase-contrast microscopic image of *Acetobacterim woodii*.** Taken from [6]. A pili-like structure [P] and subterminal flagellum [F] are indicated. *A. woodii* grows in distinct pairs of rod-shaped cells.

discovered in *A. woodii* are a hydrogen-dependent formate dehydrogenase (HDFD) [113] and an electron-bifurcating hydrogenase (EBHyd) [112]. Both are fully reversible and are essential for net positive ATP generation when growing on $\text{CO}_2\text{-H}_2$. EBHyd reduces ferredoxin (an essential reductant for ATP generation via the Rnf complex) and NAD by stripping electrons from H_2 . HDFD reduces CO_2 to formate using H_2 instead of reduced ferredoxin, thereby allowing the reduced ferredoxin to contribute towards the Na^+ gradient.

Recent sequencing of the *A. woodii* genome has answered some elusive questions about how energy is conserved during carbon fixation [92] and has allowed for Genome-scale Metabolic Modelling of its metabolism. Metabolic modelling allows for system-wide analysis of metabolism as opposed to a reductionist approach. The underpinnings of the metabolic modelling approaches used here are described in the next section. The advantages of using *A. woodii* for biotechnological applications are that it is a non-pathogenic, non-sporulating, mesophilic, non-photosynthesising microorganism readily cultivatable on defined media.

1.4 Theoretical Underpinnings

1.4.1 Mathematical Description of Metabolic Networks

A quantitative approach to modelling a metabolic system requires a mathematical description of it. A chemical reaction catalysed by an enzyme can be abstracted into the following

notation:



where A and B in this reaction are reactants (or substrates) that are catalytically converted by an enzyme, which is represented by the arrow, into the product C. The reaction proceeds in a physiologically relevant direction which is in this case, two moles of A and one mole of B are catalysed into one mole of C. This quantitative relationship between reactants and products is called the stoichiometry. The reactants and products of reactions can also be called metabolites, given that they are often intermediates in net metabolic processes. In a metabolic network, a reaction like this would be one of many hundreds or thousands for any given organism. What is not captured by this abstraction are other properties of the metabolites and enzymes such as a kinetic aspects, but this as a basic description of the reaction can be used as a foundation for layering additional properties of reactions if required.

One convenient way of mathematically describing a metabolic network is as a matrix where column entries describe a reaction and rows a metabolite [53]. Since modelling is most commonly concerned with discrete systems, a distinction between internal and external metabolites is made. The conversion of a simplified metabolic network into a stoichiometry matrix is shown in Figure 1.4.

Entries in the matrix are the stoichiometric coefficients, a positive number denotes production of a metabolite and negative a consumption of metabolite. An external stoichiometric matrix differs from the internal stoichiometric matrix by having additional rows for the external metabolites of its network. In matrix form then, it is a trivial matter for example to count the number of reactions in a metabolic network using a computer program by counting the number of columns, or counting the amount of metabolites in the network by counting the number of rows, or to identify how connected a metabolite is (in terms of how many reactions it is involved in). For example, metabolite C has a connectedness of four and metabolite F is referred to as an “orphan” metabolite because it has a connectedness of one. Orphan metabolites have a single row entry in the stoichiometry matrix. A metabolite similar but distinct to orphans is a dead-end metabolite. These are metabolites associated with more than one reaction, but associated with reactions that either only consume it or produce it. These are identifiable by assessing row entry coefficient signs i.e. row entries with only positive or only negative coefficients. Further informative features of a network that are described by the matrix and its analysis will be described in later sections.

Another way to describe the metabolic network in Figure 1.4 is as a system of ordinary differential equations:

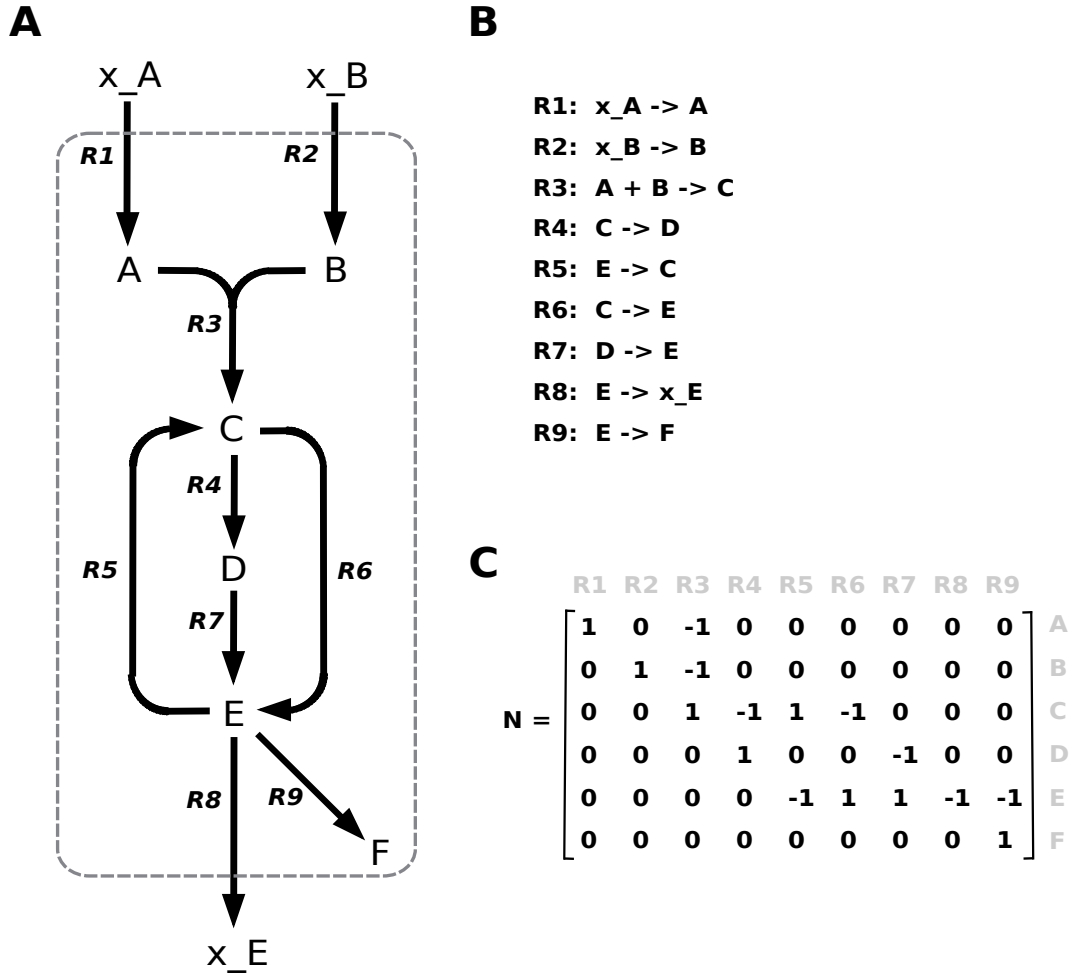


Figure 1.4: **Mathematical representation of a simple metabolic network as a stoichiometric matrix.** [A] shows a network diagram of a simple metabolic network. The “x_” prefix indicates the metabolite is external, the arrows represent enzymatic reactions and letters represent metabolites. The dashed grey line is the boundary of the network. [B] shows the metabolic network written as a series of reactions, [C] is the internal stoichiometric matrix describing this network.

$$\begin{aligned}
\frac{dA}{dt} &= v_{R_1} - v_{R_3} \\
\frac{dB}{dt} &= v_{R_2} - v_{R_3} \\
\frac{dC}{dt} &= v_{R_3} - (v_{R_4} + v_{R_6}) \\
\frac{dD}{dt} &= v_{R_4} - v_{R_7} \\
\frac{dE}{dt} &= (v_{R_6} + v_{R_7}) - (v_{R_5} + v_{R_8} + v_{R_9}) \\
\frac{dF}{dt} &= v_{R_9}
\end{aligned}$$

where v_{R_i} is the rate of reaction i that depends on metabolite concentrations and kinetic properties of the enzyme such as the Michaelis-Menten parameters. The stoichiometric matrix - as well as describing the stoichiometric structure - can also be used for the generalisation of reaction rates:

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \\ \frac{dD}{dt} \\ \frac{dE}{dt} \\ \frac{dF}{dt} \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} v_{R_1} \\ v_{R_2} \\ v_{R_3} \\ v_{R_4} \\ v_{R_5} \\ v_{R_6} \\ v_{R_7} \\ v_{R_8} \\ v_{R_9} \end{bmatrix}$$

where the rate of change of internal metabolite concentrations is the dot product of a vector of reaction rates and the internal stoichiometry matrix. This can be expressed in matrix notation as:

$$\frac{d\mathbf{s}}{dt} = \mathbf{N}\mathbf{v} \tag{1.1}$$

where $\mathbf{s}$ is a vector of internal metabolites, $\mathbf{N}$ is the internal stoichiometry matrix and $\mathbf{v}$ is a vector of reaction rates.

1.4.2 Metabolism at Steady State and the Nullspace

The concept of a metabolic steady state is one of a dynamic equilibrium maintained by a living organism that keeps the internal flux of matter and energy constant over time. This is achieved by consuming source material from the environment at a constant concentration and excreting an end product at a constant concentration via the action of metabolic pathways. This requires that the flux through these pathways remains constant, so that the net consumption and production of intermediate metabolites is balanced [39]. Laboratory experiments that involve growth of organisms over time commonly observe metabolism approaching steady state during exponential growth in batch cultures, and closer still during

continuous growth in bioreactors.

If the steady state assumption is made when modelling a metabolic network then, the rate of change of metabolite concentrations over time is zero, and so:

$$\frac{ds}{dt} = \mathbf{N}\mathbf{v} = \mathbf{0} \quad (1.2)$$

where $\mathbf{s}$, $\mathbf{N}$ and $\mathbf{v}$ are the same in equation 1.1, and $\mathbf{0}$ is a zero vector. A basis for vector $\mathbf{v}$ can now be determined and will comprise a set of linearly independent vectors that satisfy all possible steady state solutions; this is called the nullspace, or kernel ($\mathbf{K}$). Its calculation is relatively straightforward using the well known method of Gauss-Jordan Elimination. It is also independent of reaction kinetics and scalable to metabolic networks of organisms, making it a powerful tool for metabolic modelling.

$$\mathbf{K} = \begin{array}{ccc|l} 0 & 1 & 0 & \text{R1} \\ 0 & 1 & 0 & \text{R2} \\ 0 & 1 & 0 & \text{R3} \\ 0 & 0 & 1 & \text{R4} \\ 1 & -1 & 1 & \text{R5} \\ 1 & 0 & 0 & \text{R6} \\ 0 & 0 & 1 & \text{R7} \\ 0 & 1 & 0 & \text{R8} \\ 0 & 0 & 0 & \text{R9} \end{array} \quad (1.3)$$

The vectors of the kernel are not necessarily unique and can depend on the calculation method used, but nonetheless, invariant properties of the stoichiometric matrix are identified by it. Two biologically relevant properties are identified by inspection of $\mathbf{K}$ alone: so-called “dead” reactions, and reaction subsets. Dead reactions are defined as reactions not able to carry flux under any steady state condition and are identified in the kernel by a zero row vector [117]. These reactions are not however necessarily dead if a transient metabolism is considered. Reaction subsets (sometimes less precisely called enzymes subsets) are reactions that must carry the same or proportional flux in any steady state solution. Reactions in the same subset are identified by proportional row vectors in $\mathbf{K}$ [89].

So, from analysis of the structure of a metabolic network alone, emergent properties of it are readily identified that are true regardless of kinetic aspects. For example, if a reaction is dead, no matter the V_{max} of an enzyme or its affinity for a substrate (K_m), the reaction will not be relevant during steady state. The kernel in equation 1.3 shows that R9 in Figure 1.4 is a dead reaction, and two reaction subsets exist: subset 1, consisting of R1, R2, R3 and R8 and subset 2, consisting of R4 and R7. Both reaction subsets are depicted schematically in Figure 1.5A.

Any steady state reaction rates vector ($\mathbf{v}$) can be represented as a linear combination of the vectors in $\mathbf{K}$. The linear combination is a vector of arbitrary linear scalars (C) that

must contain the same number of elements as the column dimension of $\mathbf{K}$:

$$\mathbf{v} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} = \begin{bmatrix} C_2 \\ C_2 \\ C_2 \\ C_3 \\ (C_1 + C_3) - C_2 \\ C_1 \\ C_3 \\ C_2 \\ 0 \end{bmatrix} = \begin{bmatrix} v_{R_1} \\ v_{R_2} \\ v_{R_3} \\ v_{R_4} \\ v_{R_5} \\ v_{R_6} \\ v_{R_7} \\ v_{R_8} \\ v_{R_9} \end{bmatrix} \quad (1.4)$$

and so to check $\mathbf{v}$ satisfies equation 1.2, its multiplication with $\mathbf{N}$ must equal zero:

$$\mathbf{N}\mathbf{v} = 0 \Leftrightarrow \begin{array}{cccccc} 1 \cdot v_{R_1} & + & 0 \cdot v_{R_2} & + & -1 \cdot v_{R_3} & + & \cdots & + & 0 \cdot v_{R_9} & = & 0 \\ 0 \cdot v_{R_1} & + & 1 \cdot v_{R_2} & + & -1 \cdot v_{R_3} & + & \cdots & + & 0 \cdot v_{R_9} & = & 0 \\ \vdots & & \vdots & & \vdots & & & & \vdots & & \vdots \\ 0 \cdot v_{R_1} & + & 0 \cdot v_{R_2} & + & 0 \cdot v_{R_3} & + & \cdots & + & 1 \cdot v_{R_9} & = & 0 \end{array} \quad (1.5)$$

The stoichiometric matrix is often under-determined since there are often fewer metabolites than reactions in metabolic networks and so an infinite number of solutions satisfy this subspace. Relevant solutions though can be determined using other modelling tools, some of which will be discussed below.

1.4.3 Elementary Modes Analysis

Several methods allow for the further exploration of $\mathbf{K}$, some of which are Linear Programming [37], Mixed integer Linear Programming [74], Flux Variability Analysis [49] and Elementary Modes Analysis (EMA) [118]. EMA allows for a metabolic network to be decomposed into a comprehensive set of minimal reaction pathways able to operate as stand-alone sub-networks, capable of carrying steady state flux. “Minimal” means that the deletion of a reaction in an elementary mode would break a steady state flux-carrying network, and so an elementary mode cannot be further decomposed into more elementary modes. Additionally, any irreversible reaction in a elementary mode must carry either positive or zero flux thus obeying thermodynamic constraints imposed by reaction directionality. This allows for objective identification of metabolic pathways in a metabolic network; a task still abound with subjectivity [119].

EMA is a fundamental tool of structural modelling and has proven to be a valuable tool in metabolic engineering efforts [137]. It was used for example to engineer *E.coli* into an ethanol producing strain, obtaining ethanol yields close to those theoretically predicted.

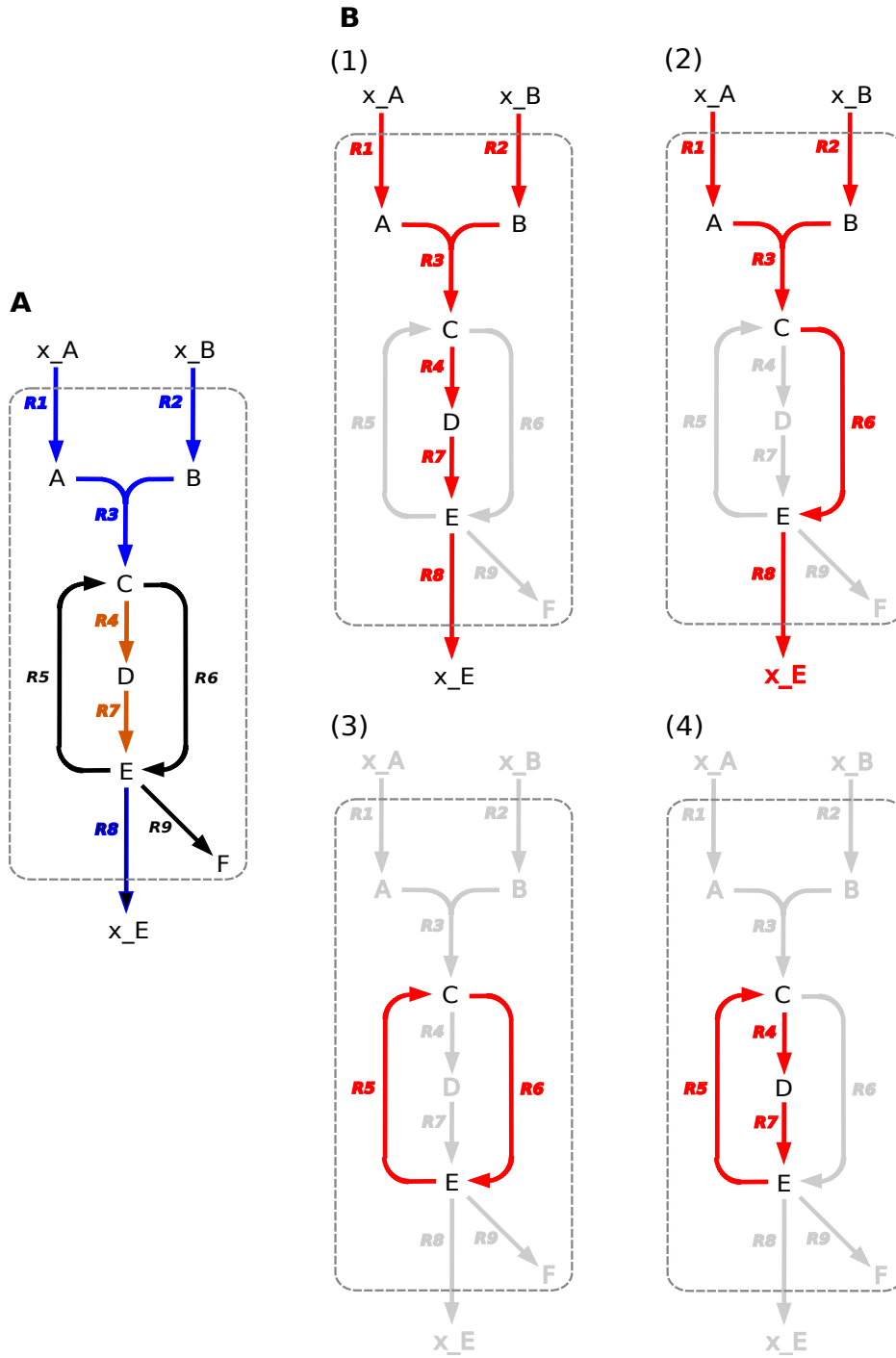


Figure 1.5: **Reaction subsets and elementary modes of a simple metabolic model.** [A] Reaction subsets of the metabolic model. Reactions in blue: R1, R2, R3 and R8 are one subset and reactions in orange: R4 and R5 are the second subset. Reactions in black are not in any subset. [B] Elementary modes of the model. Each mode is represented by a schematic and is a valid non-decomposable steady state route in this network.

This was achieved by constructing a medium sized model of *E.coli* central metabolism and computing routes to ethanol and biomass from glucose and pentoses. They identified a set of eight reactions that were not essential to ethanol and biomass production and constructed a deletion mutant with the genes coding for enzymes catalysing those eight reactions [136]. This approach has been generalised into a concept called minimal cut set analysis and has many potential applications to rational metabolic engineering [69, 68].

To demonstrate EMA, the elementary modes of the simple metabolic network from Figure 1.4 were computed. Schematics of the four possible modes are shown in Figure 1.5. Modes (1) and (2) both have a net stoichiometry of $1 A + 1 B \rightarrow 1 E$, and were routes independently identified with LP equations 1.7, 1.9 and 1.8. Modes (3) and (4) are both internal cycles with no net stoichiometry. All four meet the requirements of an elementary mode.

If one was interested in disrupting the production of say E, effective single reaction knockout targets of one of either R1, R2, R3 or R8 would achieve this, whereas other reaction targets would need a multiple reaction knockout approach. These targets also happen to be part of a single reaction subset, demonstrating further that knowledge of the contribution of individual reactions to a larger reaction network provides valuable information when attempting to engineer the network as a whole.

1.4.4 Linear Programming

Linear Programming (LP) is a mathematical optimisation method that finds uses in fields of study as wide as economics, business and telecommunications, and for modelling a diverse type of problem, which includes metabolic modelling. It can be used to explore the otherwise infinite solution space of $\mathbf{K}$, by constraining it to a smaller subspace and identifying in that subspace a particular point or facet. A point in this space is a unique optimal solution and a facet is a non-unique optimal solution. This is achieved by applying linear equality and inequality constraints, thereby reducing the feasible solution space and identifying a point at the edges of the reduced feasible space by setting a linear objective function to either be maximised or minimised.

Example of a Linear Programming Problem

The formulation of a LP problem is subjective in that it needs a function to optimise and a set of constraints. Although this may be trivial on for example a factory floor where costs of material are clear, hours of labour are known and the optimisation principle is likely to be profit, these optimality principles are not so obvious in a living organism. A simple demonstration of a LP formulation can be seen with the well known Enginola problem which is worked through in great detail in [110].

It is essentially a mixed products problem in that a television company called Enginola wants to maximise its profit from the production of two types of TV sets, “Astro” and

“Cosmo” sets, but is constrained by various aspects of the production line. The capacity of the Astro line is 60 sets per day, and Cosmo, 50 sets per day. It takes one hour of labour to produce an Astro, and two to produce a Cosmo, and 120 hours of labour per day is available to the factory. Finally, the profit contribution of each set is £20 for the Astro and £30 for the Cosmo. The question then is - what should daily production of these sets be to make the highest profits? This problem is formulated in the following way:

$$\begin{array}{ll} \text{maximise :} & 20A + 30C \quad (\text{Profit in pound sterling}) \\ \text{subject to} & \left\{ \begin{array}{ll} A \leq 60 & (\text{Astro capacity}) \\ C \leq 50 & (\text{Cosmo capacity}) \\ A + 2C \leq 120 & (\text{Labour in hours}) \end{array} \right. \end{array} \quad (1.6)$$

where A is units of Astros produced per day, and C is units of Cosmos produced per day. The first line indicates the objective function which is to maximise profits, and the subsequent lines are constraints based on the factory resources which are the two production lines and available labour hours per day. Since this is a simple problem with only two variables (A and C), it can be worked out algebraically or graphically by hand but for complex problems algorithms are necessary and more convenient, of which the Simplex method [111] is a particularly successful one. The particular solution to this problem is to make 60 Astros (the maximum amount), and 30 Cosmos. The objective value (profit in pounds) is £2100 per day.

Applications of Linear Programming to Metabolism

The application of LP to metabolic networks was reported as early as 1986 by Watson [144] who was looking for alternatives to what was - at the time - computationally demanding kinetic modelling, mostly for teaching purposes. It was Fell and Small [37] in the same year that used Linear Programming for a biological question; how was triacylglycerol synthesised from glucose in rat adipose tissue? It was found that setting the steady state constraint, assigning a flux constraint on a few reactions, and setting an objective function allowed for a realistic flux value to be assigned to every reaction in the network. It was also found that modifying constraints on intermediate metabolites was a powerful way of computing pathways through the metabolic network, and that experimentally suggested pathways could be predicted this way.

These findings can be demonstrated by formulating linear programming problems to the simple metabolic network seen in Figure 1.4. In the first example below, flux through the “E” transporter is maximised, and upper bounds are put on import of “A” and “B”:

$$\begin{aligned}
& \text{maximise :} && R8 \\
& \text{subject to} && \begin{cases} \mathbf{N}\mathbf{v} = 0 \\ R1 \leq 1 \\ R2 \leq 1 \end{cases}
\end{aligned} \tag{1.7}$$

where $R8$ is the E transporter, $R1$ and $R2$ are the transporters for A and B and $\mathbf{N}\mathbf{v} = 0$ is the steady state constraint. The solution is a vector of relative reaction fluxes: $R1:1$, $R2:1$, $R3:1$, $R6:1$, and $R8:1$. All other reactions are zero. Since the objective function was to maximise $R8$, the objective value is one. This solution represents a feasible route through this network during steady state metabolism. Any number of additional constraints can be added to the formulation, however a system too constrained with result in infeasible solutions. Conversely, a system not constrained enough will result in an unbounded solution i.e. an unlimited amount of the objective function. An additional constraint on $R6$ would change equation 1.7 to:

$$\begin{aligned}
& \text{maximise :} && R8 \\
& \text{subject to} && \begin{cases} \mathbf{N}\mathbf{v} = 0 \\ R1 \leq 1 \\ R2 \leq 1 \\ R6 \leq 0 \end{cases}
\end{aligned} \tag{1.8}$$

where $R6 \leq 0$ essentially blocks flux through this reaction. This is because, as a uni-directional reaction, no negative flux (i.e. reaction running in reverse) is possible. The solution to this LP is: $R1:1$, $R2:1$, $R3:1$, $R4:1$, $R7:1$, and $R8:1$. This solution is qualitatively different to the solution of equation 1.7, but the objective value is the same at one. This highlights a disadvantage of using an LP based approach. Only a single solution is returned per LP problem, which means that “alternative optimal” solutions will not be seen. There methods to address this issue which includes using other network analysis approaches and is discussed in the next section. In some instances, reformulating the LP problem will also return a single optima, as will be seen in the next LP. Essentially though, problems 1.7 and 1.8 both show that the maximum yield of metabolite E from metabolites A and B is one, and the solutions show metabolic routes for achieving this which is interesting in its own right. The following LP is both less constrained and results in a single optimal solution (for this network):

$$\begin{aligned}
& \text{minimise :} && \sum |v_i| \\
& \text{subject to} && \begin{cases} \mathbf{N}\mathbf{v} = 0 \\ R8 = 1 \end{cases}
\end{aligned} \tag{1.9}$$

where v is a vector of reactions for the i 'th column name of the stoichiometry matrix (i.e. reaction) and $|v_i|$ is value of all reaction fluxes. The singular optimal solution for this is:

R1:1, R2:1, R3:1, R6:1, and R8:1. One advantage of being able to use fewer constraints is that fewer fluxes have to be experimentally determined. Also, the minimisation of total flux tends to return solutions with fewer reactions i.e. shorter routes, has solutions with fewer internal cycles and overall minimises the total flux load of the network when compared to LPs that maximise a metabolic objective. It has been shown to be an effective optimisation method [59, 120, 97, 51], and is used as a proxy for minimising the investment in total enzyme synthesis. Efficient proteome allocation has also been shown to be employed in *E.coli* experimentally [7]. However, maximisation objectives are also widely used albeit at the expense of requiring additional constraints [143, 141].

1.4.5 Genome-scale Metabolic Models

The application of LP to metabolic models has since expanded to larger metabolic networks and further innovations have made it a staple of metabolic modelling. It has been rebranded Flux Balance Analysis (FBA) and the use of biomass equations [36] has been popularised to develop whole-cell metabolic analysis capabilities and cellular growth [104, 141, 142], in ways predictive of realistic growth patterns and rates [143]. These whole-cell models encompass the entire proteomic repertoire of an organisms genome and are called Genome-scale Metabolic Models (GSMs). Their construction has been continually refined [46, 38, 132, 96] and their capability to predict cellular physiology and metabolism has made them essential in interpreting aspects of whole genome sequences and for metabolic engineering [145, 81, 35]. GSMs have been constructed for both unicellular and multicellular organisms and for a great variety organisms [97, 51, 134, 133, 86], showing a greater applicability to understanding large-scale metabolic networks.

1.5 ScrumPy

ScrumPy was developed in-house in the Oxford Brookes University, Cell Systems Modelling group and is a mature metabolic modelling software package made publicly available in 2006 [95]. It is freely available and open source. It was initially made available for various Linux distributions, but as of Windows 10, version 1607 it also runs natively in a Windows environment using the Subsystem for Linux, a component that allows Linux-compatible binary software to be run in an Ubuntu-based mode environment. All metabolic modelling work was carried out using it and all auxiliary Python code was executed in it. Some design features include use of the Python language as an interface via the Integrated Development Environment (IDLE). This is a deliberate aspect of its design driven by the restrictive nature of GUIs for - including but not limited to - carrying out tasks in a reproducible way and allowing development of novel algorithms, code and scripts. An object-orientated programming approach is used to abstract models as data objects. Complete access to lower-level data structures of the model are available but also effectively hidden when unnecessary

to view. This allows minor variations to be made to standard analysis as well as development of entirely novel algorithms. Also, with it also comes an extensible set of data analysis modules and graphing facilities making it accessible as off-the-shelf software. Further design features are set out in [95].

In ScrumPy, a model is loaded from a ScrumPy formatted file which is composed of asci text. A ScrumPy model file is a collection of reaction names with associated reaction stoichiometry, as well as a directive pointing ScrumPy to the type of model being created, which, in the example below, is a structural one. Another (optional) directive is to declare certain metabolites as external:

```
Structural()
External("PROTON")
```

The command to load a model file is:

```
m = ScrumPy.Model(ScrumPyfile)
```

where ‘Model’ is a function in the ScrumPy constructor module, ‘ScrumPyFile’ is the ScrumPy formatted file, and ‘m’ is the variable the model is stored in. For an example of the syntax used to describe reactions in a ScrumPy file, see Figure 1.6 for a model snippet and Appendix A for a small model. Various methods can be called on the model object, such as computing elementary modes or reaction subsets of that model:

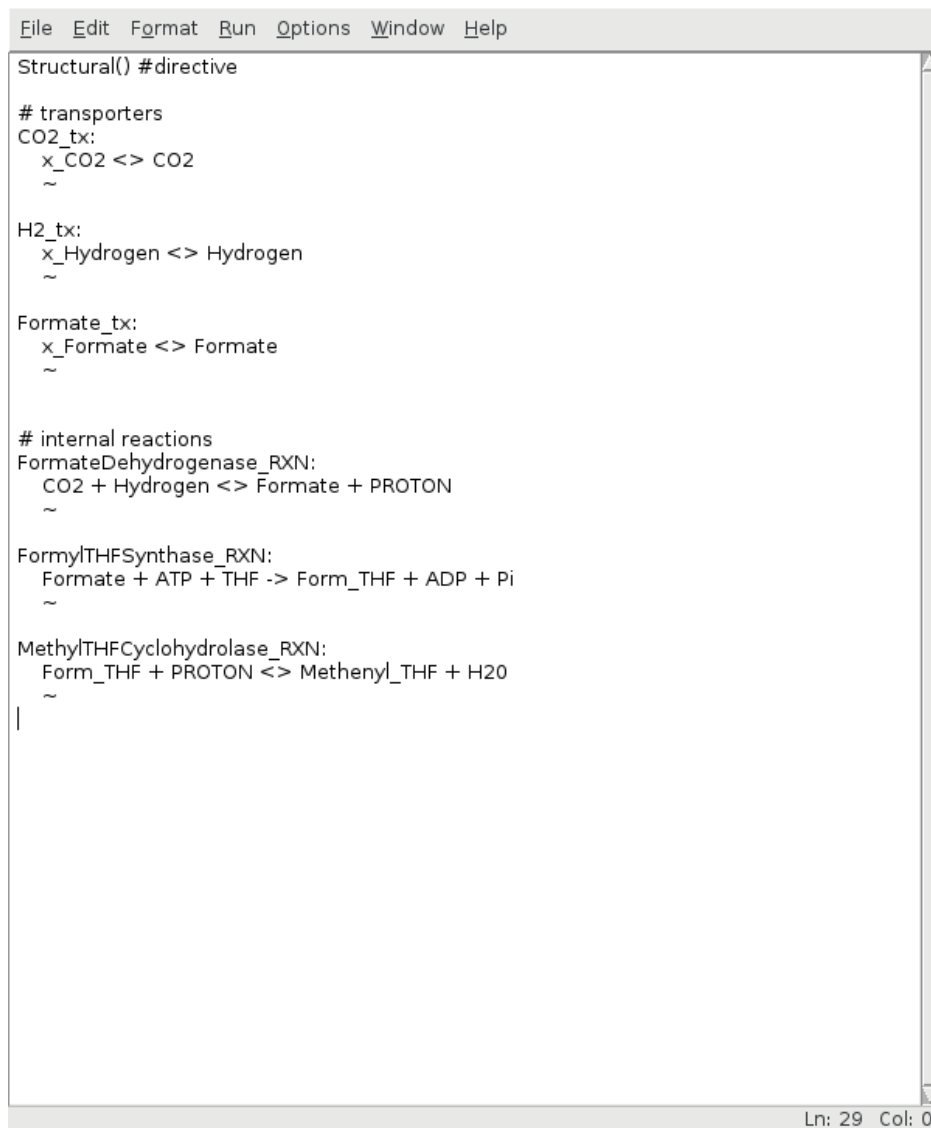
```
es = m.EnzSubsets()
el = m.ElModes()
```

where es is a variable that stores the reaction subsets of that model, and el is a variable that stores the elementary modes of that model.

ScrumPy also interfaces with bioinformatic databases. The “Bioinf” package contains “PyoCyc”, a package that stores tools for conveniently interfacing with the biochemical database BioCyc (see chapter 2 for information on biochemical databases), which includes data on reactions, compounds, enzyme commission numbers, gene associations, pathways and pathway annotations. The command to load this database and for example load the reactions in a variable:

```
import ScrumPy.Bioinf.PyoCyc
db = PyoCyc.Organism('Awoodii')
```

where ‘import’ directs python to import the PyoCyc module, ‘Organism’ is a function in the PyoCyc module that loads the database indicated in parenthesis, which in this case is an *A. woodii* database. “db” then is a variable that stores the *A. woodii* database.



```
File Edit Format Run Options Window Help
Structural() #directive

# transporters
CO2_tx:
  x_CO2 <> CO2
  ~

H2_tx:
  x_Hydrogen <> Hydrogen
  ~

Formate_tx:
  x_Formate <> Formate
  ~

# internal reactions
FormateDehydrogenase_RXN:
  CO2 + Hydrogen <> Formate + PROTON
  ~

FormylTHFSynthase_RXN:
  Formate + ATP + THF -> Form_THF + ADP + Pi
  ~

MethylTHFCyclohydrolase_RXN:
  Form_THF + PROTON <> Methenyl_THF + H2O
  ~
|

Ln: 29 Col: 0
```

Figure 1.6: **A screenshot of a snippet of a ScrumPy model file.** Lines beginning with a hash mark are comments and ignored by ScrumPy.

Another widely available and used modelling software is the COnstraint-Based Reconstruction and Analysis (COBRA) toolbox which is available in MATLAB, Python and Julia [54]. It is a mature structural modelling software with a wide range of modelling tools. ScrumPy however has the advantage of using the same framework to investigate kinetic models as well as structural and so allows for the scope of structural models to be extended to incorporate kinetic analysis. ScrumPy also allows complete access to lower level data structures while effectively hiding them when unnecessary to view. The syntax and format of ScrumPy files is also very human friendly and readable.

1.6 Aims and Thesis Outline

Given the topics discussed in the introduction then, the three primary aims of this project are:

1. Gain fundamental insights into acetogenic metabolism using metabolic modelling.
2. Construct, validate and analyse a Genome-scale Metabolic Model of *A. woodii*.
3. Direct *in silico* metabolic engineering of an industrially relevant *A. woodii* strain.

The aims reflect the nature of this project in that it is both theoretical and experimental in scope, and is concerned with gaining fundamental insights into a group of commercially relevant bacteria that has yet received little interest in way of systems-wide analysis. With this in mind, the hope is to be able to contribute to designing the upstream section of a commercial process for the production of sustainable commodity chemicals using engineered bacteria. The series of objectives tackled to achieve the stated outcomes underlies the structure of most of this thesis. The structure is as follows:

Chapter 1 provides background on the societal context of this project and the direction work undertaken here will take. It also introduces material to the various streams of this work including the computational, mathematical and biological aspects of it.

Chapter 2 deals with the methodology of constructing, validating and analysing the various models developed throughout the project.

Chapter 3 presents a metabolic model of the central metabolic pathway employed by acetogens, the Woods-Ljungdahl pathway, and an extension of it to encompass proximal catabolic pathways. Analysis of it attempts to provide explanations for conflicting reports on *A. woodii* metabolism.

Chapter 4 presents the construction of a Genome-scale Metabolic Model of *A. woodii*. Its curation is detailed as well as its general features.

Chapter 5 investigates bioenergetic aspects of *A. woodii* metabolism. A way of consistently calculating maximum ATP yields is presented and its use to construct a stand-alone bioenergetic core metabolic model.

Chapter 6 is devoted to experimental laboratory-based investigations. Characteristics of *A. woodii* growth are detailed as well as their use for further developing the analysis of metabolic models.

Chapter 7 primarily deals with surveying the feasibility of producing relevant compounds using metabolically engineered *A. woodii*, and suggests engineering approaches for this. Both small and large models developed in this project were used for investigations in this chapter.

Chapter 8 provides conclusions on the various streams of results discussed in previous chapters. An outlook for future work is also provided.

Chapter 2

Computational Methods

In this chapter, the general computational methods used to construct and analyse the metabolic models in this these will be presented. The process for building models followed the pattern used in our research group for other models and other research groups adopt a broadly similar approach. Both construction and analysis of the model was done using ScrumPy (see Introduction, section 1.5).

2.1 Building and Analysing Models

The model work cycle consisted of broadly four steps with two feedback loops: defining the system being modelled by collecting and modifying reactions, reviewing the model for general network connectivity and violations of fundamental physical laws conservation, and validating the model against phenotypic observations. During the model review or validation step, if a flaw was discovered in the model, modifications were made to the definition of the model. That ended the build process, after which the model was ready for further analysis (see Figure 2.1 for schematic).

2.1.1 Databases

BioCyc is a database collection of microbial genomes annotated with metabolic pathways [66]. Each annotation is called a Pathway/Genome Database(PGDB), and these are built automatically using a metabolic pathway prediction algorithm called PathoLogic [65]. In addition to mapping reaction pathways to genomes, PathoLogic also has a gap-filling algorithm that will insert reactions from a pathway into a PGDB if a threshold of other reactions in that pathway has been met. This leads to more complete metabolic annotations but can also erroneously add false-positive reactions. The PGDBs vary in how complete there annotations are and are categorised into three tiers based on this. Tier 1 PGDBs have received the most curation and had at least a years worth of curation from a scientist after

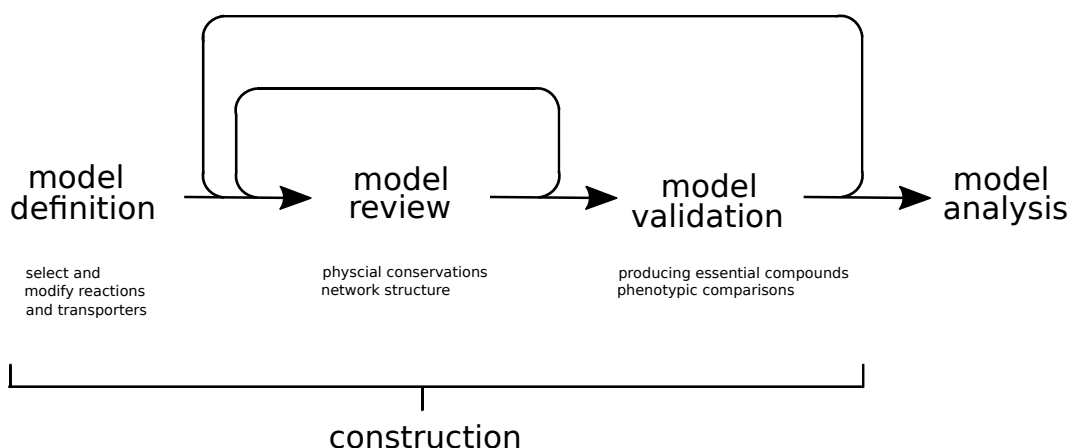


Figure 2.1: **The work cycle for models from construction to analysis.** With each successive model review and validation, changes are made to reactions and transporters of the model until the model is in a satisfactory state. The small text describes what is involved in the construction steps.

their automatic generation from PathoLogic. Tier 2 PGDBs receive on the order of months worth of curation after automated generation. Tier 3 PGDBs have received no curation have automated generation. The *A. woodii* PGDB (referred to as ‘awoodiiCyc’ from hereon) is tier 3 and was accessed via the website. The database flat file for this was also available (registration required) and awoodiiCyc20.5 was downloaded for use in building the *A. woodii* Genome-scale Model.

MetaCyc is a PGDB that has been generated using organisms from all domains of life and so is a compilation of all known metabolic reactions and pathways [19]. This database flat file was also available (registration required), and MetaCyc 18.5 was downloaded for use in building all models.

PRofils pour l’Identification Automatique du Métabolisme (PRIAM) is another tool used to annotate genomes with metabolic information [22]. It is a method for automated enzyme detection in a fully sequenced genome, based on all sequences available in the ENZYME database. It specialises in detecting enzymes based on detailed information on known enzymes and storing these as enzyme profiles. Using position-specific score matrices, these enzyme profiles are used to detect enzymes in nucleotide sequences. The profile database from March 2015 was downloaded (from <http://priam.prabi.fr>) and used for additional model curation. The threshold set for a positive PRIAM hit was a sequence profile proportion of 0.8 and an e-value of 1×10^{-30} .

BRaunschweig **E**Nzyme **D**Atabase (BRENDA) is a database collection of enzyme information [15]. All information is manually curated from the scientific literature, and stores much information on detailed enzyme kinetics, structure, optimal conditions, substrates, ligands, products and thermodynamics.

NCBI Reference Sequence (RefSeq) is a comprehensive database containing publicly available non-redundant genomic data, transcripts and proteins [98]. The database is built and maintained by NCBI. The genome sequence for *A. woodii* DSMZ1030 and its protein sequences were accessed on 4/12/2018. The protein sequences were used as inputs to PRIAM analysis, and for analysis of the *A. woodii* amino acid content.

2.1.2 Model System Definition

Of the steps required to build a model, this step is the most varied as most new models attempt to capture new systems or extend models of systems previously studied. Reactions in models described in this work derive from the databases GenBank, MetaCyc, BioCyc (specifically awoodiiCyc) and BRENDA. The tool PRIAM was used for additional genome annotation given its specificity for enzyme detection. BRENDA was used to assign reaction direction in problematic and ambiguous reactions and also in limited cases to identify enzymes based on substrate ligands. In limited cases eQuilibrator was used to assign reaction directionality.

As acetogens have relatively uncharacterised metabolism, throughout the course of this work novel enzymes catalysing novel reactions were reported in the literature and added manually to the model as these were often missing in databases and make important contributions to the model.

When modifications to a model were necessary because of additional data or on discovery of a flaw the modifications made were below:

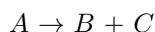
- changing the reversibility of a reaction
- changing the stoichiometry of a reaction
- removing or adding a reaction
- modifying metabolite names
- defining a new metabolite name
- adding or removing a transporter

Transporters for known growth substrates and excreted end-products were added to models manually. MetaCyc IDs were used as metabolite names. A non-MetaCyc metabolite was defined in very limited cases when it was not available in MetaCyc, which was

the case with for example an *A. woodii* specific peptidoglycan monomer. Protons and water were considered as external metabolites. Metabolites with undefined atomic formulae such as ‘Long-Chain-Polyphosphate’ and ‘Aldehydes’ and non-metabolic compounds such as charged tRNAs e.g. ‘TYR-tRNAs’, ‘MET-tRNAs’, ‘Peptides-with-Leader-Sequence’ that are included in MetaCyc/BioCyc databases are replaced with specific instances of those metabolites or removed completely, and so not found in any of the models here. The inclusion of this type of metabolite often leads to stoichiometric inconsistency in the network. This was first described by [46], who also developed an algorithm for automatic detection of such inconsistencies that does not depend on knowing atomic formulae of compounds. Essentially, stoichiometric inconsistency can be explained by viewing it as statements about a reaction stoichiometry. The following reaction:



shows both that A is catalytically converted to B but also that the atomic formulae of A is equal to B . So, if another reaction like the following exists:



then both reactions are not consistent with each other as A must now equal both B and C . The only way both reactions would be consistent with each other is if C had a mass of zero which would preclude it from being a member of the stoichiometric equation in the first place. Stoichiometric inconsistent reactions are found by analysing the left nullspace of the stoichiometric matrix and is a useful tool for model curation.

Small Models

The reactions for small metabolic models (typically less than a hundred reactions and generally of central carbon metabolism), were gathered either by manual curation from the literature, or extracted from a larger model. Depending on the analysis, small models were extended to include heterologous reactions. When analysing the impact of producing novel compounds for example, the heterologous reactions needed to produce that compound as well as a transporter for it were included as model extensions.

Large Model - Genome-scale metabolic model

Large metabolic models are typically more than a hundred reactions and can number in the thousands of reaction. The large model constructed in this work was the genome-scale metabolic model (GSM). Reactions in the GSM primarily derive from awoodiiCyc.

Biomass Composition To model growth of *A. woodii* using a GSM, biomass precursors are synthesised and exported in proportions equal to average cell. Additionally, an

energetic demand is required to meet the maintenance and growth cost. Transporters for these biomass precursors were added to the model.

The precursor proportions would ideally be specific to *A. woodii*, but some biomass constituents have not been measured and reported. The major macromolecular proportions of DNA, RNA, amino acids, cellular membrane and wall, and other cofactors have been adopted from *Bacillus subtilis*, a model gram-positive bacteria [26], in weight/weight %. DNA content was determined from the *A. woodii* genome sequence and contained 30.5% adenine, 30.2% thymine, 20.0% guanine and 19.3% cytosine. RNA was split into equal proportions of each base. Detailed data was available for cell wall components [62]. The overall composition of cell wall is 55% murein (peptidoglycan), 30% polysaccharide (techoic acids) and 15% ash. The peptidoglycan and techoic acid percentages were adjusted to 100% equally. The acid hydrolysate of isolated cell walls contained Glc:Orn:Ala:Ser:Mur:GlcNH₂:NH₃, in molar ratios of 1:2:0.6:1:1:1:0.3 and was used to determine *A. woodii* as having a group B type peptidoglycan using ornithine as an interpeptide bridge (see Figure 2.2 for cell wall schematic). The techoic acid component was mainly rhamnose, smaller amounts of glucose and galactose and small but significant amounts of glycerol. The membrane lipid content was taken from the gram positive model bacteria *Bacillus megaterium* [152], and palmitate used as the representative fatty acid. Amino acids, cofactors and soluble compounds were taken from *Bacillus subtilis* [103], but modified with data on the reported corrinoid content of *A. woodii*. The complete breakdown of the biomass composition used can be found in the appendix C.

Maintenance Costs An ATPase reaction was added to the model. During analysis, this reaction was constrained to carry a fixed flux corresponding to measured maintenance costs. The way in which maintenance costs were calculated are shown in chapter 5.

Genome-scale model organisation The model files are modular and are organised in a hierarchical structure, where the top level module includes lower level modules (see Figure 2.3). The modules are organised in the following way:

Awoodii.spy top level module which calls lower level modules.

FromBuildWoodii.spy reactions that derive from awoodiiCyc and have been parsed and processed by ScrumPy.

Biomass.spy transporters of biomass precursor metabolites.

Extras.spy extra reactions that do not appear in awoodiiCyc but have been added to the model, and include the ATPase reaction.

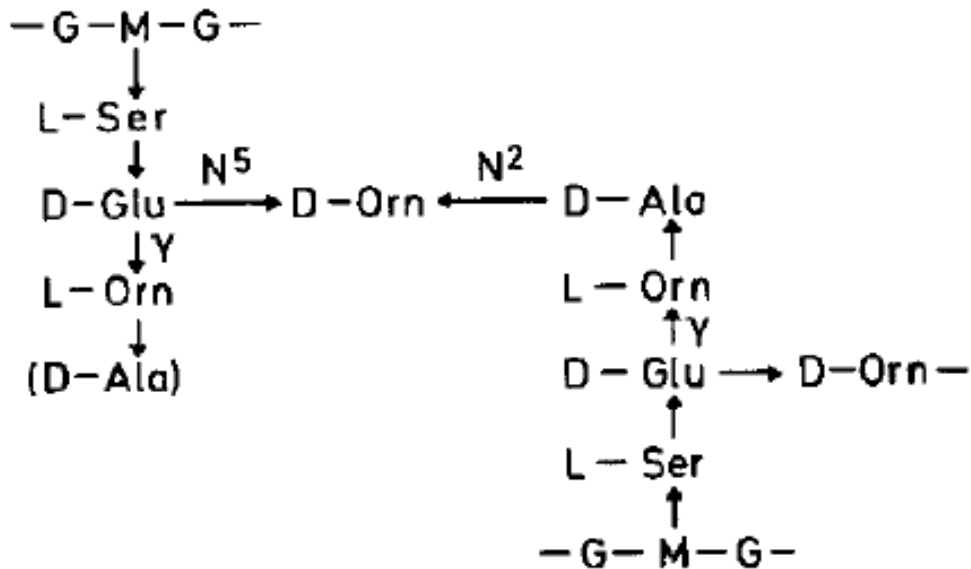


Figure 2.2: Schematic of the primary structure of a fragment of *A. woodii* peptidoglycan, reproduced from [62]. (G) is N-acetylglucosamine and (M) is N-acetylmuramic acid. The other compounds are amino acid abbreviations. Bracketed D-Alanine is missing in most cases.

ETC_core.spy a manually written set of reactions which represent the energy-transducing reactions of *A. woodii*.

Transporters.spy a set of all transporters other than biomass transporters used in the model. This includes transporters for all known growth substrates of *A. woodii*, other necessary growth media components, known by-product excretions, and passive diffusion of small molecules.

ExtraSpontaneousReactions.spy additional reactions which do not require enzymes.

2.1.3 Model Review

Before models could be meaningfully analysed, they were reviewed for mass and energy conservations and their network structure. The way in which the models were reviewed was with manual curation but also with analytical tools such as stoichiometric inconsistency tests (see section 2.1.2 for description of this). If upon review modifications to the model were necessary, one of the options outlined in section 2.1.2 were used.

```

#
##
#### Top level model build
##
#

Structural() # ScrumPy directive specifying model type
External("PROTON","WATER") # specifies external metabolites

Include(FromBuildWoodii.spy,
        Biomass.spy,
        Extras.spy,
        ETC_Core.spy,
        Transporters.spy,
        ExtraSpontaneousReactions.spy
)

```

Figure 2.3: **The top level Genome-scale Model file, Awoodii.spy.** In it are the ScrumPy directives that specify the model type and metabolites treated as external. It uses the ‘Include’ directive to assemble the other GSM files into one large composite model.

Physical Conservations Review

Mass Balancing The atoms carbon, nitrogen, phosphorus, sulphur and sodium are mass balanced in all models. Elements incorporated into cofactors such as iron (in ferredoxin) and cobalt (in cobalamin) are also mass balanced. If a reaction could not be atom balanced it was removed from the model. No mass balanced atom can be synthesised and exported in the absence of import of suitable transporters.

Frequently, unbalanced reactions included those with polymers of undefined length (and so undefined atomic formulae). For polymers of saccharides this was overcome by renaming it glycogen, assigning it a stoichiometry of a single glucose, and increasing the length of it by increasing its stoichiometric coefficient as was done in [94, 97].

Energy Balances Unconserved energy producing cycles were broken by changing the reversibility of reactions using data on the physiological operation of the enzyme in the same or similar organisms. This was done using BRENDA or in limited cases the thermodynamic feasibility of a reaction operating in a certain direction was computed with eQuilibrator. ATP cannot be synthesised with internal cycles. NAD(H), NADP(H) and ferredoxin cannot be reduced or oxidised with internal cycles.

Network Structure Review

Stoichiometric inconsistency . Once reactions were mass balanced, no stoichiometrically inconsistent metabolites were found except those relating to external metabolites protons and water, namely: hydrogen peroxide, oxygen, water, protons, and molecular hydrogen. As mentioned in the chemical conservations review, none of the internal metabolites (excluding water and protons) can be synthesised and exported without appropriate substrate import.

Reaction subsets analysis was used identify the occurrence of blocked sets of reactions due to inconsistent reaction directionality. Ten inconsistent reaction subsets were found in the initial genome-scale model. For increased network connectivity, these reactions were unblocked by modifying reaction directionality when possible, or removing reactions. After this was done, one inconsistent subset with two reactions involved in carbon tetrachloride catabolism remained. This was left in the model, for potential future extension of the model to include carbon tetrachloride metabolism.

2.1.4 Model Validation

This final step in the build process ensures the model reflects *in vivo* phenotypic observations. Validations can be split into internal modelling validations and direct comparisons to laboratory data. Both reflect *in vivo* observations.

The modelling validation done was to assess the import and export of metabolites known to be transported by *A. woodii* in an atomically and energetically balanced way. Data on *A. woodii* growth experiments from the literature was collected and compared to model analysis for the validations concerning comparisons to laboratory data and will be discussed in relevant sections.

Import and export of metabolites The catabolism of imported substrates was checked. If model analysis showed reactions for catabolism were missing, these were added where available from the literature. Models could not import substrates such as fructose and CO₂ unless appropriate export of other metabolites was allowed. The capacity of each model to export metabolites known to be produced by *A. woodii* was also checked. Models could not export end-products such as acetate unless appropriate import of other metabolites was allowed.

If models over or underestimated the metabolic capability of *A. woodii*, this was usually the result of either one of the issues below:

False positive gap-filled reactions These can be incorrectly inserted into the model by the sometimes overly liberal BioCyc PGDB gap-filling algorithm. In the pre-curated GSM for example, the pantoate beta-alanine ligase reaction was in the model. The

inclusion of this reaction allowed for pantothenate auxotrophic growth, which is in disagreement with reported phenotypes [6]. This was identified as a false positive gap-filled reaction by finding no gene annotation associated with the enzyme catalysing the reaction and so was removed.

Gaps in the metabolic network Conversely to false positive reactions, false negative reactions also occur. These were mentioned in the import and export of metabolite section for example, but these are not the only cases. When evidence for reactions in a pathway is reported to carry flux, such as reactions in the Woods-Ljungdahl pathway, but analysis of the model does not allow for that pathway to carry flux, either additional reactions were added or already present reactions were modified to align with the phenotype. An example for indirect evidence of this is induced expression of an enzyme in the presence of substrates catalysed by it along with protein purification and characterisation, and direct evidence would be radiolabelled ^{13}C tracer analysis.

2.2 Model Analysis

The theoretical basis for flux balance analysis was covered in the introduction (section-structural modelling). Using ScrumPy, the stoichiometric matrix, right and left nullspace analysis and elementary modes analysis were all used in various investigations.

Some frequently used LPs are described below. LPs for various analysis were built from base LPs using a Python module called WoodiiLP.spy. This was done with the aim of maintaining consistency in transporter constraints across analyses and to reduce code redundancy. Additional constraints were then layered on top of base LPs to tailor the LP to more specific investigations. All LPs here apply an objective coefficient weighting of one and not explicitly stated in equations below. The default transporter constraints were as follows:

$$\begin{aligned} v_{J_S} &= t_{J_{F...H}}; J_{F...H} \cdot 0 \\ v_{J_E} &= t_{J_{I...K}}; -100 \leq J_{I...K} \leq 0 \end{aligned} \quad (2.1)$$

where $J_{F...H}$ are substrate transporters used as carbon and energy sources such as fructose and methanol and $t_{J_{F...H}}$ is a vector of these transporters set to zero. $J_{I...K}$ are all transporters that are not carbon/energy sources or biomass transporters. This includes by-products like acetate and propionate, exchange metabolites like CO_2 and ethanol, and other transporters such as micronutrients and nitrogen sources and $t_{J_{I...K}}$ is a vector of these transporters with bounds allowing for export but not import.

The two base LPs are:

$$\text{subject to } \begin{cases} \mathbf{N}\mathbf{v} = \mathbf{0} \\ v_J = v_{J_S}, v_{J_E} \end{cases} \quad (2.2)$$

where $\mathbf{N}$ is the stoichiometry matrix and $\mathbf{v}$ a vector of reaction rates and:

$$\begin{aligned} & \text{minimize} && : \sum |v_i| \\ & \text{subject to} && \begin{cases} \mathbf{N}\mathbf{v} = \mathbf{0} \\ v_J = v_{J_S}, v_{J_E} \end{cases} \end{aligned} \quad (2.3)$$

where $\sum |v_i|$ is the absolute sum of fluxes, which is set to a minimisation objective and reflects the minimisation of enzyme synthesis of the network as a whole. Both LPs allow for as much flexibility as possible in transporter fluxes, while maintaining biologically feasible analysis. The difference between these base LPs is that LP 2.2 has no objective set. This was used with analysis that needed more flexibility in objective setting such as that carried out in analysis detailed in the bioenergetics chapter (chapter 5) for $Y_{ATP}^{\max}$ calculations.

Another regularly used LP is one that is used for microbial growth analysis by computing biomass production:

$$\begin{aligned} & \text{minimize} && : \sum |v_i| \\ & \text{subject to} && \begin{cases} \mathbf{N}\mathbf{v} = \mathbf{0} \\ v_{ATPase} = r_{ATP} \\ v_{J_B} = t_{J_{P...R}}; J_{P...R} \cdot \mu \\ v_J = v_{J_S}, v_{J_E}, v_{J_B} \end{cases} \end{aligned} \quad (2.4)$$

where r_{ATP} is a ATP hydrolysis reaction set to a rate equivalent to the maintenance rate of *A. woodii* in $\text{mol}(\text{g-DW} \cdot \text{h})^{-1}$, μ is a growth rate, $J_{P...R}$ is a vector of biomass transporters set to rates as found in appendix D and $t_{J_{P...R}}$ is this vector scaled by a growth rate. For the above LP to compute biomass production, the carbon and energy source of choice for the particular analysis is unconstrained as well as the essential micronutrients necessary for growth. The LP is then solved.

When analysing the model for metabolite synthesis, an additional constraint was set on top of 2.3 resulting in:

$$\begin{aligned} & \text{minimize} && : \sum |v_i| \\ & \text{subject to} && \begin{cases} \mathbf{N}\mathbf{v} = \mathbf{0} \\ v_{J_M} = t_{J_M}; J_M \leq -1 \\ v_J = v_{J_S}, v_{J_E}, v_{J_M} \end{cases} \end{aligned} \quad (2.5)$$

where J_M is the transporter of a metabolite being exported, and t_{J_M} is this transporter set to export -1 flux unit. Again, the appropriate transporters for the substrates are unconstrained before solving.

An application of FBA presented by Poolman et al. [97] was used in chapters 7, 6. It involves collecting a set of solutions generated by incrementally increasing constraints on a single reaction. The precise formulation of these reaction scans are shown in the relevant chapters.

Chapter 3

A Metabolic Model of the Wood-Ljungdahl Pathway

3.1 Introduction

To gain insights into the Wood-Ljungdahl pathway (WLP) using metabolic modelling, a structural model was constructed and analysed. This was manually constructed and consisted of reactions from the WLP particular to *A. woodii*. This is the only example of a WLP that has every reaction stoichiometry completely defined. Cofactor utilisation for WLP in other acetogens still remains unknown in terms of NAD(P)H and ferredoxin. Also, high resolution data on the transmembrane flow between Na^+ per turn of the ATPsynthase C ring has been measured Fritz et al. [41], allowing for complete accounting of ATP formation from sodium gradients. Additionally, the growth patterns of *A. woodii* on various gaseous mixes and other substrates have been reported allowing for comparisons with model analysis. Specific gases of interest to this project are $\text{CO}_2\text{-H}_2\text{-CO}$ mixtures since this is found in synthesis gas (as discussed in the Introduction, chapter 1).

However, to date, there has been no metabolic modelling of *A. woodii*'s WLP. Furthermore, several growth phenotypes still lack a coherent metabolic explanation. For example, contradictory reports exist concerning CO metabolism in *A. woodii*. Genthner and Bryant [45] report it as a viable sole growth substrate, but this was not reproduced by Ma et al. [78] and Bertsch and Müller [8]. Ma et al. [78] and Bertsch and Müller [8] did however report growth on $\text{CO}_2\text{-CO-H}_2$ mixes, and Ma et al. [78] additionally reported that carbon from CO was found in acetate using ^{14}CO . Diekert et al. [28] reported that acetate could not be produced from CO in a resting cell assay, but CO could be formed and utilised by resting cells, whereas NMR studies by Kerby et al. [67] showed that carbon from CO was incorporated into methyl and carbonyl groups of acetate.

Bertsch and Müller [8] made the latest attempt to resolve the matter with the most com-

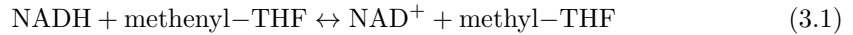
prehensive experiments on CO as sole and co-substrate. The key findings were that CO₂-CO-H₂, and CO-formate support growth of *A. woodii* and during growth, CO is utilised as a carbon source. CO as a sole substrate, CO₂-CO, fructose-CO and methanol-CO did not support growth. The growth observed using CO₂-CO-H₂ was biphasic, with the first phase being slow and the second fast. H₂ consumption rates increased when CO levels were below 3% in the gas phase. Resting cell assays also showed acetate production from CO as sole substrate and as a co-substrate with CO₂-H₂ is possible although these production rates were much slower than with CO₂-H₂ alone. The authors offer explanation for some of these observations by invoking studies on CO-dependent inhibition of the hydrogen-dependent CO₂ reductase (HDCR), which produces formate for the methyl-branch of WLP. The electron bifurcating hydrogenase (EBHyd), which can provide reducing equivalents for WLP by stripping electrons from H₂, is also inhibited by CO, but the authors argue this enzyme is unlikely to cause the observed effect because reducing equivalents can be generated by other enzymes.

To explore feasible steady states of the WLP with several gas mixes and mixotrophic substrates of CO with either formate, fructose or methanol in light of the reports outlined above, elementary modes analysis was carried out. Particular attention was directed to CO utilising modes, their potential relationship to the experimental observations, and the actions of HDCR and EBHyd in relevant modes.

3.2 Model Construction and Curation

3.2.1 Construction

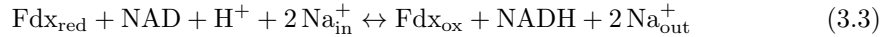
A literature review was undertaken on reactions in the WLP in acetogens and *A. woodii* specific WLP reactions. These were compiled manually into a ScrumPy formatted file. The reactions uniquely characterised in *A. woodii*'s WLP are the methenyl-THF reductase that uses NADH as reductant and has a stoichiometry of:



the HDCR with a stoichiometry of:



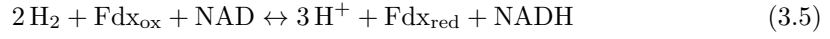
the Rnf complex, where Na^+_{in} specifies internal sodium and Na^+_{out} is external sodium:



the sodium-dependent ATP synthase:



and the EBHyd:



To represent ATP consumption, an irreversible ATP hydrolysis reaction was added to the model. In general the curation steps described in chapter 2 were used.

3.2.2 Review and Validation

There are no stoichiometric inconsistencies in the model and it is energy and mass balanced for carbon, nitrogen, sulphur, hydrogen, oxygen, phosphorus and sodium. Equation 2.5 was used to ensure acetate can be produced from $\text{CO}_2\text{-H}_2$ as reported in the literature. In general the review and validation steps described in chapter 2 were used.

3.2.3 Model Properties

A schematic of the model is shown in Figure 3.1. The model has 16 internal reactions, 27 metabolites and 7 transporters. Five transporters are bidirectional exchange reactions for CO_2 , H_2 , CO , H^+ and H_2O and two transporters export acetate and ethanol. The model has four reaction subsets, the largest of which is the methyl-branch-Acs subset (R5, R7, R8, R9, R12, R13 and R14, orange arrows). The acetate producing subset consists of phosphate acetyltransferase, acetate kinase and the acetate transporter (R10, R16 and Ac.tx, blue arrows), the ethanol producing subset consists of acetaldehyde dehydrogenase, alcohol dehydrogenase and the ethanol transporter (R11, R15 and Etoh.tx, green arrows), and the final subset is the Rnf-ATPsynthase subset consisting of the Rnf complex reaction, and the ATP synthase (R1 and R2, red arrows). Reactions catalysed by CODH (R3), EBHyd (R4), ATPase (R6) are not in any subsets, and there are no dead reactions.

Since reactions in a subset will carry a fixed proportional flux in any steady state solution, reactions in a subset can be condensed into a single net stoichiometry. The net stoichiometry of the Rnf complex (equation 3.3) and ATPsynthase (equation 3.4) for example is:



so for every five reduced ferredoxin (Fdx_{re}) oxidised by the Rnf-ATPsynthase subset, three ATP and three NADH are formed. This effectively makes NAD the electron acceptor when the sodium gradient is generated by Fdx_{re} . This subset can also operate in the reverse direction where Fdx_{ox} would be the electron acceptor when Na^+ gradients are generated by ATP and NADH. In this case three NADH and three ATP, form one Fdx_{re} (equation 3.6).

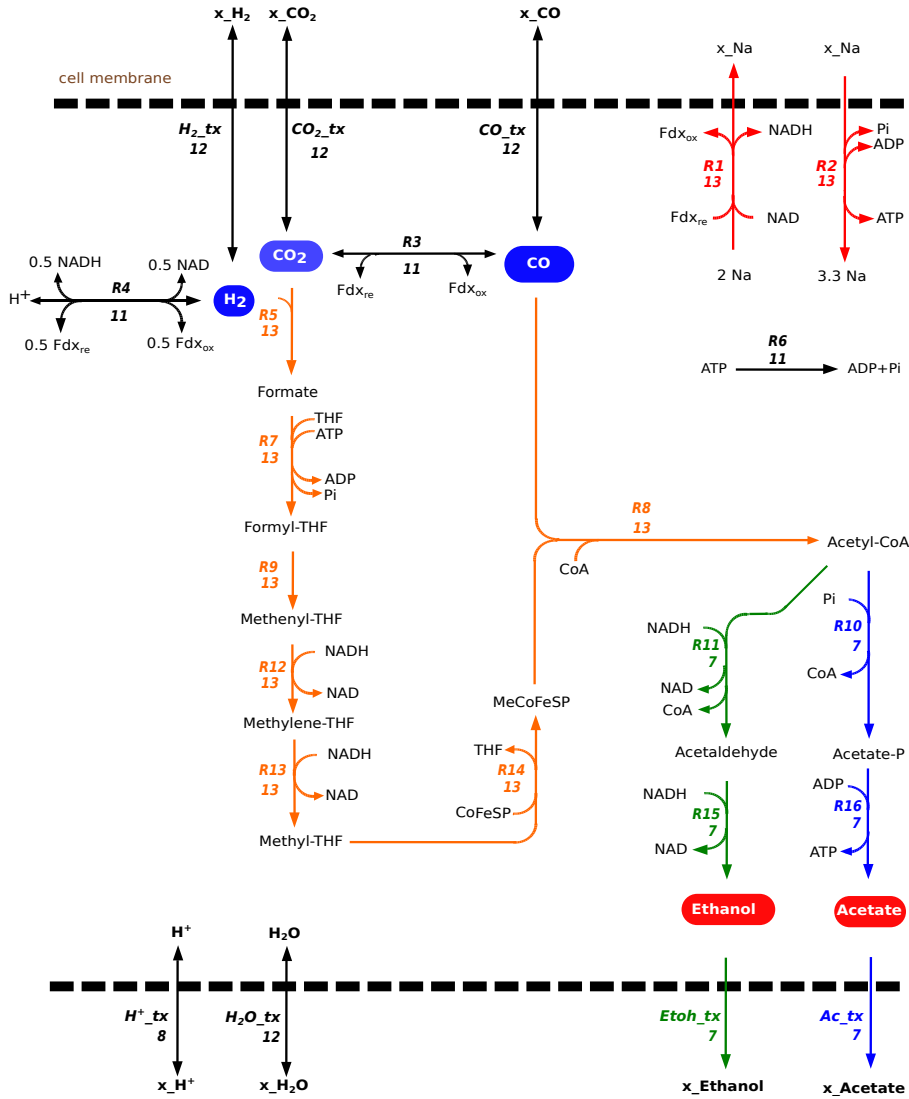


Figure 3.1: **A structural metabolic model of the Wood-Ljungdahl pathway.** Metabolites with blue backgrounds have exchange transporters and those with red backgrounds have export transporters. Dashed lines represent the cell membrane and so metabolites inside these are internal metabolites. The colour of non-black arrows represents membership to one of the four reaction subsets. Numbers under reaction labels indicate the number of elementary modes that reaction is involved in. **R1**:Rnf complex, **R2**:ATP synthase, **R3**:carbon monoxide dehydrogenase (CODH), **R4**:bifurcating hydrogenase (EBHyd), **R5**:hydrogen-dependent formate dehydrogenase (HDFD), **R6**:ATP hydrolase, **R7**:formate-THF ligase, **R8**:acetyl-CoA synthase, **R9**:methenyl-THF cyclohydro-lase, **R10**:phosphate acetyltransferase, **R11**:acetaldehyde dehydrogenase, **R12**:methylene-THF dehydrogenase, **R13**:methyl-THF reductase, **R14**:methyl-THF:corrinoid/iron-sulfur methyltransferase, **R15**:alcohol dehydrogenase, **R16**:acetate kinase, **Ac_tx**:acetate transporter, **Etoh_tx**:ethanol transporter.

Table 3.1: Elementary modes of gaseous exchange in the Wood-Ljungdahl pathway. The number of reactions involved in each mode, the Y_{ATP} achieved, and its schematic representation in a Figure is indicated. Y_{ATP} is given in $\text{mol}\cdot\text{Cmol}^{-1}$, and is rounded to two decimal places.

Mode	Stoichiometry	No. of Reactions	Y_{ATP}	Figure
1	$2 \text{ CO}_2, 4 \text{ H}_2 \rightarrow \text{Ac}, 2 \text{ H}_2\text{O}, \text{H}^+$	19	0.15	3.2 [A]
2	$26 \text{ CO}_2, 72 \text{ H}_2 \rightarrow 3 \text{ Ac}, 10 \text{ EtOH}, 36 \text{ H}_2\text{O}, 3 \text{ H}^+$	19	0.00	3.2 [B]
3	$3 \text{ CO}_2, 5 \text{ H}_2 \rightarrow \text{Ac}, \text{CO}, 3 \text{ H}_2\text{O}, \text{H}^+$	15	0.00	3.2 [C]
4	$4 \text{ CO}, 2 \text{ H}_2\text{O} \rightarrow \text{Ac}, 2 \text{ CO}_2, \text{H}^+$	18	0.38	3.3 [A]
5	$\text{CO}, \text{H}_2\text{O} \rightarrow \text{H}_2, \text{CO}_2$	9	0.30	3.3 [B]
6	$6 \text{ CO}, 3 \text{ H}_2\text{O} \rightarrow \text{EtOH}, 4 \text{ CO}_2, \text{H}^+$	18	0.29	3.3 [C]
7	$\text{CO}_2, \text{CO}, 3 \text{ H}_2 \rightarrow \text{Ac}, \text{H}_2\text{O}, \text{H}^+$	19	0.30	3.4 [A]
8	$\text{CO}_2, \text{CO}, 5 \text{ H}_2 \rightarrow \text{EtOH}, 2 \text{ H}_2\text{O}$	18	0.11	3.4 [B]
9	$17 \text{ CO}_2, 3 \text{ CO}, 57 \text{ H}_2 \rightarrow 10 \text{ EtOH}, 27 \text{ H}_2\text{O}$	18	0.00	3.4 [C]
10	$2 \text{ CO}, 2 \text{ H}_2 \rightarrow \text{Ac}, \text{H}^+$	18	0.45	3.5 [A]
11	$3 \text{ CO}, \text{H}_2, \text{H}_2\text{O} \rightarrow \text{Ac}, \text{CO}_2, \text{H}^+$	19	0.40	3.5 [B]
12	$5 \text{ CO}, \text{H}_2, 2 \text{ H}_2\text{O} \rightarrow \text{EtOH}, 3 \text{ CO}_2$	18	0.29	3.5 [C]
13	$3 \text{ CO}, 3 \text{ H}_2 \rightarrow \text{EtOH}, \text{CO}_2$	18	0.27	3.6 [D]
14	$2 \text{ CO}, 4 \text{ H}_2 \rightarrow \text{EtOH}, \text{H}_2\text{O}$	18	0.26	3.6 [E]

When considering the operation of reaction subsets, the WLP network of 16 reactions can be reduced to seven configurable metabolic units made up of the four reaction subsets and three independent reactions.

3.3 Elementary Modes Analysis

Elementary modes analysis revealed 14 modes (from hereon “mode” denotes “elementary mode”) in total as shown in table 3.1. The number of modes each reaction is involved in is shown in Figure 3.1. Three use $\text{CO}_2\text{-H}_2$ as inputs, three use CO, three use $\text{CO}_2\text{-H}_2\text{-CO}$ and five use CO-H_2 . No elementary modes were calculated that use CO_2 as sole input, H_2 as sole input or $\text{CO}_2\text{-CO}$ as inputs. The methyl-branch-Acs subset and reactions in the Rnf-ATPsynthase subset carried flux in 13 of the 14 modes. ATP hydrolysis occurred in 11 modes indicating that these are ATP generating modes. The reactions in the least number of modes are those in the ethanol producing subset, and the acetate producing subset, which are involved in 7 modes each. No reaction was involved in all modes. The Rnf subset operates in the ATP forming direction in all modes it is involved in. The only mode this subset is not involved in is mode 3 where acetate and CO are produced from $\text{CO}_2\text{-H}_2$. Across all modes, acetate production allows for higher Y_{ATP} than ethanol production because of substrate-level phosphorylation (SLP) by acetate kinase. However, modes were found that generate ATP without acetate, showing that ATP can be generated without SLP. One mode

does not produce either acetate or ethanol. Instead, it allows for the interconversion between the gases, oxidising CO to CO₂-H₂ while also forming 0.3 mol·Cmol⁻¹ ATP showing that ATP can be generated from Na⁺ gradients absent SLP or ethanol production.

3.3.1 Elementary modes consuming CO₂-H₂

Elementary modes consuming CO₂-H₂ are shown in Figure 3.2 and table 3.1. Mode 1 is the canonical path known for *A. woodii* (see chapter 1) (Fig. 3.2, A), generating 0.15 mol·Cmol⁻¹ ATP. Mode 2 produces ethanol as the major product alongside acetate (Fig. 3.2, B) and no ATP. Mode 3, exports CO alongside acetate (Fig. 3.2, C) and no ATP.

All three produce acetate, but the co-production of either ethanol or CO could be used to differentiate between them. Three different H₂:carbon consumption ratios also reflect the three different modes. Mode 3 consumes the most carbon in a 1.7:1 ratio. Mode 2 consumes an intermediate amount of carbon in a 2:1 ratio and mode 3 consumes the least carbon in a 2.8:1 ratio.

All three modes employ CODH (R2), EBHyd (R3), the methyl-branch-Acs subset and the acetate producing subset, operating them in the same reaction direction. The H₂:CO₂ ratios are reflected by the flux ratio of EBHyd:CODH, which are 1:1, 1.5:1, and 2.3:1 from highest to lowest CO₂ consumption.

Reduction of ferredoxin is solely carried out by EBHyd. NAD is reduced by either Rnf or EBHyd. In mode 1 25% NADH is produced by Rnf and 75% by EBHyd, in mode 2 35.9% is produced by Rnf and 64.1% by EBHyd, and 100% by EBHyd in mode 3.

Oxidation of ferredoxin is by either CODH or Rnf. In mode 1, 33.3% is oxidised by Rnf and 66.6% by CODH. In mode 2 55.9% is oxidised by Rnf and 44% by CODH, and in mode 3 100% of the Fdx_{re} is oxidised by CODH. NADH is oxidised by either by the methyl-branch-Acs subset or the ethanol producing subset. In modes 1 and 3 NADH is solely oxidised by the methyl-branch-Acs subset. In mode 2 56.5% is oxidised by the methyl-branch-Acs subset and 43.4% is the ethanol producing subset.

Since H₂ is the only available electron donor, Fdx_{re} can only be produced by the EBHyd. NADH is also produced by this reaction but given the transhydrogenase activity of the Rnf, NADH can also be produced from Fdx_{re} directly. If NADH is produced without the Rnf, as seen in mode 3, the two NADH needed for the methyl-branch is supplied by the EBHyd. The absence of the Na⁺ gradient results in no ATP generation.

The HDCR (R5) was shown to be able to use Fdx_{re} in place of H₂ (though at a lower rate [113]), so an alternative WLP model was constructed to ask the question whether growth would be possible if only Fdx_{re} was available to the HDCR. In the alternative model the

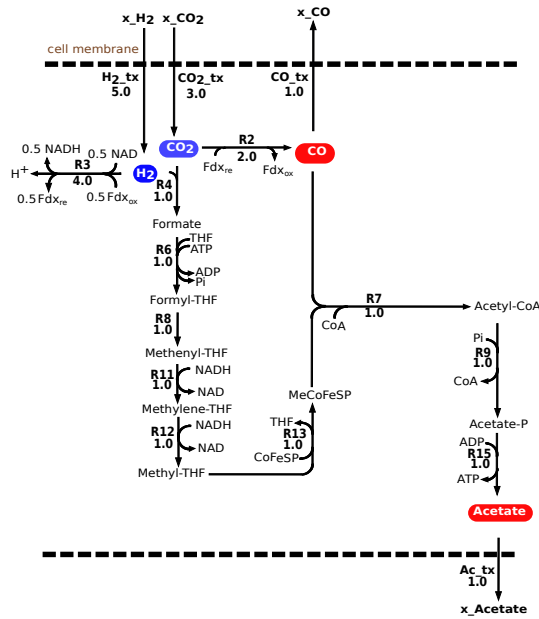
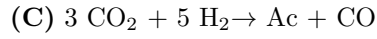
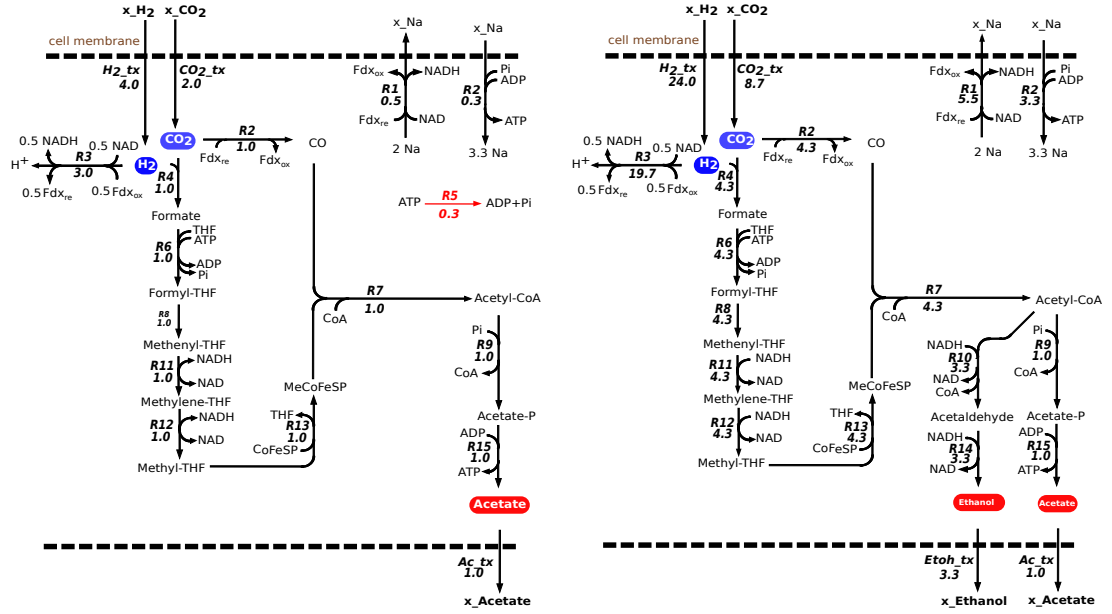
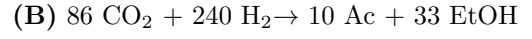
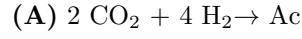
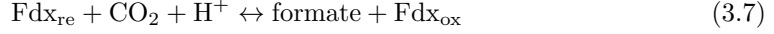


Figure 3.2: Elementary modes with $\text{CO}_2\text{-H}_2$ as inputs. Modes 1, 2 and 3 in table 3.1, correspond to figures A, B and C respectively. Proton and water balance not shown.

HDCR reaction stoichiometry (as seen in equation 3.2) was changed to the stoichiometry in equation 3.7, and the elementary mode analysis repeated.



In this model, ten modes in total were found. One mode was found that used $\text{CO}_2\text{-H}_2$ as input, and three each using CO , $\text{CO}_2\text{-H}_2\text{-CO}$ and CO-H_2 as inputs. The $\text{CO}_2\text{-H}_2$ mode could produce 1 acetate from 2 CO_2 and 4 H_2 but with no ATP. Thus, using Fdx_{re} in place of H_2 to reduce CO_2 would not support growth.

3.3.2 Elementary modes consuming CO

Elementary modes consuming CO (with H_2O) as sole carbon source are shown in Figure 3.3 and table 3.1. Mode 4, exports acetate and CO_2 and produces $0.38 \text{ ATP mol}\cdot\text{Cmol}^{-1}$ (Fig. 3.3A). Mode 5 converts CO into $\text{CO}_2\text{-H}_2$ and produces $0.30 \text{ ATP mol}\cdot\text{Cmol}^{-1}$ (Fig. 3.3B). Mode 6 exports ethanol alongside CO_2 and produces $0.29 \text{ ATP mol}\cdot\text{Cmol}^{-1}$ (Fig. 3.3C).

All modes use CODH, EBHyd, ATPase, the Rnf-Acs subset and operate them in the same direction. Additionally they export at least 50% of the carbon in CO to CO_2 , alongside an additional by-product. 100% of carbon in CO is excreted as CO_2 (with H_2) in mode 5, 66.7% in mode 6 and 50% in mode 4. All modes produce ATP but there is no correlation between CO_2 export and Y_{ATP} . When acetate is produced the highest Y_{ATP} is achieved, and ethanol production results in the lowest. All modes use CODH (R2), EBHyd (R3), and the Rnf subset (R1 and R2) in the same direction. This group of modes are the only ones which use EBHyd to evolve H_2 rather than oxidise it.

Reduction of ferredoxin in all modes is carried out by CODH. NADH is produced by Rnf in all modes.

Oxidation of ferredoxin is carried out by Rnf and EBHyd. In the mode producing acetate (mode 4) 83.3% is oxidised by the Rnf and 16.6% by EBHyd, in the gas conversion mode (mode 5) 50% is oxidised by the Rnf and 50% by EBHyd, and in the ethanol producing mode (mode 6) 90% is oxidised by Rnf and 10% by EBHyd. NADH oxidation is split between the methyl-branch-Acs subset, ethanol producing subset and EBHyd. Mode 4 oxidises 80% of NADH in the methyl-branch reactions and 20% with EBHyd. Mode 5 oxidises 100% of the NADH with EBHyd. Mode 6 uses the methyl-branch reactions to oxidise 44.4% of NADH, the ethanol producing reactions to oxidise 44.4% and EBHyd for 11.1%.

Since no H_2 is available, Fdx_{re} can only be produced by oxidation of CO . While the subsequent CO_2 can be directly exported, the Fdx_{re} cannot and so must be re-oxidised. EBHyd cannot oxidise the Fdx_{re} without NADH, but the Rnf can transhydrogenate Fdx_{re}

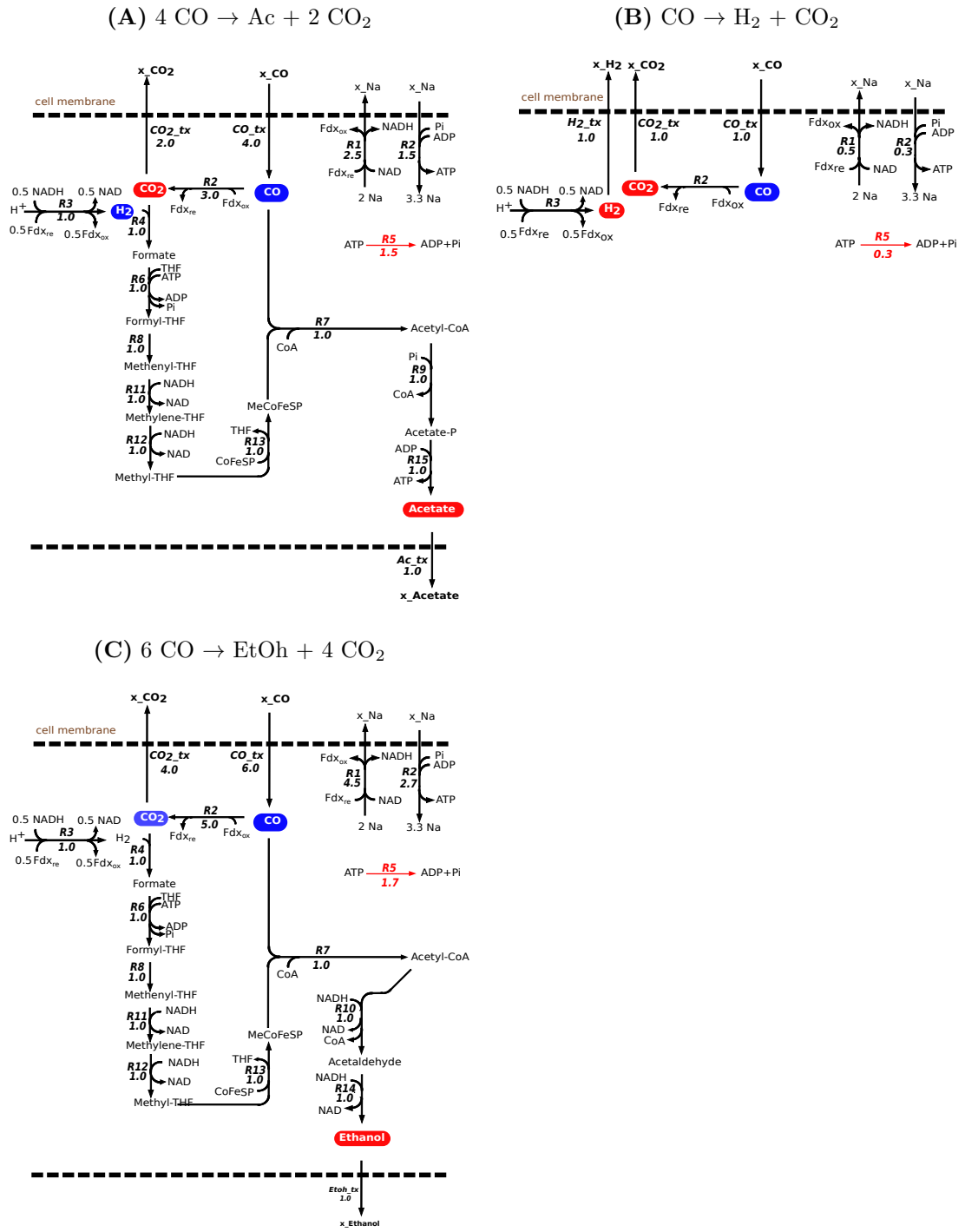


Figure 3.3: **Elementary modes with CO as input.** Modes 4, 5 and 6 in table 3.1, correspond to figures A, B and C respectively. Proton and water balance not shown.

to NADH in a 1:1 stoichiometry. The Rnf then, supplies the NADH for EBHyd using one Fdx_{re} and second Fdx_{re} for EBHyd can be supplied directly by the oxidation of another CO. The subsequent H_2 produced by EBHyd is then exported along with CO_2 as seen in mode 5, or used to reduce some of the CO_2 to formate as seen in modes 4 and 6. The Na^+ gradients in this scenario can be generated using Fdx_{re} directly from CO oxidation instead of H_2 oxidation, and so also without WLP. The joint oxidation of NADH and Fdx_{re} by EBHyd allows electrons to be accepted by H^+ for H_2 evolution, thus generating ATP with or without ethanol and acetate production.

3.3.3 Elementary modes consuming CO_2 -CO- H_2

Modes with CO_2 -CO- H_2 as joint inputs are shown in Figure 3.4. Mode 7 consumes one each of CO_2 and CO with 3 H_2 to produce acetate generating 0.6 ATP, the same amount of ATP as mode 5 and double that of mode 1 (Fig 3.4A). Mode 8 consumes two more H_2 than mode 7 and produces ethanol instead of acetate, generating 0.21 ATP (Fig 3.4B). Mode 9 more CO_2 and H_2 and less CO to produce ethanol than both modes 7 and 8 and generates no ATP (Fig 3.4C).

All modes use EBHyd, the Rnf-ATPsynthase subset and the methyl-branch-Acs subset and operate them in the same direction. No modes are found where more than one carbon-based compound was exported. Each mode has a different H_2 :carbon consumption ratio. Mode 7 has a ratio of 1.5:1, mode 8 is 2.5:1 and mode 9 is 2.85:1.

Reduction of ferredoxin is carried out by EBHyd in all modes. Mode 7 and 8 produce 50% NADH with Rnf and 50% with EBHyd. Mode 9 produces 41.25% NADH with Rnf and 58.75% with EBHyd.

Oxidation of ferredoxin is by both Rnf and EBHyd. 100% of Fdx_{re} is oxidised by Rnf in mode 7 and 8. In mode 9 Rnf oxidises 70.2% Fdx_{re} and EBHyd oxidises 29.8%. NADH is oxidised by the methyl-branch, and ethanol producing reactions. In modes 7 and 8, 100% is oxidised by the methyl-branch. In mode 9, NADH oxidation is split evenly between the methyl-branch and ethanol producing reactions.

3.3.4 Elementary modes consuming CO- H_2

Elementary modes consuming CO- H_2 are shown in Figures 3.5 and 3.6. Mode 10 consumes CO and H_2 to produce an acetate and generates $0.45 \text{ mol} \cdot \text{Cmol}^{-1}$ ATP. Mode 11 exports CO_2 alongside acetate $0.4 \text{ mol} \cdot \text{Cmol}^{-1}$ ATP. Modes 12 and 13 both export CO_2 with ethanol, generating 0.29 and $0.27 \text{ mol} \cdot \text{Cmol}^{-1}$ ATP respectively. Mode 14 exports ethanol and generates $0.26 \text{ mol} \cdot \text{Cmol}^{-1}$ ATP.

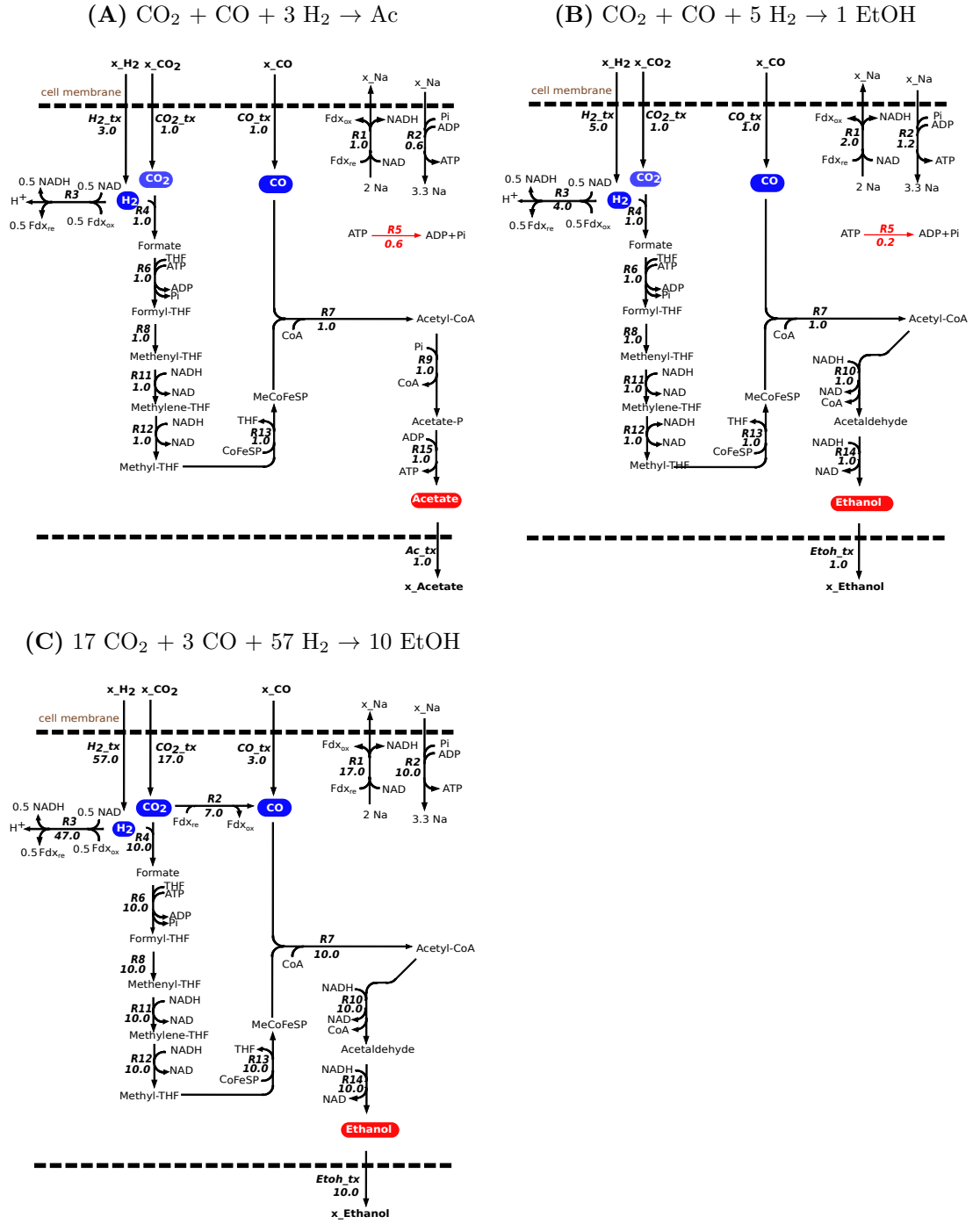


Figure 3.4: **Elementary modes with CO_2 - H_2 - CO as inputs.** Modes 7, 8 and 9 in table 3.1, correspond to figures A, B and C respectively. Proton and water balance not shown.

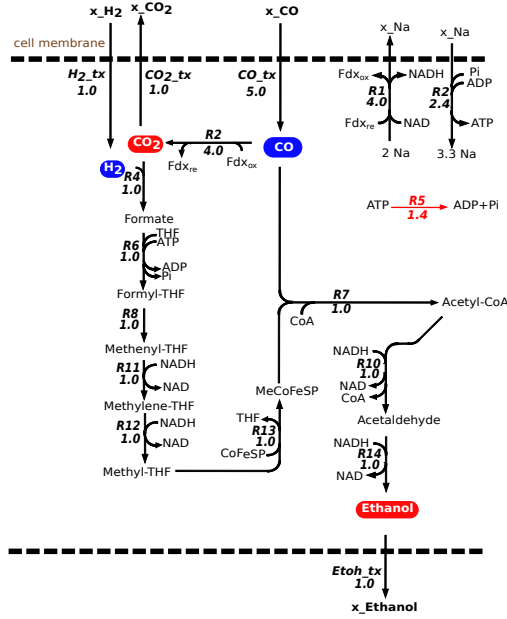
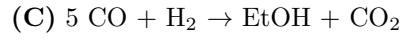
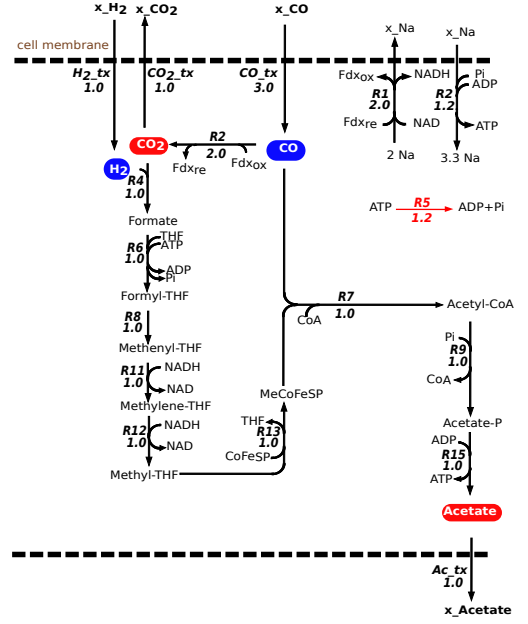
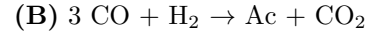
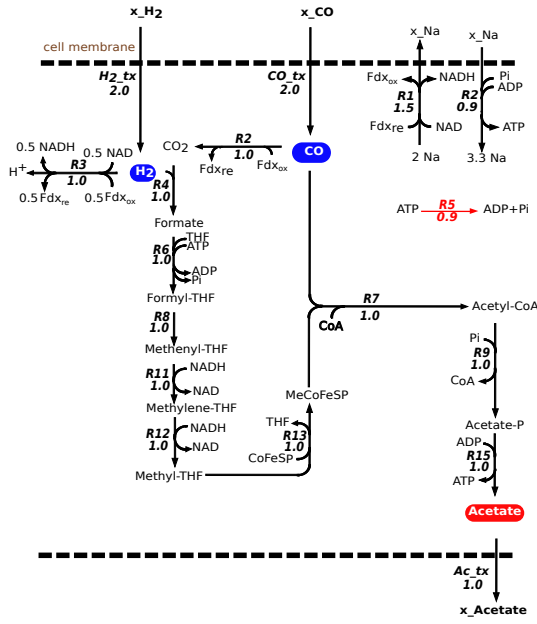
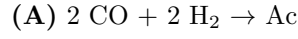


Figure 3.5: **Elementary modes with CO-H₂ as inputs (part 1).** Modes 10, 11, and 12 in table 3.1 , correspond to Figures A, B, and C respectively. Proton and water balance not shown.

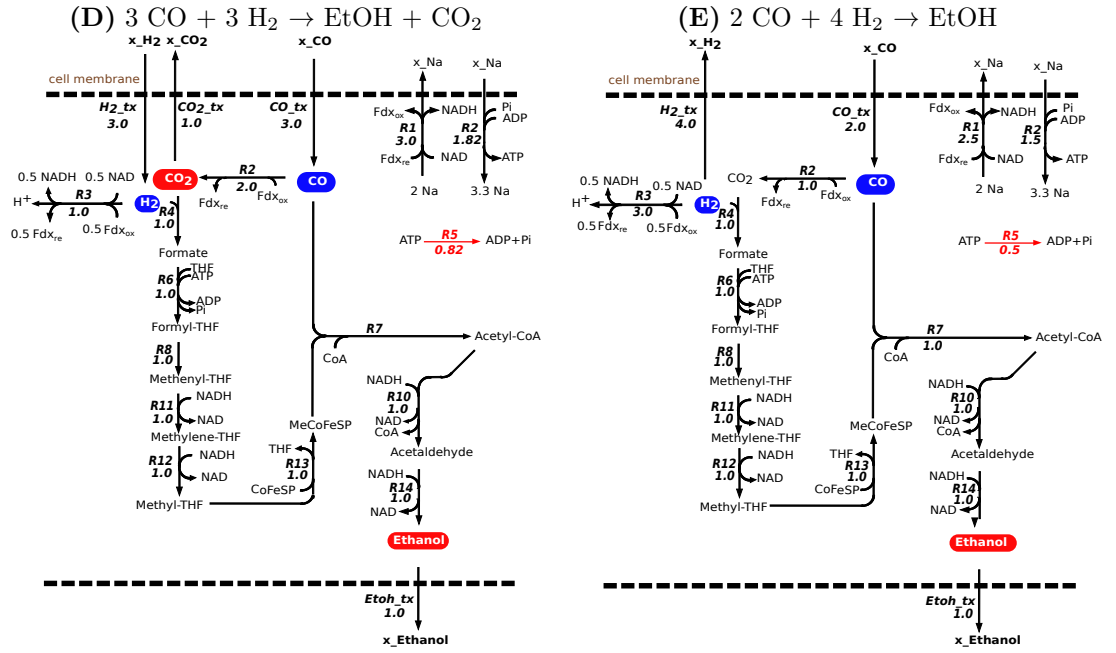


Figure 3.6: **Elementary modes with CO-H₂ as input (part 2)**. Modes 13 and 14 in table 3.1, corresponds to Figure D and E respectively. Proton and water balance not shown.

All modes use CODH, ATPase, the methyl-branch-Acs subset and the Rnf-ATPsynthase subset and operate them in the same direction. These are the only group of modes where EBHyd is used in all modes. The two highest ATP generating modes across all fourteen are in this set of modes (modes 10 and 11).

Reduction of ferredoxin is carried out by both CODH and EBHyd. 100 % of Fdx_{re} is produced by CODH in modes 11 and 12. In modes 10 and 13, 66.6% of Fdx_{re} is produced by CODH and 33.3% by EBHyd. NADH is produced by Rnf and EBHyd. In modes 11 and 12, 100% of NADH is produced by the Rnf. In modes 10 and 13, 75% of the NADH is produced by Rnf and 25% by EBHyd. Mode 14 produces 62.5% of its NADH via Rnf and 37.5% by EBHyd.

Oxidation of ferredoxin is by the Rnf in all modes. NADH is oxidised by the methyl-branch-Acs subset and ethanol producing subset. In modes 10 and 11, NADH is solely oxidised by the methyl-branch-Acs subset. In modes 12, 13 and 14, 50% is oxidised by methyl-branch-Acs subset and 50% by the ethanol-producing subset.

3.3.5 Formate, fructose and methanol metabolism (with CO)

To investigate whether the CO-dependent inhibition of HDCR and EBHyd could account for the growth patterns described in the introduction (section 3.1), an extended version of the WLP model was constructed and analysed, as seen in Figure 3.7. To incorporate formate metabolism, a bidirectional formate transporter ('Form.tx' in Figure) was added to the model. For fructose metabolism, the model was extended with the nine reactions of the Embden-Meyerhof-Parnas pathway ('EMP' in Figure) (as found in *awoodiiCyc* - see Methods for information on this database, section 2.1.2), the pyruvate-ferredoxin oxidoreductase ('PFOR' in Figure) was added, and also a fructose transporter ('Fruc.tx' in Figure).

Reactions for methanol metabolism were identified and included based on methods described in 2.1 (also see section 4.2.3 for detailed outline on curation). Four reactions in total were added: two representing the demethylation of methanol and methylation of a THF via a corrinoid methyl-transferase and two representing methanol oxidation to formaldehyde and formaldehyde ligation with THF (a spontaneous reaction) to methylene-THF.

The extended model contained 30 internal reactions, 41 metabolites and 10 transporters, and no dead reactions. The extended model has double the amount of reaction subsets. Three reactions in the methyl branch (R7, R12, and R13 in Fig. 3.1) are one subset, the two methanol oxidation routes (MDH and FTHFL, and CPMT and MTHF-MT in Fig. 3.7) form a subset each and the last new reaction subset consists of the corrinoid-iron-sulphur protein methylation and acetyl-CoA synthase (R8 and R14 in Fig. 3.1). The HDCR (R5 in Fig. 3.1), and the methyl-THF reductase (R13 in Fig. 3.1) became detached from any reaction subset and the other subsets remain the same as in the WLP model.

In total, 490 modes were found. The modes that consume either formate or fructose as sole substrates, or as co-substrates with CO are listed in table 3.2. The modes that consume methanol with CO as co-substrate are listed in table 3.3. Whether HDCR or EBHyd are present in the mode as well the net stoichiometry and Y_{ATP} of the mode is indicated in the tables. The 14 modes shown in table 3.1 are a subset of the 490 modes. 24 modes use fructose or formate, or either one of these as joint inputs with CO. An additional CO utilising mode was found (mode 24), which generates ATP and exports formate and H^+ . Thirty modes use methanol as input, and twenty five modes use methanol-CO as input. The other 411 modes, use combinations of CO_2 , CO, H_2 , fructose, methanol and formate as inputs. Mode 14 in table 3.2 is the only fully reversible mode.

Eleven modes use fructose as an input. They all generate ATP; mode 1 is the only mode to produce a sole carbon by-product in acetate, and modes 2-11 export a range of mixed products. Modes 10 and 11 do not use either HDCR and EBHyd. Two modes use fructose-CO as inputs. Both generate ATP and use HDCR and EBHyd. Mode 12 exports acetate, CO_2 , and H^+ , and mode 13 exports acetate, formate and H^+ .

Four modes use formate as input, one of which generates ATP (mode 13), and exports acetate, CO_2 and H^+ . All use HDCR and EBHyd, except for mode 14 which uses only

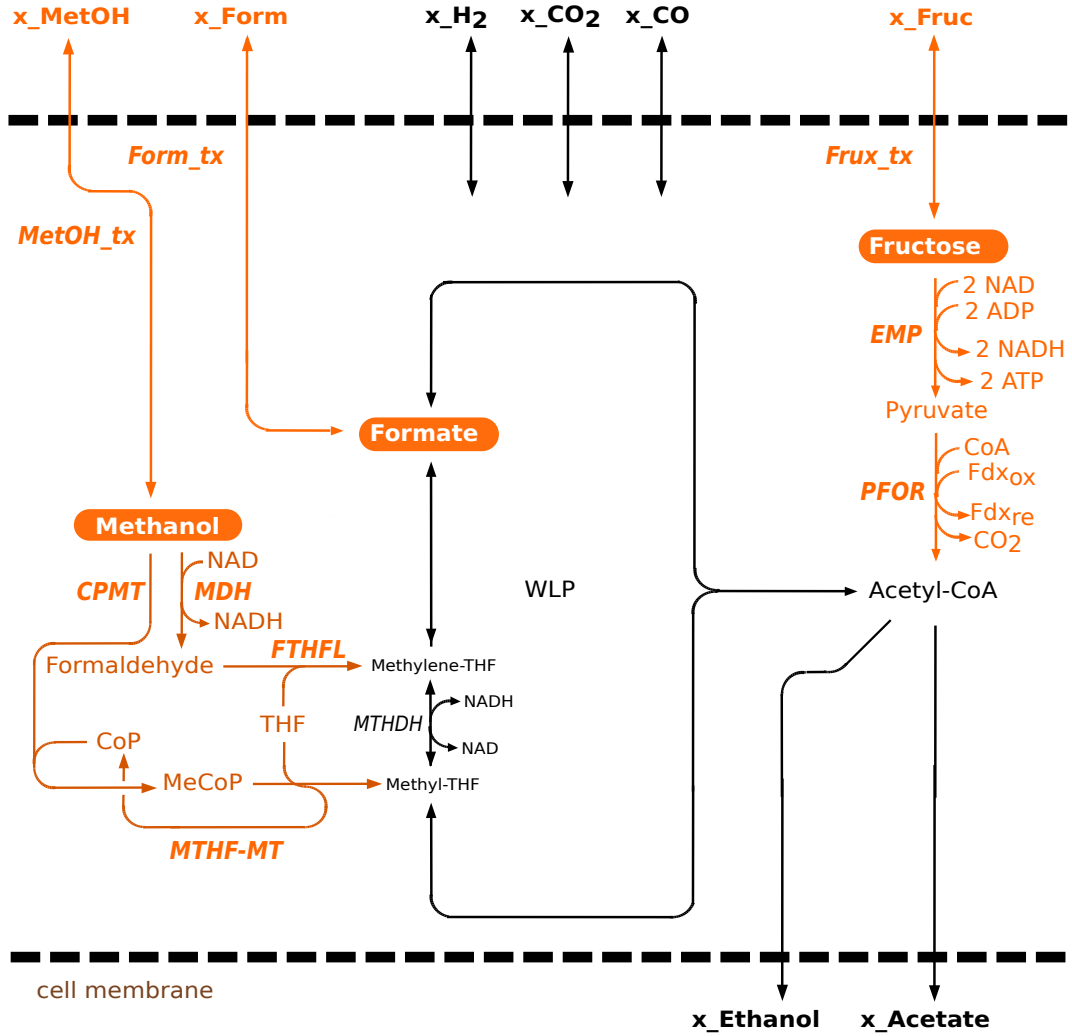


Figure 3.7: **Schematic of the Wood-Ljungdahl pathway model extended to include fructose, formate and methanol metabolism.** Orange arrows and metabolite names are additions to the WLP model. Metabolites with orange backgrounds have transporters. Reactions in the WLP are abbreviated to ‘**WLP**’, reactions in the Embden-Meyerhof-Parnas pathway are abbreviated to ‘**EMP**’. Other reaction names are as follows, ‘**PFOR**’: pyruvate-ferredoxin oxidoreductase, ‘**CPMT**’: corrinoid protein methyl-transferase, **MDH**: methanol dehydrogenase, **FTHFL**: formaldehyde THF ligase, **MTHF-MT**: methyl-THF methyl-transferase, **MTHFDH**: methylenetetrahydrofolate dehydrogenase. The formate transporter, ‘**Form_tx**’, fructose transporter ‘**Frux_tx**’ and methanol transporter ‘**MetOH_tx**’ are shown. The gaseous exchange of CO_2 , H_2 and CO are indicated with the double headed arrow and the ‘x_’ prefix. Acetate and ethanol production and export are indicated with a single arrow. The dashed line indicates the cellular membrane. Elementary modes of this extended model are shown in tables 3.2 and 3.3.

HDCR. Seven modes use formate-CO as inputs, six of which generate ATP (modes 17-22). Modes 17 and 18 do not use HDCR and EBHyd.

Of the twenty five methanol-CO modes, all generate ATP with a Y_{ATP} ranging from 0.3 to 0.5 ATP mol·Cmol⁻¹. All modes use either the formaldehyde route, the methyl-transferase route or a combination of both. The route taken in many modes does not generate a higher or lower Y_{ATP} , or change the product spectrum. The only methanol-CO modes that produce a single carbon-based by-product are ones that produce acetate (modes 1 and 2). These two modes generate the most ATP, use the least number of reactions, and bypass action of HDCR, EBHyd, CODH and Na⁺ gradients built by the Rnf or ATPsynthase. The methanol-CO modes that generate the lowest Y_{ATP} are those that produce ethanol and CO₂ as by-products. Out of the twenty five methanol-CO modes, nine do not employ the HDCR and seven use neither the HDCR or EBHyd.

3.4 Discussion

3.4.1 Introduction

Experimental results on the utilisation of mixed substrates, in particular, substrates in combination with CO are diverse and have not currently been explained in a full or consistent manner. Elementary modes analysis on the WLP was carried out with substrate mixes that had been used in experiments. In total 14 modes were found when gaseous exchange in WLP was investigated, and when this model was extended to include glycolysis, methanol catabolism, formate, fructose and methanol transport, 490 modes were found. In both analyses, it was found that a variety of substrate could feasibly be metabolised into various products. However, the modes identified need to be put into context of the conflicting reports on CO metabolism and *in vitro* enzymatic inhibition studies.

Firstly, CO-inhibition in the context of gas mixes used for growth when fructose, methanol and formate are used in combination with CO or as sole substrates will be discussed. Reports of CO production and re-utilisation will then be discussed next. Biphasic growth patterns on CO₂-CO-H₂, will be examined in light of the previous two discussions. Finally modes that have not yet been seen experimentally, and which might be of interest, will be discussed last.

3.4.2 CO-inhibited reactions in modes of the extended WLP model

Bertsch and Muller [8] reported that growth on formate and CO-formate was mostly uninhibited (until CO was above 50% of the gas phase), but no growth on CO-fructose or CO-methanol was observed. The authors also showed that H₂ uptake (when H₂ was available) was inhibited if CO was above a certain concentration. To explain these growth phenotypes, the authors argue that it is because CO inhibits the catalytic activity of hydrogen-dependent

Table 3.2: Elementary modes of formate and fructose metabolism (with or without CO) in the extended WLP model. Listed are modes that consume formate or fructose as sole substrate or as co-substrate with CO. An additional CO consuming mode found in the extended model is also shown. The reaction direction of HDCR ‘ $\rightarrow$ ’ indicates CO₂ reduction, EBHyd ‘ $\rightarrow$ ’ H₂ evolution, and ‘ $\leftarrow$ ’ indicates the reaction is not found in the mode. Y_{ATP} is given in mol.Cmol⁻¹, and is rounded to two decimal places. Groups of modes have been separated by their inputs.

Mode	Stoichiometry	No. of Reactions	Reaction direction HDCR	EBHyd	Y_{ATP}
1	Fruc $\rightarrow$ 3 Ac, 3 H ⁺	27	$\rightarrow$	$\rightarrow$	0.72
2	Fruc, H ₂ O $\rightarrow$ 2 Ac, 2 Form, 2 H ₂ , 4 H ⁺	21	$\rightarrow$	$\rightarrow$	0.67
3	Fruc, 2 H ₂ O $\rightarrow$ 2 Ac, 4 H ₂ , 2 CO ₂ , 2 H ⁺	20	-	$\rightarrow$	0.67
4	3 Fruc $\rightarrow$ 8 Ac, 2 H ₂ , 2 CO, 8 H ⁺	27	$\rightarrow$	$\rightarrow$	0.67
5	Fruc, H ₂ O $\rightarrow$ Ac, EtOH, 2 Form, 3 H ⁺	25	$\rightarrow$	$\rightarrow$	0.60
6	Fruc $\rightarrow$ 2 Ac, 2 H ₂ , 2 CO, 2 H ⁺	22	-	$\rightarrow$	0.57
7	3 Fruc $\rightarrow$ 4 Ac, 2 EtOH, 4 CO, 4 H ₂ , 2 CO ₂ , 4 H ⁺	24	-	$\rightarrow$	0.56
8	2 Fruc $\rightarrow$ 4 Ac, 2 EtOH, 2 Form, 4 CO, 2 H ₂ , 6 H ⁺	25	$\rightarrow$	$\rightarrow$	0.56
9	Fruc $\rightarrow$ Ac, EtOH, Form, CO, 2 H ⁺	26	$\rightarrow$	$\rightarrow$	0.55
10	Fruc $\rightarrow$ 2 EtOH, 2 CO ₂	18	-	-	0.54
11	Fruc $\rightarrow$ Ac, EtOH, 2 CO, H ₂ O, H ⁺	22	-	-	0.50
12	3 Fruc, 4 CO, 2 H ₂ O $\rightarrow$ 10 Ac, 2 CO ₂ , 10 H ⁺	29	$\rightarrow$	$\rightarrow$	0.66
13	Fruc, CO, H ₂ O $\rightarrow$ 3 Ac, Form, 4 H ⁺	29	$\rightarrow$	$\rightarrow$	0.66
13	4 Form, 3 H ⁺ $\rightarrow$ Ac, 2 CO ₂ , 2 H ₂ O	19	$\leftarrow$	$\leftarrow$	0.08
14	Form, H ⁺ $\leftrightarrow$ H ₂ , CO ₂	5	$\leftarrow$	-	0.00
15	72 Form, 69 H ⁺ $\rightarrow$ 3 Ac, 10 EtOH, 46 CO ₂ , 36 H ₂ O	21	$\leftarrow$	$\leftarrow$	0.00
16	5 Form, 4 H ⁺ $\rightarrow$ Ac, CO, 2 CO ₂ , 3 H ₂ O	17	$\leftarrow$	$\leftarrow$	0.00
17	Form, 3 CO, H ₂ O $\rightarrow$ Ac, 2 CO ₂	17	-	-	0.30
18	Form, 5 CO, 2 H ₂ O, H ⁺ $\rightarrow$ EtOH, 4 CO ₂	18	-	-	0.24
19	2 Form, 2 CO, H ⁺ $\rightarrow$ Ac, 2 CO ₂	19	$\leftarrow$	$\leftarrow$	0.23
20	3 Form, 3 CO, 3 H ⁺ $\rightarrow$ EtOH, 4 CO ₂	19	$\leftarrow$	$\leftarrow$	0.14
21	3 Form, CO, 2 H ⁺ $\rightarrow$ Ac, 2 CO ₂ , H ₂ O	19	$\leftarrow$	$\leftarrow$	0.15
22	5 Form, CO, 5 H ⁺ $\rightarrow$ EtOH, 4 CO ₂ , 2 H ₂ O	19	$\leftarrow$	$\leftarrow$	0.04
23	57 Form, 3 CO, 57 H ⁺ $\rightarrow$ 10 EtOH, 40 CO ₂ , 27 H ₂ O	19	$\leftarrow$	$\leftarrow$	0.00
24	CO, H ₂ O $\rightarrow$ Form, H ⁺	10	$\rightarrow$	$\rightarrow$	0.30

Table 3.3: Elementary modes of methanol-CO metabolism in the extended WLP model. Listed are modes that consume methanol as co-substrate with CO. The reaction direction of HDCR ‘ $\rightarrow$ ’ indicates CO₂ reduction, EBHyd ‘ $\rightarrow$ ’ H₂ evolution, and ‘-’ indicates the reaction is not found in the mode. Y_{ATP} is given in mol·Cmol⁻¹, and is rounded to two decimal places. Modes are ordered by their Y_{ATP} and groups of modes have been separated by their outputs.

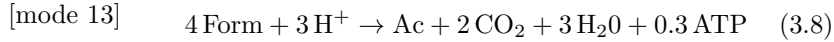
Mode	Stoichiometry	No. of Reactions	Reaction direction		Y_{ATP}
			HDCR	EBHyd	
1	MetOH, CO $\rightarrow$ Ac, H ⁺	11	-	-	0.50
2	MetOH, CO $\rightarrow$ Ac, H ⁺	12	-	-	0.50
3	2 MetOH, CO $\rightarrow$ EtOH, Form, H ⁺	16	-	-	0.33
4	2 MetOH, CO $\rightarrow$ EtOH, Form, H ⁺	16	-	-	0.33
5	2 MetOH, CO $\rightarrow$ EtOH, Form, H ⁺	17	-	-	0.33
6	MetOH, 2 CO, 3 H ₂ O $\rightarrow$ 5 H ₂ , 3 CO ₂	14	$\leftarrow$	$\rightarrow$	0.33
7	MetOH, 2 CO, 3 H ₂ O $\rightarrow$ 5 H ₂ , 3 CO ₂	15	$\leftarrow$	$\rightarrow$	0.33
8	MetOH, 2 CO, 3 H ₂ O $\rightarrow$ Form, 4 H ₂ , 2 CO ₂ , 1 H ⁺	15	-	$\rightarrow$	0.33
9	MetOH, 2 CO, 3.1 H ₂ O $\rightarrow$ Form, 4.2 H ₂ , 2 CO ₂ , H ⁺	16	-	$\rightarrow$	0.33
10	MetOH, 2 CO, 3 H ₂ O $\rightarrow$ 3 Form, 2 H ₂ , 3 H ⁺	15	$\rightarrow$	$\rightarrow$	0.33
11	MetOH, 2 CO, 3 H ₂ O $\rightarrow$ 3 Form, 2 H ₂ , 3 H ⁺	16	$\rightarrow$	$\rightarrow$	0.33
12	2 MetOH, CO $\rightarrow$ EtOH, H ₂ , CO ₂	17	$\leftarrow$	-	0.33
13	2 MetOH, CO $\rightarrow$ EtOH, H ₂ , CO ₂	17	$\leftarrow$	-	0.33
14	2 MetOH, CO $\rightarrow$ EtOH, H ₂ , CO ₂	18	$\leftarrow$	-	0.33
15	9 MetOH, 3 CO $\rightarrow$ 5 EtOH, 2 CO ₂ , 3 H ₂ O	19	$\leftarrow$	$\leftarrow$	0.33
16	9 MetOH, 3 CO $\rightarrow$ EtOH, 2 CO ₂ , 3 H ₂ O	19	$\leftarrow$	$\leftarrow$	0.33
17	9 MetOH, 6 CO $\rightarrow$ 5 EtOH, 2 CO ₂ , 6 H ₂ O	20	$\leftarrow$	$\leftarrow$	0.33
18	10 MetOH, 6 CO $\rightarrow$ 6 EtOH, 4 CO ₂ , 2 H ₂ O	21	$\leftarrow$	$\leftarrow$	0.33
19	5 MetOH, 3 CO $\rightarrow$ 3 EtOH, 2 CO ₂ , H ₂ O	20	$\leftarrow$	$\leftarrow$	0.33
20	5 MetOH, 3 CO $\rightarrow$ 3 EtOH, 2 CO ₂ , H ₂ O	20	$\leftarrow$	$\leftarrow$	0.33
21	3 MetOH, 3 CO $\rightarrow$ 2 EtOH, 2 CO ₂	20	$\leftarrow$	$\leftarrow$	0.32
22	3 MetOH, 3 CO $\rightarrow$ 2 EtOH, 2 CO ₂	20	$\leftarrow$	$\leftarrow$	0.32
23	2 MetOH, 2 CO $\rightarrow$ EtOH, 2 CO ₂	21	$\leftarrow$	$\leftarrow$	0.32
24	MetOH, 3 CO, H ₂ O $\rightarrow$ EtOH, 2 CO ₂	15	-	-	0.30
25	2 MetOH, 3 CO, H ₂ O $\rightarrow$ EtOH, 2 CO ₂	16	-	-	0.30

CO₂ reductase (HDCR) [21]. The electron bifurcating hydrogenase (EBHyd) is also inhibited by CO [112], but was not considered relevant by the authors and so this enzyme was also considered here.

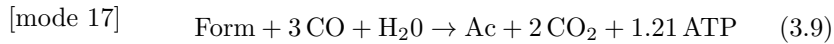
In the WLP model as seen in Figure 3.1, HDCR (R5) is involved in 13 out of 14 modes and EBHyd (R11) is involved in 11 out of 14 modes. Every mode has either one or the other reaction and most have both. The only mode that does not use HDCR is mode 5 (Fig. 3.3B), and although this mode would generate ATP, it does not produce precursor metabolites essential for growth and so would not sustain it. In the extended model (fig 3.7), nine modes that consume CO do not use either HDCR or EBHyd, two of these two modes consume formate as co-substrate, and seven consume methanol as co substrate.

Formate and formate-CO metabolism

Mode 13 (in table 3.2, and equation 3.8) uses formate as a sole substrate to form acetate and CO₂, in a stoichiometry observed in [8]. The authors did not calculate an ATP yield for this metabolism, but on the basis of elementary modes analysis, we calculate a Y_{ATP} of 0.08 ATP mol·Cmol⁻¹. This is lower than the Y_{ATP} of 0.15 ATP mol·Cmol⁻¹ allowed for with CO₂-H₂ (mode 1 in table 3.1, and fig. 3.2 A) and as far as we are aware the lowest reported growth-supporting ATP yield in *A. woodii*.



When Bertsch and Müller [8] used 100mM formate and 50% CO in the gas phase, the formate was completely consumed and 184.8mM CO was consumed, a ratio of 1:1.8, although there may have been some other limiting factor in their culture which did not allow for the residual CO to be consumed. This is closest to the ratio in mode 17, equation 3.9.



The rates of uptake during growth could be more closely examined if the raw data were available. The authors did not calculate a possible ATP yield for this either. Diender et al. [29] calculated a yield of 0.5 mol·Cmol⁻¹ ATP, but in their calculation they employ the CODH reaction operating in both the forward and reverse direction simultaneously to reduce and oxidise ferredoxin, which is thermodynamically infeasible. On the basis of elementary modes analysis presented here, we calculate a yield of 0.3mol·Cmol⁻¹.

Taken together, the stoichiometry observed during growth on formate and formate-CO is reflected by elementary modes. Investigation of the formate-CO modes show that indeed the HDCR is not involved during growth in these conditions.

Fructose and Fructose-CO metabolism

Fructose-based growth on *A. woodii* occurs with a stoichiometry seen in mode 1 of table 3.2, and produces what is claimed to be the highest known yield of $0.72\text{mol}\cdot\text{Cmol}^{-1}\text{ATP}$ for a hexose fermentation [115]. This Y_{ATP} is achieved by CO_2 re-utilisation via WLP and the establishment of Na^+ gradients.

Only one mode with fructose or fructose-CO produces acetate as a sole product and it uses HDCR and EBHyd (mode 1). Both modes 12 and 13 (the fructose-CO modes), use both enzymes also. So, acetate cannot be the sole product when growing on fructose or fructose-CO without HDCR and EBHyd. However, modes that do not use these enzymes and still produce ATP are found but this is accompanied by production of compounds other than acetate. So why growth does not occur then when fructose is available (in the presence of CO) is not clear. Mode 3 for example, in the same table does not use HDCR and allows for full fructose catabolism but to a mix of products including acetate, CO_2 and H_2 . Mode 10 does not use either HDCR or EBHyd, and catabolises fructose to equal amounts of ethanol and CO_2 . *A. woodii* is known to produce ethanol in at least small amounts and can also use it as a growth substrate [10].

Purification and characterisation of the coenzyme A-linked acetaldehyde dehydrogenase (R11 in Fig. 3.1) showed both a dependence on NADH but also inhibition by high levels it Buschhorn et al. [18]. This could potentially lead to inhibition of ethanol as a by-product when NADH concentrations are high. *A. woodii* is also known to evolve small amounts of H_2 , but again, high amounts of H_2 evolution becomes growth-limiting [14, 27]. Formate production on the other hand has been observed as an additional product to acetate in voluminous amounts under high H_2 partial pressures [63]. Although the regulation of some products may limit growth, attenuated growth could have potentially still been observed if CO induced HDCR inhibition was the reason for no growth.

Diekert and Maria [27] for example, found that growing *A. woodii* on fructose and gradually depleting amounts of nickel in the media resulted in a “soft” knockout of CODH since it is a nickel-dependent enzyme. In media with the lowest amounts of nickel, fructose was catabolised to acetate, formate, H_2 , and CO_2 using a stoichiometry similar to modes 2 and 3 in table 3.2 - modes that also do not use CODH. This was significantly detrimental to the growth rate and final OD, but in principle, shows that growth can still occur when products other than acetate are formed alongside it.

Methanol-CO metabolism

Methanol is known to be a growth supporting substrate when used in combination with CO_2 , formate and a range of aromatic acrylate groups [5, 4, 60]. Methanol-CO however does not support growth [8]. Two modes were identified that metabolise methanol-CO to acetate without action of either HDCR or EBHyd (modes 1 and 2 in table 3.3). These modes generated the highest yields of ATP ($0.5\text{ATP mol}\cdot\text{Cmol}^{-1}$), and used the smallest

number of reactions.

The route used in mode 1 was via the corrinoid-protein dependent methyl-transferase (CPMT), and the route used in mode 2 was via methanol oxidation to formaldehyde. Although this route via formaldehyde has not been shown experimentally it only requires one enzyme, a methanol dehydrogenase, which has bioinformatic evidence supporting its presence in the *A. woodii* genome (not yet shown, see section 4.2.3). This oxidative route would not affect Y_{ATP} and allows circumvention of the Methylene-THF dehydrogenase (MTHFDH in fig 3.7) having to operate in the reverse methylene-THF forming direction, which is a highly endergonic reaction with NADH ($\Delta G_0' = +23kJ/mol$) [9]. Having two catabolic routes, one demethylating and one oxidative, for methanol in one organism is not unheard of in methylotrophic organisms [20]. The oxidative intermediate, formaldehyde, has also been independently shown to support *A. woodii* growth [32].

Even without the oxidative formaldehyde route, methanol catabolism via CPMT still allows for acetate production without HDCR and EBHyd in mode 1. Additionally, other potential sites of inhibition such as CODH and the Rnf complex are also not used in this mode. This suggest that inhibition of the HDCR and EBHyd does not explain why no growth on methanol-CO was observed.

HDCR inhibition alone cannot account for observed growth phenotypes

Inhibition of HDCR may be enough to explain growth on formate-CO (at CO levels below 50%) and CO₂-CO-H₂ mixtures (albeit at low rates) but cannot explain growth phenotypes on methanol-CO, fructose (in the presence of CO) and formate-CO conditions where CO is 50% and above in the the gas phase. This suggests CO is having an inhibitory or regulatory affect on more enzymes than the HDCR and and EBHyd alone. Also, the low rates of growth on CO₂-CO-H₂ in the presence of high CO suggests that HDCR is still active, even if this activity is limited, since HDCR is present in all growth supporting modes using these gas mixtures (table 3.1).

Schuchmann and Müller [113] showed that 10% of CO in the gas phase inhibited the H₂-dependent subunit of HDCR, and around 0.01% CO in the gas phase inhibits up to 50% activity of the EBHyd in the H₂ oxidising direction [112]. This inhibition is not permanent and is fully reversible once CO disappears [21]. These enzyme characterisations were made *in vitro* and so their activity could differ *in vivo*. For example, *A. woodii* is known to regulate certain intracellular microenvironments using bacterial microcompartments when growing on 1,2-propanediol [116] to limit damage caused by toxic intermediates.

When fructose-CO are used as inputs, both HDCR and EBHyd are essential enzymes. These enzymes were not essential though on fructose alone, but if they are not present will result in products other than acetate being formed. It may be that regulation on the formation of products other than acetate is so tight that growth is inhibited completely. This however needs reconciliation with reports of *A. woodii* growth on nickel-free media

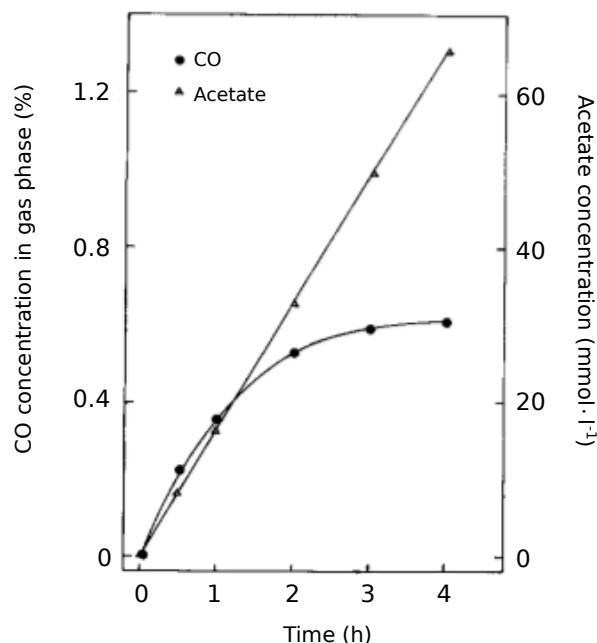


Figure 3.8: **Findings from a report on CO production and re-utilisation, reproduced from Diekert et al. [28].** CO is produced from $\text{CO}_2\text{-H}_2$ in a resting cell assay of *A. woodii*. According to elementary mode analysis carried out here, the operation of (at least) two elementary modes is necessary for this.

[27], which showed that other products can be produced alongside acetate in stoichiometric amounts, even though growth is severely limited.

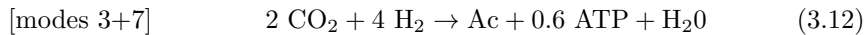
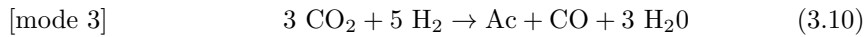
An altogether different explanation is needed for methanol-CO conditions. CO is known to be a potent inhibitor of both aerobic and anaerobic metalloenzymes [130, 102]. *A. woodii* uses several metalloenzymes and cofactors: nickel-based CODH, cobalt-based CPMT, iron-sulphur based corrinoid proteins (Acs) and ferredoxin itself consists of iron-sulphur cores. These allow for several potential sites of inhibition. Since formate-CO supports growth at high CO levels though, some of these can be ruled out. CO is also a known transcriptional regulator [123, 150], and has been implicated in autotrophic organisms which use bifunctional CODH/ACS [149, 85]. It has also been suggested it may play a role in uncoupling the expression of monofunctional CODH and bifunctional CODH/ACS [130], and so transcriptional regulation of CO in *A. woodii* should also be considered.

3.4.3 CO production and re-utilisation

CO production from CO₂-H₂

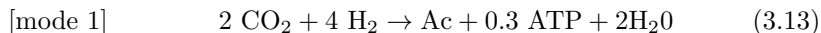
CO₂-H₂ was found to produce acetate and smaller amounts of CO in resting cells [28], see Figure 3.8 for a reproduction of these findings. The reason for this investigation in resting cells was because of the authors previous findings that small amounts of CO were produced in cells growing on CO₂-H₂. While the acetate formation rate was linear with time, and at a rate similar to its formation in growing cells, CO formation rates decreased after 3 hours from an initial rate of 3.7nmol·(min·mg)⁻¹ to zero once ≈0.6% of the gas phase was CO. Upon exogenous addition of CO to a CO₂-H₂ gas mix, CO levels decreased to 0.6%. Removing the gas atmosphere after 3 hours and replacing it with fresh CO₂-H₂, also resulted in CO production at 3nmol·(min·mg)⁻¹. Higher concentrations of CO from CO₂-H₂ were also measured when various inhibitors were used. The authors suggest then, that CO was being consumed at the same rate as it was produced after four hours.

However, their stoichiometric equation for CO₂-H₂ to CO as a sole product cannot be correct according to the enzymes available to *A. woodii* and the elementary modes found. This is because at the time of publication, insufficient data was available on the enzymes now characterised in *A. woodii*, including the CODH which uses reduced ferredoxin rather than H₂ to reduce CO₂ to CO. Instead their measurements for acetate and CO production from CO₂-H₂ can be explained using equation 3.10 as found in mode 3, table 3.1, which is consistent with their observation and *A. woodii* enzymes. This mode however, would not support growth or cell maintenance as no ATP is produced this way. This makes sense in the light of CO formation rates dropping to zero after a non-zero phase, as CO starts to be consumed. Its simultaneous consumption with CO₂-H₂ can be explained by equation 3.11 as found in mode 7 in the same table.



For these modes to be able to operate interchangeably without a lag, all reactions in common between these modes must operate in the same direction and this is the case, with the obvious exception of CO transport. The reactions that differ are the CODH and the Rnf-ATPsynthase subset which are present in mode 3 but not mode 7. So the scheme being suggested with the above modes is that CO is accumulated by mode 3 in the gas phase by diffusion out of the cell, and CO then diffuses back across the membrane as the input for mode 7 along with CO₂-H₂ once a set concentration is reached. Interestingly, the sum of both modes 3 and 7 as seen in equation 3.12, allows for more ATP mol·Cmol⁻¹ to be

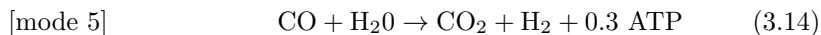
generated when compared to direct acetate formation from CO₂-H₂ (mode 1 and equation 3.13) without CO accumulation:



However, a growth strategy involving the interplay of modes 3 and 7 working in tandem would mean that all of the CO used in mode 7 would come from the CO produced in mode 3, and so over a period of time the overall amount of ATP generated by equation 3.13 and 3.12 would be the same since modes 3 and 7 would have to be operating for an equal period of time since the rate of CO production and consumption becomes equal after four hours. Alternatively, it is also feasible that although CO can be consumed its production could possibly be an artefact of a leaky CODH. More investigation on the CODH enzyme would be necessary argue in favour of active CO production. If CO, does indeed leak out of the enzyme and cell, mode 1 (3.13) is still insufficient to explain active CO consumption with CO₂-H₂.

CO re-utilisation by oxidation to CO₂-H₂

The authors also show that ATP formation can be driven by CO oxidation to CO₂-H₂. CO oxidation to CO₂-H₂ was also independently shown in [55]. Since the rate of CO oxidation was very low, they unfortunately could not accurately measure a stoichiometry for this. They did manage to measure both an increased ATP concentration, and active histidine transport (a process requiring metabolic energy) driven by CO oxidation. Elementary mode 5 (equation 3.14) describes a stoichiometry that oxidises CO to CO₂-H₂ whilst forming ATP, corroborating and consistent with their finding:



This type of CO oxidation however, needs reconciliation with what appears to be inconsistency in at least two data points reported by [112] and [8]. Bertsch and Müller [8], like Diekert et al. [28] also did a resting cell assay with CO, but observed acetate production from it instead of CO₂-H₂. This can be explained by differences in the media composition in both experiments. Diekert et al. [28] would have been unaware at the time that *A. woodii* is dependent on Na⁺ gradients for acetate production. In fact a portion of their paper is dedicated to providing evidence for proton gradients in *A. woodii* using various inhibitors, and so Na⁺ concentrations in their media was insufficient for acetate production. This would be consistent with the observation of CO production and re-utilisation from CO₂-H₂ using modes 3 and 7. As mode 3 does not involve Na⁺ gradients the flux through the Rnf complex would be lower than that used in mode 1. The CO-oxidation to CO₂-H₂ observation could potentially be replicated then by using reduced concentrations of Na⁺ in the media, or by using Na⁺ ionophores.

The second apparent inconsistency is the activity of EBHyd in mode 5 (equation 3.14). This mode employs this enzyme but Schuchmann and Müller [112] clearly demonstrate EBHyd is strongly inhibited by CO, even at levels as low as 0.01%. The assay used to characterise this enzyme inhibition used methylviologen as an electron acceptor for H₂ oxidation. No investigation was done however on the evolution of H₂ from reduced ferredoxin and NADH in the presence of CO, so this remains an open question. For CO oxidation to occur in the way [28] report, this enzyme must be carrying flux in reverse, and so suggests *in vivo* activity of this enzyme allows for this reaction to occur, at least with low activity. This is also consistent with discussion from the previous section (section 3.4.2).

The findings from Diekert et al. [28] then, when analysed in the context of elementary modes, indicate that an interplay of two modes is needed to explain both the resting and growing cell experiments in the context of CO related observations. What would dictate the interplay between these two modes, is possibly examinable with further analysis, a kinetic model of this pathway and also considerations of the shifting thermodynamic driving forces in the changing gas mixture, which are beyond the scope of the current investigation.

3.4.4 Biphasic growth patterns on CO₂-CO-H₂

In regards to report on CO₂-CO-H₂ biphasic growth, see Figure 3.9 for a reproduction of some of the key growth data [8]. Their findings have not yet been adequately explained in a metabolic context. They showed that addition of CO to a CO₂-H₂ gas mix drastically slowed down growth and resulted in a biphasic growth pattern with a slow and fast phase that are not dependent on enzyme adaption but instead on whether CO was below 3% of the gas or not. Final OD values with CO in the gas mix were higher than with CO₂-H₂ alone in a carbon limited media and CO had completely disappeared in the gas phase once stationary phase was reached indicating CO was taken up for growth. They also show that CO as a sole carbon and energy source does not support growth. In CO₂-CO-H₂ mixes the H₂ levels decreased at a much faster rate when CO was below 3% but the transition between the first phase (slow growth) and second (fast growth) did not correlate exactly with slow and fast H₂ consumption since some fast growth occurred chronologically before fast H₂ consumption, although this is difficult to see from their Figure. Resting cell assays by the authors showed that in principle acetate could be formed from pure CO although the rate of this was 2% that of formed on CO₂-H₂, and that above 10% CO in a CO₂-H₂ gas mix resulted in acetate production rate similar to that of on CO alone.

The elementary modes may offer suggestions for what metabolism is occurring. Since the second (fast) growth phase occurred with rates of H₂ consumption at 130 $\mu\text{mol} \cdot \text{h}^{-1}$, comparable to *A. woodii* grown on the CO₂-H₂ control, this growth is likely to occur with the CO₂-H₂ to acetate stoichiometry frequently observed with *A. woodii*, as represented by mode 1 (equation 3.13). The question at hand is, can the growth stoichiometry in the first (slow) growth phase, where slow H₂ consumption rates are observed, be explained with an

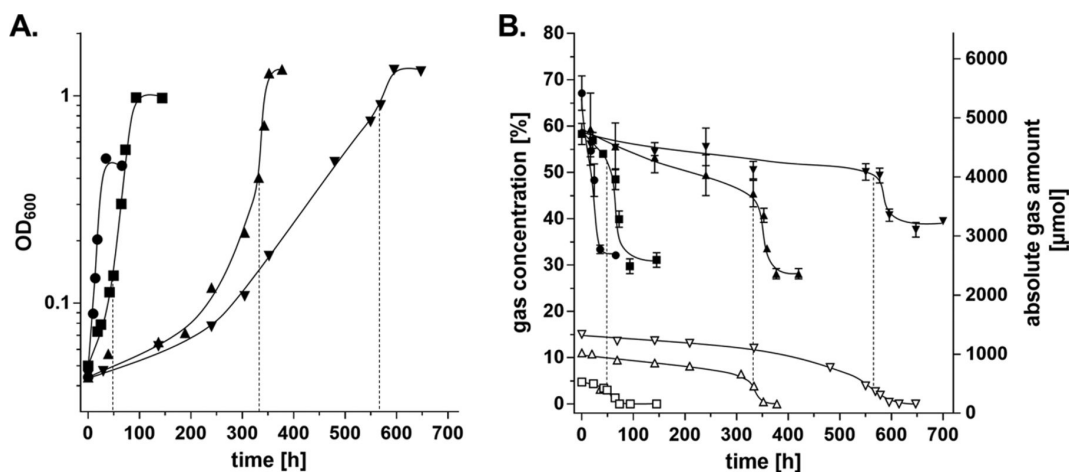


Figure 3.9: **Findings from a report on biphasic CO-based growth, reproduced from Bertsch and Müller [8].** The findings show that CO inhibits growth and H_2 consumption. The gas phase consisted of 64% H_2 , 16% CO_2 , and 5% (■), 10% (▲), or 15% (▼) CO. The control did not receive any CO (●). All cultures were grown in phosphate-buffered medium free of bicarbonate and inoculated from the same preculture adapted to growth on H_2 - CO_2 . (A) Growth was measured by following the optical density at 600 nm. (B) The concentrations of the gases CO (open symbols) and H_2 (closed symbols) in the gas phase. The dashed lines mark the time points when the CO concentration was reduced to 3%. The data shown are representative of three independent experiments.

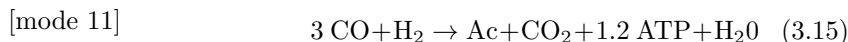
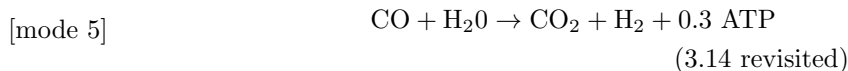
elementary mode?

Slow growth phase during growth on 15% CO

During growth on the various gas mixes (Figure 3.9), H_2 consumption was the slowest at $2\mu\text{mol}\cdot\text{h}^{-1}$ when CO was 15%, which is down from $15\mu\text{mol}\cdot\text{h}^{-1}$ and $7\mu\text{mol}\cdot\text{h}^{-1}$ when CO was 5%, 10% respectively. The reported H_2 consumption rate at 15% CO is then in conflict with the resting cell findings that acetate production was solely from CO when above 10%. The slow H_2 uptake at 15% CO is also questionable because the measurements before fast uptake all seem to be within the margin of error. This data was not available as raw data and was not subject to any statistical treatment by the authors. The potentially small amount of H_2 disappearing in the gas phase after around 550 hours (22.9 days), could be due to H_2 escaping the serum flasks during sampling which H_2 is prone to do.

To work out which of the 14 modes in table (3.1) could be active during growth in high CO levels (15% CO), some modes must be ruled out. For growth to be supported, ATP must be generated; this rules out modes 2, 3 and 9 since no ATP is generated with these. Acetate was observed as the major sole fermentation product, ruling out modes 5, 6, 8, 12 and 13 that produce compounds other than this (CO_2 measurements were not reported so modes with acetate- CO_2 products are still considered). Mode 1 already describes fast growth so is also ruled out. That leaves modes 4, 7, 10 and 11. Mode 4 produces acetate and CO_2 from CO alone, which was shown to be possible in practice using resting cells but CO as sole substrate did not support growth and so is also ruled out. Out of modes 7, 10 and 11, all consume CO and H_2 , but one also consumes CO_2 , one produces CO_2 , and one is CO_2 balanced. Additional measurements of CO_2 levels could possibly have determined which one of these modes (if any) was active. Since it is questionable that any H_2 is consumed stoichiometrically with CO before the second growth phase, none of these modes may represent the correct metabolism. If more H_2 measurements are made available and it is shown that small amounts of H_2 are consumed, mode 11 rather than mode 10 may be a good candidate mode. This is because in mode 11 only the HDCR is carrying flux whereas in mode 10 both the HDCR and the EBHyd - in the strongly inhibited H_2 oxidising direction - are carrying flux. Mode 7 is also an unlikely candidate mode here because three times as much H_2 is being consumed than CO, compared to the 3:1 CO: H_2 ratio in mode 11, which is more in keeping with the data in Figure 3.9.

If however, there is no net H_2 consumption at high CO levels, the slow growth and acetate production rates must be explained by CO consumption alone, since no CO- CO_2 mode exists and no other carbon/energy source is in the gas mix. Since acetate is produced, albeit at low rates, this indicates that the stoichiometry of mode 4 is being observed. How is this possible if CO as a sole substrate does not support growth? An alternative solution exists if mode 5 is used in conjunction with mode 11. The sum of the net stoichiometry of both modes 5 and 11 equals to the net stoichiometry in mode 4 (equation 3.16).



Mode 5 produces and exports H_2 , and mode 11 imports H_2 . None of the reaction directions are in conflict in both modes and so like in equation 3.12, the modes could operate interchangeably without a lag. These modes in combination, would allow for both the production of H_2 for immediate energetic requirements (as seen in [28] and section 3.4.3), and for the reuptake of H_2 for acetyl-CoA and acetate production. Mode 4 alone however, would require H_2 evolution for direct reduction of CO_2 in the HDCR.

This could explain why growth occurs with CO₂-CO-H₂ even if H₂ on net not consumed, but does not occur when CO is supplied as sole substrate. The concentration of H₂ might be too low to have a sufficient driving force for the HDCR with mode 4 alone, compared to the H₂ concentrations available in CO₂-CO-H₂. This is consistent with another recent finding reported at the annual C1Net conference that H₂ cycling in *A. woodii* and WLP is essential for growth [148].

Also, although methanol-CO did not support growth, acetate formation was observed (in the presence and absence of Na^+) in a resting cell assay when methanol-CO- H_2 was available, up to rates of $60.9 \text{ nmol} \cdot (\text{min} \cdot \text{mg})^{-1}$ [55], again implicating a role of H_2 when in the presence of CO. The considerably higher Y_{ATP} of $0.38 \text{ mol} \cdot \text{Cmol}^{-1}$ in equation 3.16, compared to $0.15 \text{ mol} \cdot \text{Cmol}^{-1}$ in equation 3.13, allowed by this metabolism might also support the slow growth rate of 0.0112 h^{-1} .

Slow growth phase during growth on 10% CO

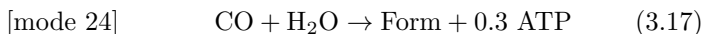
The first (slow) phase during biphasic growth on 10% CO is likely to be different to that on 15% CO. This is because H₂ was consumed at a higher rate than CO, and more H₂ in total was consumed when compared to H₂ consumed at 15% CO. Even though more H₂ was consumed at 10% CO, the same final OD was reached as the 15% CO culture. As mentioned previously in section 3.4.2, it was shown that *in vitro* the HDCR was inhibited at 10% CO [113, 99]. Levels below this might then allow faster H₂ consumption. Mode 7, already used to explain findings from [28] (section 3.4.3), would reflect CO consumption, higher H₂ consumption (three times as much), acetate production, and ATP generation for growth and so would be a good candidate to explain the slow growth phase on 10% CO.

3.4.5 Potential novel metabolism

Elementary modes analysis, as well as being useful for interpreting metabolic data, reveals potentially novel metabolism. Exchange of material between CO₂-H₂, CO and formate has been reported. CO₂-H₂ can be converted into formate for use in further metabolism or excreted into the extracellular environment. If formate is excreted, it can be taken up again for conversion back into CO₂-H₂ or itself be used for growth ([113, 67, 63] and mode 14 in table 3.2). As far as I am aware, there are no reports of *A. woodii* storing compounds as long-chain polymers intracellularly. The back and forth shuttling between gaseous and non-gaseous compounds, may provide an ecological advantage for *A. woodii* in allowing it to sequester excess carbon and energy sources into forms not utilisable by other organisms, thus capitalising the source for itself.

Here, new light is shed on gaseous exchange capabilities of *A. woodii*'s metabolism. Another exchange pathway between compounds, this time CO to CO₂-H₂ is shown that would convert a reactive and inhibitory gas into ones better suited for growth whilst also generating immediate metabolic energy in the form of 0.3 ATP mol·Cmol⁻¹, which can be used for cell maintenance (mode 5 in table 3.1). This is achieved using only five reactions.

An additional novel gas-to-solute route was found in mode 24 from table 3.2. It was shown that formate production from syngas (CO₂-CO-H₂) is possible [113], but no such elementary mode was found. Mode 24 instead shows that CO is directly converted to formate (although H₂ may be needed in the gas phase because of previously discussed H₂ cycling), and generates 0.3 ATP in the process. Formate production from CO was also detected in another acetogen, *Clostridium thermoautotrophicum*, when cultures were supplemented with nitrate [42].



Due to future energy demands, alternative energy feedstocks such as H₂ are being investigated. It has already been shown that the HDCR efficiently hydrogenates CO₂ reversibly to formate ([21], and mode 14 in table 3.2), effectively storing H₂ in solution. This is advantageous for industrial applications both because the hydrogenation occurs at ambient temperatures as opposed to the high temperatures and pressures needed for chemical hydrogenation of CO₂ [2], and because it could resolve issues around H₂ storage [107]. Formate production in mode 24 however would provide additional ATP needed for cell maintenance, which does not occur when CO₂-H₂ is converted to formate. Generating formate this way however would occur at a slower rate because of HDCR inhibition [113].

Altogether, 490 elementary modes were found in a model consisting of only 30 internal reactions. The possibility of identifying these without computational methods is unlikely and at best labour intensive. The modes discussed here have related to growth phenotypes reported, but the vast majority represent potentially achievable metabolism according to the structure of the metabolic network. Of the modes computed here, unrealisable ones *in vivo*

raise questions on the types of regulatory frameworks that govern *A. woodii*'s metabolism.

3.4.6 Conclusions

Structural modelling of WLP, and an extended version of it, has allowed for novel insights into various reports and findings on *A. woodii*. It was shown that there are potentially two ways to achieve a stoichiometry of $2 \text{ CO}_2 + 4 \text{ H}_2 \rightarrow \text{Ac}$, and that both would ultimately produce the same amount of ATP but one involves accumulation of CO (probably because of a low concentration of Na^+ in growth media). It was shown that reduced ferredoxin cannot replace H_2 in the HDCR reaction when growing on $\text{CO}_2\text{-H}_2$. The stoichiometry of substrate to product during growth on formate and formate-CO was represented by elementary modes and a Y_{ATP} of each of these are calculated and reported here.

Evidence for further roles of CO in regulation of WLP, other than inhibition of HDCR and EBHyd has been provided here using data on mixotrophic substrate conditions. Strands of seemingly inconsistent reports of CO metabolism have been reconciled into an explanation consistent with reported findings on metabolism with both CO and $\text{CO}_2\text{-CO-H}_2$.

Additionally, insights into potentially novel metabolism achievable with the WLP are also reported here including formate production from CO, and the vast majority of modes which represent metabolism not yet observed or reported on in *A. woodii*.

Chapter 4

Genome-Scale Metabolic Model Construction and Overview

4.1 Introduction

Presented here will be the construction of an *A. woodii* Genome-scale Metabolic Model (GSM) and the results of model curation on model analysis. The general steps for building GSMs and the *A. woodii* GSM is outlined in Computational Methods (chapter 2). Construction of GSMs relies on the availability of a sequenced and annotated genome, and is comprised of reactions that encompass the entire proteomic repertoire of the modelled organism. The particular *A. woodii* strain this GSM is based on is DSMZ1030 of which a genome sequence was reported [92]. The genome annotation was downloaded from BioCyc as a tier 3 database and is referred to here as awoodiiCyc. Since tier 3 databases have undergone no manual curation, many parts of the metabolic reconstruction were missing or erroneous. In this chapter the focus of GSM construction will be on curation of reactions and reaction pathways.

Model curation is necessary to ensure that on analysis of the model, solutions obey laws of mass and energy conservation. The curation process also provides fresh insight into the state of knowledge about the metabolic network being curated. The curation of an *A. woodii* GSM allowed for the identification of missing reactions, new reactions and reactions lacking evidence in awoodiiCyc as well as identifying necessary reactions needed to allow for synthesis of essential biomass precursors. The analysis of relatively unexplored parts of *A. woodii* metabolism then, allowed for novel insights into sections of metabolism that are of central importance and how they may affect more familiar sections of the network. An overview of general features of the completed GSM is also presented.

4.2 Curation

Several reactions that were not found in MetaCyc or awoodiiCyc were needed for the synthesis of the biomass components (described in chapter 2, section 2.1.2) and so written manually. Reactions for *A. woodii* specific peptidoglycan synthesis were written manually as a series of lumped reactions. A lumped reaction for techoic acids spanning the peptidoglycan was written manually. A lumped reaction was added for fatty acid synthesis using palmitate as a representative fatty acid and two reactions were added for CDP-diacylglycerol and phosphatidylethanolamine synthesis to allow for lipid synthesis. Some reactions of the Woods-Ljungdahl pathway (WLP) were incorrectly defined in awoodiiCyc and so the WLP model described in chapter 3 was used to replace the WLP found in awoodiiCyc. In general curation of the rest of the reactions was carried out according to the steps outlined in chapter 2 section 2.1.

Relevant and recent literature describing aspects of *A. woodii* metabolism was continually added manually to update the GSM. This included adding transporters for known growth supporting substrates and essential micronutrients used in growth media. The linear program used for determining biosynthesis of metabolites is described in chapter 2, section 2.2. Analysis of how reactions operate during biomass formation was done using equation 2.4 from the same section.

Using equation 2.5, biomass precursors that could not be synthesised from simple substrates were identified. Four nucleotides could not be synthesised: deoxycytidine-triphosphate, cytidine-triphosphate, thymidine triphosphate and uridine-triphosphate. Four amino acids could not be synthesized: methionine, arginine, asparagine and lysine. These were resolved using methods described in chapter 2, section 2.1.3. Lysine however, could still not be synthesised after this and it was discovered that this was due to missing reactions. Adenosylcobalamin (vitamin B12), another biomass component could also not be synthesised. Methanol, a growth supporting substrate could not be utilised as a growth substrate in the model. Additionally, reactions related to classical central metabolism were manually curated because of the lack of published data on these reactions and because of their importance. The PRIAM tool (see chapter 2, section 2.1.1) was used to identify missing reactions and to provide evidence or lack thereof of enzymes with unclear genetic annotation.

4.2.1 Lysine Biosynthesis

Evidence for lysine biosynthesis

Evidence for native *A. woodii* having a lysine biosynthesis pathway includes: being a free living organism, being able to grow on a defined - lysine free - media ([47], and Laboratory Experiments chapter 6), and furthermore, the genome contains 4.3% of lysine codons where the average amino acid representation is 5%.

Curation approach

The MetaCyc database was used to identify potential lysine biosynthesis pathways. It contains six such pathways that are divided into based on either having the intermediate metabolite diaminopimelate (DAP) or L-2-aminoadipate. In the pathways that belong to the DAP group, lysine is produced from aspartate. All DAP pathways share four upstream reactions required for conversion of L-aspartate to (S)-2,3,4,5-tetrahydropicolinate and the enzyme diaminopimelate decarboxylase (EC-4.1.1.20) which converts meso-diaminopimelate to lysine. PRIAM was used to discover if any new enzymes not already present in the model were potentially missed in genome annotations. Together with data on reaction-gene associations, reactions were added to the model if a basis for them could be established in the *A. woodii* genome.

Curation results

For the reconstructed lysine biosynthesis pathway (as found in the GSM) see figure 4.1 for a schematic. The two MetaCyc pathways that were best represented in the draft GSM were ‘L-lysine biosynthesis II’ (PWY-2941) and ‘L-lysine biosynthesis VI’ (PWY-5097) which contain nine and seven reactions respectively. Of the nine reactions in L-lysine biosynthesis II, seven were supported by either genetic evidence, PRIAM based evidence, or both. The two enzymes without supporting evidence were tetrahydropicolinate acetyltransferase (EC-2.3.1.89) and diaminopimelate epimerase (EC-5.1.1.7). Of the seven reactions in L-lysine biosynthesis VI, six had either genetic evidence, or PRIAM based evidence or both. L,L-diaminopimelate aminotransferase (EC-2.6.1.83) is unique to this pathway and was not initially in the GSM but was identified with PRIAM and so added. The enzyme with no evidence was diaminopimelate epimerase (EC-5.1.1.7). Since all other reactions in L-lysine biosynthesis VI are present in the model, this reaction was added to allow for lysine synthesis until further biochemical evidence becomes available.

Using equation 2.5, the model allows for lysine production from fructose with a net stoichiometry of:



using 36 reactions. A growth LP was solved using equation 2.4, which showed that lysine biosynthesis during both heterotrophic and autotrophic growth corresponds to the MetaCyc pathway ‘L-lysine biosynthesis VI’ (PWY-5097) as seen in figure 4.1.

4.2.2 Vitamin B12 - Cobalamin Biosynthesis

Evidence for cobalamin biosynthesis

A. woodii is a known cobalamin producer and it has been reported that 95% of its corrinoid content is cobalamin [128]. Given the complex biosynthetic pathway and difficulty in

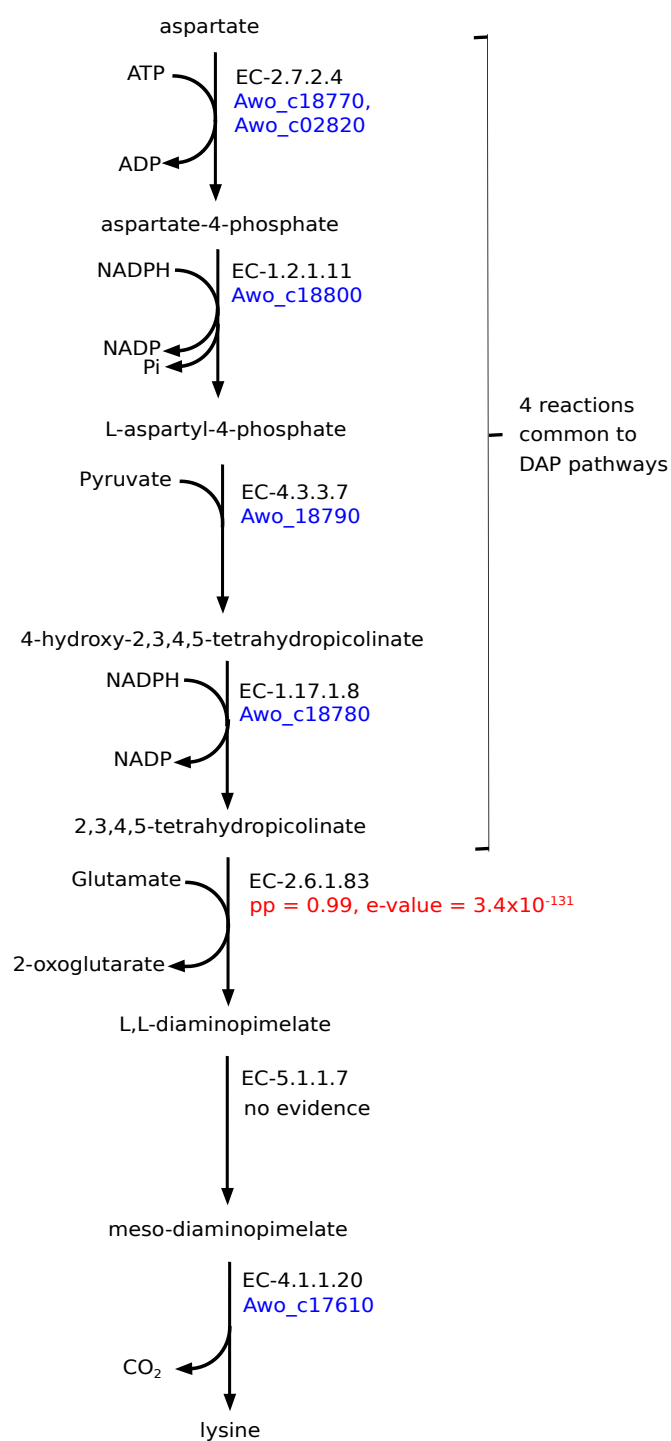


Figure 4.1: **The curated lysine biosynthesis pathway that allows for lysine synthesis and export from aspartate.** Diaminopimelate is abbreviated to DAP. Names in blue are gene locus tags, in red are the PRIAM hit scores profile proportion (pp) and its e-value.

characterising anaerobic proteins, characterisation of the anaerobic synthesis pathway has taken decades to decipher [82]. Significant advances to understanding the synthesis of 5,6-dimethylbenzimidazolyl, an essential cobalamin ligand, has been made recently in the model anaerobe *Eubacterium limosum* [52], and now allows for the reconstruction of this pathway in *A. woodii*.

Curation approach

Two *de novo* cobalamin synthesis pathways exist in the MetaCyc database: the aerobic and anaerobic variants. They both incorporate a cobalt ion in the central corrin ring and employ upper and lower ligands attached to the ring. Both pathways build tetrapyrrole uroporphyrinogen-III with glutamate after both pathways deviate. They become similar again after cob(II)yrinate a,c-diamide synthesis. What differs is the sequence of reactions leading up to a cobalt insertion. Cobalt is inserted after a shorter sequence of reactions in the anaerobic pathway.

De novo cobalamin synthesis has the most reaction steps of any B-vitamin. Given the number of reactions involved in its synthesis, the pathway was initially split into three sections based on sections described in MetaCyc. These sections were: tetrapyrrole biosynthesis from glutamate (section 1), cob(II)yrinate a,c-diamide biosynthesis (early cobalt insertion) (section 2) and adenosylcobalamin biosynthesis from cobyrinate a,c-diamide I (section 3)

Curation results

For an overview of the anaerobic pathway as found in the GSM, see Figure 4.2. No genetic evidence for any of the enzymes catalysing the eleven reactions of section two was initially found in *awoodiiCyc*. This was because the *A. woodii* genome annotations for enzymes in this pathway had been annotated erroneously in RefSeq. The aerobic version of this pathway was the version found in the model. These were removed and replaced by reactions in the anaerobic version. After this, a genetic basis could be established for 27 out of the 29 reactions in the biosynthesis pathway as a whole.

Section 1 consists of six reactions, all of which were already in the GSM.

Section 2 consists of fifteen reactions. Two reactions were missing in the GSM and lack genetic evidence for their inclusion: cobalt-precorrin 3 synthase, coded by *CbiH60* (identified in *Bacillus megaterium*) which has no enzyme association and cob(II)yrinate a,c-diamide reductase which can occur spontaneously, coded by *CobR* and *FldA* and again has no enzyme association. These two reactions were added to the model.

Although the reactions for adenosylcobalamin synthesis were now present in the model, adenosylcobalamin could not be exported using equation 2.5. Since large sections of the pathway had inconsistent enzyme subsets and dead reactions, temporary transporters were

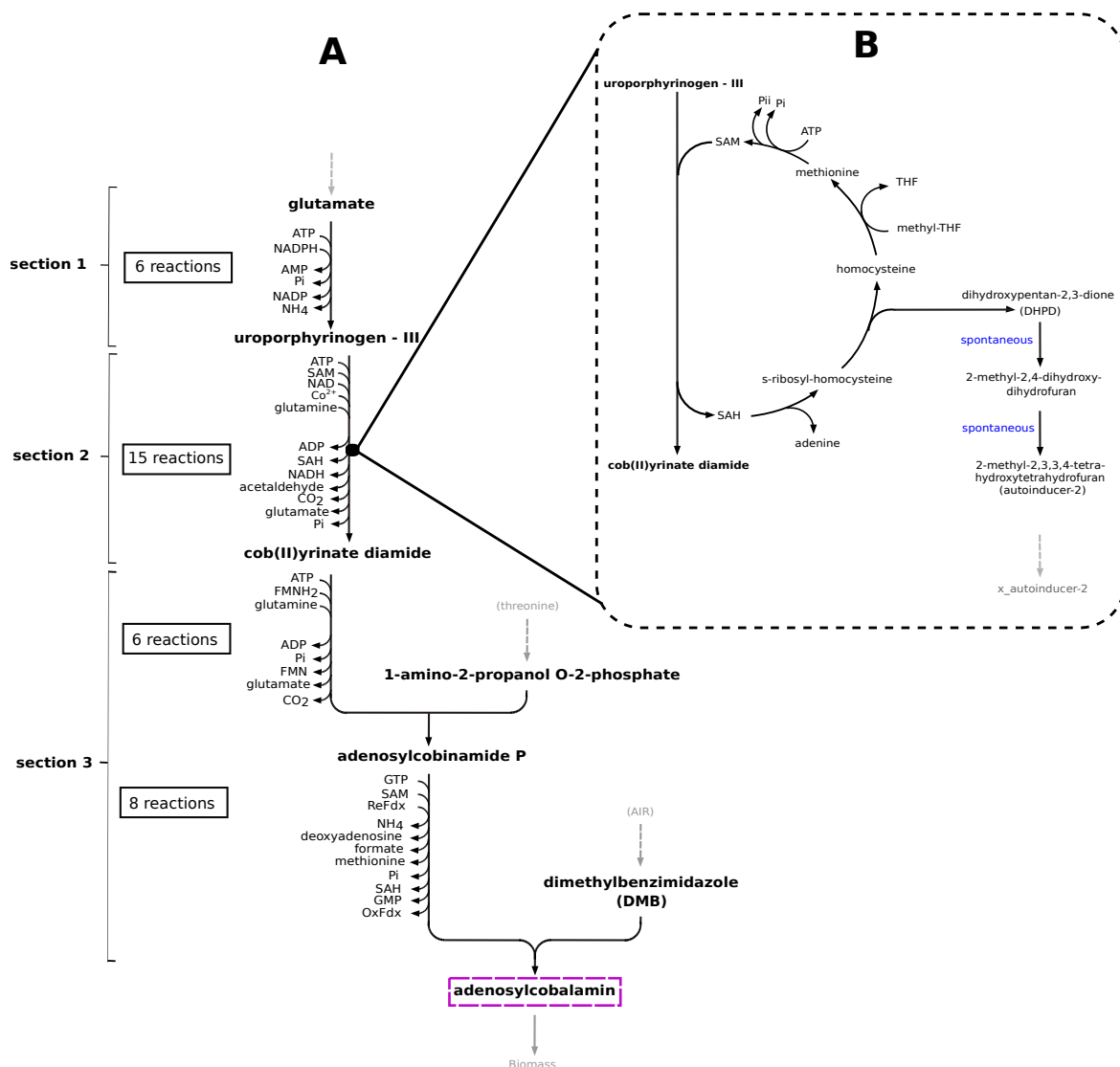


Figure 4.2: **Overview of the curated adenosylcobalamin (vitamin B12) biosynthetic pathway.** Shown in **[A]** is an overview of the anaerobic cobalamin synthesis pathway as found in the curated GSM. Important intermediates are shown, and arrows represent several reactions. The pathway was split in sections during curation which is also indicated on the figure. Highlighted in **[B]** is how the S-adenosyl methionine regeneration cycle is linked to autoinducer 2 export and adenosylcobalamin synthesis.

used to test whether intermediates of the pathway could be exported. This revealed that uroporphyrinogen-III (at the end of section 1 in Figure 4.2), the last metabolite of tetrapyrrole biosynthesis from glutamate could be exported.

The second section of the pathway, cob(II)yrinate a,c-diamide biosynthesis (early cobalt insertion) has in it eight methyltransferase reactions using S-adenosyl-L-methionine (SAM) as the methyl donor. The LP solutions for exporting the intermediates precorrin-2 and sirohydrochlorin, showed that SAM was being generated in reactions other than the SAM regeneration cycle (PWY-6151) and that these were the only source of SAM, indicating that SAM regeneration reactions could not carry flux. This was indeed the case. Upon further inspection it was found that it was dihydroxypentanedione that could not be balanced, as all reactions involved with it were dead. Dihydroxypentanedione spontaneously transforms into (2R,4S)-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran ((R)-THMF). Since (R)-THMF is a dead-end metabolite (see Introduction, section 1.4 for a definition of dead-end metabolite), and is known to be exported in other bacteria as a quorum sensing compound called autoinducer 2 (AI2), a transporter for this was added to the model. This resolved both SAM regeneration, and the rest of the reactions in cob(II)yrinate a,c-diamide biosynthesis. See Figure 4.2B for a schematic of how SAM regeneration is linked to this pathway.

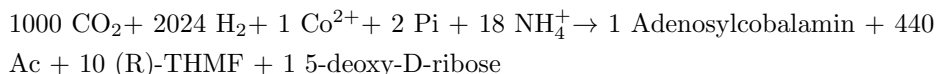
Section three the final section is adenosylcobalamin synthesis from cobyrinate a,c-diamide I, where two ligands are attached to the main corrin ring. These are an aminopropanol group which is attached to the propionate side chain of the corrin ring and the lower ligand, dimethylbenzimidazole (DMB). Phosphorylation of threonine with a threonine kinase followed by a decarboxylation reaction with threonine-phosphate decarboxylase provide the aminopropanol group. A threonine kinase reaction was not in model and no genes for it were found in *awoodii*Cyc. Using MetaCyc to search through operon based genome annotations, an operon 3684bp long with three genes in it Awo_c21560, Awo_c21570 and Awo_c21580 was found and the product of Awo_c21560 is potentially a serine/threonine-protein kinase (gene *prkC*). This was found under entry EC-2.7.11.21 in *awoodii*Cyc. The enzyme is promiscuous and acts on several substrates and had been filtered out previously by the model construction pipeline but was re-written and included in the model as a threonine kinase.

Reactions involved in anaerobic DMB synthesis described in *Eubacterium limosum* [52], were linked to 5-aminoimidazole ribotide (AIR) a precursor in purine nucleotide and thiamine synthesis. Details of the final step of 5-methoxy-6-methylbenzimidazole to DMB however were not provided and is probably catalysed by more than one enzyme. The MetaCyc ID for this reaction is RXN-17125. It was atomically balanced and added to the model as a temporary reaction. Adding the synthesis reactions for DMB from AIR however did not allow for DMB export, even though AIR could be exported. As with dihydroxypentanedione, it was found that another metabolite, 5-deoxyadenosine, could also not be balanced and reactions involving it were dead. A 5-methylthioadenosine nucleosidase gene was found

to be in the *A. woodii* genome under Awo_c03690 (*mtnN*) in a five gene operon, and a transporter was added for the product of this reaction, 5-deoxy-D-ribose. This allowed for DMB synthesis. Finally, a triphosphatase reaction was added to hydrolyse triphosphate to diphosphate and phosphate.

Complete adenosylcobalamin synthesis

Complete adenosylcobalamin is synthesised from CO₂, H₂, cobalt, phosphate and ammonium when analysing the *A. woodii* GSM. The net stoichiometry of adenosylcobalamin synthesis was computed using equation 2.5 and the net stoichiometry from CO₂ and H₂ is:



A stoichiometry this large and complex would be very difficult to establish using non-computational methods, but is trivial using FBA methods once the reaction network is constructed. The net stoichiometry is a result of 117 reactions and includes several metabolic pathways including the expected: adenosylcobalamin pathways already discussed, along with the SAM regeneration cycle and the Woods-Ljungdahl pathway. Other pathways in the solution also include: a partial pentose phosphate pathway, glutamate, glycine and serine synthesis, an adenine salvage pathway, AIR synthesis, inosine-5-phosphate synthesis and pyruvate to acetate fermentation.

Most of the metabolites involved in cobalamin synthesis that are not common redox cofactors or pathway unspecific reaction by-products (e.g. CO₂, NH₄, free phosphate groups, glutamine) have a reaction connectedness of two. They are produced by one reaction and consumed by another reaction. The average connectedness of these metabolites is 2.6. The most connected metabolite was dihydrosirohydrochlorin, being connected to five reactions and five metabolites were connected to four reactions including adenosylcobalamin. The low level of connectedness of this pathway has stretches of linear reactions and minimal branch points to other metabolism.

As a result of detailed curation and reconstruction of this pathway, it was necessary to fix broken or incomplete adjacent pathways such as SAM cycling, (R)-THMF synthesis and DMB synthesis. This process identified new transporters for the model which were essential to B12 synthesis, as well as the essential coupling of SAM regeneration to B12 synthesis and (R)-THMF export.

4.2.3 Methanol Catabolism

Evidence for methanol based growth

Methanol based growth has been described in some acetogens including *A. woodii* [32, 138, 3, 127]. Measurements of methanol uptake have been reported, as well as growth by-products.

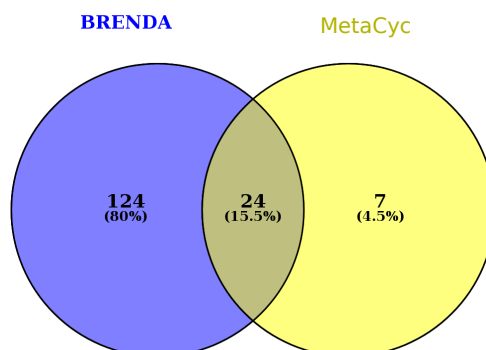


Figure 4.3: A venn diagram showing the number of unique enzymes known to be involved with methanol found in the BRENDA and MetaCyc databases.

The biochemistry however of methanol uptake into central metabolism had not been described at the time of model construction. Before specific methanol related curations were done, the GSM could not use methanol as a growth substrate after the addition of a methanol transporter. For these reasons, and for the promise of using methanol as a renewable growth feedstock for the production of valuable chemicals [25], pathways for methanol catabolism were investigated in the attempt to elucidate missing pathways.

Curation approach

Since the initial enzymes involved in the catabolism of methanol were not known, all enzymes involved with its direct metabolism were collected from MetaCyc as well as reactions involved in methanol related MetaCyc pathways. Additionally, a search for all enzymes that have methanol listed as a ligand were obtained from BRENDA via its API services. The MetaCyc and BRENDA derived enzymes were compiled into a single unique list of enzymes. Adding a reaction to the model was dependent on whether a genetic basis or PRIAM based evidence could be established for the enzyme. Additionally, the literature on methanol related catabolism was also searched for missing reactions or for modifications of existing reactions in the GSM.

Curation results

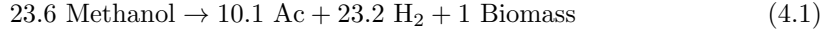
A venn diagram of enzymes collected from both BRENDA and MetaCyc databases are shown in Figure 4.3. The roles of methanol as a ligand were categorised in BRENDA as the following: inhibitor (106 enzymes), product (117 enzymes), substrate (48 enzymes), activator (57 enzymes), or reaction(21 enzymes). The number of unique enzymes that list methanol as a product, substrate or reaction is 148. From MetaCyc, 31 enzymes involved with methanol or present in methanol pathways were found. One enzyme, a methanol dehydrogenase (EC-1.1.1.244) was identified in this search that was also identified in the *A. woodii* genome with PRIAM.

Separately, the literature review found that the acetogen *Sporomusa ovata* had been studied for methanol catabolism, and a report detailing this described a catabolic pathway using demethylating corrinoid-proteins [127]. The genes for these protein were found in the *A. woodii* genome and so these reactions were added to the model. Additionally, characterisation of the methylene-THF reductase was shown to operate in a reversible direction when purified and experimented on *in vitro* [9]. The reversibility of this reaction was initially irreversible in the GSM, but was changed to a bidirectional one based on these findings.

Altogether, three new reactions were added to the model and one reaction was modified. The reactions are as follows:

1. methanol + corrinoid-protein $\rightarrow$ methylated-corrinoid-protein (added based on literature review, genetic information)
2. methylated-corrinoid-protein + THF $\rightarrow$ methyl-THF + methylated-corrinoid protein (added based on literature, genetic information)
3. methanol + NAD $\rightarrow$ formaldehyde + NADH (added based database mining and PRIAM)
4. methylene-THF + NADH $\leftrightarrow$ methyl-THF + NAD (modified based on literature review)

A spontaneous formaldehyde THF ligating reaction that produces methylene-THF was already present in the GSM, linking formaldehyde to WLP. After these changes, the model could utilise methanol as a growth substrate both as a sole carbon and energy source and as a co-substrate using equation 2.4. The net stoichiometry when using methanol for growth as sole carbon and energy source or as a co-substrate with CO₂ is:



The units are in $\text{mmol}(\text{g-DW} \cdot \text{h})^{-1}$ and $\text{g-DW} \cdot \text{h}^{-1}$ for biomass. The modifications made to the model then, allow the funnelling of methanol into WLP for energy generation and biomass production. The objective values of equation 4.1 and 4.2 are 272.7 and 232.7 respectively, showing the addition of CO_2 reduces the flux load of the network as a whole. The LP solutions for both CO_2 and CO_2 -free conditions showed that the demethylation route (reaction 1) and the oxidative route (reaction 3) are utilised simultaneously during growth.

An investigation of the essentiality of the new reactions (reactions 1-3) was carried out by serially setting a zero flux constraint on each reaction and pairs of reactions whilst solving for biomass production (data not shown). It was found that knocking out either reaction 1 or 3 did not inhibit growth, but knocking out both simultaneously did. This shows that the demethylating route via corrinoid proteins and the oxidative route via formaldehyde can theoretically function as independent growth support routes for methanol catabolism.

4.2.4 Central Metabolic Reactions

Reasons for Curation

Although much progress has been made in the characterisation of WLP and the catabolism of substrates other than CO_2 , H_2 , and CO in recent years [115], much of central acetogen metabolism remains uncharacterised. Reports on how reactions traditionally relating to the tricarboxylic acid cycle (TCA) and pentose phosphate pathway (PPP) operate in acetogens remain almost complete gaps in the scientific literature. Additionally, the question of how NADPH is generated in *A. woodii* has not been addressed since the reactions in WLP exclusively use NAD(H) and ferredoxin. Given the unusual metabolism employed by *A. woodii* and the number of novel enzymes already discovered [113, 135, 92, 146, 108], standard reconstructions of textbook versions of these pathways may be inadequate representations. The MetaCyc database holds data on TCA cycle, PPP and transhydrogenase reactions from all known metabolism, and so is a useful database for examining alternative versions of these pathways and characterising if any such pathways can be identified in *A. woodii*.

Curation approach

Reaction-associated EC numbers were collected into a unique list and checked against predictable PRIAM enzymes. Those which could be checked with PRIAM against the *A. woodii* genome were. Information on genetic evidence and whether reactions were spontaneous or not was also collected. Additionally, genes for transhydrogenase enzymes were blasted against the *A. woodii* genome but this generated no new potential hits in the genome.

Reactions were not only added to the model in these curations, but also removed if they were found to have a lack of both genetic and PRIAM based evidence. This can sometimes be the case because of BioCyc's PGDB gap-filling algorithms. In this way the model was

pruned to include reactions which had only direct evidence. See table 4.1 for a summary of curated reactions.

Tricarboxylic acid cycle curation results

There were 26 MetaCyc entries of TCA variations and 58 unique reactions. Out of the 26 pathways representing full, partial or related-to TCA pathways, none are fully represented in the *A. woodii* genome and consequently the GSM. Two pathways that are the best partially represented are ‘TCA cycle VIII’ (with MetaCyc ID: REDCITCYC and nine reactions in the pathway), and the ‘incomplete reductive TCA cycle’ (with MetaCyc ID: P42-PWY and seven reactions in the pathway).

A reaction not in the GSM but part of the afore mentioned pathways was menaquinone-dependent fumarate reductase. TCA cycle VIII additionally employs a menaquinone-dependent malate oxidoreductase that is not in the GSM. An ATP dependent succinyl-CoA synthetase (EC-6.2.1.5), employed in the incomplete reductive TCA cycle is also missing in the GSM. In total, seven of the nine reactions in TCA cycle VIII and five of the seven reactions in the incomplete reductive TCA cycle are in the GSM.

Fructose and CO₂-H₂ dependent growth were analysed using equation 2.4 and reactions related to the TCA cycle were tracked. The TCA reactions employed during growth were found to be the same and operate in the same direction in both growth conditions. Of the two partially represented pathways previously mentioned, four of the seven reactions in the incomplete reductive TCA cycle were utilised. Furthermore, the four reactions were not utilised in the same physiological direction. These reactions identified and the directions in which they carry flux are depicted schematically in Fig. 4.4 and annotations on these reactions are listed in table 4.2.

PPP curation results

There were 11 entries of PPP variations in MetaCyc with 51 unique reactions. After removal of unsubstantiated PPP reactions, shown in table 4.1, the oxidative part of the PPP became completely absent in the GSM. As in the TCA analysis, fructose and CO₂-H₂ dependent growth was analysed using equation 2.4. Again the two growth conditions utilised the same reactions and in the same directions. The non-oxidative part of PPP carried flux in the direction towards ribulose-5-phosphate synthesis.

To test whether non-glycolytic growth could occur using fructose as a growth substrate before and after removal of the oxidative PPP reactions, a constraint of zero flux was applied to phosphofructokinase whilst also solving for biomass production. This simulated a glycolysis knockout during growth. Before curation growth is possible with this knockout, and after curation it is not. That makes phosphofructokinase essential in the post-curated model when using fructose as a growth substrate.

Table 4.1: Reactions in central metabolism that were either removed or added to the model

Action Taken	Reaction ID	Description	Association	Evidence
Removed	GLU6PDEHYDROG-RXN	glucose-6-phosphate dehydrogenase	PPP	none
	XYLISOM-RXN	xylose isomerase	PPP	none
	6PGLUCONOLACT-RXN	6-phosphogluconate-lactonase	PPP	none
	RXN-8642	isocitrate dehydrogenase	TCA	none
	PEPCARBOX-RXN	phosphoenolpyruvate carboxylase	TCA	none
	RXN-9951	isocitrate dehydrogenase	TCA	none
Added	RXN0-5116	ribulokinase	PPP	PRIAM
	L-XYLULOSE-REDUCTASE-RXN	xylulose reductase	PPP	PRIAM
	RXNI-2	succinyl-CoA: acetoacetate CoA-transferase	TCA	PRIAM
	RXN-8457	2-oxoglutarate carboxylase	TCA	PRIAM
	ISOCITRATE-DEHYDROGENASE-NAD	isocitrate dehydrogenase (NAD)	TCA	PRIAM
	SUCCSEMIALDDEHYDROG-RXN	succinate semialdehyde dehydrogenase (NADP ⁺)	TCA	PRIAM
	PYRUVDEH-RXN	pyruvate dehydrogenase	TCA	PRIAM
	2-OXOGLUTARATE-SYNTHASE-RXN	2-oxoglutarate synthase	TCA	PRIAM

Table 4.2: Information on reactions related to the classical tricarboxylic acid cycle that are in the *A. woodii* GSM. How these reactions operate during growth is shown in Figure 4.4. '-' indicates no specific locus tag was identified as the reaction was identified using PRIAM.

Reaction no.	Name	EC number	Genes	
			locus tag	name
R1	phosphoenolpyruvate carboxykinase	4.1.1.49	-	<i>Pck</i>
R2	malate dehydrogenase	1.1.1.37	Awo_c16770	<i>Mdh</i>
R3	fumarase	4.2.1.2	Awo_c04220 Awo_c13650 Awo_c13660	<i>Fum, FumA</i> <i>FumB</i>
R4	aspartate aminotransferase	2.6.1.1	Awo_c08010 Awo_c17620	<i>AspB1</i> <i>AspB2</i>
R5	dihydroorotate dehydrogenase	1.3.1.14	Awo_c16190 Awo_c16200	<i>PyrK, PyrD</i>
R6	2-oxoglutarate synthase	1.2.7.3	-	<i>Oor, Kor</i>
R7	orotate reductase	1.3.1.14	Awo_c16190	<i>PyrK, PyrD</i>
R8	succinyl-CoA: acetoacetate CoA-transferase	2.8.3.5	Awo_c10980 Awo_c15700	<i>ScoA, ScoB</i>
R9	acetoacetyl-CoA transferase	2.8.3.9	-	<i>CtfA</i> <i>CtfB</i>

Transhydrogenases curation results

There were 14 unique reactions associated with transhydrogenase activity in MetaCyc. No NADPH related transhydrogenases mapped to the *A. woodii* genome

During growth on either fructose or CO₂-H₂, reactions that produce NADPH were tracked during growth (as in TCA and PPP analysis). The primary source of NADPH is via oleoyl lysophosphatidate breakdown to dihydroxyacetone phosphate by a lysophosphatidase. Lysophosphatidase however, is not essential for growth and if knocked out during growth, two other routes for NADPH generation are found. Both routes involve the reconfiguration of TCA reactions.

The first generates NADPH through a NADP⁺ dependent succinate semialdehyde dehydrogenase (reaction added to GSM during curation, see table 4.1). Succinic acid is then metabolised as seen in Figure 4.4, except that fumaric acid is no longer oxidised to succinic acid and R8 and R9 carry an increased flux. The succinate semialdehyde is a product of a 4-aminobutyrate aminotransferase (Awo_c17310, Awo_c12080) from 4-aminobutyrate and 2-oxoglutarate.

The second alternative route is achieved through an additional limb of the TCA reactions which resembles the 'start' of a classic TCA cycle. Citric acid is synthesised from oxaloacetic acid and acetyl-CoA via a citrate synthase (Awo_c19610). Aconitase (Awo_c03540) then converts citric acid to isocitric acid and isocitric acid undergoes decarboxylated dehydrogenation (Awo_c13980) forming 2-oxoglutarate, CO₂ and NADPH.

In summary of NADPH generation, at least three routes are available in the GSM during growth on fructose and CO₂, H₂:

- oleoyl lysophosphatidate breakdown
- succinic acid synthesis from succinate semialdehyde
- 2-oxoglutaric acid synthesis from isocitric acid

Further investigation into the activity of these pathways *in vivo* will help elucidate whether they play a role in NADPH generation in *A. woodii*.

4.3 Model Overview

In Figure 4.5, a high level overview of the GSM is shown. The model is mass balanced, conserves energy and can produce precursors for an *A. woodii* cell using known growth-substrates. The known growth substrates have dedicated transporters. Essential nutrients that were found to be necessary for any biomass production computation also have dedicated transporters. These nutrients are phosphorus, sulphur, ammonium, cobalt and pantothenate. Cobalt is essential for adenosylcobalamin synthesis and pantothenate for CoA and fatty

acid synthesis. Other features of the completed GSM are presented here. The PyoCyc interface to awoodiiCyc was used to gather data on how well the model encompasses genetic information. The connectedness of a metabolite and reaction subset analysis was done using ScrumPy features described in chapter 1, section 1.5.

4.3.1 General Metrics

A. woodii GSM contains 841 reactions, 867 metabolites and 98 transporters. 375 of these reactions are dead (39.9%) and 326 metabolites are orphans (37.6%). Of these transporters, 23 are growth substrates, 15 are by-products (e.g. acetate and hydrocaffeate), 6 are essential micronutrients, and 57 are transporters for biomass precursors. 5 transporters are treated as both growth substrates and by-products (e.g. ethanol, CO₂, and formate). Of the 3657 genes reported for the *A. woodii* genome in RefSeq, 951 are involved in metabolic reactions and found in awoodiiCyc and after curation and model construction 568 are represented in the GSM.

The top 15 connected metabolites and the number of reactions they are involved in are as follows: CO₂, 57, ammonium, 50, glutamate, 50, pyruvate, 29, 2-oxoglutarate, 29, S-adenosylmethionine, 22, S-adenosyl-L-homocysteine, 22, glutamine, 21, UDP, 19, acetate, 18, CTP, 16, GTP, 15, 5-phosphoribosyl diphosphate and glucose 1-phosphate, 13. The following metabolites were excluded from the above list: ATP, ADP, inorganic phosphates, NAD(P)H and NAD(H). If included, ATP would be the most connected, being involved in 136 reactions.

The number of reaction subsets in the GSM were also computed and are as follows:

- two reaction subset: 64
- three reaction subset: 21
- four reaction subset: 4
- five reaction subset: 5
- more than five reaction subset: 9

The largest reaction subset had in it 25 reactions. Four of the reactions are in the tetrapyrole biosynthesis pathway, two are in aminopropanol-phosphate biosynthesis and twelve are in cobyrinate-diamide biosynthesis. All but seven reactions are related to adenosylcobalamin synthesis. A nine reaction subset contains all five DMB biosynthesis reactions and also includes adenosylcobalamin export. The low level of metabolite connectedness mentioned in the previous section on adenosylcobalamin synthesis (section 4.2.2) is the reason for the size of this 25 reaction subset.

One of the five reaction subsets has in it four reactions of the seven reaction isoleucine biosynthesis pathway (from threonine, MetaCyc ID: ILEUSYN-PWY). A three reaction

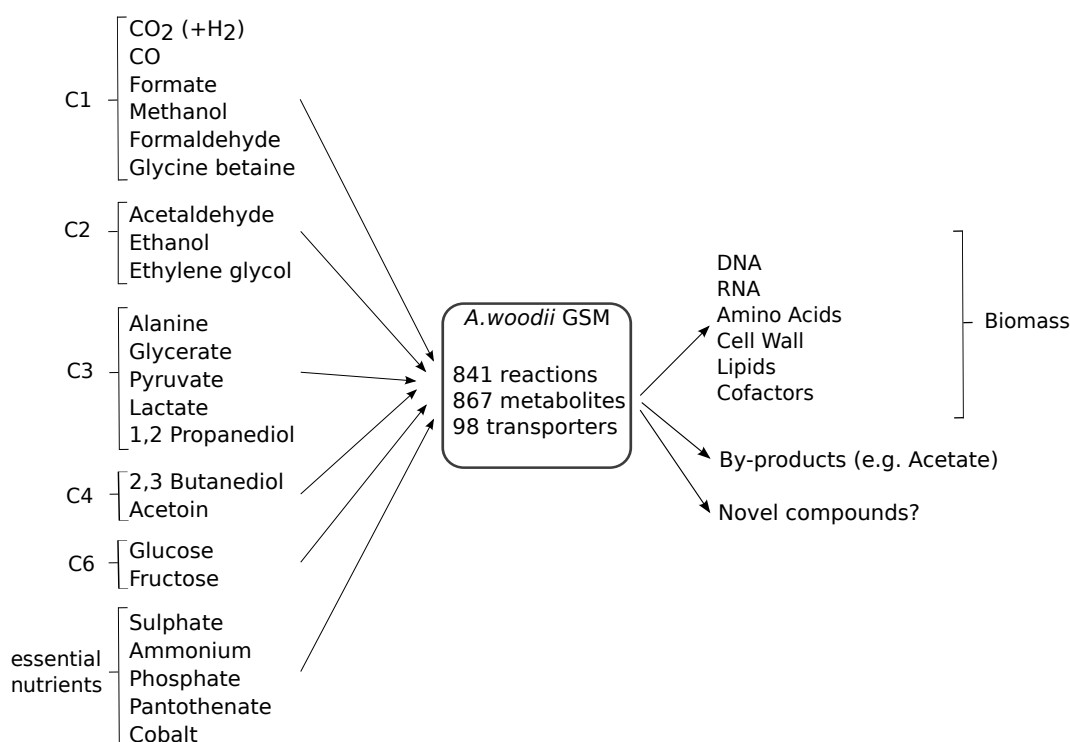


Figure 4.5: **A high level view of *A. woodii* Genome-scale Model (GSM).** The possible inputs to the model are on the left and outputs are on the left. The black box encompasses GSM metrics. Essential micronutrients are needed to produce biomass components.

subset that independently maps to methionine degradation II (PWY-701), overlaps with reactions in the isoleucine pathway. So in effect, reactions related to methionine degradation are linked to isoleucine synthesis. Thirteen of the 21 three reaction subsets do not match to any MetaCyc pathway. These are reactions that can be treated as a single metabolic unit and so behave, in essence, like a pathway but are not linked to any defined pathway. However, some reaction subsets do map well to MetaCyc pathways. The five, five reaction subsets include ones relating to tryptophan, peptidoglycan, autoinducer-2 synthesis and export, and 5-aminoimidazole ribonucleotide (AIR) biosynthesis.

Unlike in the reaction subsets of the model discussed in chapter 3, reactions in the WLP are not in any reactions subsets. indicating branching of the metabolic network in this pathway to other parts of metabolism.

4.4 Discussion

4.4.1 Overview

The result of the detailed curation process presented in this chapter is a complete GSM of *A. woodii*. It is validated in that it meets internal model consistency requirements by conserving mass and energy, and that it can use known growth substrates to support biomass production (see [115] for a review of growth substrates). Several recognisable synthesis pathways were revealed by reaction subset analysis, indicating that growth solutions are rooted in established biochemistry. The variation in curation approaches shown here is an example of why it is difficult to fully automate accurate GSM reconstructions. In the case of lysine synthesis, the issue of not having a lysine biosynthesis pathway present in *A. woodii* had to be resolved. For adenosylcobalamin several issues had to be addressed including erroneous gene annotations and missing transporters. Elucidating catabolic methanol pathways required a two-pronged approach of database mining and a literature review, and pruning central metabolic reactions made a significant difference to biomass production solutions. Additionally, the process of curation highlighted gaps in scientific literature related to acetogen and specifically *A. woodii* metabolism. Since dead reactions are effectively reactions disconnected from flux carrying networks, the number of dead reactions in the model also highlights gaps in the biochemical characterisation of many reactions and the opportunity for novel discoveries. The analysis of metabolite connectedness revealed the top fifteen connected metabolites. Examination of metabolism adjacent to these highly connected metabolites would provide further insight into central metabolism.

Lysine synthesis

Pathways of *de novo* lysine biosynthesis have not been elucidated in any acetogen. The pathway defining enzyme in the MetaCyc pathway ‘L-lysine biosynthesis VI’ is L,L-diaminopimelate aminotransferase, which was first discovered in green plants but also identified in cyanobacteria and archaeobacteria, namely methanobacteria and methnocoeci [75, 80].

Although lysine can now be synthesised in the GSM, the addition of a temporary reaction catalysed by diaminopimelate epimerase (EC-5.1.1.7) was necessary. Since no complete pathway could be mapped to the *A. woodii* genome, further biochemical investigation is needed. No identification of a suitable diaminopimelate epimerase enzyme might suggest a novel reaction or pathway may be responsible in *A. woodii* for lysine synthesis.

Adenosylcobalamin (vitamin B12) synthesis

During reconstruction of the adenosylcobalamin pathway it was found that export of (R)-THMF - also known as autoinducer 2 (AI2) because of its involvement in quorum sensing (QS) in some bacteria - is essential for adenosylcobalamin synthesis. Several B12 yields

Table 4.3: Vitamin B12 yields in *A. woodii* when using different carbon and energy sources for growth.

Substrate	Yield (nMol/gDW)	reference
CO ₂ -H ₂	650	[129]
Fructose, N ₂	1210	[129]
Methanol-CO ₂	2672	[5]
Methanol-Formate	3564	[4]

have been reported in different growth conditions and are shown in table 4.3. It seems that methanol based growth stimulates more B12 production than fructose or autotrophic growth. If (R)-THMF is involved in QS for *A. woodii* as it is in other bacteria, then the stoichiometry above indicates that an increasing B12 yield would result in increased (R)-THMF export. Whether these increases could be significant has not been explored but it raises the question whether the availability of growth substrates could induce communication of *A. woodii* within a microbial community. Recent investigations on quorum sensing in the acetogen *Clostridium autoethanogenum* [87], revealed that the Agr QS system regulates the WLP and may do so to manage carbon-source utilisation. Based on the findings described here on B12 synthesis, the same group were able to detect (R)-THMF in *A. woodii* growing on vitamin B12 deficient media (personal communication).

The average connectedness of metabolites in the adenosylcobalamin pathway was 2.6. When comparing this to metabolites in a pathway like glycolysis, which has an average connectedness of 9.4 (with pyruvate unsurprisingly being the most connected to 30 reactions), it becomes clear how linear much of adenosylcobalamin synthesis is. It would be of interest to investigate the genetic organisation and transcriptional regulation of it to see if the regulation of the expression of enzymes is in any way indicative of its metabolic structure.

Tricarboxylic acid cycle in *A. woodii*

None of the 26 variants of the MetaCyc TCA cycles mapped to the *A. woodii* genome. The closest pathways were ‘TCA cycle VIII’ (MetaCyc ID: REDCITCYC) found in *Helicobacter pylori* [91] and ‘incomplete reductive TCA cycle’ (MetaCyc ID: P42-PWY), found in the methanogen *Methanococcus maripaludis* [124]. Both of these pathways employ quinones which are missing in *A. woodii* [108].

Analysis of the TCA cycle during biomass production shows that the reactions do not operate as a cycle. They instead converge on 2-oxoglutarate synthesis. Oxalacetic acid and glutamate are converted to 2-oxoglutarate and aspartate, bypassing the classical route via citrate synthase [152]. 2-oxoglutarate is also synthesised from succinyl-CoA in a reaction dependent on reduced ferredoxin. Acetyl-CoA is the donor for the CoA group needed to convert succinate into succinyl-CoA, and the acetyl group is excreted as acetate. Interestingly, the only fumarate oxidase, found in the *A. woodii* genome is a dihydroorotate dependent

one. This is utilised to form succinate and the product of this reaction, orotate, is used to form uridine monophosphate (UMP). As a consequence of inosine monophosphate (IMP) synthesis, itself a precursor metabolite for purines, some fumarate is produced. According to the reactions in 4.4, two NADH and one reduced ferredoxin are oxidised, and one ATP is formed if 2-oxoglutarate is considered an exit point. This is because, much of the classical TCA cycle reactions operate in reverse.

Pentose-phosphate pathway in *A.woodii*

The incorrect inclusion of the oxidative pentose-phosphate pathway in the awoodiiCyc database is a result of over a liberal pathway annotation by the gap-filling PathogLogic algorithm. It is a pathway that allows for NADPH generation and so, when removed from the model, an alternative pathway for NADPH generation is needed. It is also an alternative way to oxidise hexoses in many cell types [152] and so an alternative glycolytic route is not present in the GSM, which makes 6-phosphofructokinase an essential reaction when growing on fructose. By having full representation of the non-oxidative PPP interconversion between 3, 4 and 5 carbon skeletons is possible for synthesis of important biomass precursor metabolites.

Transhydrogenases in *A.woodii*

The generation of NADPH is of central importance to biosynthetic reactions but it has not been investigated in *A. woodii*. No additional transhydrogenases were found in this investigation. The only known transhydrogenases in *A. woodii*: the Rnf complex (a ferredoxin:NAD oxidoreductase) and the electron-bifurcating hydrogenase, are both ferredoxin and NAD⁺ specific [11, 112] and already in the GSM. The methylene-THF dehydrogenase reaction in WLP is also NAD⁺ specific in *A. woodii* [9]. The GSM is a useful first step in hypothesising how NADPH could be metabolised. By investigating its generation, it was revealed that the TCA reactions which operate the same in both autotrophic and heterotrophic growth, are capable of being reconfigured to meet NADPH demands for growth if an NADPH constraint is set in the LP. Three pathways in total are identified for further investigation into NADPH metabolism.

Methanol catabolism

Many enzyme reactions that are found uniquely in BRENDA because it contains details about enzyme promiscuity, not often included in MetaCyc descriptions of the enzyme [15]. For example, an alcohol dehydrogenase (EC-1.1.5.5), listed in BRENDA as being able to oxidise methanol to formaldehyde with several different redox cofactors but with very low activity, is found in MetaCyc as a reaction converting ethanol to acetaldehyde using ubiquinol and ubiquinone, the preferred substrates and products for this enzyme. An example of re-

action unique to MetaCyc is a deoxypodophyllotoxin oxidative demethylase reaction that is inferred from experiment and where methanol production is hypothesised but not observed [72].

Since attempting to reconstruct methanol catabolic pathways bioinformatically here, a report was published on methanol catabolism in *A. woodii* [71]. It provides transcriptomic data for the demethylating route via corrinoid proteins in a way that is consistent with the demethylation route described here. It did not however investigate the potential for an oxidative route via formaldehyde. Having only this demethylating route would be problematic because of the high endergonic reaction of methylene-THF reductase operating in reverse (already discussed in chapter 3, section 3.4.2).

There are no reports of *A. woodii* growing on methanol as a sole carbon and energy source, and indications from experimental scientists suggest it does not occur (personal communication). According to the stoichiometry in equation 4.1, although methanol would support growth, nearly as many moles of methanol consumed are released as moles of H_2 . This is a very wasteful use of electrons which could otherwise be used to synthesis ATP, and is a result of not being able to balance redox intermediates. If CO_2 is introduced into the system, no electrons are wasted as H_2 because CO_2 allows for electrons to be used for its reduction in WLP. The increased efficiency of producing biomass when CO_2 is present is reflected in the decreased objective value of 232.7 from 272.7. Since the objective here was to minimise the sum of flux through the metabolic network, which is a proxy for minimising the proteomic investment of the cell, the lower value when CO_2 is present indicates the cell would need less enzyme protein to produce the same amount of biomass. However, modelling analysis still shows that methanol could in theory be a sole growth substrate. Why this does not appear to occur *in vivo* would need further investigation.

Potential further development of the GSM

Although the GSM has undergone successful validation here, it is possible to further develop it to encapsulate additional features of *A. woodii* metabolism. An interesting area of development could be the addition of *A. woodii* specific transporters. The only transporters researched in detail have been the Rnf complex and the ATP synthase [56, 79]. The transporter identification algorithms and transporter-centric databases in TransporterDB2.0 have recently been significantly improved and further automated [34]. This has made it possible to bioinformatically identify transporters from genetic sequences in web-based tools. A preliminary query into transporters present in *A. woodii* show that the *A. woodii* genome contains the multicomponent phosphotransferase system (PTS). An efflux transporter for autoinducer 2 (AI-2E family of transporters) was also found, adding to the evidence already presented on this above.

Chapter 5

Bioenergetics of *A. woodii* Metabolism

5.1 Introduction

For insights into bioenergetic aspects of *A. woodii* metabolism, the Genome-scale model (GSM) described in chapter 4 was analysed. It was shown by elementary modes analysis in chapter 3 that the WLP can be comprehensively reduced to smaller networks, that represent routes through it, by elementary modes analysis. In this way, the array of product and ATP yields potentially feasible by that network is revealed. This type of analysis however is currently only possible on networks smaller than those typically found in a GSM because of limitations in computer processing power. Since GSMs represent the entire metabolic repertoire of an organism, its reliance is less on subjective curation of ‘independent’ metabolic pathways and analysis of it can identify potentially novel reactions involved in metabolism. Also, the wide range of substrates available to the *A. woodii* GSM, allows for bioenergetic analysis across a wider spectrum of relevant growth conditions.

In an attempt to gain insight into the bioenergetics of *A. woodii* across many conditions, the GSM was analysed. In particular, attention was directed at comparing ATP yields calculable with a model to those reported in the literature, and whether an available carbon source and being able to generate ATP alone is enough to sustain growth. Further attention was directed to whether these ATP yields are enough to support growth in light of reports on *A. woodii* maintenance costs.

5.2 Y_{ATP}^{MAX} Calculations

Y_{ATP} can be calculated using various LP formulations. Additionally, when searching for a Y_{ATP}^{max} of a given substrate, the unit used to quantify the ATP becomes crucial. In the

literature, various units are used to describe ATP quantities, including ATP per acetate, ATP per ethanol, ATP per glucose etc. This variation in units makes comparing these ATP quantities difficult. To consistently evaluate $Y_{\text{ATP}}^{\text{max}}$ and compare Y_{ATP} across many conditions, $\text{ATP} \cdot \text{Cmol}^{-1}$ was used in ATP yield calculations. Four LP formulations were then compared to see how frequently each LP calculated the $Y_{\text{ATP}}^{\text{max}}$ using the specified units. The four LPs were formulated using the base LP (equation 2.2) outlined in chapter 2, section 2.2. Specifically, these were:

$$\begin{array}{ll}
 \text{[miniC]} & \begin{array}{l} \text{minimise} \quad : \quad v_j \cdot c \\ \text{subject to} \quad \left\{ \begin{array}{l} \mathbf{N}\mathbf{v} = \mathbf{0} \\ v_{j_{\text{nonS}}} = t_{j_{\text{nonS}}} - 100 \leq t_{j_{\text{nonS}}} \leq 0 \\ v_{j_{\text{-sx}}} = 0 \\ v_{\text{ATPase}} = 1 \end{array} \right. \end{array}
 \end{array} \quad (5.1)$$

where: v are model reactions; j_{sx} are growth substrates not relevant to the analysis; $v_{j_{\text{nonS}}}$ are essential micronutrients; v_{ATPase} is the rate of the ATP hydrolysis reaction; and c is a carbon weighting cost. This 5.1 also minimises substrate import but with an additional weighting for the amount of carbon in that substrate, and then sets the ATPase flux to one.

$$\begin{array}{ll}
 \text{[miniT]} & \begin{array}{l} \text{minimise} \quad : \quad v_j \\ \text{subject to} \quad \left\{ \begin{array}{l} \mathbf{N}\mathbf{v} = \mathbf{0} \\ v_{j_{\text{nonS}}} = t_{j_{\text{nonS}}} - 100 \leq t_{j_{\text{nonS}}} \leq 0 \\ v_{j_{\text{-sx}}} = 0 \\ v_{\text{ATPase}} = 1 \end{array} \right. \end{array}
 \end{array} \quad (5.2)$$

This minimises substrate import and sets the ATPase flux to one.

$$\begin{array}{ll}
 \text{[maxi]} & \begin{array}{l} \text{maximise} \quad : \quad v_{\text{ATPase}} \\ \text{subject to} \quad \left\{ \begin{array}{l} \mathbf{N}\mathbf{v} = \mathbf{0} \\ v_{j_{\text{nonS}}} = t_{j_{\text{nonS}}} - 100 \leq t_{j_{\text{nonS}}} \leq 0 \\ v_{j_{\text{-sx}}} = 0 \\ v_{J_{\text{-s}}} = \text{None}, 1 \end{array} \right. \end{array}
 \end{array} \quad (5.3)$$

This sets a maximisation objective on the ATPase reaction and puts upper bounds on substrate import.

$$\begin{array}{ll}
\text{minimise} & : |v| \\
\text{[mini] subject to} & \left\{ \begin{array}{l} \mathbf{N}\mathbf{v} = \mathbf{0} \\ v_{j_{nonS}} = t_{j_{nonS}} \quad -100 \leq t_{j_{nonS}} \leq 0 \\ v_{j_{-sx}} = 0 \\ v_{\text{ATPase}} = 1 \end{array} \right. \quad (5.4)
\end{array}$$

This minimises total flux and sets the ATPase flux to one.

To assess which of the LP formulations most consistently computes the $Y_{\text{ATP}}^{\text{max}}$ of any given substrate, the $Y_{\text{ATP}}^{\text{max}}$ of many substrate combinations was calculated and compared using the *A. woodii* genome-scale model. The 23 growth substrates available in the model, and all paired combinations of these were used for $Y_{\text{ATP}}^{\text{max}}$ calculations, which was 276 conditions in total.

Once these equations are solved $\text{ATP} \cdot \text{Cmol}^{-1}$ was calculated from the solution. For substrates that are not used as either energy or carbon source, such as caffeate, the carbon content was not considered to the carbon weighting on equation 5.1 or to $\text{ATP} \cdot \text{Cmol}^{-1}$ calculations. Also, although glycine betaine is a five carbon compound, only one of its methyl groups is used as a carbon/energy source, and the resultant dimethylglycine is excreted [73] and so this compound's carbon content was considered as one.

5.2.1 $Y_{\text{ATP}}^{\text{max}}$ using four LPs and substrate combinations

Shown in Figure 5.1 are the number of substrates each particular LP was able to calculate the $Y_{\text{ATP}}^{\text{max}}$. All equations calculated at least one unique $Y_{\text{ATP}}^{\text{max}}$ and no single equation was able to calculate the $Y_{\text{ATP}}^{\text{max}}$ across all conditions. When maximising the ATPase (equation 5.3), the $Y_{\text{ATP}}^{\text{max}}$ was calculated in 106 conditions. When minimising total flux (equation 5.4), the $Y_{\text{ATP}}^{\text{max}}$ was calculated in 109 conditions. When minimising substrate import (equation 5.2), the $Y_{\text{ATP}}^{\text{max}}$ was calculated in 207 conditions. When minimising substrate import with a carbon weighting (equation 5.1), the $Y_{\text{ATP}}^{\text{max}}$ was calculated in 251 conditions.

Minimising substrate import with a carbon weighting calculated the $Y_{\text{ATP}}^{\text{max}}$ across most conditions and also calculated the most number of unique $Y_{\text{ATP}}^{\text{max}}$'s, totalling 25. The conditions include acetoin-CO (0.38 $\text{ATP} \cdot \text{Cmol}^{-1}$), 2,3-butanediol-methanol (0.55 $\text{ATP} \cdot \text{Cmol}^{-1}$) and alanine-formaldehyde (0.85 $\text{ATP} \cdot \text{Cmol}^{-1}$).

Minimising substrate import calculated the $Y_{\text{ATP}}^{\text{max}}$ in the second most conditions. The eight conditions where the unique $Y_{\text{ATP}}^{\text{max}}$ was calculated all used glycine betaine as a co-substrate, including glycine betaine-glycerate (0.46 $\text{ATP} \cdot \text{Cmol}^{-1}$), glycine betaine-1,2-propanediol (0.23 $\text{ATP} \cdot \text{Cmol}^{-1}$) and glycine betaine-acetaldehyde (0.35 $\text{ATP} \cdot \text{Cmol}^{-1}$).

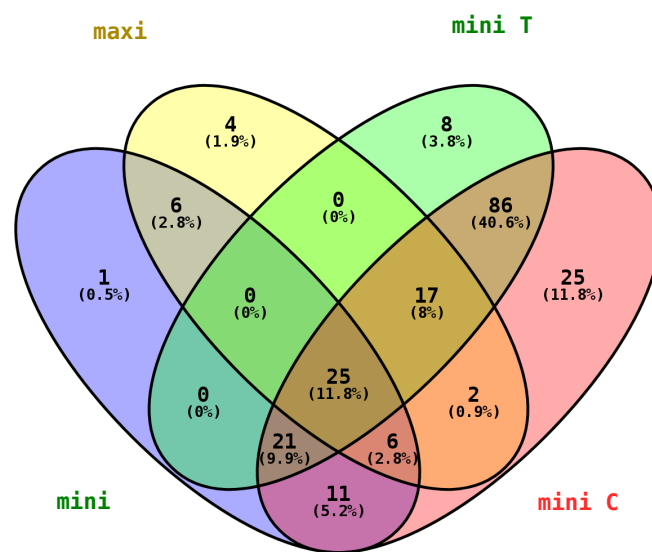


Figure 5.1: Venn diagram showing the success rate of four ATP LPs at calculating $Y_{ATP}^{\max}$ of 276 substrates and substrate combinations. The labels correspond to the equation labels found in text.

Table 5.1: ATP yields of ethanol-CO₂, calculated with the four ATP LPs. The growth stoichiometry was calculated using the *A. woodii* GSM at a growth rate of 0.1 h⁻¹ using equation 2.4

Stoichiometry	LP	No. Reactions	Y _{ATP}
EtOH, CO ₂ → Form, Ac, H ₂	maxi	14	0.13
2 EtOH, 2 CO ₂ → 3 Ac	mini	19	0.18
EtOH → Ac, 2 H ₂	miniT, miniC	11	0.20
15.16 EtOH, 14.96 CO ₂ → 20.90 Ac, Biomass	growth	313	-

Maximising ATPase calculated the $Y_{ATP}^{\max}$ in the third most conditions. The ten conditions where the unique $Y_{ATP}^{\max}$ was calculated include acetoin-H₂ (0.35 ATP·Cmol⁻¹), CO-H₂ (0.45 ATP·Cmol⁻¹) (also in table 3.1 by mode 10), and formaldehyde-caffeate (1.68 ATP·Cmol⁻¹).

Minimising total flux calculated the $Y_{ATP}^{\max}$ in the least number of conditions, but still calculated it in 106 of them. It also calculated one unique $Y_{ATP}^{\max}$ which was glycine betaine-formate (0.1 ATP·Cmol⁻¹).

ATP yields of Ethanol-CO₂

The four ATP LPs were used to calculate Y_{ATP}'s of ethanol-CO₂. Three Y_{ATP}'s were found, one each by minimising total flux and maximising ATPase, and one by minimising substrate and carbon import, see table 5.1.

The LP that calculated the smallest Y_{ATP} is the one maximising ATPase (equation 5.3). It uses ethanol and CO₂ in equal measure to produce formate, acetate and H₂ in equal amounts. This is done by ethanol oxidation to acetyl-CoA via acetaldehyde. ATP is then generated by acetate production from acetyl-CoA. The ATP and NADH produced is then used to generate the sodium gradient, producing reduced ferredoxin. NADH and reduced ferredoxin then generates H₂ with the electron bifurcating hydrogenase, half of which is exported and half is used for formate production with CO₂. In this way 0.13 ATP·Cmol⁻¹ is produced

The LP that calculated the second highest amount of Y_{ATP} is the one minimising total flux (equation 5.4). It also uses ethanol and CO₂ in equal measure but produces acetate as the sole product. Ethanol is again oxidised to acetyl-CoA. Additional acetyl-CoA is produced via the WLP. The H₂ needed for reducing CO₂ to formate is generated from the electron bifurcating hydrogenase. Thus 0.18 ATP·Cmol⁻¹ or 0.36 ATP per acetate is produced.

Both LPs minimising import (equations 5.2 and 5.1) calculated the highest amount of Y_{ATP}, the $Y_{ATP}^{\max}$. In this solution, acetate is produced solely from ethanol oxidation to acetyl-CoA, forgoing CO₂ import. The ATP and NADH produced powers the sodium

Table 5.2: Various substrate $Y_{\text{ATP}}^{\text{max}}$ calculated using the *A. woodii* GSM and ATP LPs. The substrates here are a subset of the full table in Appendix B and were chosen based on the considerable attention they have recieved in the literature. The $Y_{\text{ATP}}^{\text{max}}$ figures for each substrate and whether it is growth supporting or not is compared to literature reports. PC refers to personal communication.

Substrate	$Y_{\text{ATP}}^{\text{max}}$ model	Y_{ATP} literature	Growth Supporting		Ref
			model	literature	
MetOH, Caffeate	3.12	3	yes	yes	[60, 138]
Formaldehyde	0.85	-	yes	yes	[32]
Fruc	0.72	0.72, 0.83	yes	yes	[115, 138, 47]
Lact, Fruc	0.72	-	yes	yes	[109]
Glycerate, CO ₂	0.46	-	yes	yes	[6]
Alanine	0.38	0.38	yes	mixed	[3, 31]
Ethylene-Glycol, CO ₂	0.35	0.31	yes	yes	[106, 135]
Acetaldehyde, CO ₂	0.35	-	yes	yes	PC
1,2-PDO, CO ₂	0.23	-	yes	yes	[116]
CO ₂ , EtOH	0.2	0.18	yes	yes	[17, 10]
Lact	0.18	0.13	yes	yes	[88, 146]
CO ₂ , H ₂	0.15	0.15, 0.25-0.5, 0.75-1	yes	yes	[6, 43, 114, 138]
Glycine-Betaine, CO ₂	0.10	0.37-0.42	yes	yes	[73]
Form	0.08	0.5	yes	yes	[67, 8, 29]
Acetoin	0.35	-	no	no	[57]
2,3-BDO	0.27	-	no	no	[57]
1,2-PDO	0.23	0.25	no	-	[116]
EtOH	0.2	-	no	no	[10]
CO ₂	0.0	-	no	-	

gradient which allows for reduced ferredoxin production and ultimately H₂ production and export. Thus 0.2 ATP·Cmol⁻¹ or 0.4 ATP per acetate is formed.

When computing biomass using the GSM, the growth stoichiometry used is most similar to substrate to product stoichiometry produced by the LP minimising total flux. This stoichiometry is the one observed *in vivo* [10]. This objective then, appropriate to compute growth, but which calculates a lower ATP than possible indicates that maximisation of ATP is not the only optimisation during growth. The ethanol-CO₂ catabolic route, via the acetaldehyde and WLP is also the one found in the biomass solution and described in the literature. Biomass is produced with 313 total reactions, comprising the rest of the reactions needed for growth other than catabolism of ethanol-CO₂.

5.2.2 ATP yields table

The $Y_{\text{ATP}}^{\text{max}}$ was calculated for the 276 conditions described in section 5.2.1, and if available, compared to a Y_{ATP} reported in the literature. Whether the substrate or substrate pair supported biomass production in the GSM was also compared to literature reports on growth where possible. Substrates that have received considerable attention in the literature are presented in table 5.2, and the full table can be found in appendix B. Other non- $Y_{\text{ATP}}^{\text{max}}$ ATP yields were also calculated for substrates in a fashion similar to those presented in table 5.1.

Of the 276 substrates, 212 could be used to generate ATP and support growth, 45 could generate ATP but not support growth, 16 generated no ATP and could not support growth, and a $Y_{\text{ATP}}^{\text{max}}$ could not be calculated for 3 but these did not support growth. Of the 276 substrates, 53 had some reporting in the literature, and 51 had specific reports on being either growth or non-growth supporting. Of these 51 conditions, growth or no-growth was in concordance with GSM predictions for 48 substrate conditions (94% accuracy). In conditions where growth reports were mixed, with alanine as sole substrate and CO as sole substrate, the model predicted growth was feasible. The 45 substrates that could be used to generate ATP but did not support growth, could support growth if CO₂ was supplied as an additional input, although CO₂ was only consumed in trace amounts. The 16 substrates that could not generate ATP or support growth, did so because those substrates did not supply either carbon or energy. The 3 conditions for which no $Y_{\text{ATP}}^{\text{max}}$ could be calculated were H₂ paired with either caffeate, ferulate or coumarate, and this was because ATP could be generated by H₂ oxidation without utilising carbon which is problematic when defining ATP yields in units mol·Cmol⁻¹.

A higher $Y_{\text{ATP}}^{\text{max}}$ was calculated than a Y_{ATP} reported in the literature in four of the conditions shown in table 5.2. These were methanol-caffeate, ethylene glycol-CO₂, ethanol-CO₂, and lactate. Caffeate is reduced to hydrocaffeate in all of the methanol-caffeate solutions, and methanol is oxidised to CO₂ in the solution with the $Y_{\text{ATP}}^{\text{max}}$, and either acetate or formate in other ATP yielding solutions. In the ethylene glycol-CO₂ solutions, either acetate, acetate and ethanol or acetate and H₂ is produced. In the ethanol-CO₂ solutions, either acetate, acetate and H₂ or acetate and H₂ and formate are produced. In both ethylene glycol-CO₂ and ethanol-CO₂ solutions, the ones with H₂ production allowed for the highest Y_{ATP} . The lactate solutions all produced the same Y_{ATP} and all produced acetate.

A lower $Y_{\text{ATP}}^{\text{max}}$ was calculated than a Y_{ATP} reported in the literature in four of the conditions shown in table 5.2. These were CO₂-H₂, formate, glycine betaine-CO₂, and 1,2-propanediol. All the CO₂-H₂ solutions were, produced acetate and generated 0.15 ATP mol·Cmol⁻¹. In the glycine betaine-CO₂ solutions either acetate and dimethylglycine or acetate, dimethylglycine and H₂ were produced. All the formate solutions, generated 0.08 ATP mol·Cmol⁻¹ which is in concordance with the yield reported in the discussion in chap-

ter 3 (section 3.4.2), and produced either acetate and CO₂, or acetate, CO₂ and bicarbonate. In the 1,2-propanediol solutions, either propionate and propanol or propionate and H₂ were produced. the Y_{ATP}^{max} was generated when H₂ was the co-product.

The Y_{ATP}^{max} was the same as the Y_{ATP} reported in the literature in three conditions shown in table 5.2. These were for fructose, alanine and CO₂-H₂. All fructose solutions produced acetate, but one generated a lower ATP yield of 0.67 than the Y_{ATP}^{max} shown. All alanine solutions generated the Y_{ATP}^{max} and produced acetate and ammonium.

No Y_{ATP} was reported for eight of the substrates shown in table 5.2, although growth experiments data was available. These were formaldehyde, lactate-fructose, glycerate-CO₂, acetaldehyde-CO₂, 1,2-propanediol-CO₂, acetoin, 2,3-butanediol and ethanol. Formaldehyde was found to generate the highest Y_{ATP} of any single substrate, higher even than fructose. Although 1,2-propanediol generated no extra ATP with CO₂ present, CO₂ was needed to support growth. Generally, the substrates in this list which do support growth consisted of solutions that produced either acetate as sole product or with CO₂ as co-product, with the exception of 1,2-propanediol which is known to support non-acetogenic growth [116]. The substrate which did not support growth in this list produced other products such as H₂, formate or ethanol alongside acetate.

Overall, calculation of a higher Y_{ATP}^{max} than the Y_{ATP} reported, indicates that either more ATP is produced than estimated, or that maximum ATP generation is not pursued during growth. Calculation of a lower Y_{ATP}^{max} , indicates an infeasible amount of ATP was estimated to be produced. Calculation of a Y_{ATP}^{max} the same as the Y_{ATP} reported indicates the maximum amount of ATP can produced from that substrate in a way consistent with supporting growth.

5.3 Calculating maintenance costs

Cell maintenance is the energetic cost a cell must expend to grow and maintain essential cellular functions. These consist of energy spent on growth and non-growth associated processes (GAM and NGAM), and are used to parameterise GSM based analysis. Although ATP is the main energetic currency of a cell, maintenance costs are sometimes given as an amount of growth substrate or by-product formed making it difficult to compare cellular maintenance across cell types, and across conditions.

To calculate GAM and NGAM, the specific substrate catabolising rate, q_E (in mmol(g-DW·h)⁻¹) is plotted against an increasing dilution rate. [47] showed that the specific acetate formation rate, q_{Ac} (in mmol(g-DW·h)⁻¹), can also be used to calculate maintenance. This is because the growth yield per acetate (gDW·mol acetate⁻¹), when converted to amounts

Table 5.3: Six maintenance costs using different growth substrates are compared here in amounts ATP. These data were originally collected from the literature in amounts acetate, and then converted into an equivalent amount of ATP by calculating an appropriate ATP yield per acetate for the specific substrate (see in text for further explanation and references).

Substrate	Growth Rate (h^{-1})	GAM (ATP mmol(g-DW · h) $^{-1}$)	NGAM (ATP mmol(g-DW · h) $^{-1}$)
Fruc, CO ₂	0.1	6.31	0.61
MetOH, CO ₂	0.1	9.88	0.59
MetOH, Form	0.1	4.97	0.47
CO ₂ , H ₂	0.1	3.94	0.11
Lact, CO ₂ , H ₂	0.1	3.77	0.06
Lact, CO ₂	0.1	3.76	0.03

of fructose (using the stoichiometry $\text{Fruc} \rightarrow 3 \text{ Ac}$), is similar to the growth yield per fructose when corrected for cell maintenance.

The ATP gained per acetate however is dependent on the substrate used to produce it. Using substrate specific Y_{ATP} 's calculated with the LPs described in section 5.2.1, the amount of acetate produced can be converted into amounts of ATP with equation:

$$q_{\text{ATP}} = Y_{\text{ATP}}^{\text{Ac}} \cdot q_{\text{Ac}} \cdot 2 \quad (5.5)$$

where $Y_{\text{ATP}}^{\text{Ac}}$ is the ATP yield of the the substrate in carbon mole, and q_{ATP} is the specific ATP generation rate.

5.3.1 Maintenance costs with five growth substrates

In order to calculate maintenance costs for *A. woodii* in terms of ATP, data on Y_{ATP} was applied to reports from the literature. Four independent reports provide *A. woodii*'s maintenance cost data in five conditions: CO₂-H₂ [88], fructose-CO₂ [47], lactate-CO₂ [88], lactate-CO₂-H₂ [88], methanol-CO₂ [5], and methanol-formate [4]. Either specific substrate uptake rates, or specific acetate production rates were reported (q_{Ac}). Where possible q_{Ac} was taken for maintenance costs calculations. Godley et al. [47] did not directly provide q_{Ac} but this could be calculated using q_E , the ratio of acetate formed per fructose consumed (R_{FA}) using equation 5.6 below:

$$q_{\text{Ac}} = D \cdot \frac{q_E}{R_{FA}} \quad (5.6)$$

where D is the dilution rate. The amount of ATP generated per acetate from a given substrate was calculated with the ATP LPs (section 5.2.1) was calculated. The Y_{ATP} found when using a substrate to product stoichiometry observed in the reports, rather than the $Y_{\text{ATP}}^{\text{max}}$ in appendix B was used. This Y_{ATP} and the q_{Ac} then allowed for maintenance costs

to be calculated in units of ATP.

The maintenance in terms of amounts acetate and amounts ATP are shown in Figure 5.2 ([A] and [B] respectively), and in table 5.3. If maintenance costs were compared using acetate alone, conditions with fructose-CO₂ would have the lowest GAM and NGAM, methanol-CO₂ would have the highest NGAM and CO₂-H₂ the highest GAM. Once converted to amounts ATP however, the NGAMs in the conditions sort themselves from highest to lowest according to the Y_{ATP} of the substrates with the exception of CO₂-H₂ which has 0.08 higher ATP mmol(g-DW · h)⁻¹ generated than the lowest NGAM condition lactate-CO₂. Fructose-CO₂ has the highest Y_{ATP} and thus the highest NGAM of 0.61 ATP mmol(g-DW · h)⁻¹.

5.4 Bioenergetic Core

As has been shown here, it is possible to generate several solutions that consist of reactions catabolising a substrate input into a by-product and ATP in a genome scale metabolic model (GSM). This was done using various linear programming formulations. However, it is not possible to capture the full extent of metabolic routes through a network with this approach alone. Elementary modes analysis, identifies all non-decomposable routes through a given network but is limited to networks smaller than those found in GSMs. For analysis concerning the effect of producing novel compounds on *A. woodii* metabolism a smaller condensed network of relevant reactions was necessary for elementary modes analysis.

Also, given the wide array of substrates that *A. woodii* can use to generate ATP and grow, it is likely that a large set of reactions is involved in catabolism. It was of interest to represent these catabolic reactions as a discrete network, and to identify potentially novel reactions involved in catabolism.

Thus given the diversity of substrates available in the GSM, and the usefulness of the ATP LPs in finding several catabolic routes, a novel algorithm was applied to the GSM to extract a bioenergetic core of *A. woodii*. This was then characterised and used for further analysis.

To investigate how some substrates can be used to generate ATP but not support growth, the essentiality of reactions in the GSM as a whole and more specifically to the bioenergetic core was analysed.

5.4.1 Methodology

Extraction of the core

A sub-model was extracted from the GSM using the ATP LPs in section 5.2.1 and the range of substrates available to the GSM in an attempt to generate a condition-independent bioenergetic core. For each substrate used, the four ATP LPs were used to compute a Y_{ATP} . Reactions found at least once in these solutions were collected and ultimately used to con-

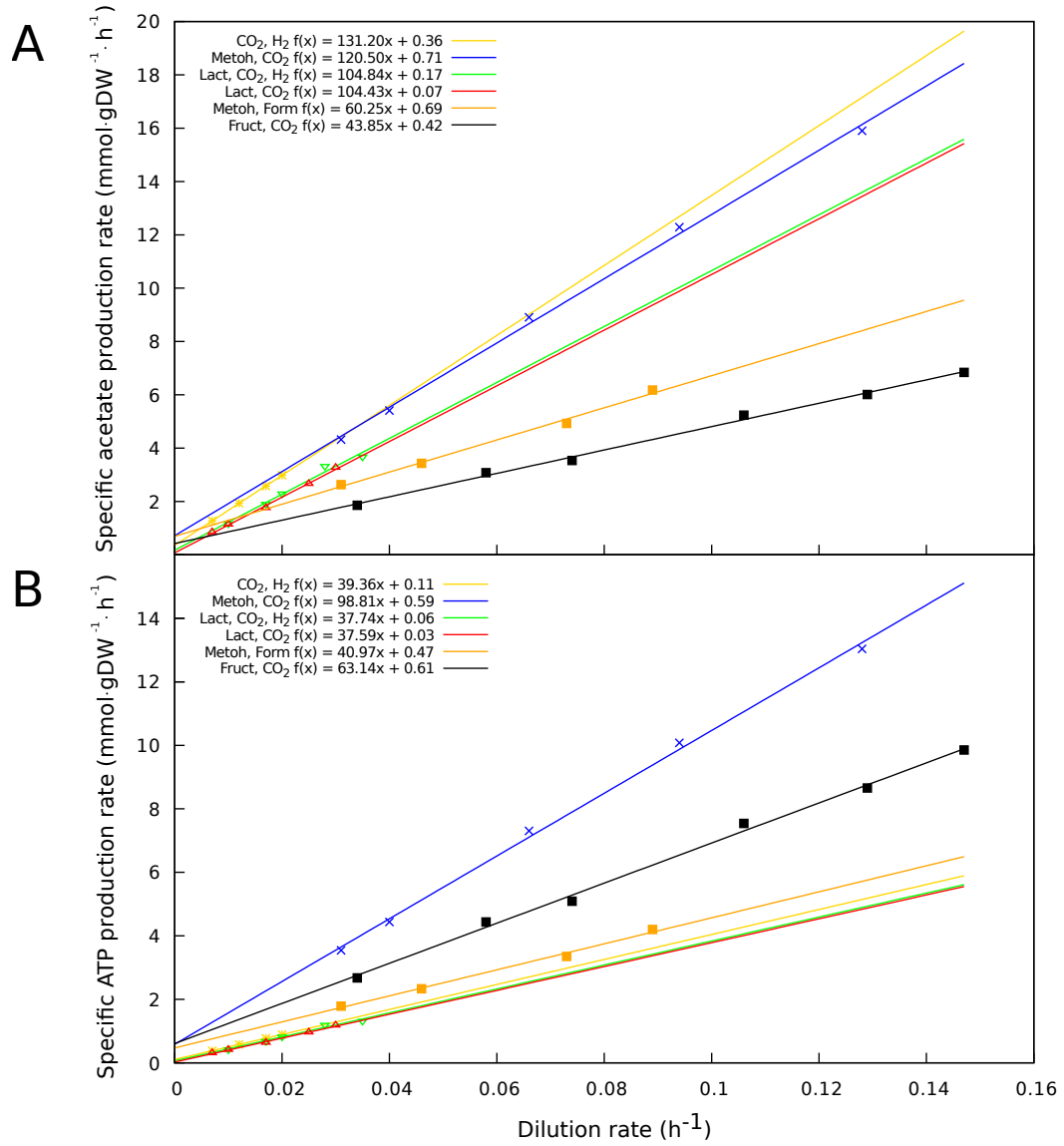


Figure 5.2: Maintenance costs in amounts acetate ATP, using five different growth substrates. Shown in [A] is maintenance costs in acetate and the data for these were collected from [88, 47, 5, 4] and replotted here. Each acetate data point was converted into an equivalent amount ATP using analysis done here and plotted in [B].

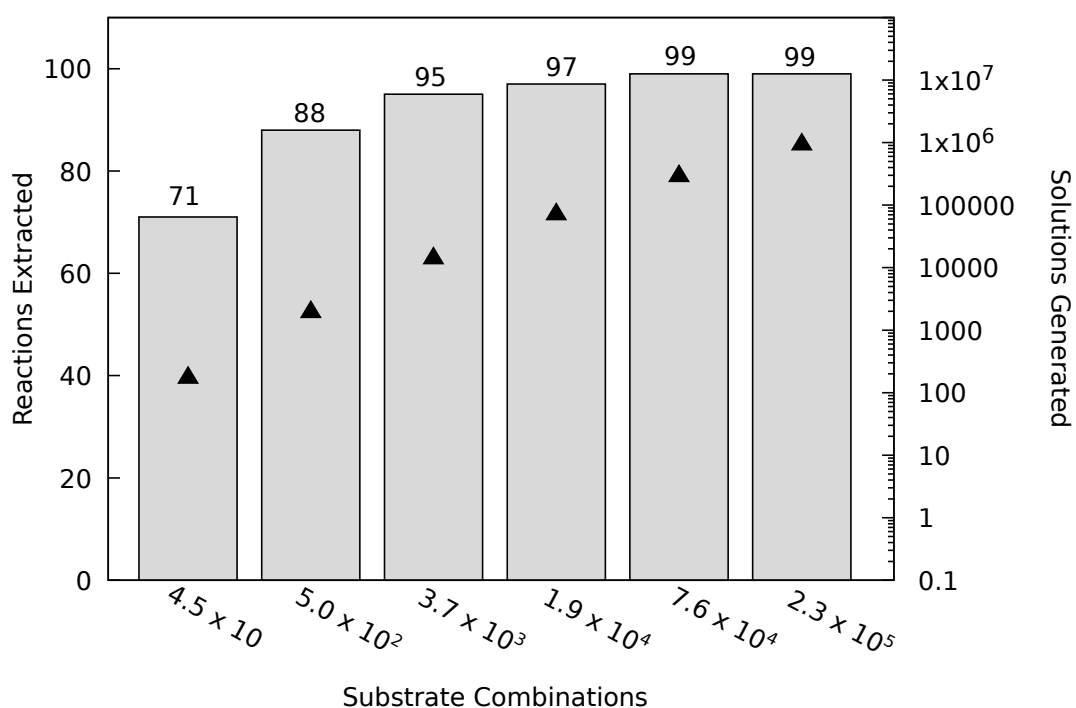


Figure 5.3: **Using an increasing number of substrate combinations to extract a condition independent bioenergetic core from the Genome-scale Model.** Grey bars with numbers indicate number of reactions extracted for the number of substrate combinations shown, and filled triangle indicates the number of solutions generated.

struct the core model. Only substrates which supported growth were selected. Since it was found that some substrates could be used to generate ATP but not support growth without CO₂ (see table in appendix 8.3), substrates were used as they were (i.e. lactate-acetoin), and also with CO₂ supplemented (i.e. lactate-acetoin-CO₂). Initially, single substrates and paired substrate combinations were used, which when adding the CO₂ condition, totalled to 507 substrate conditions (not excluding non-growth supporting substrates).

The data generated during extraction of the core is shown in Figure 5.3. It was found that the 2.028×10^3 solutions generated with the 507 substrates, did not comprehensively identify new reactions. Using triplets (3.665×10^3 conditions) and quadruplicates (1.9019×10^4 conditions) of substrate combinations, identified additional reactions. No new reactions were identified after quintuplets of substrate combinations (7.5942×10^4 conditions), were used. The number of reactions found with this number of combinations was 99, the same number as that found with sextuplet combinations (2.42934×10^5 conditions). The number of solutions generated by sextuplet combinations was 9.71736×10^5 , 1.46885×10^5 of which were unique.

In terms of time taken to generate these datasets, single substrates and up to triplet combinations could be generated within a day with a debian operating system and a 2200 MHz processor. Solutions for sextuplet combinations took 551 hours (23 days).

Reaction Essentiality

This was done by iteratively setting the flux to each reaction. not including transporters and dead reaction, in the model to 0 (i.e. knocking that reaction out), and then solving for biomass production using equation 2.4 for the 507 substrate combinations. In this way, the effect of every reaction knockout is tested in 507 conditions.

5.4.2 Bioenergetic core metrics and properties

The 99 reactions found during model extraction were used to construct the bioenergetic core as shown in Figure 5.4. This consists of 68 internal reactions, 86 metabolites, and 31 transporters. 18 transporters were substrates, 6 transporters are for exchange (both substrates and by-products), and 7 are by-products. These reactions in sum represent a core of reactions available to *A. woodii* to generate ATP from any feasible substrate combination.

In total the core contains 27 reaction subsets, two of which have five reactions in them. One forms part of the glycolytic reactions and the other includes valine-pyruvate aminotransferase (R45 in Figure 5.4), and a branched-chain amino acid aminotransferase (R56 in same figure). No reaction was found in all solutions. The ATP hydrolysis reaction was found in the most number of solutions (4.9201×10^4 solutions), followed by acetate kinase and phosphotransacetylase in solutions (4.9119×10^4), followed by the Rnf complex and ATP-synthase (4.8157×10^4), and then the electron bifurcating hydrogenase reaction (4.7950

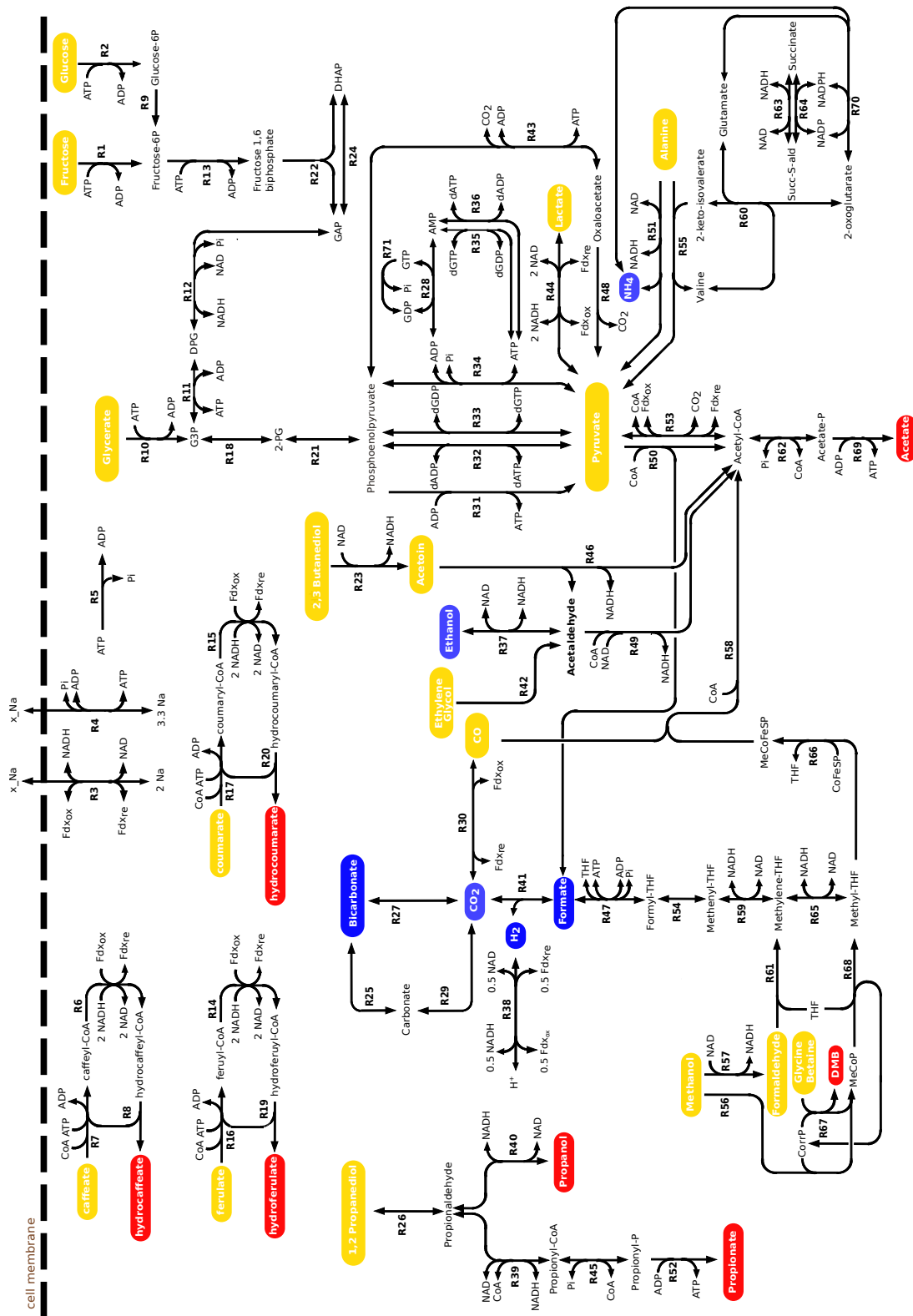


Figure 5.4: The bioenergetic core of *A. woodii*. Highlighted names in yellow are growth substrates (inputs), in red are by-products (outputs), and blue are exchange (inputs or outputs). Dashed line indicates cell membrane barrier. Reaction direction is as found in the model.

$\times 10^4$). Phosphoenol transphosphorylase (R24 in Figure 5.4) and guanosine-diphosphate kinase (R26) were found in only one solution and pyruvate-formate lyase (R41) was found in only three solutions.

Potentially novel reactions utilising deoxyribonucleosides in reactions involving pyruvate and PEP were identified. Also, an additional route to pyruvate from PEP was identified in phosphoenolpyruvate carboxykinase (R43), which generates an ATP, produces CO_2 as well as oxaloacetate, a metabolite involved in TCA related reactions as seen in chapter 4. An alternative alanine catabolic route was also identified.

15 pathways found in metaCyc are fully represented in the core model. Five of these pathways are single reactions, two have two reactions, four pathways have three reactions, one pathway has five reactions (1,2-propanediol degradation), one pathway has eight reactions and is the WLP, and one pathway has ten reactions (Embden-Meyerhof-Parnas pathway).

Pathways not fully represented in the core model were also identified. Twenty pathways with one reaction missing in were found. These include, L-threonine degradation IV (PWY-5436), which uses an aldolase to split threonine into acetaldehyde and glycine, CO_2 fixation into oxaloacetate (PWYQT-4429), which uses a phosphoenolpyruvate carboxylase to form oxaloacetate and inorganic phosphate, and acetylene degradation which uses a hydratase to convert acetylene into acetaldehyde.

5.4.3 Reaction essentiality

Of the 507 substrate combinations used, 437 were found to support growth and so used for the rest of the algorithm. The algorithm took in total 405 hours (17 days) to complete, was run in six parts and later compiled into a single dataset.

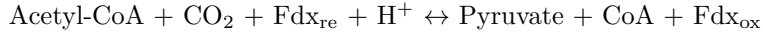
Substrate-independent essential reactions

Of the 841 reactions in the GSM, 158 proved to be essential with all substrate combinations. Half of these reactions were involved in non-redundant biosynthetic pathways of essential biomass precursors and central metabolism. This includes reactions in the pentose-phosphate pathway, amino acid biosynthesis, chorismate biosynthesis, cardiolipin biosynthesis, coenzyme A, AIR and IMP biosynthesis, the SAM cycle, nucleotide synthesis and flavin synthesis. Only one of the TCA related reactions presented in chapter 4, table 4.2, figure 4.4 was found to be essential in all conditions - aspartate aminotransferase (R4 in same table and figure). None of these essential reactions are found in the bioenergetic core.

Substrate-dependent essential reactions in the bioenergetic core

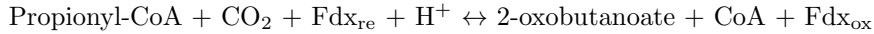
Another 44 reactions were found to be essential depending on what substrate was used for growth. Of these, 38 are in the bioenergetic core. That makes 56% of the reactions in the bioenergetic core model substrate-dependent essential reactions.

This substrate-dependent essentiality data was used to investigate why substrates in table 5.2 and appendix B, that were found to be energy-generating but not-growth-supporting (EGNGS), become growth-supporting in the presence of CO₂. In total out of the 45 EGNGS substrates, 35 had one CO₂ consuming reaction (when CO₂ was supplied as co-substrate) in common - pyruvate-ferredoxin oxidoreductase (PFOR, R53 in Fig. 5.4), which has a stoichiometry of:



When these 35 EGNGS substrates are used, PFOR is the only reaction that allows for pyruvate production, making CO₂ an essential metabolite for building biomass precursors. It is possible for some CO₂ to be produced internally. For example ethanol as a sole substrate (and member of EGNGS) can produce threonine, leucine, isoleucine and glycine by producing CO₂ internally to power PFOR in the pyruvate forming direction, but to accomplish this, reactions important for producing other biomass precursors are set in a conflicting non-growth supporting direction. When it is possible to produce pyruvate by reactions other than PFOR such as by the electron-bifurcating lactate dehydrogenase (EC-1.2.7.7 and R44 in Fig. 5.4) when lactate is supplied as co-substrate, by alanine dehydrogenase (R51 in Fig. 5.4) with alanine as co-substrate, pyruvate kinase (R31 in Fig. 5.4) with hexoses, or when pyruvate is supplied directly as co-substrate, growth is supported (see appendix B).

The 10 EGNGS substrates where PFOR was not essential but CO₂ still was, all used 1,2 propanediol as a co-substrate. With CO₂ supplementation it was found that during growth, as well as PFOR, a second CO₂ consuming enzyme, 2-oxobutanoate synthase (see stoichiometry below), could compensate for a PFOR knockout by producing pyruvate via an alternate pathway:



The propionyl-CoA derives from 1,2-propanediol via propionaldehyde (R26 and R39 in Fig. 5.4). 2-Oxobutanoate then is converted to serine via five reactions (not detailed here) which is then deaminated to pyruvate by serine deaminase (EC-4.3.1.17).

Finally, reaction essentiality of the Woods-Ljungdahl pathway was investigated. Below are enumerated reactions of WLP with their reaction number in Fig.5.4 and the number of substrates combinations that reaction was essential for:

1. formate-THF ligase (R47): 37
2. hydrogen-dependent formate dehydrogenase (R41): 59
3. carbon monoxide dehydrogenase (R30): 83
4. acetyl-CoA synthase (R58): 86
5. methyl-THF:corrinoid/iron-sulfur methyltransferase (R66): 86

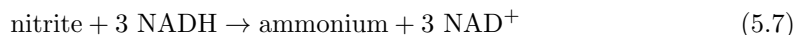
6. methenyl-THF cyclohydrolase (R54): 89
7. methylene-THF dehydrogenase (R59): 89
8. methylene tetrahydrofolate reductase (R65): 368

The range of essentiality in WLP reactions, is a characteristic of this pathway not existing as a reaction subset in the GSM (see section 4). Methylene tetrahydrofolate reductase was essential in the most number of conditions, and was only non-essential when substrate combinations included either methanol or glycine betaine. Tracking methyl-THF consumption and production when fructose is used, revealed that methyl-THF is essential for methionine synthesis by methionine synthase (EC-2.1.1.13), which itself is used a protein building block and for methyl charging of s-adenosyl moieties to s-adenosylmethionine (SAM). Methanol and glycine betaine are able to provide methyl-THF via an alterante route using the cobalt-dependent methyl-transferase (R4 in Fig 5.4).

5.5 Using alternative electron acceptors changes ATP yields

To investigate how redox balances affect the Y_{ATP} of substrates, a hypothetical NADH oxidase or reduced ferredoxin oxidase was added to the model and the $Y_{\text{ATP}}^{\text{max}}$ of 276 substrate combinations recalculated (for substrates in appendix B). Adding these oxidases did not affect model conservation. When the reduced ferredoxin oxidase was added to the model, the $Y_{\text{ATP}}^{\text{max}}$ of substrates did not show any increase. Adding an NADH oxidase did increase the $Y_{\text{ATP}}^{\text{max}}$ of substrates, and increased it for every substrate that was not in combination with either caffeate, ferulate and coumarate. This is shown in figure 5.5. Three of the 276 substrates were excluded from the $Y_{\text{ATP}}^{\text{max}}$ recalculation. $\text{CO}_2\text{-H}_2$, formate- H_2 and $\text{HCO}_3^- \text{-H}_2$. This is because once an NADH oxidase was added to the model, these substrate combinations could generate ATP using only H_2 , and so an ATP yield in $\text{mol}\cdot\text{Cmol}^{-1}$ becomes misleading.

A. woodii is known to use a several electron acceptors for growth, including derivatives of ligning such as caffeate 8.3. Other acetogens are also known to use electron acceptors such as fumarate 8.3, various nitrogen and sulphur sources 8.3. A putative nitrite reductase gene was identified in *awoodiiCyc* (see methods) at gene locus *Awo_c10170*, as well as a putative formate-nitrite transporter (FNT). A putative equation for nitrite reductase was added to the model based on this according to the stoichiometry in 5.7, as well as a nitrite transporter.



This would make nitrite a potent source of NADH oxidation if used as an alternative nitrogen source to ammonium, as it oxidises three NADH per mole where as CO_2 would

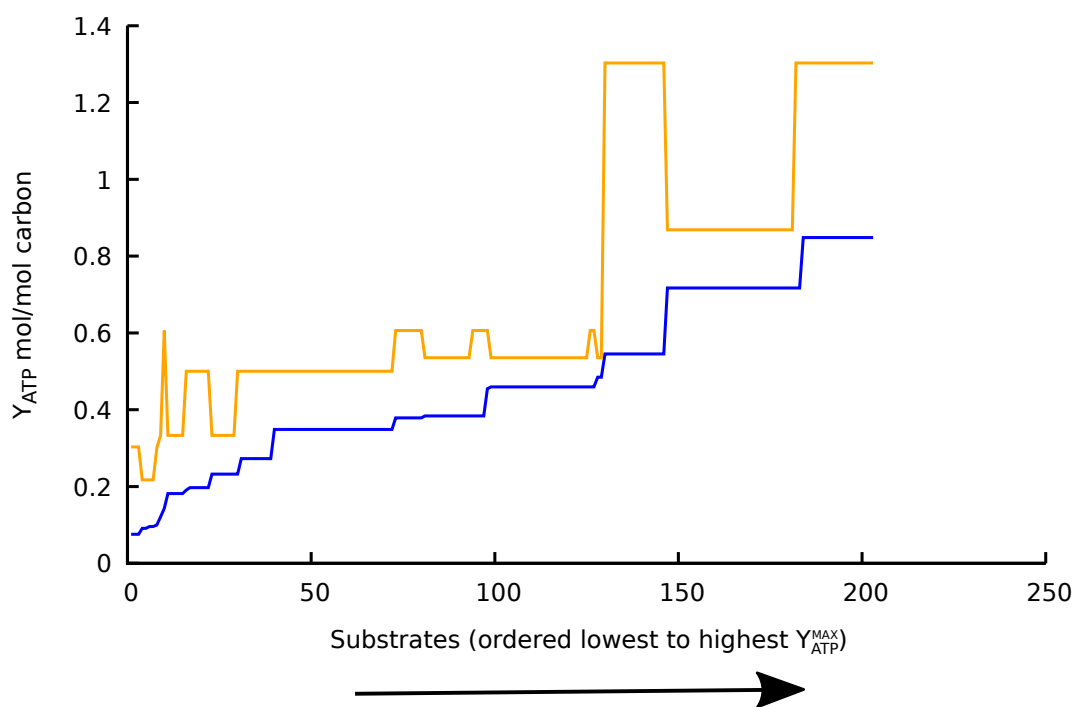


Figure 5.5: **The effect of introducing a hypothetical NADH sink on $Y_{\text{ATP}}^{\text{max}}$ calculations.** Blue indicates $Y_{\text{ATP}}^{\text{max}}$ of substrate pairs in native *A. woodii* metabolic network (using the genome-scale model, as found in appendix B). Orange line indicates the $Y_{\text{ATP}}^{\text{max}}$ of the same substrates after an NADH oxidase reaction was added to the metabolic network. Substrates are ordered from lowest to highest ATP yields. Substrates that cannot be used to generate ATP were omitted as well as substrates paired with either caffeate, ferulate or coumarate

Table 5.4: The effect of various electron acceptors on the Y_{ATP} of various C1 substrates and ethanol. Dashed line indicates, no additional substrate was supplied. Asterisks indicates no additional substrate was supplied but the Y_{ATP} was calculated after adding an NADH oxidase to the model.

Substrate	-	CO ₂	NADH oxidase*	Nitrite	Caffeate
CO	0.38	0.38	0.61	0.61	1.21
Methanol	0.55	0.55	1.30	1.30	3.12
Formate	0.08	0.08	0.30	0.30	0.91
Ethanol	0.20	0.20	0.50	0.50	1.11

allow for the oxidation of two. The stoichiometry for caffeate reduction is seen in equation 5.8.



The effect of CO₂, NADH oxidase, nitrite and caffeate have as electron acceptors is summarised in table 5.4. The smallest ATP yield across the four conditions tested was both with the substrate with CO₂, and the substrate on its own. The addition of CO₂ however changed the product spectrum however, allowing for acetate as the sole by-product.

Overall, the effect of nitrite oxidation is equivalent to the unlimited potential of NADH oxidation with an NADH oxidase. Since Caffeate allows for both oxidation of excess NADH and reduction of additional ferredoxin, the highest Y_{ATP} is found when it is used as the electron acceptor.

5.6 Discussion

For this study on the bioenergetics of *A. woodii*, four novel datasets in total were collected from analysis of the GSM and a bioenergetic core model was constructed to compare model predictions with laboratory ones, and to gain novel insights into growth and ATP generation. These datasets are: a comprehensive set of Y_{ATP} data in 276 conditions, maintenance costs in ATP units, condition-dependent and independant data on reaction essentiality, and Y_{ATP} of substrates in the face of a metabolic perturbation - a NADH or ferredoxin oxidase knock in.

ATP yields are often calculated with a single LP in structural models, but here, taking advantage of both the diversity of *A. woodii* metabolism and the GSM which reflects this, it was found that no single LP was able to consistently calculate a $Y_{\text{ATP}}^{\text{max}}$ across all conditions. Furthermore, the metabolic pathway used when calculating when a $Y_{\text{ATP}}^{\text{max}}$ was not necessarily found in the biomass production solution, indicating objectives other than ATP maximisation are at play during growth. Minimising total flux, a proxy for minimising total

protein investment, calculated a $Y_{\text{ATP}}^{\text{max}}$ in only 70 of the 276 substrate combinations but its use in biomass production still reflects growth stoichiometries observed in the laboratory when growing on substrates such as ethanol and CO_2 .

In terms of growth on carbon sources, the model correctly predicted growth with 94% accuracy in 51 reported conditions. The fact that many theoretical $Y_{\text{ATP}}^{\text{max}}$ are higher than those observed during growth, again highlights a yet unidentified regulatory framework that constricts the main products of *A. woodii* growth to acetate which is in agreement with findings in chapter 3.

The use of these ATP LPs allowed for the extraction of a bioenergetics core from the GSM which is amenable to elementary modes analysis. Potentially novel reactions involved in catabolism were also identified including a link to the TCA related reaction - phosphoenolpyruvate carboxylase.

Being able to calculate the $Y_{\text{ATP}}^{\text{max}}$ consistently across many conditions also allowed for its recalculation after metabolic perturbations. The finding that constraints in redox balancing, specifically the oxidation of NADH, imposes a limitation on the amount of ATP being produced. This was reflected by yields of ATP changing based on how much NADH oxidation was allowed for by the electron acceptor. The highest ATP yields were found when ferredoxin is reduced and two NADH are oxidised, but NADH oxidation alone sufficed in increasing a substrates Y_{ATP} . This may be useful for the design of a metabolic pathway for production of industrial compounds of interest, especially when kinases are not an available option.

Reports of substrates such as ethanol, 2,3 butanediol or 1,2 propandiol need CO_2 as a co-substrate to support growth even though they themselves are a source of carbon and energy [116, 57, 10]. These are observations made both in *in vivo* and with the GSM here *in silico*. A systems level analysis of *A. woodii* network was needed to explain this observation. In isolation, reactions exist to produce important metabolites like pyruvate, but in the context of genome-wide metabolism operating in synchrony, the sum of these reactions cannot feasibly allow for the synthesis of essential biomass precursors needed for growth. Furthermore, the specific enzymes responsible for this phenotype were identified here to be PFOR and 2-oxobutanoate synthase. Furthermore, the essentiality of reactions in WLP revealed that methylene tetrahydrofolate reductase is needed for supplying methyl-THF for methionine synthesis in all conditions where methanol and glycine betaine are not available, highlighting a role of WLP in process other than carbon fixation and energy conservation.

The maintenance costs calculated were all found to be greater than zero and thus the ATP yields provided by the various substrates investigated are enough to support growth. The maintenance costs also provide additional parameterisation of the GSM for more accurate predictions. NGAM takes into account energy spent by the cell for processes such as macromolecular compound turnover, motility and maintenance of a chemiosmotic potential, GAM refers to energy spent on synthesising machinery and cellular constituents to form

new cells. These are however, approximations and whether these are truly mathematical constants has been called into question more recently [12]. Here a range of maintenance costs were found that depended on the growth conditions used. The size of the maintenance costs correlated with the energetic richness of the substrate (except for $\text{CO}_2\text{-H}_2$). The higher the Y_{ATP} available from that substrate, the higher the NGAM was found to be. This was also true for GAM with the exception of methanol- CO_2 . This supports the argument that it may not be useful to think of maintenance costs as constants but instead as functions of bacterial cells redistributing its energetic currency to adapt to different growth conditions. Additionally the maintenance costs in ATP units, when compared to maintenance costs of other organisms is very low [126], which would be advantageous from an industrial application standpoint.

Chapter 6

Experimental Growth of *A. woodii*

6.1 Chapter Introduction

In vivo studies on *A. woodii* were done at Denmark Technical University in collaboration with Dr Sheila Ingemann Jensen to complement modelling work. To justify biologically meaningful interpretations of model analysis, a validation of the models predictive power is necessary. Also, parameterisation of the model using measured uptake, growth, production and energetic maintenance rates allows for better alignment between *in silico* and *in vivo* metabolism. The plans of the investigations were informed by modelling results on biomass production and bioenergetic analysis on alternative nitrogen sources discussed in chapter 5.

6.2 Aims

- Validation of the models predictive capability against *in vivo* measurement, and refine model parameters for better predictions.
- Laboratory investigations of using alternative nitrogen sources with *A. woodii*.

6.3 Methods

6.3.1 Laboratory data analysis

Culture

A. woodii strain 1030 was collected from Deutsch Sammlung von Mikroorganismen und Zellkulturen (DSMZ). To ensure axenic cultures, the culture collection was streaked and

incubated, after which a single colony was picked and used throughout subsequent experiments. Verification of *A. woodii* was done by myself using microscopy and with colony PCR by Dr Jensen.

Growth media

Growth media were prepared using anaerobic techniques. The standard undefined growth medium contained in ($\text{g}\cdot\text{l}^{-1}$): KH_2PO_4 , 0.2; NH_4Cl , 0.25; NaCl , 1.16; KCl , 0.5; yeast-extract, 1; $\text{MgSO}_4\cdot 7\text{H}_2\text{O}$, 1.45; CaCl_2 , 0.11, cysteine-HCl, 0.5; NaHCO_2 , 5. The media also contained 0.5ml resazurin, 1ml trace element solution SL9 and 0.2ml vitamin B solution. If fructose (20mM) was used $3.6\text{g}\cdot\text{l}^{-1}$ was added and if ethanol (50mM) was used $2.3\text{g}\cdot\text{l}^{-1}$ was added. Defined media was the same above but with yeast-extract removed. Media with NaHCO were degassed with $\text{CO}_2\text{:N}_2$ 80:20 (vol:vol). NaHCO was removed in CO_2 -free media and instead buffered with $4.77\text{g}\cdot\text{l}^{-1}$ HEPES, and degassed with 100% N_2 . Media with nitrite contained either 4, 6, 8 or 10mM of Na-nitrite. Whether NH_4Cl was used in nitrite supplemented media is indicated in text. The pH was adjusted to 7.0 with NaOH or HCl unless stated otherwise.

Microscopy

Images shown of *A. woodii* were taken using phase contrast microscopy with a Leica DM4000 B. Cells were taken during the growth phase in fructose based media. A representative section of the image is shown in figure 6.4.

Growth experiments

Cultures were grown in quadruplicates of 40ml media in 100ml volume rubber-stopped and crimped serum flasks. Cultures were inoculated to an initial OD_{600} of 0.02 from a fructose pre-culture. Samples were taken with an anaerobic syringe for centrifugation every 3/4 hours or more frequently, and the supernatant stored at -80°C . After each sampling, the head space of the serum flask was re-pressurised to 1 bar with $\text{N}_2\text{-CO}_2$ 80:20 (vol/vol). OD_{600} readings over 0.3 were diluted.

Substrate and product analysis

Samples were analysed for fructose and acetate concentrations analytically using high-performance liquid chromatography (HPLC) by Dr Jensen. 1mL of cell suspension was collected and centrifuged at 6500g for 5mins. 0.8mL of the supernatant was transferred into an Eppendorf tube and stored at -20°C until further analysis. The thawed supernatant was analysed using an Ultimate 3000 HPLC equipped with a Biorad Aminex HPX-87H column. Samples (run for 30 min.) were analysed using 5mM sulfuric acid as mobile phase

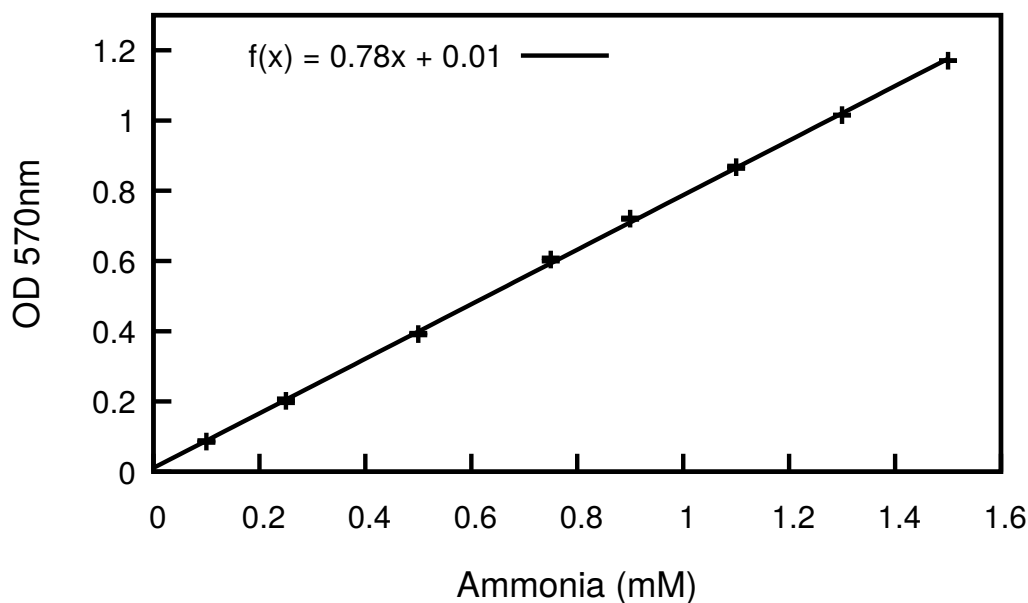


Figure 6.1: The calibration curve established for ammonia measurements

at a flow rate of 0.6mL min⁻¹, and a column temperature of 30°C. Fructose and acetic acid concentrations were estimated using a refractive index detector (RI).

Ammonium concentrations were determined spectrophotometrically according to a modified method of [13]. A calibration curve was established using stock concentrations (figure 6.1). Samples were stored on ice before and after incubation, and incubation was at 37°C for 15 minutes. The OD at 570nm was read. Two technical replicates were used across 8 ammonium concentrations. Ammonium concentration increased linearly with OD₅₇₀ until an OD₅₇₀ of 1.17.

Cell dry-weight assay

Cells were grown to late log phase on fructose based media until an OD₆₀₀ of 1.3. Cells were then harvested into pre-dried 50ml falcon tubes in a series of increasing concentrations and centrifuged at 4000rpm. Cells were resuspended and washed with fresh media and centrifuged again. Pelleted cells were then incubated at 80°C until the weight no longer changed. See figure 6.3 for the calibration curve.

Statistics

The growth data presented for the fructose cultures are means of four biological replicates. For other cultures, two biological replicates were used. Error bars represent the standard error.

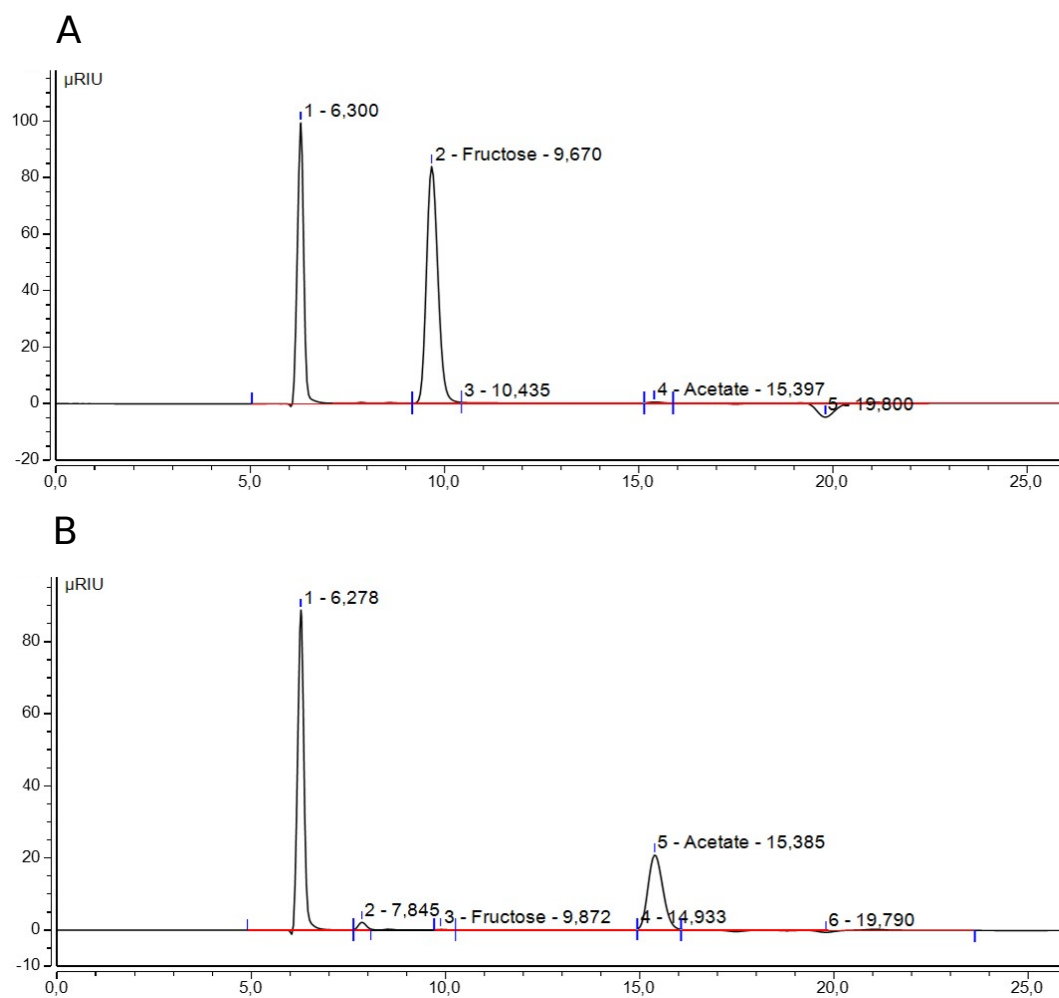


Figure 6.2: HPLC chromatograms. [A][B]

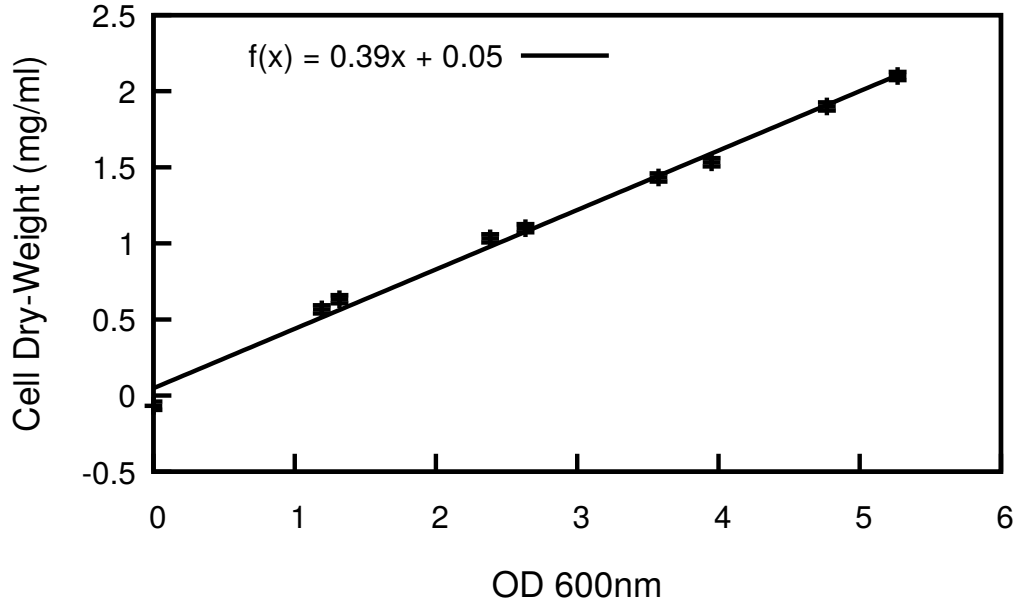


Figure 6.3: The calibration curve established for dry weight measurements

Yield calculations

Growth yields on substrates or products was done using biomass formation rates and consumption production rates observed during exponential growth between eleven and thirty one hours. No reports could be found on the ash percentage of biomass so an average of 6% (wt/wt) was used as seen in [126]. To calculate growth yields from specific substrate uptake rates and growth rates (as was done with growth yields calculated from model predicted rates), the specific rate equation in [90] was rearranged to:

$$Y_{XE} = \frac{\mu}{|q_E|} \quad (6.1)$$

where μ is growth rate and $|q_E|$ is the absolute specific uptake or production rate.

6.3.2 Modelling data generation

Comparisons were made between observed specific rates of substrate uptake or product formation and equivalent data from the model. Unless specified otherwise the biomass LP from (methods and materials) was used to set up the LP using with a specific growth rate.

Growth Rate Scan

To compare specific rates across observed growth, a growth rate scanning algorithm was designed and implemented. The LP used for this was a modification of equation 2.4 and is:

$$\begin{array}{ll} \text{minimize} & : |v_i| \\ \text{subject to} & \left\{ \begin{array}{l} \mathbf{N}\mathbf{v} = \mathbf{0} \\ v_{ATPase} = r_{ATP} \\ v_{J_B} = t_{J_{P...R}}; J_{P...R} \cdot 0.001 \leq \mu \leq 0.19 \\ v_J = v_{J_S}, v_{J_E}, v_{J_B} \end{array} \right. \end{array} \quad (6.2)$$

where μ (the growth rate) is fixed to a set value. All other symbols have the same meaning as in Equation 2.4 (chapter 2, section 2.2). The analysis was performed by repeatedly solving the LP, where all the constraints remained fixed except for μ which is successively fixed to a value between the upper and lower bound.

6.3.3 Carbon mass balance

Elemental composition and the Cmol stoichiometry of biomass was calculated using the biomass composition compiled for the *A. woodii* model. This was calculated by using the atomic composition of each component and multiplying it by its flux requirement. This was then used to calculate biomass concentrations on a Cmol basis and used with the chemical formula of fructose and acetate to calculate the carbon balance.

6.4 Laboratory Results

6.4.1 Fructose batch growth

The observed growth curve on fructose, carbon recoveries and specific substrate uptake and by-product secretion rates are shown in figure 6.5. *A. woodii* grew to a final OD₆₀₀ of 1.87 ± 0.051 , which corresponds to a biomass DW of $0.78\text{mg/ml} \pm 0.02$ from an initial OD₆₀₀ of 0.03 ± 0.004 and $0.06 \text{ mg/ml} \pm 0.001$. During exponential growth, a doubling time of 6.24 h and growth rate of 0.11h^{-1} was observed. The consumption of fructose and ammonium occurred with a rate of $0.823 \text{ mmol}\cdot\text{hr}^{-1}$ and $0.204 \text{ mmol}\cdot\text{hr}^{-1}$ respectively. Acetate production rates were $1.714 \text{ mmol}\cdot\text{hr}^{-1}$. Ethanol was measured but not observed to be produced in any significant amounts throughout the fermentation. Thus, homoacetogenic growth was observed as acetate was the only by-product detected. The ratio of acetate to fructose was 2.1, and increased to 2.7 if corrected for biomass carbon. The ratio observed by [47] at the same growth rate was also 2.1.



Figure 6.4: Characteristic *Acetobacterium woodii* rod shaped cells growing in pairs. Cells were grown on fructose- CO_2 , and taken during log phase for imaging.

Growth yields

Growth yields for fructose, ammonium and acetate were $38.5\text{g-DW}\cdot\text{mol}^{-1}$, $155.8\text{g-DW}\cdot\text{mol}^{-1}$ and $18.5\text{g-DW}\cdot\text{mol}^{-1}$ respectively. These are comparable to yields reported in continuous cultures under similar conditions which were reported to be $42.1\text{g-DW}\cdot\text{mol}^{-1}$ for fructose and $20.2\text{g-DW}\cdot\text{mol}^{-1}$ for acetate [47]. These growth yields were used to calculate specific uptake and production rates seen in figures 6.5[B] and 6.7.

Carbon recovery

Carbon recovery can be seen in 6.5[C]. Recovery rates remained above 90% for the duration of the fermentation and above 95% for 27 hours. After 39 hours 90.5% carbon could be recovered suggesting some of the carbon's destination was not acetate or biomass. Carbon recovery was higher during exponential growth at 93.5%. During exponential growth 22.8% of carbon in fructose could be recovered in biomass and 69.4% in acetate. Carbon recovery in [47] at the same growth rate was lower at 89.4%.

6.4.2 Using nitrite as nitrogen source

Whether nitrite could be utilised by *A. woodii* as a nitrogen source was investigated by using it in conjunction with ethanol (50mM)- CO_2 , and fructose(20mM)- CO_2 . Cultures were grown in duplicates and inoculated from a fructose pre-culture at late log phase.

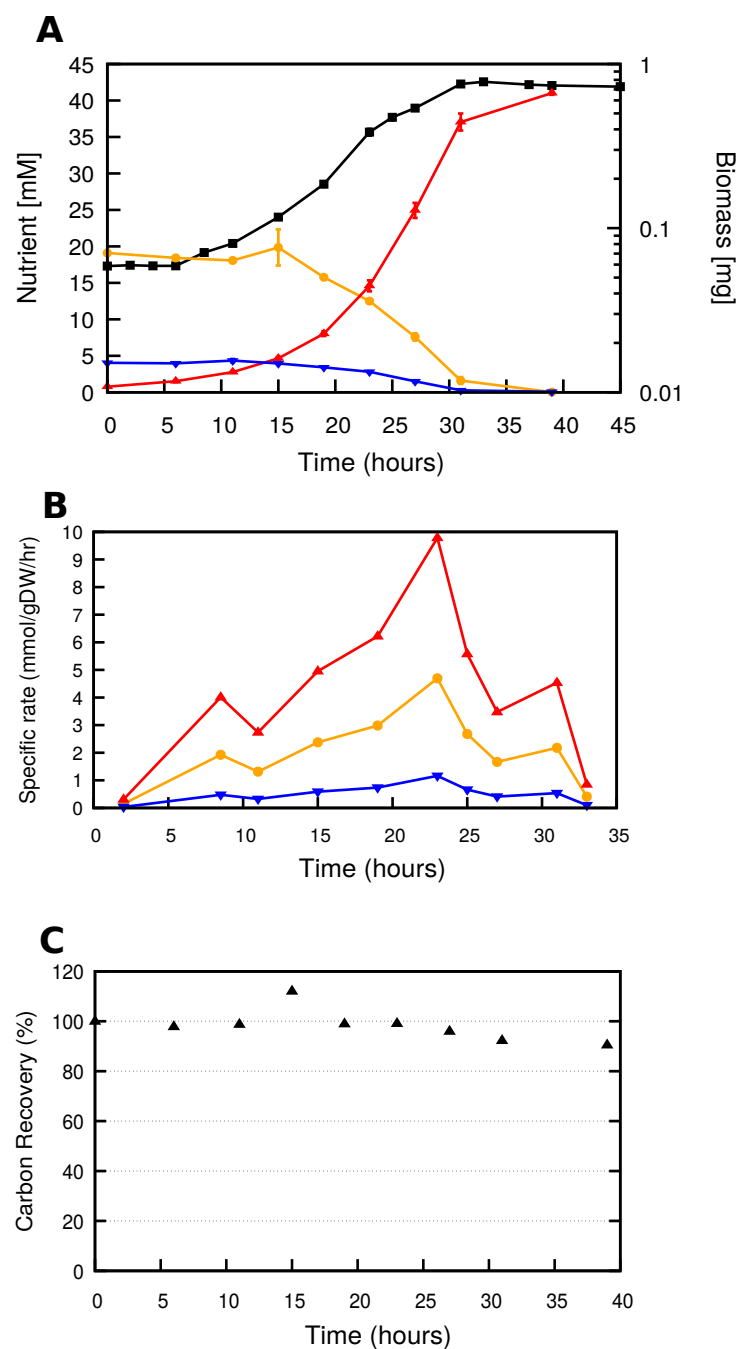


Figure 6.5: Fructose growth experiments. Black squares and lines are biomass measurements, yellow circles and lines are fructose, red upwards triangles and lines are acetate and blue downward triangles and lines are ammonium [A] shows the culture growth curve on semi-log axes. Error bars indicated standard error of four biological replicates. [B] shows specific substrate uptake rates and specific acetate production rates. [C] shows carbon recovery rates for the duration of the fermentation.

No growth could be determined when 1mM, 3mM or 9mM nitrite was added with fructose-CO₂ in the media, indicating that nitrite in these conditions inhibits growth.

Using ethanol as carbon and energy source

A. woodii reached a cell density of OD₆₀₀ 0.34±0.013 after 75.4 hours in standard media with ethanol-CO₂, and 0.11±0.003 after 57.7 hours with ethanol as sole carbon source, from an initial 0.04 OD₆₀₀.

A series of conditions with nitrite were set up to investigate whether growth would still occur. All media contained yeast extract. Concentrations of nitrite at 4mM, 6mM, 8mM and 10mM were used. Media with ammonium removed, media with CO₂ removed, media with both of these removed, and media with a range of pH's from 6.44 to 7.32 were also used. From an initial OD₆₀₀ of 0.04, cultures with nitrite at 6mM and 8mM, cell densities of OD₆₀₀ 0.13±0.08 and 0.13±0.000 after 56 and 75 hours respectively were recorded. With ammonium removed from the media and 5mM nitrite, an OD₆₀₀ of 0.13 ±0.014 was recorded after 77 hours.

Cell density in media with a slightly elevated pH of 7.3, with 5mM nitrite and CO₂ removed, reached an OD₆₀₀ of 0.16 ±0.006 after 184 hours from an initial OD₆₀₀ of 0.02. Small yellow aggregates and a cell suspension could be seen in these cultures. Cells were also observed under the microscope in a sample. It cannot be ruled out however that the small amount of growth observed was not a result of carry over nutrients from starter cultures.

6.5 Comparing *in vivo* observations to modelling predictions

6.5.1 Comparing the batch fructose fermentation to model predictions

The specific rates of fructose, ammonium and acetate during exponential growth were compared with the rates predicted by the model, see table 6.1. The specific uptake rates during exponential growth were 2.882 mmol(g-DW· h)⁻¹ for fructose, and 0.713mmol(g-DW· h)⁻¹ for ammonium. The specific acetate production rate was 6.004 mmol(g-DW· h)⁻¹. The model specific uptake rates were 2.616mmol(g-DW· h)⁻¹ and 0.897mmol(g-DW· h)⁻¹ for fructose and ammonium, and 5.65mmol(g-DW· h)⁻¹ acetate.

Using the model to calculate biomass yields resulted in yields of 42.4g-DW·mol⁻¹, 123.7g-DW·mol⁻¹ and 19.7g-DW·mol⁻¹ for fructose, ammonium and acetate respectively. These are comparable to the yields from the batch culture but even better aligned with the yields observed in the continuous culture [47]. Biomass yields on ammonium had the largest discrepancy, with the model underestimating the yield by 32.1g-DW·mol⁻¹ when compared to the yield seen in 6.4.1.

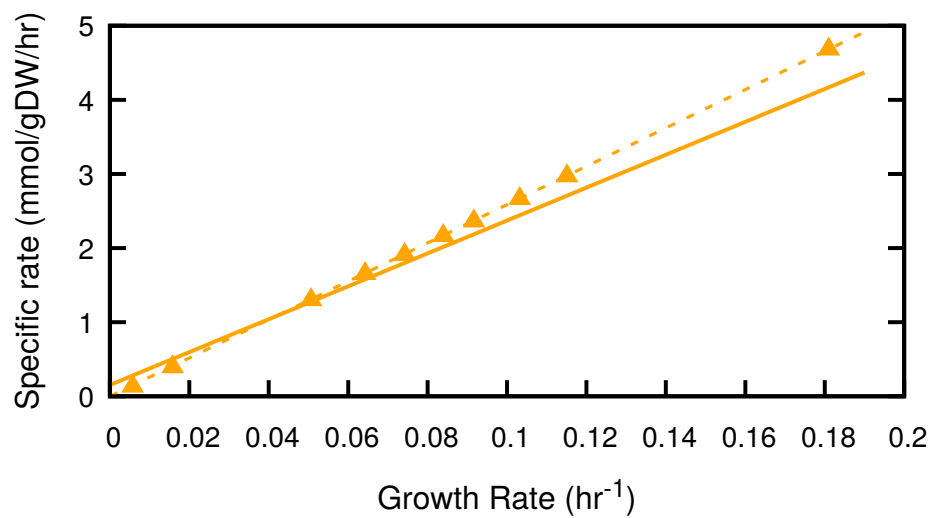


Figure 6.6: Comparing fructose specific uptake rates calculated from batch culture against model predictions. Triangles are rates calculated from experiment, solid line is rates predicted by model using maintenance costs from continuous culture and dashed line is rates predicted by model using maintenance costs calculated from exponential phase of batch culture.

In the batch culture a total of 92.2% of carbon from fructose was found in either biomass or acetate, meaning 7.8% of carbon had an unknown fate. Using the model to track carbon fates resulted in 24.62% being found in biomass and 75.38% in acetate, similar to the amounts in the batch culture as well as a further 3.4% carbon being found as formate, closing the *in silico* carbon balance to 100%.

6.5.2 Comparing model predictions with different maintenance costs

The effect of maintenance cost parameterisation on model predictions were tested using costs from the continuous culture (as found in chapter 5), costs estimated from the batch culture here, and a zero cost, see figures 6.6 and 6.7. Costs from the batch culture were calculated using biomass acetate yields during exponential growth and the observed growth rates throughout the fermentation. This resulted in a growth-associated maintenance (GAM) of $54.05 \text{ mmol(g-DW} \cdot \text{h)}^{-1}$ acetate and a non-growth associated maintenance (NGAM) of $0.00 \text{ mmol(g-DW} \cdot \text{h)}^{-1}$. These were converted into amounts ATP (as detailed in 5), resulting in a GAM of $77.83 \text{ mmol(g-DW} \cdot \text{h)}^{-1}$ ATP and an NGAM $0.00 \text{ mmol(g-DW} \cdot \text{h)}^{-1}$. It should be noted that NGAM cannot actually be 0.00, and is a reflection of the fact that measuring NGAM from batch cultures was not possible. Using this, specific uptake rates of fructose were compared across observed batch culture, model predictions with the continuous culture maintenance and model predictions with the batch culture, see figure 6.6. It was found that model predictions of fructose uptake could be better aligned to batch culture observations using the batch culture maintenance cost.

Growth rates observed during exponential phase ranged between 0.06 h^{-1} , to 0.18 h^{-1} , with a median rate of 0.097 h^{-1} . At the lowest, median and highest growth rate the q_{Fruc} predicted by the model with the continuous culture maintenance cost were 1.6, 2.1 and $4.2 \text{ mmol(g-DW} \cdot \text{h)}^{-1}$. The predicted q_{Fruc} using maintenance costs estimated from batch culture were 1.7, 2.5 and $4.7 \text{ mmol(g-DW} \cdot \text{h)}^{-1}$. For the same growth rates the q_{Ac} predicted by the model using continuous culture maintenance were 3.5, 5.0, $8.9 \text{ mmol(g-DW} \cdot \text{h)}^{-1}$, and 3.7, 5.6, $10.5 \text{ mmol(g-DW} \cdot \text{h)}^{-1}$ using batch culture maintenance. This data is presented in figure 6.7[A] for all growth rates, and at the growth rate of 0.11 h^{-1} in 6.7[B]. Overall, the highest discrepancy was $1.6 \text{ mmol(g-DW} \cdot \text{h)}^{-1}$ acetate at the highest growth rate and the lowest discrepancy was $0.1 \text{ mmol(g-DW} \cdot \text{h)}^{-1}$ fructose at the lowest growth rate, and an overall average of $0.57 \text{ mmol(g-DW} \cdot \text{h)}^{-1}$ discrepancy across all growth rates for fructose and acetate. This shows that the specific rates predicted using both sets of maintenance costs are within good range of each other. A maintenance cost of zero reduced acetate production to $0.04 \text{ mmol(g-DW} \cdot \text{h)}^{-1}$, and fructose consumption to $0.68 \text{ mmol(g-DW} \cdot \text{h)}^{-1}$ at a 0.11 h^{-1} growth rate, well outside *in vivo* ranges. Ammonium, phosphate, sulphate and other micronutrient uptake was unaffected by varying maintenance.

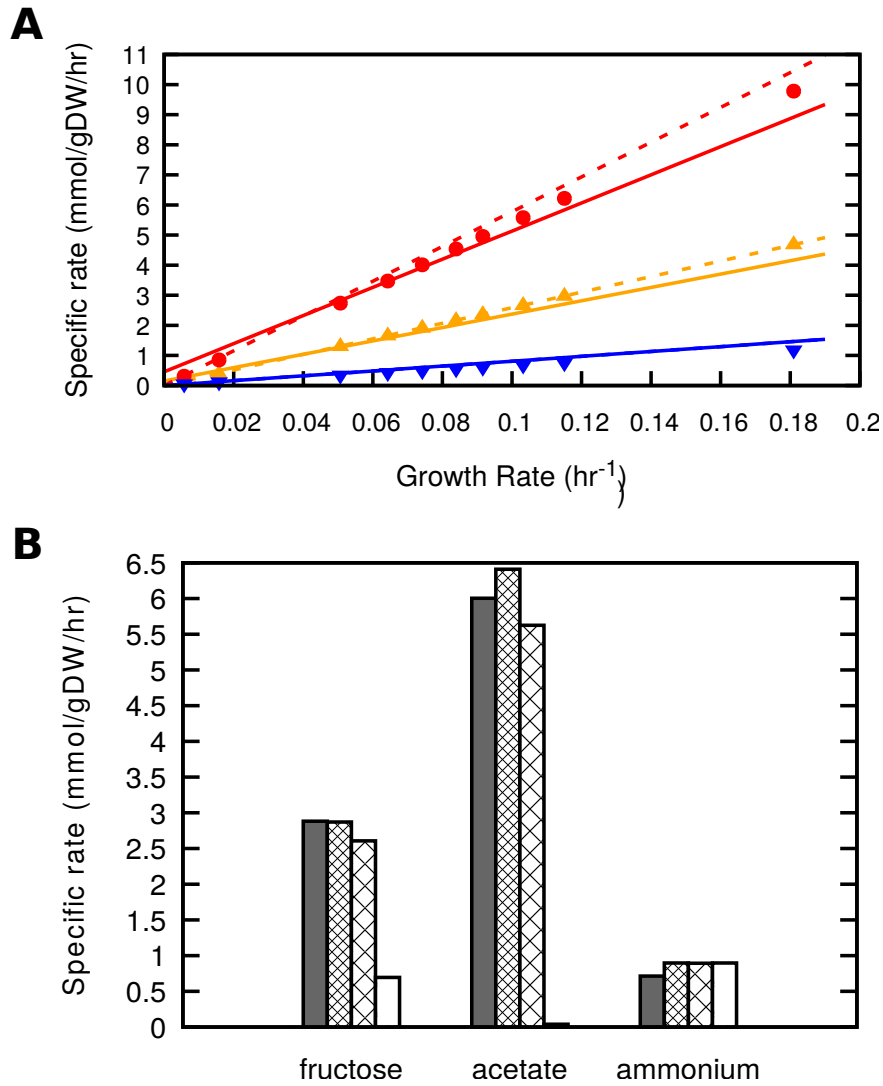


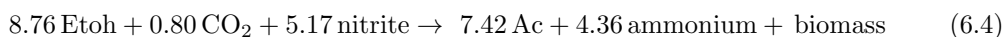
Figure 6.7: Comparisons of model specific rates with different maintenance costs to specific rates from the the fructose experiment. continuous culture maintenance rates were taken from [47]. **[A]** Specific uptake/production rates across different growth rates. Yellow circles and lines indicate acetate production, red upward triangles and lines indicate specific fructose consumption and downward blue triangles and lines indicate specific ammonium uptake. Solid lines indicate model predictions using maintenance costs from continuous cultures, and dashed lines from costs derived here using specific rates from the fructose batch culture. **[B]** Specific uptake/production rates at a growth rate of 0.11 h^{-1} using different maintenance costs. Filled bars are rates determined experimentally. Finely-crossed patterned bars indicate model predictions when parameterised with maintenance cost data from fructose experiments done here. Loosely-crossed patterned bars indicate model predictions when parameterised with maintenance cost data from continuous fructose cultures. Empty bars are model predictions with zero maintenance costs.

Table 6.1: Comparing model predicted and laboratory observed transport of metabolites at a growth rate of 0.11h^{-1} . Units are in $\text{mmol}(\text{g-DW} \cdot \text{h})^{-1}$ unless specified otherwise. AI2 is autoinducer 2, also known as (S)-4,5-dihydroxy-2,3-pentanedione. DRP is 5-deoxy- α -ribose 1-phosphate.

Chemical Compound	Specific rates ($\text{mmol}(\text{g-DW} \cdot \text{h})^{-1}$)	
	model	observed
Fructose	-2.616	-2.882
Ammonium	-0.897	-0.713
Phosphate	-0.144	nd
Sulphate	-0.015	nd
Pantothenate	>-0.001	nd
Cobalt	>-0.001	nd
Acetate	5.646	6.004
Formate	0.540	nd
H ₂	0.090	nd
AI2	>0.001	nd
DRP	>0.001	nd
Biomass	1g	1g

6.5.3 Model analysis of ethanol-nitrite based growth

The model analysis was set up for biomass production in two scenarios - both allowed for ethanol-CO₂ as carbon and energy sources. The first used ammonium as sole nitrogen source (see stoichiometry in 6.3). This solution is identical to that shown in chapter 5 section 5.2.1 but the ammonium uptake is shown here. The second used nitrite as sole nitrogen source (see stoichiometry in 6.4).



The biomass yield on ethanol increases when nitrite is used from $6.6\text{g-DW} \cdot \text{mol}^{-1}$ to $11.4\text{g-DW} \cdot \text{mol}^{-1}$. This is reflected by the percentage of carbon from ethanol that is found in biomass which increases from 7.7% to 19.9%. When nitrite is used, only trace amounts of CO₂ is needed for building biomass precursors. Acetyl-CoA is not produced from Woods-Ljungdahl pathway, but instead from acetaldehyde. The methyl-branch of WLP still carries low levels of flux. Methylene-THF dehydrogenase for example carries flux for methionine synthesis from methyl-THF. The nitrite both supplies nitrogen for biomass and is used as a redox sink, using electrons from ethanol oxidation. The amount of ammonium produced is exactly proportional to the amount of nitrite oxidation needed for energetic purposes and

Table 6.2: Elemental breakdown of genome-scale model biomass composition is in the range of other bacteria and of specific reported *A. woodii* measurements. Elemental composition is in % weight/weight

	C	H	O	N	P	S
Cmol stoichiometry	1	1.923	0.607	0.232	0.037	0.004
Elemental composition (%)	42.62	6.83	34.49	11.53	4.07	0.45

so also proportional to the amount of NADH oxidation needed for energetic purposes.

6.5.4 Comparing other reported data to modelling predictions

Biomass equation validation and atomic balances

The *A. woodii* GSM biomass composition as a Cmol stoichiometry and the elemental composition is shown in table 6.2. The elemental composition of biomass has been reported to be 41.8%-48.6% carbon, 10.6%-13% nitrogen and 6.1%-6.9% hydrogen in percentage weight/weight [47], depending on growth conditions such as growth rate, presence or absence of CO₂ and presence or absence of yeast extract. Biomass proportions in the model are well within these ranges. They are also within ranges of elemental compositions for several microorganisms in [126]. The Cmol biomass equation was used to close the nitrogen balance in the fructose fermentation to 0.05 cmol·mol⁻¹ ammonium (101.00% recovery).

Acetate to substrate ratio comparisons

To validate model predictions during growth in a range of conditions, data on acetate/substrate ratios were taken from the literature and acetate/fructose from here and compared to model predictions in equivalent conditions, see figure 6.8. The substrates compared were fructose-CO₂, CO₂-H₂, lactate-CO₂, lactate-CO₂-H₂, methanol-formate and methanol-CO₂ [4, 88, 47, 5]. As well as different substrates, the ratio across a range of 37 growth rates in total were used. Three distinct bands of ratio were found that corresponded to the number of carbons in the substrate. Single carbon substrates formed ≈ 0.5 mol/mol acetate, three carbon substrates formed ≈ 1.5 mol/mol acetate and six carbon substrates formed ≈ 2.2 mol/mol acetate. The residual sum of squares was 0.119 indicating a tight fit between the model and data.

6.6 Chapter Discussion

6.6.1 Fructose metabolism in experiments and modelling

Using axenic cultures of *A. woodii* DSMZ 1030 for batch growth with fructose as the carbon and energy source in a CO₂-N₂ gas phase, resulted in rates of growth and nutrient exchange

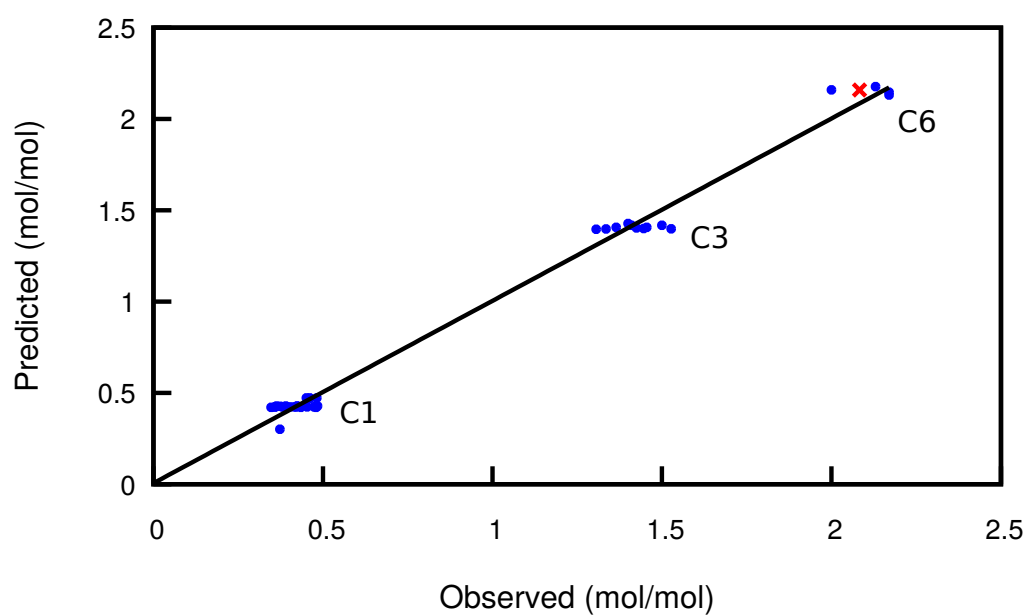


Figure 6.8: Reported acetate/substrate ratios 37 conditions reported in the literature [4, 88, 47, 5] compared to ratios predicted by the model in equivalent conditions. Red 'x' marks the acetate/fructose ratio observed in the fructose fermentation experiment here. The number of carbons in the substrates are indicated by 'C' followed by a number. A tight fit between model and data was found with a residual sum of squares of 0.119.

comparable to those reported in similar conditions under continuous growth. The biomass yields achieved during exponential growth were only slightly lower than the ones from continuous cultures, but the ratios of acetate to fructose were all in range of those reported. The biomass yields predicted by the model were also similar to the ones in batch and continuous growth but closer to continuous growth yields. For example at a growth rate of 0.11h^{-1} the biomass yield on fructose batch growth was $38.5\text{g-DW}\cdot\text{mol}^{-1}$; in continuous culture at the same growth rate $42.1\text{ g-DW}\cdot\text{mol}^{-1}$, and in the model at this growth rate $42.4\text{g-DW}\cdot\text{mol}^{-1}$.

Carbon recovery rates were above 95% for the majority of the fermentation and 93.5% during exponential growth. This was better than the 89.4% recovery rate in [47], showing that the HPLC, and dry-weight assay and Cmol stoichiometry from the model biomass composition were all consistent and in alignment. Indeed the elemental composition of the model biomass composition was in close alignment with the reported composition [47] and it was also possible to close the nitrogen balance to 101%, showing that all consumed nitrogen is found within biomass. Model analysis also showed that some carbon would not be found in either acetate or biomass, but instead accumulated as formate. Given that in [47] CO_2 was measured and found to be consumed in trace amounts, the lost carbon is unlikely to be evolved CO_2 . Additionally, no significant ethanol production was detected here. Inducible formate accumulation has been reported in *A. woodii* [113, 27] and other acetogens [42]. However, no indication in HPLC tracer charts suggested formate was accumulating. There could potentially be a change in biomass composition in later stages of growth that causes the roughly 10% discrepancy, or there is accumulation of a yet unidentified compound.

Since metabolism during batch growth is dynamic with several transitory phases, calculating rates essential for genome-scale model parametrisation is preferably done using data from continuous cultures. However, maintaining several continuous cultures is more technically demanding and laborious. Being able to have good estimations of rates from batch culture experiments thus would be desirable. Given the reproducibility of the various biomass yields, specific uptake and production rates were calculated using the biomass yields at exponential phase and the observed growth rates throughout the fermentation. This allowed for the estimation of maintenance cost parameters from a batch culture. Since true maintenance cost parameters were available for *A. woodii* from continuous cultures as well as a genome-scale model of *A. woodii*, it was possible to evaluate whether estimations of maintenance cost parameters from a batch growth allowed for reliable model predictions. This was indeed the case at the range of growth rates observed during exponential growth for fructose, acetate and ammonium, and likely also true for sulphate and phosphate uptake rates.

6.6.2 Alternative nitrogen sources in acetogens

An initial exploratory investigation of the effect nitrite has on *A. woodii* growth was carried out here. It was found that nitrite at concentrations as low as 1mM inhibited fructose

dependent growth. Using ethanol as carbon and energy source showed some promise, but cell densities were also detrimentally affected by nitrite supplementation. Potential nitrite dependent growth was observed however between 5mM and 8mM nitrite concentrations. Ammonium and CO₂ were not essential for this growth. CO₂ has been reported to be essential for ethanol based growth [10]. However, the modelling done here shows that CO₂ would only be needed in trace amounts as the energetic requirements would be fulfilled by nitrite reduction and compounds in yeast-extract could provide the cell with utilisable carbon. Alanine for example, has been shown to be converted to pyruvate [31], and able to replace the function of CO₂ (see chapter 5).

To date, two acetogens have been reported to consume nitrate for growth in ammonium-free media, *Clostridium ljungdahlii* and *Moorella thermoacetica* [121, 83]. *A. woodii* was one of five acetogens reported to not grow on nitrate supplemented ammonium-free media, but growth on nitrite was not investigated. *M. thermoacetica* was shown to dissimilate nitrate to nitrite and ammonium for energy-conservation and growth. Furthermore, nitrite also served as an electron-accepting energy-conserving nitrogen source, resulting in ammonium excretion. A key feature of using nitrate or nitrite was the marked increases in biomass yields, when compared to using ammonium [122]. Interestingly, nitrate supplementation completely abolished acetate production when using CO₂ in combination with either H₂, CO, formate, vanillate, vanillate-CO, glycolate or propanol [42, 122]. Ethanol however, was the only substrate tested that still produced stoichiometric amounts of acetate, albeit at lower yields [122]. Inhibition of WLP enzymes is postulated for these observations. Finally, all experiments showed that yeast extract was an essential component on ammonia-free media supplemented with either nitrate or nitrite.

The higher biomass yields are reflected by the modelling analysis. Authors in [121] show that nitrate is clearly the superior electron sink in terms of available Gibbs free energy but also postulate whether the increased biomass yields are due to quantitative differences in the amount of reductant transformed. The modelling done here supports the latter argument. Reports from several other microbes growing on nitrate/nitrite have also shown increased molar yields when growing anaerobically [44, 93, 50].

It is also of interest that yeast extract was an essential media component in all experiments with nitrate or nitrite supplementation. Since evidence was presented that WLP activity is suppressed by nitrate-dependant inhibition of WLP enzymes [42], a role for WLP other than in energy conservation and acetyl-CoA synthesis could be speculated. Reaction essentiality analysis of WLP enzymes was carried out in chapter 5 and showed that WLP played a key role in methionine synthesis via methylene-THF dehydrogenase. This enzyme's was found to be essential in all conditions that did not use methanol or glycine betaine as substrates. This enzymes activity was also one of the most suppressed by nitrate amongst WLP enzymes decreasing from 2.8 to 0.8 U/mg in the presence of nitrate [42]. Together, this provides evidence for an expanded role played by WLP.

Chapter 7

Producing Value-Added Compounds

7.1 Introduction

The potential of using *A. woodii* for production of commodity chemicals is reported in this chapter. Products of biotechnological interest were identified from several sources. Our collaborators at DTU were interested in 3-hydroxypropionate, past collaborators at ZuvaSyntha were interested in 1,3-butanediol, and meetings at C1Net (now CCNet), provided much discussion on products of interest. Compounds of interest identified by the DoE in the United States for valuable commodity chemicals were also considered [147]. Many of the compounds have existing native or engineered biosynthetic routes available in patent publications or scientific literature. In regards to engineering *A. woodii* metabolism, to produce these compounds, in most cases heterologous enzymes would need to be knocked-in. Since knocking-in enzymes (as reactions) is trivial in a computational model, investigating the production of a range of compounds, using a range of biosynthesis pathways was a suitable and interesting analysis with the computational models developed here. The Genome-scale Metabolic Model presented in chapter 4 was used for growth based analysis, and the bioenergetic core presented in chapter 5 was used for elementary modes analysis.

7.2 Method

7.2.1 Construction of heterologous pathways

Pathways to allow the synthesis of heterologous compounds were collected from three sources; scientific literature, publicly available patent literature, or suggestions from collaborators. Compounds that did not need heterologous pathways were 2,3-butanediol and acetoin.

Six pathways were constructed here as novel pathways. Information from MetaCyc and BRENDA on known enzymes was used to do this. In total 24 heterologous pathways were constructed and are discussed in the text.

7.2.2 Construction and analysis of *in silico* engineered *A. woodii* models

General workflow

The general workflow of constructing engineered *A. woodii* models is shown in Figure 7.1 and consists of four main steps:

1. A substrate feedstock of choice was selected. Three single carbon substrates were used for this study: formate, CO₂-H₂ and methanol-CO₂. Although CO₂-CO-H₂ is of biotechnology interest in several acetogens, *A. woodii* has been shown to grow poorly in CO rich gas and so CO₂-H₂ gas mixtures as one type of substrates was used instead.
2. A metabolic configuration is chosen. Two *A. woodii* networks were available for this; the bioenergetic core and the genome-scale metabolic model, and 24 heterologous compound pathways. In cases where the compound of interest could be synthesised natively (e.g. acetoin), only the native *A. woodii* network was selected.
3. The transporter for the compound of interest was selected. This was the final step in model construction, after which the complete model is compiled.
4. Analysis of the model depended on the size of the model. Elementary modes analysis could be carried out on small core but not large models and growth analysis (using flux balance analysis) could be done on the large GSM but not the small core model. Analyses included computing maximum theoretical yields of compounds, ATP yields and the effect of reaction knockouts during biomass production.

Model modifications

Since adding heterologous reactions to already validated models changed the structure, each new model was subject to *in silico* validation before analysis to ensure mass and energy conservation. This resulted in changing some reaction directions and taking care not to duplicate reactions in the GSM. This was true for the following reactions in the butanol synthesis pathway: acetoacetyl-CoA reductase; 3-hydroxybutyryl-CoA dehydratase; butanal dehydrogenase and aldehyde dehydrogenase (reactions R2, R3, R6 and R10 in Figure 7.10), which were already present in the GSM.

If two acetaldehyde dehydrogenases were present in a constructed model, as would be the case for 1,3-butanediol synthesis (route 1 in Figure 7.6), the native one was removed.

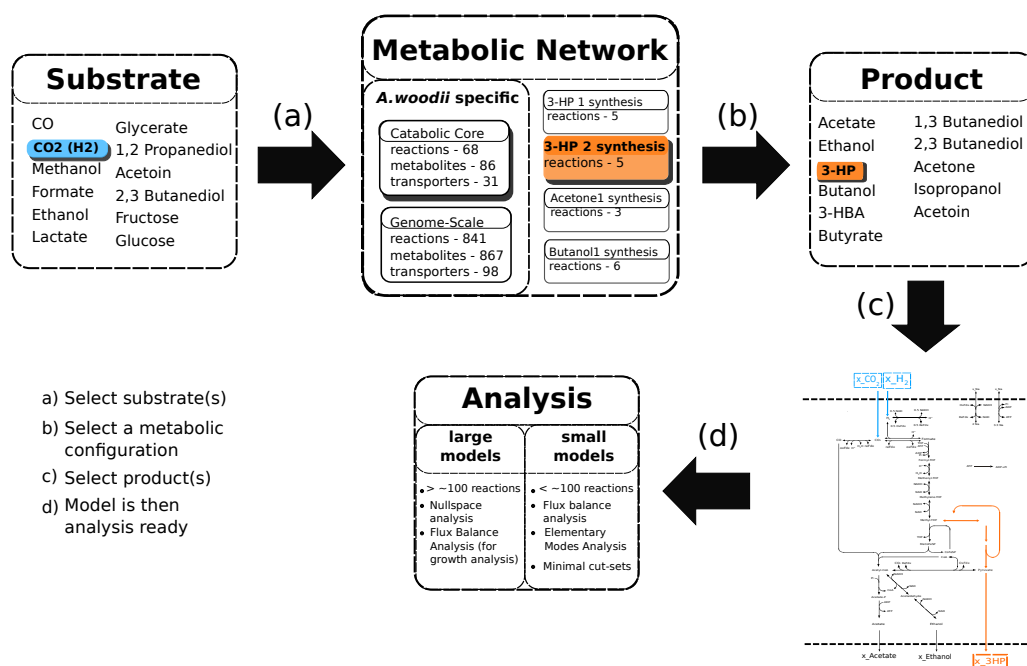


Figure 7.1: **The general workflow used to investigate the impact of producing compounds on the *A. woodii* metabolic networks.** The schematic is illustrative of the method but not comprehensive in the networks or substrates available, or type of analysis possible.

This was because the two reactions caused an unconserved energy generating loop. Two additional changes were made to all models based on an initial canvas of the impact pyruvate-formate lyase (PFL) had on C1 metabolism. The enzyme has the following locus tags: Awo_c14920, Awo_c14930 and Awo_c14960, and was identified as part of the bioenergetic core. Generally, this enzyme facilitates pyruvate lysis to formate and CO₂, and so the GSM reflects this. However, some recent reports have shown this enzyme can operate in reverse [151], and preliminary results (not shown here) showed that PFL in the pyruvate forming direction had a positive effect on product yields when C1 substrates were used. This change however, caused an unconserved energy producing loop with pyruvate ferredoxin oxidoreductase (PFOR), and so in all models constructed in this chapter PFOR and PFL are operating in the pyruvate forming direction.

7.2.3 Growth analysis of *in silico* engineered *A. woodii*

Growth analysis was carried out using the biomass production LP outlined by equation 2.4 in the computational methods (chapter 2), with additional constraint to allow export of the compound of interest. A growth rate of 0.05h⁻¹ was used for all biomass production LPs.

A reaction knockout was simulated by setting a fixed flux of zero for the reaction.

7.2.4 Yield calculations

Product compound yields are calculated as a percent of carbon converted. A 100% yield means that a 1:1 Cmol·Cmol⁻¹ conversion from substrate to product was achieved.

7.2.5 Metabolite turnover rate calculations

Metabolite turnover was computed to analyse 3-hydroxypropionate modes. A steady constraint assumes that the production of any internal metabolite is equal to the consumption of that metabolite in any given solution. This is also true for transport of mass conserved atoms i.e. the amount of carbon imported is equal to the amount of carbon exported. It was possible then to calculate a turnover rate using either the production or consumption of that metabolite for any given solution. This turnover was then normalised on a Cmol basis. This is described by defining total consumption and production as:

$$C_{\text{met}} - P_{\text{met}} = 0 \quad (7.1)$$

where C_{met} is the net consumption of that metabolite in the reactions of that solution, which is calculated by multiplying the stoichiometric coefficient by reaction flux and P_{met} is the net production of that metabolite.

$$C_{\text{met}}/J_{\text{carbon}} = T_{\text{met-carbon}} \quad (7.2)$$

where J_{carbon} is the net import of carbon, and $T_{\text{met-carbon}}$ is the turnover rate of that metabolite on a Cmol basis in mol/Cmol.

7.2.6 ATP scan

To compute growth across an increasing demand for ATP, an ATP scanning algorithm was implemented. The LP used for this was a modification of equation 2.4 and is:

$$\begin{aligned} & \text{minimize} && : |v_i| \\ & \text{subject to} && \begin{cases} \mathbf{N}\mathbf{v} = \mathbf{0} \\ v_{ATPase} = r_{ATP}; 1 \leq r_{ATP} \leq 10 \\ v_{J_B} = t_{J_{P...R}}; J_{P...R} \cdot \mu \\ v_J = v_{J_S}, v_{J_E}, v_{J_B} \end{cases} \end{aligned} \quad (7.3)$$

where r_{ATP} , the ATP requirement, is fixed to a set value. All other symbols have the same meaning as in equation 2.4 (chapter 2, section 2.2). The analysis was performed by repeatedly solving the LP, where all the constraints remained fixed except for r_{ATP} which is successively fixed to a value between the upper and lower bound.

7.3 Results - producing higher-value added compounds

The analysis done on *in silico* engineered *A. woodii* is reported in this section. In total 42 model variants were used to assess the feasibility of producing 10 compounds. 24 pathways distinct synthesis pathways were used to do this by merging them with native *A. woodii* models. Elementary modes analysis of 24 small models was carried out and flux balance analysis on the GSM using 18 metabolically engineered variants of it. All models were tested using single carbon substrates $\text{CO}_2\text{-H}_2$, formate and methanol- CO_2 . Special attention was given to 3-HP and 1,3-BDO because of interest in these products of collaborators and of butanol because of the promising features of the butanol producing pathways.

7.3.1 3-Hydroxypropionate production from $\text{CO}_2\text{-H}_2$

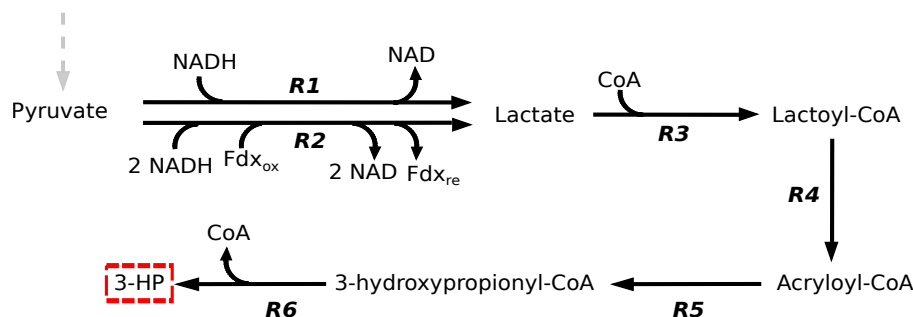


Figure 7.2: Two routes to 3-hydroxypropionate (3-HP) synthesis from pyruvate. **Route 1** is: R1, R3, R4, R5 and R6. Route provided by collaborators at DTU. **Route 2** is: R2, R3, R4, R5 and R6. The reaction numbers correspond to the following enzymes, **R1**: Lactate dehydrogenase, **R2**: Lactate dehydrogenase/Electron transferring flavoprotein complex, **R3**: Lactoyl-CoA transferase, **R4**: Lactoyl-CoA dehydratase, **R5**: 3-Hydroxypropionyl-CoA dehydratase, **R6**: 3-Hydroxypropanonyl-CoA hydrolase.

Modes producing 3-Hydroxypropionate

The routes to 3-HP from pyruvate are shown in Figure 7.2. The routes differ in the lactate dehydrogenase (LDH) used; one uses a NADH-dependent LDH (route 1) and the other utilises the native Lactate dehydrogenase/electron transferring flavoprotein (LDH/Etf) complex (route 2). Route 1 was suggested to us by collaborators at DTU.

Elementary modes analysis was carried on the two alternative 3-Hydroxypropionate (3-HP) modes. Four modes were identified that produced 3-HP in varying yields, see figure 7.3. The metabolic flux in these modes were normalised on a per carbon basis, and are ordered from highest to lowest in terms of $Y_{3\text{HP}}$. Modes 3 and 4 used the NADH-dependent

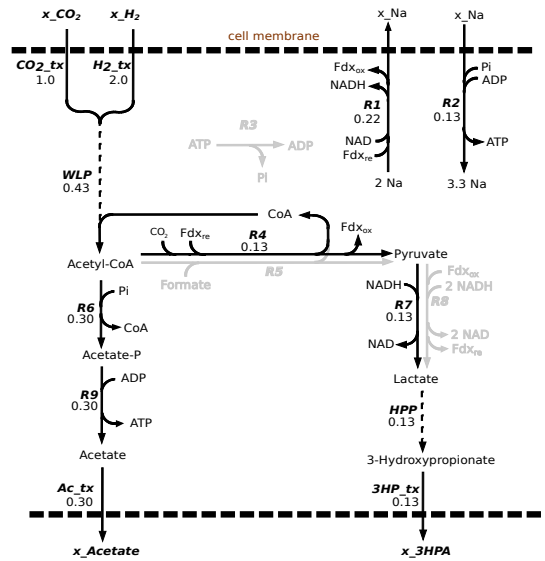
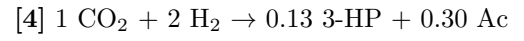
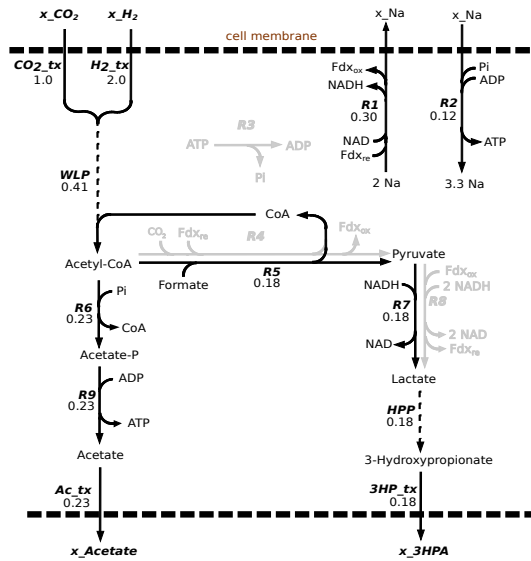
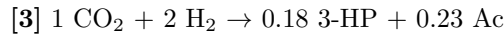
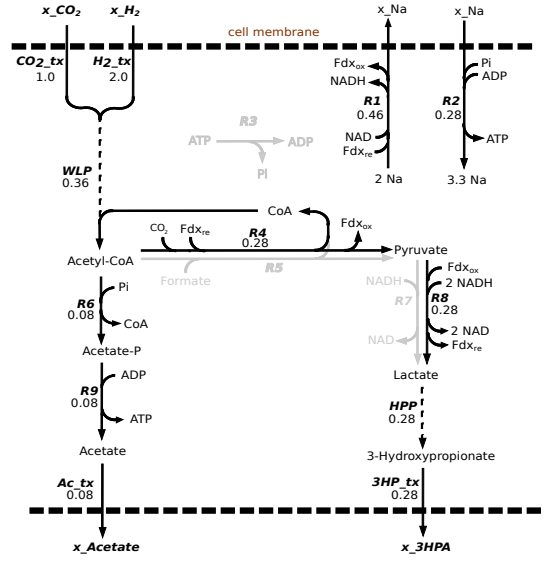
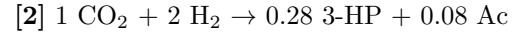
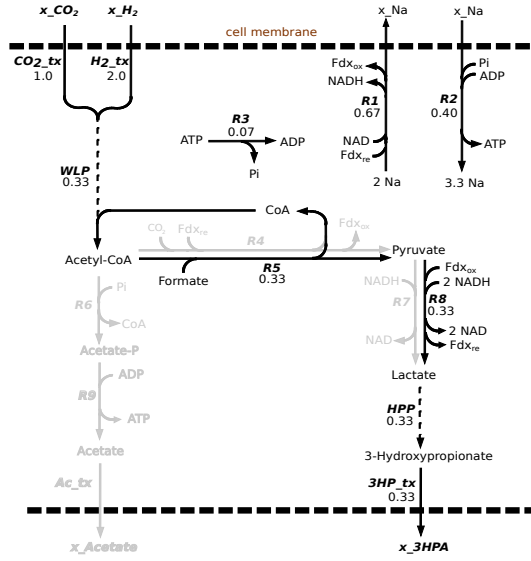
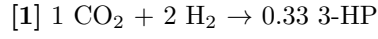


Figure 7.3: Elementary modes of 3-hydroxypropionate synthesis when using CO₂-H₂ as substrate feedstocks. Modes are ordered according the highest (mode 1) to lowest (mode 4), 3-HP yields. The mode number also corresponds to the mode numbers in table 7.1.

Table 7.1: **Elementary modes of 3-HP production using CO₂-H₂.** The number of reactions involved in each mode, Y_{HP} and the Y_{ATP} achieved is shown. Yields are given in mol·Cmol⁻¹, and are rounded to two decimal places.

Mode	Y _{HP} (% carbon)	Y _{ATP} (mol/Cmol)	Reactions	Turnover rates (mol/Cmol)			
				NADH	Fdx _{re}	Acetate	Na ⁺
1	100.0	0.07	21	1.33	1.00	0.00	1.33
2	83.3	0.00	23	1.28	1.10	0.08	0.92
3	53.6	0.00	23	1.00	0.71	0.23	0.59
4	39.5	0.00	23	1.00	0.78	0.30	0.43

LDH (route 1), and modes 1 and 2 used the native bifurcating LDH (route 2). Additionally, Modes 1 and 3 use PFL, and modes 2 and 4 use PFOR.

Analysis of the 3-Hydroxypropionate modes

A summary of the features of the four 3-HP modes are shown in table 7.1. A theoretical maximum yield of 100% (Cmol·Cmol⁻¹) was achievable by mode 1, as well as a small but positive ATP yield of 0.07 mol·Cmol⁻¹ (0.21 mol·mol⁻¹ 3-HP). This Y_{ATP} is 0.01 mol·Cmol⁻¹ lower than what is achieved with the formate to acetate conversion (see chapter 5), which is currently the lowest Y_{ATP} thought to support growth in *A. woodii*.

To gain insight into factors changing in throughout these modes as Y_{3HP} increases, turnover of important intermediates was investigated. Turnover rates of metabolites in the four modes are shown in Figure 7.4. It was found that in each successive mode that increased the Y_{3HP} there was a decrease in substrate level phosphorylation (SLP) by acetate kinase (Ack) and an increase in Na⁺ turnover by the respiratory Rnf complex. An increase in respiratory activity would be necessary in scenarios where the opportunity for SLP decreases and is achieved by redistributing redox reactions.

NADH turnover trended upwards between modes 1-3, whereas Fdx_{re} both increases and decreases. This indicates that the synthesis of a product without SLP needs an increased amount of ATP generated by electron transfer and that an increase in electron transfer is facilitated by increasing NADH turnover through the Rnf complex. Since NADH is produced by the Rnf complex (if Na⁺ is being exported as is occurring here), increased opportunity for oxidation of this NADH allows for an increase in its turnover. This is seen in mode 1 which has the highest NADH and Na⁺ turnover. This is supported by a reduced need for Fdx_{re}, which allows for more NADH production to be shifted to the Rnf and away from H₂-dependent electron bifurcation. In this way, exactly half of the NADH produced by Rnf is oxidised by WLP and half is oxidised by 3-HP production.

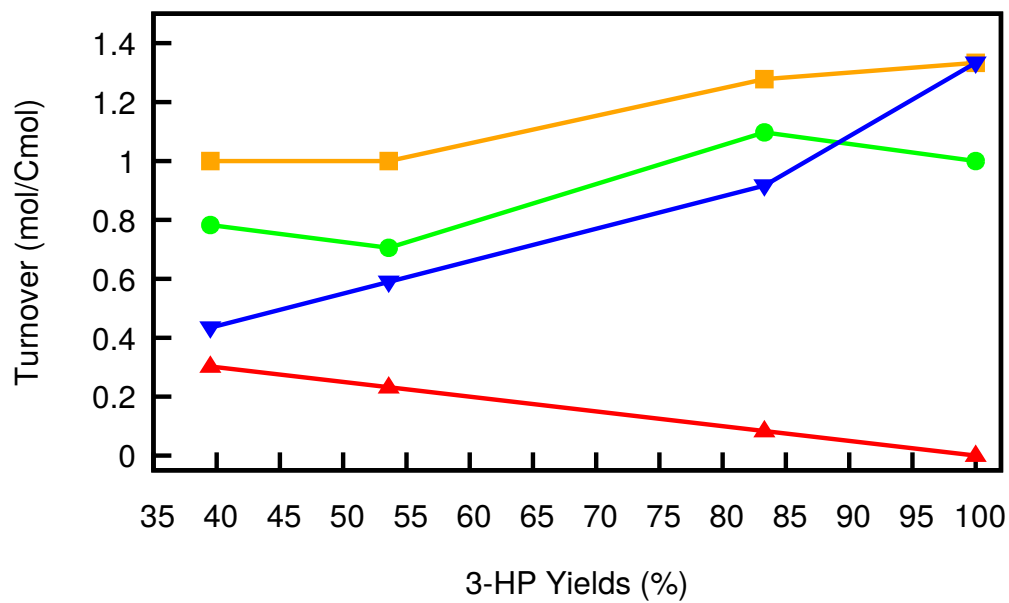


Figure 7.4: **How turnover rates of important intermediates change as 3-hydroxypropionate (3-HP) yields increase across the four elementary modes.** Red downward triangles and lines is substrate level phosphorylation turnover rates according to acetate synthesis, blue upward triangles and lines is respiratory turnover rates according to Na^+ turnover, orange squares and lines is NADH turnover and green circles and lines is Fdx_{re} .

7.3.2 A survey of higher-value added chemical production with single carbon substrates

Description of heterologous pathways

Schematics of the 24 heterologous pathways investigated here are shown in Figures 7.2 and 7.5 - 7.10. The pathways were either taken from scientific reports, published patents, or novel pathways constructed as part of this work here.

3-Hydroxypropionate (3-HP) cannot be synthesised natively. Routes to synthesis of this compound has already been discussed in the previous section.

1,3-Butanediol (1,3-BDO) cannot be synthesised natively. Routes to synthesis of this compound is shown in Figure 7.6, and were taken from [77] which utilises a variant of a deoxyribose phosphate aldolase (DERA) enzyme EC-4.1.2.4. The three routes differ in their acetaldehyde synthesis. Route 1 uses an AOR, route 2 a pyruvate decarboxylase and route 3 a native acetaldehyde dehydrogenase from acetyl-CoA.

2,3-Butanediol (2,3-BDO) can be synthesised natively. Routes to synthesis of this compound can be seen in Figure 5.4, in chapter 5

Acetoin can be synthesised natively. Routes to synthesis of this compound can be seen in Figure 5.4, in chapter 5

Acetone cannot be synthesised natively. Routes to synthesis of this compound is shown in Figure 7.8. Route 1 is native to other acetogens and was taken from [61]. Route 2 is an engineered pathway and taken from [76], and uses the phosphate butyryltransferase (Ptb) and butyrate kinase (Buk) reactions (Ptb-Buk). Route 3 was constructed here. It consists of the native pathway with the acetoacetate decarboxylase (R4 in Figure) substituted with an acetone carboxylase (EC-6.4.1.6) (R6 in figure).

Butanol cannot be synthesised natively. Routes to synthesis of this compound are shown in Figure 7.10. Route 1 is engineered and taken from [40]. Route 2 is also engineered and taken from [76] using the Ptb-Buk reactions. Routes 3 and 4 have been constructed here and are the same as routes 1 and 2 but with crotonyl-CoA reductase replaced with the bifurcating butanoyl-CoA dehydrogenase (EC-1.3.1.109).

Butyrate cannot be synthesised natively. Routes to synthesis of this compound are shown in Figure 7.9. Routes 1 and 2 native to other acetogens and was taken from [61] and [131] respectively. Routes 3 and 4 have been constructed here and are the same as routes 1 and 2

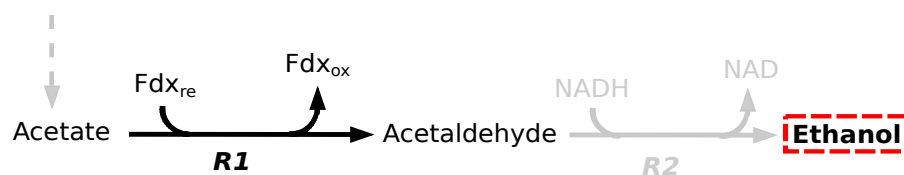


Figure 7.5: **Route to ethanol synthesis constructed here.** Grey arrow indicates the enzyme is native to *A. woodii*. Reaction numbers correspond to following enzymes, **R1**: Aldehyde:ferredoxin oxidoreductase and **R2**: Ethanol dehydrogenase

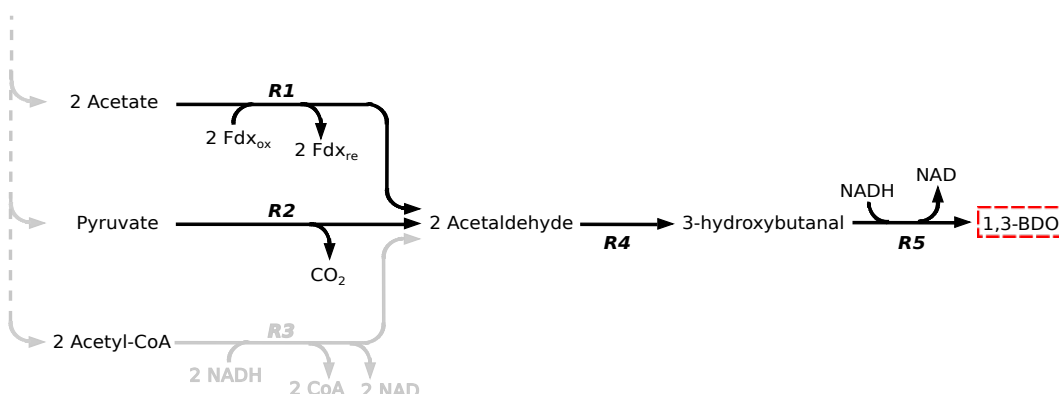


Figure 7.6: **Routes to 1,3-butanediol (1,3-BDO) synthesis taken from [77].** **Route 1** is: R1, R4 and R5. **Route 2** is: R2, R4 and R5. **Route 3** is: R3, R4 and R5. Reaction numbers correspond to following enzymes, **R1**: Aldehyde:ferredoxin oxidoreductase, **R2**: Pyruvate decarboxylase, **R3**: Acetaldehyde dehydrogenase, **R4**: Deoxyribosephosphate aldolase (DERA) variant and **R5**: Alcohol dehydrogenase.

but with crotonyl-CoA reductase replaced with the bifurcating butanoyl-CoA dehydrogenase (EC-1.3.1.109).

Ethanol can be synthesised natively, but ethanol synthesis after the addition of an AOR (EC-1.2.7.5) was investigated here, see Figure 7.5.

3-Hydroxybutyrate (3-HBA) cannot be synthesised natively. Routes to synthesis of this compound are shown in Figure 7.7. Route 1 is engineered and taken from [139]. Route 2 is also engineered and taken from [76] (Ptb-Buk reactions).

Isopropanol cannot be synthesised natively. Routes to synthesis of this compound are the same as those for acetone and shown in the same Figure, with an additional propan-2-ol dehydrogenase. Route 1 was taken from [70], route 2 from [76] (Ptb-Buk reactions) and route 3 constructed here using the acetone carboxylase (EC-6.4.1.6) enzyme.

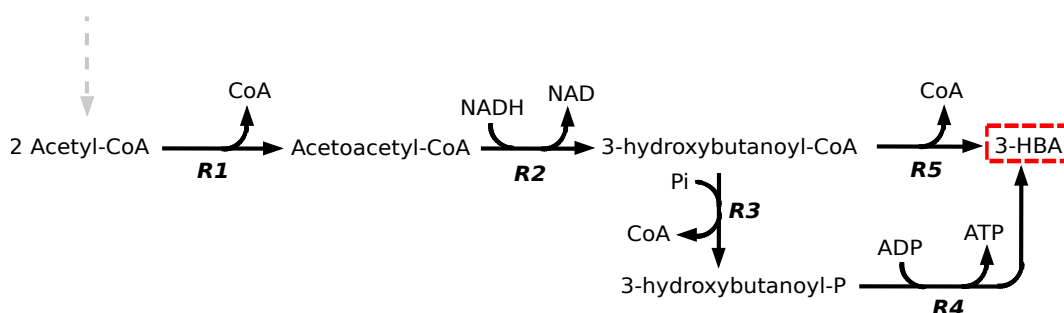


Figure 7.7: **Routes to 3-hydroxybutyrate (3-HBA) synthesis.** **Route 1** is: R1, R2 and R5 taken from [139]. **Route 2** is: R1, R2, R3 and R4 taken from [76]. Reaction numbers correspond to following enzymes, **R1**: Acetyl-CoA acetyltransferase, **R2**: Acetoacetyl-CoA reductase, **R3**: Phosphotransbutyrylase, **R4**: Butyrate kinase, **R5**: Acyl-coA hydrolase

Heterologous product yields

The product (Y_P) and ATP (Y_{ATP}) yields of heterologous compounds across C1 substrates is shown in table 7.2. Compounds that returned good Y_P and Y_{ATP} using multiple routes were butanol and butyrate. For both of these, it was found that substituting crotonyl-CoA reductase with a bifurcating butanoyl-CoA dehydrogenase, resulted in a better Y_{ATP} , and an increase to 100% Y_P if not already achieved by the non-bifurcating version of the pathway.

3-HBA returned low Y_P , and Y_{ATP} for both routes. 100% 1,3-BDO yields with $\text{CO}_2\text{-H}_2$, and methanol- CO_2 as well as positive Y_{ATP} for all three substrates when the route via acetate (route 1) was used. 2,3-BDO and acetoin both returned low yields and zero Y_{ATP} .

Routes 1 and 2 for acetone and isopropanol use quite different enzymes but returned the same Y_P , and Y_{ATP} , across all substrates. The Ptb-Buk reactions did not seem to confer advantageous yields even though additional SLP is thought to be available through it. Upon closer inspection of the elementary modes (not shown here), it was seen that this is due to the demand of the two acetyl-CoA by both pathways and in reality only a single SLP step. Essentially one ATP is required per acetyl-CoA. Since two acetyl-CoA are needed, at least two ATP are needed for a net Y_{ATP} of zero, making the ATP formed by butyrate kinase insufficient for 100% acetone and isopropanol synthesis. In route 1 this ATP is supplied through the formation of acetate. The addition of a acetoacetyl-CoA transferase (R2 in Figure 7.8) to route 2 resulted in more elementary modes but not better yields either (data not shown). However, use of route 3 for both acetone and isopropanol, allowed for a second SLP step and so supported 100% yields and 0.2 and 0.3 ATP/Cmol respectively when using $\text{CO}_2\text{-H}_2$.

It was also found that isopropanol, consistently returned higher yields than acetone. This is in keeping with the results from the previous section on 3-HP analysis that showed additional opportunity for NADH oxidation steps allowed for more flux through the Rnf and

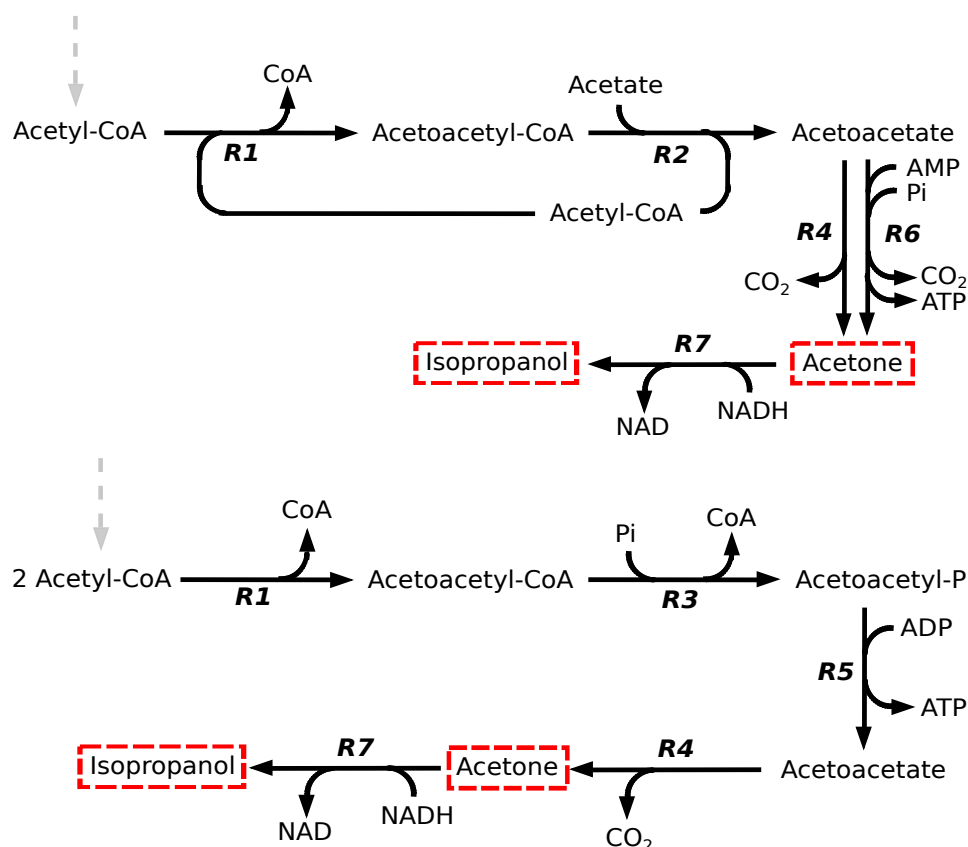


Figure 7.8: **Routes to acetone and isopropanol synthesis.** **Route 1 to acetone** is: R1, R2 and R4 taken from [61]. **Route 2 to acetone** is: R1, R2 and R6 taken from [76]. **Route 3 to acetone** is: R1, R2 and R6 constructed here. **Routes 1-3 to isopropanol** are the same as routes to acetone with the addition of R7. Reaction numbers correspond to following enzymes, **R1**: Acetyl-CoA acetyltransferase, **R2**: Acetoacetyl-CoA transferase, **R3**: Phosphotransbutyrylase, **R4**: Acetoacetate decarboxylase, **R5**: Butyrate kinase, **R6**: Acetone carboxylase, **R7**: Propan-2-ol dehydrogenase

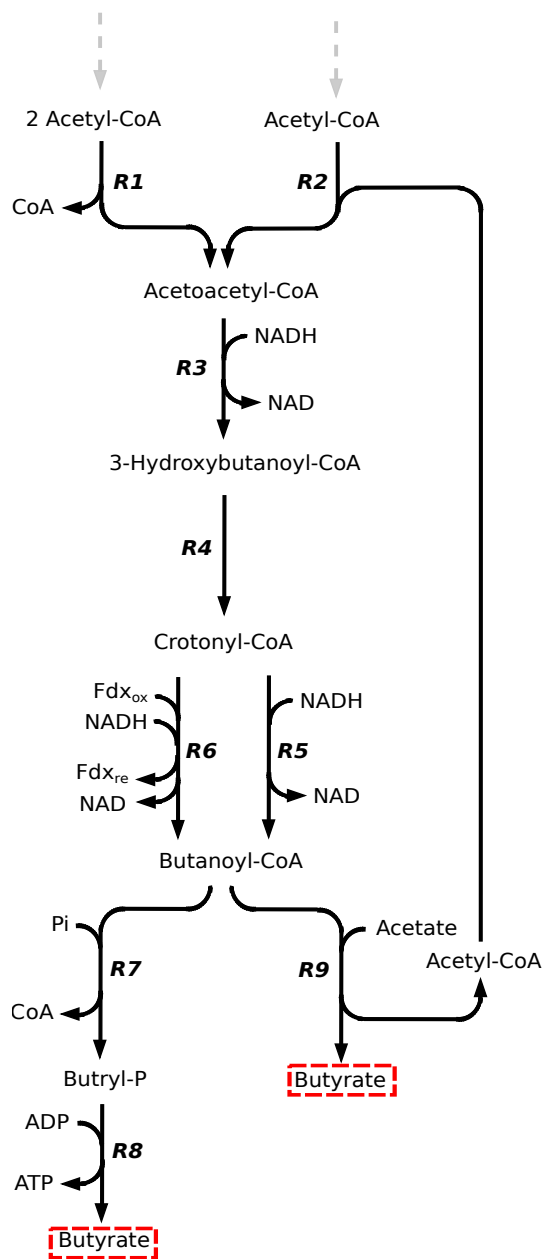


Figure 7.9: **Routes to butyrate synthesis.** **Route 1** is: R1, R3, R4, R5, R7 and R8 from [61]. **Route 2** is: R2, R3, R4, R5 and R9 from [131]. **Route 3** is: R1, R3, R4, R6, R7 and R8 constructed here. **Route 4** is: R2, R3, R4, R6 and R9 constructed here. Reaction numbers correspond to following enzymes, **R1**: Acetyl-CoA acetyltransferase, **R2**: Acetyl-CoA acetyltransferase, **R3**: Acetoacetyl-CoA reductase, **R4**: 3-Hydroxybutyryl-CoA dehydratase, **R5**: Crotonyl-CoA reductase, **R6**: Bifurcating butanoyl-CoA dehydrogenase, **R7**: Phosphotransbutyrylase, **R8**: Butyrate kinase, **R9**: Acetate CoA-transferase

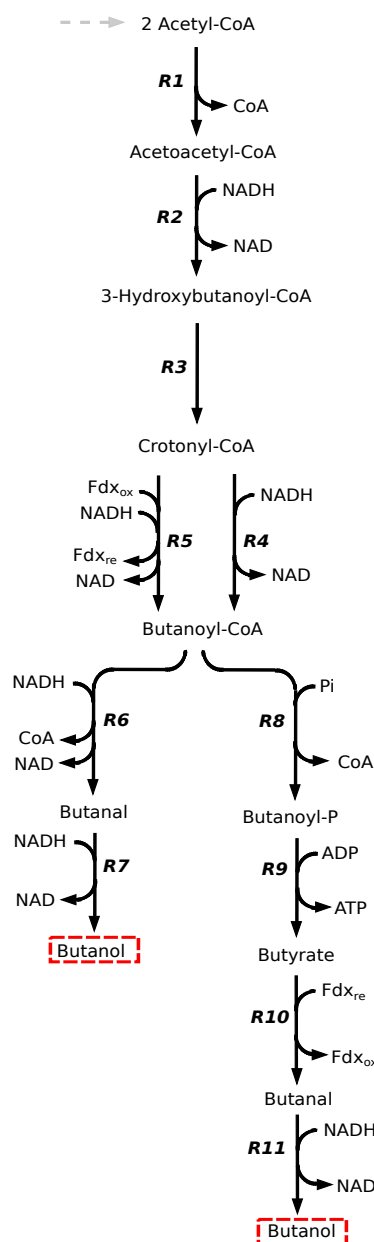


Figure 7.10: **Routes to butanol synthesis.** **Route 1** is: R1, R2, R3, R4, R6 and R7 from [40]. **Route 2** is: R1, R2, R3, R4, R8, R9, R10 and R11 from [76], **Route 3** is: R1, R2, R3, R5, R6 and R7 constructed here. **Route 4** is: R1, R2, R3, R5, R8, R9,, R10 and R11 constructed here. Reaction numbers correspond to following enzymes, **R1**: Acetyl-CoA acetyltransferase, **R2**: Acetoacetyl-CoA reductase, **R3**: 3-Hydroxybutyryl-CoA dehydratase, **R4**: Crotonyl-CoA reductase, **R5**: Bifurcating butanoyl-CoA dehydrogenase, **R6**: Aldehyde dehydrogenase, **R7**: Alcohol dehydrogenase, **R8**: Phosphotransbutyrylase, **R9**: Butyrate kinase, **R10**: Butanal dehydrogenase, **R11**: Alcohol dehydrogenase

so more ATP formed through electron transfer.

Comparing C1 substrates affect on yields

In terms of comparing the effectiveness of producing compounds from different C1 substrates, using methanol-CO₂ allowed for 100% yields from every route to product tested, see Figure 7.11. All of these routes were also accompanied with a positive Y_{ATP} ranging from 0.02 (3-HBA, route 1) to 0.7 (isopropanol, route 3).

Formate performed the worst in terms of Y_P. This was because the carbon found in products was split between acetate, CO₂, and compound of interest or CO₂ and compound of interest. This highest Y_P, achieved was by 3-HP, route 2, where carbon was split equally between 3-HP and CO₂.

CO₂-H₂ allowed for 100% yields for 1,3-BDO, 3-HP, acetone, butanol, butyrate, ethanol and isopropanol, in at least one route. It was found that ethanol synthesis would allow for the same Y_{ATP} as acetate synthesis from CO₂-H₂ if AOR was added to *A. woodii* metabolism. The additional Fdx_{re} needed to reduce acetate to acetaldehyde is produced by additional H₂ bifurcation, and the resultant additional NADH is oxidised by ethanol dehydrogenase.

7.3.3 Growth analysis of *A. woodii* using the genome-scale metabolic model

Growth feasibility of *in silico* engineered *A. woodii*

Using the GSM, whether biomass could be produced whilst also producing the compound of interest was investigated, see table 7.3. In an attempt to couple products to growth, a single Δ Ac_tx (acetate transporter) and double Δ Ac_tx+ Δ EtOH_tx (ethanol transporter) knock-out model were constructed for 3-HP, 1,3-BDO, butanol and native *A. woodii*, and tested using C1 substrates.

In native *A. woodii*, Δ Ac_tx was found to be fatal if grown on CO₂-H₂ and formate, but not methanol-CO₂. This simplified growth-coupling for these two substrates as the greater robustness seen with methanol-CO₂ results in a wider spectrum of products that is able to support growth.

For 3-HP, route 2 allowed for growth in both knockout mutants across all three substrates; using formate resulted in CO₂ production, methanol-CO₂ allowed for complete 3-HP synthesis (but not growth-coupled synthesis), and CO₂-H₂ resulted in complete 3-HP synthesis which was also strictly coupled to growth.

The same was true for route 1 of 1,3-BDO synthesis across all substrates. However, a double Δ Ac_tx+ Δ EtOH_tx was needed to achieve this. This was because the AOR present in this pathway also facilitated efficient ethanol synthesis, as is seen in the product survey table 7.2.

Table 7.2: The best elementary mode for producing higher-value added compounds, in terms of Y_P and Y_{ATP} . Three different single carbon substrates were tested with each product pathway. The citations for product pathways are given in text.

Product Pathway	CO ₂ -H ₂			Formate			Methanol-CO ₂			Fig
	Y _P	Y _{ATP}	Reactions	Y _P	Y _{ATP}	Reactions	Y _P	Y _{ATP}	Reactions	
3-Hydroxypropionate 1	53.60	0.00	23	26.80	0.00	23	100.00	0.13	21	7.2
3-Hydroxypropionate 2	100.00	0.07	21	50.00	0.04	21	100.0	0.33	19	
1,3-Butanediol 1	100.00	0.08	20	36.40	0.03	20	100.0	0.44	22	7.6
1,3-Butanediol 2	55.60	0.00	20	23.00	0.00	20	100.0	0.24	20	
1,3-Butanediol 3	55.60	0.00	20	23.00	0.00	20	100.00	0.24	20	
2,3 Butanediol	55.60	0.00	20	23.00	0.00	20	100.0	0.24	20	5.4
Acetoin	43.50	0.00	19	19.60	0.00	19	100.00	0.13	19	5.4
Acetone 1	53.60	0.00	20	22.70	0.00	20	100.00	0.22	22	7.8
Acetone 2	53.60	0.00	21	22.70	0.00	21	100.00	0.22	21	
Acetone 3	100.00	0.20	22	37.50	0.08	22	100.00	0.55	24	
Butanol 1	76.90	0.00	23	27.80	0.00	23	100.00	0.35	22	7.10
Butanol 2	100.00	0.05	23	33.30	0.02	23	100.00	0.45	24	
Butanol 3	100.00	0.11	21	33.30	0.04	21	100.00	0.50	20	
Butanol 4	100.00	0.20	23	33.30	0.07	23	100.00	0.60	24	
Butyrate acid 1	100.00	0.05	21	40.00	0.02	21	100.00	0.38	23	7.9
Butyrate acid 2	100.00	0.05	22	40.00	0.02	22	100.00	0.38	24	
Butyrate 3	100.00	0.20	21	40.00	0.08	21	100.00	0.53	23	
Butyrate 4	100.00	0.20	22	40.00	0.08	22	100.00	0.53	24	
Ethanol	100.00	0.15	19	33.30	0.05	19	100.00	0.55	20	7.5
3-Hydroxybutyrate 1	35.70	0.00	20	17.10	0.00	20	100.00	0.02	20	7.7
3-Hydroxybutyrate 2	87.00	0.00	21	39.20	0.00	21	100.00	0.27	21	
Isopropanol 1	83.30	0.00	21	29.40	0.00	21	100.00	0.36	21	7.8
Isopropanol 2	83.30	0.00	22	29.40	0.00	22	100.00	0.36	22	
Isopropanol 3	100.00	0.30	23	33.30	0.10	21	100.00	0.70	24	

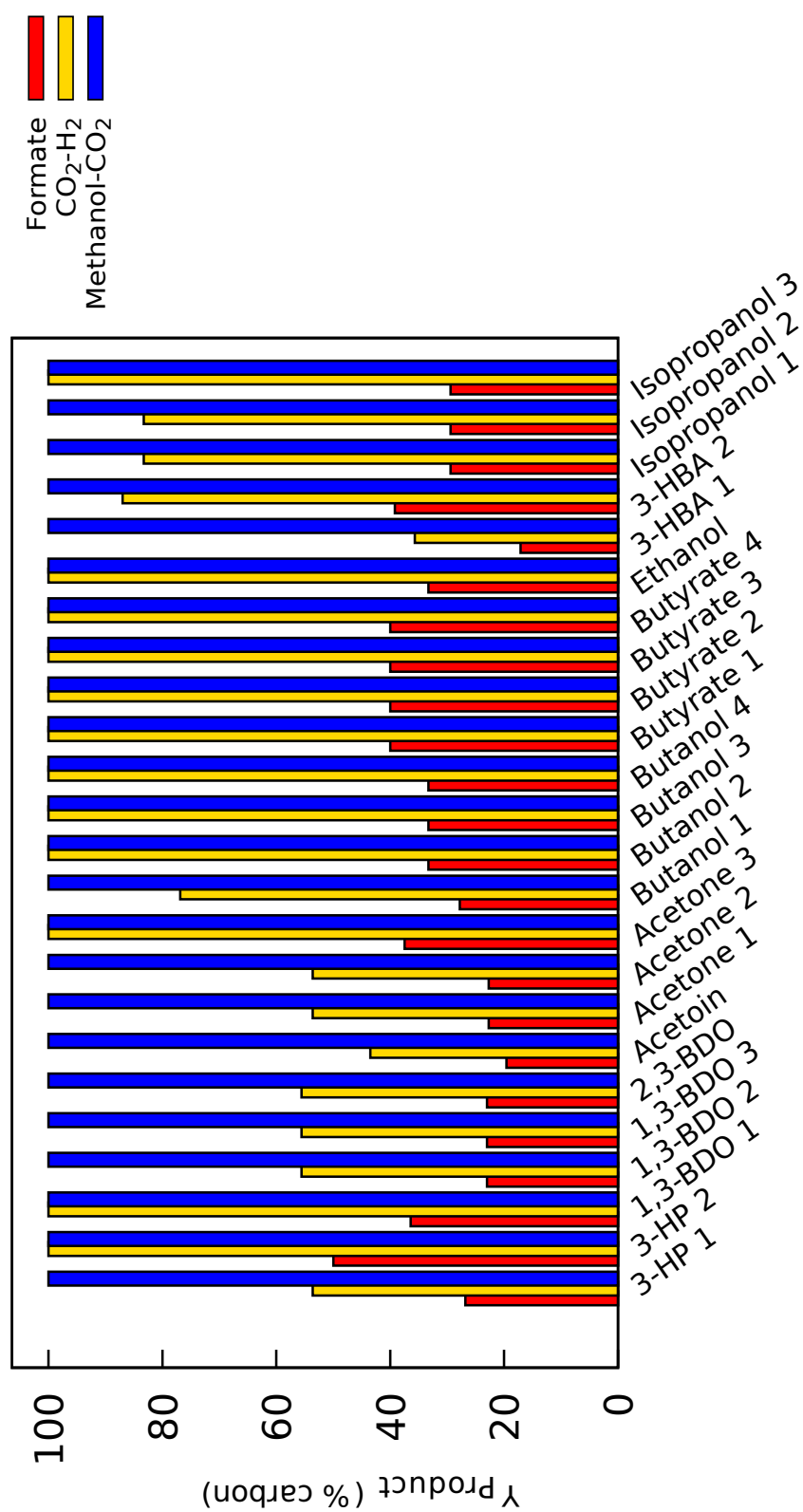


Figure 7.11: A survey of producing valuable compounds using single-carbon substrates in engineered *A. woodii*.

For butanol, only mode 1 did not allow for its synthesis across all substrates. It was found that when using methanol-CO₂, use of the bifurcating enzymes made butanol the preferential product during growth. Without this enzyme, H₂ and formate are produced as side products. As with the previous products, CO₂-H₂ and formate based growth was coupled to butanol formation in routes 2, 3 and 4.

Increasing 1,3-Butanediol yields during growth by ATP wasting

Of the model constructs discussed in the previous section above, route 1 of 1,3-BDO synthesis was both coupled to growth and the sole product when using CO₂-H₂, and the double $\Delta\text{Ac_tx}+\Delta\text{EDH}$ knockout. The growth stoichiometry for this was:



which gives a biomass yield of 5.03 g-DW·mol⁻¹ 1,3-BDO. Since this growth is 3-BDO dependent, a further decrease in biomass yield was sought to increase the overall amount of product made. Given that the nature of this growth coupling was for a supply of ATP for growth, an ATP wasting approach was employed, the result of which is shown in Figure 7.12. A linear relationship was found between ATP usage, substrate uptake, flux through WLP (as seen by flux through acetyl-CoA synthase (AcS)) and flux through the 1,3-BDO pathway (as seen by flux through DERA). Doubling the ATP requirement during growth gave a decreased biomass yield of 2.96 g-DW·mol⁻¹ 1,3-BDO, thus increasing the 1,3-BDO yield.

7.4 Discussion

As part of exploring the feasibility of producing higher-value added compounds with *A. woodii*, a comprehensive survey was completed here that investigated both the feasibility of producing a variety of compounds and alternative routes for their synthesis. Much variation was found in Y_P when using different C1 substrates. Methanol has received interest for decades as a potential feedstock for biotechnological processes [30] because of its ease of handling and low cost, and this interest has been renewed because of its potential as a renewable source [25]. Methanol-CO₂, would allow for the widest spectrum of products, but would need more metabolic engineering to couple products to growth than CO₂-H₂, and formate. Gaseous based product fermentation from CO₂-H₂, allowed for seven out ten compounds to achieve theoretical maximums of 100% yields. However, it was shown that careful consideration of the synthesis pathway must be taken into account for these yields to be met. Formate did not allow for 100% yields of any compounds and could generate ATP for only a small set of compounds.

The central issue with using formate remains that its oxidation can only be performed

Table 7.3: Computing growth in mutant *A. woodii* GSM designed for production of higher-value added compounds. Biomass production was computed at a 0.05 h⁻¹ growth rate

Product Pathway	Substrate	Growth By-Products	
		Δ Ac.tx	Δ Ac.tx + Δ Etoh.tx
None	Formate	infeasible	infeasible
	CO ₂ -H ₂	infeasible	infeasible
	Methanol-CO ₂	Formate, EtOH	H ₂ , Formate
3-Hydroxypropionate 1	Formate	infeasible	infeasible
	CO ₂ -H ₂	infeasible	infeasible
	Methanol-CO ₂	Formate, EtOH	H ₂ , Formate
3-Hydroxypropionate 2	Formate	CO ₂ , 3-HP	CO ₂ , 3-HP
	CO ₂ -H ₂	3-HP	3-HP
	Methanol-CO ₂	3-HP	3-HP
1,3-Butanediol 1	Formate	CO ₂ , EtOH	CO ₂ , 1,3-BDO
	CO ₂ -H ₂	EtOH	1,3-BDO
	Methanol-CO ₂	EtOH	H ₂ , 1,3-BDO, Formate
1,3-Butanediol 2	Formate	infeasible	infeasible
	CO ₂ -H ₂	infeasible	infeasible
	Methanol-CO ₂	Formate, H ₂	Formate, 1,3-BDO
1,3-Butanediol 3	Formate	infeasible	infeasible
	CO ₂ -H ₂	infeasible	infeasible
	Methanol-CO ₂	H ₂ , EtOH, Formate	H ₂ , 1,3-BDO, Formate
Butanol 1	Formate	infeasible	infeasible
	CO ₂ -H ₂	infeasible	infeasible
	Methanol-CO ₂	Formate, EtOH	Formate, Butanol
Butanol 2	Formate	CO ₂ , Butanol	CO ₂ , Butanol
	CO ₂ -H ₂	Butanol	Butanol
	Methanol-CO ₂	H ₂ , Butanol, Formate	H ₂ , Butanol, Formate
Butanol 3 (bifurc)	Formate	CO ₂ , Butanol	CO ₂ , Butanol
	CO ₂ -H ₂	Butanol	Butanol
	Methanol-CO ₂	Butanol	Butanol
Butanol 4 (bifurc)	Formate	CO ₂ , Butanol	CO ₂ , Butanol
	CO ₂ -H ₂	Butanol	Butanol
	Methanol-CO ₂	Butanol	Butanol

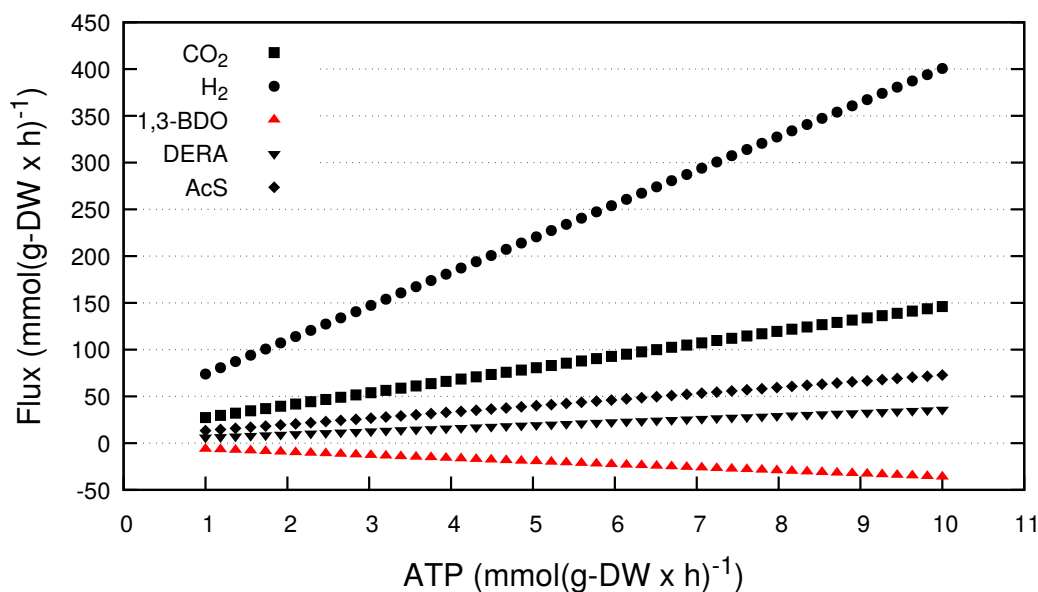


Figure 7.12: **ATP scan in a 1,3-Butanediol engineered *A. woodii*.** Increasing the ATP demand during growth increases the 1,3-Butanediol yield. Scan was done at using a 0.05 h⁻¹ growth rate. The negative 1,3-BDO flux indicates its export.

by a single enzyme, H₂-dependent formate dehydrogenase [113], which also produces CO₂. The advantage of formate over the other substrates is that it can be co-utilised with CO [8], a major component of syngas. Although this would allow for a wider product spectrum, due to more favourable bioenergetics, 50% of carbon is still likely to be lost (see chapter 3 for elementary modes of formate and formate-CO pairs). This is because, CO oxidation also results in CO₂ formation.

It was shown here that products without SLP steps in their synthesis pathways could still be produced with high yields and ATP generation if enough opportunity for NADH oxidation was provided, by analysis of elementary modes synthesising 3-HP. NADH turnover was correlated with increases in respiratory ATP generation. This is line with findings of NADH sinks allowing for higher ATP yields of substrates from chapter 5.

Promising compounds for production include butanol and butyrate. Both could be synthesised from multiple pathways with high yields with both methanol-CO₂, and CO₂-H₂. The use of recently discovered bifurcating enzymes [16] showed that yields could be increased further. Ethanol and 1,3-BDO also showed high yields across the same two substrates, and needed minimal metabolic engineering to be produced. 3-HP was found to achieve maximum yields similar to that of 1,3-BDO. It was dependent though on a reversible PFL and a reversible electron confurcating-LDH. Some evidence for this exists [146, 151] but it is

as of yet unclear, how well they would operate in the direction needed for 3-HP synthesis. Two separate routes for acetone and isopropanol synthesis returned the same yields. On the face of it, it seems that the Ptb-Buk route would allow for higher yields because of the SLP, but it was shown that this is not enough for complete single product synthesis. A third route constructed here using an acetone carboxylase (EC-6.4.1.6) in the ATP forming direction was presented here. The Y_P were promising, but further investigation is needed in using a highly regulated metabolite such as AMP [39] for high-throughput flux such as would be necessary for bulk product generation. Although butanol and butyrate engineering strategies are shown here to be promising, one must take into account their relatively lower commercial value than other compounds even though butanol in its own right can be used as a biofuel and also blended with diesel [147]. Given this, butanol and butyrate could be feasibly used as platform compounds to form higher value ones.

Chapter 8

Conclusions and Outlook

8.1 Overview

In this work, the main aims of the project have been addressed by constructing, validating and analysing three primary structural metabolic models and carrying out complementary laboratory experiments. The first model presented was of the Wood-Ljungdahl pathway (WLP), which allowed for gaseous inputs to be converted to acetate and ethanol. It was then extended to allow for fructose, formate and methanol catabolism (chapter 3). The second model was a Genome-scale Metabolic Model (GSM) of *A. woodii*. It was shown that the GSM could be used to model growth using a diverse range of substrates (chapters 4 and 5) and was predictive of various rates observed *in vivo* (chapter 6). The final model was extracted from the GSM and represents a bioenergetic core of *A. woodii* (chapter 5). This model was combined with 24 heterologous pathways, thus forming 24 model variants. Additionally, nine of the heterologous pathways were also combined with the GSM to form nine GSM variants (chapter 7). Altogether the three primary models were used as a basis for constructing 35 model variants. All models underwent validation, and their analysis centred around gaining fundamental insights into *A. woodii* metabolism and investigating metabolic engineering approaches to modifying *A. woodii* into an industrial microbe capable of producing bulk chemicals of commercial interest.

8.2 Summary of Results

In chapter 3, elementary modes analysis provided novel insight on CO metabolism. In chapter 5, novel ways to compute ATP yields showed that no single linear programming (LP) formulation provided a consistently maximum yield, and instead a combination of LPs was needed. These were used to analyse the bioenergetics of *A. woodii* metabolism. The feasibility of producing an array of higher value-added compounds was investigated in

7 using both genome-scale and small scale models. Analysis of the WLP model showed that numerous stoichiometries observed *in vivo* were each represented by an elementary mode including: acetate production; CO₂-H₂, fructose, methanol, formate and formate-CO consumption; formate accumulation from CO₂-H₂, or from CO, and CO accumulation from CO₂-H₂. Having an accurate stoichiometric description of the *A. woodii* WLP allowed for the construction of a model that could represent many growth phenotypes, and present new, yet to be discovered ones.

The GSM underwent extensive curation and validation. In terms of curation, the biomass composition encompasses the unique group B ornithine type cell wall, specific nucleic acid content and adenosylcobalamin. During curation it was found that cobalamin synthesis is linked to the synthesis of a potential quorum sensing compound. Curation also highlighted a reaction gap in lysine synthesis, an additional methanol catabolic pathway and insights into a unique central metabolism. The elemental composition of the biomass corresponded very well to an experimentally observed composition and the model could be used to accurately predict growth or no growth with 51 different substrates. When compared to laboratory data collected on secondment, specific consumption and production rates observed during fructose-based growth could also be accurately predicted by the model.

Having established four ATP calculating LPs, a basis for consistently calculating Y_{ATP}^{max} was attained. GSMs or structural models generally tend to be analysed for ATP yields using a single LP, but it was shown here no single LP could do this. These ATP LPs were used to extract reactions from the GSM to construct the bioenergetic core model. This underwent *in silico* validation but since it was a result of analysing the GSM, and since a bioenergetic core of acetogens has not been reported, it has not yet been experimentally validated apart from encompassing parts of recognisable *A. woodii* metabolism reported in the literature. It allowed for the identification of potentially novel reactions involved in ATP generation. It was also essential in constructing models for elementary modes analysis of heterologous compound synthesis using different single carbon compounds.

Several results reported here have provided novel insights into *A. woodii* metabolism. Elementary modes analysis (EMA) allowed for the reconciliation of disparate and conflicting data on CO metabolism into a coherent and consistent theory based on pairs of elementary modes working in tandem. A previously unreported Y_{ATP} for formate and formate-CO here, and is the lowest Y_{ATP} so far known to support *A. woodii* growth. The elementary modes helped to determine that H₂-dependent formate dehydrogenase (HDCR) inhibition was not enough to explain the reported bi-phasic growth on CO₂-CO-H₂, since methanol-CO₂, and fructose-CO based growth was also not observed. EMA also suggests that a likely regulatory framework, yet undiscovered, prevents the production of compounds other than acetate even if cell carbon and ATP can be generated. This will need consideration in the light of attempting to engineer *A. woodii* for production of compounds other than acetate.

ATP generation in WLP has been discussed as being dependent on the amount of Fdx_{re}

available, since this is the reductant that facilitates Na^+ export during autotrophic growth. However, several computational analysis indicate that it is not the availability of Fdx_{re} that constrains the ATP yield of a substrate, but instead the amount of NADH oxidation allowed for by the network as a whole. First, it was shown in chapter 5 that the $Y_{\text{ATP}}^{\text{max}}$ of every substrate pair that did not include caffeate, ferulate or coumarate, increased if a hypothetical NADH sink was added to the metabolic network. Then, in the same chapter it was shown that more oxidised electron acceptors also increased the $Y_{\text{ATP}}^{\text{max}}$. In chapter 7, it was shown that increased NADH turnover, and not Fdx_{re} , resulted in increased respiratory dependent ATP generation. It seems that Fdx_{re} is freely available if H_2 is present as a growth substrate, since there is no limit on electron bifurcation. What imposes the constraint on ATP formation instead is the potential accumulation of NADH not re-oxidised to NAD.

Advantages of using *A. woodii* for industrial applications include that it is a non-spore forming, mesophilic bacteria that grows on defined media. It is the most biochemically characterised acetogen and has genetic engineering tools available to it, as well as being able to utilise a wide-range use of substrates for homoacetogenic growth. In chapter 7, it was revealed that $\text{CO}_2\text{-H}_2$ and methanol- CO_2 allowed for a wide spectrum of products to be produced at high yields while simultaneously producing ATP for growth. For $\text{CO}_2\text{-H}_2$ however this was highly dependent on the synthesis pathway used. Butanol, butyrate, 1,3-butanediol (1,3-BDO), ethanol were particularly promising compounds, but this could be extended to acetone, isopropanol and 3-HP depending on enzyme efficiencies in operating reverse directions. The novel biosynthesis pathways constructed here included promising ones for acetone and isopropanol production. It was shown that the coupling of product synthesis to growth was possible using the GSM and that this allowed for 1,3-BDO yields to be increased when using an ATP wastage approach.

8.3 Outlook

Modelling the metabolism of *A. woodii* has provided novel insights into its basic metabolic architecture and provided insights into potentially useful engineering approaches for applied industrial technologies. Additionally, both the collation of material in the scientific literature needed to build metabolic models, and the novel data produced by their analysis provide new explanations for incomplete or inconsistent hypothesis, and areas of research which still require attention if extended theories on acetogen metabolism are to develop.

Although many results have been generated in this work, many questions still remain. Perhaps the most important biotechnologically relevant question is of the little understood regulatory framework that prevents the production of compounds other than acetate in high amounts. Implementing the engineering strategies identified here, will present new challenges of which understanding metabolic regulation is one. Using metabolic modelling for understanding this would provide an effective tool for robustly testing hypothesis.

Other interesting future investigations could target some of the questions raised here. Does quorum sensing play a role in acetogens and is this affected by vitamin B12 synthesis? Would additional opportunity for NADH oxidation allow for enhanced ATP generation? Can this be used to widen the product spectrum available to *A. woodii*? Can alternative nitrogen sources provide opportunities for some additional NADH oxidation? How is lysine synthesised, and does methanol have an additional oxidative catabolic route? These are all directly testable questions that would have, because of the models constructed here, a theoretical framework to be integrated into.

Within this work, the essential upstream generation of industrially relevant acetogens was investigated. Underlying bioenergetic principles, co-factor utilisation, substrates, products and other growth requirements are all elements taken into account when modelling the metabolism of *A. woodii*, and this has allowed for novel insights. What occurs downstream of this research would be the implementation of findings from here into a commercial-scale process, which would require a continued long-term effort to genetically modify *A. woodii* and analysing its phenotype in an iterative process. Growing it in a pilot scale bioreactor, and eventually in a commercial-scale bioreactor would present new challenges which would benefit from applying both theoretical and experimental approaches.

The genetic tools for genetic modification of *A. woodii* have seen progress since the start of this project [58], but highlighted in this work is much is left to be understood about its regulatory framework. Much interest and research is occurring on *A. woodii* and the metabolic models developed here will allow for an integrated framework of incorporating new findings and exploring further features of acetogenic metabolism.

Bibliography

- [1] J. Abrini, H. Naveau, and E. J. Nyns. *Clostridium autoethanogenum*, sp. nov., an anaerobic bacterium that produces ethanol from carbon monoxide. *Archives of Microbiology*, 161(4):345–351, Apr 1994.
- [2] N. M. Aslam, M. S. Masdar, S. K. Kamarudin, and W. R. W. Daud. Overview of direct formic acid fuel cells (DFAFCs) as an energy source. *APCBEE Procedia*, 3: 33–39, Jan 2012.
- [3] R. Bache and N. Pfennig. Selective isolation of *Acetobacterium woodii* on methoxylated aromatic acids and determination of growth yields. *Archives of Microbiology*, 130:255–261, Jul 1981.
- [4] A. E. Bainotti and N. Nishio. Growth kinetics of *Acetobacterium* sp on methanol-formate in continuous culture. *Journal of Applied Microbiology*, 88:191–201, Jun 2000.
- [5] A. E. Bainotti, K. Yamaguchi, Y. Nakashimada, and N. Nishio. Kinetics and Energetics of *Acetobacterium* sp. in chemostat culture on methanol-CO₂. *Journal of Fermentation and Bioengineering*, 85(2):223–229, Nov 1998.
- [6] W. E. Balch, S. Scherberth, R. S. Tanner, and R. S. Wolfe. *Acetobacterium*, a new genus of hydrogen-oxidising, carbon dioxide-reducing, anaerobic bacteria. *International Journal of Systematic Bacteriology*, 27(4):355–361, Oct 1977.
- [7] M. Basan, S. Hui, H. Okano, Z. Zhang, Y. Shen, R. Williamson, J., and T. Hwa. Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature*, 528:99–104, Dec 2015.
- [8] J. Bertsch and V. Müller. CO metabolism in the acetogen *Acetobacterium woodii*. *Applied and Environmental Microbiology*, 81(17):5949–5956, Sept 2015.
- [9] J. Bertsch, C. Öppinger, V. Hess, J. D. Langer, and V. Müller. Heterotrimeric NADH-oxidising methylenetetrahydrofolate reductase from the acetogenic bacterium *Acetobacterium woodii*. *Journal of Bacteriology*, 197(9):1681–9, May 2015.

- [10] J. Bertsch, A. L. Siemund, F. Kremp, and V. Müller. A novel route for ethanol oxidation in the acetogenic bacterium *Acetobacterium woodii*: the acetaldehyde/ethanol dehydrogenase pathway. *Environmental Microbiology*, 18(9):1462–2920, Sept 2015.
- [11] E. Biegel and V. Müller. Bacterial NA^+ -translocating ferredoxin: NAD^+ oxidoreductase. *PNAS*, 107(42):18138–18142, Oct 2010.
- [12] P. Bodegom. Microbial maintenance: A critical review on its quantification. *Microbial Ecology*, 53(4):513–523, Mar 2007.
- [13] W. T. Bolleter, C. J. Bushman, and P. W. Tidwell. Spectrophotometric determination of ammonia as indophenol. *Anal. Chem.*, 33(4):592–594, Apr 1961.
- [14] K. Braun and G. Gottschalk. Effect of molecular hydrogen and carbon dioxide on chemo-organotrophic growth of *Acetobacterium woodii* and *Clostridium aceticum*. *Archives of Microbiology*, 128:294–298, Sept 1981.
- [15] BRENDA database. BRENDA: Enzyme DB. online database url: <http://brenda.bc.uni-koeln.de>, 2004. BRENDA team: Prof. Dr. Dietmar Schomburg et al.
- [16] W. Buckel and R. K. Thauer. Flavin-based electron bifurcation, a new mechanism for biological energy coupling. *Chemical Reviews*, 118(118):3862–3886, Dec 2018.
- [17] H. Buschhorn, P. Dürre, and G. Gottschalk. Production and utilization of ethanol by the homoacetogen *Acetobacterium woodii*. *Applied and Environmental Microbiology*, 55(7):1835–1840, Apr 1989.
- [18] H. Buschhorn, P. Dürre, and G. Gottschalk. Purification and properties of the coenzyme A-linked acetaldehyde dehydrogenase of *Acetobacterium woodii*. *Archives of Microbiology*, 158:132–138, Jan 1992.
- [19] R. Caspi, R. Billington, L. Ferrer, H. Foerster, C. A. Fulcher, I. M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, A. Muller. L., Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver, and P. D. Karp. The MetaCyc databases of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research*, 44(D1):D471–D480, Jan 2016.
- [20] L. Chistoserdova, M. G. Kalyuzhnaya, and M. E. Lidstrom. The expanding world of methylotrophic metabolism. *Annu Rev Microbiol*, 63:477–499, Feb 2009.
- [21] P. Ciccaldi, K. Schuchmann, V. Müller, and S. Elliott. The hydrogen dependent CO_2 reductase: the first completely CO tolerant FeFe-hydrogenase. *Energy & Environmental Science*, 10:503, Oct 2016.

- [22] C. Claudel-Renard, C. Chevalet, T. Faraut, and D. Kahn. Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Research*, 31(22):6633–6639, Jan 2003.
- [23] European Commision. A sustainable bioeconomy for Europe: strengthening the connection between economy, society and the environment, Oct 2018.
- [24] A Cornish-Bowden. *The Pursuit of Perfection: Aspects of Biochemical Evolution*. Oxford University Press, 2004.
- [25] C. A. Cotton, N. J. Claassens, S. Benito-Vaquerizo, and A. Bar-Even. Renewable methanol and formate as microbial feedstocks. *Curr. Opin. Biotechnol.*, 13(62):168–180, Nov 2019.
- [26] M. Dauner and U. Sauer. Stoichiometric growth model for riboflavin-producing *Bacillus subtilis*. *Biotechnology and Bioengineering*, 76(2), Sept 2001.
- [27] G. Diekert and R. Maria. Nickel Requirement of *Acetobacterium woodii*. *Journal of Bacteriology*, 151(2):1043–1045, Aug 1982.
- [28] G. Diekert, E. Schrader, and W. Harder. Energetics of CO formation and CO oxidation in cell suspensions of *Acetobacterium woodii*. *Archives of Microbiology*, 144:386–392, Jan 1986.
- [29] M. Diender, A. J. M. Stams, and D. Z. Sousa. Pathways and bioenergetics of anaerobic carbon monoxide fermentation. *Frontiers in Microbiology*, 6:1275, Nov 2015.
- [30] L. Dijkhuizen, T. A. Hansen, and W. Harder. Methanol, a potential feedstock for biotechnological processes. *Trends in Biotechnology*, 3(10):262–267, Oct 1985.
- [31] J. Dönig and V. Müller. Alanine, a novel growth substrate for the acetogenic bacterium *Acetobacterium woodii*. *Applied and Environmental Microbiology*, 84(23):e02023–18, Dec 2018.
- [32] H. L. Drake, editor. *Diversity, ecology and isolation of acetogenic bacteria*, pages 197–235. Springer, July 1994. ISBN 1461357160.
- [33] H. L. Drake, A. S Gößner, and S. L. Daniel. Old acetogens, new light. *Ann. N.Y. Acad. Sci.*, 1125:100–128, Jan 2008.
- [34] D. H. Elbourne, L., G. Tetu, S., A. Hassan, K., and T. Paulsen, I. TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Research*, 45:320–324, Nov 2016.
- [35] S. Evangelos and N. D. Price. Genome-scale modeling for metabolic engineering. *J. Ind. Microbiol. & Biotechnol.*, 42(3):327–38, Mar 2015.

- [36] A. M. Feist and B. Ø. Palsson. The biomass objective function. *Curr. Opin. Microbiol.*, 13:344–349, Apr 2010.
- [37] D. A. Fell and J. R. Small. Fat synthesis in adipose tissue. *Biochem. J.*, 238:781–786, Sept 1986.
- [38] D. A. Fell, M. G. Poolman, and A. Gevorgyan. Building and analysing genome-scale metabolic models. *Biochem. Soc. Trans.*, 38:1197–1201, 2010.
- [39] D.A. Fell. *Understanding the Control of Metabolism*. Portland Press, London, 1997.
- [40] L. Fontaine, I. Meynial-Salles, L. Girbal, X. Yang, C. Croux, and P. Soucaille. Molecular characterization and transcriptional analysis of adhE2, the gene encoding the NADH-dependent aldehyde/alcohol dehydrogenase responsible for butanol production in alcohologenic cultures of *Clostridium acetobutylicum* ATCC 824. *Journal of Bacteriology*, 184(3):821–30, Feb 2002.
- [41] M. Fritz, L. Klyszejko. A., N. Morgner, J. Vonck, B. Brutschy, D. J. Muller, T. Meier, and V. Müller. An intermediate step in the evolution of ATPases - a hybrid F₀-V₀ rotor in a bacterial Na⁺ F₁F₀ ATP synthase. *FEBS*, 275:1999–2007, Feb 2008.
- [42] J. M. Fröstl, C. Seifritz, and H. L. Drake. Effect of nitrate on the autotrophic metabolism of the acetogens *Clostridium thermoautotrophicum* and *Clostridium thermoaceticum*. *Journal of Bacteriology*, 178(15):4597–4603, Aug 1996.
- [43] G. Fuchs. CO₂ fixation in acetogenic bacteria: variations on a theme. *FEMS Microbiology Letters*, 39(3):181–213, Aug 1986.
- [44] L. V. Gent-Ruijters, M., W. DeVries, and H. Southamer, A. Influence of nitrate on fermentation pattern, molar growth yields and synthesis of cytochrome B in *Propionibacterium Pentosaceum*. *J. Gen. Microbiol.*, 88:36–48, May 1975.
- [45] B. R. S. Genthner and M. P. Bryant. Additional characteristics of one-carbon-compound utilization by *Eubacterium limosum* and *Acetobacterium woodii*. *Applied and Environmental Microbiology*, 53(3):471–476, Mar 1987.
- [46] A. Gevorgyan, M. G. Poolman, and D. A. Fell. Detection of stoichiometric inconsistencies in biomolecular models. *Bioinformatics*, 24(19):2245–2251, October 2008. ISSN 1460-2059. URL <http://view.ncbi.nlm.nih.gov/pubmed/18697772>.
- [47] A. A. Godley, P. E. Linnet, and J. P. Robinson. The effect of carbon dioxide on the growth kinetics of fructose-limited chemostat cultures of *Acetobacterium woodii* DSM1030. *Archives of Microbiology*, 154:5–11, Feb 1990.
- [48] HM Government. Industrial Strategy: Growing the bioeconomy, a national bioeconomy strategy to 2030, Dec 2018.

- [49] S. Gudmundsson and I. Thiele. Computationally efficient flux variability analysis. *BMC Bioinformatics*, 11(489):<https://doi.org/10.1186/1471-2105-11-489>, Sept 2010.
- [50] M. Hahnke, S., P. Moosmann, J. Erb, T., and M. Strous. An improved medium for the anaerobic growth of *Paracoccus denitrificans* Pd1222. *Front. Microbiol.*, 5(18):5–18, Jan 2014.
- [51] H. B. Hartman, D. A. Fell, S. Rossell, P. Ruhdal Jensen, M. J. Woodward L. Thorn-dahl, L. Jelsbak, J. Elmerdahl Olsen, A. Raghunathan, S. Daefer, and M. G. Poolman. Identification of potential drug targets in *Salmonella enterica* sv. Typhimurium using metabolic modelling and experimental validation. *Microbiology*, 160:1252–1266, 2014.
- [52] A. B. Hazra, A. W. Han, A. P. Mehta, K. C. Mok, V. Osadchiy, T. P. Begley, and M. E. Taga. Anaerobic biosynthesis of the lower ligand of vitamin B12. *PNAS*, 112(34):10792–10797, Aug 2015.
- [53] R. Heinrich and S. Shuster. *The Regulation of Cellular Systems*. Chapman and Hall, New York, 1996.
- [54] L. Heirendt. Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0. *Nature Protocols*, 14(3):639–702, Mar 2019.
- [55] R. Heise, V. Müller, and G. Gottschalk. Sodium dependence of acetate formation by the acetogenic bacterium *Acetobacterium woodii*. *Journal of Bacteriology*, 171(10):5473–5478, Oct 1989.
- [56] V. Hess, K. Schuchmann, and V. Müller. The ferredoxin:NAD⁺ oxidoreductase (Rnf) from the acetogen *Acetobacterium woodii* requires Na⁺ and is reversibly coupled to the membrane potential. *J. Biol. Chem*, 288(44):31496–31502, Jan 2013.
- [57] V. Hess, O. Oyrik, D. Trifunović, and V. Müller. 2,3-Butanediol metabolism in the acetogen *Acetobacterium woodii*. *Applied and Environmental Microbiology*, 81(14):4711–4719, Jul 2015.
- [58] S. Hoffmeister, M. Gerdorf, F. R. Bengelsdorf, S. Linder, S. Flüchter, H. Öztürk, W. Blümke, A. May, R. J. Fischer, H. Bahl, and P. Dürre. Acetone production with metabolically engineered strains of *Acetobacterium woodii*. *Metab. Eng.*, 36:37–47, Jul 2016.
- [59] H.-G. Holzhütter. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur. J. Biochem.*, 271:2905–2922, 2004.
- [60] F. Imkamp, E. Biegel, E. Jayamani, W. Buckel, and V. Müller. Dissection of the caffeate respiratory chain in the acetogen *Acetobacterium woodii*: identification of an Rnf-type NADH dehydrogenase as a potential coupling site. *Journal of Bacteriology*, 189(22):8145–8153, Nov 2007.

- [61] D. T. Jones and D. R. Woods. Acetone-butanol fermentation revisited. *Microbiology Reviews*, 50(4):484–524, Dec 1986.
- [62] O. Kandler and S. Schoberth. Murein structure of *Acetobacterium woodii*. *Archives of Microbiology*, 120:181–183, Oct 1978.
- [63] C. Kantzow and Weuster-Botz. D. Effects of hydrogen partial pressure on autotrophic growth and product formation of *Acetobacterium woodii*. *Bioprocess Biosys Eng*, 39(8):1325–1330, Aug 2016.
- [64] S. C Karekar, K. Srinivas, and B. K. Ahring. Kinetic study on heterotrophic growth of *Acetobacterium woodii* on lignocellulosic substrate. *Fermentation*, 5(17):1–12, Jan 2019.
- [65] P. D. Karp, M. Latendresse, and R. Caspi. The Pathway Tools Pathway Prediction Algorithm. *Stand Genomic Sci*, 5:doi: 10.4056/sigs.1794338, 2011.
- [66] P. D. Karp, R. Billington, R. Caspi, C. A. Fulcher, M. Latendresse, A. Kothari, I. M. Keseler, M. Krummenacker, P. E. Midford, Q. Ong, W. K. Ong, S. Paley, and P. Subhraveti. The BioCyc collection of microbial genomes and metabolic pathways. *Briefings in Bioinformatics*, 20(4):1085–1093, Jul 2017.
- [67] R. Kerby, W. Niemczura, and J. G. Zeikus. Single carbon catabolism in Acetogens: analysis of carbon flow in *Acetobacterium woodii* and *Butyribacterium methylotrophicum* by fermentation and ^{13}C nuclear magnetic resonance measurement. *Journal of Bacteriology*, 155(3):1208–1218, June 1983.
- [68] S. Klamt. Generalized concept of minimal cut sets in biochemical networks. *BioSystems*, 83:233–247, 2006.
- [69] S. Klamt and E.D. Gilles. Minimal cut sets in biochemical reaction networks. *Bioinformatics*, 20(2):226–234, 2004.
- [70] T. Kotani, T. Yamamoto, H. Yurimoto, Y. Sakai, and N. Kato. Propane monooxygenase and NAD^+ -dependent secondary alcohol dehydrogenase in propane metabolism by *Gordonia* sp. strain TY-5. *Journal of Bacteriology*, 185(24):7120–8, Dec 2003.
- [71] F. Kremp, A. Poehlein, R. Daniel, and V. Müller. Methanol metabolism in the acetogenic bacterium *Acetobacterium woodii*. *Environmental Microbiology*, 20(12):4369–4384, Jul 2018.
- [72] W. Lau and E. S. Sattely. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science*, 349(6253):1224–8, Sep 2015.

- [73] M. Lechtenfeld, J. Heine, J. Sameith, F. Kremp, and V. Müller. Glycine betaine metabolism in the acetogenic bacterium *Acetobacterium woodii*. *Environmental Microbiology*, 20(12):4512–4525, Oct 2018.
- [74] S. Lee, C. Phalakornkule, M. M. Domach, and I. E. Grossmann. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Computers and Chemical Engineering*, 24:711–716, 2000.
- [75] Y. Liu, R. H. White, and W. B. Whitman. Methanococci use the diaminopimelate aminotransferase (DapL) pathway for lysine biosynthesis. *Journal of Bacteriology*, 192(13):3304–10, Jul 2010.
- [76] Lanzatech New Zealand LTD. [Patent] Genetically engineered bacterium comprising energy-generating fermentation pathway, Aug 2017. US 9,738,875 B2.
- [77] ZuvaSyntha LTD. [Patent] Modified Enzyme, Mar 2017. WO 2017/168161 A1.
- [78] K. Ma, S. Siemon, and G. Diekert. Carbon monoxide metabolism in cell suspensions of *Peptostreptococcus productus* strain Marburg. *FEMS Microbiology Letters*, 43:367–371, May 1987.
- [79] D. Matthies, W. Zhou, L. Klyszejko, A., C. Anselmi, O. Yildiz, K. Brandt, V. Müller, and D. Faraldo-Gomez, J. High-resolution structure and mechanism of an F/V-hybrid rotor ring in a sodium-coupled ATP synthase. *Nature Communications*, 5:5286, Nov 2015.
- [80] A. J. McCoy, N. E. Adams, A. O. Hudson, C. Gilvarg, T. Leustek, and A. T. Maurelli. L,L-diaminopimelate aminotransferase, a trans-kingdom enzyme shared by Chlamydia and plants for synthesis of diaminopimelate/lysine. . *PNAS*, 103(47):17909–14, Nov 2006.
- [81] C. B. Milne, P. J. Kim, J. A. Eddy, and N. D. Price. Accomplishments in genome-scale *in silico* modelling for industrial and medical biotechnology. *Biochem. J*, 4(12): 1653–70, Dec 2009.
- [82] S. J. Moore, A. D. Lawrence, R. Biedendieck, E. Deery, S. Frank, M. J. Howard, S. E. J. Rigby, and M. J. Warren. Elucidation of the anaerobic pathway for the corrin component of cobalamin (vitamin B12). *PNAS*, 110(37):14906–14911, Sep 2013.
- [83] H. Nagarajan, M. Sahin, J. Nogales, H. Latif, D. Lovley, A. Ebrahim, and K. Zengler. Characterising acetogenic metabolism using a genome-scale metabolic reconstruction of *Clostridium ljungdahlii*. *Microbial Cell Factories*, 12(118), Jan 2013.
- [84] F.C. Neidhardt, J.L. Ingraham, and M Schaechter. *Physiology of the bacterial cell. A molecular approach*. Sunderland: Sinauer Associates, 1990.

- [85] H. Nonaka, G. Keresztes, Y. Shinoda, Y. Ikenaga, M. Abe, K. Naito, K. Inatomi, K. Furukawa, M. Inui, and H. Yukawa. Complete genome sequence of the dehalorespiring bacterium *Desulfitobacterium hafniense* Y51 and comparison with *Dehalococcoides ethenogenes*. *Journal of Bacteriology*, 188(6):2262–2274, Mar 2006.
- [86] E. J. O’Brien, J. M. Monk, and B. Ø. Palsson. Using Genome-scale Models to Predict Biological Capabilities. *Cell*, 161(5):971–987, May 2015.
- [87] P. Pawel. *Characterisation of Agr quorum sensing in Clostridium autoethanogenum*. PhD thesis, University of Nottingham, 2018.
- [88] V. Peters., P. H. Janssen, and R. Conrad. Efficiency of hydrogen utilization during unitrophic and mixotrophic growth of *Acetobacterium woodii* on hydrogen and lactate in the chemostat. *FEMS Microbiology Ecology*, 26:317–324, Mar 1998.
- [89] T. Pfeiffer, I. Sanchez-Valdenebro, J.C. Nuno, F. Montero, and S. Schuster. META-TOOL: for studying metabolic networks. *Bioinformatics*, 15(3):251–257, 1999.
- [90] S. J. Pirt. *Principles of Microbe and Cell Cultivation*. Blackwell Scientific, 1975. ISBN 9780632081509.
- [91] S. M. Pitson, G. L. Mendz, S. Srinivasan, and S. L. Hazell. The tricarboxylic acid cycle of *Helicobacter pylori*. *Eur. J. Biochem.*, 260(1):258–67, Feb 1999.
- [92] A. Poehlein, S. Schmidt, A. Kaster, M. Goenrich, J. Vollmers, A. Thürmer, J. Bertsch, K. Schuchmann, B. Voigt, M. Hecker, R. Daniel, R. K. Thauer, G. Gottschalk, and V. Müller. An ancient pathway combining carbon dioxide fixation with the generation and utilizing of a sodium ion gradient for ATP synthesis. *PLoS ONE*, 7(3):e33439, Mar 2012.
- [93] W. Polcyn and R. Lucinski. Aerobic and anaerobic nitrate and nitrite reduction in free living cells of *Bradyrhizobium* sp. *FEMS Microbiology Letters*, 226(2):331–337, Sept 2003.
- [94] M. G. Poolman. ScrumPy - metabolic modelling with Python. *IEE Proceedings Systems Biology*, 153(5):375–378, 2006.
- [95] M. G. Poolman. ScrumPy - metabolic modelling with Python. *IEE Proceedings Systems Biology*, 153(5):375–378, 2006.
- [96] M. G. Poolman, B. K. Bonde, A. Gevorgyan, H. H. Patel, and D. A. Fell. Challenges to be faced in the reconstruction of metabolic networks from public databases. *IEE Proceedings Systems Biology*, 153(5):379–384, 2006.

- [97] M. G. Poolman, L. Miguët, L. J. Sweetlove, and D. A. Fell. A genome-scale metabolic model of *Arabidopsis thaliana* and some of its properties. *Plant Physiol.*, 151(3): 1570–1581, 2009.
- [98] K. D. Pruitt, T. Tatusova, and R. Maglott. D. NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 33:D501–D504, Jan 2005.
- [99] S. W. Ragsdale and L. G. Ljungdahl. Hydrogenase from *Acetobacterium woodii*. *Archives of Microbiology*, 139:361–365, Jan 1984.
- [100] S. W. Ragsdale and E. Pierce. Acetogenesis and the Wood-Ljungdahl pathway of CO₂ fixation. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1784(12): 1873–1898, Dec 2008.
- [101] C. Ratledge and B. Kristiansen. *Basic Biotechnology*. Cambridge University Press, 2001.
- [102] F. L. Rodkey, J. D. O’Neal, H. A. Collison, and D. E. Uddin. Relative affinity of hemoglobin S and hemoglobin A for carbon monoxide and oxygen. *Clinical Chemistry*, 20:83–84, Jan 1974.
- [103] U. Sauer, V. Hatzimanikatis, H. Hohmann, M. Manneberg, A. P. G. M. van Loon, and J. E. Bailey. Physiology and metabolic fluxes of wild-type and riboflavin-producing *Bacillus subtilis*. *Applied and Environmental Microbiology*, 62(10):3687–3696, Oct 1996.
- [104] J. M. Savinell and B. Ø. Palsson. Network analysis of intermediary metabolism using linear optimisation. I. Development of mathematical formalism. *J. Theor. Biol.*, 154(4):421–54, Feb 1992.
- [105] B. Schink. *Clostridium magnum* sp. nov., a non-autotrophic homoacetogenic bacterium. *Archives of Microbiology*, 137:250–255, Nov 1983.
- [106] B. Schink and M. Stieb. Fermentative degradation of polyethylene glycol by a strictly anaerobic, gram-negative, nonsporeforming bacterium, *Pelobacter venetianus* sp. nov. *Applied and Environmental Microbiology*, 45(6):1905–1913, Jun 1983.
- [107] L. Schlabach and A. Züttel. Hydrogen-storage materials for mobile applications. *Nature*, 414:353–358, Nov 2001.
- [108] S. Schmidt, E. Biegel, and V. Müller. The ins and outs of Na⁺ bioenergetics in *Acetobacterium woodii*. *Biochimica et Biophysica Acta*, 1787:691–696, Jan 2009.

- [109] M. M. Schoelmerich, A. Katsyv, W. Sung, V. Mijic, A. Wiechmann, P. Kottenhahn, J. Baker, N. P. Minton, and V. Müller. Regulation of lactate metabolism in the acetogenic bacterium *Acetobacterium woodii*. *Environmental Microbiology*, 20(12): 4587–4595, Sept 2018.
- [110] L. Schrage. *Optimization Modeling with LINDO*. Duxbury Press, 5 edition, Feb 1997. ISBN 0534348572.
- [111] L. Schrage. *Optimization Modeling with LINDO*. Duxbury Press, 5 edition, Chapter 19, 432–459, Feb 1997. ISBN 0534348572.
- [112] K. Schuchmann and V. Müller. A bacterial electron-bifurcating hydrogenase. *Journal of Biological Chemistry*, 287(37):31165–31171, Sept 2012.
- [113] K. Schuchmann and V. Müller. Direct and Reversible Hydrogenation of CO₂ to Formate by a Bacterial Carbon Dioxide Reductase. *Science*, 342:1382, Dec 2013.
- [114] K. Schuchmann and V. Müller. Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nature Reviews Microbiology*, 12:809–821, Dec 2014.
- [115] K. Schuchmann and V. Müller. Energetics and applications of heterotrophy in acetogenic bacteria. *Applied and Environmental Microbiology*, 82(14):4056, Jul 2016.
- [116] K. Schuchmann, S. Schmidt, A. M. Lopez, C. Kaberline, M. Kuhns, W. Lorenzen, H. B. Bode, F. Joos, and V. Müller. Nonacetogenic growth of the acetogen *Acetobacterium woodii* on 1,2-propandiol. *Journal of Bacteriology*, 197(2):382–391, Jan 2015.
- [117] S. Schuster and R. Schuster. Detecting strictly detailed balanced subnetworks in open chemical-reaction networks. *J.Math.Chem.*, 6(1):17–40, 1991.
- [118] S. Schuster, T. Dandekar, and D.A. Fell. Detection of Elementary Flux Modes in Biochemical Networks: a Promising Tool for Pathway Analysis and Metabolic Engineering. *Trends. Biotech.*, 17(2):53–60, 1999.
- [119] S. Schuster, D.A. Fell, and T. Dandekar. A General Definition of Metabolic Pathways Useful for Systematic Organization and Analysis of Complex Metabolic Networks. *Nature Biotech.*, 18:326–332, 2000.
- [120] S. Schuster, T. Pfeiffer, and D. A. Fell. Is maximization of molar yield in metabolic networks favoured by evolution? *J. Theor. Biol.*, 252:497–504, 2008.
- [121] C. Seifritz, S. L. Daniel, A. Gößner, and H. L. Drake. Nitrate as a preferred electron sink for the acetogen *Clostridium thermoaceticum*. *Journal of Bacteriology*, 175(24): 8008–8013, Dec 1993.

- [122] C. Seifritz, J. M. Fröstl, H. L. Drake, and S. L. Daniel. Glycolate as a metabolic substrate for the acetogen *Moorella thermoacetica*. *FEMS Microbiology Letters*, 170: 399–405, Jan 1999.
- [123] D. Shelver, R. L. Kerby, Y. He, and G. P. Roberts. Carbon monoxide-induced activation of gene expression in *Rhodospirillum rubrum* requires the product *cooA*, a member of the cyclic AMP receptor protein family of transcriptional regulators. *Journal of Bacteriology*, 177:2157–2163, Jan 1995.
- [124] J. S. Shieh and W. B. Whitman. Pathway of acetate assimilation in autotrophic and heterotrophic methanocci. *Journal of Bacteriology*, 169(11):5327–9, Nov 1987.
- [125] F. L. Sousa, T. Thiergart, G. Landan, S. Nelson-Sathi, I. A. C. Pereira, J. F. Allen, N. Lane, and W. F. Martin. Early bioenergetic evolution. *Phil. Trans. R. Soc. Biol.*, 368:20130088, Jan 2013.
- [126] G. N. Stephanopoulos, A. A. Aristidou, and J. Nielsen. *Metabolic Engineering: Principles and Methodologies*. Academic Press, Oct 1998. ISBN 9780080536286.
- [127] E. Stupperich and R. Konle. Corrinoid-dependent methyl transfer reactions are involved in methanol and 3,4-dimethoxybenzoate metabolism by *Sporomusa ovata*. *Appl Environ Microbiol.*, 59(9):3110–3116, Sept 1993.
- [128] E. Stupperich, H. J. Eisinger, and B. Kräutler. Diversity of corrinoids in acetogenic bacteria. P-cresolcycobamide from *Sporomusa ovata*, 5-methoxy-6-methylbenzimidazolcycobamide from *Clostridium formicoaceticum* and vitamin B12 from *Acetobacterium woodii*. *Eur. J. Biochem.*, 172(2):459–64, Mar 1988.
- [129] R. S. Tanner, R. S. Wolfe, and L. G. Ljungdahl. Tetrahydrofolate enzyme levels in *Acetobacterium woodii* and their implication in the synthesis of acetate from CO₂. *Journal of Bacteriology*, 134(2):668–670, May 1978.
- [130] S. M. Techtmann, A. S. Colman, and F. T. Robb. ‘That which does not kill us only makes us stronger’: the role of carbon monoxide in thermophilic microbial consortia. *Environmental Microbiology*, 11(5):1027–1037, Nov 2009.
- [131] R. K. Thauer, K. Jungermann, H. Henninger, J. Wenning, and K. Decker. The energy metabolism of *Clostridium kluyveri*. *Eur. Jour. Biochem.*, 4(2):173–80, Apr 1968.
- [132] I. Thiele and B. Ø. Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc.*, 5:93–121, 2010.
- [133] I. Thiele, N. Swainston, R. Flemindg, and Al Et. A community-driven global reconstruction of human metabolism. *Nature Biotech.*, 31:419–425, Dec 2013.

- [134] G. H. Thomas, J. Zucker, S. J. Macdonald, S. J. Sorokin, I. Goryanin, and A. E. Douglas. A fragile metabolic network adapted for cooperation in the symbiotic bacterium *Buchnera aphidicola*. *BMC Systems Biology*, 21:3–24, Feb 2009.
- [135] D. Trifunović, K. Schuchmann, and V. Müller. Ethylene glycol metabolism in the acetogen *Acetobacterium woodii*. *Journal of Bacteriology*, 198(7):1058–1065, Apr 2016.
- [136] C.T. Trinh, P. Unrean, and F. Sreenc. Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. *Appl Environ Microbiol.*, 74:3634–3643, 2008.
- [137] C.T. Trinh, A. Wlaschin, and F. Sreenc. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl Environ Microbiol.*, 81:813–826, 2009.
- [138] A. Tschech and N. Pfennig. Growth yield increase linked to caffeate reduction in *Acetobacterium woodii*. *Archives of Microbiology*, 137:163–167, Oct 1984.
- [139] H. C. Tseng, C. H. Martin, D. R. Nielsen, and K. L. Prather. Metabolic Engineering of *Escherichia coli* for enhanced production of (R)- and (S)-3-hydroxybutyrate. *Applied and Environmental Microbiology*, 75(10):3137–45, May 2009.
- [140] Department of Economic and Social Affairs Population Division United Nations. World Population Prospects (2019): Data Booklet(ST/ESA/SER.A.424), 2019.
- [141] A. Varma and B. Ø Palsson. Metabolic Capabilities of *Escherichia coli*: II. Optimal Growth Patterns. *J. Theor. Biol.*, 165:503–522, 1993.
- [142] A. Varma and B. Ø. Palsson. Metabolic Flux Balancing: Basic concepts, scientific and practical use. *Bio/Technology*, 12:994–998, 1994.
- [143] A. Varma and B. Ø. Palsson. Stoichiometric Flux Balance Models Quantitatively Predict Growth and Metabolic By-Product Secretion in Wild-Type *Escherichia coli* W3110. *Applied and Environmental Microbiology*, 60(10):3724–3731, Oct 1994.
- [144] M. R. Watson. A discrete model of bacterial metabolism. *Comp. Appl. Biosci.*, 2(1): 23–27, 1986.
- [145] D. S. Weaver, I. M. Keseler, A. Mackie, I. T. Paulsen, and P. D. Karp. A genome-scale metabolic flux model of *Escherichia coli* K-12 derived from the EcoCyc database. *BMC Systems Biology*, 8(1):79, Jun 2014. ISSN 1752-0509. doi: 10.1186/1752-0509-8-79. URL <https://doi.org/10.1186/1752-0509-8-79>.
- [146] M. C. Weghoff, J. Bertsch, and V. Müller. A novel mode of lactate metabolism in strictly anaerobic bacteria. *Environmental Microbiology*, 17(3):670–677, Apr 2015.

- [147] T. (Ed) Werpy and G. (Ed) Petersen. Top value added chemicals from Biomass (results of screening for potential candidates from sugars and synthesis gas, Vol 1). *US Department of Energy*, jan 2004.
- [148] A. Wiechmann, S. Ciurus, and V. Müller. Hydrogen cycling is essential for growth of the acetogen *Acetobacterium woodii*, *C1Net Conference 4*, Jan 2019.
- [149] M. Wu, Q. Ren, A. S. Durkin, S. C. Daugherty, L. M. Brinkac, R. J. Dodson, R. Madupu, S A. Sullivan, J. F. Kolonay, W. C. Nelson, L. J. Tallon, K. M. Jones, L. E. Ulrich, J. M. Gonzalez, I. B. Zhulin, F. T. Robb, and J. A. Eisen. Life in hot carbon monoxide: the complete genome sequence of *Carboxydotherrnus hydrogenoformans* Z2901. *PLoS Genet*, 1(5):e65, Nov 2005.
- [150] H. Youn, R. L. Kerby, M. Conrad, and G. P. Roberts. Functionally critical elements of CoxA-related CO sensors. *Journal of Bacteriology*, 186:1320–1329, Jan 2004.
- [151] L. Zelcbuch, S. N. Lindner, Y. Zegman, Slutskin I. Vainberg, N. Antonovsky, S. Gleizer, and R. Milo. Pyruvate formate-lyase enables efficient growth of *Escherichia coli* on acetate and formate. *Biochemistry*, 55(17):2423–6, May 2016.
- [152] G. Zubay. *Biochemistry*. Wm. C. Brown, Oxford, 3rd edition, 1993.

Appendix A

Metabolic Model of Woods-Ljungdahl Pathway

Model file used for elementary modes analysis of gas mixtures in WLP presented in chapter 3.

```
Structural()

#
##
### transporters , exchange

H2_sf_tx:
  1 x_h2 <= 1 "HYDROGEN-MOLECULE"
  ~

CO2_sf_tx:
  1 x_co2 <= 1 "CARBON-DIOXIDE"
  ~

CO_s_tx:
  1 x_co <= 1 "CARBON-MONOXIDE"
  ~

H2O_tx:
  1 x_water <= 1 "WATER"
  ~

PROTON_tx:
  1 x_proton <= 1 "PROTON"
  ~

#
##
### transporters , out

ACETATE_f_tx:
  1 x_ACETATE <= 1 "ACET"
```

```

~

ETHANOL_sf_tx:
  1 x_ETHANOL <- 1 "ETOH"
~

#
##
### internal reactions

rnfComplex_etc:
  2 x_Na_o + 1 "NADH" + 1 "Oxidized-2-4Fe-4S-Ferredoxins" <> 1 "NAD" + 2 "Na_i
    " + 1 "PROTON" + 1 "Reduced-2-4Fe-4S-Ferredoxins"
~

elecBifurc_etc:
  1 "NADH" + 3 "PROTON" + 1 "Reduced-2-4Fe-4S-Ferredoxins" <> 2 "HYDROGEN-
    MOLECULE" + 1 "NAD" + 1 "Oxidized-2-4Fe-4S-Ferredoxins"
~

"ACETATEKIN-RXN":
  1 "ACETYL-P" + 1 "ADP" <> 1 "ATP" + 1 "ACET"
~

"FORMATETHFLIG-RXN":
  1 "THF" + 1 "FORMATE" + 1 "ATP" <> 1 "10-FORMYL-THF" + 1 "ADP" + 1 "Pi"
~

"PHOSACETYLTRANS-RXN":
  1 "ACETYL-COA" + 1 "Pi" <> 1 "ACETYL-P" + 1 "CO-A"
~

"CARBON-MONOXIDE-DEHYDROGENASE-RXN":
  1 "CARBON-DIOXIDE" + 2 "PROTON" + 1 "Reduced-2-4Fe-4S-Ferredoxins" <> 1 "
    Oxidized-2-4Fe-4S-Ferredoxins" + 1 "CARBON-MONOXIDE" + 1 "WATER"
~

"1.5.1.15-RXN":
  1 "5-10-METHENYL-THF" + 1 "NADH" <> 1 "NAD" + 1 "METHYLENE-THF"
~

"ACETYLSYNCLTH-RXN":
  1 "Methylated-corrinoid-Fe-S-Proteins" + 1 "CARBON-MONOXIDE" + 1 "CO-A" -> 1
    "ACETYL-COA" + 1 "PROTON" + 1 "Corrinoid-Fe-S-proteins"
~

"ADENOSINETRIPHOSPHATASE-RXN":
  1 "WATER" + 1 "ATP" -> 1 "PROTON" + 1 "ADP" + 1 "Pi" + 1 x_ATPwork
~

```

```

atp-synth-etc:
  33/10 x_Na_o + 1 "ADP" + 1 "Pi" + 1 "PROTON" <> 33/10 "Na_i" + 1 "ATP" + 1 "
  WATER"
  ~

"METHENYLTHFCYCLOHYDRO-RXN":
  1 "10-FORMYL-THF" + 1 "PROTON" <> 1 "5-10-METHENYL-THF" + "WATER"
  ~

"ACETALD-DEHYDROG-RXN":
  1 "CO-A" + 1 "NAD" + 1 "ACETALD" <> 1 "NADH" + 1 "ACETYL-COA" + "PROTON"
  ~

"1.5.1.20-RXN":
  1 "METHYLENE-THF" + 1 "NADH" + "PROTON" <> 1 "NAD" + 1 "5-METHYL-THF"
  ~

"ALCOHOL-DEHYDROG-RXN":
  1 "ETOH" + 1 "NAD" <> 1 "NADH" + 1 "ACETALD" + "PROTON"
  ~

"METHCOCLTH-RXN":
  1 "5-METHYL-THF" + 1 "Corrinoid-Fe-S-proteins" + "PROTON" -> 1 "Methylated-
  corrinoid-Fe-S-Proteins" + 1 "THF"
  ~

"FHLMULTI-RXN":
  1 "CARBON-DIOXIDE" + 1 "HYDROGEN-MOLECULE" <> "PROTON" + 1 "FORMATE"
  ~

##"1.2.1.2-RXN": #alternative reaction for "FHLMULTI-RXN"
## 1 "FORMATE" + 1 "Oxidized-2-4Fe-4S-Ferredoxins" <> 1 "CARBON-DIOXIDE" + 1
  "Reduced-2-4Fe-4S-Ferredoxins"
## ~

```

WLP model extension file used for elementary modes analysis of fructose and formate as well as gas mixtures in WLP presented in chapter 3.

```

Structural()
#ANAGLYCOLYSIS-PWY
#
##
### transports in

FRUCTOSE_s-tx:
  1 x-fruc <> 1 "CPD-15382"
  ~

```

```

FORMATE_s_tx:
  1 x_formate < 1 "FORMATE"
  ~

#
##
### internal

"F16ALDOLASE-RXN":
  1 "FRUCTOSE-16-DIPHOSPHATE" < 1 "DIHYDROXY-ACETONE-PHOSPHATE" + 1 "GAP"
  ~

"TRIOSEPIISOMERIZATION-RXN":
  1 "GAP" < 1 "DIHYDROXY-ACETONE-PHOSPHATE"
  ~

"RXN-15513":
  1 "G3P" < 1 "2-PG"
  ~

"PHOSGLYPHOS-RXN":
  1 "ATP" + 1 "G3P" < 1 "DPG" + 1 "ADP"
  ~

"6PFRUCTPHOS-RXN":
  1 "ATP" + 1 "FRUCTOSE-6P" -> 1 "FRUCTOSE-16-DIPHOSPHATE" + 1 "ADP" + 1 "
    PROTON"
  ~

"2PGADEHYDRAT-RXN":
  1 "2-PG" < 1 "PHOSPHO-ENOL-PYRUVATE" + 1 "WATER"
  ~

"FRUCTOKINASE-RXN":
  1 "ADP" + 1 "FRUCTOSE-6P" + 1 "PROTON" <- 1 "ATP" + 1 "CPD-15382"
  ~

"PYRUFLAVREDUCT-RXN":
  1 "PYRUVATE" + 1 "Oxidized-2-4Fe-4S-Ferredoxins" + 1 "CO-A" < 1 "ACETYL-COA"
    + 1 "CARBON-DIOXIDE" + 1 "Reduced-2-4Fe-4S-Ferredoxins" + 1 "PROTON"
  ~

"PEPDEPHOS-RXN":
  1 "ADP" + 1 "PHOSPHO-ENOL-PYRUVATE" + 1 "PROTON" -> 1 "PYRUVATE" + 1 "ATP"
  ~

"GAPOXNPHOSPHN-RXN":
  1 "DPG" + 1 "NADH" + 1 "PROTON" < 1 "NAD" + 1 "GAP" + 1 "Pi"
  ~

```

Appendix B

Substrate Combinations: ATP Yields and Growth Phenotypes

Table 8.1: For all substrate pairs included in the GSM, an ATP yield is calculated for them as well as whether they support growth. Comparisons are made with reports in the scientific literature. No distinction is made between CO₂ and its carbonate forms. ‘PC’ is personal communication.

Substrate	Y _{ATP} ^{max} model	Y _{ATP} literature	Growth model	Supporting literature	Ref
MetOH, Caffeate	3.12	3	yes	yes	[60, 138]
MetOH, Coumarate	3.12	-	yes	yes	[60]
MetOH, Ferulate	3.12	-	yes	yes	[60]
Formaldehyde, Caffeate	2.52	-	yes	-	
Formaldehyde, Ferulate	2.52	-	yes	-	
Formaldehyde, Coumarate	2.52	-	yes	-	
Caffeate, Glc	1.27	-	yes	-	
Fruc, Ferulate	1.27	-	yes	yes	[60]
Glc, Ferulate	1.27	-	yes	-	
Caffeate, Fruc	1.27	-	yes	yes	[60]
Coumarate, Glc	1.27	-	yes	-	
Coumarate, Fruc	1.27	-	yes	yes	[60]
CO, Ferulate	1.21	-	yes	-	
CO, Caffeate	1.21	-	yes	-	
CO, Coumarate	1.21	-	yes	-	
Alanine, Caffeate	0.94	-	yes	-	
Alanine, Ferulate	0.94	-	yes	-	
Alanine, Coumarate	0.94	-	yes	-	
Caffeate, Form	0.91	-	yes	yes	[60]

Coumarate, Form	0.91	-	yes	yes	[60]
Ferulate, Form	0.91	-	yes	yes	[60]
Formaldehyde, Form	0.85	-	yes	-	
Formaldehyde, Ethylene-Glycol	0.85	-	yes	-	
Formaldehyde, 2,3-BDO	0.85	-	yes	-	
Formaldehyde, H ₂	0.85	-	yes	-	
Glycerate, Formaldehyde	0.85	-	yes	-	
Formaldehyde, HCO ₃ ⁻	0.85	-	yes	-	
Formaldehyde, EtOH	0.85	-	yes	-	
Formaldehyde, CO ₂	0.85	-	yes	-	
Formaldehyde, MetOH	0.85	-	yes	-	
Formaldehyde, 1,2-PDO	0.85	-	yes	-	
Formaldehyde, Lact	0.85	-	yes	-	
Formaldehyde, Acetaldehyde	0.85	-	yes	-	
Formaldehyde, Acetoin	0.85	-	yes	-	
Formaldehyde, Pyruv	0.85	-	yes	-	
Formaldehyde, Fruc	0.85	-	yes	-	
CO, Formaldehyde	0.85	-	yes	-	
Formaldehyde, Glc	0.85	-	yes	-	
Formaldehyde, Glycine-Betaine	0.85	-	yes	-	
Alanine, Formaldehyde	0.85	-	yes	-	
Formaldehyde	0.85	-	yes	yes	[32]
Glycerate, Coumarate	0.74	-	yes	-	
Pyruv, Caffeate	0.74	-	yes	-	
Lact, Coumarate	0.74	-	yes	-	
Lact, Ferulate	0.74	-	yes	-	
Glycerate, Ferulate	0.74	-	yes	-	
Lact, Caffeate	0.74	-	yes	-	
Pyruv, Coumarate	0.74	-	yes	-	
Glycerate, Caffeate	0.74	-	yes	-	
Pyruv, Ferulate	0.74	-	yes	-	
Acetaldehyde, Glc	0.72	-	yes	-	
Fruc, H ₂	0.72	-	yes	-	
Fruc	0.72	0.72, 0.83	yes	yes	[115, 138]
Glc, Form	0.72	-	yes	-	
Lact, Fruc	0.72	-	yes	yes	[109]
Alanine, Fruc	0.72	-	yes	-	
Glc, Ethylene-Glycol	0.72	-	yes	-	
CO, Fruc	0.72	-	yes	no	[8]

Glc, 1,2-PDO	0.72	-	yes	-	[115, 138]
Pyruv, Glc	0.72	-	yes	-	
Alanine, Glc	0.72	-	yes	-	
CO, Glc	0.72	-	yes	-	
Glycerate, Fruc	0.72	-	yes	-	
Glc, Glycine-Betaine	0.72	-	yes	-	
Fruc, Ethylene-Glycol	0.72	-	yes	-	
2,3-BDO, Fruc	0.72	-	yes	-	
Glc	0.72	0.72, 0.83	yes	yes	
Pyruv, Fruc	0.72	-	yes	-	
Glc, Fruc	0.72	-	yes	-	
Glc, EtOH	0.72	-	yes	-	
Fruc, Acetoin	0.72	-	yes	-	
Fruc, Glycine-Betaine	0.72	-	yes	-	
Acetaldehyde, Fruc	0.72	-	yes	-	
MetOH, Fruc	0.72	-	yes	-	[6, 3, 17, 64]
Glycerate, Glc	0.72	-	yes	-	
Fruc, Form	0.72	-	yes	-	
Glc, CO ₂	0.72	-	yes	yes	
Lact, Glc	0.72	-	yes	-	
Fruc, EtOH	0.72	-	yes	-	
Glc, H ₂	0.72	-	yes	-	
Fruc, CO ₂	0.72	-	yes	yes	
Glc, HCO ₃ ⁻	0.72	-	yes	yes	
Fruc, 1,2-PDO	0.72	-	yes	-	[6, 3, 17, 138]
MetOH, Glc	0.72	-	yes	-	
Fruc, HCO ₃ ⁻	0.72	-	yes	yes	
Glc, Acetoin	0.72	-	yes	-	
2,3-BDO, Glc	0.72	-	yes	-	
MetOH, Lact	0.55	-	yes	-	
MetOH, H ₂	0.55	-	yes	-	
Glycerate, MetOH	0.55	-	yes	-	
Acetaldehyde, MetOH	0.55	-	yes	-	
MetOH, EtOH	0.55	-	yes	-	PC
MetOH	0.55	-	yes	no	
MetOH, 1,2-PDO	0.55	-	yes	-	
Alanine, MetOH	0.55	-	yes	-	
MetOH, HCO ₃ ⁻	0.55	0.42, 0.75-1	yes	yes	
MetOH, CO ₂	0.55	0.42, 0.75-1,	yes	yes	[71, 138, 5]

		0.5			
MetOH, Form	0.55	0.65	yes	yes	[4]
MetOH, Glycine-Betaine	0.55	-	yes	-	
MetOH, Acetoin	0.55	-	yes	-	
Pyruv, MetOH	0.55	-	yes	-	
2,3-BDO, MetOH	0.55	-	yes	-	
CO, MetOH	0.55	-	yes	no	[8]
MetOH, Ethylene-Glycol	0.55	-	yes	-	
Coumarate, Glycine-Betaine	0.52	-	yes	-	
Glycine-Betaine, Ferulate	0.52	-	yes	-	
Caffeate, Glycine-Betaine	0.52	-	yes	-	
Pyruv, H ₂	0.48	-	yes	-	
Glycerate, H ₂	0.48	-	yes	-	
Glycerate, Ethylene-Glycol	0.46	-	yes	-	
Pyruv, Ethylene-Glycol	0.46	-	yes	-	
CO, Pyruv	0.46	-	yes	-	
Pyruv, Acetoin	0.46	-	yes	-	
Glycerate, Pyruv	0.46	-	yes	-	
Pyruv, Form	0.46	-	yes	-	
Glycerate, Glycine-Betaine	0.46	-	yes	-	
Pyruv, 1,2-PDO	0.46	-	yes	-	
Glycerate, Acetaldehyde	0.46	-	yes	-	
Glycerate, CO	0.46	-	yes	-	
Alanine, Pyruv	0.46	-	yes	-	
Pyruv	0.46	-	yes	yes	[3]
Glycerate, 2,3-BDO	0.46	-	yes	-	
Pyruv, Lact	0.46	-	yes	-	
2,3-BDO, Pyruv	0.46	-	yes	-	
Glycerate, 1,2-PDO	0.46	-	yes	-	
Glycerate, HCO ₃ ⁻	0.46	-	yes	-	
Acetaldehyde, Pyruv	0.46	-	yes	-	
Glycerate, CO ₂	0.46	-	yes	yes	[6]
Glycerate, EtOH	0.46	-	yes	-	
Glycerate, Alanine	0.46	-	yes	-	
Pyruv, CO ₂	0.46	-	yes	yes	[3, 105]
Glycerate, Lact	0.46	-	yes	-	
Pyruv, Glycine-Betaine	0.46	-	yes	-	
Glycerate, Acetoin	0.46	-	yes	-	
Pyruv, EtOH	0.46	-	yes	-	

Glycerate	0.46	-	yes	-	[3, 105]
Glycerate, Form	0.46	-	yes	-	
Pyruv, HCO_3^-	0.46	-	yes	yes	
CO , H_2	0.45	-	yes	-	
Alanine, Glycine-Betaine	0.38	-	yes	-	
Alanine, HCO_3^-	0.38	-	yes	-	
Alanine, H_2	0.38	-	yes	-	
Alanine, 2,3-BDO	0.38	-	yes	-	
Alanine, EtOH	0.38	-	yes	-	
Alanine, Form	0.38	-	yes	-	
CO , Acetaldehyde	0.38	-	yes	-	
CO , Acetoin	0.38	-	yes	-	
CO , Alanine	0.38	-	yes	-	
Alanine, Ethylene-Glycol	0.38	-	yes	-	[3, 31]
Alanine	0.38	0.38	yes	mixed	
Alanine, Lact	0.38	-	yes	-	
Alanine, 1,2-PDO	0.38	-	yes	-	
Alanine, Acetaldehyde	0.38	-	yes	-	
Alanine, Acetoin	0.38	-	yes	-	
Alanine, CO_2	0.38	-	yes	-	
CO , Ethylene-Glycol	0.38	-	yes	-	
CO , EtOH	0.38	-	yes	-	
CO , 1,2-PDO	0.38	-	yes	-	
CO , Lact	0.38	-	yes	-	
CO , Form	0.38	0.5	yes	yes	[8, 29]
CO , HCO_3^-	0.38	-	yes	-	
CO , 2,3-BDO	0.38	-	yes	-	
CO	0.38	-	yes	mixed	
CO , CO_2	0.38	-	yes	-	[67, 45, 8]
Acetaldehyde, Form	0.35	-	yes	-	
Ethylene-Glycol, Form	0.35	-	yes	-	
Acetaldehyde, HCO_3^-	0.35	-	yes	yes	
Acetaldehyde, Glycine-Betaine	0.35	-	yes	-	PC
Acetoin, Glycine-Betaine	0.35	-	yes	-	
Acetaldehyde, Lact	0.35	-	yes	-	
Lact, Acetoin	0.35	-	yes	-	
Ethylene-Glycol, Glycine-Betaine	0.35	-	yes	-	[57]
Acetoin, HCO_3^-	0.35	0.31	yes	yes	
Acetoin, Form	0.35	-	yes	-	

Acetaldehyde, CO ₂	0.35	-	yes	yes	PC
Lact, Ethylene-Glycol	0.35	-	yes	-	
Ethylene-Glycol, HCO ₃ ⁻	0.35	0.31	yes	yes	[106, 135]
Ethylene-Glycol, CO ₂	0.35	0.31	yes	yes	[106, 135]
Acetoin, CO ₂	0.35	0.31	yes	yes	[57]
2,3-BDO, HCO ₃ ⁻	0.27	0.24	yes	yes	[57]
2,3-BDO, Glycine-Betaine	0.27	-	yes	-	
2,3-BDO, Lact	0.27	-	yes	-	
2,3-BDO, Form	0.27	-	yes	-	
2,3-BDO, CO ₂	0.27	0.24	yes	yes	[57]
Lact, 1,2-PDO	0.23	-	yes	-	
1,2-PDO, HCO ₃ ⁻	0.23	-	yes	yes	[116]
1,2-PDO, CO ₂	0.23	-	yes	yes	[116]
1,2-PDO, Glycine-Betaine	0.23	-	yes	-	
1,2-PDO, Form	0.23	-	yes	-	
CO ₂ , EtOH	0.20	0.18	yes	yes	[17, 10]
HCO ₃ ⁻ , EtOH	0.20	0.18	yes	yes	[17, 10]
Lact, EtOH	0.20	-	yes	-	
Form, EtOH	0.20	-	yes	-	
Glycine-Betaine, EtOH	0.19	-	yes	-	
Lact, CO ₂	0.18	-	yes	yes	[6, 88]
Lact, Form	0.18	-	yes	-	
Lact	0.18	0.13	yes	yes	[88, 146]
Lact, H ₂	0.18	-	yes	yes	[88]
Lact, HCO ₃ ⁻	0.18	-	yes	-	
Form, H ₂	0.15	-	yes	-	
HCO ₃ ⁻ , H ₂	0.15	0.15, 0.25-0.5, 0.75-1	yes	yes	[6, 43, 114, 138]
CO ₂ , H ₂	0.15	0.15, 0.25-0.5, 0.75-1	yes	yes	[6, 43, 114, 138]
CO, Glycine-Betaine	0.14	-	yes	-	
Lact, Glycine-Betaine	0.12	-	yes	-	
Glycine-Betaine, Form	0.10	-	yes	-	
Glycine-Betaine, HCO ₃ ⁻	0.10	-	yes	yes	[73]
Glycine-Betaine, CO ₂	0.10	0.37-0.42	yes	yes	[73]
Glycine-Betaine	0.09	-	yes	no	[73]
Glycine-Betaine, H ₂	0.09	-	yes	-	

Form, HCO_3^-	0.08	0.75-1	yes	yes	[6, 3, 138]
CO_2 , Form	0.08	0.75-1	yes	yes	[6, 3, 138]
Form	0.08	0.5	yes	yes	[67, 8, 29]
Ferulate, EtOH	1.11	-	no	-	
Caffeate, EtOH	1.11	-	no	-	
Coumarate, EtOH	1.11	-	no	-	
2,3-BDO, Ferulate	0.95	-	no	-	
2,3-BDO, Coumarate	0.95	-	no	-	
2,3-BDO, Caffeate	0.95	-	no	-	
Caffeate, Ethylene-Glycol	0.80	-	no	-	
Coumarate, Acetoin	0.80	-	no	-	
Acetaldehyde, Ferulate	0.80	-	no	-	
Ethylene-Glycol, Ferulate	0.80	-	no	-	
Acetaldehyde, Coumarate	0.80	-	no	-	
Acetaldehyde, Caffeate	0.80	-	no	-	
Caffeate, Acetoin	0.80	-	no	-	
Acetoin, Ferulate	0.80	-	no	-	
Coumarate, Ethylene-Glycol	0.8	-	no	-	
Caffeate, 1,2-PDO	0.54	-	no	-	
Coumarate, 1,2-PDO	0.54	-	no	-	
1,2-PDO, Ferulate	0.54	-	no	-	
Acetoin, 1,2-PDO	0.35	-	no	-	
Acetaldehyde, 1,2-PDO	0.35	-	no	-	
Acetoin, H_2	0.35	-	no	-	
Ethylene-Glycol, H_2	0.35	-	no	-	
Acetaldehyde, Ethylene-Glycol	0.35	-	no	-	
Acetaldehyde, EtOH	0.35	-	no	-	
Acetoin, EtOH	0.35	-	no	-	
1,2-PDO, Ethylene-Glycol	0.35	-	no	-	
Acetaldehyde, H_2	0.35	-	no	-	
Ethylene-Glycol, EtOH	0.35	-	no	-	
Ethylene-Glycol	0.35	-	no	-	
Acetaldehyde	0.35	-	no	-	
2,3-BDO, Acetaldehyde	0.35	-	no	-	
2,3-BDO, Acetoin	0.35	-	no	-	
Acetoin, Ethylene-Glycol	0.35	-	no	-	
2,3-BDO, Ethylene-Glycol	0.35	-	no	-	
Acetaldehyde, Acetoin	0.35	-	no	-	

Acetoin	0.35	-	no	no	[57]
2,3-BDO, H ₂	0.27	-	no	-	
2,3-BDO	0.27	-	no	no	[57]
2,3-BDO, 1,2-PDO	0.27	-	no	-	
2,3-BDO, EtOH	0.27	-	no	-	
1,2-PDO, H ₂	0.23	-	no	-	
1,2-PDO	0.23	0.25	no	-	[116]*
1,2-PDO, EtOH	0.23	-	no	-	
EtOH	0.2	-	no	no	[10]
H ₂ , EtOH	0.2	-	no	-	
CO ₂ , HCO ₃ ⁻	0.0	-	no	-	
Ferulate, HCO ₃ ⁻	0.0	-	no	-	
Caffeate	0.0	-	no	-	
Coumarate	0.0	-	no	-	
Ferulate	0.0	-	no	-	
Caffeate, CO ₂	0.0	-	no	-	
Caffeate, HCO ₃ ⁻	0.0	-	no	-	
Ferulate, CO ₂	0.0	-	no	-	
Caffeate, Coumarate	0.0	-	no	-	
HCO ₃ ⁻	0.0	-	no	-	
H ₂	0.0	-	no	-	
Caffeate, Ferulate	0.0	-	no	-	
Coumarate, Ferulate	0.0	-	no	-	
Coumarate, HCO ₃ ⁻	0.0	-	no	-	
CO ₂	0.0	-	no	-	
Coumarate, CO ₂	0.0	-	no	-	
Coumarate, H ₂	NC	-	no	-	
Caffeate, H ₂	NC	-	no	-	
Ferulate, H ₂	NC	-	no	-	

Appendix C

Biomass Composition

Table 8.2: Macromolecular composition adopted from [26]

Component	Content%(w/w)
Protein	52.84
RNA	6.55
DNA	2.6
Lipids	7.6
Cell Wall	25.46
Other Compounds	4.94
Sum	100.00

Table 8.3: Breakdowns of macromolecular biomass composition

Component	Concentration (mmol · g-DW ⁻¹)	Reference
<i>Amino Acids</i>		
Methionine	0.085	[103]
Asparagine	0.174	[103]
Serine	0.215	[103]
Threonine	0.216	[103]
Proline	0.164	[103]
Lysine	0.252	[103]
Valine	0.308	[103]
Histidine	0.060	[103]
Aspartate	0.172	[103]
Isoleucine	0.237	[103]
Leucine	0.302	[103]
Alanine	0.718	[103]

Phenylalanine	0.119	[103]
Tyrosine	0.074	[103]
Arginine	0.122	[103]
Glycine	0.710	[103]
Glutamine	0.235	[103]
Cysteine	0.053	[103]
Glutamate	0.233	[103]
Tryptophan	0.017	[103]

Nucleic Acids

AMP	$2.12 \cdot 10^{-2}$	[26, 103]
ADP	$9.70 \cdot 10^{-3}$	[26, 103]
ATP	$3.25 \cdot 10^{-2}$	[26, 103]
dATP	$1.63 \cdot 10^{-2}$	[92]
CMP	$5.00 \cdot 10^{-3}$	[26, 103]
CTP	$3.58 \cdot 10^{-2}$	[26, 103]
CDP	$1.70 \cdot 10^{-3}$	[26, 103]
dCTP	$1.08 \cdot 10^{-2}$	[92]
GTP	$3.26 \cdot 10^{-2}$	[26, 103]
GMP	$2.20 \cdot 10^{-3}$	[26, 103]
GDP	$1.80 \cdot 10^{-3}$	[26, 103]
dGTP	$1.03 \cdot 10^{-2}$	[92]
TTP	$1.64 \cdot 10^{-2}$	[92]
UTP	$3.41 \cdot 10^{-2}$	[26, 103]

Cell Wall

Techoic Acid	$1.07 \cdot 10^{-2}$	[62]
Peptidoglycan	$1.30 \cdot 10^{-1}$	[62]

Lipids

Phosphatidylglycerol	$3.27 \cdot 10^{-2}$	[152]
Cardiolipin	$1.57 \cdot 10^{-3}$	[152]
Phosphatidylethanolamine	$6.50 \cdot 10^{-2}$	[152]

Other compounds

FAD	$4.0 \cdot 10^{-4}$	[26, 103]
NADP	$1.9 \cdot 10^{-3}$	[26, 103]
NADH	$4.0 \cdot 10^{-4}$	[26, 103]

NAD	$3.84 \cdot 10^{-2}$	[26, 103]
CoA	$5.0 \cdot 10^{-4}$	[26, 103]
Adenosylcobalamin	$3.0 \cdot 10^{-4}$	[129]