Flux Balance Techniques for Modelling Metabolic Networks and Comparison with Kinetic Models

Matthew Coleman

September 2017

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

Abstract

A variety of techniques used to model metabolic networks are examined, both kinetic (ODE) models and flux balance (FB) models. These models are applied to a case study network describing CO and CO_2 metabolism in *Clostridium autoethanogenum*, bacteria which can produce both ethanol and butanediol from a source of carbon monoxide. ODE and FB methods are also used to model a variety of simpler networks. By comparing the results from these simpler networks, the strengths and weaknesses of each examined method are highlighted, and ultimately, insight is gained into the conclusions that can be drawn from each model.

ODE models have commonly been used to model metabolism in both *in vivo* and *in vitro* contexts, allowing the dynamic behaviour of wildtype bacteria to be examined, as well as that of mutants. An ODE model is formed for the *C. autoethanogenum* network. By exploring a range of parameter schemes, the possible long timescale behaviours of the model are fully determined. The model is able to exhibit both steady states, and also states in which metabolite concentrations grow indefinitely in time. By considering the scalings of these concentrations in the long timescale, six different non-steady behaviours are categorised and one steady. For a small range of parameter schemes, the model is able to exhibit both steady and unsteady behaviours in the long timescale, depending on initial conditions.

FB methods are also applied to the same network. First flux balance analysis (FBA) is used to model the network in steady state. By imposing a range of constraints on the model, limits on levels of flux in the network that are required for a steady-state are found. In particular, boundaries on the ratio of inputs into the network are calculated, outside of which steady states cannot exist. Comparing the steady state regions predicted by FBA and our ODE model, it is found that the FBA model predicts a wider range of conditions leading to steady state. FBA is only able to observe a network in steady state, so an extension of FBA, known as dynamic flux balance analysis (dFBA), is used to examine non-steady-state behaviours. dFBA predicts similar long term non-steady behaviour to the ODE models, with states in which concentrations of some metabolites are able to grow indefinitely in time. These dFBA states do not precisely match those found by the ODE model, and states that cannot be observed in the ODE model are also found, suggesting that other ODE models for the same network could exhibit different long timescale behaviours. The examples considered clarify the strengths and weaknesses of each approach and the nature of insight into metabolic behaviour each provides.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Metabolism

Metabolism is, most simply, a set of chemical reactions taking place within a living organism. These reactions include the uptake of nutrients into a system, the processing of these nutrients into products the organism needs, and the removal of waste products from the system. Most of these reactions are catalysed by enzymes. Often the chemicals used in metabolic reactions are referred to as metabolites. Metabolism can be broadly split into two categories, anabolism and catabolism. Catabolism deals with the breaking down of larger metabolites into smaller ones, whilst anabolism deals with the combining of smaller metabolites into larger products a system needs, such as ATP or amino acids. Generally catabolism produces energy an organism needs, whilst anabolism consumes it. Metabolism is an ongoing process within living cells, with multiple reactions occurring at any one time. In order to visualize the reactions happening in a cell, we create metabolic networks. Metabolic networks are effectively maps we can examine to trace the reactions that lead from an input to an output [60]. The study of metabolism is of interest in a number of different fields. Drug metabolism [23] is the study of how the body can metabolise compounds that are not native to the system. This is especially useful in the study of how the human body can metabolise different drugs, but also has relevance in non pharmaceutical contexts. The study of metabolism in tumours is a field of much activity, with application to identification of cancer strains [2] and the treatment of cancer [55]. Other areas of interest, especially with regards to metabolic engineering, are genetically modified (GM) crops, food security, and the production of biofuels.

1.2 Metabolic engineering

According to Bailey [6], metabolic engineering is the improvement of cellular activities by manipulation of enzymatic, transport, and regulatory functions of the cell. That is, by the modification of the DNA of an organism, we aim to improve its metabolic function. The goals of metabolic engineering of an organism were broadly be divided into five categories by Cameron and Tong [9]. Firstly, increasing the yields of chemicals already produced by the organism, such as biofuels naturally produced as a waste product. Secondly, introducing new metabolic pathways to allow the organism to grow on a wider variety of food sources. Next,

adding new pathways to allow the organism to metabolise and remove chemicals that are toxic to it, increasing survivability. Fourth, the introduction of pathways to the system that allow it to produce additional products that it previously did not. This is especially useful in the production of various drugs and antibiotics. Finally, we can also use metabolic engineering to modify the physical properties of a cell, such as its level of growth. We are especially interested in the first goal, the improvement of yields of already existing products, and to a lesser extent, the fourth, in which we cause an organism to produce new products.

Whilst metabolic engineering can lead to the improvement of an organism, it can also lead to a greater understanding of the pathways within a system. Lee et al. [40] investigated the ATP production of *Escherichia coli* in a mutant strain. This mutant had a deficiency in its TCA cycle, that caused carbon from glycolysis to not enter into it. It was suspected that in an organism growing on glucose, the TCA cycle was highly suppressed and was primarily used for biosynthesis. The flux distributions for this mutant and the wildtype showed very similar amounts of ATP production from the glycolytic pathway, despite carbon not being able to enter the TCA cycle in the mutant type. This confirmed the idea of a repressed TCA cycle.

Chao and Liao [11] performed an experiment in which *E. coli* was modified to over express two specific enzymes, which caused a new cycle between two metabolites. This lead to futile cycling between the two metabolites. That is, oxygen and glucose uptakes increased suggesting a higher rate of respiration, however growth rates of the bacteria decreased. There was also an increase in the production of the waste products pyruvate and acetate. These effects were not observed when only one of the enzymes was overexpressed.

The production of biofuels such as ethanol is a key area of research for metabolic engineering. Normally, an organism that already produces the desired product is chosen, and then further modified to increase the production of this product. Sometimes this is done by optimizing and improving already existing pathways, sometimes it can be done by adding in pathways that were not present in the original wildtype. Often a mixture of the two approaches is used.

Zhang et al. [81] took an ethanol producing bacteria, *Zymomonas mobilis*, and modified it to include a new pathway. This pathway allowed the bacteria to effectively metabolise a new carbon source. In this case the new carbon source was xylose, a pentose sugar. This new mutant strain was able to ferment xylose to create ethanol. In addition, the modified bacteria did not lose the ability to ferment glucose. The ability to grow on xylose, glucose, and a combination of the two allowed it to efficiently produce ethanol.

Ethanol is not the only biofuel we may wish to increase the production of. More complex alcohols may prove more useful as biofules due to the higher energy densities present within them. However there is no organism that natively produces these more complex alcohols in an efficient manner. Some bacteria, however, may produce minimal amounts of these alcohols as byproducts.

By introducing new genes into *E. coli* to direct metabolites usually used to form amino acids into alcohol production, Atsumi et al. [4] were able to stimulate various alcohol productions in *E. coli*. By inserting genes associated with the reactions, the bacteria could be made to produce various alcohols, such as isobutanol, 1-butanol and 1-propanol. The pathways were then further optimized by

inhibiting other reactions that used the metabolites required for the production of these alcohols, by removing genes that produced enzymes associated with these reactions. This further increased the production rates of the desired biofuels. In the case of isobutanol, one strain managed to produce 0.35g of isobutanol per gram of glucose uptake. This was 86 percent of the theoretical maximum. They further examined the production of 1-butanol in *E. coli* [3]. In anaerobic conditions production of 0.21g of 1-butanol per gram of glucose uptake was achieved, where previously there was negligible production. In aerobic conditions with a small amount of oxygen, this production was increased; however, large amounts of oxygen led to the stimulation of the TCA cycle, causing the carbon to be directed away from alcohol production.

Due to the wealth of genomic information on *E. coli*, Hanai et al. [29] chose to engineer it to produce isopropanol rather than attempt to increase the yield of other bacteria that naturally produce it, such as various members of the clostridia family. This was achieved by first adding in genes normally found in bacteria of the clostridia family in order to stimulate acetone production. The yields of acetone were at a similar level to the bacteria from which the genes were initially taken. To further produce isopropanol from this acetone another gene was added to cause the transformation of acetone into the desired isopropanol. The maximum yield for these genetically engineered strains was 0.43 grams of isopropanol per gram of glucose.

E. coli has also been engineered to produce 2-methyl-1-butanol [10] and 3-methyl-1-butanol [14] separately. *E. coli* is often favoured as an organism due to the unusually complete knowledge of its genes and metabolism [29].

More recently, as knowledge and techniques have been created for the genetic engineering of other organisms, there has been more work into the genetic engineering of members of the clostridium family. Mann was able to increase butanol production within *Clostridium acetobutylicum* via the overexpression of a specific enzyme [45]. Others have found methods of changing the main production of *C. acetobutylicum* from acetone, butanol and ethanol (known as ABE fermentation) into isopropanol, butanol and ethanol (known as IBE fermentation)[**?**, **?**, **?**]. The desire to alter metabolic activity in an organism raises some problems. Whilst it is possible to add and delete genes to modify the DNA of a bacteria and attempt to modify its metabolism, we may not fully understand the effects of these changes. It may be that removing a gene causes the bacteria to no longer be able to produce necessary biological products for survival. Alternatively, we may find that adding in a gene boosts the creation of a specific product, but to much lower levels that we would have hoped. This is where the ability to mathematically model the metabolism of an organism becomes useful. By modelling metabolic networks, we aim to gain a greater understanding into the dynamics of metabolism. In particular, these models help us predict the effects of any changes we may make to metabolism through the course of metabolic engineering. We may also be able to find and highlight specific targets for metabolic engineering, by simulating them in the model first. An obvious choice for the modelling of metabolism is the theory of ordinary differential equations (ODEs).

1.3 Kinetic modelling of metabolism

When modelling metabolism, we must first consider how to model reaction rates. Reaction rates are the rate at which a reaction occurs and are functions of the concentrations of metabolites within the network. A simple method to model these reaction rates would be using the law of mass action, assuming that the reaction rates are proportional to the concentrations of metabolites consumed by the reaction. Since the majority of reactions are enzyme catalysed, we could also choose to model them using the standard Michaelis-Menten kinetics [48]. There are also various transport processes that would need to be considered, as well as uptake rates, that are also taken to be metabolic reactions in a model. Then the rate of change of a metabolite concentration would be equal to the sum of all reaction rates that produce that specific metabolite minus the sum of all reaction rates that consume it, i.e.

$$\frac{dc_n}{dt} = \sum_n f_n(\mathbf{c}) - \sum_m g_m(\mathbf{c}), \tag{1.3.1}$$

where $\mathbf{c}$ is a vector of all metabolite concentrations, c_n is a specific element of that vector, representing a single metabolite concentration, $f_n(\mathbf{c})$ are the reaction rates of the reactions that produce the metabolite c_n, and $g_m(\mathbf{c})$ are the reaction rates of the reactions that consume c_n.

Once an ODE model is formulated, we have a number of parameters that must be estimated. The estimation of parameters that fit experimental data is in itself a difficult task. Identifying best fitting parameters requires solving a non-linear optimisation problem, with derivatives as constraints. Methods for solving these problems often have the issue of only finding local optima, and there is no deterministic method that finds global optima for every problem of this form in finite time. Stochastic methods are also used, though due to the random nature of these methods, there is still no guarantee that these methods find a global optimum. Mole et al. [49] compared various global optimum searching methods, and concluded that only one stochastic method, known as evolution strategies, could reliably find solutions to parameter estimation problems for kinetic models of metabolism.

Rodriguez-Fernandez et al. [58] proposed a hybrid method for finding best fitting parameters. By using a stochastic global optimum searching technique initially, and then switching to a quicker local optimum searching technique when in the basin of attraction for the global optimum, the computation time is decreased drastically. Whilst the point at which the methods are switched must be tuned for each model individually, this still provides better results than a local optimum searching method alone. This hybrid technique was also able to deal with noisy data (due to experimental errors). However, there is still an issue with identifiability. As examined by Audoly et al. [5], parameters may not be globally identifiable. That is, there may be multiple combinations of parameter values leading to the same optimum value for the optimisation problem. These parameters can not then be identified uniquely and may require additional data in order to achieve accurate parameterisations of a model. In spite of these difficulties, full kinetic models for various organisms have been developed.

Chassagnole et al. [12] designed a model for the central carbon metabolism of *E. coli*. This was the first model to link central carbon metabolism with the

sugar transport system within the organism. The model was compared with experimental data of intracellular concentrations of metabolites to attempt to validate the structure. The data was used further to attempt to estimate the kinetic parameters in the model to provide a more accurate model.

Usada et al. [73] created a more complex model, based on the one put forward by Chassagnole et al. [12], expanding it to include the regulatory system of *E. coli*, through to the production of amino acids. Using previously measured kinetic parameters a model was formed and compared to experimental data from a culture grown on fructose. The uptake rate of fructose in the experiment agreed with the prediction from the dynamic model, as did the CO_2 excretion rate, however the external acetate concentrations predicted by the model were somewhat higher than the predicted values. This may be explained by the model not allowing the uptake and reutilization of acetate.

A further model for *E. coli* was developed by Peskov et al. [56], in which enzyme regulatory properties were included when formulating the reaction rates This was done with the aim of increasing the predictive power of the model. By including detailed biochemical properties of enzymes, both *in vivo* and *in vitro*, experimental data were able to estimate parameters to verify the model. The model was used to describe metabolite kinetics across a timescale of several minutes, but beyond that enzyme production and degradation would also need to be taken into account for an accurate model.

Kinetic modelling has also been applied to the metabolism of plants, as reviewed by Rohwer [59]. In particular, many models have been developed to model the process of photosynthesis, in both C_3 plants (plants which use the C_3 metabolic pathway for photosynthesis), as described by Gross et al. [28] and Kirschbaum [37], and in C_4 plants (plants which use the C_4 metabolic pathway for photosynthesis), by Chen et al. [13].

Formulating ODE models for metabolic networks containing large amounts of reactions and metabolites is difficult. A lack of data on kinetic parameters, and the difficulty in obtaining them has lead to the development of simplified ODE models being developed, using approximate reaction rates, reducing the number of parameters needing to be estimated. These approximate reaction rates are based in the theory of metabolic control analysis (MCA) as described by Fell [**?**, **?**]. MCA allows us to calculate the "elasticities" of reaction rates. These elasticities quantify how much a reaction rate increases or decreases when a specific metabolite concentration is increased or decreased from a reference state. Similarly, elasticities can also be calculated for the perturbation of enzyme levels. Then using these elasticities, along with the reference concentrations, approximate reaction rate equations are formulated. A common form for these approximate reactions is linlog kinetics, as seen in studies by Visser et al. [76, **?**], and Nikerel et al. [51]. Visser et al. [77] also presented another form of approximate reaction rates, termed "tendency modelling", based on generalised mass action power laws.

Other methods aim to bypass the difficulty of estimating parameters. Ensemble modelling, developed by Tran et al. [72] and used by Tan et al. [71] seeks to build up an ensemble of models that all reach the same steady state. By using elementary reactions (in which chemical reactions are split into their constituent steps), each step is modelled using the law of mass action. Then by

screening models using a similar technique to finding elasticities in MCA, measuring sensitivity of reactions to perturbations in enzyme levels, this ensemble of models is screened down to a smaller set. Thermodynamic data can also be added to provide extra limitations on the parameters of these models. By forming many models with different parameters and screening down to those that match the data, rather than building a single model and attempting to find the best matching parameters, the computational complexity of parameterising a model is decreased, though the data on perturbations may not be readily available.

Whilst kinetic models are able to provide information on the dynamic behaviour of a metabolic network, it may be cost prohibitive to generate the data required to parameterise such a model. In other cases, we may find difficulty finding a structure for the model. It is not uncommon that whilst we have a great deal of information about the network of reactions forming the metabolism in an organism, we may have very little information on the exact kinetics. In these cases, it is useful to use techniques that only require information about the structure of the metabolic network (known as the stoichiometry). The class of techniques that rely on this stoichiometry data are often referred to as COBRA (constraint based reconstruction and analysis) methods, laid out in the software package, the COBRA toolbox [64]. Of these methods, the subset of most interest to us are flux balance techniques.

1.4 Flux balance techniques

The modelling of metabolic networks is often a difficult task. Due to the difficulty in measuring enzyme reaction rates and various other kinetic parameters, accurate ODE models are costly to produce. In order to deal with these problems, techniques have been developed that bypass the need for kinetic parameters entirely. Rather than focusing on dynamic models, these techniques instead observe systems in a steady state, and examine possible flux values (reaction rates) in these steady states. In order to fix a system to its steady-state, the condition of mass balance is imposed, requiring that the sum of fluxes producing a metabolite balances with the sum of fluxes consuming the metabolite. These techniques are therefore known as flux balance techniques.

This mass balance produces a set of equations, $S\mathbf{v} = \mathbf{0}$, where S is the stoichiometric matrix of a metabolic network, and $\mathbf{v}$ is a vector containing values for all the fluxes in a system, called the flux vector. This set of equations is often underdetermined, meaning we have less equations than unknown elements of $\mathbf{v}$. Hence we need some technique to include additional constraints to find the most relevant solutions from the infinitely many prescribed by $S\mathbf{v} = \mathbf{0}$

On such technique is stoichiometric metabolic flux analysis (MFA). MFA seeks to fill in the gaps by fixing fluxes to experimentally measured values in order to fully determine the system. In practice, the measuring of internal fluxes is difficult, so often the only fluxes we are able to measure are external fluxes. As described by Wittmann and Portais [80], by measuring these external fluxes, we gain some idea of the internal fluxes by extension. However it is often the case that due to internal loops and cycles, we are unable to find a unique solution for the system of equations described, and therefore techniques for measuring the internal fluxes are necessary.

A commonly used method for measuring internal fluxes is ^{13}C metabolic flux analysis. In this, marked carbon isotopes are entered into the metabolic system, and then using mathematical modelling techniques and concentrations of marked carbons in different forms, we estimate the internal fluxes. Software has been developed to perform this analysis [78]. We can then take these estimated fluxes and include them in our system of equations to provide more accurate steady state solutions. With enough measured fluxes, the system becomes fully determined, and a completely unique solution is found.

The solution found in this manner is subject to experimental error in the measurement of fluxes. In some cases, an experimental error on a flux could lead to a large variation in an entire flux vector found by applying MFA. By finding condition numbers for each flux in a network, Savinell and Palsson [62] were able to find an upper bound for a networks sensitivity to perturbations in each flux. The fluxes with the lowest condition numbers would be the best target for experimental measurement, allowing minimal sensitivity to experimental error in the solutions. Savinell and Palsson [61] applied this technique to *E. coli* finding that experimental errors could, in the worst case, be amplified up to 60 fold. In a hybridoma cell, these amplifications of experimental error could be orders of magnitude higher (up to 10^5). By measuring the fluxes with the lowest condition numbers, the sensitivity to experimental error would be much lower, and measuring additional fluxes past the number required to fully determine the MFA system would also lower the effects of these errors.

There are other methods of extracting relevant solutions from the set of those that solve $S\mathbf{v} = \mathbf{0}$. The most prevalent uses linear programming to find the solution that maximises a specific objective function. The idea of using linear programming to extract an optimal state for a metabolic network was first used by Papoutsakis [54], and this idea was eventually extended into a technique known as stoichiometric flux balancing by Varma et al. [74]. This technique later came to be known as flux balance analysis (FBA).

FBA, as described in recent review papers by Orth et al. [53] and Kauffman et al. [33], is a technique based on constraining the fluxes within a metabolic network to produce biologically viable steady states. The first constraint used in this technique is the previously described mass balance constraint, requiring $S\mathbf{v} = \mathbf{0}$. We then set an objective function we wish to find an optimal solution for. Generally, this objective function is of the form $\mathbf{a}^T\mathbf{v}$, making it some linear combination of fluxes in the network. One common choice would be maximising biomass production. In this case, the objective function maximises the production of metabolites that are used to produce biomass. Other possible choices include maximising the flux into a specific product of the network (such as ethanol, in the case of biofuel production), or ATP production. Finally we must place upper bounds on our flux values, to prevent solutions that are infinitely large in size. To this end all fluxes are constrained to being between a maximum and a minimum value. Where experimentally measured fluxes exist, we constrain these fluxes to being the measured value plus or minus an error term. In this way we find a steady state flux vector that maximises our objective function. A basic FBA problem is written in the form

$$\max_{\mathbf{v}} \quad \mathbf{a}^T\mathbf{v} \quad \text{such that: } S\mathbf{v} = \mathbf{0} \text{with,}$$

$$\mathbf{v}_{min} \leq \mathbf{v} \leq \mathbf{v}_{max}, \tag{1.4.1}$$

where $\mathbf{v}_{min}$ and $\mathbf{v}_{max}$ are vectors containing the minimum and maximum values for each flux respectively. In general if this problem has solutions it is feasible, and if the constraints permit no possible solutions it is called infeasible. Farkas' lemma tells us that exactly one of the problems,

$$A\mathbf{v} \leq \mathbf{b} \tag{1.4.2}$$

and

$$A^T\mathbf{y} = \mathbf{0}, \mathbf{b}^T\mathbf{y} \leq 0 \tag{1.4.3}$$

has solutions. Our linear programming constraints can be written in this first form, noting that our equality constraint, $S\mathbf{v} = \mathbf{0}$ can be rewritten as

$$S\mathbf{v} \leq \mathbf{0}, S\mathbf{v} \geq \mathbf{0} \tag{1.4.4}$$

and

$$\mathbf{v}_{min} \leq \mathbf{v} \Rightarrow -\mathbf{v} \leq -\mathbf{v}_{min}. \tag{1.4.5}$$

We can then construct a matrix A containing the left hand side of all our new inequality constraints, and a column vector b containing the right hand side for all of these constraints. Then if there exists some vector y that satisfies equation (1.4.3), our linear programming problem is infeasible. We still cannot guarantee a unique solution. Similarly to when estimating parameters, it is possible for multiple flux vectors to optimise this objective function. We must be aware of the possibility of multiple optimal solutions when performing FBA.

FBA is a simple and fast technique to obtain information on possible steady state flux distributions for a metabolic network. In particular, solving a linear programming problem is not computationally expensive, allowing even problems with large stoichiometric matrices, such as those for genome scale models, to be solved quickly.

1.4.1 Applications of flux balance analysis

FBA has been applied with great success to *E. coli* bacteria. Due to the wealth of biochemical data, and the a completely sequenced genome, Edwards and Palsson [18] were able to construct the entire metabolic network of the organism. In this case, the stoichiometric matrix used was 436x720. Flux balance analysis was then applied to this network, with upper bounds on the glucose and oxygen uptakes to find the steady state in which growth was maximal. To find a maximal growth state, a reaction in which the precursors to biomass were gathered and synthesized into biomass was added, and this was then used as the objective function to be maximised.

Continuing on from this, gene knockouts were simulated for each individual gene in the genome by constraining flux across reactions connected to those genes to be zero. In this way we observed which genes are essential to the continued growth of the organism. The maximal growth rates after these deletions was compared to that before to see if there was a significant decrease from the deletion.

To validate the results found, the predicted results were compared to experimental results for known mutants. Answering the question of whether or not a gene was essential for growth, the model predictions were correct in 86 percent (68/79) of cases.

More recently, metabolic networks have been constructed to some degree of completion for *Plasmodium falciparum*. This is one of the types of parasite that causes malaria in humans. Plata [57] constructed a model containing 1001 reactions across 616 metabolites. The reactions involved include numerous metabolite transformations and transport processes. As with previous studies of large metabolic networks for known organisms, gene deletions were simulated. Also simulated were drug inhibitions, when a drug is used to inhibit a reaction to try and kill the organism. Whilst the data on *P. falciparum* is not as extensive as that for *E. coli*, 14 gene deletions and 25 drug inhibitions were able to be compared with existing data. The model correctly predicted whether or not a gene being deleted was essential in 100 percent of cases. For drug inhibitions, the prediction rate was lower at 70 percent. All of the cases in which FBA gave the incorrect result were cases in which FBA predicted the organism would survive, but the drug in question had been experimentally proven to be lethal to the organism. In particular there were no cases of FBA predicting the organism dying when it would actually survive.

FBA has also been applied in the field of predictive microbiology. Predictive microbiology aims to describe the growth and inactivation of pathogens and microbes within food. Osmotic shock is a state in which the concentration of a soluble material around a cell changes rapidly, causing a rapid change in movement across a cell's membrane. Under these circumstances, there is a large change in the growth rate of the cell. Metris et al. [47] applied FBA to cells undergoing osmotic shock in order to explain the reason behind it. It was found that a change in biomass composition could not explain the lowered growth rate. It would also take a very high change in energy required for cell maintenance (both growth and non-growth related) to explain such a growth rate decrease. In this case, it could be said that the factors that cause the limited growth rate are those not modelled by FBA, such as intracellular crowding (space limitations within the organism).

Beg et al. [8] provided a way to incorporate intracellular crowding into FBA models. Here we consider how much space each enzyme takes up within a cell and consider it a constraint on the overall flux. The first step is to constrain the number of enzyme macromolecules present to the volume of the cell. If v_i is the volume of the enzyme present in the ith reaction, n_i is the number of macromolecules of that enzyme present, and V is the overall volume of the cell, we have

$$\sum_i v_i n_i \leq V.$$

Further, we divide by cell mass to instead consider the enzyme concentrations rather than the number of macromolecules. In this way we find

$$\sum_i v_i E_i \leq \frac{1}{C},$$

where C is the cytoplasmic density of a cell and E_i is the concentration of the enzyme present in the ith reaction. We now have constraints on enzyme concentrations. If we go one step further, we consider an enzyme concentration E_i

corresponding to a specific flux $q_i = b_i E_i$, where b_i is a rate coefficient for the reaction. Then our constraint is written as

$$\sum_i a_i q_i \leq 1.$$

Here, we have q_i being the ith flux, and a_i is defined as

$$a_i = \frac{C v_i}{b_i}.$$

In this way we have converted a constraint on the number of enzyme molecules into a constraint on fluxes. An obvious downside to this extension is we now need some knowledge of the rate coefficients of each reaction. It is possible to estimate these parameters by comparing predicted results with experimental data and fitting the parameter to the data.

1.4.2 Extensions to flux balance analysis

The strength of FBA is in its relative simplicity. It requires little prior knowledge about the system and computing solutions via LP methods is computationally cheap. Its simplicity is also a weakness of the technique, and many extensions and additions to FBA have been applied successfully in order to improve the results found and the situations it is applicable to.

In addition to mass balance constraints, we can also add energy balance constraints to the system. Energy balance allows us to further constrain our solution space to only include thermodynamically feasible solutions. Beard et al. [7] laid out two additional constraints we can add to an FBA problem in order to incorporate energy balance. First we consider a vector $\mathbf{\Delta u}$ of chemical potential differences associated with each reaction in a system. The first constraint is there must exist some $\mathbf{\Delta u}$ such that $v_n \Delta u_n \leq 0$, where v_n is the nth flux. This constraint corresponds to the second law of thermodynamics, and ensures entropy production is non-negative for each reaction. For the second constraint, we consider the stoichiometric matrix S. If we remove all columns corresponding to outside fluxes from this matrix (those that add or remove metabolites from the system), we have a modified stoichiometric matrix S'. We then find the null space of this matrix. We then store the null space vectors as rows in a matrix K. It is seen that if we take all the reactions in the system and multiply them by the elements of either of these null space vectors, and then sum them, we obtain a perfectly balanced chemical equation. Similarly, if we multiply our chemical potential differences by these elements and sum them, we should obtain a perfect energy balance. That is to say that $K\mathbf{\Delta u} = \mathbf{0}$. So in order to constrain our system so that it is balanced for energy, there must exist a $\mathbf{\Delta u}$ such that $K\mathbf{\Delta u} = \mathbf{0}$. Combining this with the previous constraint, we require there to exist some $\mathbf{\Delta u}$ such that both constraints hold. This restricts our solution space to only thermodynamically feasible solutions. This method has been seen to improve the predictive power of the effects of gene knockouts on *E. coli* [7]. A possible problem with these additional constraints is the potential for numerical error when calculating the nullspace vectors that form the matrix K.

More recently, the idea of thermodynamic feasibility has been investigated further. Zhu et al. [82] proposed a method to not only restrict the solution space

to a thermodynamically feasible one, but to actively seek the most thermodynamically optimal solution. It has previously been observed by Martyushev et al. [46] that metabolic networks tend to end up in a steady state that maximises entropy production (also known as the maximum entropy production principle or MEPP), so it would be sensible to consider this in FBA. In thermodynamic optimum searching, the Gibbs free energy production for each reaction must firstly be estimated to a range. The maximum and minimum flux values for each reaction are set based on these ranges. Then using these maxima and minima along with mass balance, a standard FBA problem is solved to maximise biomass to find some optimal value. The final step is another optimisation, based on a non-linear objective function to maximise entropy production per unit energy. In this, the mass balance constraints are maintained, but biomass production is fixed at its optimal value observed in the previous step. In this way a thermodynamically optimal solution is found.

For some FBA problems, there are multiple solutions corresponding to the same maximal value for the objective function. We may not be aware of these alternative optima by simply performing FBA. Mixed Integer Linear Programming (MILP) methods have been developed to find these alternate optima. Lee et al. [41] presented an algorithm guaranteed to find all alternative optima. These methods are iterative and the number of steps required grow exponentially with the number of alternate optima. Some systems may contain many alternative optima, particularly those with internal flux loops, causing these methods to become increasingly computationally intensive, to the point of infeasibility. Indeed, many problems may have infinitely many optimal solutions, severely limiting the usefulness of trying to find all alternative optima.

Rather than trying to find all these alternative optima, it may be of more use to examine how much a given flux can vary whilst still maintaining the same maximised value for the objective function. Flux variability analysis (FVA) is a technique used to examine this. The first step is to perform FBA as usual. In this, we obtain a maximised value for our objective function $\mathbf{a}^T\mathbf{v} = A_{max}$, where $\mathbf{a}^T\mathbf{v}$ is our objective function and A_{max} is the maximised value found. We then set up two additional linear programming (LP) problems for each flux within the system to determine their maximum and minimum values whilst maintaining the same objective function value. The problem is

$$\begin{aligned} \underset{\mathbf{v}}{max/min} \quad v_n \quad \text{such that: } & S\mathbf{v} = \mathbf{0}, \text{with} \\ & \mathbf{a}^T\mathbf{v} = A_{max}, \text{and} \\ & \mathbf{v_{min}} \leq \mathbf{v} \leq \mathbf{v_{max}}. \end{aligned} \tag{1.4.6}$$

By performing $2n$ additional LP problems, where n is the number of fluxes in the system, we find how much a flux can be varied whilst still maintaining the same maximised objective function. Mahadevan [44] presented the effects that these alternate solutions could have on gene knockouts within a system. Also explored was the idea of sub-optimal flux distributions, in which the objective function was not maximised, but was at least a certain percentage of the optimised value. Here, the first constraint in the linear programming problem is changed to

$$\mathbf{a}^T\mathbf{v} \geq (1-\delta)A_{max},$$

where δ is a small number. This allows us to examine the variability of fluxes in sub-optimal cases. These may be of interest to studies after gene knockouts or after a change of carbon source. In these situations the network may not necessarily immediately optimise itself in its new conditions, and instead remain in a sub-optimal flux distribution that is closer to its original state.

FBA has been used to simulate the effect of gene knockouts on a system. That is, removing a gene from the organism causing one or more reactions to no longer occur within it. Performing FBA on a system with a gene knockout provides a new optimal flux distribution, however a system may not necessarily default to this new optimal state. It is hypothesised that the system attempts to come to a steady state that is as close to its previous steady state as possible, even if this results in sub-optimal fluxes. To account for this, when simulating the effects of a gene knockout, rather than setting an objective function to maximise a desired product, we should instead set it to minimise the change between the original steady state and the steady state of this new system. Segrè [70] found success in minimising the Euclidean distance between the two flux distributions. Here the new objective function is

$$\sqrt{(\sum_i((x_i - v_i)^2)}$$

and we seek to minimise this. Here, x_i is the ith flux from the flux distribution before the knockout, and v_i is the ith flux in the FBA problem. Since the square root is a monotonic function, minimising $\sqrt{x}$ is equivalent to minimising x. We also ignore any constant terms in this minimisation as minimising $x - c$ where c is constant, is the same as minimising x. Then our objective function is reduced to

$$\sum_i (v_i^2 - 2x_i v_i).$$

We can no longer use linear programming methods to solve this, as our objective function is no longer linear, however quadratic programming (QP) methods exist for minimising a quadratic objective function with linear constraints, as described by Frank and Wolfe[22].

One of the defining features of FBA is that it can only be used to predict the steady states of a system. Dynamic flux balance analysis (dFBA), however, can be used to model a system transitioning between two steady states. Mahadevan applied two methods, both producing similar results [43] . The first is the dynamic optimization approach, in which the entire time period is looked at as a whole, and the rates of changes of fluxes, growth rate and metabolic concentrations are limited by their time derivatives. Then the problem is solved as a single non-linear programming problem. The second is the static optimization approach, where the time period is divided into N subintervals. Rates of changes are then constrained by limiting how much they can change from their value in the previous subinterval. Then the system is solved as N linear programming problems. We then have flux vectors, concentration vectors and a growth rate prediction at each of these time intervals. We continue iterating this process until the system has reached a new steady state. The key advantage of this technique is it allows us not only predict how the growth of an organism changes after modification, but also how the concentration vector does, which is normally ignored by FBA.

Dynamic flux balance analysis has been applied with some success to larger scale systems. Grafahrend-Belau et al. [27] used a multiscale modelling approach in order to produce a dFBA model on a whole plant scale. By modelling on a large scale, they were able to present a metabolic analysis of the interactions between both seed and stem biomass in barley, which could later allow targeted metabolic engineering to improve crop yields.

Covert et al. [15] were also able to extend FBA to provide dynamic results. By incorporating boolean constraints to represent regulatory networks, they were able to examine diauxic growth of an example carbon metabolism network on complex mediums. In particular they examined growth on a mixture of two carbon sources, as well as a switch from aerobic to anaerobic growth. Whilst the method used here did not provide estimates of internal concentration vectors, the addition of temporary regulatory constraints was a useful addition to flux balance analysis.

Biological systems generally require more than a single nutrient to grow. Commonly, a system requires at least a carbon source and an oxygen source. It is often of interest to find an optimal ratio for the uptake of these multiple compounds. Phenotype Phase Plane Analysis (PhPP analysis) as described by Edwards [19] utilizes the results of FBA to describe the effects of different nutrient uptake rates on the system. In order to generate the phase plane, first we must define shadow prices. The shadow price, Y_i of a metabolite c_i is given by

$$Y_i = -\frac{dA_{max}}{dc_i},$$

where A_{max} is the value of the objective function obtained in FBA. This defines how the addition of a metabolite affects the objective function, and suggests how useful a metabolite is to the system. These shadow prices remain fixed for a phenotype. That is, in each region on the phase plane they remain fixed, and only change when the uptakes change the system from one phenotype to another. Using this knowledge, we find critical ratios between the uptake rates of the nutrients that cause the system to shift. Often after a shift between phenotypes, different pathways are utilized in different ways. For example, if the system has an excess of one nutrient, it may need to activate a pathway to get rid of it in a different way.

We can also use isoclines to obtain qualitative information about the system. Isoclines are lines where the objective function $\mathbf{a}^T\mathbf{v} = A_{max}$ takes the same value. The slopes of isoclines are given by $\rho = -Y_i/Y_j$, that is the negative ratio of shadow prices for the nutrients involved. As the slopes are found by the shadow prices which are different in each region of the phase plane, the slopes themselves are also different in each region. Regions with a positive ρ are said to be "futile", as increasing the uptake rate of one nutrient has a negative effect on the objective function. In this way, we characterise the different possible phenotypes in a system and see at which points the system shifts from one to another.

A key difficulty in FBA is choosing an objective function. Whilst biomass production is a common choice, we can't always assume it leads to the best results. Schuetz examined the results from various objective functions on an *E. coli* network and compared them to actual *in vivo* measured fluxes [67]. Surprisingly it was found when modelling batch cultures biomass production was

not the most accurate predictor. ATP production per unit flux was found to be the best objective function in this case. ATP production per unit flux is a non linear objective function given as the ATP production divided by the sum of fluxes squared. This is a non convex objective function, so it is possible that it was only finding local optima, however this was examined by reformulating the objective function in a convex for and found to not be the case. In continuous chemostat cultures, it was found that ATP production and biomass production as objective functions yielded accurate predictions of the internal fluxes. In cases for other objective functions, it was found that adding additional constraints could still provide accurate results, but these constraints required additional knowledge about the organism to impose, and therefore it is preferable to avoid them when possible. Ibarra et al. [32] showed that whilst, initially, a strain of *E. coli* exhibited a suboptimal growth rate compared to the theoretical maximum predicted by an FBA model, after forty generations of adaptive evolution, this same strain exhibited the optimal rate predicted by FBA. This data provides further support to using biomass production as an objective function.

1.5 Stochastic models

All the types of models considered so far are deterministic, however it is also possible to use stochastic methods to model metabolism. The chemical master equation, explicitly derived by Gillespie [24], is a system of differential equations that describes the evolution of a metabolic network as a stochastic process. The solution to these differential equations provides the probability density function for the network at a given time. This is sometimes called the propensity function. By using a method such as the stochastic simulation algorithm (SSA) proposed by Gillespie [25], or the software package STOCHSIM, developed by Le Novere and Shimizu [52], this stochastic process is simulated. It is seen that cases with large quantities of metabolites, these solutions converge to those seen in deterministic models. Stochastic models, whilst useful tools for modelling metabolism, particularly in cases where there are low concentrations of metabolites, are generally outside the scope of this work, and will not be considered further.

1.6 Overview

There are many different methods that can be used to model metabolism, and each has its own strengths and weaknesses. Throughout the course of this work, we aim to model metabolic networks using a variety of techniques. In particular, we will comprehensively examine flux balance techniques, specifically flux balance analysis, and its dynamic counterpart, dynamic flux balance analysis. We will also examine a kinetic model of a metabolic network for *Clostridium autoethanogenum*, and seek to compare results from these methods. We will also examine the previously mentioned technique of ensemble modelling as a method of parameterising a kinetic model, using data from both experimental sources, and provided by flux balance methods. By examining these methods in detail, we are able to identify the advantages of each class of model, and determine under which situations they might be of most use. Specifically, we determine

what knowledge would be needed of a metabolic network to in order to use each method, and what information could then be obtained. We also examine which long timescale behaviours can be modelled by each method in order to further highlight which methods would be most applicable.

In Chapter 2 we examine flux balance methods in detail. First, we apply MFA to small networks, before moving on to applying FBA to similar small networks. By applying FBA to these toy networks, we look to gain an understanding of the key features of the technique, and how it behaves under a variety of circumstances. We will examine the links between the flux balance models and kinetic ODE models, before also examining two variations of FBA, flux variability analysis (FVA), and elementary modes analysis (EMA). Finally we apply all these techniques to a metabolic network for *C. autoethanogenum*.

In Chapter 3, we formulate an ODE model for the metabolism of *C. autoethanogenum*. We note that this model exhibits both steady states, and states with concentrations exhibiting unbounded growth. By considering long timescale asymptotics, we are able to quantify every long timescale behaviour this model is able to display. We also find conditions on parameters that will lead to each long term behaviour. In particular, we find the ratio of inputs into the network is important for determining which behaviour is seen.

In Chapter 4, we compare the steady state results from both the FBA model seen in Chapter 2, and the ODE model seen in Chapter 3. Further, by comparing certain flux ratios in the FBA model, we form a steady state region that is equivalent to the steady state region seen for the ODEs. We also consider a slight change to the network, replacing one reaction with two others, and see how this affects the steady state region seen in both cases. We find this leads to a region in parameter space that permits the existence of both steady state solutions, and a state in which concentrations grow linearly in time, allowing our ODE model to exhibit multiple long timescale behaviours for the same parameter set.

Chapter 5 examines the possibility of using flux balance methods to predict the existence of multiple steady states, such as those found at the end of Chapter 4. We introduce a well studied autocatalytic network that is known to permit multiple steady states for a range of parameter values and attempt to find the same results using FBA.

We introduce the technique of dynamic flux balance analysis (dFBA) in Chapter 6. Initially applying this technique to small networks, including the autocatalytic network introduced in Chapter 5, we attempt to again find the existence of multiple steady states using this new dynamic extension to FBA. We also apply the technique to a metabolic network analogous to the Lotka-Volterra predator prey equations to attempt to model dynamic long term behaviour in the form of periodic solutions. Finally we apply the technique to our *C. autoethanogenum* case study, to observe the linear growing states found in Chapter 3.

In Chapter 8, we examine ensemble modelling. We apply this technique first to a single reaction, before moving on to the ODE model for *C. autoethanogenum* metabolism introduced in Chapter 3. First we attempt to parameterise the model using data from an equivalent ODE model, before attempting to parameterise an ODE model using data from the dFBA model presented in Chapter 7. In this way, we aim to produce a kinetic model that matches the dFBA model, and is

therefore optimal according to an objective function.

Our conclusions are presented in Chapter 9. In this chapter we first apply previously examined techniques to a simple network containing a loop, to compare the difference in the level of information each technique is able to provide. We also examine the possibility of combining multiple techniques to provide additional insight.

Chapter 2

Using flux balance methods to model metabolic networks

2.1 Introduction

2.1.1 Modelling metabolic networks

Modelling the dynamics of metabolic networks is traditionally a difficult task. Forming complete ODE models for metabolic networks requires extensive knowledge, not only of the structure of the network, but the exact dynamics of each reaction. Knowledge of regulatory systems, and other dynamic processes can be difficult to attain, and even if a complete model is formed, we are left with many parameters to estimate from experimental data, which often proves highly expensive. Rather than relying on these complete dynamic models, in recent years attention has turned to flux balancing methods. These methods rely only on the structure of the network to provide us with information about how a metabolic network behaves at steady-state. Where an ODE model would provide solutions in terms of dynamic concentration profiles in time, flux balance techniques instead provide us with rates of each reaction of a network at steady-state. These reaction rates, commonly referred to as fluxes, are given in the form of flux vectors, denoted as $\mathbf{v}$. Each element of this vector contains the flux across one specific reaction at steady-state. By using these flux balance techniques, we seek to find information on steady-state reaction rates of a metabolic network.

2.1.2 The stoichiometric matrix

Flux balance techniques use information about the reactions within a system to form a stoichiometric matrix. This matrix contains all the data about internal metabolites and reactions in the system. Let us first consider a set of reactions that describe a simple network and generate the stoichiometric matrix for this system.

Glycolysis is a common biological process, in which glucose or other sugars, are transformed by a series of metabolic reactions into pyruvate. Pyruvate is a key metabolite for several other metabolic pathways, most notably the production of ATP in the citric acid cycle. A reduced set of reactions for glycolysis, listing only the key metabolites is given in Table 2.1 and a network diagram is shown

Glucose	$\xrightarrow{v_1}$	Glucose 6-phosphate
Glucose 6-phosphate	$\xrightarrow{v_2}$	Fructose 6-phosphate
Fructose 6-phosphate	$\xrightarrow{v_3}$	Fructose 1,6-biphosphate
Fructose 1,6-biphosphate	$\xrightarrow{v_4}$	Phosphoglyceraldehyde
Phosphoglyceraldehyde	$\xrightarrow{v_5}$	Pyruvate
External input	$\xrightarrow{Q_G}$	Glucose
Pyruvate	$\xrightarrow{Q_P}$	External output

Table 2.1: Set of reactions used in the reduced glycolysis pathway seen in Figure 2.1. Each reaction only includes the key metabolites used, neglecting additional metabolites such as NADH and hydrogen.

in Figure 2.1. The reactions are labelled with v_n representing internal reactions, such as the transformation of one metabolite into another, and Q_n representing the external reactions, such as the uptake of food sources and the output of products.

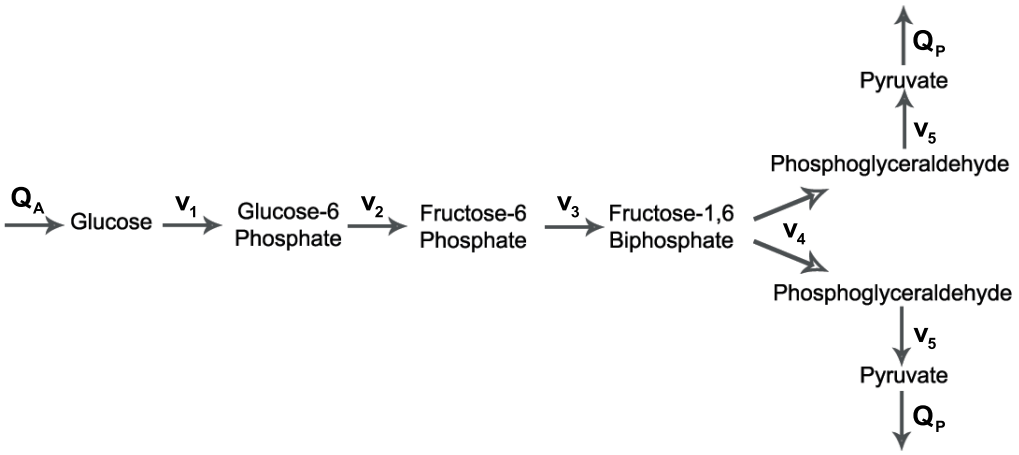

Figure 2.1: Simplified network diagram for the reactions involved in glycolysis described in Table 2.1. The diagram shows only the primary metabolites used in the metabolic pathway.

When we form our stoichiometric matrix, each column in the matrix contains the information for a single reaction, and each row contains the information for a single metabolite. We have six key metabolites in this network, and, including the uptake of glucose and output of pyruvate, seven reactions. Therefore the stoichiometric matrix for this network should be a 6×7 matrix. By looking at the reactions we are able to fill in the elements of the matrix. For example, the first reaction, v_1 consumes one glucose and produces one glucose 6-phosphate. Therefore, the column in our stoichiometric matrix representing v_1 should have a '-1' on the row corresponding to glucose, and a '$+1$' on the row corresponding to glucose 6-phosphate. All other metabolites are not used in this reaction, so the

elements in the v_1 column representing the remaining metabolites should equal zero. Filling in the remaining columns in the same manner provides us with the stoichiometric matrix for the glycolysis network,

$$S = \begin{array}{c} \begin{matrix} v_1 & v_2 & v_3 & v_4 & v_5 & Q_G & Q_P \end{matrix} \\ \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{pmatrix} \end{array} \begin{array}{l} \text{Glucose} \\ \text{Glucose 6-phosphate} \\ \text{Fructose 6-phosphate} \\ \text{Fructose 1,6-biphosphate} \\ \text{Phosphoglyceraldehyde} \\ \text{Pyruvate.} \end{array} \tag{2.1.1}$$

We note that the reaction v_4 uses one fructose 1,6-biphosphate to produce two phosphoglyceraldehyde, so the element of the stoichiometric matrix relating to the production of phosphoglyceraldehyde, in this case $S_{4,5}$, is equal to two.

We are able to freely rearrange the rows in this stoichiometric matrix with no effect on the network that the matrix represents. We could also rearrange the order of the columns, however this would have an effect on our flux balance methods. The order of the columns directly relates to the order of elements in the flux vector. If we encoded the v_1 reaction in the second column of the stoichiometric matrix, then the second element in the flux vector would be v_1. As long as the flux vector is consistent with the stoichiometric matrix, then the order of the columns in the stoichiometric matrix can also be freely changed whilst still representing the same network.

2.1.3 Mass balance

In a metabolic network, for a metabolite to satisfy the condition of mass balance, we must have that the flux in to the metabolite is equal to the flux out. That is, the amount of the metabolite produced by various reactions in the network is exactly equal to the amount used up by other reactions. In this way, the concentration of the metabolite remains constant in time. If all metabolites in a system satisfy mass balance, then the system is in steady-state. For example, from the reactions seen in Table 2.1, we see that glucose is involved in two reactions. One is the uptake of glucose, Q_G, and the other is the use of glucose to produce glucose 6-phosphate, v_1. If these two reactions balance, so that the flux in, Q_G, is equal to the flux out, v_1, then glucose is in mass balance, i.e. $Q_G = v_1$, or $Q_G - v_1 = 0$. If we extend this concept to the rest of the metabolites in the system, we obtain a set of six equations (one for each metabolite) that must be true for the whole network to be in mass balance. It turns out that this set of equations is exactly equal to $S\mathbf{v} = \mathbf{0}$, where S is the stoichiometric matrix seen in equation (2.1.1), and $\mathbf{v}$ is the flux vector. Now that we have introduced the concepts that are key to flux balance techniques, we seek to use these techniques to find steady-state solutions for metabolic networks.

2.2 Metabolic flux analysis

Metabolic flux analysis (MFA), as described by Wittmann and Portais [80], is a simple flux balance technique, using the stoichiometric matrix to find a steady-state solution for the fluxes in a metabolic network. Using the mass balance equation, $S\mathbf{v} = \mathbf{0}$, we generate a set of simultaneous equations that must be true for the system in steady-state. Generally there are less equations than unknown fluxes, so the system is underdetermined, having infinitely many possible solutions. In MFA, we add additional data in the form of experimentally measured flux values in order to fully determine the system and obtain a solution.

2.2.1 Applying MFA to glycolysis

To illustrate this technique, we apply it to our previously described glycolysis system. We use the stoichiometric matrix we have already found in equation (2.1.1), and combine it with the mass balance condition so that $S\mathbf{v} = \mathbf{0}$. Here the flux vector, $\mathbf{v}$ is given as

$$\mathbf{v} = (v_1, v_2, v_3, v_4, v_5, Q_G, Q_P). \tag{2.2.1}$$

Then we have the system of six simultaneous equations we solve to find a steady-state solution for the flux vector, v. These equations are

$$Q_G - v_1 = 0, \tag{2.2.2}$$
$$v_1 - v_2 = 0, \tag{2.2.3}$$
$$v_2 - v_3 = 0, \tag{2.2.4}$$
$$v_3 - v_4 = 0, \tag{2.2.5}$$
$$2v_4 - v_5 = 0, \tag{2.2.6}$$
$$v_5 - Q_P = 0. \tag{2.2.7}$$

Since we have only six equations for seven unknowns, the system is underdetermined, and does not have a unique solution. As this system is a simple chain of reactions, by adding in a single measured flux value, the system becomes fully determined and we obtain a unique solution. External fluxes are easier to experimentally measure than internal fluxes, so we should choose either the uptake rate of glucose, or the output rate of pyruvate as our measured flux value. In this case, we choose to include a measured value for the glucose uptake rate, such that

$$Q_G = x. \tag{2.2.8}$$

We write down expressions for each unknown in terms of this measured flux x, so that

$$v_1 = x, \tag{2.2.9}$$
$$v_2 = x, \tag{2.2.10}$$
$$v_3 = x, \tag{2.2.11}$$
$$v_4 = x, \tag{2.2.12}$$
$$v_5 = 2x, \tag{2.2.13}$$

$$Q_G = x, \tag{2.2.14}$$
$$Q_P = 2x. \tag{2.2.15}$$

Then our steady-state flux vector is $\mathbf{v} = (x, x, x, x, 2x, x, 2x)$, or equivalently,

$$\mathbf{v} = x(1, 1, 1, 1, 2, 1, 2), \tag{2.2.16}$$

and this steady-state is unique for a given value of x. In this case, we only needed to measure one additional flux to fully determine the system, and we had two external fluxes to choose from. If we had values for both external fluxes, then the system would become overdetermined. In that situation, if the two measured values were not consistent, the model would then have no valid solutions. That is, if Q_G was equal to x, but Q_P was not equal to $2x$, then the system of equations would become inconsistent and have no solutions. This could still potentially be a useful result, as it would tell us that there was either a problem with our experimentally measured flux values, or perhaps with our stoichiometry.

Measuring external fluxes is easier than measuring internal fluxes, and being able to fully determine the system by only measuring these external fluxes would be a useful property. However, whilst we were able to fully determine the glycolysis system using only measured external flux values, this is often not the case.

2.2.2 When external flux knowledge is not enough

Examples such as the glycolysis system seen previously, where the system is fully determined by its external fluxes are, unfortunately, the exception rather than the rule. In general, with larger biological networks, we are unable to fully determine a system simply by measuring its external fluxes. We may also find a system that behaves in a way that, even if we should have enough external fluxes measured to theoretically fully determine the system, we are still unable to find a unique solution. One example of such a network is that shown in Figure 2.2. This network has an internal loop, in which $A \leftrightarrow B$ through two separate equations, as well as the formation of A and output of B. By analysing the reactions as described in Section 2.1.2, we are able to find the stoichiometric matrix for this network,

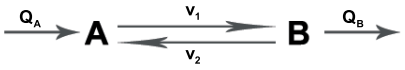

Figure 2.2: Network diagram for an example metabolic network featuring a feedback loop. In this network, A is transformed into B in one reaction, and B is in turn transformed back into A though another reaction.

$$S = \begin{pmatrix} -1 & 1 & 1 & 0 \\ 1 & -1 & 0 & -1 \end{pmatrix}. \tag{2.2.17}$$

The system, $S\mathbf{v} = \mathbf{0}$, has two equations for four unknown fluxes. We should be able to fully determine this system by measuring the two external fluxes. If we find $v_3 = a$ and $v_4 = b$, then we have the system of equations

$$-v_1 + v_2 + Q_A = 0 \tag{2.2.18}$$

A	$\xrightarrow{v_1}$	B
B	$\xrightarrow{v_2}$	A
External input	$\xrightarrow{Q_A}$	A
B	$\xrightarrow{Q_B}$	External output

Table 2.2: Set of reactions describing the example metabolic network seen in Figure 2.2. The reactions transforming A into B, and B into A are considered separate irreversible reactions, rather than a single reversible reaction.

$$v_1 - v_2 - Q_B = 0 \tag{2.2.19}$$
$$Q_A = a \tag{2.2.20}$$
$$Q_B = b \tag{2.2.21}$$

Then we have $v_2 - v_1 = -a$ and $v_1 - v_2 = b$. Since $v_2 - v_1 = -(v_1 - v_2)$, we have $a = b$. As long as $a = b$, these two equations are always true, and there are infinitely many solutions. If $a \neq b$ then the system of equations is inconsistent and permits no solutions. In particular, if we wish to have a fully determined system of linear equations, we need as many linearly independent equations as we have variables. We see that, in the case that $a = b$, summing equations (2.2.18) and (2.2.19), to find $Q_A - Q_B = 0$, and substituting in the value $Q_A = a$ from equation (2.2.20), we have exactly equation (2.2.21). This means we only have three linearly independent equations, and our system remains underdetermined. With only information about the external fluxes, we are unable to fully determine this system to find a unique solution. Specifically, we are unable to determine the values of the fluxes in the loop. Internal loops like the one seen in this are common in metabolic networks, so we cannot rely on MFA to fully determine systems when we only have data for the external flux values. Either we need to use data of internal flux values to fully determine the system, or find some other method of extracting relevant solutions from our stoichiometric data. Since obtaining data on internal reactions may not be feasible, we should instead look for other methods, such as flux balance analysis.

2.3 Flux balance analysis (FBA)

2.3.1 Introduction to flux balance analysis

Flux balance analysis (FBA) follows the same principles as the previously examined MFA. We begin with a stoichiometric matrix describing a metabolic network, and assume mass balance for this network giving us the equations $S\mathbf{v} = \mathbf{0}$, where S is the stoichiometric matrix, and $\mathbf{v}$ is the flux vector. Since this system is, again, likely to be underdetermined, and permitting infinitely many solutions, we must seek some way to extract relevant possible solutions. The idea of using linear programming to seek optimal product yields was first proposed by Papoutsakis [54], with regards to fermentation in butyric acid producing bacteria. Later, this

idea was also used by Fell and Small [21] to study fat synthesis. The technique we use in the coming section was first laid out by Varma et al. [74] and was initially known as stoichiometric flux balancing, but is now known as flux balance analysis (FBA) [33]. Whilst with MFA we sought to find a unique solution by taking measured values of fluxes, with FBA we wish to find the solution that maximises a specific objective function. Most commonly, this objective function is some linear combination of fluxes from the flux vector, i.e. of the form $\mathbf{a}^T\mathbf{v}$. This combination is typically chosen to represent some specific biological process, such as maximising the production of biomass (which leads to growth), or some other useful product, such as ATP (used for providing energy in reactions).

2.3.2 Glycolysis example

To illustrate this method, we re-examine the glycolysis example seen in Section 2.2.1. Since the key product in the glycolysis network is pyruvate, it would make sense to choose our objective function to be maximising the flux that produces this, v_5. Then, with $\mathbf{v}$ as given in equation (2.2.1), we have $\mathbf{a} = (0, 0, 0, 0, 1, 0, 0)$, which we wish to maximise. Along with our mass balance constraint, $S\mathbf{v} = \mathbf{0}$, we form a linear programming problem,

$$\max_{\mathbf{v}} \quad v_5 \quad \text{such that: } S\mathbf{v} = \mathbf{0}, \tag{2.3.1}$$

where S is the stoichiometric matrix seen in equation (2.1.1), and $\mathbf{v}$ is the flux vector given in equation (2.2.1). This optimisation problem, however, is not enough to find a relevant solution. Whilst there are feasible solutions, it is not possible to find an optimum as the problem is not bounded above. We could say that the optimal solution is to have all fluxes be infinite, but this is not biologically relevant. In order to restrict our optimisation problem to find relevant solutions we require inequality constraints on some or all of our fluxes. In general, we place constraints all our fluxes, but, when we find a solution, we often see that only a select few take values on their boundary, meaning we do not necessarily require inequality constraints for all our fluxes. Inequality constraints are also required to maintain biological relevance in the case of irreversible reactions. If a reaction is irreversible, we cannot have a negative flux through it, so we set the constraint that it must be greater than or equal to zero.

By looking at out network, we see that for steady-state, we need all fluxes in the same direction, so the question of reversibility is not particularly important for this example. We consider all fluxes to be irreversible, so that they take a minimum value of zero, and have a maximum value of one. By adding this set of constraints to the linear programming problem seen in equation (2.3.1), we form the full FBA problem,

$$\begin{gathered}\max_{\mathbf{v}} \quad v_5 \quad \text{such that: } S\mathbf{v} = \mathbf{0}, \\ 0 \leq v_1, v_2, v_3, v_4, v_5, Q_G, Q_P \leq 1.\end{gathered}$$

As described in Chapter 1, we can use Farkas' Lemma to determine if such an optimisation problem has solutions. Assuming solutions exist, the problem can be solved using simplex methods, or using numerical algorithms such as the linprog function in Matlab leading to the solution $\mathbf{v}^T = (0.5, 0.5, 0.5, 0.5, 1, 0.5, 1)$. In

this solution only v_5 and Q_P have reached their maximum possible values. We conclude that, under these conditions, the reactions relating to these fluxes are the rate-limiting steps in the system. As Q_P is simply a process by which pyruvate is removed, and in this example, not a biological one, we discount it somewhat and consider only v_5 as the rate-limiting step. If we increased the maximum permitted values for v_5 and Q_P, the fluxes for the rest of the system would also increase up until the point when they also reached their maxima and became rate-limiting steps.

After performing FBA for this particular network with varied maximum values for fluxes, we note that every solution is simply a linear multiple of previously found solutions. In particular, they are all of the form seen in equation (2.2.16), which is the solution found for MFA. In simple straight line networks lacking branches, it is fairly obvious that this would occur. We see that the mass balance equations for this network, (2.2.2)–(2.2.7), simply reduce to $Q_G = v_1 = v_2 = v_2 = v_3 = v_4$, and $2v_4 = v_5 = Q_P$. For one-to-one reactions such as these, the mass balance equations always reduce in this way. In a larger network it would be possible, and in many cases beneficial, to simplify large chains like this down to a single reaction. In this case, we would represent it as a reaction that produces half as much pyruvate as it uptake glucose, i.e. the reaction

$$2\,Glucose \rightarrow 1\,Pyruvate. \tag{2.3.2}$$

We should also note that it is possible to choose our minima and maxima in such a way as to provide the system with no solutions. For example, if we set the minimum flux for v_1 to be 1, but the maximum flux for v_2 to be less than 1, there would be no way to satisfy the mass balance whilst adhering to these constraints, yielding no viable solutions. Whilst for a simple network like this, it would be obvious when there are conflicting constraints, with larger networks, such conflicts could be far less obvious, so care must always be taken when placing these constraints on fluxes. One final thing to note about this network is that changing the objective function has very little effect on the solution provided. As long as we attempt to maximise some flux, or positive sum of fluxes, we always find the same solution. In the case that we attempt to minimise a flux, then we obtain the solution in which all fluxes are zero.

2.3.3 Moving from an ODE model to an FBA model

Until now we have built our stoichiometric matrices simply from the list of reactions in a system, however it is also possible to generate this matrix from an ODE system. This is also useful to examine, as it allows us to directly compare aspects of our FBA models with corresponding parts of an ODE model. To do this, we first formulate an ODE model for our glycolysis network seen in Figure 2.1. We know the reactions that are included in our model, from Table 2.1, so all we need is some way to model the reaction rates. Since all of the reactions in this network are enzyme catalysed, it is appropriate to use Michaelis-Menten kinetics to model them. In this model we use c_X (X $\in$ [A,B,C,D,E,F]) to represent concentrations of metabolites with the notation

$$[\text{Glucose}] = c_A,$$

$$
\begin{aligned}
[\text{Glucose 6-phosphate}] &= c_B, \\
[\text{Fructose 6-phosphate}] &= c_C, \\
[\text{Fructose 1,6-biphosphate}] &= c_D, \\
[\text{Phosphoglyceraldehyde}] &= c_E, \\
[\text{Pyruvate}] &= c_F. \qquad (2.3.3)
\end{aligned}
$$

We have several parameters in this model. Here $V_{X_{max}}$ refers to the maximum reaction rate for the reaction involving X, whilst K_{X_m} is the Michaelis coefficient for the reaction involving X. There are also two reactions transporting metabolites in and out of the system. These are simply labelled as $f_{in}(c_A)$ for the uptake of glucose and $f_{out}(c_F)$ for the output of pyruvate. Then our ODE model is

$$
\begin{aligned}
\frac{dc_A}{dt} &= f_{in}(c_A) - \frac{V_{Amax}c_A}{K_{Am} + c_A}, \qquad (2.3.4) \\
\frac{dc_B}{dt} &= \frac{V_{Amax}c_A}{K_{Am} + c_A} - \frac{V_{Bmax}c_B}{K_{Bm} + c_B}, \qquad (2.3.5) \\
\frac{dc_C}{dt} &= \frac{V_{Bmax}c_B}{K_{Bm} + c_B} - \frac{V_{Cmax}c_C}{K_{Cm} + c_C}, \qquad (2.3.6) \\
\frac{dc_D}{dt} &= \frac{V_{Cmax}c_C}{K_{Cm} + c_C} - \frac{V_{Dmax}c_D}{K_{Dm} + c_D}, \qquad (2.3.7) \\
\frac{dc_E}{dt} &= 2\frac{V_{Dmax}c_D}{K_{Dm} + c_D} - \frac{V_{Emax}c_E}{K_{Em} + c_E}, \qquad (2.3.8) \\
\frac{dc_F}{dt} &= \frac{V_{Emax}c_E}{K_{Em} + c_E} - f_{out}(c_F). \qquad (2.3.9)
\end{aligned}
$$

For this set of ODEs, we assume that there exists a vector of concentrations, $\mathbf{c}^* = (c_A^*, c_B^*, c_C^*, c_D^*, c_E^*, c_F^*)$ in which the system attains a steady-state. That is, for which all time derivatives are equal to zero. By substituting these values in, we obtain the equations

$$
\begin{aligned}
0 &= f_{in}(c_A*) - \frac{V_{Amax}c_A^*}{K_{Am} + c_A^*}, \qquad (2.3.10) \\
0 &= \frac{V_{Amax}c_A^*}{K_{Am} + c_A^*} - \frac{V_{Bmax}c_B^*}{K_{Bm} + c_B^*}, \qquad (2.3.11) \\
0 &= \frac{V_{Bmax}c_B^*}{K_{Bm} + c_B^*} - \frac{V_{Cmax}c_C^*}{K_{Cm} + c_C^*}, \qquad (2.3.12) \\
0 &= \frac{V_{Cmax}c_C^*}{K_{Cm} + c_C^*} - \frac{V_{Dmax}c_D^*}{K_{Dm} + c_D^*}, \qquad (2.3.13) \\
0 &= 2\frac{V_{Dmax}c_D^*}{K_{Dm} + c_D^*} - \frac{V_{Emax}c_E^*}{K_{Em} + c_E^*}, \qquad (2.3.14) \\
0 &= \frac{V_{Emax}c_E^*}{K_{Em} + c_E^*} - f_{out}(c_F^*). \qquad (2.3.15)
\end{aligned}
$$

It would be possible, at this point, to place constraints on our parameters and steady-state concentrations, and then perform an optimisation problem analogous to the one used in FBA, however since these functions are non-linear, we would require non-linear programming to find an optimal value, and this is undesirable. Instead, we also note that each Michaelis-Menten term is a constant, and could

be replaced by a single flux value, like those seen in our earlier mass balance equations. For example, in equation (2.3.14), we replace the two non-linear terms with

$$\frac{V_{Dmax}c_D^*}{K_{Dm}+c_D^*} = v_4, \tag{2.3.16}$$

$$\frac{V_{Emax}c_E^*}{K_{Em}+c_E^*} = v_5, \tag{2.3.17}$$

so that the equation becomes $0 = 2v_4 - v_5$. By replacing each flux term in this manner, we obtain exactly the set of equations (2.2.2)–(2.2.7), which is our set of mass balance constraints.

From here, we set the minimum and maximum values for the fluxes as usual. Notably, if we had $V_{X_{max}}$ parameters for our initial ODEs, we could use these as the maximum values for our fluxes in FBA, as in Michaelis-Menten kinetics, these represent the maximum possible flux value. Finally, we choose an objective function for this optimisation. In this way, we have formed an FBA model and it is clear now how a flux balance model relates to an ODE model for the same network.

2.3.4 FBA in terms of the ODEs

When we perform FBA, we are finding an optimal steady-state for our metabolic network. Having clearly illustrated the link between an ODE model and the stoichiometric matrix used in FBA, it would be worth attempting to use equivalent optimisation methods on the steady-state of our ODE model, seen in equations (2.3.10)–(2.3.15) to try and achieve the same results. The method of Lagrange multipliers could be used to find the concentrations that correspond to the steady-state in which a flux (or fluxes) specified by an objective function are maximal. For now we assume all parameters in the equations to be constant, and we maximize across the concentrations only. For our glycolysis example, we can set up this problem with relative ease. First we choose an objective function to maximise. We choose the same objective function as in our previous flux balance mode seen in equation (2.3.2), v_5. In our ODE model, we have that v_5 is equal to $\frac{V_{Emax}c_E}{K_{Em}+c_E}$, so we choose that to be our objective function. For constraints, we simply use the steady-state equations for our ODE, already derived in equations (2.3.10)–(2.3.15). For ease of notation, from here we denote the Michaelis-Menten functions as

$$g_X(c_X) = \frac{V_{Xmax}c_X}{K_{Xm}+c_X}, \tag{2.3.18}$$

for $X \in (A, B, C, D, E, F)$.

Then we wish to maximize $g_E(E)$, subject to the constraints

$$f_{in}(A) - g_A(c_A) = 0, \tag{2.3.19}$$

$$g_A(c_A) - g_B(c_B) = 0, \tag{2.3.20}$$

$$g_B(c_B) - g_C(c_C) = 0, \tag{2.3.21}$$

$$g_C(c_C) - g_D(c_D) = 0, \tag{2.3.22}$$

$$2g_D(c_D) - g_E(c_E) = 0, \tag{2.3.23}$$

$$g_E(c_E) - f_{out}(F) = 0. \tag{2.3.24}$$

Using the method of Lagrange multipliers, we find a solution to the problem where

$$\nabla\Lambda(A, B, C, D, E, F, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6) = 0. \tag{2.3.25}$$

The lagrangian is

$$\begin{aligned}&\Lambda(A, B, C, D, E, F, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6) = \\ &g_E(c_E) + \lambda_1(f_{in}(A) - g_A(c_A)) + \lambda_2(g_A(c_A) - g_B(c_B)) + \lambda_3(g_B(c_B) - g_C(c_C)) + \\ &\lambda_4(g_C(c_C) - g_D(c_D)) + \lambda_5(2g_D(c_D) - g_E(c_E)) + \lambda_6(g_E(c_E) - f_{out}(F)),\end{aligned} \tag{2.3.26}$$

and the operator, ∇, is

$$\nabla = (\partial c_A, \partial c_B, ..., \partial c_F, \partial\lambda_1, \partial\lambda_2, ..., \partial\lambda_6). \tag{2.3.27}$$

This problem should theoretically lead us to the same solutions that we found by the FBA model shown in equations (2.3.2). Without attempting to solve the problem presented, it is clear that whilst $f_E(E)$ has a theoretical maximum of V_{Emax}, this is only achieved at an infinite concentration of E, regardless of all other values, which is not biologically feasible. We would also need to consider maximum concentration values in order to find useful solutions using this method. We also note that the minimum solution, in which all concentrations (and therefore fluxes) are equal to zero would also be a valid solution to equation (2.3.25), so we would need to be careful that we are indeed obtaining the required maximum.

In the problem formulated in equation (2.3.25), we attempt to maximise the flux v_5 by finding optimal concentrations and treating our parameters as constant. Effectively we have found a steady-state for a system defined uniquely by the parameters in the ODEs, though in this case it requires infinite concentrations. It may also be of interest to perform our maximisation from an alternative perspective. We seek optimal flux distributions whilst treating the concentrations as constant and maximising over the parameters. We set up another Lagrange multipliers problem to find solutions in this way, in which Λ becomes a function of the parameters V_{Xmax} and K_{Xm} rather than the concentrations.

In the example presented, this maximisation is trivial. We have 2 parameters for each flux, V_{Xmax} and K_{Xm}. In order to maximise a specific flux, we simply set the V_{Xmax} to be arbitrarily high, and the K_{Xm} to be arbitrarily close to zero. As in the previous problem, and indeed in our FBA models, we see that we would need to set maximum values for our parameters to prevent infinite fluxes.

In actuality, it is not possible to separate the concentrations and parameters in FBA in this way, and the optimisation would instead need to be considered as acting on both. For our glycolysis example, this gives us ten parameters and six concentrations to maximise over (not including any parameters that might be included in our general functions for the input and output fluxes). This starts to become very difficult to solve using Lagrange methods for optimisation. Indeed, one of the strengths of FBA is that it turns these complicated non-linear maximisation problems into much simpler linear programming problems that are, computationally, far easier to solve. Having now thoroughly examined this straight line glycolysis system, we next move to examine a metabolic network with two branches.

A	$\xrightarrow{v_1}$	B
A	$\xrightarrow{v_2}$	C
External input	$\xrightarrow{Q_A}$	A
B	$\xrightarrow{Q_B}$	External output
C	$\xrightarrow{Q_C}$	External output

Table 2.3: Set of reactions describing the example metabolic network seen in Figure 2.3. There metabolite A is transformed into either B or C in separate reactions.

2.4 A branching network

Consider a network in which a single metabolite A, is changed into B or C, depending on which reaction it undergoes, as shown in Figure 2.3, with the reactions listed in Table 2.3. This is a simple example of a branching network. To form an

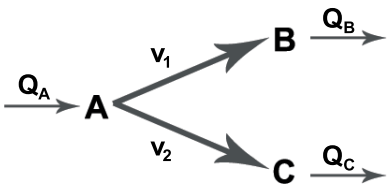

Figure 2.3: Network diagram for the branching network, in which A is transformed into either B or C.

ODE model for this network, we could use Michaelis-Menten kinetics to model the reaction rates, as we did in the prior example, or we could instead use the law of mass action for a simpler set of ODEs. For now, we refer to the fluxes as general functions of the concentrations, as the exact kinetics are not required to form FBA models. The system of ODEs is given as

$$\frac{dc_A}{dt} = -v_1(c_A) - v_2(c_A) + v_{in}, \tag{2.4.1}$$

$$\frac{dc_B}{dt} = v_1(c_A) - v_{B-out}, \tag{2.4.2}$$

$$\frac{dc_C}{dt} = v_2(c_A) - v_{C-out}, \tag{2.4.3}$$

where v_1 and v_2 are functions that depend on the concentration of A, and v_{in}, v_{B-out} and v_{C-out} are external reactions that add A and remove the products B and C respectively. These could be constants or also functions of concentrations.

2.4.1 FBA

To perform FBA on this network, we find the stoichiometric matrix, whether by inspection of the network, or by converting the ODE model (2.4.1)–(2.4.3)

through the process described in Section 2.3.3. Either method produces the same stoichiometric matrix,

$$S = \begin{pmatrix} -1 & -1 & 1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \end{pmatrix}, \tag{2.4.4}$$

for the flux vector, $\mathbf{v} = (v_1, v_2, Q_A, Q_B, Q_C)$. For the purposes of FBA, we constrain all the fluxes to be between zero and one. It is worth considering two separate objective functions for this system. In the first we maximise flux towards B and in the second we maximise flux towards C. In the first case, we maximise v_1, and in the second we maximise v_2. Then we formulate our linear programming problem

$$\begin{aligned} \max_{\mathbf{v}} \quad \mathbf{a}^T\mathbf{v} \quad \text{such that:} \quad & -v_1 - v_2 + v_3 = 0, \\ & v_1 - v_4 = 0, \\ & v_2 - v_5 = 0, \\ & 0 \leq v_1, v_2, Q_A, Q_B, Q_C \leq 1. \end{aligned} \tag{2.4.5}$$

Here, the first three constraints are the expanded form of $S\mathbf{v} = \mathbf{0}$, where S is the stoichiometric matrix given in equation (2.4.4), and $\mathbf{v} = (v_1, v_2, Q_A, Q_B, Q_C)$. When maximising v_1, we use $\mathbf{a}^T = (1, 0, 0, 0, 0)$, and when maximising v_2 we use $\mathbf{a}^T = (0, 1, 0, 0, 0)$. The unique solutions to these FBA problems are

$$\begin{array}{rccccc} & v_1 & v_2 & Q_A & Q_B & Q_C \\ \mathbf{v}_B = & (1, & 0, & 1, & 1, & 0\ ), \\ \mathbf{v}_C = & (0, & 1, & 1, & 0, & 1\ ), \end{array} \tag{2.4.6}$$

where $\mathbf{v}_B$ is the solution maximising flux towards B, and $\mathbf{v}_C$ is the solution maximising flux towards C. In the case of maximising flux towards B, we see that v_2 has been completely inhibited in order to maximise flux through v_1. In the case of maximising flux towards C, the opposite has happened, and v_1 has been inhibited to maximise flux through v_2. It is obvious to see that any flux distribution that satisfies steady-state conditions must be some linear combination of these two solutions. This example highlights two things. Firstly, the results are heavily dependent on the objective function chosen by the modeler, and are subject to their bias. To illustrate the second, we consider the ODEs, (2.4.1)–(2.4.3), again.

2.4.2 Effects on the ODEs

Let us examine in more detail the first state, in which we maximised v_1 at the expense of v_2. Equation (2.4.1) tells us how the concentration of A changes in time. If we try and fit our FBA solution into these equations, we find that in the steady-state, we have $0 = -v_1(c_A) - v_2(c_A) + v_{in}$, where $v_{in} = 1$, $v_1 = 1$ and $v_2 = 0$ as given by FBA. We see potential difficulties here with fluxes being completely inhibited by FBA. We would require some value of c_A such that both $v_1(c_A) = 1$ and $v_2(c_A) = 0$. Let us first consider the law of mass action. In this case we have $v_1 = k_1 c_A = 1$ and $v_2 = k_2 c_A = 0$. It is clear that in order for both

these equations to hold true, we must have $k_2 = 0$. In other words, the second reaction must be completely inhibited at all times in order for it to have zero flux, when the first reaction has a non-zero flux. We could also consider modelling the fluxes with Michaelis-Menten kinetics. In this case, we would have

$$v_1 = \frac{V_{max_1} c_A}{K_1 + c_A}, \tag{2.4.7}$$

$$v_2 = \frac{V_{max_2} c_A}{K_2 + c_A}. \tag{2.4.8}$$

Here we find similar results. In order for the second flux, v_2 to be zero when v_1 is non-zero, we require V_{max_2} to be equal to zero.

Flux balance analysis has the potential to provide us with solutions where certain reactions are completely inhibited, but biologically we know that they cannot be. They may be required to provide other products that the organism needs to grow or survive. In cases like these, the obvious solution for FBA would be to have a non-zero lower bound for fluxes in branching paths, but it may be very difficult to provide one which is suitable. We also note that if we are performing FBA and considering it to only be acting on the concentrations of metabolite, and not on the parameters within an ODE system, we may stumble across biologically impossible solutions. For our simple branching problem, if we found a solution where $v_1 = 0.6$ and $v_2 = 0.4$, for example, there is no guarantee that there exists a corresponding c_A in an ODE system that would provide the same solution.

Examining this branching network has provided two key points we must be aware of when using FBA. The first is that different objective functions lead to different solutions. Since the choice of objective function has such a large effect on the results provided, we must be careful to choose one that provides relevant results. Schuetz et al. [67] evaluated a range of different objective functions for the same network in an attempt to find one that provided the results most accurate to experimental data, however it should be noted that there is no single objective function that provides the best results in every situation.

Secondly, we learned that the results provided by FBA may not be feasible in an equivalent ODE model. In particular FBA provides results that would require the complete inhibition of reactions, which may not always be biologically feasible. In many cases, the theoretical optimum values for objective functions provided by FBA will not be possible in a real organism.

Now that we have examined a system with a branch, we next examine a network containing a loop. That is a network containing a reaction that transforms one metabolite, A, into another metabolite B, but also contains a second reaction transforming B back into A. We have already seen such a network in our MFA analysis seen in Section 2.2.2. Rather than examining this same system again, we consider a network with the additional feature of combining two metabolites.

2.5 A network with a feedback loop

Consider the network shown in Figure 2.4. Here we have a system which uptakes a source of A. This A is then transformed into B, and B is, in turn, transformed back into A, creating a loop. Additionally the two metabolites, A and B are

combined to form C. Finally this C is output. These reactions are given in Table 2.4.

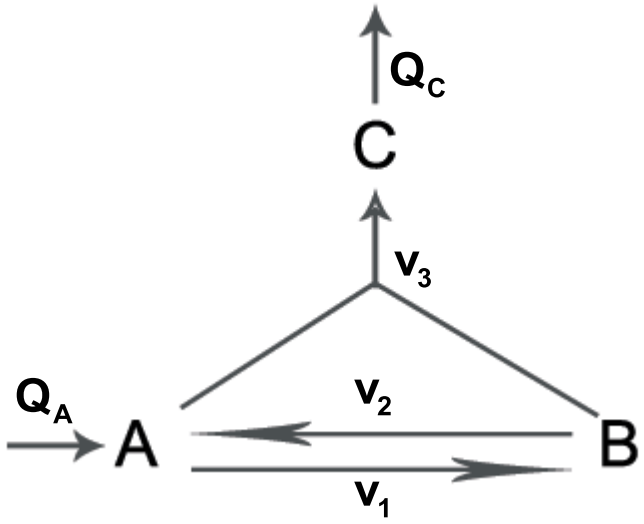

Figure 2.4: Network diagram for the looping network in which A is transformed into B. The two are then combined to form C, which is then outputted through a separate reaction.

We form an ODE model for this network. External reactions are given as general functions, f_{in} and f_{out}. Whilst remaining reaction rates are modelled using the law of mass action. As we have three metabolites, our ODE model consists of three equations,

$$\frac{dc_A}{dt} = f_{in}(c_A) - k_1 c_A + k_2 c_B - k_3 c_A c_B, \tag{2.5.1}$$

$$\frac{dc_B}{dt} = k_1 c_A - k_2 c_B - k_3 c_A c_B, \tag{2.5.2}$$

$$\frac{dc_C}{dt} = k_3 c_A c_B \ - f_{out}(c_C). \tag{2.5.3}$$

2.5.1 FBA

Suppose we wish to find the steady-state that maximises the production of C, that is, maximising the flux v_3. To do this, we perform FBA on the network.

A	$\xrightarrow{v_1}$	B
B	$\xrightarrow{v_2}$	A
A+B	$\xrightarrow{v_3}$	C
External input	$\xrightarrow{Q_A}$	A
C	$\xrightarrow{Q_C}$	External output

Table 2.4: Set of reactions describing the example metabolic network seen in Figure 2.4.

First we form the stoichiometric matrix

$$S = \begin{pmatrix} -1 & 1 & -1 & 1 & 0 \\ 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 \end{pmatrix}. \tag{2.5.4}$$

The flux vector for this matrix is $\mathbf{v} = (v_1, v_2, v_3, Q_A, Q_C)^T$. Next we define an objective function, $\mathbf{a}^T\mathbf{v} = v_3$, where v_3 is the flux in to C that we wish to maximise. As in the previous examples, we constrain all fluxes to be between zero and one. We now formulate the linear programming problem for FBA

$$\max_{\mathbf{v}} \quad v_3 \quad \text{such that:} \quad -v_1 + v_2 - v_3 + Q_A = 0, \tag{2.5.5}$$

$$v_1 - v_2 - v_3 = 0,$$

$$v_3 - Q_C = 0,$$

$$0 \leq v_1, v_2, v_3, Q_A, Q_C \leq 1. \tag{2.5.6}$$

One possible solution for this optimisation problem is $\mathbf{v} = (0.75, 0.25, 0.5, 1.0.5)$. However, we find that this solution is not unique. Indeed, other solutions, such as $\mathbf{v}_2 = (0.5, 0, 0.5, 1, 0.5)$, also satisfy the optimisation problem. The values for v_3, Q_A and Q_C remain fixed across all solutions, but there is variation in v_1 and v_2. We would like to see how much these two fluxes are able to vary, or indeed, if any of the other fluxes are variable, whilst still providing a solution to the optimisation problem (2.5.6).

2.5.2 Flux variability analysis

Flux variability analysis (FVA), first set out by Mahadevan et al. [44], is a technique to investigate these variations in solutions to FBA problems. By performing a secondary optimisation problem, we find the range of variation in a flux. Specifically, we learn how much a flux can be varied whilst still maintaining an optimal steady-state (according to our objective function). FVA is a two step process. In the first step, we solve an FBA problem to find an optimal value for some objective function, so that $\mathbf{a}^T\mathbf{v} = A_{max}$. In the second step, we form a new optimisation problem, with the additional constraint that $\mathbf{a}^T\mathbf{v} = A_{max}$. Then all solutions of this new optimisation problem maximise the original objective function. The objective function for the new optimisation problem is to maximise or minimise a specific flux, so that $\mathbf{b}^T\mathbf{v}$ is a single flux from our flux vector. By repeating this step for each individual flux, we see the range of variability allowed for each flux whilst still maintaining an optimal solution for our initial FBA problem. This process is summarised as

Maximise $\mathbf{a}^T\mathbf{v}$ such that: $S\mathbf{v} = \mathbf{0}$,
$\mathbf{v}_{min} \leq \mathbf{v} \leq \mathbf{v}_{max}$.
$\max_{\mathbf{v}} \mathbf{a}^T\mathbf{v} = A_{max}$
for All elements of $\mathbf{v}, v_n$ **do**
    Maximise and minimise v_n such that: $S\mathbf{v} = \mathbf{0}$,
    $\mathbf{v}_{min} \leq \mathbf{v} \leq \mathbf{v}_{max}$,
    $\mathbf{a}^T\mathbf{v} = A_{max}$
end for

For our looping network, the maximal value for our initial objective function is $v_3 = 0.5$. Then by forming four additional optimisation problems with $v_3 = 0.5$ as an additional constraint, we find the maximal and minimal values that v_1 and v_2 are able to take whilst still maintaining this maximal production of C. These optimisation problems are of the form

$$\begin{aligned} \underset{\mathbf{v}}{max/min} \quad v_1/v_2 \quad \text{such that:} \quad & -v_1 + v_2 - v_3 + Q_A = 0, \\ & v_1 - v_2 - v_3 = 0, \\ & v_3 - Q_C = 0, \\ & v_3 = 0.5, \\ & 0 \leq v_1, v_2, v_3, Q_A, Q_C \leq 1. \end{aligned} \tag{2.5.7}$$

Maximising either v_1 or v_2 both lead to the same solution, $\mathbf{v}^T = (1, 0.5, 0.5, 1, 0.5)$. Similarly, minimising either v_1 or v_2 both lead to the same solution where $\mathbf{v}^T = (0.5, 0, 0.5, 1, 0.5)$. We see that v_1 and v_2 have a variability range of 0.5, found by subtracting their maximal value from their minimal one. We also apply FVA for the three other fluxes, v_3, Q_A, and Q_C, but find that their range of variabilities are all zero.

The second constraint in our optimisation problem, (2.5.6), is $v_1 - v_2 - v_3 = 0$. Rearranged, this gives $v_1 - v_3 = v_2$. In our optimal solutions, v_3 is always equal to 0.5, so we substitute that value in to find $v_1 - 0.5 = v_2$. That is, as long as v_1 is 0.5 larger than v_2, we have an optimal steady-state. Combining this with the fact that no fluxes other than v_1 or v_2 can vary in an optimal state, we are able to write down the general form of a solution to the FBA problem (2.5.6),

$$\mathbf{v}^T = (0.5, 0, 0.5, 1, 0.5) + a(1, 1, 0, 0, 0), \tag{2.5.8}$$

where $0 \leq a \leq 0.5$ (given by the variability range of v_1 and v_2). FBA alone is not able to give us a value for a, as we any solution of this form satisfies the optimisation problem.

We find now that, even among our optimal solutions according to FBA, we still have some degeneracy, allowing multiple solutions to attain the same maximal value for the objective function. This means that, even with FBA, we may not be able to find a unique steady-state solution. By applying additional techniques to explore these alternative optimal solutions, such as FVA, we are still able to quantify the behaviour of the network in an optimal state. In particular we are able to write the flux vector $\mathbf{v}$ as some positive sum of other flux vectors, in order to attain a general solution to the FBA problem. Every solution of this form provides an optimal value for the objective function. We are always able to find a general form for the optimal solution in this way, though with larger networks, simply doing this by inspection proves difficult. In the next section we examine a network where the general behaviour is not immediately apparent.

2.6 A network with multiple combining steps

Consider another network shown in the network diagram in Figure 2.5. This example has some similarities to our previous looping example, where we had

A	$\xrightarrow{v_1}$	B
B	$\xrightarrow{v_2}$	C
A+C	$\xrightarrow{v_3}$	D
B+D	$\xrightarrow{v_4}$	E
External input	$\xrightarrow{Q_A}$	A
E	$\xrightarrow{Q_E}$	External output

Table 2.5: Set of reactions describing the example metabolic network seen in Figure 2.5. Here we have two reactions combining metabolites, v_3 and v_4.

two metabolites combining to form a product, however here there are two stages in which metabolites combine, and no feedback loops. The reactions for this network are given in Table 2.5. As before we shall examine the behaviour of this system in a steady-state by performing FBA. In this case, it is not immediately obvious how the fluxes are distributed in a steady-state, so FBA should provide some useful insight.

The stoichiometric matrix S for this network is

$$S = \begin{pmatrix} -1 & 0 & -1 & 0 & 1 & 0 \\ 1 & -1 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \end{pmatrix}, \tag{2.6.1}$$

corresponding to the flux vector, $\mathbf{v} = (v_1, v_2, v_3, v_4, Q_A, Q_E)$. Seeking to maximise the production of E, via the flux combining C and D, we choose the objective function $\mathbf{a}^T\mathbf{v} = v_4$. Combining our mass balance constraint $S\mathbf{v} = \mathbf{0}$ and our usual minimum and maximum flux values (zero and one respectively) with this objective function gives us an FBA problem,

$$\begin{aligned} \max_{\mathbf{v}} v_4 \quad \text{such that:} \quad & -v_1 - v_3 + Q_A = 0, \\ & v_1 - v_2 - v_3 = 0, \\ & v_2 - v_3 = 0, \end{aligned}$$

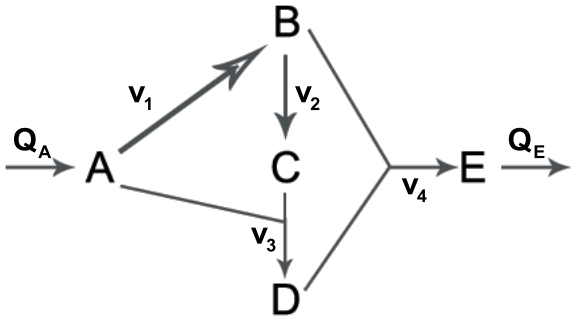

Figure 2.5: Network diagram for a more complex branching network.

$$
\begin{aligned}
v_3 - v_4 &= 0, \\
v_4 - Q_E &= 0, \\
0 \leq v_n, Q_n &\leq 1.
\end{aligned} \tag{2.6.2}
$$

Solving this optimisation problem, we find a solution where

$$\mathbf{v} = (2/3, 1/3, 1/3, 1/3, 1, 1/3).$$

Performing FVA on this network, as described in Section 2.5.2, we find that every flux has a variability range of zero. That is, we cannot vary any flux in this network from these values whilst still maintaining a state that maximises v_4. Therefore, this solution is unique. Since this network is more complex than previous examples, we cannot see so easily if other objective functions could potentially yield different optimal flux vectors. For previous networks, we were able to write down a general form of steady-state flux vectors for a network and we would like to be able to do the same for this network.

2.6.1 Elementary Mode Analysis

In previous networks, we have found general ways of writing down possible flux distributions for the examined networks in steady-states in terms of simpler flux distributions. In our first FBA example in Section 2.3, we found that all solutions were simply multiples of the first solution we found. For our system with an internal loop, in Section 2.5, we were able to split the solutions into a path through the system and an internal loop. In the branching example seen in Section 2.4, flux distributions could trivially be written as the sum of the fluxes through each branch. This is known as elementary modes analysis, as described by Schuster et al. [68]. By decomposing solutions into these "elementary modes", we gain additional knowledge of how metabolic networks behave in steady-state.

In general a flux vector $\mathbf{v}$ categorises its elements as either reversible or irreversible. Whilst reversible reactions have negative flux values, irreversible reactions must maintain a flux value $v_n \geq 0$. These are known as reversibility conditions. Every flux vector $\mathbf{v}$ that satisfies these reversibility conditions is an allowable flux state for a network. A flux vector $\mathbf{v}$ that satisfies these reversibility conditions and also satisfies $S\mathbf{v} = \mathbf{0}$, i.e. satisfies the mass balance condition, is known as a flux mode. Flux modes represent possible steady-state flux distributions for a network. A solution to a FBA problem is a flux mode, but not all flux modes are solutions to FBA problems. Often a flux mode is able to be written as the positive sum of two or more other modes that are not scalar multiples of each other, that is, it can be written as $\sum_i a_i \mathbf{v}_i$, where $\mathbf{v}_i$ is another flux mode, and a_i are coefficients. These coefficients must be positive, that is we must have $a_i \geq 0 \, \forall i$. This is to preserve the idea of irreversibility in reactions, and allows us to keep a biological relevance. A mode which can be written as the sum of other modes in this way is called decomposable. Some modes are not decomposable. These modes are referred to as "elementary modes". In the case of the branching network described in Section 2.4, there are two elementary modes,

$$\mathbf{m}_1 = (1, 0, 1, 1, 0)^T, \tag{2.6.3}$$

$$\mathbf{m}_2 = (0, 1, 1, 0, 1)^T, \tag{2.6.4}$$

in which the first mode, $\mathbf{m}_1$, represents the flux through the top branch, and $\mathbf{m}_2$ represents the flux through the second branch.

Every metabolic network has a finite set of elementary modes. For a given stoichiometric matrix and set of reversibility conditions, we form a set of elementary modes that are found from this network. These elementary modes satisfy both $S\mathbf{v} = \mathbf{0}$ and the irreversibility conditions. This set of elementary modes forms a spanning set for the solution space of a linear programming problem using these conditions. Every solution for any given objective function can be written as a positive sum of these elementary modes. This minimal set of elementary modes is unique up to scalar multiples, and all the elementary modes are called systematically independent, which means no elementary mode is able to be written as a positive sum of the other elementary modes. If it could, it would be considered decomposable and would, therefore, not be an elementary mode. However, whilst the elementary modes are systemically independent, they are not necessarily linearly independent. Whilst they do form a spanning set, it is not a necessary requirement that this set is also a basis. Often we find more elementary modes than would be required to form a basis. In the case of larger biological systems, it is common for there to be more elementary modes than there are fluxes within the network. Schuster and Hilgetag [69] showed that these elementary modes are analogous to a concept in linear programming known as extreme pathways. When we are performing linear programming, the constraints we add restrict our solution space to a polyhedral cone. This cone has a set of generating vectors called extreme pathways. These extreme pathways are the minimal set of systemically independent vectors that are used to describe this polyhedral cone. They are also unique up to scalar multiplication. It is more common to refer to extreme pathways as elementary modes when dealing with FBA.

Schilling et al. [65] described the method of using FBA to provide solutions in terms of these elementary modes. An earlier paper also laid out an algorithm to find the elementary modes for a network by Schilling et al. [66]. Applying this algorithm, manually or using software packages such as CellNetAnalyzer[38], to our current example described in equation (2.6.2) presents us with a single elementary mode, $\mathbf{m} = (2, 1, 1, 1, 3, 1)^T$. As this is the only elementary mode for this system, any solution to our flux balance analysis problem must be some multiple of this mode. That is, $\mathbf{v} = \lambda(2, 1, 1, 1, 3, 1)^T$ for some λ. Finding the elementary mode for this network has provided us with additional insight into how the network behaves in steady-state. We find that there is a single route metabolites take to get through the network, and the output of the system is only a third of the input. Looking towards optimising any flux in this network, we see that if we increase the flux in any step, we must also increase all the other fluxes proportionally in order to maintain our mass balance. For any objective function, the optimal solution always takes the same form, and there is no freedom of choice for our flux vectors.

Elementary modes analysis allows us to write general solutions to an FBA problem in terms of other non-decomposable solutions. These elementary modes can be seen as pathways, or routes, that flow takes through the network from inputs to outputs. By decomposing a solution into the sum of these modes we quickly gain an obvious physical interpretation of how the flux is distributed in

a particular solution. A decomposition of a solution may not be unique, but still provides a physical insight into the behaviour of a network in an optimal state.

2.7 A network with multiple branches

All the previous examples of metabolic networks we have studied share one thing in common. Each network contained only a single elementary mode that led from each input to each output. For the glycolysis example in Section 2.3.2, and the combining example in Section 2.6 this was a single mode leading from input to output. The looping example seen in Section 2.5 had a second mode containing the reactions in the loop, allowing some freedom in flux levels for the two reactions contained in it. Our branching example from Section 2.4 had a single mode to each of the two outputs. Let us now instead examine a network that has multiple inputs, multiple outputs, and multiple elementary modes leading from each input to each output. This network is shown in the network diagram, Figure 2.6, and the reactions are listed in Table 2.6. We form an ODE model for this

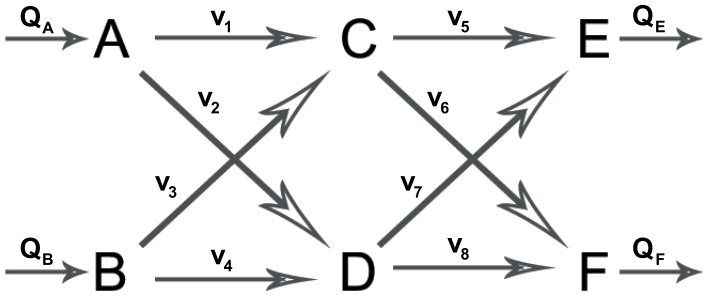

Figure 2.6: Network diagram for an example with multiple branches. Here we have two inputs, A and B, with two outputs, E, and F. Each intermediate metabolite, C, and D, is created by both input, and used to create either output.

network, using the law of mass action to model internal reaction rates. This model consists of six equations, where the variables c_X (X $\in$ [A,B,C,D,E,F]) represent the concentrations of the metabolites,

$$\frac{dc_A}{dt} = Q_A - k_1 c_A - k_2 c_A, \tag{2.7.1}$$

$$\frac{dc_B}{dt} = Q_B - k_3 c_B - k_4 c_B, \tag{2.7.2}$$

$$\frac{dc_C}{dt} = k_1 c_A + k_3 c_B - k_5 c_C - k_6 c_C, \tag{2.7.3}$$

$$\frac{dc_D}{dt} = k_2 c_A + k_4 c_B - k_7 c_D - k_8 c_D, \tag{2.7.4}$$

$$\frac{dc_E}{dt} = k_5 c_C + k_7 c_D - Q_E, \tag{2.7.5}$$

$$\frac{dc_F}{dt} = k_6 c_C + k_8 c_D - Q_F. \tag{2.7.6}$$

In all the previous networks we examined, we had a single uptake flux. In this example, we have two. If the maximum values of these uptake fluxes are constrained to be one, then the maximum overall uptake of the network is two. In

A	$\xrightarrow{v_1}$	C
A	$\xrightarrow{v_2}$	D
B	$\xrightarrow{v_3}$	C
B	$\xrightarrow{v_4}$	D
C	$\xrightarrow{v_5}$	E
C	$\xrightarrow{v_6}$	F
D	$\xrightarrow{v_7}$	E
D	$\xrightarrow{v_8}$	F
External input	$\xrightarrow{Q_A}$	A
External input	$\xrightarrow{Q_B}$	B
E	$\xrightarrow{Q_E}$	External output
F	$\xrightarrow{Q_F}$	External output

Table 2.6: Set of reactions describing the example metabolic network seen in Figure 2.6. Here we have two inputs and two outputs, and multiple reactions producing C, D, E, and F.

order to accommodate for this, where we would usually constrain the maximum values of our output fluxes to be one, we instead constrain them to be two. In this way, if we wish to maximise the production of either E or F, this optimal value is not limited by the maximum amount the network is able to output, allowing more freedom in our FBA solutions. Then our new maximum and minimum flux values are

$$\begin{aligned} 0 \leq v_n, Q_A, Q_B \leq 1, \\ 0 \leq Q_E, Q_F \leq 2. \end{aligned} \tag{2.7.7}$$

The stoichiometric matrix for this network is

$$S = \begin{pmatrix} -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & -1 \end{pmatrix}, \tag{2.7.8}$$

for the flux vector $\mathbf{v} = (v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, Q_A, Q_B, Q_E, Q_F)$. First, we consider maximising the flux in to E. Due to the symmetry of the system, we could also consider maximising in to F, but the results would be nearly identical. In this example, the flux towards E is the sum of two fluxes, v_5 and v_7, so our objective function becomes $\mathbf{a}^T\mathbf{v} = v_5 + v_7$. Then our FBA problem is

$$\begin{aligned} \max_{\mathbf{v}}(v_5 + v_7) \quad \text{such that: } S\mathbf{v} = \mathbf{0}, \\ 0 \leq v_n, Q_A, Q_B \leq 1, \end{aligned}$$

$$0 \leq Q_E, Q_F \leq 2. \tag{2.7.9}$$

Using the linprog function in Matlab, we find an example solution of

$$\mathbf{v} = (0.5, 0.5, 0.5, 0.5, 1, 0, 1, 0, 1, 1, 2, 0),$$

though we know this is not unique. In this solution, the flux between A and B and C and D is shared equally across all v_1, v_2, v_3 an v_4, then all the flux is directed from C and D towards E, with none of it going to F. Performing a further flux variability analysis upon these fluxes leads us to find that it is possible for v_1, v_2, v_3, and v_4 to be varied between zero and one whilst maintaining the maximised steady-state, however the remaining fluxes have zero variability whilst maintaining the maximised steady-state. Hence we see that it doesn't matter which route the metabolites take through the left half of the system, but in order to achieve a maximised steady-state, as much of the flux as possible must be sent to E rather than F in the right half.

We now examine the effects of flux balance analysis upon the ODE model of the system. Let us continue to examine the situation in which we wish to maximise the flux in to E. In this network, this means we wish to maximise $k_5 c_C + k_7 c_D$.

Rearranging equations (2.7.3) and (2.7.4) and setting the time derivatives to zero gives

$$k_1 c_A + k_3 c_B = c_C(k_5 + k_6) \tag{2.7.10}$$
$$k_2 c_A + k_3 c_B = c_D(k_7 + k_8). \tag{2.7.11}$$

From these equations we see that to increase $k_5 c_C$ we must either lower $k_6 c_C$ or increase $k_1 c_A + k_3 c_B$ to maintain mass balance. Similarly to increase $k_7 c_D$ we must either lower $k_8 c_D$ or increase $k_2 c_A + k_3 c_B$. Decreasing $k_6 c_C$ or $k_8 c_D$ is effectively redirecting the flux to F towards E. Increasing the other fluxes increases the overall flow through the system, and would also require us to increase the fluxes into the system. If we are unable to increase the uptake rates, then the only possible optimisation is redirecting the flux to F towards E instead.

2.7.1 Elementary modes

As in our previous example, in Section 2.6.1, we find the set elementary modes for this network using CellNetAnalyzer [38]. For the current network, seen in Figure 2.6, we find 8 such modes, each representing a different route through the system. Labelling these modes $\mathbf{m}_n$ for $n = 1$–8, we have

$$\mathbf{m}_1 = (1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0),$$
$$\mathbf{m}_2 = (0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0),$$
$$\mathbf{m}_3 = (0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0),$$
$$\mathbf{m}_4 = (0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0),$$
$$\mathbf{m}_5 = (1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1),$$
$$\mathbf{m}_6 = (0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1),$$
$$\mathbf{m}_7 = (0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 1),$$

$$\mathbf{m}_8 = (0,0,0,1,0,0,0,1,0,1,0,1). \tag{2.7.12}$$

Here, the first four modes are the modes that produce E as a product, and the second four are the modes producing F. The modes taking up A are $\mathbf{m}_1$, $\mathbf{m}_2$, $\mathbf{m}_5$ and $\mathbf{m}_6$. The rest are the modes that take up B. Any flux distribution that satisfies the mass balance condition is a linear combination of these elementary modes, and we examine what effects our optimisation would have on these modes. For example, it is obvious that maximising the flux to E would require us to increase the activity through any of $\mathbf{m}_1$ to $\mathbf{m}_4$. Increasing any of these modes increases the flux towards E, and this is the same as simply increasing the amount of flux throughout the network as described earlier.

To redirect flux from F to E also requires lowering the activity of another pathway. Modes $\mathbf{m}_2$ and $\mathbf{m}_6$ direct flux from A towards D, and then from D to either product, E or F. One way to redirect flux from F to E would be to lower the activity of $\mathbf{m}_6$ and increase that of $\mathbf{m}_2$. Each mode, $\mathbf{m}_1$ to $\mathbf{m}_4$ has a corresponding mode in $\mathbf{m}_5$ to $\mathbf{m}_8$ with the same relationship, where increasing the activity in one whilst decreasing the activity in the other represents redirecting the flux from F to E

This network has eight pathways for twelve fluxes and six metabolites, so we still have less pathways than fluxes overall. If we were to add an additional pair of fluxes representing a reversible reaction between C and D we would have fourteen fluxes (assuming we model the reversible reaction as two separate irreversible fluxes) and six metabolites. Calculating the elementary modes for this new network, we gives us seventeen modes. Even in a network this size, it is entirely possible for the number of elementary modes to exceed the number of fluxes. In a genome scale model, the number of modes becomes so large that the complete set of elementary modes does not provide a useful basis for analysis, and instead we would restrict ourselves to examining a specific subset of modes, such as all the modes that produce a certain product.

2.8 More efficient routes

In the previous example there were multiple routes from an input to an output, but each was equally efficient i.e. each mode had the same level of output flux per unit input flux. In the network shown in Figure 2.7, with reactions given in Table 2.7 we have an example of a network with routes of different efficiencies. Specifically, the upper route produces two units output flux per unit input flux, whilst the lower route only produces one unit output flux per unit input flux. Here, A is changed into either two B or one C, then each B and C are changed into D. In this network it is obvious that the route through B is more efficient at producing D, but we should still check how FBA would behave with such a system. We form an FBA problem as usual, with the stoichiometric matrix

$$S = \begin{pmatrix} -1 & -1 & 0 & 0 & 1 & 0 \\ 2 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & -1 \end{pmatrix}, \tag{2.8.1}$$

for the flux vector $\mathbf{v} = (v_1, v_2, v_3, v_4, Q_A, Q_D)$. We choose an objective function $\mathbf{a}^T\mathbf{v} = v_3 + v_4$ to maximise the production of D. Initially, we constrain each flux

A	$\xrightarrow{v_1}$	2B
A	$\xrightarrow{v_2}$	C
B	$\xrightarrow{v_3}$	D
C	$\xrightarrow{v_4}$	D
External input	$\xrightarrow{Q_A}$	A
D	$\xrightarrow{Q_D}$	External output

Table 2.7: Set of reactions describing the example metabolic network seen in Figure 2.6. Here, the upper branch produces twice as much D as the lower branch for the same input A.

to be between zero and one. Then the FBA problem is

$$\begin{aligned} \max_{\mathbf{v}} v_3 + v_4 \quad \text{such that:} \quad -v_1 - v_2 + Q_A &= 0, \\ 2v_1 - v_3 &= 0, \\ v_2 - v_3 &= 0, \\ v_3 + v_4 - Q_D &= 0, \\ 0 \leq v_n, Q_n &\leq 1, \end{aligned} \tag{2.8.2}$$

which has a non unique solution, provided by linprog,

$$\mathbf{v} = (0.2833, 0.4334, 0.566, 0.4334, 0.7167, 1).$$

Of interest here is performing flux variability analysis: we find that if we try to minimise the input flux Q_A whilst maintaining a maximal steady-state, Q_A has a minimal value of 0.5. In this case all the flux is sent through B, and the flux distribution becomes $(0.5, 0, 1, 0, 0.5, 1)$. We also try maximising the input Q_A whilst maintaining the maximised steady-state. We find its maximal value is one, and the flux distribution becomes $(0, 1, 0, 1, 1, 1)$. We see that the pathway through B produces twice as much D as it takes up of A, and the pathway through

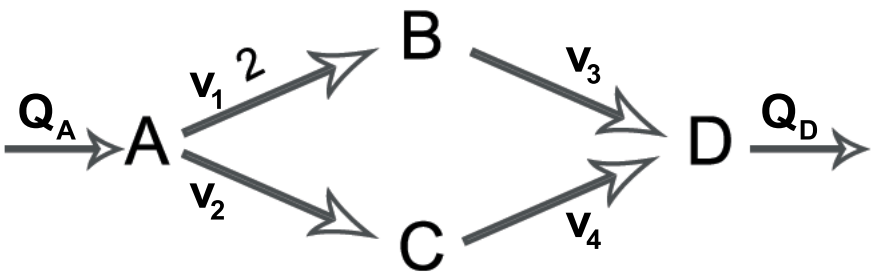

Figure 2.7: Network diagram for a network with two routes to a product. The top route is much more efficient than the bottom route.

C produces the same level of B as it takes up of A. In this case we would hope that flux balance analysis would divert as much flux towards B as possible before using the second pathway through C to provide maximal output.

We now remove entirely the upper bound on Q_D, to represent a much higher capacity to remove the product D. Performing FBA again without this upper bound we find our maximised solution becomes $(0.5, 0.5, 1, 0.5, 1, 1.5)$. In this case, the limiting fluxes are v_3 and Q_A, both having reached their maxima. We now keep the same maximum of Q_A, but increase the maximum value for v_3 to two, so that $0 \leq v_3 \leq 2$. This allows a greater flux to pass through the more efficient pathway ($A \to B \to D$), providing a higher output level. We find that under these conditions, the new flux distribution is $(1, 0, 2, 0, 1, 2)$. Here, all of the flux travels through the more efficient pathway, providing an output of $Q_D = 2$.

Let us examine now the FBA problem with the raised maximum for v_3, and no upper bound on Q_D. We consider also, a fixed value for $Q_A = z$, so we have control over the amount of A taken up by the network. This is described by the FBA problem

$$\begin{aligned} \max_{\mathbf{v}} v_3 + v_4 \quad \text{such that:} \quad & -v_1 - v_2 + Q_A = 0, \\ & 2v_1 - v_3 = 0, \\ & v_2 - v_3 = 0, \\ & v_3 + v_4 - Q_D = 0, \\ & 0 \leq v_1, v_2, v_4 \leq 1, \\ & 0 \leq Q_D, \\ & 0 \leq v_3 \leq 2, \\ & Q_A = z. \end{aligned} \tag{2.8.3}$$

We repeat the optimisation for multiple values of z to see how the optimal steady-state behaviour changes as the amount of A available increases.

Starting with $z = 0$, we find with no uptake, there is zero activity in the network. As we increase z up to one, all flux through the system is directed through the more optimal route, through v_1 and v_3. When z is equal to one, these two fluxes, v_1 and v_3 become equal to their maximum values. Further increasing z results in flux being directed through the less efficient route, through v_2 and v_4. When z reaches two, this less optimal route reaches its maximum capacity, with v_2 and v_4 at their maximum values. Increasing z any further past this results in the optimisation problem, (2.8.3) becoming infeasible, as the network cannot utilise the amount of A being taken up, and it is not possible to achieve steady-state. These results are summarised in Figure 2.8. We see that the network first uses the more efficient route, until v_1 or v_3 reach their maximum levels, before diverting any additional uptake down the less efficient route. In this way, we see FBA provides solutions that utilise more efficient routes first, before using less efficient ones, as we would expect.

Now that we have examined a number of simple networks using FBA, we begin examining a network for the metabolism of *Clostridium autoethanogenum*.

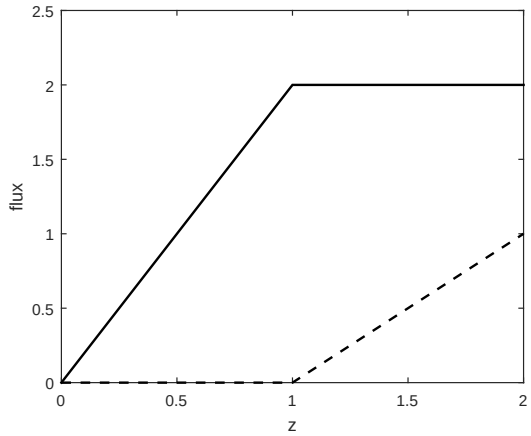

Figure 2.8: Plot of v_3 and v_4 in the FBA problem described in equation (2.8.3) for increasing values of z. The solid line shows the level of flux for v_3, representing the amount of D being produced by the more efficient pathway. The dashed line is the flux level for v_2, representing the amount of D produced in the less efficient pathway. We see v_3 increases to its maximum of two, before v_4 begins increasing to its maximum of one. If z was increased further, FBA would provide no solutions, as the optimisation problem becomes infeasible.

2.9 A metabolic network for *C. autoethanogenum*

Clostridium autoethanogenum (*C. autoethanogenum*) is a strictly anaerobic bacterium that is able to grow on carbon monoxide and carbon dioxide sources. The key products it forms are acetate, ethanol and butanediol. We also have production of lactate, valine, and a reaction leading to biomass production. A metabolic network for *C. autoethanogenum* is shown in Figure 2.9. For now, when forming our stoichiometric matrix, we only consider the key metabolites used in each reaction, and neglect additional metabolites such as ATP and water.

2.9.1 An ODE model

Let $\mathbf{c}$ be the vector of concentrations of metabolites within the network. Then the network is described by system of differential equations given in the generalised form of $\dot{\mathbf{c}} = \mathbf{f}(\mathbf{c})$ where $\mathbf{f}$ is some vector operator. We further specify the system by defining $\mathbf{f}(\mathbf{c}) = S\mathbf{v}(\mathbf{c})$ where S is the stoichiometric matrix for the network and $\mathbf{v}(\mathbf{c})$ is another vector operator. Using the law of mass action to define our $\mathbf{v}(\mathbf{c})$, we formulate a system of 21 ODEs to define this network, given as

$$\frac{dc_{CO}}{dt} = Q_{CO} - k_1 c_{CO}, \tag{2.9.1}$$

$$\frac{dc_{CO_2}}{dt} = Q_{CO_2} + k_1 c_{CO} - k_2 c_{CO_2} - k_8 c_{CO_2}, \tag{2.9.2}$$

$$\frac{dc_{Formate}}{dt} = k_2 c_{CO_2} - k_3 c_{Formate} - k_{15} c_{Formate} c_{Acetyl-CoA}, \tag{2.9.3}$$

$$\frac{dc_{Formyl-THF}}{dt} = k_3 c_{Formate} - k_4 c_{Formyl-THF}, \tag{2.9.4}$$

$$\frac{dc_{Methenyl-THF}}{dt} = k_4 c_{Formyl-THF} - k_5 c_{Methenyl-THF}, \tag{2.9.5}$$

$$\frac{dc_{Methelyne-THF}}{dt} = k_5 c_{Methenyl-THF} - k_6 c_{Methylene-THF}, \tag{2.9.6}$$

$$\frac{dc_{Methyl-THF}}{dt} = k_6 c_{Methylene-THF} - k_7 c_{Methyl-THF}, \tag{2.9.7}$$

$$\frac{dc_{Methyl-CoFeSp}}{dt} = k_7 c_{Methyl-THF} - k_9 c_{Methyl-CoFeSp} c_{CO_c}, \tag{2.9.8}$$

$$\frac{dc_{CO_c}}{dt} = Q_{CO_c} + k_8 c_{CO_2} - k_9 c_{Methyl-CoFeSp} c_{CO_c}, \tag{2.9.9}$$

$$\begin{aligned}\frac{dc_{Acetyl-CoA}}{dt} =& k_9 c_{Methyl-CoFeSp} c_{CO_c} - k_{10} c_{Acetyl-CoA} \\ & - k_{13} c_{Acetyl-CoA} - k_{15} c_{Formate} c_{Acetyl-CoA} \\ & - k_{16} c_{Acetyl-CoA} c_{Pyruvate},\end{aligned} \tag{2.9.10}$$

$$\frac{dc_{Acetyl-P}}{dt} = k_{10} c_{Acetyl-CoA} - k_{11} c_{Acetyl-P}, \tag{2.9.11}$$

$$\frac{dc_{Acetate}}{dt} = k_{11} c_{Acetyl-P} - k_{12} c_{Acetate} - out_{Acetate}, \tag{2.9.12}$$

$$\frac{dc_{Acetaldehyde}}{dt} = k_{12} c_{Acetate} + k_{13} c_{Acetyl-CoA} - k_{14} c_{Acetaldehyde}, \tag{2.9.13}$$

$$\frac{dc_{Ethanol}}{dt} = k_{14} c_{Acetaldehyde} - out_{Ethanol}, \tag{2.9.14}$$

$$\begin{aligned}\frac{dc_{Pyruvate}}{dt} =& k_{15} c_{Formate} c_{Acetyl-CoA} - k_{16} c_{Acetyl-CoA} c_{Pyruvate} \\ & - k_{17} c_{Pyruvate} - 2k_{18} c_{Pyruvate},\end{aligned} \tag{2.9.15}$$

$$\frac{dc_{Biomass}}{dt} = k_{16} c_{Acetyl-CoA} c_{Pyruvate} - out_{Biomass}, \tag{2.9.16}$$

$$\frac{dc_{Lactate}}{dt} = k_{17} c_{Pyruvate} - out_{Lactate}, \tag{2.9.17}$$

$$\frac{dc_{Acetolactate}}{dt} = k_{18} c_{Pyruvate} - k_{19} c_{Acetolactate} - k_{20} c_{Acetolactate}, \tag{2.9.18}$$

$$\frac{dc_{Valine}}{dt} = k_{19} c_{Acetolactate} - out_{Valine}, \tag{2.9.19}$$

$$\frac{dc_{Acetoin}}{dt} = k_{20} c_{Acetolactate} - k_{21} c_{Acetoin}, \tag{2.9.20}$$

$$\frac{dc_{2,3-Butanediol}}{dt} = k_{21} c_{Acetoin} - out_{2,3-Butanediol}, \tag{2.9.21}$$

where the k_n are rate coefficients for their associated reactions, and the c_n are variables for the concentration of each metabolite. It is important to note that CO and CO_c are treated as separate pools, so that reactions using CO do not access the pool of CO_c and vice versa. In this model we have 21 unknown rate coefficients, and nine other external fluxes representing the addition of the gases CO and CO_2 and the removal of products which would also require some consideration in how we choose to model them. Due to these unknowns, rather than trying to model this system dynamically using ODEs, we shall consider the system in a steady-state using the previously described FBA method.

2.9.2 A Simplified System

First we consider the generalised system $\dot{\mathbf{c}} = 0$. As before, we use $\dot{\mathbf{c}} = S\mathbf{v}(\mathbf{c})$, so that in a steady-state $S\mathbf{v}(\mathbf{c}) = 0$. We then consider $\mathbf{v}$ to simply be a vector of fluxes, neglecting any dependence on $\mathbf{c}$. Then we have the basis for our flux

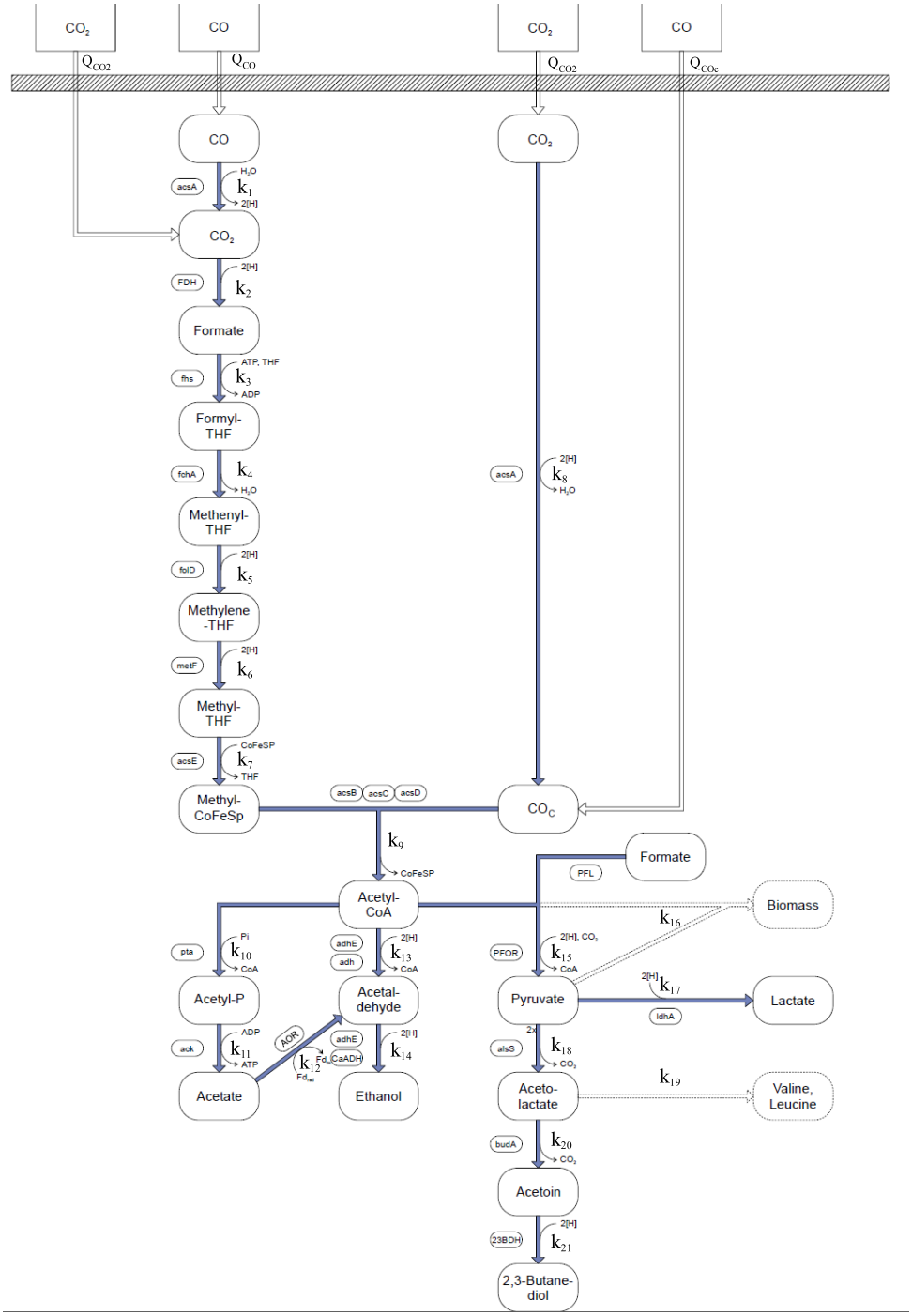

Figure 2.9: Network diagram for part of the metabolism of *C. autoethanogenum*. We have the uptake of both carbon dioxide and carbon monoxide (including a secondary carbon monoxide pool, labelled as CO_c. These two gases are then used to produce acetyl-CoA, which is used to produce on the left, acetate and ethanol, or on the right, biomass, lactate, valine, or butanediol.

balance analysis $S\mathbf{v} = 0$. In this case, we have 21 metabolites, and 30 fluxes, including nine external fluxes. This is much larger than any stoichiometric matrix we examined earlier in the chapter. To simplify the problem we consider a reduced network that keeps the same general structure as the larger one. To do this we model long chains of one-to-one reactions as single reactions. We keep only reactions that combine multiple metabolites or branch out into multiple metabolites. In this example we reduce the network seen in Figure 2.9 to the one shown in Figure 2.10, represented by the stoichiometric matrix

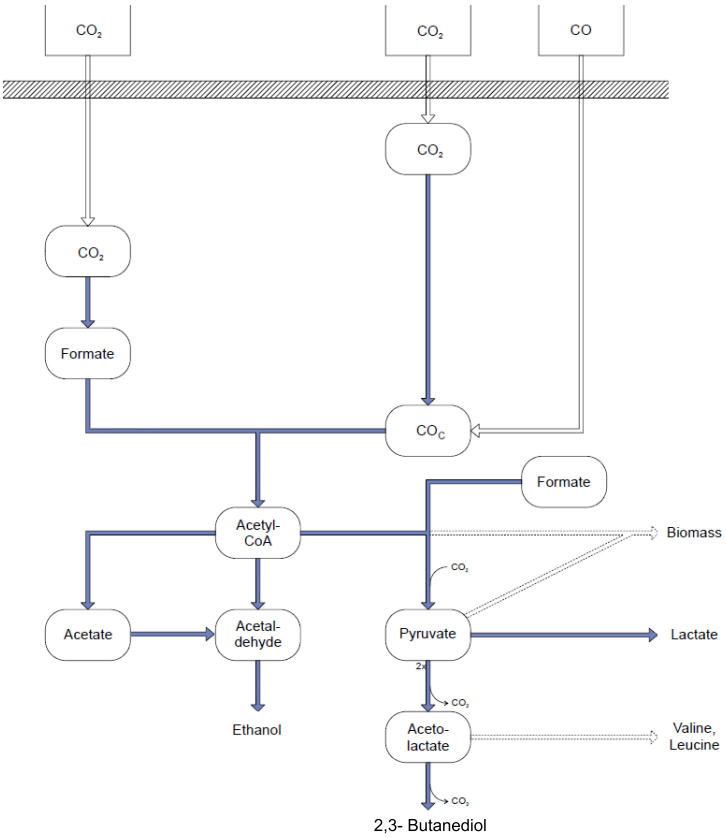

Figure 2.10: Simplified network diagram for *C. autoethanogenum*, in which long chains of reactions are simplified into single reactions.

$$
S = \begin{array}{c}
\\
CO_2 \\
CO_c \\
Formate \\
Acetyl-CoA \\
Pyruvate \\
Acetolactate \\
Acetaldehyde \\
Acetate
\end{array}
\begin{array}{c}
\begin{array}{cccccccccccccccc}
v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & Q_{CO_2} & Q_{CO_c} & out_{bio} & out_{lac} & out_{val} & out_{but} & out_{eth} & out_{ace}
\end{array} \\
\left(\begin{array}{cccccccccccccccc}
-1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & -1 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -2 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1
\end{array}\right)
\end{array}
. \tag{2.9.22}
$$

The reactions for this simplified network are given in Table 2.8. This has reduced our stoichiometric matrix from 21×30 to 8×16, which is simpler to work with.

For now, let us apply FBA in the usual manner, using maximising the produc-

CO_2	$\xrightarrow{v_1}$	Formate
CO_2	$\xrightarrow{v_2}$	CO_c
CO_c + Formate	$\xrightarrow{v_3}$	Acetyl-CoA
Formate + Acetyl-CoA	$\xrightarrow{v_4}$	Pyruvate
2 Pyruvate	$\xrightarrow{v_5}$	Acetolactate
Acetyl-CoA	$\xrightarrow{v_6}$	Acetaldehyde
Acetyl-CoA	$\xrightarrow{v_7}$	Acetate
Acetaldehyde	$\xrightarrow{v_8}$	Acetate
External input	$\xrightarrow{Q_{CO_2}}$	CO_2
External input	$\xrightarrow{Q_{CO_c}}$	CO_c
Acetyl-CoA + Pyruvate	$\xrightarrow{out_{bio}}$	Biomass output
Pyruvate	$\xrightarrow{out_{lac}}$	Lactate output
Acetolactate	$\xrightarrow{out_{val}}$	Valine output
Acetolactate	$\xrightarrow{out_{but}}$	Butanediol output
Acetaldehyde	$\xrightarrow{out_{eth}}$	Ethanol output
Acetate	$\xrightarrow{out_{ace}}$	Acetate output

Table 2.8: Set of reactions for the simplified *C. Autoethanogenum* network, seen in Figure 2.10, described by the stoichiometric matrix, (2.9.22).

tion of biomass as our objective function. In our reduced network, maximising the flux towards biomass is the same as maximising the external flux that combines acetyl-CoA and pyruvate. In this notation, this flux is referred to as out_{bio}. As in previous examples, we set the upper bound on all fluxes to be one, and the lower bound to be zero. Then our FBA problem is

$$\max_{\mathbf{v}} out_{bio} \quad \text{such that: } S\mathbf{v} = \mathbf{0}, \\ 0 \leq v_n, Q_n, out_n \leq 1. \tag{2.9.23}$$

This optimisation problem has a solution,

$$\mathbf{v} = (1, 0, 2/3, 1/3, 0, 0, 0, 0, 1, 2/3, 1/3, 0, 0, 0, 0, 0).$$

In this case, the biomass production we are optimising takes a value of 1/3. The inputs of carbon dioxide and carbon monoxide are 1 and 2/3 respectively. We see the optimal biomass production in this network uses 5/3 total input of carbon dioxide and carbon monoxide to produce 1/3 units of biomass. We also see that in order to maximise biomass production all other output fluxes have become zero. This causes any pathway not leading to biomass to be inhibited completely so that the internal fluxes v_4, v_5, v_6, v_7, and v_8 are also set to zero.

Raising the upper bounds on the input flux for CO_2, Q_{CO_2}, and performing flux variability analysis shows us that we could maintain this optimal biomass production with a sole input of CO_2, provided the uptake rate was high enough. On the other hand, this model says it would not be possible to grow solely on the pool of CO_c, as the system would be unable to produce any acetyl-CoA required for any of the products to form.

This particular steady-state predicted by using biomass as our objective function has led to all other products not being formed at all. Biologically we know this is not possible, as the system is producing other products. There are many methods by which our model could be refined to provide more accurate results.

First, we consider changing the objective function. Since we wish to move the system towards ethanol production, we could set that as our objective function. It is clear to see that merely setting the objective function to the production of ethanol shuts off biomass production entirely. We alleviate this by also raising the lower bound on biomass production to 0.15, which is a little less than half the optimal value. This provides us with a new solution to the optimisation problem

$$\mathbf{v} = (1, 0, 0.85, 0.15, 0, 0.2488, 0.3012, 0.3012, 1, 0.85, 0.15, 0, 0, 0.55, 0).$$

This tells us that whilst maintaining a biomass production of 0.15, it is possible to also make up to 0.55 units of ethanol. We also see that the carbon monoxide and carbon dioxide uptakes, Q_{CO_c} and Q_{CO_2}, have both increased. However, we still have no formation of other products, as there is not enough flux through the system as a whole to maximise production of either ethanol or biomass whilst still maintaining production of another product.

In both these examples, the limiting factor for production has been v_1, as it is the only internal flux which attained its maximum value. v_1 represents the production of formate. In this network formate is used to both produce acetyl-CoA, which in turn combines with another formate to make pyruvate. Then

pyruvate and acetyl-CoA combine to form biomass. This means three units of formate are required to make one unit of biomass. We note that in our optimal biomass producing state, we had a maximum input of CO_2 equal to one, allowing us to produce one unit of formate. This in turn limited our biomass production to be 1/3 unit. In the second example we also have a higher CO_c intake, approaching its limit. This is explained by the production of ethanol requiring less formate, so we are able to approach the full capacity of the uptakes into system as it shifts towards ethanol production rather than biomass.

2.9.3 Weighted objective functions

Rather than considering only objective functions for single products, we instead consider objective functions for multiple products. When doing this, we also weight the products in the objective function to make one more important than the other. We consider a weighted objective function of two products; biomass and ethanol. The flux towards biomass is out_{bio}, and the flux towards ethanol is out_{eth}. Then we consider objective functions of the form

$$f = \frac{d}{10} out_{bio} + \frac{10-d}{10} out_{eth}$$

for $d = 0$ to 10. We perform flux balance analysis for each value of d with the usual constraints of mass balance all fluxes being between zero and one. Then we have the flux balance analysis problem

$$\max_{\mathbf{v}} \ \frac{d}{10} out_{bio} + \frac{10-d}{10} out_{eth} \quad \text{such that: } S\mathbf{v} = \mathbf{0}, \\ 0 \leq v_n, Q_n, out_n \leq 1. \tag{2.9.24}$$

By solving this problem for values of d between zero and ten, we observe the point at which the system switches from ethanol production to biomass production. Table 2.9 shows these results. We note that there is an ethanol producing state and a biomass producing state. All other output fluxes were equal to zero across the entire range of d. This allows us to plot the feasible region for this linear programming problem in 2D, with one axis being biomass production, and the other being ethanol production. This feasible region is shown in Figure 2.11 We also see that the critical value for d at which the system chances from ethanol production to biomass production is somewhere between seven and eight. These correspond to objective functions of $0.7out_{bio} + 0.3out_{eth}$ and $0.8out_{bio} + 0.2out_{eth}$ respectively. We see that the system is able to maintain a production of one unit of ethanol whilst maximising its uptake rates. Whilst in the biomass producing state, the network showed a lower uptake flux for CO_c, lowering from one in the ethanol producing state, to 2/3 in the biomass producing state.

2.9.4 CO_2 production within the system

Up until now we have only been considering the main metabolites used in the metabolic reactions; however, we know there are additional chemicals produced and utilised within the system. In particular there is additional production and utilization of CO_2 within the internal fluxes, specifically v_4, v_5 and out_{but} shown

d	Q_{CO_2}	Q_{CO_c}	Biomass out	Ethanol out
0	1	1	0	1
1	1	1	0	1
2	1	1	0	1
3	1	1	0	1
4	1	1	0	1
5	1	1	0	1
6	1	1	0	1
7	1	1	0	1
8	1	0.6667	0.3333	0
9	1	0.6667	0.3333	0
10	1	0.6667	0.3333	0

Table 2.9: Solutions to the FBA problem seen in equation (2.9.24) for values of d between zero and ten. We see for $d \leq 7$ we have the same result in which ethanol is produced, whilst for $d \geq 8$ there is a second state in which biomass is produced.

in Figure 2.10. We now include these additional CO_2 terms in our network by replacing the first row of our stoichiometric matrix with

$$\begin{pmatrix} -1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}. \tag{2.9.25}$$

We now perform flux balance analysis with the same weighted objective functions as before using this new stoichiometric matrix, for d between zero and ten. The results are shown in Table 2.10. Interestingly we see that including these additional CO_2 requirements lowers the amount of biomass produced in the biomass producing state, however ethanol production remains at the same level. We also see at $d = 8$, we have an intermediate steady-state between the optimal biomass

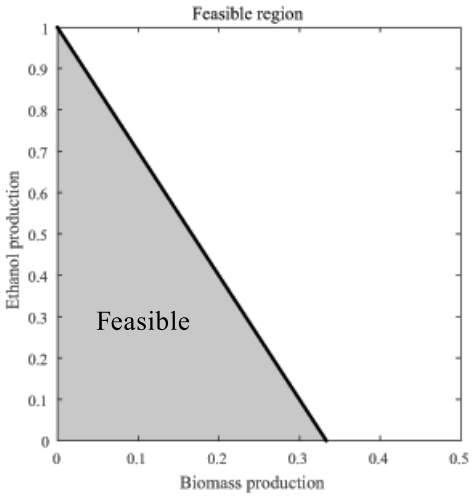

Figure 2.11: Feasible region in two dimensions for the linear programming problem, (2.9.24). We see the two vertices are the points at which ethanol production is one and biomass production is zero, and the point at which ethanol production is zero and biomass production is 1/3.

d	Q_{CO_2}	Q_{CO_c}	Biomass out	Ethanol out
$0-7$	1	1	0	1
8	1	0.6904	0.1548	0.3809
$9-10$	1	0.5	0.25	0

Table 2.10: Table showing the results of FBA problem (2.9.24), where the first row of S is replaced by (2.9.25) to include internal CO_2 production. For d between zero and seven, we have the ethanol producing state, whilst for d between nine and ten, we have the biomass producing state. For $d = 8$ there is production of both biomass and ethanol.

d	Q_{CO_2}	Q_{CO_c}	Biomass out	Ethanol out
7.999	1	1	0	1
8	1	0.6904	0.1548	0.3809
8.001	1	0.5	0.25	0

Table 2.11: As in Table 2.10, with d chosen close to the critical value of eight. We see the switch from producing ethanol to producing biomass.

production state and the optimal ethanol production state. We examine values of d close to eight, to see if this exists as a separate steady-state for certain values of d, or if it is coincidentally the exact critical value at which the states switch. The results, shown in Table 2.11, show that $d = 8$ is indeed the critical value. When the optimal state for the objective function produces biomass, it produces 0.25 units, whilst when it produces one unit of ethanol. With the objective function set as $0.8out_{bio} + 0.2out_{eth}$, we find that a steady-state in which biomass production is 0.25 and ethanol production is zero gives us the same value for the objective function as one in which ethanol production is one and biomass production is zero, showing that this is the critical value at which the optimisation problem switches from one solution type to another. By plotting the feasible region for this new FBA problem, with the production of ethanol and biomass, shown in Figure 2.12, we see that the optimal point for our objective function when $d < 8$ is the upper vertex, when ethanol production is one, and biomass production is zero. When $d > 8$, the optimal point is the lower vertex, when biomass production is 0.25, and ethanol production is zero. When $d = 8$, the optimal point becomes any point along the line between the two vertices, permitting infinitely many solutions.

2.9.5 Hydrogen Balance

Another key metabolite to consider in a refined model is hydrogen. Hydrogen is used in many of the reactions within the system as seen in Figure 2.9, and it is worthwhile to consider including it in our model. In order to include hydrogen reactions properly, we must move to a slightly more complex model seen in Figure 2.13. The stoichiometric matrix for this new model considers the initial CO pool as well as the CO_c pool, and includes a new hydrogen balance row. We also have two additional fluxes representing the uptake into the CO pool, and an uptake of

$CO + H_2O$	$\xrightarrow{v_1}$	$CO_2 + 2H$
$CO_2 + 2H$	$\xrightarrow{v_2}$	Formate
$CO_2 + 2H$	$\xrightarrow{v_3}$	CO_c
CO_c + Formate + 4H	$\xrightarrow{v_4}$	Acetyl-CoA
CO_2 + Formate + Acetyl-CoA + 2H	$\xrightarrow{v_5}$	Pyruvate
2 Pyruvate	$\xrightarrow{v_7}$	Acetolactate + CO_2
Acetyl-CoA + 2H	$\xrightarrow{v_8}$	Acetaldehyde
Acetyl-CoA	$\xrightarrow{v_9}$	Acetate
Acetate	$\xrightarrow{v_{10}}$	Acetaldehyde
External input	$\xrightarrow{Q_{CO}}$	CO
External input	$\xrightarrow{Q_{CO_2}}$	CO_2
External input	$\xrightarrow{Q_{CO_c}}$	CO_c
External input	$\xrightarrow{Q_H}$	H
Acetyl-CoA + Pyruvate	$\xrightarrow{out_{bio}}$	Biomass output
Pyruvate + 2H	$\xrightarrow{out_{lac}}$	Lactate output
Acetolactate	$\xrightarrow{out_{val}}$	Valine/Leucine output
Acetolactate + 2H	$\xrightarrow{out_{but}}$	Butanediol output + CO_2
Acetaldehyde + 2H	$\xrightarrow{out_{eth}}$	Ethanol output
Acetate	$\xrightarrow{out_{ace}}$	Acetate output

Table 2.12: The set of reactions for the network seen in Figure 2.13.

hydrogen. We also include the internal CO_2 production and usage added in by equation (2.9.25). This new stoichiometric matrix, T, can be split into two parts. The part containing the internal reactions is

$$T_{in} = \begin{array}{r} \\ CO \\ CO_2 \\ CO_c \\ Formate \\ Acetyl-CoA \\ Pyruvate \\ Acetolactate \\ Acetaldehyde \\ Acetate \\ Hydrogen \end{array} \begin{array}{c} \begin{array}{ccccccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_9 \end{array} \\ \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 2 & -2 & -2 & -4 & -2 & 0 & -2 & 0 & 0 \end{pmatrix} \end{array}, \tag{2.9.26}$$

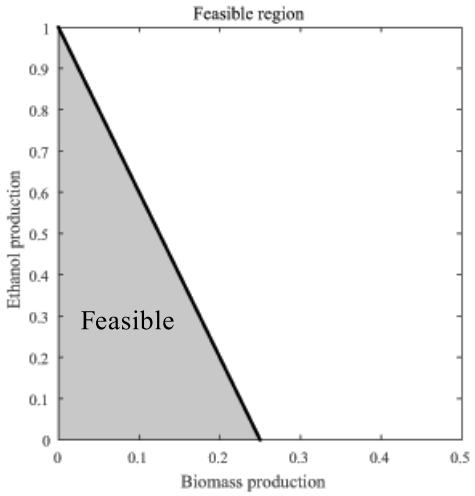

Figure 2.12: Feasible region in two dimensions for the linear programming problem, (2.9.24) with the first row of the stoichiometric matrix replaced with (2.9.25. We see the two vertices are the points at which ethanol production is one and biomass production is zero, and the point at which ethanol production is zero and biomass production is 0.25.

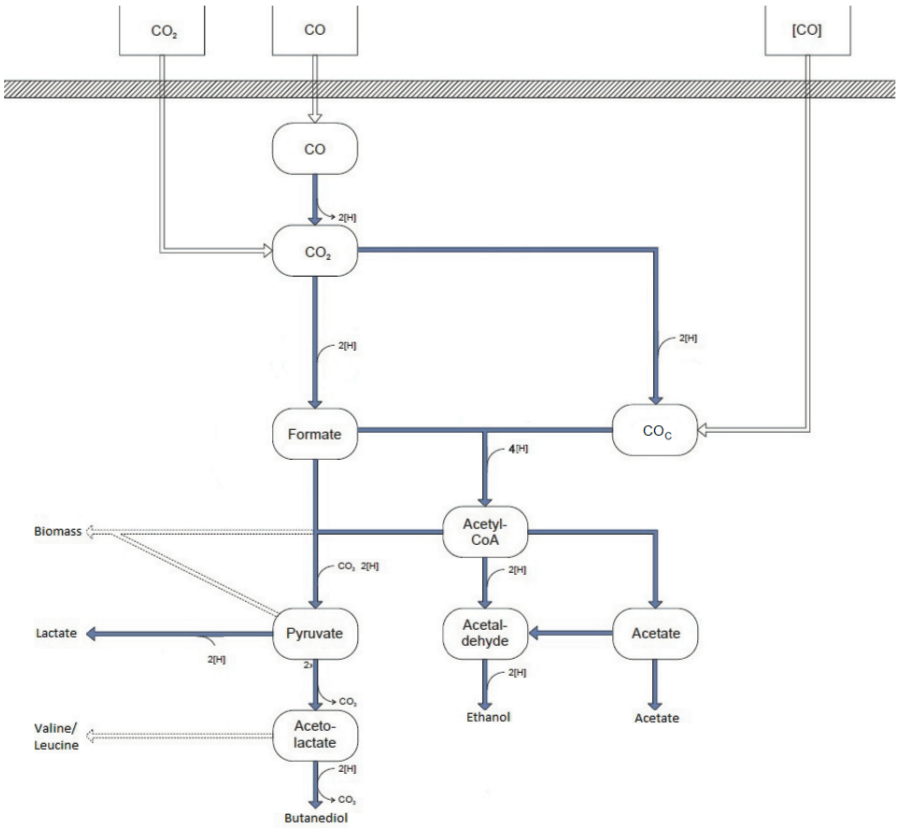

Figure 2.13: *C. autoethanogenum* network, as seen before in Figure 2.10, with the addition of hydrogen and a second separate uptake of carbon monoxide.

and the part containing the external reactions is

$$T_{ex} = \begin{array}{r} \\ CO \\ CO_2 \\ CO_c \\ Formate \\ Acetyl-CoA \\ Pyruvate \\ Acetolactate \\ Acetaldehyde \\ Acetate \\ Hydrogen \end{array} \begin{array}{c} \begin{array}{cccccccc} Q_{CO} & Q_H & out_{bio} & out_{lac} & out_{val} & out_{but} & out_{eth} & out_{ace} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & -2 & 0 & -2 & -2 & 0 \end{pmatrix} \end{array}. \tag{2.9.27}$$

So that

$$T = (T_{in}|T_{ex}), \tag{2.9.28}$$

where Q_H represents the hydrogen uptake, and the last row represents the hydrogen balance. There are two ways to approach the flux balance analysis here. Firstly we could treat the hydrogen uptake as we treat all the other fluxes in the system, however it should be clear that the hydrogen requirements for this system are higher than everything else, so limiting the uptake to be between zero and one would limit activity in the system quite heavily. Instead we consider Q_H to have no upper bound. In this way we observe what level of hydrogen the system requires to maintain its optimal

d	Q_{CO}	Q_{CO_2} in	Q_{CO_c} in	Q_H	Biomass out	Ethanol out
0	0.5901	0.7924	0.6174	8.7612	0	1
1	0.5891	0.7936	0.6174	8.7776	0	1
2	0.5881	0.7952	0.6167	8.8046	0	1
3	0.5877	0.7970	0.6153	8.8513	0	1
4	0.5877	0.7995	0.6128	8.9064	0	1
5	0.5884	0.8016	0.6100	8.9157	0	1
6	0.5881	0.8046	0.6072	8.9239	0	1
7	0.5872	0.8067	0.6060	8.9022	0	1
8	0.6700	0.8586	0.4714	4.3838	0.3333	0
9	0.6728	0.8608	0.4664	4.3882	0.3333	0
10	0.6742	0.8614	0.4644	4.3894	0.3333	0

Table 2.13: Results for the FBA problem described in equation (2.9.29) for integer values of d between zero and ten. We have an ethanol producing state for $d \leq 7$, and a biomass producing state for $d \geq 8$. Where previously all ethanol producing states were the same, we now have some variation in CO and CO_2 inputs.

steady-states. We use the same weighted objective functions as the previous section again. Then we set up the linear programming problem

$$\begin{aligned} \max_{\mathbf{v}} \ \frac{d}{10} out_{bio} + \frac{10-d}{10} out_{eth} \quad \text{such that: } T\mathbf{v} = \mathbf{0}, \\ 0 \leq v_n, out_n, Q_{CO}, Q_{CO_2}, Q_{CO_c} \leq 1, \\ -\infty \leq Q_H \leq \infty. \end{aligned} \tag{2.9.29}$$

We solve this optimisation problem for d in the range 0–10. The results are summarised in Table 2.13 Here we again have two distinct steady-states for the ethanol production and the biomass production, however unlike in our previous examples, we also see some minimal variation within those steady-states on the uptake rates. This shows there is some freedom of choice in which ratio of gases are utilized within the system whilst still maintaining an optimal steady-state. Another interesting result from this table is the hydrogen usage. The amount of hydrogen used in the system halves when switching from an ethanol producing state to a biomass producing state. This suggests that a way to stimulate the system to produce more ethanol and other products may be to increase the amount of hydrogen within the system. In biological terms, this might suggest that the bacteria growing in an environment with a lowered pH might produce more ethanol and other products than one in an environment with a higher pH, as the bacteria in the low pH environment is likely to uptake more hydrogen due to the higher concentration in the environment.

2.9.6 The ratio of hydrogen input to carbon source input

As we have seen that different products require different amounts of hydrogen to produce, we could further consider which ratios of hydrogen:carbon source inputs produce which products. We examine this using FBA. We do this by setting up an optimization problem in which all internal fluxes and output fluxes are only constrained to be above zero, and we fix our input fluxes at specific values. We then vary the hydrogen input

Q_{CO}	Q_H	Q_H/Q_{CO}	Biomass	Ethanol
6	8	1.3333	1	0
6	9	1.5	0.9375	0.1875
6	10	1.6667	0.875	0.375
6	11	1.8333	0.8125	0.5625
6	12	2	0.75	0.75
6	13	2.1667	0.6875	0.9375
6	14	2.3333	0.625	1.125
6	15	2.5	0.5625	1.3125
6	16	2.6667	0.5	1.5
6	17	2.8333	0.4375	1.6875
6	18	3	0.375	1.875
6	19	3.1667	0.3125	2.0625
6	20	3.3333	0.25	2.25
6	21	3.5	0.1875	2.4375
6	22	3.6667	0.125	2.625
6	23	3.8333	0.0625	2.8125
6	24	4	0	3

Table 2.14: Table of solutions to the FBA problem (2.9.30) with $a = 6$, and $b = c = 0$ for a range of values of d. With $a = 6$, we see the maximum biomass production is one. We have the input ratio, and the levels of product formation for biomass and ethanol. No other products are formed.

whilst keeping the carbon source input constant. For now, we keep biomass production as the objective function.

$$\begin{aligned} \max_{\mathbf{v}} \ out_{bio} \quad \text{such that: } & S\mathbf{v} = \mathbf{0}, \\ & v_n,\, out_n \geq 0, \\ & Q_{CO} = a, \\ & Q_{CO_2} = b, \\ & Q_{CO_c} = c, \\ & Q_H = d. \end{aligned} \tag{2.9.30}$$

We only examine the system with a single carbon source of either CO or CO_2. First we examine the system with only a CO carbon source. For this we set $a = 6$, $b = c = 0$ and then vary d. The results are given in Table 2.14.

The only two products that are formed in this are biomass and ethanol. All other possible products have zero output fluxes at all values of hydrogen input. The hydrogen input has been varied between eight and 24. If the hydrogen input is set to any value less than eight, or higher than 24, then the FBA problem cannot be solved. The system cannot reach a steady-state if there is a higher or lower hydrogen input whilst still maintaining the same CO input.

Peak biomass production is achieved when the hydrogen input flux is eight. At this value, the only product formed is Biomass. As we have no internal loops in our system, we are confident that there is no way to have a higher biomass production with this level of CO input. This is confirmed by setting up another FBA problem in which we maximise biomass production, whilst allowing the linear programming algorithm to

Q_{CO_2}	Q_H	Q_H/Q_{CO_2} in	Biomass	Ethanol
6	20	3.3333	1	0
6	21	3.5	0.9375	0.1875
6	22	3.6667	0.875	0.375
6	23	3.8333	0.8125	0.5625
6	24	4	0.75	0.75
6	25	4.1667	0.6875	0.9375
6	26	4.3333	0.625	1.125
6	27	4.5	0.5625	1.3125
6	28	4.6667	0.5	1.5
6	29	4.8333	0.4375	1.6875
6	30	5	0.375	1.875
6	31	5.1667	0.3125	2.0625
6	32	5.3333	0.25	2.25
6	33	5.5	0.1875	2.4375
6	34	5.6667	0.125	2.625
6	35	5.8333	0.0625	2.8125
6	36	6	0	3

Table 2.15: As in Table 2.14 except $a = c = 0$, and $b = 6$.

vary the hydrogen input. The problem

$$\begin{aligned} \max_{\mathbf{v}} \ out_{bio} \quad \text{such that: } & S\mathbf{v} = \mathbf{0}, \\ & v_n,\, out_n\, Q_H \geq 0, \\ & Q_{CO} = 6, \\ & Q_{CO_2} = 0, \\ & Q_{CO_c} = 0, \end{aligned} \tag{2.9.31}$$

gives us a maximised biomass production when $Q_H = 8$, that is, when the hydrogen input is eight. Above this peak value, we see that the change in product formation rates is linear. As the hydrogen input rises from eight to 24, the biomass production drops linearly from one to zero, and the ethanol production rises linearly from zero to three. We therefore split this into three sections dependant on the value of $k = \frac{Q_H}{Q_{CO}}$. For $k < 4/3$, the system cannot be at a steady-state. For $4/3 \leq k \leq 4$, the system produces some combination of biomass and ethanol, with $\Delta k \propto -\Delta out_{bio}$ and $\Delta k \propto \Delta out_{eth}$. Then for $k > 4$ the system again cannot be at a steady-state. At the critical points of $k = 4/3$ and $k = 4$, the steady-states are maximised towards producing a single product. for $k = 4/3$, this product is biomass whilst at $k = 4$ the sole product is ethanol.

We perform the same analysis for a system where CO_2 is the only carbon source. We use the FBA setup from equation (2.9.30) with $a = 0$, $b = 6$, $c = 0$ and varying d. The results are shown in Table 2.15 The results seen for a system with CO_2 as its only carbon source are very similar to those for the system with CO only. The patterns are exactly the same, only shifted to a higher hydrogen input value. If $k = \frac{Q_H}{Q_{CO}}$ and $l = \frac{Q_H}{Q_{CO_2}}$, then the system is split into the same three sections with $l = k + 2$. We see that the system growing only on CO_2 requires a higher hydrogen input to have steady-state solutions. If the system is grown only on CO, then the first reaction using the

CO transforms it into CO_2, which produces another two units of hydrogen. Without this reaction, the hydrogen levels required are higher overall.

Next we examine what happens when we allow the system to use both CO and CO_2, as long as the sum of those input fluxes remains fixed at six, and we allow hydrogen to vary as before. First we define the FBA problem

$$\begin{aligned} \max_{\mathbf{v}} \ out_{bio} \quad \text{such that: } S\mathbf{v} &= \mathbf{0}, \\ v_n,\, out_n &\geq 0, \\ Q_{CO} + Q_{CO_2} &= 6, \\ Q_{CO_c} &= 0, \\ Q_H &= d, \end{aligned} \tag{2.9.32}$$

where we again let d vary.

Table 2.16 shows the results. We see similar patterns to the previous two sets of simulations, where for the low input ratios we see production of biomass, and as the hydrogen input increases, the network switches from producing biomass to ethanol. We note that between $Q_H = 8$ and $Q_H = 20$ the network manages to maintain its maximised level of biomass production by shifting from using CO to using CO_2. This shows the network dealing with the higher hydrogen input by using more CO_2, so less hydrogen is produced within the network. We therefore split this network in to five sections. Let $m = \frac{Q_H}{Q_{CO}+Q_{CO_2}}$. Then for $m < 4/3$ the network cannot achieve a steady-state. For $4/3 \leq m \leq 10/3$, the network maintains maximised production of biomass, using a mixture of both the CO and CO_2 carbon sources. For $10/3 \leq m \leq 6$ the network produces a mixture of ethanol and biomass using only the CO_2 carbon source. Finally, for $m > 6$ the network again cannot achieve a steady-state. At the critical value of $m = 6$ the network only produces ethanol.

2.9.7 Elementary modes analysis for the simplified system

When examining the weighted objective functions in the previous section, we saw some variation of the carbon sources, CO and CO_2 whilst maintaining the same biomass or ethanol production levels. One way to examine this variation further is to explore whether or not there are specific ratios of input gases, or if the cell could grow and maintain a steady-state on either gas. We do this by returning to elementary modes analysis, and calculating the set of elementary modes for this system. These modes are shown in Table 2.17. Each mode only produces a single product, with a flux of one. The output flux shows which product is formed by that mode, where out_{bio} is biomass, out_{lac} is lactate, out_{val} is valine, out_{but} is butanediol, out_{eth} is ethanol and out_{ace} is acetate.

We split these elementary modes into four groups of seven, based on active uptake fluxes for carbon, Q_{CO}, Q_{CO_2}, and Q_{CO_c}. The hydrogen uptake flux, Q_H, is active in all four groups. Within each group, there is a single mode producing each biomass, lactate, valine, butanediol, and acetate, and two modes that produce ethanol, one utilizing acetyl-CoA to produce acetaldehyde through v_8, and one using acetate to produce acetaldehyde through v_{10}.

The first group of seven modes use only the standard carbon monoxide input, Q_{CO}. In this group, the modes use two units of CO and four of hydrogen to create one unit of acetyl-CoA. Any further uptake of carbon or hydrogen are required only for a specific product and unique to the mode. We see specifically that production of biomass, valine, and butanediol all require two units of acetyl-CoA to be produced. We see this in the

Q_{CO}	Q_{CO_2}	Q_H	$Q_H/(Q_{CO}+Q_{CO_2})$	Biomass	Ethanol
6	0	8	1.3333	1	0
5.5	0.5	9	1.5	1	0
5	1	10	1.6667	1	0
4.5	1.5	11	1.8333	1	0
4	2	12	2	1	0
3.5	2.5	13	2.1667	1	0
3	3	14	2.3333	1	0
2.5	3.5	15	2.5	1	0
2	4	16	2.6667	1	0
1.5	4.5	17	2.8333	1	0
1	5	18	3	1	0
0.5	5.5	19	3.1667	1	0
0	6	20	3.3333	1	0
0	6	21	3.5	0.9375	0.1875
0	6	22	3.6667	0.875	0.375
0	6	23	3.8333	0.8125	0.5625
0	6	24	4	0.75	0.75
0	6	25	4.1667	0.6875	0.9375
0	6	26	4.3333	0.625	1.125
0	6	27	4.5	0.5625	1.3125
0	6	28	4.6667	0.5	1.5
0	6	29	4.8333	0.4375	1.6875
0	6	30	5	0.375	1.875
0	6	31	5.1667	0.3125	2.0625
0	6	32	5.3333	0.25	2.25
0	6	33	5.5	0.1875	2.4375
0	6	34	5.6667	0.125	2.625
0	6	35	5.8333	0.0625	2.8125
0	6	36	6	0	3

Table 2.16: Table of solutions to the FBA problem (2.9.32) for a range of values of d. We have the input ratio, and the levels of product formation for biomass and ethanol. No other products are formed.

Mode	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	Q_{CO}	Q_{CO_2}	Q_{CO_c}	Q_H	Output
1	2	1	1	1	0	0	1	0	0	2	0	0	8	out_{eth}
2	2	1	1	1	0	0	0	1	1	2	0	0	6	out_{eth}
3	6	3	2	2	1	0	0	0	0	6	0	0	8	out_{bio}
4	4	2	1	1	1	0	0	0	0	4	0	0	6	out_{lac}
5	7	4	2	2	2	1	0	0	0	7	0	0	10	out_{val}
6	6	4	2	2	2	1	0	0	0	6	0	0	14	out_{but}
7	2	1	1	1	0	0	0	1	0	2	0	0	4	out_{ace}
8	1	1	0	1	0	0	1	0	0	1	0	1	8	out_{eth}
9	1	1	0	1	0	0	0	1	1	1	0	1	6	out_{eth}
10	4	3	0	2	1	0	0	0	0	4	0	2	8	out_{bio}
11	3	2	0	1	1	0	0	0	0	3	0	1	6	out_{lac}
12	5	4	0	2	2	1	0	0	0	5	0	2	10	out_{val}
13	4	4	0	2	2	1	0	0	0	4	0	2	14	out_{but}
14	1	1	0	1	0	0	0	1	0	1	0	1	4	out_{ace}
15	0	1	1	1	0	0	1	0	0	0	2	0	12	out_{eth}
16	0	1	1	1	0	0	0	1	1	0	2	0	10	out_{eth}
17	0	3	2	2	1	0	0	0	0	0	6	0	20	out_{bio}
18	0	2	1	1	1	0	0	0	0	0	4	0	14	out_{lac}
19	0	4	2	2	2	1	0	0	0	0	7	0	24	out_{val}
20	0	4	2	2	2	1	0	0	0	0	6	0	26	out_{but}
21	0	1	1	1	0	0	0	1	0	0	2	0	8	out_{ace}
22	0	1	0	1	0	0	1	0	0	0	1	1	10	out_{eth}
23	0	1	0	1	0	0	0	1	1	0	1	1	8	out_{eth}
24	0	3	0	2	1	0	0	0	0	0	4	2	16	out_{bio}
25	0	2	0	1	1	0	0	0	0	0	3	1	12	out_{lac}
26	0	4	0	2	2	1	0	0	0	0	5	2	20	out_{val}
27	0	4	0	2	2	1	0	0	0	0	4	2	22	out_{but}
28	0	1	0	1	0	0	0	1	0	0	1	1	6	out_{ace}

Table 2.17: Table of elementary modes for the *C. autoethanogenum* network described by the stoichiometric matrix, (2.9.28). Each mode includes a single output flux which is listed in the final column. Every mode is scaled so it produces exactly one unit of its associated product.

modes, as v_4 is equal to two for those modes and one for all other pathways in this group. Then we write the input requirements for a mode in this group as

$$v_4(2CO + 4[H]) + a_n CO + b_n[H],$$

where the a_n and b_n are found by subtracting the CO and hydrogen used in the creation of acetyl-CoA from the overall input requirements of the system. This gives us the CO and hydrogen requirements unique to each pathway.

The second group of modes are modes 8-14. These modes use two inputs, Q_{CO} and Q_{CO_c}, representing the CO and CO_c inputs, which are treated as separate in this model. The only change in these modes compared to those in the first group is that rather than taking two CO, and then changing one of them to CO_c via CO_2, they directly take up one unit of CO_c. We see that acetyl-CoA production requires one CO, one CO_c and four hydrogen. As before, the modes producing either biomass, valine, or butanediol all require two units of acetyl-CoA. We split the input requirements as before into

$$v_4(CO + CO_c + 4[H]) + a_n CO + b_n[H].$$

Here the coefficients a_n and b_n are the same as those for the first group.

The third group of modes are pathways 15-21. These modes use only Q_{CO_2}, which is the CO_2 input. This group has higher hydrogen requirements. In the previous groups, CO was used to create CO_2, producing some of the hydrogen required by the system. With the system using CO_2 directly from the input, this hydrogen producing reaction no longer occurs, causing the hydrogen input requirements to increase. This group uses two units of CO_2 and eight of hydrogen to produced a unit of acetyl-CoA. Biomass, valine and butanediol production all still require two units of acetyl-CoA. The input requirements are split as

$$v_4(2CO_2 + 8[H]) + c_n CO_2 + d_n[H].$$

We find that the c_ns, representing the additional CO_2 requirements for each pathway are the same as the a_ns used in the previous two groups. However the additional hydrogen requirements, d_n are distinct from the hydrogen requirements b_n from the previous two groups.

The fourth and final group of modes are modes 22-28. These modes use both Q_{CO_2} and Q_{CO_c}, which are the CO_2 and CO_c inputs. Similarly to the difference between groups one and two, the difference between the third and fourth groups are the usage of CO_c rather than a second unit of CO_2 in the production of acetyl-CoA. In this case, this change also lowers the hydrogen requirement of the system. The group then uses one unit of CO_2, one unit of CO_c and six units of hydrogen (rather than eight) to produce a unit of acetyl-CoA. We split the input requirements as before to

$$v_4(CO_2 + CO_c + 6[H]) + c_n CO_2 + d_n[H],$$

where the c_n and d_n are the same as in group three.

These modes also provide insight into the results of our analysis of input ratios in Section 2.9.6. Table 2.18 shows us the hydrogen/carbon input ratio for each mode in the system. We see that the modes with the lowest ratio are $\mathbf{m}_3$ and $\mathbf{m}_{10}$. These modes are in groups one and two respectively, meaning $\mathbf{m}_3$ has only Q_{CO} active, whilst $\mathbf{m}_{10}$ also has Q_{CO_c} active. Neither have Q_{CO_2} active. This shows that the network has steady-states for lower input ratios if the sole carbon source is carbon monoxide, and not carbon dioxide. Both of these modes produce biomass, as shown in Table 2.17. Similarly, we see that the mode that has the highest input ratio is mode $\mathbf{m}_{15}$, which produces ethanol. This mode is in group three, meaning the only active carbon uptake

flux is Q_{CO_2}. This shows that the network maintains steady-state for higher input ratios, if the sole carbon source is carbon dioxide. In particular, we note that for group one, where the sole carbon uptake flux is Q_{CO} the lowest input ratio is 4/3, whilst the highest is four, which agrees with the range shown by the results in Table 2.14. Similarly in group 3, the range is between 10/3 and six, which agrees with the range shown by the results in Table 2.15 These results are exactly in line with those seen in Section 2.9.6. Examining our elementary modes has allowed us to draw additional insight into the range of input ratios our *C. autoethanogenum* network is able to attain a steady-state in.

Group 1							
Mode	1	2	3	4	5	6	7
Hydrogen uptake	8	6	8	6	10	14	4
Total carbon uptake	2	2	6	4	7	6	2
Ratio	4	3	4/3	3/2	10/7	7/3	2
Group 2							
Mode	8	9	10	11	12	13	14
Hydrogen uptake	8	6	8	6	10	14	4
Total carbon uptake	2	2	6	4	7	6	2
Ratio	4	3	4/3	3/2	10/7	7/3	2
Group 3							
Mode	15	16	17	18	19	20	21
Hydrogen uptake	12	10	20	14	24	26	8
Total carbon uptake	2	2	6	4	7	6	2
Ratio	6	5	10/3	7/2	24/7	13/3	4
Group 4							
Mode	22	23	24	25	26	27	28
Hydrogen uptake	10	8	16	12	20	22	6
Total carbon uptake	2	2	6	4	7	6	2
Ratio	5	4	8/3	3	20/7	11/3	3

Table 2.18: Table of input ratios for each elementary mode of the *C. autoethanogenum* network described by the stoichiometric matrix (2.9.28).

2.10 Conclusions

We have applied FBA to a wide variety of networks in order to find steady-state flux vectors for each of them, according to specific objective functions. These simple examples illustrated several key points of FBA, namely the dependence of solutions on the choice of objective function, the ability of FBA to predict biologically infeasible solutions, the possibility of multiple optimal solutions for an objective function, and in particular, the ability to write FBA solutions as a sum of elementary modes in order to provide a physical representation of solutions. We then formed a simplified metabolic network for *C. autoethanogenum*, and applied FBA to this. By using a weighted objective function, and altering the weightings, we observed the solution switching from producing one product to another, with no intermediate period of producing both, except for in the case of the critical value. Then by adding more features back in

to this network, specifically balancing hydrogen, and including internal reactions that produced CO_2, we observed how these changes to the network modified the solutions provided by FBA. We also employed FVA and elementary modes analysis to draw further insight into the behaviour of these steady-states. Elementary modes analysis allowed us to find a biologically relevant set of flux modes that form a spanning set for solutions to the set of equations $S\mathbf{v} = \mathbf{0}$. Whilst for genome-scale metabolic networks, the size of this set becomes unwieldy, by focusing on specific subsets of modes that share common properties, we are still able to identify key properties of the network that are required for the production of certain products. In the *C. autoethanogenum* case, we explored the range of hydrogen/carbon input ratios that permitted steady-state solutions to exist. By examining the elementary modes, we found that this range was between 4/3 and six, as the lowest input ratio on an elementary mode was 4/3, and the highest was six. There cannot exist solutions that utilize a higher or lower input ratio, as every steady-state solution for this network must be some positive sum of these modes. In particular for the group of modes which used carbon monoxide as their sole input, the maximum input ratio was four, as modes in this group had lower input ratios than those which used carbon dioxide.

Now that we have used FBA to examine this *C. autoethanogenum* network in steady-state, it would be worth looking at an ODE model for the same network to see if we similar results are obtained. We will form a model for the network seen in Figure 2.13, as the bulk of our FBA was performed on this network. In the next chapter this ODE model will be explored and we will analyse the dynamic behaviour. In particular, we consider what happens when we have a hydrogen/carbon input ratio outside of those where we are able to find FBA solutions. We expect to find steady-state solutions within the bounds predicted by FBA, but outside we expect to see non-steady long timescale behaviours.

Chapter 3

An ODE Model, and asymptotic analysis, for *C. autoethanogenum*

3.1 Introduction

In Chapter 2, we used FBA methods to examine our *C. autoethanogenum* network. We found that there was a specific ratio for the carbon monoxide and hydrogen inputs that was required for steady-states to exist (namely $4/3 \leq Q_H/Q_{CO} \leq 4$). Outside of this ratio we knew that steady-states could not exist, but FBA is not able to provide any insight into the behaviour of the model in these regions. By forming a full ODE model for this network, we seek to examine the long term behaviour seen when the system is unable to maintain a steady-state. Additionally, an ODE model is also able to model the short term dynamic behaviour that FBA is also unable to predict.

ODE models, sometimes called kinetic models with reference to metabolism, have been formulated for numerous different organisms, including *E. coli*, as studied by Chassagnole et. al [12], *Penicillium chrysogenum*, in particular the pathway the produces penicillin, by de Noronha Pissara et. al [17], and the pentose phosphate pathway of *Saccharomyces cerevisiae* by Vaseghi et. al [75].

In this chapter we form an ODE model for *C. autoethanogenum*. This model is simplistic, using the law of mass action to model reaction rates, allowing us to keep the number of parameters to a minimum. Rather than attempting to estimate these parameters, we instead use asymptotic analysis to quantify the possible long timescale behaviours of this system. We already found from FBA that there is a range of input ratios that permit the existence of steady-states, but outside this range we find non-steady long term behaviours. By using asymptotic analysis to reduce the ODE models in each region, we seek parameter values that are required for a long term behaviour to exist. In particular, we examine the ratio of inputs, Q_H/Q_{CO} found to be important by FBA, and a second ratio of parameters identified to have a large effect on which long term behaviour the model displays. By performing this analysis, we are able to provide insight into the behaviour of the model, without requiring estimates of parameters.

3.2 Forming an ODE model

First, let us formulate an ODE model for the network seen in Figure 3.1 of the form

$$\dot{\mathbf{c}} = S\mathbf{v}(\mathbf{c}), \tag{3.2.1}$$

where $\mathbf{c}$ is the vector of concentrations and S is the 10×19 stoichiometric matrix described in Chapter 2, seen in equation (2.9.28). For simplicity we use the law of mass

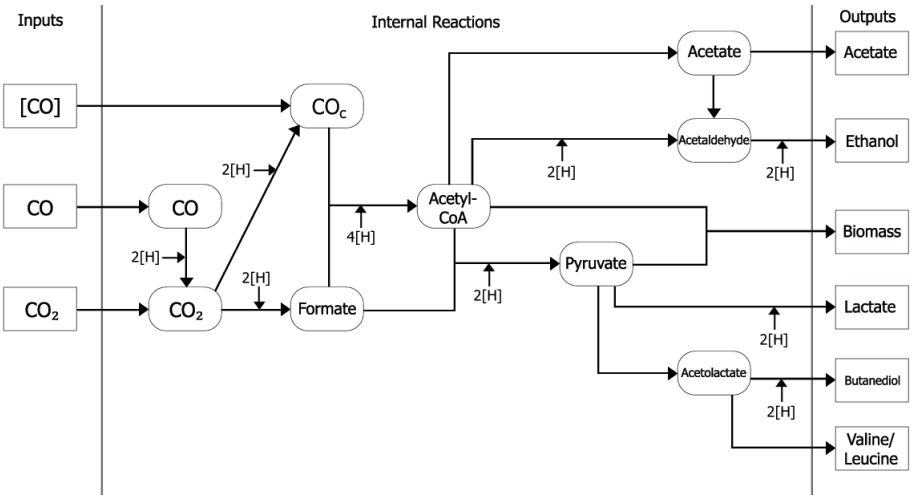

Figure 3.1: Network diagram for the metabolism of clostridium autoethanogenum

action to model reaction rates, leading to the system of ten ODEs

$$\frac{dc_{CO}}{dt} = k_{10}({c_{CO}}^* - c_{CO}) - k_1 c_{CO}, \tag{3.2.2}$$

$$\begin{aligned}\frac{dc_{CO_2}}{dt} &= k_{11}({c_{CO_2}}^* - c_{CO_2}) + k_1 c_{CO} - k_2 c_{CO_2} {c_H}^2 - k_3 c_{CO_2} {c_H}^2 \\ &\quad - k_5 c_{CO_2} c_F c_{ACA} {c_H}^2 + k_6 {c_{Py}}^2 + k_{17} c_{Al} {c_H}^2,\end{aligned} \tag{3.2.3}$$

$$\frac{dc_{CO_c}}{dt} = k_{12}({c_{CO_c}}^* - c_{CO_c}) + k_3 c_{CO_2} {c_H}^2 - k_4 c_{CO_c} c_F {c_H}^4, \tag{3.2.4}$$

$$\frac{dc_F}{dt} = k_2 c_{CO_2} {c_H}^2 - k_4 c_{CO_c} c_F {c_H}^4 - k_5 c_{CO_2} c_F c_{ACA} {c_H}^2, \tag{3.2.5}$$

$$\begin{aligned}\frac{dc_{ACA}}{dt} &= k_4 c_{CO_c} c_F {c_H}^4 - k_5 c_{CO_2} c_F c_{ACA} {c_H}^2 - k_7 c_{ACA} {c_H}^2 \\ &\quad - k_8 c_{ACA} - k_{14} c_{ACA} c_{Py},\end{aligned} \tag{3.2.6}$$

$$\frac{dc_{Py}}{dt} = k_5 c_{CO_2} c_F c_{ACA} {c_H}^2 - 2k_6 {c_{Py}}^2 - k_{14} c_{ACA} c_{Py} - k_{15} c_{Py} {c_H}^2, \tag{3.2.7}$$

$$\frac{dc_{Al}}{dt} = k_6 {c_{Py}}^2 - k_{16} c_{Al} - k_{17} c_{Al} {c_H}^2, \tag{3.2.8}$$

$$\frac{dc_{Ad}}{dt} = k_7 c_{ACA} {c_H}^2 + k_9 c_A - k_{18} c_{Ad} {c_H}^2, \tag{3.2.9}$$

$$\frac{dc_A}{dt} = k_8 c_{ACA} - k_9 c_A - k_{19} c_A, \tag{3.2.10}$$

$$\begin{aligned}\frac{dc_H}{dt} &= k_{13}({c_H}^* - c_H) + 2k_1 c_{CO} - 2k_2 c_{CO_2} {c_H}^2 - 2k_3 c_{CO_2} {c_H}^2 \\ &\quad - 4k_4 c_{CO_c} c_F {c_H}^4 - 2k_5 c_{CO_2} c_F c_{ACA} {c_H}^2 - 2k_7 c_{ACA} {c_H}^2 \\ &\quad - 2k_{15} c_{Py} {c_H}^2 - 2k_{17} c_{Al} {c_H}^2 - 2k_{18} c_{Ad} {c_H}^2,\end{aligned} \tag{3.2.11}$$

with the initial conditions of all concentrations being equal to zero,

$$\mathbf{c}(0) = \mathbf{0}. \tag{3.2.12}$$

Whilst all internal reactions and output reactions are modelled using the law of mass action, uptake rates are instead modelled as proportional to the difference between the concentration within the cell and the concentration outside the cell. Table 3.1 explains

which variables represent which concentrations, and Table 3.2 explains the parameters within the system. Whilst these parameters are not dimensionless, due to the lack of an example set of parameters, the choice of units is arbitrary.

Variable	Metabolite	Chemical Composition
c_{CO}	Carbon monoxide	CO
c_{CO_2}	Carbon dioxide	CO_2
c_{CO_c}	Carbon monoxide (Carbonyl branch)	CO
c_F	Formate	HCO_2^-
c_{ACA}	Acetyl-CoA	$C_{23}H_{38}N_7O_{17}P_3S$
c_{Py}	Pyruvate	$C_3H_4O_3$
c_{Al}	Acetolactate	$C_5H_8O_4$
c_{Ad}	Acetaldehyde	C_2H_4O
c_A	Acetate	$C_2H_3O_2^-$
c_H	Hydrogen	H

Table 3.1: Table of variables within the system, the key metabolite they represent the concentration of, and the chemical composition of the metabolite.

Parameter	Description	Dimensions
k_1	Rate coefficient for CO being transformed into CO_2	Time^{-1}
k_2	Rate coefficient for CO_2 being transformed into formate	$\text{Concentration}^{-2}\text{Time}^{-1}$
k_3	Rate coefficient for CO_2 being transformed into CO_c	$\text{Concentration}^{-2}\text{Time}^{-1}$
k_4	Rate coefficient for formate combining with CO_c to make acetyl-CoA	$\text{Concentration}^{-5}\text{Time}^{-1}$
k_5	Rate coefficient for acetyl-CoA combining with formate to make pyruvate	$\text{Concentration}^{-4}\text{Time}^{-1}$
k_6	Rate coefficient for two pyruvates combining to make acetolactate	$\text{Concentration}^{-1}\text{Time}^{-1}$
k_7	Rate coefficient for acetyl-CoA being transformed into acetaldehyde	$\text{Concentration}^{-2}\text{Time}^{-1}$
k_8	Rate coefficient for acetyl-CoA being transformed into acetate	$\text{Concentration}^{-1}\text{Time}^{-1}$
k_9	Rate coefficient for acetate being transformed into acetaldehyde	$\text{Concentration}^{-1}\text{Time}^{-1}$
k_{10}	Rate coefficient for the uptake of CO	Time^{-1}
k_{11}	Rate coefficient for the uptake of CO_2	Time^{-1}
k_{12}	Rate coefficient for the uptake of CO_c	Time^{-1}
k_{13}	Rate coefficient for the uptake of hydrogen	Time^{-1}
k_{14}	Rate coefficient for the combining of acetyl-CoA and pyruvate to output biomass	$\text{Concentration}^{-1}\text{Time}^{-1}$
k_{15}	Rate coefficient for the creation and output of lactate from pyruvate	$\text{Concentration}^{-2}\text{Time}^{-1}$
k_{16}	Rate coefficient for the creation and output of valine from acetolactate	$\text{Concentration}^{-1}\text{Time}^{-1}$
k_{17}	Rate coefficient for the creation and output of butanediol from acetolactate	$\text{Concentration}^{-2}\text{Time}^{-1}$
k_{18}	Rate coefficient for the creation and output of ethanol from acetaldehyde	$\text{Concentration}^{-2}\text{Time}^{-1}$
k_{19}	Rate coefficient for the output of acetate	Time^{-1}
${c_{CO}}^*$	Concentration of CO outside the cell	Concentration
${c_{CO_2}}^*$	Concentration of CO_2 outside the cell	Concentration
${c_{CO_c}}^*$	Concentration of CO_c outside the cell	Concentration
${c_H}^*$	Concentration of hydrogen outside the cell	Concentration

Table 3.2: Table of parameters in the system of (3.2.2)–(3.2.11).

Having formed an ODE model for our *C. autoethanogenum* network, we begin simulating it with various parameter schemes to observe the long term behaviour. Since we lack an example set of parameters to use, and there are no relevant parameters available in the literature, we start by choosing all parameters setting all parameters to equal one. Whilst this allows us to examine the possible behaviours of the model, it is important to note that different parameter choices may lead to different results, so whilst the different behaviours we find may be more generally applicable, the exact steady-state values will not be. Figure 3.2 shows the log time plots of concentrations with the choice

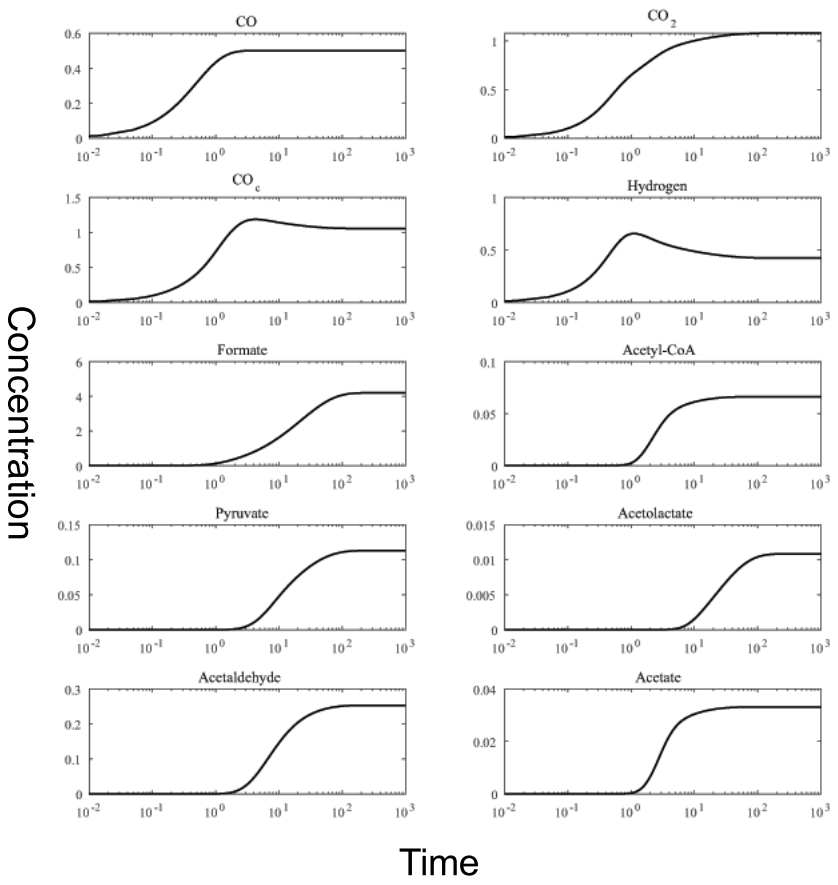

Figure 3.2: Plots of concentrations of metabolites in the ODE model in equations (3.2.2)–(3.2.11) with all parameters set to one, and initial condition $\mathbf{c} = 0$.

of all parameters equal to one. We see all metabolites increasing from their initial condition of zero to their final steady-state over a short time period. Most concentrations are monotonically increasing until this point, except for those of CO_c and hydrogen, which increase past their final steady-state values before decaying slightly again. These steady-state metabolite concentrations are summarised in Table 3.3. We also calculate the values for the fluxes in each steady-state from these concentration. Table 3.4 shows these steady-state flux values for each flux.

In particular, Table 3.4 shows us the output fluxes for the system, representing product formation rates. We see that the biomass production rate, given by $out_{bio} = k_14c_{ACA}c_{Py}$, is 0.0076, from a net carbon input of 0.3752. This means that the system is producing 0.02 units of biomass per unit of carbon. Our FBA models predicted that this network could produce up to $1/6 = 0.1667$ units of biomass per unit of carbon taken up, suggesting that, with a different choice of parameters, the system could produce a much larger amount of biomass. Similarly, it would be possible to choose parameters to maximise the formation of other products.

Also of note are the values of Q_{CO_2} and Q_{CO_c}, the uptakes of carbon dioxide and the secondary carbon monoxide source respectively. These have both taken negative values, indicating that the system is actually removing CO_2 and CO_c in these reactions rather than taking them up. Since the internal concentrations of these metabolites are

Metabolite	Concentration
CO	0.5000
CO_2	1.0712
CO_c	1.0536
Formate	4.0521
Acetyl-CoA	0.0678
Pyruvate	0.1127
Acetolactate	0.0107
Acetaldehyde	0.2542
Acetate	0.0339
Hydrogen	0.4266

Table 3.3: Steady-state concentrations of the metabolites seen in the simulation in Figure 3.2

Flux	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	
Value	0.5000	0.1950	0.1950	0.1414	0.0536	0.0127	0.0123	0.0678	0.0339	
Flux	Q_{CO}	Q_{CO_2}	Q_{CO_c}	Q_H	out_{bio}	out_{lac}	out_{val}	out_{but}	out_{eth}	out_{ace}
Value	0.5000	-0.0712	-0.0536	0.5734	0.0076	0.0205	0.0107	0.0020	0.0463	0.0339

Table 3.4: Values of fluxes for each reaction in the network shown in Figure 3.1 in the steady-state from the simulation shown in Figure 3.2.

greater than the external concentrations, the net flux through these reactions is negative, representing an output flux, rather than an uptake flux. In FBA, we constrained all our fluxes to be positive, so this result could not occur in our FBA simulations. Since these results could not occur directly in our FBA model, we should change our ODE model to correspond more closely to the FBA model, to find results that are directly comparable.

3.3 Constant Inputs

When examining the *C. autoethanogenum* network using FBA, we had fixed input fluxes of CO and hydrogen (Q_{CO} and Q_H respectively). We also neglected the CO_2 and CO_c input reactions entirely. By using only two input reactions we were able to examine how extreme the ratio of these two input fluxes could be before the system was unable to maintain steady-state. If there was no solution to the FBA problem for a given set of inputs, we deduced that the network would be unable to tend to a steady-state. In Section 2.9.6, it was found that a steady-state is only reached if the hydrogen input is greater than $\frac{4}{3}$ of the CO input and less than four times the CO input, that is if $\frac{4}{3}Q_{CO} \leq Q_H \leq 4Q_{CO}$. Equivalently, we say $4/3 \leq Q_H/Q_{CO} \leq 4$. We now modify the ODE model to see if these FBA results hold true for our ODE model. First, we completely inhibit the uptake rates for CO_2 and CO_c, by setting k_{11} and k_{12} to equal zero. Secondly, we replace the uptake fluxes for CO and hydrogen with the constant values from FBA, so that the ODEs (3.2.2) and (3.2.11) become

$$\frac{dc_{CO}}{dt} = Q_{CO} - k_1 c_{CO}, \tag{3.3.1}$$

$$\begin{aligned}\frac{dc_H}{dt} &= Q_H + 2k_1c_{CO} - 2k_2c_{CO_2}{c_H}^2 - 2k_3c_{CO_2}{c_H}^2 - 4k_4c_{CO_c}c_F{c_H}^4 \\ &\quad - 2k_5c_{CO_2}c_Fc_{ACA}{c_H}^2 - 2k_7c_{ACA}{c_H}^2 - 2k_{15}c_{Py}{c_H}^2 \\ &\quad - 2k_{17}c_{Al}{c_H}^2 - 2k_{18}c_{Ad}{c_H}^2. \end{aligned} \tag{3.3.2}$$

Then the ODE model, with k_{11} and k_{12} equal to zero, becomes

$$\frac{dc_{CO}}{dt} = Q_{CO} - k_1c_{CO}, \tag{3.3.3}$$

$$\begin{aligned}\frac{dc_{CO_2}}{dt} &= k_1c_{CO} - k_2c_{CO_2}{c_H}^2 - k_3c_{CO_2}{c_H}^2 - k_5c_{CO_2}c_Fc_{ACA}{c_H}^2 \\ &\quad + k_6{c_{Py}}^2 + k_{17}c_{Al}{c_H}^2, \end{aligned} \tag{3.3.4}$$

$$\frac{dc_{CO_c}}{dt} = k_3c_{CO_2}{c_H}^2 - k_4c_{CO_c}c_F{c_H}^4, \tag{3.3.5}$$

$$\frac{dc_F}{dt} = k_2c_{CO_2}{c_H}^2 - k_4c_{CO_c}c_F{c_H}^4 - k_5c_{CO_2}c_Fc_{ACA}{c_H}^2, \tag{3.3.6}$$

$$\begin{aligned}\frac{dc_{ACA}}{dt} &= k_4c_{CO_c}c_F{c_H}^4 - k_5c_{CO_2}c_Fc_{ACA}{c_H}^2 - k_7c_{ACA}{c_H}^2 - k_8c_{ACA} \\ &\quad - k_{14}c_{ACA}c_{Py}, \end{aligned} \tag{3.3.7}$$

$$\frac{dc_{Py}}{dt} = k_5c_{CO_2}c_Fc_{ACA}{c_H}^2 - 2k_6{c_{Py}}^2 - k_{14}c_{ACA}c_{Py} - k_{15}c_{Py}{c_H}^2, \tag{3.3.8}$$

$$\frac{dc_{Al}}{dt} = k_6{c_{Py}}^2 - k_{16}c_{Al} - k_{17}c_{Al}{c_H}^2, \tag{3.3.9}$$

$$\frac{dc_{Ad}}{dt} = k_7c_{ACA}{c_H}^2 + k_9c_A - k_{18}c_{Ad}{c_H}^2, \tag{3.3.10}$$

$$\frac{dc_A}{dt} = k_8c_{ACA} - k_9c_A - k_{19}c_A, \tag{3.3.11}$$

$$\begin{aligned}\frac{dc_H}{dt} &= Q_H + 2k_1c_{CO} - 2k_2c_{CO_2}{c_H}^2 - 2k_3c_{CO_2}{c_H}^2 - 4k_4c_{CO_c}c_F{c_H}^4 \\ &\quad - 2k_5c_{CO_2}c_Fc_{ACA}{c_H}^2 - 2k_7c_{ACA}{c_H}^2 - 2k_{15}c_{Py}{c_H}^2 \\ &\quad - 2k_{17}c_{Al}{c_H}^2 - 2k_{18}c_{Ad}{c_H}^2, \end{aligned} \tag{3.3.12}$$

and initial conditions

$$\mathbf{c}(0) = \mathbf{0}. \tag{3.3.13}$$

To test the predictions of our FBA models, we now simulate this ODE model with various values of Q_H/Q_{CO}. We fix all remaining rate coefficients, k_n to equal one. We also set the carbon monoxide input, $Q_{CO} = 0.1$. Then by choosing a range of values for Q_H, we examine the behaviour of the network for a range of values of Q_H/Q_{CO} and see if the steady-state is reached as predicted by FBA. Figures 3.3, 3.4, and 3.5 show simulations of this model with Q_H equal to 0.2, 0.3, and 0.4. FBA predicts that all three of these choices of Q_H/Q_{CO} would lead to steady-state, however in all cases for our ODE model this is not the case and we find one or more of the metabolites showing unbounded growth. The nonlinearities within the system cause the behaviour to be different to the behaviour predicted by the linear equations used in FBA. It turns out that with all rate parameters k_n equal to one, the ODE model (3.3.3)–(3.3.12) cannot reach a steady-state for any value of Q_H/Q_{CO}. We would expect the network to be able to obtain steady-state for some values within the range predicted by FBA, so there must be some other conditions on our parameters that must be satisfied in order for the ODE model to reach steady-state that could not be deduced from FBA alone. We would like to find these conditions.

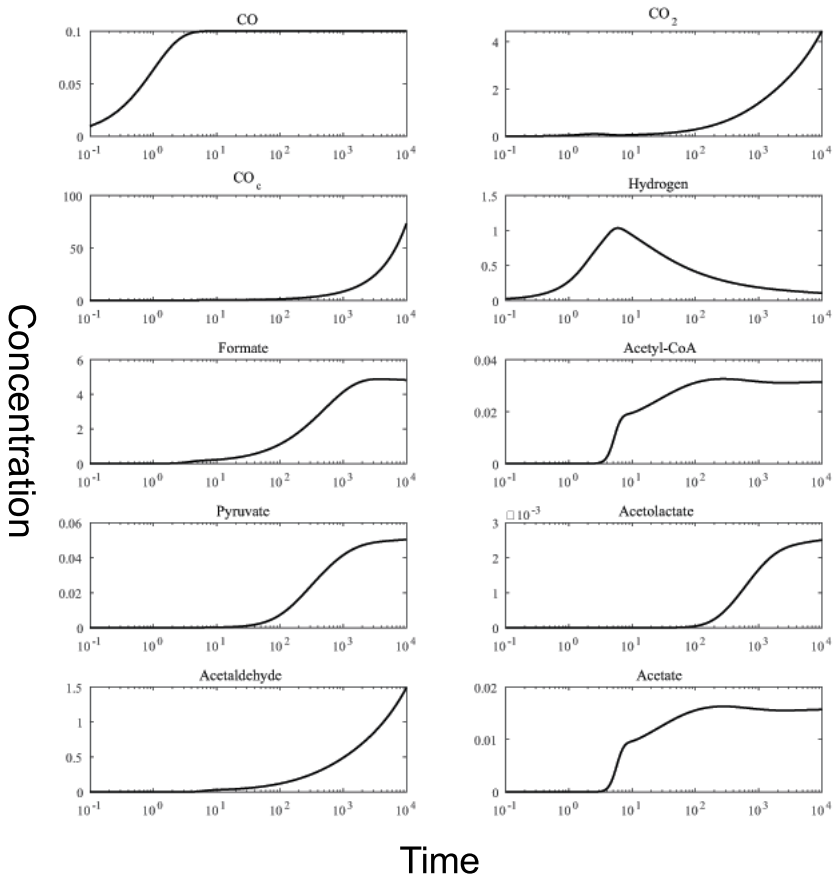

Figure 3.3: Log time plot of the constant input ODE model seen in equations (3.3.3)–(3.3.12), with all parameters set to equal one, except $Q_H = 0.2$. The simulation was ran until the long timescale behaviour was established, and there are no further changes in behaviour past the times shown in this simulation.

3.4 Long timescale asymptotics

One way to find these conditions would be to simply set all time derivatives in the ODEs (3.3.3)–(3.3.12) to zero and analyse the resulting equations. In this way we could find relationships between parameters that must be true in order for this steady-state to exist with non negative concentrations. Due to the large number of parameters within these equations, this proves infeasible. Since we have already seen that we have non steady-states, we instead analyse these states in the long timescale. In the long timescale, some reactions become negligible, and the terms representing them in the ODEs can therefore be discarded. By discarding these terms from the ODEs, we aim to simplify them down to a more easily analysed form. From these simplified forms, we are able to find critical values for parameters which lead the system to switch from one long timescale behaviour to another. By comprehensively examining these long timescale behaviours, we seek the critical parameter values which border the steady-state region, allowing us to quantify the parameter values required for steady-state, without having to directly examine the steady-state equations themselves.

Let us first consider the state shown in Figure 3.4, in which $Q_H = 0.3$. In this state we find the concentration of CO_c tending to infinity, whilst the concentrations of

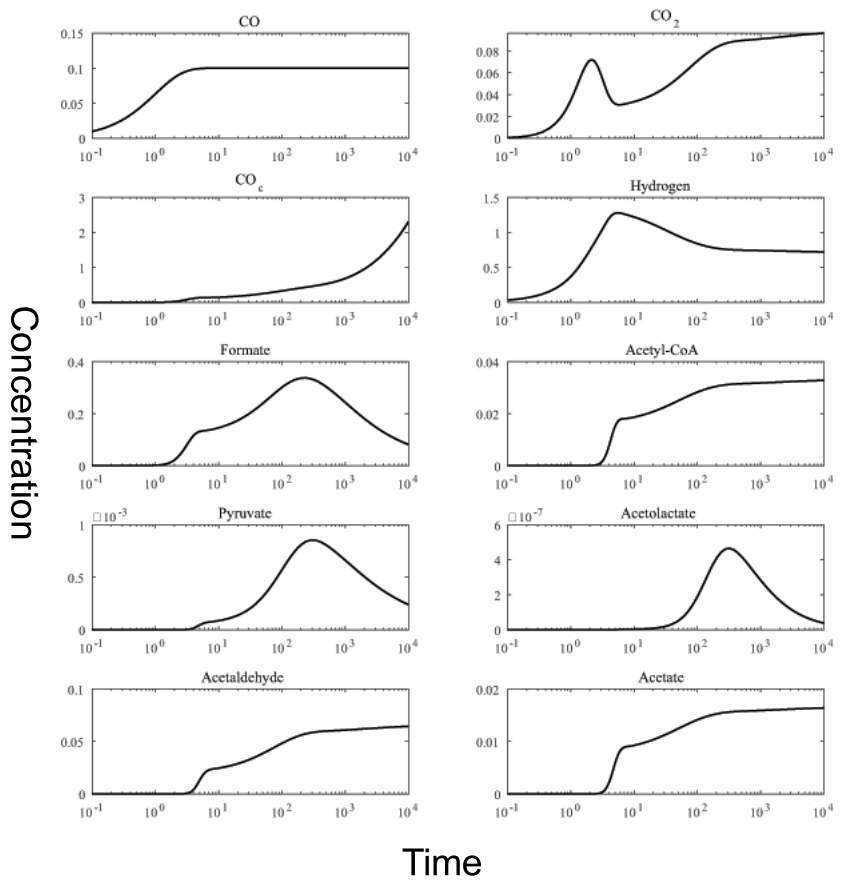

Figure 3.4: As in Figure 3.3, except $Q_H = 0.3$

formate, pyruvate and acetolactate tend towards zero. All other concentrations have a non-zero steady-state. Log-log plots of the four metabolites not tending to a non-zero steady-state are shown in Figure 3.6. Since these log-log plots appear linear, we know that the concentrations of these metabolites behave proportionally to t^x in the long time scale. Measuring the gradients of these log-log plots allows us to numerically obtain values of x for these metabolites. We also know that metabolites that tend to a non zero steady-state scales like t^0 in the long timescale. In this way we obtain scalings for every metabolite in the system,

$$\begin{aligned} c_{CO} &\sim \overline{c_{CO}} t^0, & c_{Py} &\sim \overline{c_{Py}} t^{-0.5}, \\ c_{CO_2} &\sim \overline{c_{CO_2}} t^0, & c_{Al} &\sim \overline{c_{Al}} t^{-1}, \\ c_{CO_c} &\sim \overline{c_{CO_c}} t^{0.5}, & c_{Ad} &\sim \overline{c_{Ad}} t^0, \\ c_F &\sim \overline{c_F} t^{-0.5}, & c_A &\sim \overline{c_A} t^0, \\ c_{ACA} &\sim \overline{c_{ACA}} t^0, & c_H &\sim \overline{c_H} t^0. \end{aligned} \tag{3.4.1}$$

These scalings are substituted back into the ODEs (3.3.3)–(3.3.12), so that each metabolite concentration c_n is replaced by $\bar{c}_n t^x$, where $\bar{c}_n$ is some unknown coefficient, and x is chosen according to equation (3.4.1). Since our concentrations are non-negative, these coefficients, $\bar{c}_n$ must also be non negative. The equations with

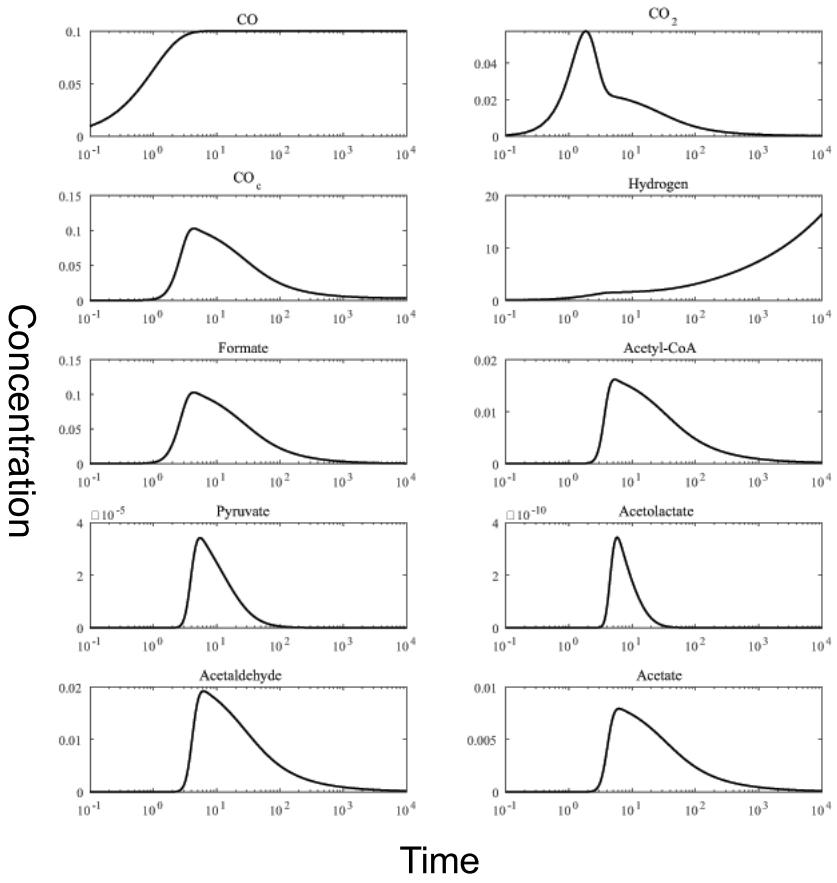

Figure 3.5: As in Figure 3.3, except $Q_H = 0.4$

these scalings substituted in are

$$0 = Q_{CO} - k_1 \bar{c}_{CO} t^0, \tag{3.4.2}$$

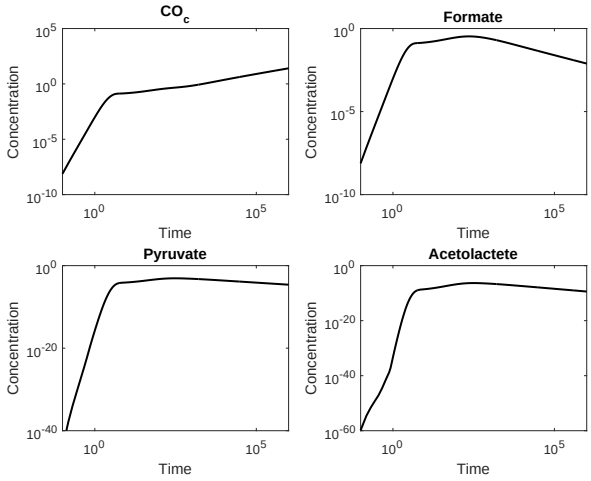

Figure 3.6: Log-log plots of metabolites not tending to nonzero steady-states when $Q_H = 0.3$

$$0 = k_1\bar{c}_{CO}t^0 - k_2\bar{c}_{CO_2}t^0(\bar{c}_H t^0)^2 - k_3\bar{c}_{CO_2}t^0(\bar{c}_H t^0)^2 - k_5\bar{c}_{CO_2}t^0\bar{c}_F t^{-0.5}\bar{c}_{ACA}t^0(\bar{c}_H t^0)^2 + k_6(\bar{c}_{Py}t^{-0.5})^2 + k_{17}\bar{c}_{Al}t^{-1}(\bar{c}_H t^0)^2, \tag{3.4.3}$$

$$0.5\bar{c}_{CO_c}t^{-0.5} = k_3\bar{c}_{CO_2}t^0(\bar{c}_H t^0)^2 - k_4\bar{c}_{CO_c}t^{0.5}\bar{c}_F t^{-0.5}(\bar{c}_H t^0)^4, \tag{3.4.4}$$

$$-0.5\bar{c}_F t^{-1.5} = k_2\bar{c}_{CO_2}t^0(\bar{c}_H t^0)^2 - k_4\bar{c}_{CO_c}t^{0.5}\bar{c}_F t^{-0.5}(\bar{c}_H t^0)^4 - k_5\bar{c}_{CO_2}t^0\bar{c}_F t^{-0.5}\bar{c}_{ACA}t^0(\bar{c}_H t^0)^2, \tag{3.4.5}$$

$$0 = k_4\bar{c}_{CO_c}t^{0.5}\bar{c}_F t^{-0.5}(\bar{c}_H t^0)^4 - k_5\bar{c}_{CO_2}t^0\bar{c}_F t^{-0.5}\bar{c}_{ACA}t^0(\bar{c}_H t^0)^2 - k_7\bar{c}_{ACA}t^0(\bar{c}_H t^0)^2 - k_8\bar{c}_{ACA}t^0 - k_{14}\bar{c}_{ACA}t^0\bar{c}_{Py}t^{-0.5}, \tag{3.4.6}$$

$$-0.5\bar{c}_{Py}t^{-1.5} = k_5\bar{c}_{CO_2}t^0\bar{c}_F t^{-0.5}\bar{c}_{ACA}t^0(\bar{c}_H t^0)^2 - 2k_6(\bar{c}_{Py}t^{-0.5})^2 - k_{14}\bar{c}_{ACA}t^0\bar{c}_{Py}t^{-0.5} - k_{15}\bar{c}_{Py}t^{-0.5}(\bar{c}_H t^0)^2, \tag{3.4.7}$$

$$-\bar{c}_{Al}t^{-2} = k_6(\bar{c}_{Py}t^{-0.5})^2 - k_{16}\bar{c}_{Al}t^{-1} - k_{17}\bar{c}_{Al}t^{-1}(\bar{c}_H t^0)^2, \tag{3.4.8}$$

$$0 = k_7\bar{c}_{ACA}t^0(\bar{c}_H t^0)^2 + k_9\bar{c}_A t^0 - k_{18}\bar{c}_{Ad}t^0(\bar{c}_H t^0)^2, \tag{3.4.9}$$

$$0 = k_8\bar{c}_{ACA}t^0 - k_9\bar{c}_A t^0 - k_{19}\bar{c}_A t^0, \tag{3.4.10}$$

$$0 = Q_H + 2k_1\bar{c}_{CO}t^0 - 2k_2\bar{c}_{CO_2}t^0(\bar{c}_H t^0)^2 - 2k_3\bar{c}_{CO_2}t^0(\bar{c}_H t^0)^2 - 4k_4\bar{c}_{CO_c}t^{0.5}\bar{c}_F t^{-0.5}(\bar{c}_H t^0)^4 - 2k_5\bar{c}_{CO_2}t^0\bar{c}_F t^{-0.5}\bar{c}_{ACA}t^0(\bar{c}_H t^0)^2 - 2k_7\bar{c}_{ACA}t^0(\bar{c}_H t^0)^2 - 2k_{15}\bar{c}_{Py}t^{-0.5}(\bar{c}_H t^0)^2 - 2k_{17}\bar{c}_{Al}t^{-1}(\bar{c}_H t^0)^2 - 2k_{18}\bar{c}_{Ad}t^0(\bar{c}_H t^0)^2. \tag{3.4.11}$$

We combine the powers of t in each term. Terms which do not contain the highest power of t in each equation are discarded, as for large t, each equation is dominated by the terms with the highest powers of t, and all other terms become negligible. In this case, all terms on the left hand side of the equations are found to be insignificant, along with several terms on the right. Powers of t are then cancelled, as all remaining terms in each equation have the same power of t, leaving us with a set of simultaneous equations for the coefficients $\bar{c}_n$ and the parameters k_n. For our current set of scalings, these equations are

$$0 = Q_{CO} - k_1\bar{c}_{CO}, \tag{3.4.12}$$

$$0 = k_1\bar{c}_{CO} - k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - k_3\bar{c}_{CO_2}(\bar{c}_H)^2, \tag{3.4.13}$$

$$0 = k_3\bar{c}_{CO_2}(\bar{c}_H)^2 - k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4, \tag{3.4.14}$$

$$0 = k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4, \tag{3.4.15}$$

$$0 = k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4 - k_7\bar{c}_{ACA}(\bar{c}_H)^2 - k_8\bar{c}_{ACA}, \tag{3.4.16}$$

$$0 = k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H)^2 - k_{14}\bar{c}_{ACA}\bar{c}_{Py} - k_{15}\bar{c}_{Py}(\bar{c}_H)^2, \tag{3.4.17}$$

$$0 = k_6(\bar{c}_{Py})^2 - k_{16}\bar{c}_{Al} - k_{17}\bar{c}_{Al}(\bar{c}_H)^2, \tag{3.4.18}$$

$$0 = k_7\bar{c}_{ACA}(\bar{c}_H)^2 + k_9\bar{c}_A - k_{18}\bar{c}_{Ad}(\bar{c}_H)^2, \tag{3.4.19}$$

$$0 = k_8\bar{c}_{ACA} - k_9\bar{c}_A - k_{19}\bar{c}_A, \tag{3.4.20}$$

$$0 = Q_H + 2k_1\bar{c}_{CO} - 2k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - 2k_3\bar{c}_{CO_2}(\bar{c}_H)^2 - 4k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4 - 2k_7\bar{c}_{ACA}(\bar{c}_H)^2 - 2k_{18}\bar{c}_{Ad}(\bar{c}_H)^2. \tag{3.4.21}$$

We see immediately, by equating (3.4.14) and (3.4.15), that one condition for the existence of this particular state is $k_3 = k_2$. This means our parameter scheme of all k_n equalling one is a critical value for the system. We find that when we set k_3 to be slightly lower than k_2, still with $Q_H = 0.3$ the system tends to a steady-state. If k_3 is set to be slightly higher than k_2, the system instead tends to another distinct state in which some concentrations grow linearly in time. This second behaviour is examined

more closely in Section 3.5.2. In biological terms, this means if the reaction transforming CO_2 into CO_c happens at a greater or equal to rate as the reaction transforming CO_2 into formate, the system does not achieve steady-state with these constant CO and hydrogen inputs. If the reaction to formate happens at a greater rate, the system achieves a steady-state.

This set of equations, (3.4.12)–(3.4.21), can be analytically solved for the coefficients, $\bar{c}_n$. We write $\bar{c}_{CO}$ and $\bar{c}_H$ strictly in terms of the rate parameters k_n and the input fluxes Q_{CO} and Q_H. We then further write $\bar{c}_{CO_2}$ and $\bar{c}_{ACA}$ in terms of the parameters and $\bar{c}_H$. We then write $\bar{c}_{Ad}$ and $\bar{c}_A$ in terms of the parameters, $\bar{c}_{ACA}$ and $\bar{c}_H$, so that we have

$$\bar{c}_{CO} = \frac{Q_{CO}}{k_1}, \tag{3.4.22}$$

$$\bar{c}_{CO_2} = \frac{Q_{CO}}{\bar{c}_H(k_2+k_3)}, \tag{3.4.23}$$

$$\bar{c}_{ACA} = \frac{k_2 Q_{CO}}{(k_2+k_3)(k_2(\bar{c}_H)^2+k_8)}, \tag{3.4.24}$$

$$\bar{c}_{Ad} = \bar{c}_{ACA}\left(\frac{k_7}{k_{18}} + \frac{k_8 k_9}{k_{18}(\bar{c}_H)^2(k_9+k_{19})}\right), \tag{3.4.25}$$

$$\bar{c}_A = \bar{c}_{ACA}\frac{k_8}{k_9+k_{19}}, \tag{3.4.26}$$

$$\bar{c}_H = \sqrt{\frac{-k_8(Q_H - Q_{CO}(2+k_9/(k_9+k_{19})))}{k_7(Q_H-4Q_{CO})}}. \tag{3.4.27}$$

However, since equations (3.4.14) and (3.4.15) simplify into a single equation, $k_2 = k_3$, we only have nine equations for the ten coefficients. We need to find some other balance in order to find explicit equations for all of the coefficients. To obtain another equation, we return to the full set of ODEs with scalings substituted in, (3.4.2)–(3.4.11) and subtract (3.4.4) from (3.4.5) to obtain the equation

$$0.5\bar{c}_{CO_c}t^{-0.5} + 0.5\bar{c}_F t^{-1.5} = k_5\bar{c}_{CO_2}t^0\bar{c}_F t^{-0.5}\bar{c}_{ACA}t^0(\bar{c}_H t^0)^2. \tag{3.4.28}$$

This difference allows us to identify a secondary balance. All terms which scale like t^0 have been cancelled out, leaving us only with terms of a lower order. The dominant power of t in this new balance is $t^{-0.5}$, and we cancel out all terms with a lower power of t to find

$$0.5\bar{c}_{CO_c} = k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H)^2. \tag{3.4.29}$$

Now using equations (3.4.12)–(3.4.21) and (3.4.29), we are able to determine expressions for the remaining coefficients,

$$\bar{c}_{CO_c} = \frac{2k_5 Q_{CO}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H)^2}{k_2+k_3}, \tag{3.4.30}$$

$$\bar{c}_F = \sqrt{\frac{k_2}{2k_4\bar{c}_{ACA}(\bar{c}_H)^2}}, \tag{3.4.31}$$

$$\bar{c}_{Py} = \frac{k_5 Q_{CO}\bar{c}_F\bar{c}_{ACA}}{(k_2+k_3)(k_{14}\bar{c}_{ACA}+k_{15}(\bar{c}_H^2))}, \tag{3.4.32}$$

$$\bar{c}_{Al} = (\bar{c}_{Py})^2\frac{k_6}{k_{16}+k_{17}(\bar{c}_H)^2}. \tag{3.4.33}$$

We have now found equations for all ten of our coefficients. As noted previously, these coefficients must be positive, since our concentrations must be positive, and they

must also be real. Of note is the equation for $\bar{c}_H$, (3.4.21). We see that in order for $\bar{c}_H$ to be real, the expression inside the square root must be positive. We also have that as the denominator tends to zero, then $\bar{c}_H$ tends to infinity. We find that if

$$\frac{Q_H}{Q_{CO}} < 2 + \frac{k_9}{k_9 + k_{19}}, \tag{3.4.34}$$

then the numerator is positive, but the denominator is negative. Similarly, if

$$\frac{Q_H}{Q_{CO}} > 4, \tag{3.4.35}$$

then the numerator is negative, but the denominator is positive, both cases leading to a complex value for $\bar{c}_H$. If Q_H/Q_{CO} remains in the range

$$2 + \frac{k_9}{k_9 + k_{19}} \leq \frac{Q_H}{Q_{CO}} \leq 4, \tag{3.4.36}$$

then $\bar{c}_H$ is positive and real, so this set of scalings exists. As Q_H/Q_{CO} tends the upper bound, $\bar{c}_H$ tends to infinity, whilst as Q_H/Q_{CO} tends towards the lower bound, $\bar{c}_H$ tends towards zero. At these boundary values, the system switches from one long term behaviour to another, and our scalings are no longer applicable.

As we have found a bifurcation that depends only on the value of Q_H/Q_{CO}, it is of interest to observe how the remaining coefficients vary with this ratio. By fixing Q_{CO} and varying Q_H, we see how these coefficients behave as this ratio is varied. We also compare our analytically found values for the coefficients of the metabolites in steady-state with their numerical values to see if they match up as we would expect.

As seen in Figure 3.7, the numerically computed steady-states match up exactly with the coefficients found.

3.5 Off the bifurcation

We found that the parameter regime, $k_2 = k_3$, was a bifurcation. As such the behaviour examined in the previous section is only observable for that specific set of parameters. This behaviour is non generic so we should instead consider parameter regimes where $k_2 < k_3$ and $k_3 < k_2$ separately.

3.5.1 Region 1, $k_3 < k_2$

The region in which Q_{CO}/Q_H fits the limits shown in equation (3.4.36) with $k_3 < k_2$ is the region in which the concentrations of all metabolites tend to nonzero steady-states. If we were to apply the same techniques to this region, we would find that all metabolites would scale like t^0 in the long timescale, and the resulting simultaneous equations would be the equivalent to the full set of ODEs with derivatives set to zero. Since the equations are not simplified at all in this case, we do not examine it any more rigorously.

3.5.2 Region 2, $k_3 > k_2$

To examine the system in this region, we keep hydrogen and CO inputs fixed at $Q_H = 0.3$ and $Q_{CO} = 0.1$. All other coefficients are set to unity except $k_2 = 0.9$. Plots for the coefficients with these values are seen in Figure 3.8. In this region CO_c shows unbounded growth, the concentrations of formate, pyruvate and acetolactate decay to zero, and all other concentrations tend to nonzero steady-states.

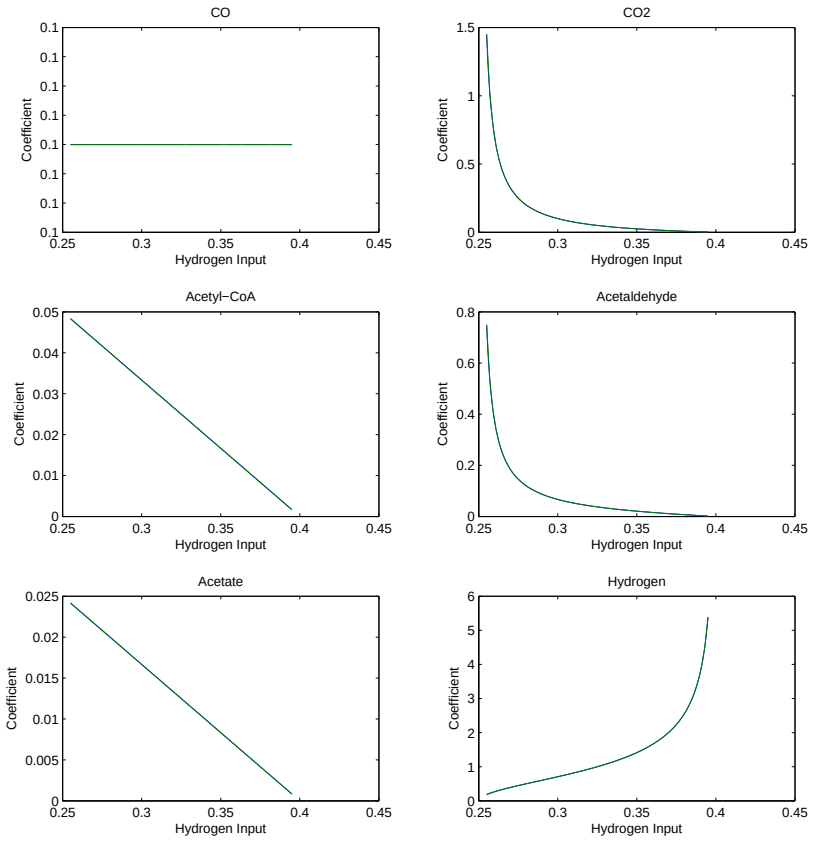

Figure 3.7: Coefficients of the six metabolites in steady-state plotted on the same axis as the numerically computed steady-states of those metabolites as we vary Q_H for the full ODEs. These two lines overlap exactly.

Scalings are obtained for this region in the same way as in Section 3.4, by measuring the gradient of the log-log plots for metabolites not tending to zero. From these measured gradients we find the scalings

$$\begin{aligned} c_{CO} &\sim \overline{c_{CO}}t^0, \quad & c_{Py} &\sim \overline{c_{Py}}t^{-1}, \\ c_{CO_2} &\sim \overline{c_{CO_2}}t^0, \quad & c_{Al} &\sim \overline{c_{Al}}t^{-2}, \\ c_{CO_c} &\sim \overline{c_{CO_c}}t^1, \quad & c_{Ad} &\sim \overline{c_{Ad}}t^0, \\ c_F &\sim \overline{c_F}t^{-1}, \quad & c_A &\sim \overline{c_A}t^0, \\ c_{ACA} &\sim \overline{c_{ACA}}t^0, \quad & c_H &\sim \overline{c_H}t^0. \end{aligned} \tag{3.5.1}$$

Substituting these scalings into the original system of ODEs and dropping non-dominant terms gives us the new set of ten simultaneous equations

$$0 = Q_{CO} - k_1\bar{c}_{CO}, \tag{3.5.2}$$

$$0 = k_1\bar{c}_{CO} - k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - k_3\bar{c}_{CO_2}(\bar{c}_H)^2, \tag{3.5.3}$$

$$\bar{c}_{CO_c} = k_3\bar{c}_{CO_2}(\bar{c}_H)^2 - k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4, \tag{3.5.4}$$

$$0 = k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4, \tag{3.5.5}$$

$$0 = k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4 - k_7\bar{c}_{ACA}(\bar{c}_H)^2 - k_8\bar{c}_{ACA}, \tag{3.5.6}$$

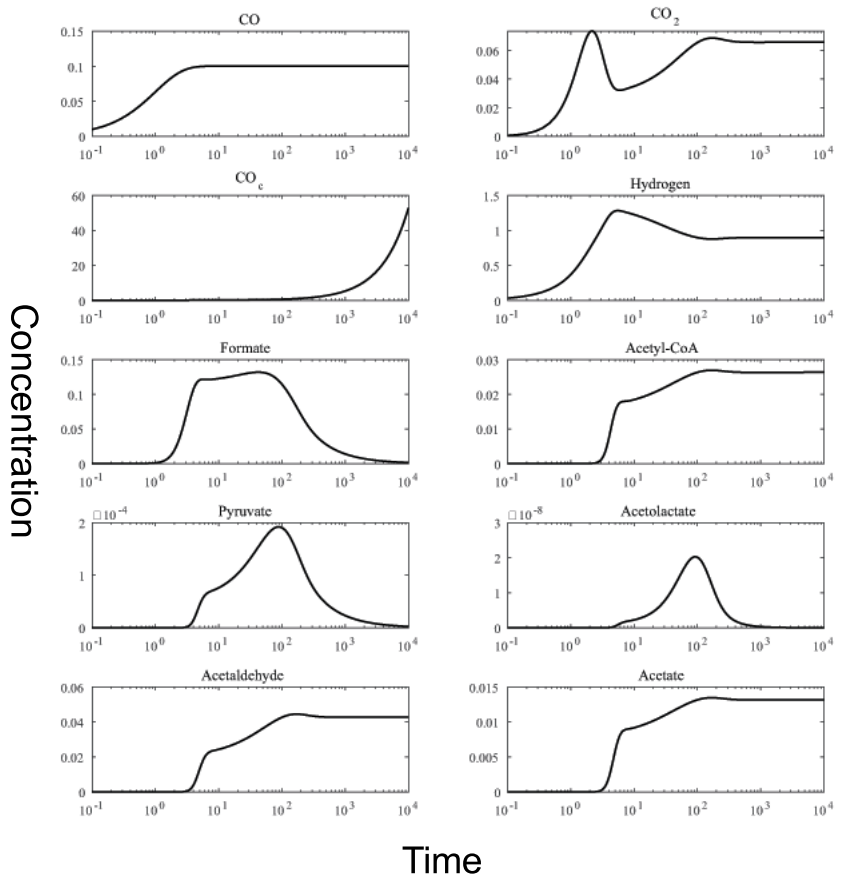

Figure 3.8: Plots of metabolite concentrations against log time inside region 2, with $Q_H = 0.3$, $Q_{CO} = 0.1$, and $k_2 = 0.9$. All other parameters set to equal one.

$$0 = k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H)^2 - k_{14}\bar{c}_{ACA}\bar{c}_{Py} - k_{15}\bar{c}_{Py}(\bar{c}_H)^2, \tag{3.5.7}$$

$$0 = k_6(\bar{c}_{Py})^2 - k_{16}\bar{c}_{Al} - k_{17}\bar{c}_{Al}(\bar{c}_H)^2, \tag{3.5.8}$$

$$0 = k_7\bar{c}_{ACA}(\bar{c}_H)^2 + k_9\bar{c}_A - k_{18}\bar{c}_{Ad}(\bar{c}_H)^2, \tag{3.5.9}$$

$$0 = k_8\bar{c}_{ACA} - k_9\bar{c}_A - k_{19}\bar{c}_A, \tag{3.5.10}$$

$$\begin{aligned} 0 = {} & Q_H + 2k_1\bar{c}_{CO} - 2k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - 2k_3\bar{c}_{CO_2}(\bar{c}_H)^2 \\ & - 4k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4 - 2k_7\bar{c}_{ACA}(\bar{c}_H)^2 - 2k_{18}\bar{c}_{Ad}(\bar{c}_H)^2. \end{aligned} \tag{3.5.11}$$

We solve these equations analytically to find expressions for each of the coefficients $\bar{c}_n$ in terms of the parameters k_n, Q_H and Q_{CO}. For ease of notation we write the coefficients in terms of these parameters and also $\bar{c}_H$ and $\bar{c}_{ACA}$. $\bar{c}_{Py}$ also uses $\bar{c}_{CO_2}$, $\bar{c}_F$ and $\bar{c}_{Al}$ is given in terms of $\bar{c}_{Py}$. Then the expressions are

$$\bar{c}_{CO} = \frac{Q_{CO}}{k_1}, \tag{3.5.12}$$

$$\bar{c}_{CO_2} = \frac{Q_{CO}}{(\bar{c}_H)^2(k_2 + k_3)}, \tag{3.5.13}$$

$$\bar{c}_{CO_c} = \frac{(k_3 - k_2)Q_{CO}}{k_2 + k_3}, \tag{3.5.14}$$

$$\bar{c}_F = \frac{k_2}{(k_2 + k_3)k_4(\bar{c}_H)^2}, \tag{3.5.15}$$

$$\bar{c}_{ACA} = \frac{k_2 Q_{CO}}{(k_2+k_3)(k_7\bar{c}_H^2+k_8)}, \tag{3.5.16}$$

$$\bar{c}_{Py} = \frac{k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H)^2}{k_{14}\bar{c}_{ACA}-k_{15}(\bar{c}_H)^2}, \tag{3.5.17}$$

$$\bar{c}_{Al} = \frac{k_6(\bar{c}_{Py})^2}{k_{16}+k_{17}(\bar{c}_H)^2}, \tag{3.5.18}$$

$$\bar{c}_{Ad} = \bar{c}_{ACA}\left(\frac{k_7}{k_{18}}+\frac{k_8k_9}{k_{18}(\bar{c}_H)^2(k_9+k_{19})}\right), \tag{3.5.19}$$

$$\bar{c}_A = \bar{c}_{ACA}\frac{k_8}{k_9+k_{19}}, \tag{3.5.20}$$

$$\bar{c}_H = \sqrt{\frac{-k_8(Q_H-Q_{CO}(2k_2/k_2+k_3)(2+k_9/(k_9+k_{19})))}{k_7(Q_H-8(k_2/k_2+k_3)Q_{CO})}}. \tag{3.5.21}$$

We note from equation (3.5.14) that in order for a positive coefficient for CO_c, we require $k_3 > k_2$ which is the restriction we have already placed on our parameter scheme. This shows that $k_2 = k_3$ is a critical value for this region as we would expect. As with the previous behaviour, by examining the expression for $\bar{c}_H$, we find conditions on Q_H/Q_{CO} that are required for this behaviour to exist. We find that in order for c_H to take a positive, real, finite value, we must have

$$\frac{2k_2}{k_2+k_3}\left(2+\frac{k_9}{k_9+k_{19}}\right) < \frac{Q_H}{Q_{CO}} < \frac{8k_2}{k_2+k_3}. \tag{3.5.22}$$

If $k_2 = k_3$ then this restriction takes the same values as it would on the bifurcation line explored in Section 3.4. Rather than the bifurcations only depending on the ratio of hydrogen and CO inputs, it also depends on the parameters k_2 and k_3. We now explore further regions to see if there are other bifurcations.

3.6 Region 3, low Hydrogen

In the behaviour examined in Section 3.4 we found a bifurcation at

$$\frac{Q_H}{Q_{CO}} = 2+\frac{k_9}{k_9+k_{19}}. \tag{3.6.1}$$

We now explore the system below this bifurcation at lower values of the input ratio, Q_H/Q_{CO}. As before we find numerical solutions for the system of equations (3.3.3)–(3.3.12) with all rate coefficients set to equal one, the CO input $Q_{CO} = 0.1$ and the hydrogen input $Q_H = 0.2$. With these parameter values, we have $Q_H/Q_{CO} = 2$, and $k_9/(k_9+k_{19})+2 = 2.5$, so we have our input ratio, Q_H/Q_{CO} lower than the critical value. This choice of parameter values was simulated previously and is shown in Figure 3.3. We see for this set of parameters, we have CO_2, CO_c and acetaldehyde concentrations tending to infinity in the long timescale whilst the hydrogen concentration tends to zero. All other metabolites tend towards non zero steady-states. We again obtain possible scalings for these variables by measuring the gradients of the log-log plots, seen in Figure 3.9 From these measured gradients we obtain scalings for the metabolites in the long timescale.

$$\begin{aligned}
c_{CO} &\sim \overline{c_{CO}}t^0, & c_{Py} &\sim \overline{c_{Py}}t^0,\\
c_{CO_2} &\sim \overline{c_{CO_2}}t^{0.5}, & c_{Al} &\sim \overline{c_{Al}}t^0,\\
c_{CO_c} &\sim \overline{c_{CO_c}}t^1, & c_{Ad} &\sim \overline{c_{Ad}}t^{0.5},\\
c_F &\sim \overline{c_F}t^0, & c_A &\sim \overline{c_A}t^0,\\
c_{ACA} &\sim \overline{c_{ACA}}t^0, & c_H &\sim \overline{c_H}t^{-0.25}.
\end{aligned} \tag{3.6.2}$$

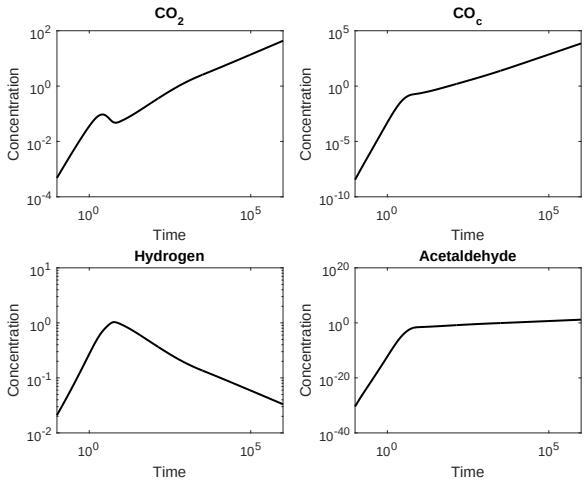

Figure 3.9: Log-log plots of metabolites not tending to non zero steady-states when $Q_H = 0.2$ enlarged to show the relevant long timescale

Substituting these into the ODEs and discarding terms of non dominant powers gives us a new set of ten simultaneous equations.

$$0 = Q_{CO} - k_1\bar{c}_{CO}, \tag{3.6.3}$$

$$0 = k_1\bar{c}_{CO} - k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - k_3\bar{c}_{CO_2}(\bar{c}_H)^2 - k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H^2) + k_6(\bar{c}_{Py})^2, \tag{3.6.4}$$

$$\bar{c}_{CO_c} = k_3\bar{c}_{CO_2}(\bar{c}_H)^2 - k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4, \tag{3.6.5}$$

$$0 = k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4 - k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H^2), \tag{3.6.6}$$

$$0 = k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4 - k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H^2) - k_8\bar{c}_{ACA} - k_{14}\bar{c}_{ACA}\bar{c}_{Py}, \tag{3.6.7}$$

$$0 = k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H)^2 - 2k_6\bar{c}_{Py}^2 k_{14}\bar{c}_{ACA}\bar{c}_{Py}, \tag{3.6.8}$$

$$0 = k_6(\bar{c}_{Py})^2 - k_{16}\bar{c}_{Al}, \tag{3.6.9}$$

$$0 = k_9\bar{c}_A - k_{18}\bar{c}_{Ad}(\bar{c}_H)^2, \tag{3.6.10}$$

$$0 = k_8\bar{c}_{ACA} - k_9\bar{c}_A - k_{19}\bar{c}_A, \tag{3.6.11}$$

$$0 = Q_H + 2k_1\bar{c}_{CO} - 2k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - 2k_3\bar{c}_{CO_2}(\bar{c}_H)^2 - 4k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4 - k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H)^2 - 2k_{18}\bar{c}_{Ad}(\bar{c}_H)^2. \tag{3.6.12}$$

Unlike the equations seen for region 2, these equations are not solvable analytically, however it is possible to solve them numerically. We solve them across a range of Q_H with rate coefficients set to equal one and $Q_{CO} = 0.1$ as in our numerical solution. This allows us to check that the computed coefficients agree with the ODEs. Since we know that this behaviour only exists when

$$\frac{Q_H}{Q_{CO}} < 2 + \frac{k_9}{k_9 + k_{19}}, \tag{3.6.13}$$

we only need to test the ranges where this is true, in this case when $Q_H < 0.25$. We find solutions for the range $0.12 < Q_H < 0.25$. Figure 3.10 shows the coefficient for formate, $\bar{c}_F$ and the coefficient for CO_2, $\bar{c}_{CO_2}$ rapidly increasing as the hydrogen input decreases. We also see a rapid drop in the coefficient for Acetaldehyde, $\bar{c}_{Ad}$. When a coefficient drops to zero, the scaling becomes invalid (as we would have a concentration scaling like $0t^x$). Similarly, when a coefficient tends to infinity, we have a concentration scaling

like $\propto t^x$, suggesting the concentration will scale faster than our current behaviour. The rapidly increasing and decreasing coefficients seen on the lower bound of Q_H suggest the existence of another critical value somewhere close to $Q_H/Q_{CO} = 1.2$, proving the existence of further behaviour for even lower values for Q_H.

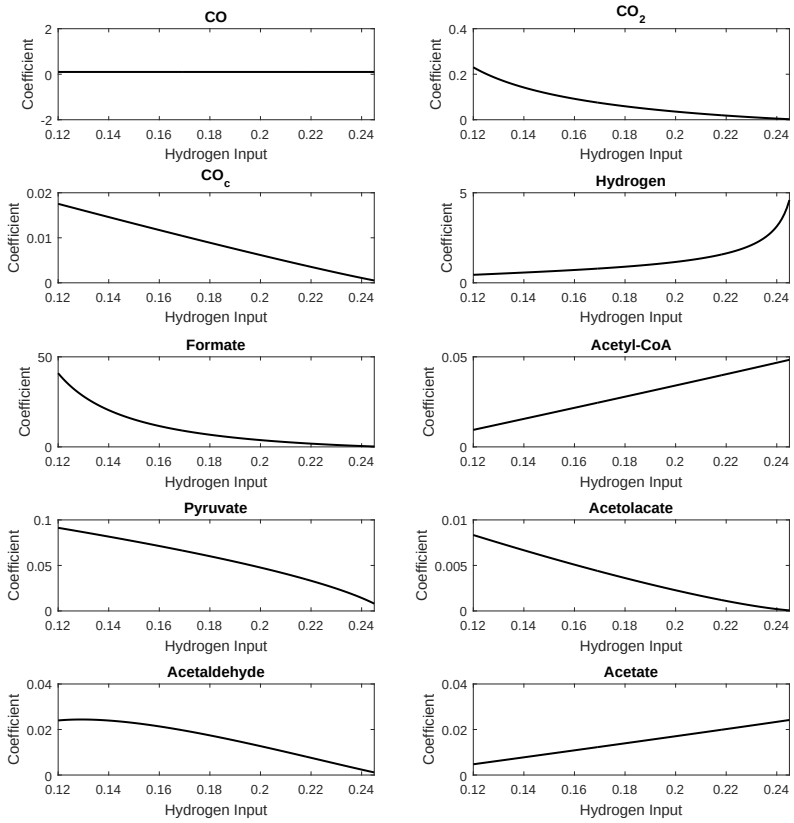

Figure 3.10: Numerical solutions for equations (3.6.3)-(3.6.12) with all rate coefficients, $k_n = 1$, and $Q_{CO} = 0.1$, for Q_H varying between 0.1 and 0.25.

3.7 Region 4, lower hydrogen levels

To check for the existence of another behaviour at even lower hydrogen levels, we simulate the system with a hydrogen input $Q_H = 0.1$. This gives us a input ratio of $Q_H/Q_{CO} = 1$, lower than the approximate value suggested in the previous section. Figure 3.11 shows the plots for this simulation. We see that there is indeed a behaviour separate to the ones seen previously. Here we have the concentrations of CO_2, CO_c and formate tending towards infinity, whilst those of hydrogen, acetyl-CoA and acetate tend to zero. All other metabolites tending to non zero steady-states. This behaviour is different to behaviour three, suggesting there is indeed a critical value for the input ratio somewhere in the range $1 < Q_H/Q_{CO} < 1.2$.

We examine this behaviour in the same way as the previous two examples, first taking the gradients of the log-log plots to find how the concentrations are scaling in the long time scale. Figure 3.12 shows the log-log plots of the concentrations that tend

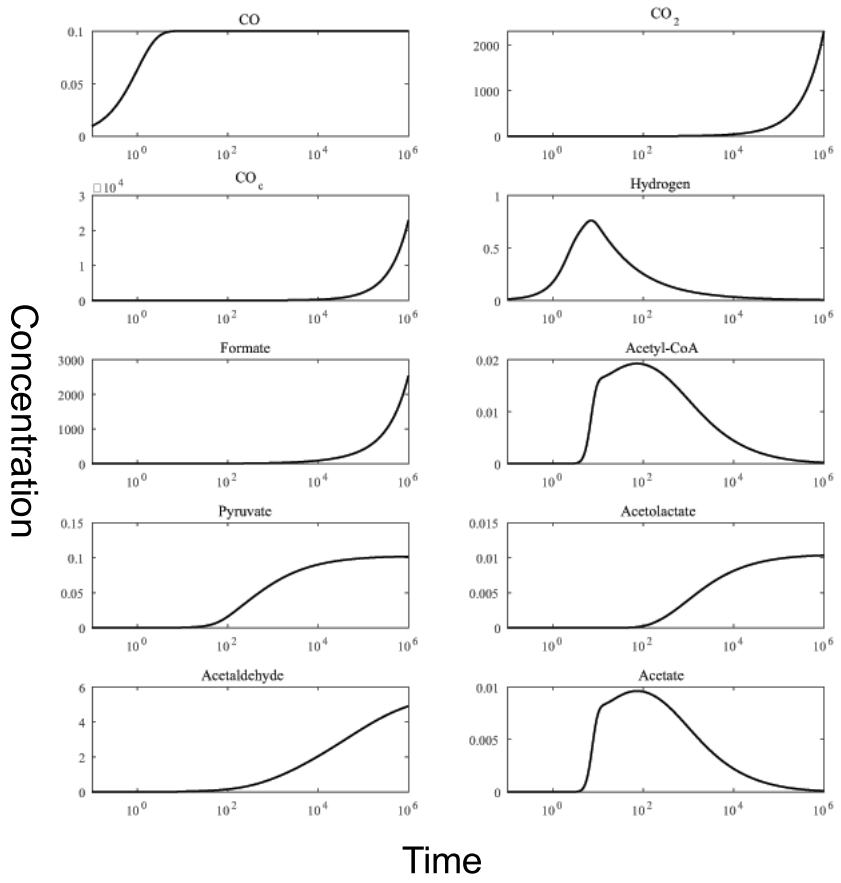

Figure 3.11: Numerical simulation of the ODE system, (3.3.3)–(3.3.12) in region 4 with rate coefficients set to one and an input ratio of $Q_H/Q_{CO} = 1$.

to infinity or zero. Then the scalings are

$$\begin{aligned} c_{CO} &\sim \overline{c_{CO}}t^0, & c_{Py} &\sim \overline{c_{Py}}t^0, \\ c_{CO_2} &\sim \overline{c_{CO_2}}t^1, & c_{Al} &\sim \overline{c_{Al}}t^0, \\ c_{CO_c} &\sim \overline{c_{CO_c}}t^1, & c_{Ad} &\sim \overline{c_{Ad}}t^0, \\ c_F &\sim \overline{c_F}t^1, & c_A &\sim \overline{c_A}t^{-1}, \\ c_{ACA} &\sim \overline{c_{ACA}}t^{-1}, & c_H &\sim \overline{c_H}t^{-0.5}. \end{aligned} \tag{3.7.1}$$

Comparing these scalings to the previous scalings, we see that whilst CO_2 and hydrogen are tending to the same limits as they did in the previous behaviour (infinity and zero respectively), they are scaling much faster than before. CO_c still scales at the same rate as it did previously. We see acetaldehyde now tends to a steady-state, where before it grew in time. Formate now grows in time, whilst before it tended to steady-state, and acetyl-CoA and acetate both now decay to zero where before they had tended to steady-state. As before we substitute in these scalings to the ODEs and then drop the non-dominant terms to find a set of simultaneous equations,

$$0 = Q_{CO} - k_1\bar{c}_{CO}, \tag{3.7.2}$$

$$\bar{c}_{CO_2} = k_1\bar{c}_{CO} - k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - k_3\bar{c}_{CO_2}(\bar{c}_H)^2 - k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H^2) + k_6(\bar{c}_{Py})^2, \tag{3.7.3}$$

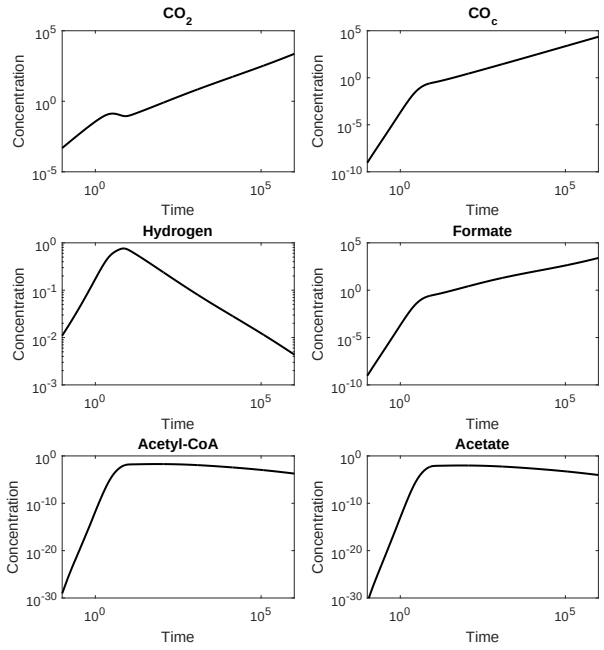

Figure 3.12: Log-log plots of metabolites not tending to non zero steady-states when $Q_H = 0.1$ zoomed on the relevant long timescale

$$\bar{c}_{CO_c} = k_3\bar{c}_{CO_2}(\bar{c}_H)^2 - k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4, \tag{3.7.4}$$

$$\bar{c}_F = k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4 - k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H)^2, \tag{3.7.5}$$

$$0 = k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4 - k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H)^2, \tag{3.7.6}$$

$$0 = k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H)^2 - 2k_6(\bar{c}_{Py})^2, \tag{3.7.7}$$

$$0 = k_6(\bar{c}_{Py})^2 - k_{16}\bar{c}_{Al}, \tag{3.7.8}$$

$$0 = k_9\bar{c}_A - k_{18}\bar{c}_{Ad}(\bar{c}_H)^2, \tag{3.7.9}$$

$$0 = k_8\bar{c}_{ACA} - k_9\bar{c}_A - k_{19}\bar{c}_A, \tag{3.7.10}$$

$$\begin{aligned} 0 = {} & Q_H + 2k_1\bar{c}_{CO} - 2k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - 2k_3\bar{c}_{CO_2}(\bar{c}_H)^2 \\ & - 4k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4 - k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H)^2. \end{aligned} \tag{3.7.11}$$

As with the equations for behaviour 3, seen in Section 3.6, these are not solvable analytically. Instead we solve these equations numerically across a range of Q_H, with other parameters fixed to equal one as in the previous example. Figure 3.13 shows the values of the coefficients for the range of hydrogen input $0 \leq Q_H \leq 1$. Since we are only considering a positive hydrogen input, we do not need to examine negative values, though since no coefficient appears to decay to zero or tend to infinity at $Q_H = 0$, there is evidence to suggest that these equations would hold with negative values.

We see that as Q_H increases, the coefficients for multiple metabolites start to increase rapidly, also suggesting the bifurcation somewhere between

$$1 < Q_H/Q_{CO} < 1.2.$$

We could either examine the region in which we know the bifurcation exists in more detail to try and narrow down where it is, or try to find the bifurcation point analytically with one of the sets of equations (3.6.3)–(3.6.12) or (3.7.2)–(3.7.11).

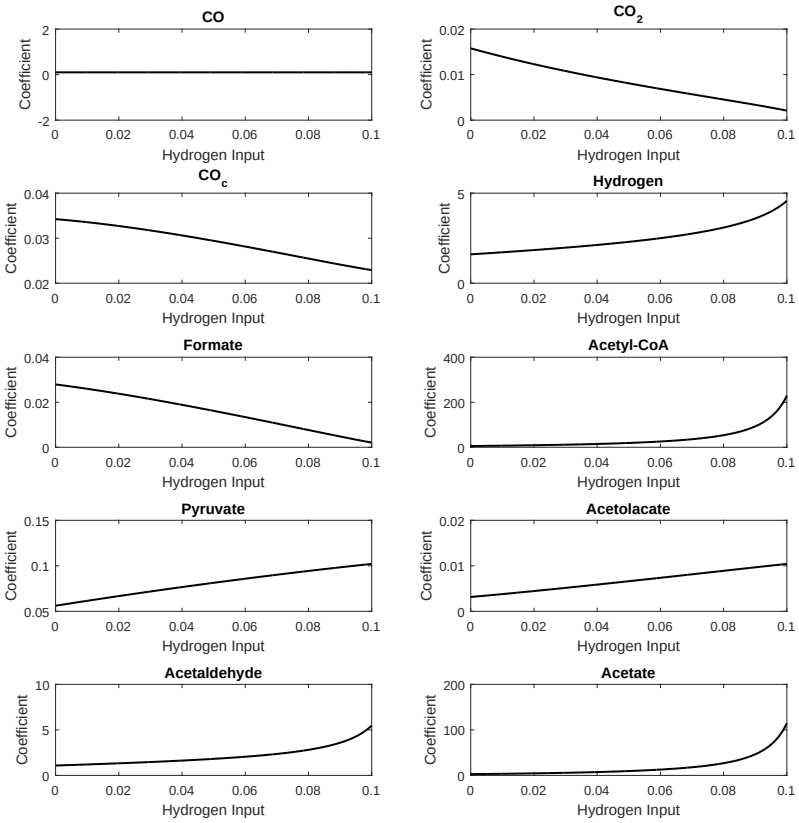

Figure 3.13: Numerical solutions for equations (3.7.2)-(3.7.11) with all $k_n = 1$, $Q_{CO} = 0.1$, and Q_H varying between zero and 0.1

3.8 Higher Levels of Hydrogen

We also have interest in the case where they hydrogen input is much higher than the CO input, that is when $Q_H/Q_{CO} > 4$. Figure 3.5 shows plots for the ODEs, (3.3.3)–(3.3.12), when $Q_H = 0.4$ and $Q_{CO} = 0.1$. We see that the transient behaviour happens on a very small timescale, with most metabolites experiencing a slight spike in concentration before rapidly decaying to zero, whilst the concentration of CO maintains a non-zero steady-state and the concentration of hydrogen increases rapidly. Examining the behaviour of the metabolites on the long timescale and substituting scalings in as before leads us to find that, as for behaviours 1 and 2, the parameter scheme in which $k_2 = k_3$ is a bifurcation. Rather than examining this specific behaviour more closely, we instead examine the behaviours when $k_3 < k_2$, and $k_3 > k_2$ separately.

3.8.1 Region 5, $k_2 > k_3$

By setting $Q_H = 0.41$ and $k_3 = 0.9$, with all other parameters set to one as before, we are able to examine the system with high hydrogen input. Here the system favours formate production over CO_c production. Figure 3.14 shows the numerical simulation of the system in this parameter regime. We see the concentrations of most metabolites tend to zero, but we have the concentrations of formate and hydrogen growing indef-

initely in time. Through a combination of balancing and numerically measuring the gradients of the log-log plots in the long timescale, we obtain a set of scalings for this behaviour,

$$\begin{aligned} c_{CO} &\sim \overline{c_{CO}} t^0, & c_{Py} &\sim \overline{c_{Py}} t^{-3}, \\ c_{CO_2} &\sim \overline{c_{CO_2}} t^{-2}, & c_{Al} &\sim \overline{c_{Al}} t^{-8}, \\ c_{CO_c} &\sim \overline{c_{CO_c}} t^{-5}, & c_{Ad} &\sim \overline{c_{Ad}} t^{-2}, \\ c_F &\sim \overline{c_F} t^1, & c_A &\sim \overline{c_A} t^{-2}, \\ c_{ACA} &\sim \overline{c_{ACA}} t^{-2}, & c_H &\sim \overline{c_H} t^1. \end{aligned} \tag{3.8.1}$$

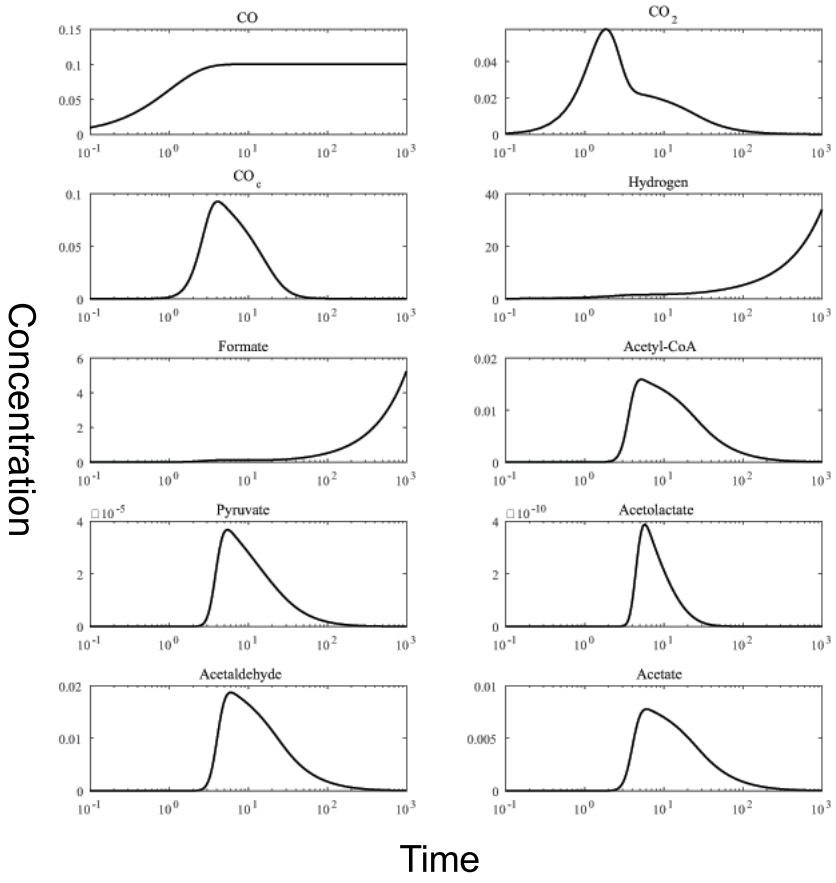

Figure 3.14: Numerical simulation of the ODE system in region 5, with rate coefficients set to one except k_3 set to 0.9, $Q_{CO} = 0.1$ and $Q_H = 0.41$

Examining the system with these scalings, we see CO_c is decreasing much more rapidly than before, and acetolactate is also decreasing at a much higher rate than in any previous system. This quicker decrease for CO_c balances the increasing formate and hydrogen in the reaction that creates acetyl-CoA from hydrogen, CO_c and formate, maintaining a constant production level in the long timescale.

As before, we now substitute these scalings back into the original set of differential equations (3.3.3)–(3.3.12) and drop the non-dominant terms to give the set of simultaneous equations

$$0 = Q_{CO} - k_1 \bar{c}_{CO}, \tag{3.8.2}$$

$$0 = k_1\bar{c}_{CO} - k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - k_3\bar{c}_{CO_2}(\bar{c}_H)^2 \tag{3.8.3}$$

$$0 = k_3\bar{c}_{CO_2}(\bar{c}_H)^2 - k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4, \tag{3.8.4}$$

$$\bar{c}_F = k_2\bar{c}_{CO_2}(\bar{c}_H)^2) - k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4, \tag{3.8.5}$$

$$0 = k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4 - k_7\bar{c}_{ACA}(\bar{c}_H)^2, \tag{3.8.6}$$

$$0 = k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H)^2) - k_{15}\bar{c}_{Py}(\bar{c}_H)^2, \tag{3.8.7}$$

$$0 = k_6(\bar{c}_{Py})^2 - k_{17}\bar{c}_{Al}(\bar{c}_H)^2, \tag{3.8.8}$$

$$0 = k_7\bar{c}_{ACA}(\bar{c}_H)^2 - k_{18}\bar{c}_{Ad}(\bar{c}_H)^2, \tag{3.8.9}$$

$$0 = k_8\bar{c}_{ACA} - k_9\bar{c}_A - k_{19}\bar{c}_A, \tag{3.8.10}$$

$$\bar{c}_H = Q_H + 2k_1\bar{c}_{CO} - 2k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - 2k_3\bar{c}_{CO_2}(\bar{c}_H)^2 - 4k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4 - 2k_7\bar{c}_{ACA}(\bar{c}_H)^2 - 2k_{18}\bar{c}_{Ad}(\bar{c}_H)^2. \tag{3.8.11}$$

Unlike the equations found for the two low hydrogen input behaviours seen in Sections 3.6 and 3.7, these equations are relatively easy to solve analytically, allowing us to write down values for all of them in terms of only the parameters. For ease of notation, we show them in terms of the parameters, Q_{CO}, Q_H, k_n, and the coefficient $\bar{c}_H$. The coefficients are

$$\bar{c}_{CO} = \frac{Q_{CO}}{k_1}, \tag{3.8.12}$$

$$\bar{c}_{CO_2} = \frac{Q_{CO}}{(k_2 + k_3)}, \tag{3.8.13}$$

$$\bar{c}_{CO_c} = \frac{k_3}{(k_2 - k_3)k_4(\bar{c}_H)^4}, \tag{3.8.14}$$

$$\bar{c}_F = \frac{Q_{CO}(k_2 - k_3)}{k_2 + k_3}, \tag{3.8.15}$$

$$\bar{c}_{ACA} = \frac{k_3 Q_{CO}}{(k_2 + k_3)k_7(\bar{c}_H)^2}, \tag{3.8.16}$$

$$\bar{c}_{Py} = \frac{(k_2 - k_3)k_3k_5{Q_{CO}}^3}{(k_2 + k_3)^3k_7k_{15}(\bar{c}_H)^4}, \tag{3.8.17}$$

$$\bar{c}_{Ad} = \frac{(k_2 - k_3)^2k_3^2k_5^2k_6{Q_{CO}}^6}{(k_2 + k_3)^6k_7^2k_{15}^2k_{17}(\bar{c}_H)^{10}}, \tag{3.8.18}$$

$$\bar{c}_{Al} = \frac{k_3 Q_{CO}}{(k_2 + k_3)k_18(\bar{c}_H)^2}, \tag{3.8.19}$$

$$\bar{c}_A = \frac{k_3k_8Q_{CO}}{(k_2 + k_3)k_7(k_9 + k_19)(\bar{c}_H)^2}, \tag{3.8.20}$$

$$\bar{c}_H = Q_H - \frac{8k_3Q_{CO}}{k_2 + k_3}. \tag{3.8.21}$$

In our first set of equations, (3.4.12)–(3.4.21) we found a bifurcation at $Q_H/Q_{CO} = 4$. However, this was on the bifurcation $k_2 = k_3$, so for $k_2 \neq k_3$ the bifurcation may occur at other values of Q_H/Q_{CO}. In the region where $k_3 < k_2$, the expression for $\bar{c}_H$, equation (3.8.21) tells us the critical value is actually

$$\frac{Q_H}{Q_{CO}} = \frac{8k_3}{k_2 + k_3}. \tag{3.8.22}$$

When $k_2 = k_3$, this agrees with the previously found critical value seen in equation 3.4.35. In addition, since $k_3 < k_2$, $k_3/(k_2 + k_3) < 0.5$, so the bifurcation point for Q_H/Q_{CO} is never greater than four. We also note that from the expression for $\bar{c}_{CO_c}$,

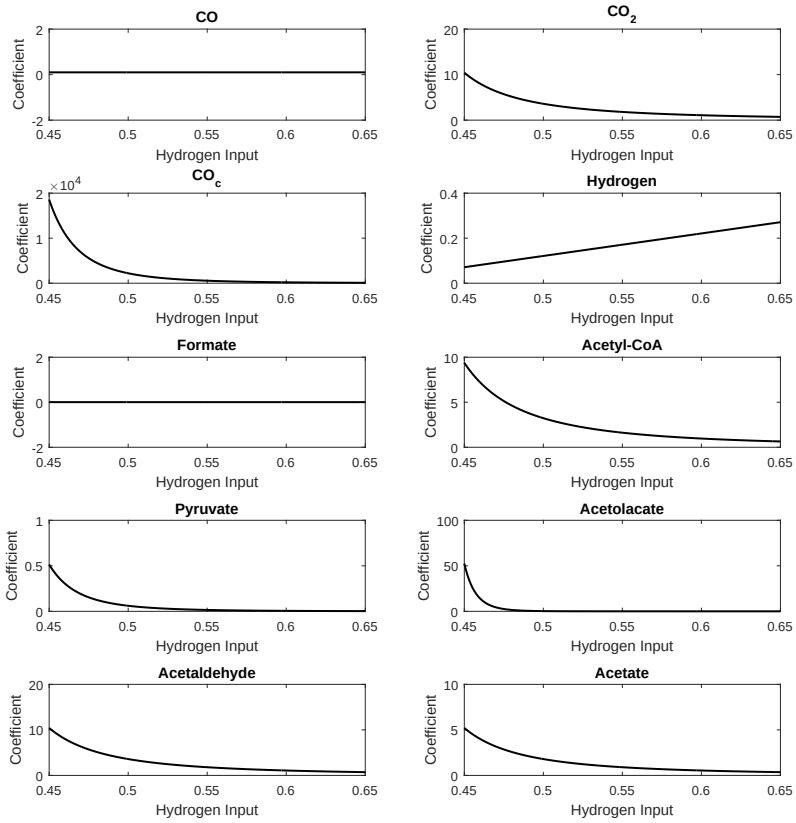

Figure 3.15: Graph of solutions to equations (3.8.12)–(3.8.21) with $k_n = 1$, $Q_{CO} = 0.1$, and Q_H in the range $0.45 < Q_H \leq 0.65$

equation (3.8.14), $k_2 = k_3$ is a bifurcation again, since as k_3 tends to k_2, $\bar{c}_{CO_c}$ tends to infinity.

We plot these equations for varying Q_H with a fixed k_3. Plots for $k_3 = 0.9$ are shown in Figure 3.15. We see that many coefficients tend to infinity as $\bar{c}_H$ tends to zero. This is expected as the next region past this bifurcation is the previously observed steady-state region, so we expect the metabolites that tend to zero in this region to have coefficients that tend to infinity towards the bifurcation to allow for a steady-state. However we notice that $\bar{c}_F$ does not tend to zero. If the region past the critical value was the steady-state region, we would expect this coefficient to tend to zero, allowing the concentration of formate to scale to like a lower power of t. Since it does not, formate should continue scale proportionally to t^1 past this critical value. This suggests a further region between the steady-state region and this high hydrogen region in which most metabolites tend to either steady-state or at least scale like a higher power of t than in this region, but where formate still scales like t^1, and we will examine this region in Section 3.9. First however, we examine the region in which $k_3 > k_2$.

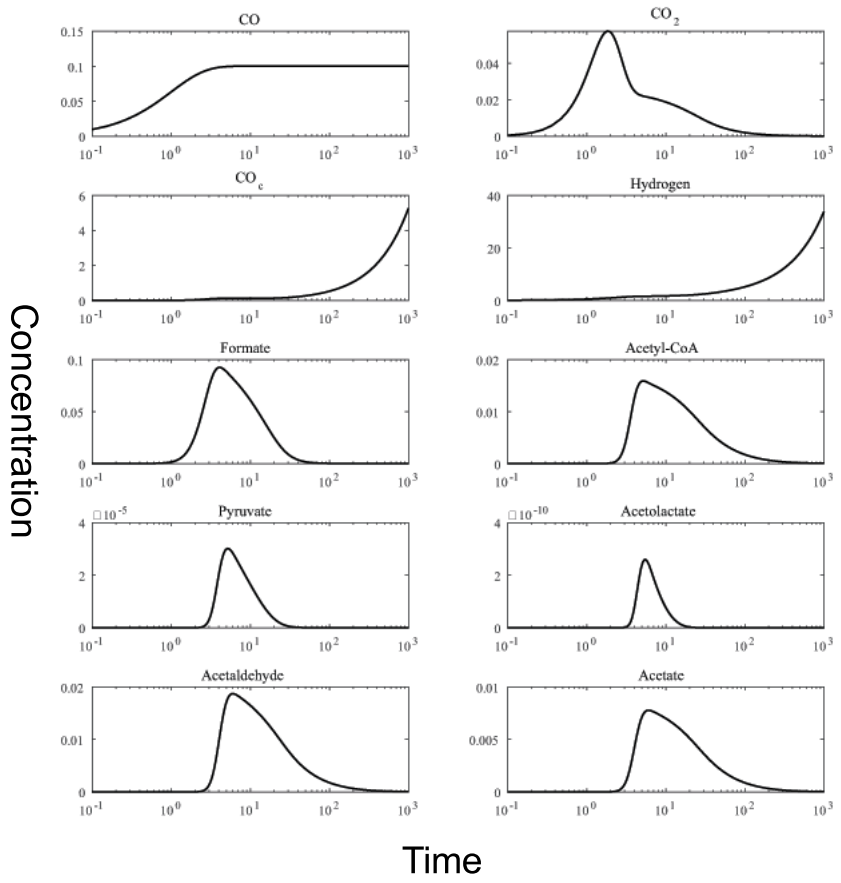

Figure 3.16: Numerical simulation of the ODE system in region 6, with rate coefficients set to one except k_2 set to 0.9, $Q_{CO} = 0.1$ and $Q_H = 0.41$

3.8.2 Region 6, $k_3 > k_2$

To examine this region, we set $Q_H = 0.41$ as before and all other parameters to equal one except $k_2 = 0.9$. This represents a regime in which the system favours CO_c production over formate production. Figure 3.16 shows the simulations for this parameter scheme. We notice that in this regime, we have CO_c tending to infinity, whilst formate decays to zero which is the opposite of the previous regime. As the system is biased towards producing CO_c, it makes sense that there is an excess of this, whereas when the system was biased towards formate production, there was an excess of formate. As before we examine the long timescale behaviour and find scalings for the metabolites with a combination of asymptotic balancing and numerical measurement. The scalings for this parameter regime are

$$\begin{aligned} c_{CO} &\sim \overline{c_{CO}} t^0, & c_{Py} &\sim \overline{c_{Py}} t^{-9}, \\ c_{CO_2} &\sim \overline{c_{CO_2}} t^{-2}, & c_{Al} &\sim \overline{c_{Al}} t^{-20}, \\ c_{CO_c} &\sim \overline{c_{CO_c}} t^1, & c_{Ad} &\sim \overline{c_{Ad}} t^{-2}, \\ c_F &\sim \overline{c_F} t^{-5}, & c_A &\sim \overline{c_A} t^{-2}, \\ c_{ACA} &\sim \overline{c_{ACA}} t^{-2}, & c_H &\sim \overline{c_H} t^1. \end{aligned} \tag{3.8.23}$$

We see that acetolactate has an even sharper decline than before. This is shown

on the graphs as acetolactate undergoes a very rapid decrease (within $t = 100$). We also notice that the scalings for CO_c and formate have swapped as we crossed the bifurcation $k_2 = k_3$. By substituting these scalings into the ODEs and dropping non-dominant terms, we find the set of simultaneous equations

$$0 = Q_{CO} - k_1\bar{c}_{CO}, \tag{3.8.24}$$
$$0 = k_1\bar{c}_{CO} - k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - k_3\bar{c}_{CO_2}(\bar{c}_H)^2, \tag{3.8.25}$$
$$\bar{c}_{CO_c} = k_3\bar{c}_{CO_2}(\bar{c}_H)^2 - k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4, \tag{3.8.26}$$
$$0 = k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4, \tag{3.8.27}$$
$$0 = k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4 - k_7\bar{c}_{ACA}(\bar{c}_H)^2, \tag{3.8.28}$$
$$0 = k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H)^2 - k_{15}\bar{c}_{Py}(\bar{c}_H)^2, \tag{3.8.29}$$
$$0 = k_6(\bar{c}_{Py})^2 - k_{17}\bar{c}_{Al}(\bar{c}_H)^2, \tag{3.8.30}$$
$$0 = k_7\bar{c}_{ACA}(\bar{c}_H)^2 - k_{18}\bar{c}_{Ad}(\bar{c}_H)^2, \tag{3.8.31}$$
$$0 = k_8\bar{c}_{ACA} - k_9\bar{c}_A - k_{19}\bar{c}_A, \tag{3.8.32}$$
$$\bar{c}_H = Q_H + 2k_1\bar{c}_{CO} - 2k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - 2k_3\bar{c}_{CO_2}(\bar{c}_H)^2 - 4k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4 - 2k_7\bar{c}_{ACA}(\bar{c}_H)^2 - 2k_{18}\bar{c}_{Ad}(\bar{c}_H)^2. \tag{3.8.33}$$

Interestingly, these equations are nearly the same as the previous set of equations, (3.8.2)–(3.8.11), but we have kept a $\bar{c}_{CO_c}$ on the left hand side of the third equations, rather than a $\bar{c}_F$ on the left hand side of the fourth. The resulting expressions for the coefficients will therefore also likely be similar. Solving equations (3.8.24)–(3.8.33) analytically gives us

$$\bar{c}_{CO} = \frac{Q_{CO}}{k_1}, \tag{3.8.34}$$
$$\bar{c}_{CO_2} = \frac{Q_{CO}}{(k_2 + k_3)}, \tag{3.8.35}$$
$$\bar{c}_{CO_c} = \frac{Q_{CO}(k_3 - k_2)}{k_2 + k_3}, \tag{3.8.36}$$
$$\bar{c}_F = \frac{k_2}{(k_3 - k_2)k_4(\bar{c}_H)^4}, \tag{3.8.37}$$
$$\bar{c}_{ACA} = \frac{k_2 Q_{CO}}{(k_2 + k_3)k_7(\bar{c}_H)^2}, \tag{3.8.38}$$
$$\bar{c}_{Py} = \frac{k_2^2 k_5 {Q_{CO}}^2}{(k_2 + k_3)^2(k_3 - k_2)k_{15}(\bar{c}_H)^8}, \tag{3.8.39}$$
$$\bar{c}_{Ad} = \frac{k_2^4 k_5^2 k_6 {Q_{CO}}^4}{(k_2 + k_3)^4(k_3 - k_2)^2 k_{15}^2 k_{17}(\bar{c}_H)^{18}}, \tag{3.8.40}$$
$$\bar{c}_{Al} = \frac{k_2 Q_{CO}}{(k_2 + k_3)k_1 8(\bar{c}_H)^2}, \tag{3.8.41}$$
$$\bar{c}_A = \frac{k_2 k_8 Q_{CO}}{(k_2 + k_3)k_7(k_9 + k_{19})(\bar{c}_H)^2}, \tag{3.8.42}$$
$$\bar{c}_H = Q_H - \frac{8k_2 Q_{CO}}{k_2 + k_3}. \tag{3.8.43}$$

Most of these expressions are the same as in region 5, with k_3 and k_2 swapped. The quantities $\bar{c}_{CO_c}$ and $\bar{c}_F$ have also swapped, whilst $\bar{c}_{Py}$ and $\bar{c}_{Al}$ differ more. We see from the equation for $\bar{c}_F$ that $k_2 = k_3$ is still a bifurcation. We also see from the equation for $\bar{c}_H$ a bifurcation when $Q_H/Q_{CO} = \frac{8k_2}{k_2+k_3}$. What we originally though was a single

bifurcation at $Q_H/Q_{CO} = 4$ is actually a piecewise function of k_2 and k_3, that is written as

$$Q_H/Q_{CO} = \begin{cases} \dfrac{8k_3}{k_2+k_3} & k_2 > k_3 \\ 4 & k_2 = k_3 \\ \dfrac{8k_2}{k_2+k_3} & k_2 < k_3 \end{cases} \tag{3.8.44}$$

Whilst the region where Q_H is greater than this bifurcation has been referred to as the high hydrogen region, if $k_3 \ll k_2$ or $k_2 \ll k_3$, this region extends well into levels of hydrogen that are much lower.

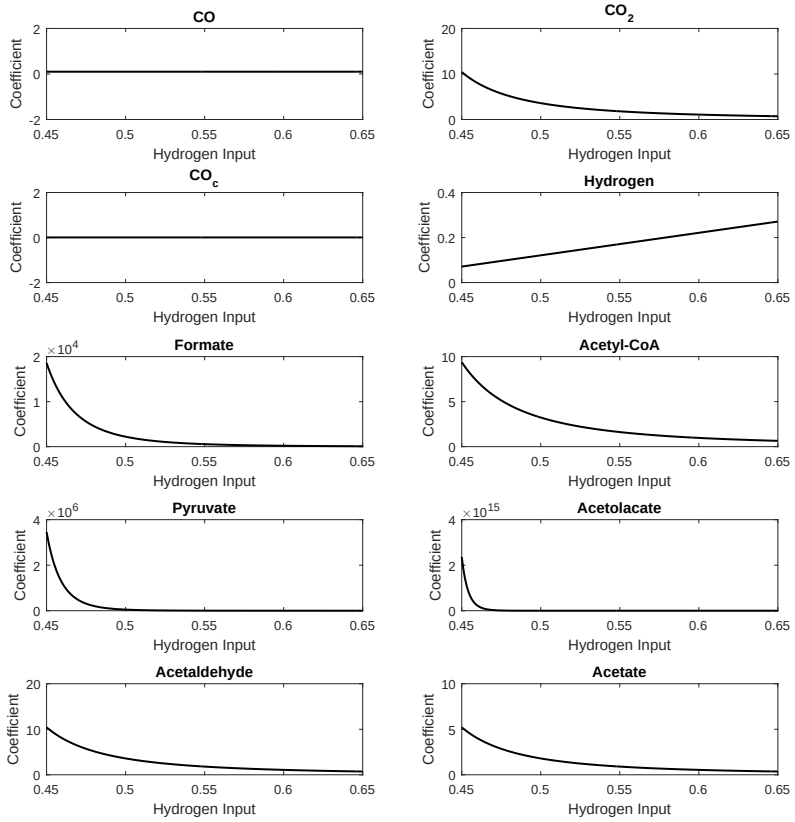

Figure 3.17: Graph of solutions to equations (3.8.34)–(3.8.43) with $k_n = 1$, $Q_{CO} = 0.1$, and Q_H in the range $0.45 < Q_H \leq 0.65$

We plot the coefficients for a fixed value of k_2 across varying Q_H, with all other parameters set to one. Figure 3.17 shows these plots for $k_2 = 0.9$. We see again the majority of coefficients tending to infinity as $\bar{c}_H$ tends to zero. We observe $\bar{c}_{CO_c}$ remaining constant and not tending to zero. This is not a problem as $\bar{c}_F$ was in the previous example, as the next region past the bifurcation as Q_H decreases was not identified to be a steady-state.

3.9 Region 7. A region between high hydrogen and steady-state

We have deduced that there is another region in the parameter space between the high hydrogen region where $k_2 > k_3$ examined in Section 3.8.1, and the steady-state region seen when $k_2 > k_3$ and $Q_H/Q_{CO} < \frac{8k_2}{k_2+k_3}$. However, as we are able to describe the shape of the steady-state region analytically, it is not clear how large this intermediate region may be. We investigate the area near the boundary of the region and try to find a new behaviour that is different to that seen in Section 3.8.1. A set of parameters that provides us with this new behaviour is $Q_H = 0.34$, $Q_{CO} = 0.1$, $k_3 = 0.8$, and all other parameters set to one. Plots for the system with these parameter values are given in Figure 3.18. We obtain scalings for this behaviour in the usual way. The scalings are

$$\begin{aligned} c_{CO} &\sim \overline{c_{CO}}t^0, & c_{Py} &\sim \overline{c_{Py}}t^{-1}, \\ c_{CO_2} &\sim \overline{c_{CO_2}}t^{-1}, & c_{Al} &\sim \overline{c_{Al}}t^{-3}, \\ c_{CO_c} &\sim \overline{c_{CO_c}}t^{-3}, & c_{Ad} &\sim \overline{c_{Ad}}t^{-1}, \\ c_F &\sim \overline{c_F}t^1, & c_A &\sim \overline{c_A}t^{-1}, \\ c_{ACA} &\sim \overline{c_{ACA}}t^{-1}, & c_H &\sim \overline{c_H}t^{0.5}. \end{aligned} \tag{3.9.1}$$

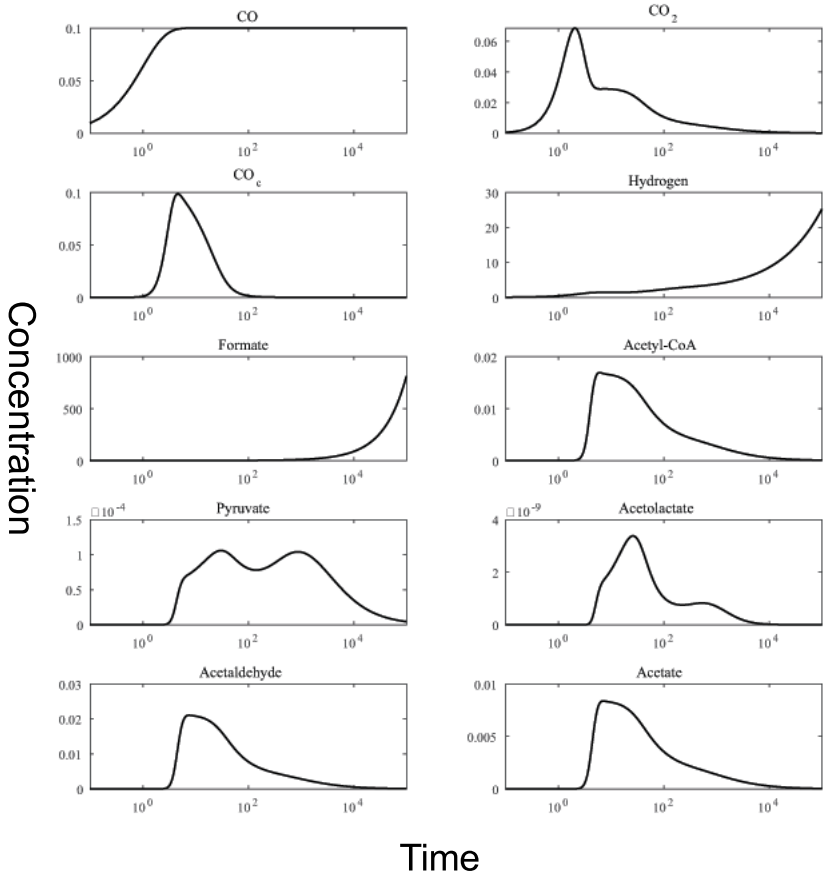

Figure 3.18: Numerical simulation of the ODE system in the region 7, with rate coefficients set to one except k_3 set to 0.8, $Q_{CO} = 0.1$ and $Q_H = 0.34$)

We see that whilst the same metabolites are tending to zero or infinity as in behaviour 5, their exponents are closer to zero. In particular we observe hydrogen now scales like $t^{0.5}$ rather than t^1. Formate still scales like t^1, as predicted by equation (3.8.14) for behaviour 5. We substitute these scalings into the original ODEs, (3.3.3)–(3.3.12), and drop non dominant terms to find the set of simultaneous equations for this region,

$$0 = Q_{CO} - k_1\bar{c}_{CO}, \tag{3.9.2}$$
$$0 = k_1\bar{c}_{CO} - k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - k_3\bar{c}_{CO_2}(\bar{c}_H)^2 - k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H)^2, \tag{3.9.3}$$
$$0 = k_3\bar{c}_{CO_2}(\bar{c}_H)^2 - k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4, \tag{3.9.4}$$
$$\bar{c}_F = k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4 - k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H)^2, \tag{3.9.5}$$
$$0 = k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4 - k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H)^2 - k_7\bar{c}_{ACA}(\bar{c}_H)^2, \tag{3.9.6}$$
$$0 = k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H)^2 - k_{15}\bar{c}_{Py}(\bar{c}_H)^2, \tag{3.9.7}$$
$$0 = k_6(\bar{c}_{Py})^2 - k_{17}\bar{c}_{Al}(\bar{c}_H)^2, \tag{3.9.8}$$
$$0 = k_7\bar{c}_{ACA}(\bar{c}_H)^2 - k_{18}\bar{c}_{Ad}(\bar{c}_H)^2, \tag{3.9.9}$$
$$0 = k_8\bar{c}_{ACA} - k_9\bar{c}_A - k_{19}\bar{c}_A, \tag{3.9.10}$$
$$\begin{aligned} 0 = {} & Q_H + 2k_1\bar{c}_{CO} - 2k_2\bar{c}_{CO_2}(\bar{c}_H)^2 - 2k_3\bar{c}_{CO_2}(\bar{c}_H)^2 - 4k_4\bar{c}_{CO_c}\bar{c}_F(\bar{c}_H)^4 \\ & - 2k_5\bar{c}_{CO_2}\bar{c}_F\bar{c}_{ACA}(\bar{c}_H)^2 - 2k_7\bar{c}_{ACA}(\bar{c}_H)^2 - 2k_{15}\bar{c}_{Py}(\bar{c}_H)^2 - 2k_{18}\bar{c}_{Ad}(\bar{c}_H)^2. \end{aligned} \tag{3.9.11}$$

These equations are not solvable analytically, but we are able to solve them numerically. We keep k_3 fixed whilst varying the hydrogen input Q_H to obtain some idea of the boundaries of this region as Q_{CO}/Q_H changes. Figure 3.19 shows the numerical solutions between $Q_H = 0.32$ and $Q_H = 0.355$. The upper boundary was chosen as it is just below the known bifurcation point at $Q_H/Q_{CO} = \frac{8k_3}{k_2+k_3}$ for $k_3 = 0.8$ and $k_2 = 1$. The lower boundary is where the solver stopped converging. We see from the graphs that as we approach the lower boundary, both $\bar{c}_H$ and $\bar{c}_F$ tend to zero as we would hope, and the remaining coefficients other than $\bar{c}_{CO}$ tend to infinity. This suggests that the state past this bifurcation is indeed be the steady-state, and there is no evidence to suggest further possible behaviour between the two.

3.10 Conclusions

We have found and identified seven behaviours of the system that arise from different parameter regimes, characterised by different long timescale scalings for the concentrations. These scalings are summarised in Table 3.5. We have found four possible parameters which heavily influence which long timescale behaviour the model displays. These parameters are Q_{CO}, Q_H, k_2, and k_3. In particular, the values of the ratios Q_{CO}/Q_H and k_3/k_2 have been found to have critical values where the model switches from one long timescale behaviour to another. For each behaviour we have described a set of ten equations that are true in the long timescale e.g. equations (3.9.2)–(3.9.11) for behaviour seven. By attempting to solve these sets of equations for a variety of values of Q_{CO}/Q_H and k_3/k_2, we find approximations of the critical values for these parameter ratios for each behaviour. Whilst the equations have solutions with coefficients greater than or equal to zero, then the behaviour exists, otherwise it does not. When solutions cease to exist, we have found an approximate boundary. We have performed this numerical analysis for all behaviours to build up a complete 2D representation of where in parameter space each behaviour exists. A diagram showing these regions is given in Figure 3.20. We were also able to obtain a number of boundaries analytically

from these equations, those of behaviours 2, 5, and 6, therefore numerical estimates were not required for these behaviours.

3.10.1 Behaviour 1

Behaviour 1 is a steady-state behaviour. The concentrations of all metabolites tend to nonzero steady-states in the long timescale. As a result, we find production of all six key products from the system. All reactions remain active. The network diagram for this region in the long timescale is the same as the full network in Figure 3.1.

3.10.2 Behaviour 2

This behaviour exists with moderate hydrogen input, and is in the area where $k_2 > k_3$. This represents a bias in production towards CO_c over formate, which causes the concentration of CO_c to tend to infinity whilst the concentration of formate decays to zero. We see in the long timescale the production of pyruvate tending to zero due the lack of formate, which causes no production of the lower fours products in the system. The upper branch remains active however, and both acetate and acetaldehyde are produced, leading to the production of ethanol with this behaviour. Figure 3.21 shows the effective long timescale metabolism for this behaviour.

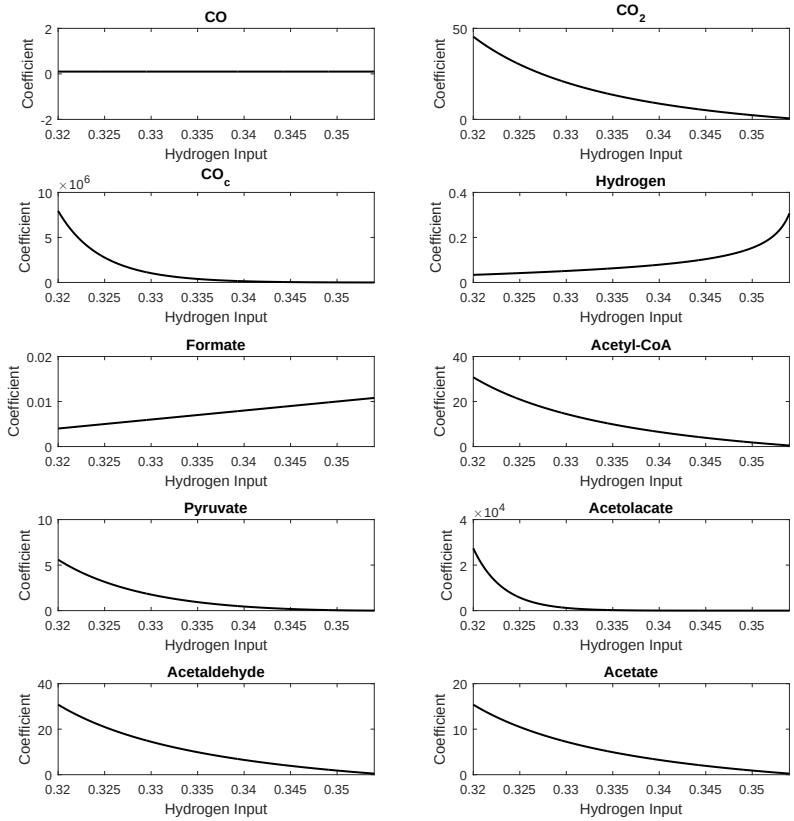

Figure 3.19: Numerical solutions to equations (3.9.2)–(3.9.11) over the range $0.32 \leq Q_H \leq 0.355$ with $k_3 = 0.8$, $Q_{CO} = 0.1$, and all other $k_n = 1$.

Metabolite	Time-scaling by behaviour						
	1	2	3	4	5	6	7
CO	t^0	t^0	t^0	t^0	t^0	t^0	t^0
CO_2	t^0	t^0	$t^{0.5}$	t^1	t^{-2}	t^{-2}	t^{-1}
CO_c	t^0	t^1	t^1	t^1	t^{-5}	t^1	t^{-3}
Formate	t^0	t^{-1}	t^0	t^1	t^1	t^{-5}	t^1
Acetyl-CoA	t^0	t^0	t^0	t^{-1}	t^{-2}	t^{-2}	t^{-1}
Pyruvate	t^0	t^{-1}	t^0	t^0	t^{-3}	t^{-9}	t^{-1}
Acetolactate	t^0	t^{-2}	t^0	t^0	t^{-8}	t^{-20}	t^{-3}
Acetaldehyde	t^0	t^0	$t^{0.5}$	t^0	t^{-2}	t^{-2}	t^{-1}
Acetate	t^0	t^0	t^0	t^{-1}	t^{-2}	t^{-2}	t^{-1}
Hydrogen	t^0	t^0	$t^{-0.25}$	$t^{-0.5}$	t^1	t^1	$t^{0.5}$

Table 3.5: Table of scalings for each metabolite as $t \to +\infty$ in each regime. A scaling of t^0 represents a metabolite tending to a nonzero steady-state. An exponent greater than zero represents a metabolite showing unbounded growth. A negative exponent represents a concentration decaying to zero.

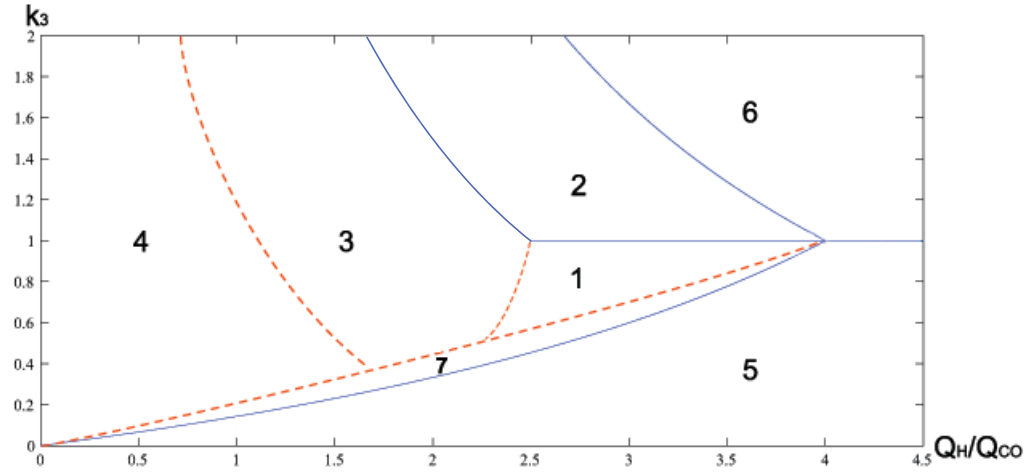

Figure 3.20: The regions in parameter space for the system of ODEs (3.3.3)–(3.3.12) with regions numbered by observed behaviour. The boundaries with solid lines are known analytically, whilst the dotted ones are obtained numerically. The horizontal axis is Q_H/Q_{CO}, with Q_H being varied, and the vertical axis is k_3. All other parameters are fixed to be one.

3.10.3 Behaviour 3

This behaviour has a lower hydrogen input than behaviours 1 or 2. With this behaviour we find the concentrations of CO_2, CO_c and acetaldehyde tending to infinity in the long timescale whilst the concentration of hydrogen decays to zero. It is obvious the decay in concentration of hydrogen is due to the reduced input, however, due to the increase of multiple other metabolite concentrations, even in the long timescale the majority of reactions remain active. The production of acetaldehyde from acetyl-CoA is not present but acetaldehyde is still formed from acetate, so there is still an ethanol production. The only product not formed in this region is butanediol. The network diagram for

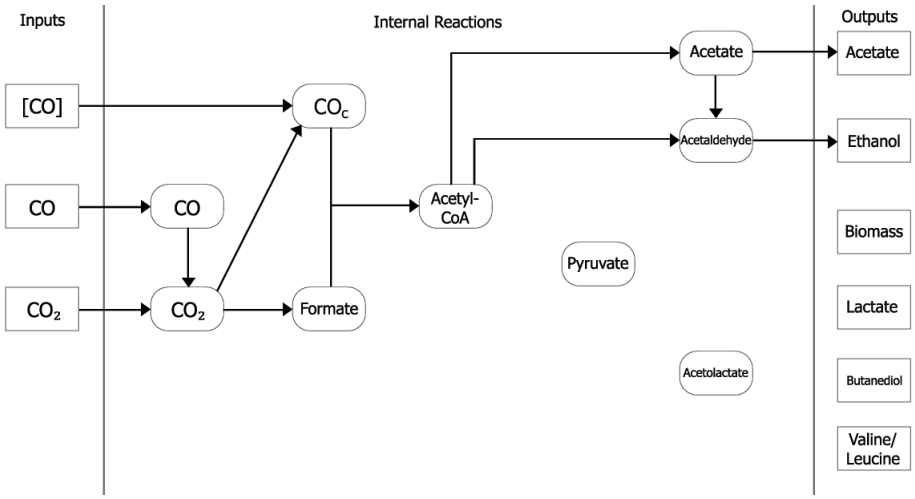

Figure 3.21: Network diagram for the effective long timescale metabolism of behaviour 2. Only reactions which are present in the long timescale are included on the diagram.

this behaviour is seen in Figure 3.22

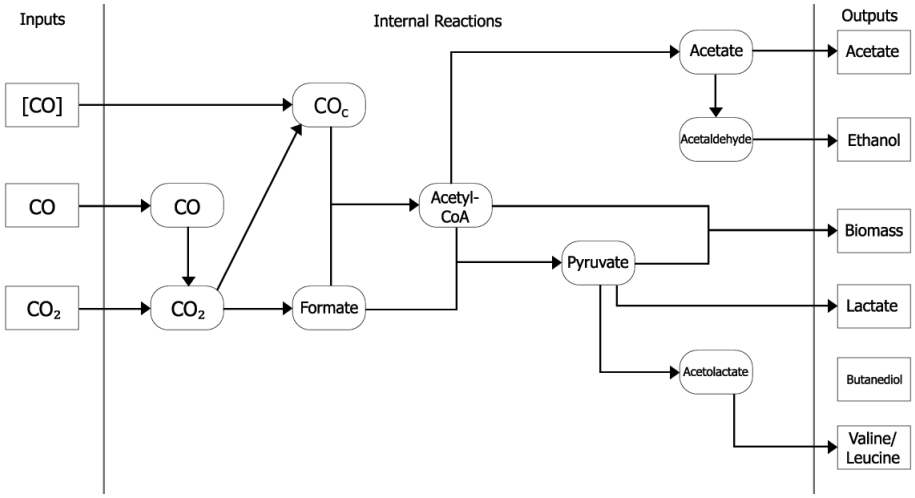

Figure 3.22: As in Figure 3.21, but for behaviour 3.

3.10.4 Behaviour 4

Behaviour 4 is present when there is a very low hydrogen input. For this behaviour, CO_2, CO_c, and formate have concentrations that tend to infinity. The concentrations of hydrogen, acetyl-CoA, and acetate decay to zero. The top branch is shut off completely in this region, and there is only production of valine in the lower branch. This is because valine does not require any additional hydrogen to produce, whilst both lactate and butanediol do. The decay in concentration of acetyl-CoA prevents the production of biomass in the long timescale. The network diagram for this behaviour is given in Figure 3.23

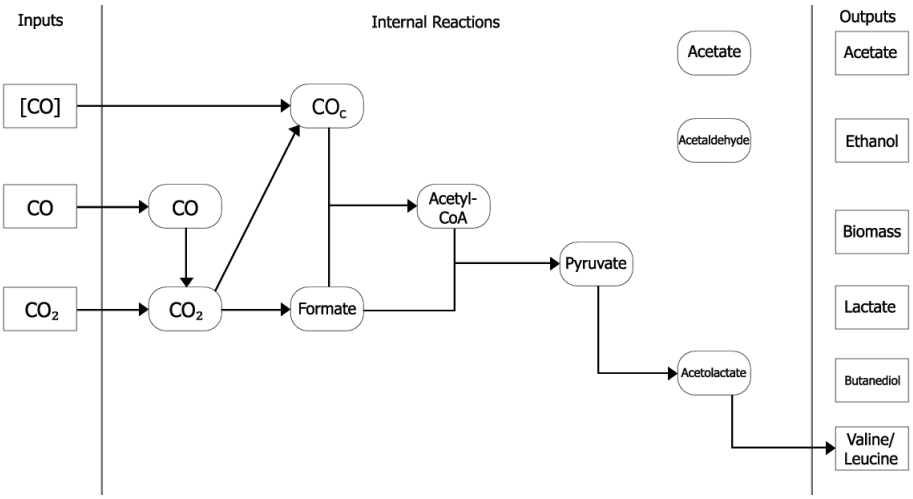

Figure 3.23: As in figure 3.21, but for behaviour 4.

3.10.5 Behaviour 5

Behaviour 5 is present in a high hydrogen input region with $k_3 < k_2$. Here, the majority of concentrations of metabolites decay to zero in the long timescale. Only the concentration of CO remains at a nonzero steady-state, whilst the concentrations of hydrogen and formate tend to infinity. The system in this region favours formate production over CO_c production which explains the boundless growth of formate. The growth of hydrogen is also obviously due to the high constant hydrogen input in the region. Due to the majority of concentrations tending to zero, the majority of reactions also become inactive in the long timescale. Only the production of ethanol remains in this region, with the flux taking a direct route from acetyl-CoA through acetaldehyde into ethanol, which is the route using the most hydrogen. There is no activity in the lower branch, and no production of acetate. The network diagram for this behaviour is shown in Figure 3.24

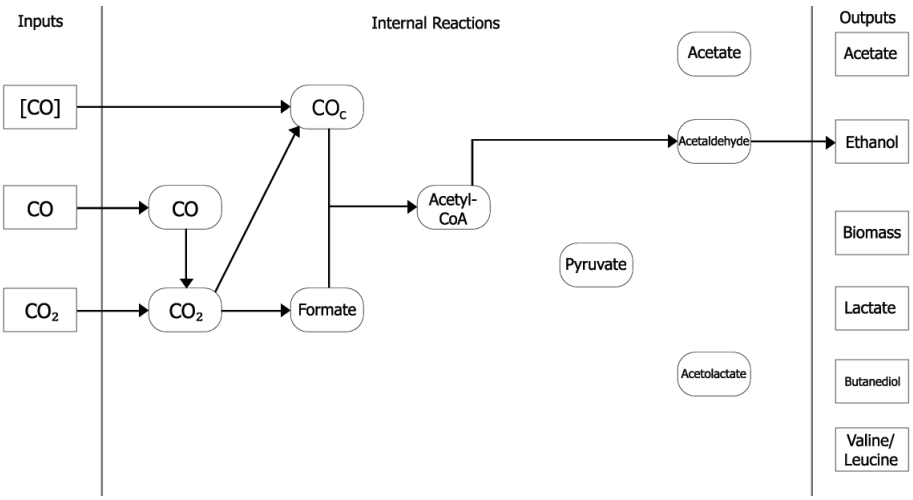

Figure 3.24: As in Figure 3.21, but for behaviours 5 and 6.

3.10.6 Behaviour 6

Behaviour 6 is found in the high hydrogen input region with $k_2 < k_3$. Here, the network has similar behaviour to that seen in region 5, with the majority of metabolite concentrations decaying to zero in the long timescale. We have a switch in the long term behaviours of CO_c and formate. In this region CO_c has unbounded growth whilst the concentration of formate decays to zero. This is due to the system now having a bias towards the production of CO_c rather than formate. This was also observed in behaviour 2. Despite the switch in behaviours of the two metabolites, we find that, as in behaviour 5, only ethanol production remains active in the long timescale. The network diagram for this behaviour is the same as the diagram for behaviour 5, seen in Figure 3.24

3.10.7 Behaviour 7

Behaviour 7 is found in a region in the parameter space between behaviours 5 and 1. The concentrations of metabolites behave in the same way as in behaviour 5, with the majority of concentrations decaying to zero, while CO remains constant and the concentrations of hydrogen and formate tend to infinity. The decays and growths however are slower than in region 5, which allows the system to maintain a production of pyruvate in the lower branch. We still have no biomass production, however lactate is produced in this region. The top branch remains with only ethanol production active. The network diagram for this behaviour is shown in 3.25

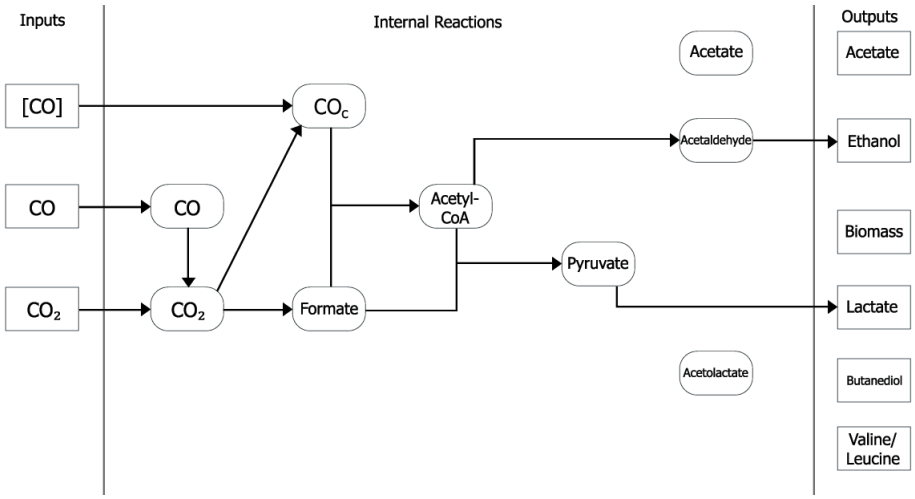

Figure 3.25: As in Figure 3.21, but for behaviour 7.

3.10.8 Conclusions

Our asymptotic analysis has allowed us to find and quantify seven different long term behaviours for this network. Of these behaviours, 6 are non-steady. Only behaviour 1 is a steady-state. We note that the range of parameter schemes that lead to this behaviour were relatively small. We also note that only behaviours 1 and 3 have a biomass production in the long timescale. In all other states the organism would stagnate and die due to a lack of biomass production required for growth. We have also found that there is butanediol production in the steady-state behaviour but no production in any other regions. It is also observed that in behaviours with extreme

hydrogen inputs, whether high or low, the ODEs predict only a single product being formed. In the high hydrogen case this is ethanol, and the network reduces to the first steps in which acetyl-CoA is formed, and then a single route to ethanol through acetaldehyde. In the low hydrogen case this single product is valine, and the network reduces to the first steps in which acetyl-CoA is formed, and then a single route to valine through pyruvate.

Within these non-steady behaviours, there are some reactions that become insignificant in the long timescale, which we are able to identify using power laws. By neglecting the terms that correspond to these insignificant reactions we are able to form sets of ODEs that approximate these behaviours in the long timescale. In particular, in Figure 3.7, we showed that we are able to compute exact values for the steady-state concentrations using these simplified ODEs. We use these simplified ODEs to find relationships in parameters that lead to each specific long timescale behaviour. We find that the input ratio Q_H/Q_{CO} and the ratio of rate parameters k_3/k_2, relating to the utilization of CO_2 within the network, were key to predicting which behaviour would be exhibited by the ODEs for a given parameter set. Whilst the boundaries between behaviours do not always have explicit analytic representations, we are always able to numerically approximate them.

By examining the behaviours that surround the steady-state region, we are able to find approximate bounds on parameter values that lead to the network tending to steady-state. This steady-state region is comparable to the steady-state region predicted by flux balance analysis. Whilst we are not able to directly compare the non-steady-state behaviours from ODEs with FBA, we are able to compare the boundaries of the steady-state region, to see if FBA predicts non-steady long timescale behaviours for the same parameter schemes.

Chapter 4

Comparing steady-state predictions of FBA and ODE models.

4.1 Introduction

We have now analysed our *C. autoethanogenum* network using both FBA and ODE models. We now compare the results from these two different models. FBA is only able to describe systems in steady-state. It does not allow us to examine the dynamics of the network prior to reaching steady-state, nor does it allow us to observe the behaviour of the network when it does not tend to a steady-state. Hence, we are only able to compare steady-state results. In particular, in section 2.9.6, we used FBA to predict bounds on the ratio of inputs to the network, Q_H/Q_{CO} that were required for the network to reach steady-state. We were able to describe similar bounds on our ODE model in Chapter 3. In addition, for our ODE model, we were able to describe conditions on an additional parameter ratio, k_3/k_2 that also determined whether the network was able to reach steady-state, allowing us to describe a 2D region in parameter space where steady-states were possible. In this chapter, we extend our FBA analysis to build up an equivalent region in flux space where steady-states are possible. In the second half of the chapter, we modify the network to examine how this affects the steady-state region predicted by FBA. By comparing the results of this steady-state region with the results for non-steady-state regions from our ODE models, we are able to find parameter schemes in which the ODE model exhibits multiple long timescale behaviours.

4.2 Steady-states in FBA

Using FBA in section 2.9.6, we determined a range of input ratios, Q_H/Q_{CO}, that allow the system to reach a steady-state. This range is

$$\frac{4}{3} \leq \frac{Q_H}{Q_{CO}} \leq 4. \tag{4.2.1}$$

For input ratios that lay outside this range, the constraints for our optimisation problem become infeasible, and steady-state solutions cannot be found. Outside this range we expect there to no longer be possible steady-states. In particular, using elementary modes analysis, in section 2.9.7, we showed that it was not possible for steady-states to exist outside this range, as there were no elementary modes that utilised input ratios

outside of this range. We overlay this range on the bifurcation diagram produced for the ODE models in Figure 3.20. This overlayed range is shown in Figure 4.1

We produced an equivalent range for the input ratios in our ODE model, (3.3.3)–(3.3.12). By numerically solving these equations in steady-state for a range of input ratios, Q_H/Q_{CO}, we find a numerical estimate for this range. We see from Figure 4.1 that this range is about

$$2.2 \leq \frac{Q_H}{Q_{CO}} \leq 4. \tag{4.2.2}$$

In section 3.4, we analysed the bifurcation between the steady-state region, region 1, and a linear growth state region, region 2, where $k_2 = k_3$. We found the behaviour on this bifurcation required the input ratio to be within the range

$$2 + \left(\frac{k_9}{k_9 + k_{19}}\right) \leq \frac{Q_H}{Q_{CO}} \leq 4. \tag{4.2.3}$$

The lower bound for this range takes values between 2 and 2.5 depending on the values of k_9 and k_{19}, generally agreeing with our numerical observation.

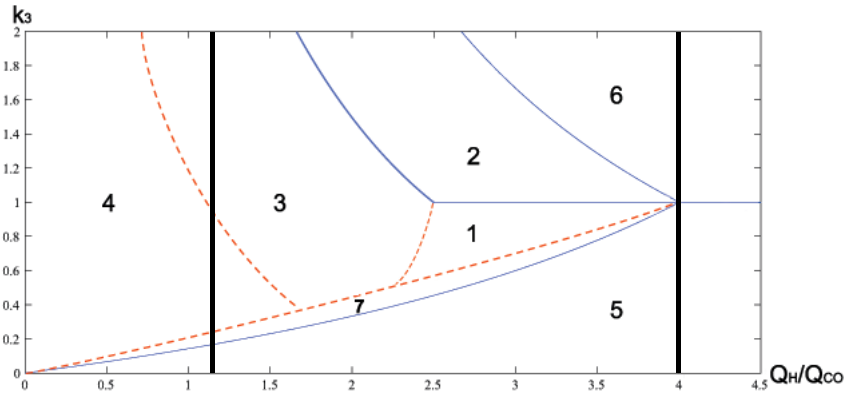

Figure 4.1: The bifurcation diagram seen in Figure 3.20, with the range of Q_H/Q_{CO} values for which FBA predicts steady-states existing marked. The region between the thick vertical lines is the region where FBA predicts steady-states, whilst outside this region FBA predicts no steady-states can be found.

The range of possible values predicted by FBA is significantly larger than those we have found by analysing the ODEs. Whilst the two methods agree exactly on the upper bound, the lower bound predicted by FBA is 4/3, whilst the lower bound predicted by the ODE model is approximately 2.2. It is possible that our particular ODEs are not able to attain a steady-state in the lower input region due to the specific choices for modelling reaction rates, but it is also possible that specific parameter choices could lead to steady-states being found in this region.

Whilst our ODE steady-state region shown in Figure 4.1 is 2D, we have so far only produced a 1D range of flux ratios that must be satisfied for steady-states to exist. We now consider an additional flux ratio in order to create an equivalent 2D region in flux space that is required for the existence of steady-states in FBA.

4.2.1 The ratio of v_2 and v_3

From our analysis in Chapter 3, we determined that the values of the parameters k_2 and k_3 have an effect on which long term behaviour the ODE model exhibits. In particular,

we found for steady-states to be possible, we must have $k_2 > k_3$, that is $k_3/k_2 > 1$. The fluxes governed by these parameters are $v_2 = k_2 c_{CO_2} c_H^2$ and $v_3 = k_3 c_{CO_2} c_H^2$. We see that the ratio of these two fluxes is

$$\frac{v_3}{v_2} = \frac{k_3 c_{CO_2} c_H^2}{k_2 c_{CO_2} c_H^2} = \frac{k_3}{k_2}. \tag{4.2.4}$$

This tells us that constraining the ratio of fluxes, $\frac{v_3}{v_2}$, in our FBA model is exactly equivalent to constraining the ratio of the rate parameters, $\frac{k_3}{k_2}$, in our ODE model. As we can directly compare these two ratios, we can test our results from the ODE model about k_2 and k_3 with flux balance analysis to see if the same behaviour is observed with FBA.

We wish to find the minimum and maximum values for $\frac{v_3}{v_2}$ across different input ratios $\frac{Q_H}{Q_{CO}}$. To do this we set up a new FBA problem. We include the original constraints of $S\mathbf{v} = \mathbf{0}$, where S is the stoichiometric matrix presented in equation (2.9.28), and $\mathbf{v}_{min} \leq \mathbf{v} \leq \mathbf{v}_{max}$, but we also include a new constraint fixing $\frac{Q_H}{Q_{CO}} = r$ for some r within the range that permits the existence of steady-states. Rearranging this constraint gives $Q_H - rQ_{CO} = 0$. In order to find the maximum and minimum values for the ratio v_2/v_3, we add the constraint $v_2 = 1$, and set our objective function to be maximising or minimising v_3. We set minimum values for all fluxes to be zero, but we do not need to set maximum values, as the overall flux through the system is constrained by the constraint $v_2 = 1$. We continue to assume that the carbon dioxide and secondary carbon monoxide inputs Q_{CO_2} and Q_{CO_c} are equal to zero. Then our FBA problem is

$$\begin{aligned} \max_{\mathbf{v}} / \min_{\mathbf{v}} \; v_3 \quad \text{such that: } & S\mathbf{v} = \mathbf{0}, \\ & v_n,\, out_n \geq 0, \\ & Q_H - rQ_{CO} = 0, \\ & Q_{CO_2} = 0, \\ & Q_{CO_c} = 0, \\ & v_2 = 1. \end{aligned} \tag{4.2.5}$$

for $4/3 \leq r \leq 4$. Since we have $v_2 = 1$, the maximum and minimum values of v_3 provided by this linear programming are the same as the maximum and minimum values of $\frac{v_3}{v_2}$. The results are shown in Table 4.1 and plotted in Figure 4.2.

We see that the maximum value for v_3/v_2 is one. This matches our findings from the ODEs, where a steady-state could only exist if $k_3 < k_2$. We also see a clear lower bound at $v_3/v_2 = 0.5$. For higher ratios of Q_H/Q_{CO}, the lower bound for v_3/v_2 is higher. Similarly, as we approach the lower bound of Q_H/Q_{CO}, the lower and upper bound for v_3/v_2 converge to $\frac{2}{3}$.

By examining the network, and the stoichiometric matrix (2.9.28) we are able to determine why the upper bound of one and the lower bound of 0.5 exist. In steady-state, we have

$$v_3 = v_4, \tag{4.2.6}$$

$$v_2 = v_4 + v_5, \tag{4.2.7}$$

$$v_4 = v_5 + v_7 + v_8 + out_{bio}. \tag{4.2.8}$$

If we have $v_3/v_2 > 1$, so that $v_3 > v_2$, then we have $v_4 > v_2$, and equation (4.2.7) cannot be satisfied, as all fluxes are positive. Since all fluxes are positive, equations (4.2.8) tells us that $v_4 \geq v_5$, and substituting this into equation (4.2.7) tells us that

Q_H/Q_{CO}	Minimum v_3/v_2	Maximum v_3/v_2
1.3333	0.6667	0.6667
1.5	0.5	0.75
1.6667	0.5	0.8333
1.8333	0.5	0.9167
2	0.5	1
2.1667	0.5	1
2.3333	0.5	1
2.5	0.5385	1
2.6667	0.5789	1
2.833	0.6216	1
3	0.6667	1
3.1667	0.7143	1
3.3333	0.7647	1
3.5	8182	1
3.6667	0.8750	1
3.8333	0.9355	1
4	1	1

Table 4.1: Minimum and maximum possible values for v_3/v_2 in the flux balance analysis problem described in (4.2.5) for different values of $r = Q_H/Q_{CO}$.

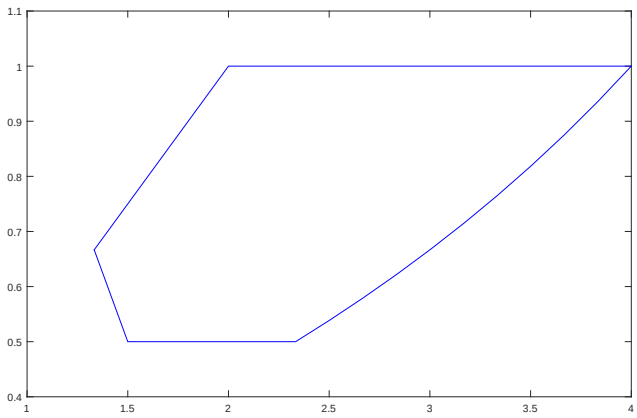

Figure 4.2: Plot of the values from Table 4.1. The x axis is Q_H/Q_{CO}, and the y axis is v_3/v_2. The upper line is the maximum possible values for v_3/v_2 and the lower line is the minimum. The region between these two lines is the area where steady-states are possible.

$v_2 \geq 2v_4$. Finally substituting in equation (4.2.6) gives us $v_2/2 \geq v_3$, giving us exactly the lower bound, $v_3/v_2 \geq 1/2$.

We see from Figure 4.2, if the ratio of Q_H/Q_{CO} is close to the maximum or minimum values allowing steady-states, the possible range for v_3/v_2 is lower than these maximum and minimum values of one and 0.5. As Q_H/Q_{CO} tends to four, the lower bound for v_3/v_2 tends to one, whilst the upper bound stays the same. As Q_H/Q_{CO} decreases

Pathway	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	in_1	in_2	in_3	in_4	Output
1	2	1	1	1	0	0	1	0	0	2	0	0	8	out_{eth}
2	2	1	1	1	0	0	0	1	1	2	0	0	6	out_{eth}
3	6	3	2	2	1	0	0	0	0	6	0	0	8	out_{bio}
4	4	2	1	1	1	0	0	0	0	4	0	0	6	out_{lac}
5	7	4	2	2	2	1	0	0	0	7	0	0	10	out_{val}
6	6	4	2	2	2	1	0	0	0	6	0	0	14	out_{but}
7	2	1	1	1	0	0	0	1	0	2	0	0	4	out_{ace}

Table 4.2: Table of elementary modes for FBA when there are only CO and hydrogen inputs. Each mode produces only a single output which is listed in the last column. All other outputs are equal to zero.

past two, the upper and lower bounds converge to $\frac{2}{3}$ as Q_H/Q_{CO} reaches $\frac{4}{3}$. In section 2.9.7, we generated a set of elementary modes for this network. Initially we found 28 possible modes, however only seven of these use only the Q_{CO} and Q_H inputs, with the other 21 also using Q_{CO_2} and Q_{CO_c}. Hence, only the first seven could be active in these simulations, and these are reproduced in Table 4.2. When the input ratio, Q_H/Q_{CO} is in the middle of its range then we have some freedom in generating a solution from combinations of these modes. As the input ratio tends to more extreme values the choices become more restricted, until finally at the most extreme points the solution has to be some multiple of a single mode. When the ratio is equal to four, only the first mode in Table 4.2 can be active, as all other modes have a lower input ratio. So for the highest possible input ratio, the network can only produce ethanol if it is in a steady-state. When the input ratio is equal to $\frac{4}{3}$, then only the third mode can be active, as all other modes have a higher input ratio. We also see looking at the table that modes with $v_2/v_3 = 1$ have an input ratio between two (mode seven) and four (mode one). If our input ratio is less than two, then we must include modes in our solution that have $v_2/v_3 < 1$ in order to have a solution. Similarly, the modes that have $v_2/v_3 = 0.5$ have input ratios between 1.5 (mode three) and $\frac{7}{3}$ (mode six), so outside of the range $1.5 \leq Q_H/Q_{CO} \leq 7/3$, we must include other modes to have a valid steady-state solution leading to a higher ratio, v_2/v_3. This matches the results seen in Table 4.1 and Figure 4.2.

4.2.2 k_9 and k_{19}

In our ODE model, k_9 is the rate parameter for the production of acetaldehyde from acetate and k_{19} is the rate parameter for the removal of acetate as a product from the system. The two fluxes, $v_9 = k_9 c_A$, and $out_{ace} = k_{19} c_A$, have the same property as v_2 and v_3, in that analysing their ratios is the same as analysing the ratio of the rate coefficients in the ODE system. However, we cannot examine k_9 and k_{19} in the same way. For v_2 and v_3, we were able to find a maximum and minimum ratio, but for v_9 and out_{ace}, there is no equivalent maximum. In particular, the ratio v_9/out_{ace} can take any value greater than or equal to zero. Rather than examining this ratio on its own, we instead constrain it to take specific values, and then produce the same 2D regions of flux space that permit the existence of steady-states, as generated by equation 4.2.5, seen in Figure 4.2. We constrain the v_9/out_{ace} ratio in the same way we constrain the input ratio, so that we have $v_9 - r_2 out_{ace} = 0$ for various values of r_2. Then the new

FBA problem is

$$\begin{aligned} \max_{\mathbf{v}} / \min_{\mathbf{v}} \; v_3 \quad \text{such that: } & S\mathbf{v} = \mathbf{0}, \\ & v_n,\, out_n \geq 0, \\ & Q_H - r_1 Q_{CO} = 0, \\ & Q_{CO_2} = 0, \\ & Q_{CO_c} = 0, \\ & v_2 = 1, \\ & v_9 - r_2 out_{ace} = 0. \end{aligned} \tag{4.2.9}$$

We solve this FBA problem for the range $4/3 \leq r_1 \leq 4$, and various values of r_2. Figure 4.3 shows these steady-state regions for r_2 equal to 0, 0.5, 1, and 1000. This ratio, v_9/out_{ace} has no effect on the lower bound for v_3/v_2. We see that it does affect the upper bound, particularly for low input ratios where the maximum ratio of v_3/v_2 becomes lower. Specifically, the minimum input ratio Q_H/Q_{CO} required for $v_3/v_2 = 1$ to allow steady-states increases. In our ODE models, we have always used $k_9 = k_{19}$, so

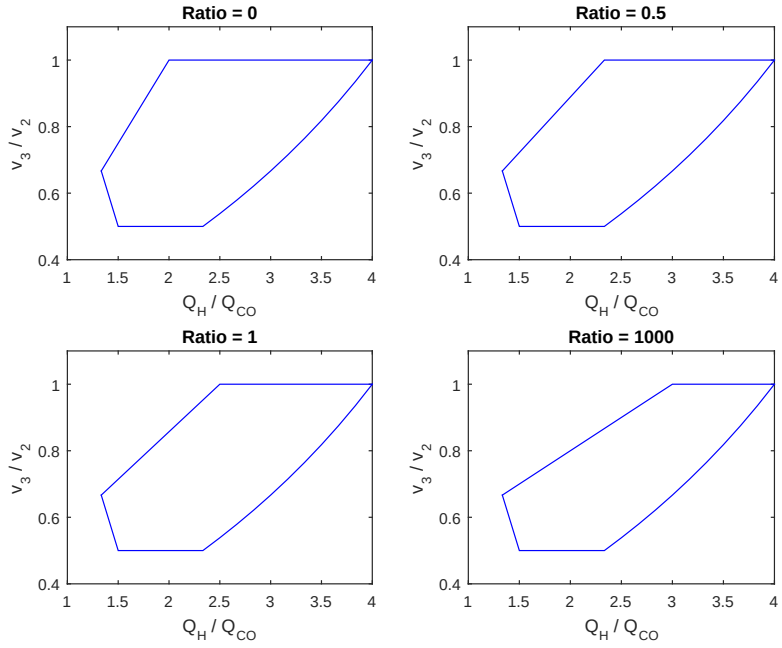

Figure 4.3: Plots of the boundaries of the region in which FBA solutions exist with v_9/out_{ace} set to 0, 0.5, 1, and 1000.

the ratio of the two is equal to one. In FBA, this is equivalent to adding the constraint $v_9 - out_{ace} = 0$. This is shown in the third graph in Figure 4.3. For $v_9 = out_{ace}$, we find that the input ratio required for $v_3/v_2 = 1$ to allow steady-states is $Q_H/Q_{CO} = 2.5$. This matches the results found from our ODE model. Specifically equation (4.2.3) told us that the minimum value for Q_H/Q_{CO} required was $Q_H/Q_{CO} = 2 + \frac{k_9}{k_9+k_{19}} = 2.5$ when $k_9 = k_{19}$. In fact, comparing our results from the FBA model, (4.2.9), we find that range of values Q_H/Q_{CO} that allow a steady-state when $v_3/v_2 = 1$ for a given v_9/out_{ace} ratio match exactly the range predicted by equation (4.2.3).

4.3 Comparing the steady-state regions of the two models

We compare the steady-state region predicted by FBA from (4.2.5) with the steady-state region predicted by our ODE model, (3.3.3)–(3.3.12). Figure 4.4 shows both these regions. We note that the region of steady-states predicted by FBA is much larger than the region predicted by ODEs. The upper boundary on k_3/k_2 matches exactly, with both methods predicting no steady-states when $k_3 > k_2$. The lowest possible values of k_3/k_2 predicted by the ODE model is quite close to the lowest possible value predicted by FBA, however for most values of Q_H/Q_{CO} these boundaries are not in agreement. In particular, we note that whilst larger values of Q_H/Q_{CO} give a closer agreement between the two methods, for lower values of Q_H/Q_{CO} there is much more variation between the two methods. Indeed, we see that our ODEs do not predict states at all for values of Q_H/Q_{CO} less than near 2.3, whilst FBA predicts them as low as $Q_H/Q_{CO} = 4/3$. The upper bound of Q_H/Q_{CO} for both methods, however, is in agreement, with the maximum value allowed for the existence of steady-states being $Q_H/Q_{CO} = 4$. We see that the steady-state region predicted by FBA extends into regions where our ODE model predicts non-steady-states. In particular, it extends into the regions where we see behaviours 3 and 7. Figure 4.5 shows the region in which our ODE model predicts the behaviour 3, whilst the FBA model predicts the possibility of steady-states, and Figure 4.6 shows the region in which the ODE model predicts behaviours 7, whilst FBA predicts steady-state. This latter region is of particular interest, as it allows us to provide an upper bound on values of k_3/k_2 that lead to behaviour 7 in the long timescale. The lower bound for region 7 is analytically calculated from the long timescale asymptotics for region 5, seen in section 3.8.1, but the upper bound is only numerically estimated for a specific parameter set. The lower bound of the FBA steady-state region provides a general upper bound for behaviour 7.

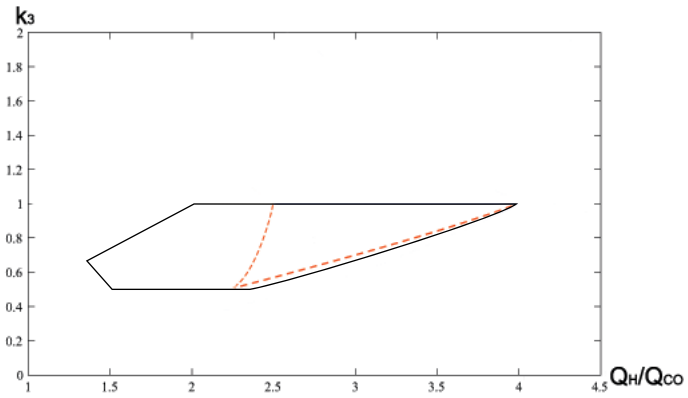

Figure 4.4: Comparison of regions predicting steady-state by the ODE model, (3.3.3)–(3.3.12), and the FBA, as found using equation (4.2.5). The region enclosed by the solid line is the steady-state region predicted by FBA, whilst the region enclosed by the dashed line shows the steady-state region calculated by ODEs.

Now that we have compared the steady-state regions predicted by our ODE model, and those predicted by our FBA model, it would be worth examining how these regions would react to a slight modification to the network. We next examine a network for the same organism, *C. autoethanogenum*, in which one reaction has been modified.

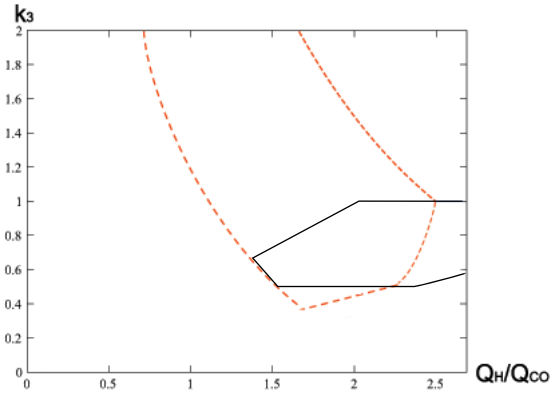

Figure 4.5: Plot of steady-state region predicted by FBA from equation (4.2.5), and the region numerically calculated from the ODE mode, (3.3.3)–(3.3.12), to exhibit behaviour 3. The region enclosed by the solid line is the FBA steady-state region, and the region enclosed by the dashed line is the region where behaviour 3 is exhibited by the ODEs.

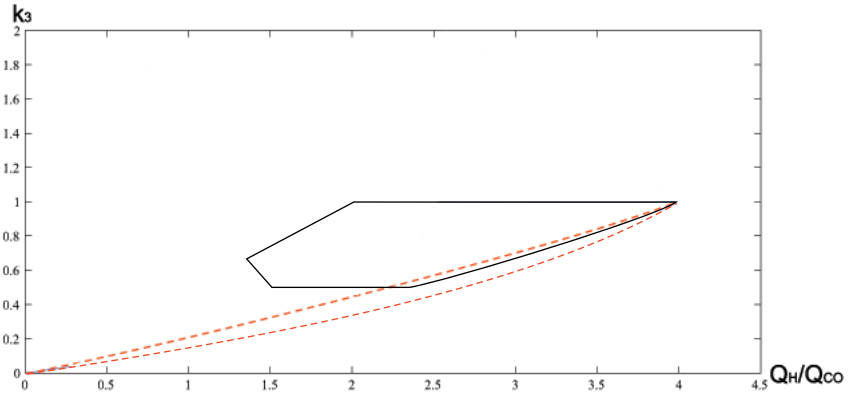

Figure 4.6: Plot of steady-state region predicted by FBA from equation (4.2.5), and the region numerically calculated from the ODE mode, (3.3.3)–(3.3.12), to exhibit behaviour 7. The region enclosed by the solid line is the FBA steady-state region, and the region enclosed by the dashed line is the region where behaviour 7 is exhibited by the ODEs.

4.4 A modified *C.autoethanogenum* network

A closer inspection of our *C. autoethanogenum* network reveals a problem. The reaction

$$\begin{aligned} \text{Carbon dioxide} + \text{Formate} + \text{Acetyl-CoA} + 2\text{H} &\xrightarrow{v_5} \text{Pyruvate} + \text{CoA} \\ \text{CO}_2 + \text{HCO}_2\text{-} + \text{C}_{23}\text{H}_{38}\text{N}_7\text{O}_{17}\text{P}_3\text{S} + 2\text{H} &\xrightarrow{v_5} \text{C}_3\text{H}_4\text{O}_3 + \text{C}_{21}\text{H}_{36}\text{N}_7\text{O}_{16}\text{P}_3\text{S}, \end{aligned}$$

which produces pyruvate does not balance. In particular, the carbon balance is incorrect, with 25 carbon on the left hand side, and only 24 on the right hand side. It turns out that this reaction is actually an approximation of two different reactions that both produce pyruvate. To make the model more biologically accurate, we split this one reaction into the two separate reactions. The first combines carbon dioxide and hydrogen with acetyl-CoA, the second combines formate with acetyl-CoA, both creating

pyruvate as the result, so that we have

$$\begin{aligned} \text{Carbon dioxide + Acetyl-CoA + 2H} &\xrightarrow{v_5} \text{Pyruvate + CoA}, \\ \text{Formate + Acetyl-CoA} &\xrightarrow{v_6} \text{Pyruvate + CoA}. \end{aligned}$$

Both of these equations balance carbon (1 carbon + 23 carbon $\longrightarrow$ 3 carbon + 21 carbon in both cases), and the new network should be more biologically correct than the previous one. A network diagram for this new modified network is given in Figure 4.7. We form a new stoichiometric matrix for this modified network. This matrix is very similar to the previous one seen in Chapter 2, equation (2.9.28), but the column representing v_5 is changed, with an additional column inserted after it to represent the new v_6 reaction combining formate and acetyl-CoA. The stoichiometric matrix concerning the internal reactions for our *C. autoethanogenum* is

$$S_{in} = \begin{array}{r} \\ CO \\ CO_2 \\ CO_c \\ Formate \\ Acetyl-CoA \\ Pyruvate \\ Acetolactate \\ Acetaldehyde \\ Acetate \\ Hydrogen \end{array} \begin{array}{c} \begin{array}{cccccccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_9 & v_{10} \end{array} \\ \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 2 & -2 & -2 & -4 & -2 & 0 & 0 & -2 & 0 & 0 \end{pmatrix} \end{array}, \tag{4.4.1}$$

and the stoichiometric matrix for the uptake and output reactions is

$$S_{ext} = \begin{array}{c} \begin{array}{cccccccc} Q_{CO} & Q_H & out_{bio} & out_{lac} & out_{val} & out_{but} & out_{eth} & out_{ace} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & -2 & 0 & -2 & -2 & 0 \end{pmatrix} \end{array}. \tag{4.4.2}$$

Then the complete stoichiometric matrix is the augmented matrix

$$S = (S_{in}|S_{ext})\,. \tag{4.4.3}$$

This new modified network allows us to examine the effects of slightly modifying a network on our previous analysis. By performing both our long timescale asymptotic analysis from Chapter 3, and our FBA from Chapter 2 again, we are able to see the differences between the long timescale behaviours of both networks. To begin, we perform the long timescale asymptotic analysis on an ODE model for this new network.

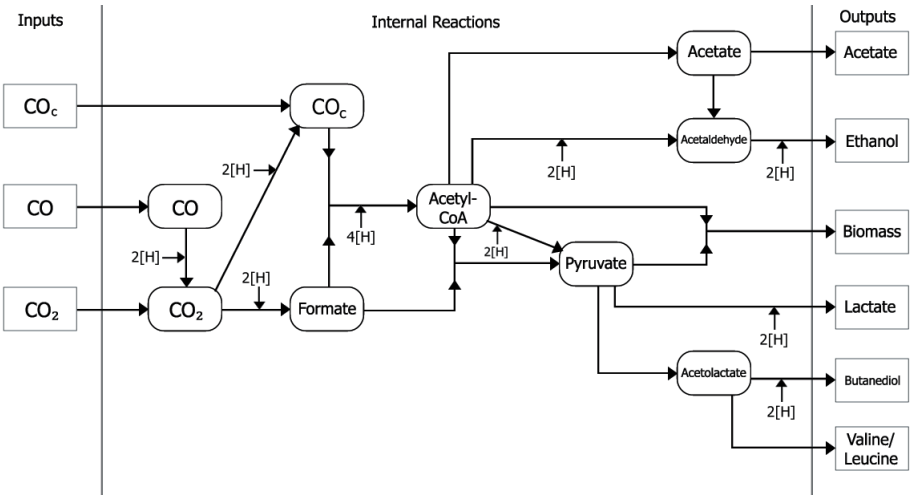

Figure 4.7: Full network diagram for the metabolism of *C. autoethanogenum*. Note the new second reaction directly from acetyl-CoA to pyruvate, representing the new reaction combining CO_2, acetyl-CoA and hydrogen.

4.5 The new model

We first form an ODE model for the modified network. We base this on the ODE model used in the Chapter 3, with constant inputs of hydrogen and carbon monoxide. This new model is again formed of ten ODEs,

$$\frac{dc_{CO}}{dt} = Q_{CO} - k_1 c_{CO}, \tag{4.5.1}$$

$$\begin{aligned}\frac{dc_{CO_2}}{dt} &= k_1 c_{CO} - k_2 c_{CO_2} {c_H}^2 - k_3 c_{CO_2} {c_H}^2 \\ &\quad - k_5 c_{CO_2} c_{ACA} {c_H}^2 + k_7 {c_{Py}}^2 + k_{18} c_{Al} {c_H}^2,\end{aligned} \tag{4.5.2}$$

$$\frac{dc_{CO_c}}{dt} = k_3 c_{CO_2} {c_H}^2 - k_4 c_{CO_c} c_F {c_H}^4, \tag{4.5.3}$$

$$\frac{dc_F}{dt} = k_2 c_{CO_2} {c_H}^2 - k_4 c_{CO_c} c_F {c_H}^4 - k_6 c_F c_{ACA}, \tag{4.5.4}$$

$$\begin{aligned}\frac{dc_{ACA}}{dt} &= k_4 c_{CO_c} c_F {c_H}^4 - k_5 c_{CO_2} c_{ACA} {c_H}^2 - k_6 c_F c_{ACA} - k_8 c_{ACA} {c_H}^2 \\ &\quad - k_9 c_{ACA} - k_{15} c_{ACA} c_{Py},\end{aligned} \tag{4.5.5}$$

$$\begin{aligned}\frac{dc_{Py}}{dt} &= k_5 c_{CO_2} c_{ACA} {c_H}^2 + k_6 c_F c_{ACA} - 2k_7 {c_{Py}}^2 - k_{15} c_{ACA} c_{Py} \\ &\quad - k_{16} c_{Py} {c_H}^2,\end{aligned} \tag{4.5.6}$$

$$\frac{dc_{Al}}{dt} = k_7 {c_{Py}}^2 - k_{17} c_{Al} - k_{18} c_{Al} {c_H}^2, \tag{4.5.7}$$

$$\frac{dc_{Ad}}{dt} = k_8 c_{ACA} {c_H}^2 + k_{10} c_A - k_{19} c_{Ad} {c_H}^2, \tag{4.5.8}$$

$$\frac{dc_A}{dt} = k_9 c_{ACA} - k_{10} c_A - k_{20} c_A, \tag{4.5.9}$$

$$\begin{aligned}\frac{dc_H}{dt} &= Q_H + 2k_1 c_{CO} - 2k_2 c_{CO_2} {c_H}^2 - 2k_3 c_{CO_2} {c_H}^2 - 4k_4 c_{CO_c} c_F {c_H}^4 \\ &\quad - 2k_5 c_{CO_2} c_{ACA} {c_H}^2 - 2k_8 c_{ACA} {c_H}^2 - 2k_{16} c_{Py} {c_H}^2 - 2k_{18} c_{Al} {c_H}^2 \\ &\quad - 2k_{19} c_{Ad} {c_H}^2.\end{aligned} \tag{4.5.10}$$

We renumber the rate coefficients k_n, so that the rate coefficients for the two new reactions are k_5 and k_6, and the rate coefficients k_7 to k_{20} are the equivalent of k_6 to k_{19} in the old model. The parameters for this new model are described in Table 4.3.

Parameter	Description	Equivalent from previous model
k_1	Rate coefficient for CO being transformed into CO_2	k_1
k_2	Rate coefficient for CO_2 being transformed into formate	k_2
k_3	Rate coefficient for CO_2 being transformed into CO_c	k_3
k_4	Rate coefficient for formate combining with CO_c to make acetyl-CoA	k_4
k_5	Rate coefficient for acetyl-CoA combining with CO_2 to make pyruvate	k_5
k_6	Rate coefficient for acetyl-CoA combining with formate to make pyruvate	–
k_7	Rate coefficient for two pyruvates combining to make acetolactate	k_6
k_8	Rate coefficient for acetyl-CoA being transformed into acetaldehyde	k_7
k_9	Rate coefficient for acetyl-CoA being transformed into acetate	k_8
k_{10}	Rate coefficient for acetate being transformed into acetaldehyde	k_9
Q_{CO}	Constant uptake rate for carbon monoxide	Q_{CO}
Q_H	Constant uptake rate for hydrogen	Q_H
k_{15}	Rate coefficient for the combining of acetyl-CoA and pyruvate to output biomass	k_{14}
k_{16}	Rate coefficient for the creation and output of lactate from pyruvate	k_{15}
k_{17}	Rate coefficient for the creation and output of valine from acetolactate	k_{16}
k_{18}	Rate coefficient for the creation and output of butanediol from acetolactate	k_{17}
k_{19}	Rate coefficient for the creation and output of ethanol from acetaldehyde	k_{18}
k_{20}	Rate coefficient for the output of acetate	k_{19}

Table 4.3: Table of parameters in the system of equations (4.5.1)–(4.5.10). Where parameters have an equivalent in the old model, this is listed.

4.5.1 Asymptotic analysis

Now that we have an ODE model, we repeat the asymptotic analysis from Chapter 3. Rather than finding new scalings from scratch, we instead check the old scalings in the new model to see if they are still consistent. Since we have only changed one reaction into two, we only need to examine balances for the ODEs that contain those reactions. In this case, the ODEs we need to check are those for CO_2, CO_c, formate, acetyl-CoA, pyruvate, and hydrogen. If these equations still balance with the old scalings, then the scalings are still consistent across the entire system, and do not need changing. If there is a problem, then we must find a new set of scalings for that region, whether by running an ODE simulation, or by inspection.

The steady-state region is trivially verified, we see that when all concentrations tend to a constant, all equations balance. For the second behaviour, in which $k_3 > k_2$, we find a problem with the equation for pyruvate. With the previous scalings we find equation (4.5.6) becomes

$$\begin{aligned} -\bar{c}_{Py}t^{-2} = k_5\bar{c}_{CO_2}\bar{c}_{ACA}\bar{c}_H^2t^0 + k_6\bar{c}_F\bar{c}_{ACA}t^{-1} - 2k_7\bar{c}_{Py}^2t^{-2} \\ - k_{15}\bar{c}_{ACA}\bar{c}_{Py}t^{-1} - k_{16}\bar{c}_{Py}\bar{c}_H^2t^{-1}. \end{aligned} \tag{4.5.11}$$

where we see there is a single term of order t^0 with nothing to balance it. Previously, the pyruvate creating reaction combining CO_2, formate, acetyl-CoA and hydrogen would disappear in the long timescale, and the amount of pyruvate in the system decayed because of this. Now that we have a pyruvate-creating reaction that is active in the long timescale, it is possible that the concentration of pyruvate tends to a steady-state rather than decaying to zero. Indeed, if we take pyruvate to be tending to a steady-state in the long timescale, so that $c_{Py} \sim \overline{c_{Py}}t^0$, the k_7, k_{15}, and k_{16} reactions now balance with the k_5 reaction. By doing this, however, we introduce an imbalance in the acetolactate reaction (4.5.7), where the k_7 reaction, which was before of the order t^{-2} is now of order t^0. This is rectified in the same way, by setting the long timescale

behaviour of acetolactate to be tending to a steady-state, i.e. $c_{Al} \sim \overline{c_{Al}} t^0$, rather than decaying to zero. Changing the scalings of pyruvate and acetolactate to be t^0 is the only way to modify the set of scalings so that they maintain balances across all reactions, and this is backed up by the simulation shown in Figure 4.8. Here we have set parameters so that we are simulating the network in the region where behaviour two exists, as seen in Figure 4.9, and the concentrations scale exactly as predicted by our modifications, with $c_{Py} \sim \overline{c_{Py}} t^0$ and $c_{Al} \sim \overline{c_{Al}} t^0$. With our new scalings, we obtain

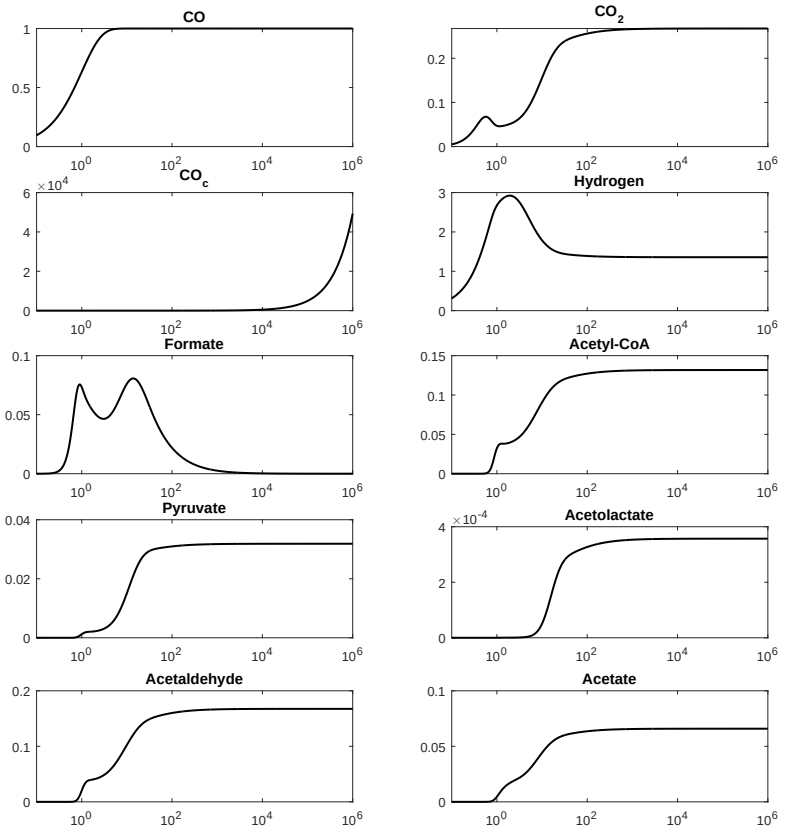

Figure 4.8: Metabolite concentration plots against log time of the ODEs (4.5.1)–(4.5.10), with all parameters set to equal one, with the exception of $k_3 = 0.9$, and $Q_H = 3$, so that the model resides in the region where we expect to find behaviour 2.

the set of simultaneous equations for our coefficients,

$$k_1 \bar{c}_{CO} = Q_{CO}, \tag{4.5.12}$$

$$k_2 \bar{c}_{CO_2} \bar{c}_H^2 + k_3 \bar{c}_{CO_2} \bar{c}_H^2 = k_1 \bar{c}_{CO} - k_5 \bar{c}_{CO_2} \bar{c}_{ACA} \bar{c}_H^2 + k_7 \bar{c}_{Py}^2 + k_{18} \bar{c}_{Al} \bar{c}_H^2, \tag{4.5.13}$$

$$\bar{c}_{CO_c} = k_3 \bar{c}_{CO_2} \bar{c}_H^2 - k_4 \bar{c}_{CO_c} \bar{c}_F \bar{c}_H^4, \tag{4.5.14}$$

$$k_4 \bar{c}_{CO_c} \bar{c}_F \bar{c}_H^2 = k_2 \bar{c}_{CO_2}, \tag{4.5.15}$$

$$k_4 \bar{c}_{CO_c} \bar{c}_F \bar{c}_H^4 = k_5 \bar{c}_{CO_2} \bar{c}_{ACA} \bar{c}_H^2 + k_8 \bar{c}_{ACA} \bar{c}_H^2 + k_9 \bar{c}_{ACA}, \tag{4.5.16}$$

$$k_5 \bar{c}_{CO_2} \bar{c}_{ACA} \bar{c}_H^2 = k_7 \bar{c}_{Py}^2 + k_{15} \bar{c}_{ACA} \bar{c}_{Py} + k_{16} \bar{c}_{Py} \bar{c}_H^2, \tag{4.5.17}$$

$$k_7\bar{c}_{Py}^2 = k_{17}\bar{c}_{Al} + k_{18}\bar{c}_{Al}\bar{c}_H^2, \tag{4.5.18}$$

$$k_{19}\bar{c}_{Ad}\bar{c}_H^2 = k_8\bar{c}_{ACA}\bar{c}_H^2 + k_{10}\bar{c}_A, \tag{4.5.19}$$

$$k_9\bar{c}_{ACA} = k_{10}\bar{c}_A + k_{20}\bar{c}_A, \tag{4.5.20}$$

$$\begin{aligned} Q_H + 2k_1\bar{c}_{CO} = {} & 2k_2\bar{c}_{CO_2}\bar{c}_H^2 + 2k_3\bar{c}_{CO_2}\bar{c}_H^2 + 4k_4\bar{c}_{CO_c}\bar{c}_F\bar{c}_H^4 \\ & + 2k_5\bar{c}_{CO_2}\bar{c}_{ACA}\bar{c}_H^2 + 2k_8\bar{c}_{ACA}\bar{c}_H^2 + 2k_{16}\bar{c}_{Py}\bar{c}_H^2 \\ & + 2k_{18}\bar{c}_{Al}\bar{c}_H^2 - 2k_{19}\bar{c}_{Ad}\bar{c}_H^2. \end{aligned} \tag{4.5.21}$$

These equations differ from those seen in section 3.5.2 for behaviour two in the old network. Whilst we were able to analytically solve those equations to obtain expressions for the coefficients in terms of the parameters, we are unable to do the same for these equations; however, numerically solving these equations for a range of parameter values suggests that the region in which this behaviour is observable remains largely the same.

Both behaviours 3 and 4 are consistent in the new network, with no modifications needing to be made. In behaviour 3, we have the concentrations of CO_2, the secondary carbon monoxide source, CO_c and acetaldehyde growing in time, with the concentration of hydrogen decaying. In behaviour 4 we have the concentrations of CO_2, CO_c, and formate growing in time, with the concentrations of acetyl-CoA, acetate and hydrogen decaying. For the old network, these behaviours both maintained production of pyruvate in the long timescale, and in the new network they are both able to maintain at least one of the pyruvate producing k_5 or k_6 reactions in the long timescale. Notably, in behaviour 3, both of these reactions are active in the long timescale, whilst with behaviour 4, only the k_6 reaction is maintained.

Behaviour 5, in which the concentrations of formate and CO_c grow in time, and all other concentrations except that of CO decay to zero, maintains consistency with this new network. Previously, the pyruvate producing reaction was not active in the long timescale, decaying to zero with order t^{-1}. In the new network, the k_6 reaction combining formate and acetyl-CoA also decays to zero with order t^{-1}, whilst the k_5 reaction decays more quickly, with order t^{-2}. Since we still have a t^{-1} reaction and the other additional reaction is of lower order, the balance is maintained with no differences.

Behaviour 6, like behaviour 2, also encounters problems with the pyruvate reaction. In behaviour 6, we have the concentrations of CO_c and hydrogen growing in time, with all other concentrations except that of CO decaying to zero. In the previous network, the sole pyruvate creating reaction used formate, which decays very quickly in this behaviour (scaling like $c_F \sim \overline{c_F}t^{-5}$), meaning the production of pyruvate decayed rapidly, and therefore so did the concentration of pyruvate. In the new network, the k_5 reaction does not use formate, causing it to decay more gradually (to the order of t^{-2}). With the previous scalings, the ODE for pyruvate becomes

$$\begin{aligned} -9\bar{c}_{Py}t^{-10} = {} & k_5\bar{c}_{CO_2}\bar{c}_{ACA}\bar{c}_H^2t^{-2} + k_6\bar{c}_F\bar{c}_{ACA}t^{-7} - 2k_7\bar{c}_{Py}^2t^{-18} \\ & - k_{15}\bar{c}_{ACA}\bar{c}_{Py}t^{-11} - k_{16}\bar{c}_{Py}\bar{c}_H^2t^{-7}. \end{aligned} \tag{4.5.22}$$

We see the k_5 reaction is the leading order and does not balance with any other terms. As in behaviour two, the correction we make is changing the scaling of pyruvate to be higher. To have a term that balances with the k_5 reaction, we change the pyruvate scaling to $c_{Py} \sim \overline{c_{Py}}t^{-4}$, so that the k_{16} reaction is of the order t^{-2} (as the concentration of hydrogen grows linearly in time), and balances with the k_5 reaction. This in turn causes problems with the ODE for acetolactate, necessitating a change in the scaling of that concentration too. The equation for acetolactate with the new pyruvate scaling is

$$-20\bar{c}_{Al}t^{-21} = k_7\bar{c}_{Py}^2t^{-8} - k_{17}\bar{c}_{Al}t^{-20} - k_{18}\bar{c}_{Al}\bar{c}_H^2t^{-18}. \tag{4.5.23}$$

We see the k_7 reaction that produces acetolactate does not balance with any other terms in the equation. By changing the scaling of acetolactate so that $c_{Al} \sim \overline{c_{Al}} t^{-10}$, the k_{18} reaction balances with the k_7 reaction, allowing this behaviour to exist. Despite the change in the scalings for pyruvate and acetolactate, there are no other changes in reactions which remain active in the long timescale, unlike behaviour 2, which had the production of pyruvate, acetolactate, and their related products in the new network that were not present in the old network.

Finally, behaviour 7, in which we have the same concentrations growing and decaying in time as behaviour 5, remains consistent with the new network, as the k_6 reaction remains active in the long timescale, allowing all reactions to maintain balance. Overall, with this slight change to the network, only two behaviours have changed their scalings, and only one of those has caused any further difference to the reactions that remain active in the long timescale. The scalings that describe the possible long timescale behaviours for this new modified network are summarised in Table 4.4.

Metabolite	Time-scaling by behaviour						
	1	2	3	4	5	6	7
CO	t^0	t^0	t^0	t^0	t^0	t^0	t^0
CO_2	t^0	t^0	$t^{0.5}$	t^1	t^{-2}	t^{-2}	t^{-1}
CO_c	t^0	t^1	t^1	t^1	t^{-5}	t^1	t^{-3}
Formate	t^0	t^{-1}	t^0	t^1	t^1	t^{-5}	t^1
Acetyl-CoA	t^0	t^0	t^0	t^{-1}	t^{-2}	t^{-2}	t^{-1}
Pyruvate	t^0	$\boldsymbol{t^0}$	t^0	t^0	t^{-3}	$\boldsymbol{t^{-4}}$	t^{-1}
Acetolactate	t^0	$\boldsymbol{t^0}$	t^0	t^0	t^{-8}	$\boldsymbol{t^{-10}}$	t^{-3}
Acetaldehyde	t^0	t^0	$t^{0.5}$	t^0	t^{-2}	t^{-2}	t^{-1}
Acetate	t^0	t^0	t^0	t^{-1}	t^{-2}	t^{-2}	t^{-1}
Hydrogen	t^0	t^0	$t^{-0.25}$	$t^{-0.5}$	t^1	t^1	$t^{0.5}$

Table 4.4: Table of scalings for each metabolite as $t \to +\infty$ in each regime. A scaling of t^0 represents a metabolite tending to a nonzero steady-state. An exponent greater than zero represents a metabolite showing unbounded growth. A negative exponent represents a concentration decaying to zero. Scalings which have changed with the new network are in bold.

It is notable that when answering the question of whether or not a set of scalings is still consistent, we only needed to consider the new pyruvate producing reactions, k_5, and k_6. The behaviours of these pyruvate producing reactions in both the old and new models are summarised in Table 4.5. For most behaviours, if the previous models k_5 reaction was active in the long timescale, then one or both of the new reactions was active. If this was the case, then the set of scalings would generally be consistent, otherwise (in the case of behaviour 2), changes needed to be made. The only exception is behaviour 6, in which two scalings (pyruvate and acetaldehyde) need to be changed, despite neither of the new reactions being present in the long timescale. However, this did not lead to a change in which reactions were active in the long timescale. This is in contrast to behaviour 2, where the k_5 reaction being present in the long timescale caused a difference to the effective metabolism in the long timescale, allowing it to produce all the networks products rather than just acetate and ethanol. The change to the network has not caused any change to the regions in parameter space that exhibit these behaviours, as shown in Figure 4.9.

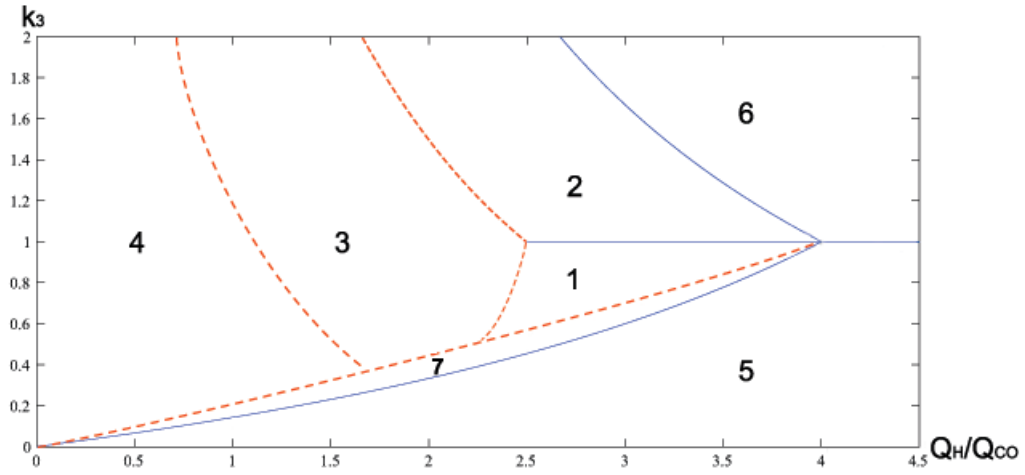

Figure 4.9: The regions in parameter space for the system of ODEs (4.5.1)–(4.5.10) with regions numbered by observed behaviour. The boundaries with solid lines are known analytically, whilst the dotted ones are obtained numerically. The horizontal axis is Q_H/Q_{CO}, with Q_H being varied, and the vertical axis is k_3. All other parameters are fixed to be one. The only difference between this diagram and that of the previous network, shown in Figure 3.20 is the boundary between behaviours 2 and 3 is now numerically estimated.

Behaviour	Reaction producing pyruvate in original model	Reaction producing pyruvate in modified model
1	k_5	k_5, k_6
2	None	k_5
3	k_5	k_5, k_6
4	k_5	k_6
5	None	None
6	None	None
7	k_5	k_6

Table 4.5: Table listing which reactions in the long timescale produce pyruvate for each behaviour. We see that the only behaviour that switches from no production of pyruvate to some production of pyruvate in the long timescale is behaviour 2. There is no behaviour which switches from some production to no production.

4.6 Flux balance analysis

In section 4.2, we constructed a two dimensional section of parameter space where steady-states exist. The two key values we considered were the input ratio, Q_H/Q_{CO}, and the ratio of two other fluxes, v_3/v_2. These two ratios have both been seen in ODE analysis to have an effect on the long term behaviour of the model. In particular, every boundary that was analytically derived uses some or all of the rate coefficients for these fluxes. The v_3/v_2 ratio is especially useful to observe because, as shown in the previous chapter, $v_3/v_2 = k_3/k_2$, so the ratio of these two fluxes is exactly equal to the ratio of the two rate coefficients. We also examined the effects of the v_9/v_{19} ratio (now the v_{10}/v_{20} ratio after renumbering for the new model), though this was less significant in determining the boundaries of steady-state regions.

We now build up an equivalent region for our new network where steady-states exist.

By using the same FBA problem described in equation (4.2.5), finding a maximum and minimum value of v_3/v_2 for a range of Q_H/Q_{CO} ratios, we find a complete 2D section where steady-states exist. Figure 4.10 shows this region, along with the region found for the old network for comparison. First we note that the new network requires a higher input ratio, Q_H/Q_{CO} than the previously examined network for the existence of steady-states. The previous network attained steady-state with an input ratio $Q_H/Q_{CO} \geq 4/3$, however this new modified network does not reach steady-state unless the input ratio is at least 1.6. The maximum input ratio remains the same at 4. Whilst there is a slight decrease in the input ratios, Q_H/Q_{CO} that allow steady-states to exist, there is generally much more freedom in the value of v_3/v_2. Whilst the overall maximum and minimum values for this ratio remain the same, the range for the new network is always at least as high as the old network. In particular, the full range $0.5 \leq v_3/v_2 \leq 1$ is possible for input ratios in the range $2 \leq Q_H/Q_{CO} \leq 3.5$, where previously the maximum input ratio for this full range was $Q_H/Q_{CO} = 7/3$. In general, the new network allows higher values of v_3/v_2 for low input ratios, and lower values for high input ratios. The old network, on the other hand, allows a lower overall input ratio, and lower values of v_3/v_2 for low input ratios.

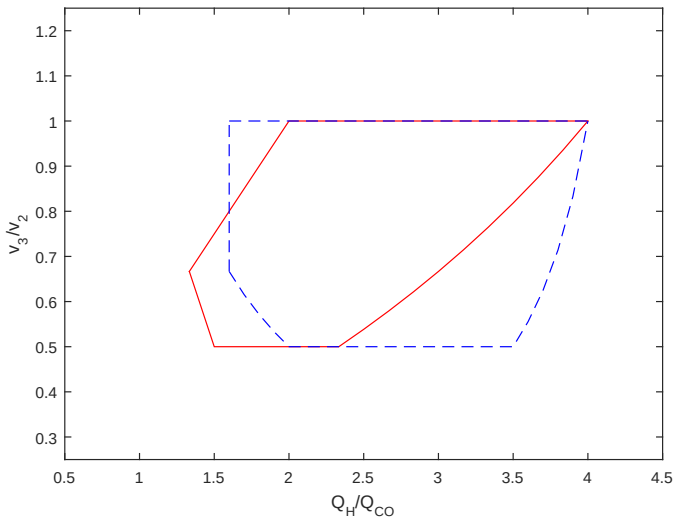

Figure 4.10: Diagram showing the region where FBA predicts the existence of steady-states. The solid line encloses the region permitting steady-states for the old network, shown in Figure 3.1, discussed in Chapters 2 and 3. The dashed line encloses the region where the new network, shown in Figure 4.7, permits the existence of steady-states.

Previously we also considered the effects of the v_{10}/out_{ace} ratio on the boundaries of the steady-state region. This ratio also had the property of being exactly equal to a ratio of rate coefficients in the ODE model, making it a useful ratio to examine. By fixing the value $v_{20}/out_{ace} = k_{10}/k_{20}$ and generating the 2D region where steady-states exist for each value, we investigate whether this ratio has an effect on the steady-state region. In the old network, whilst the range of input ratios allowed did not change, the maximum values for the v_3/v_2 ratio did change. As the v_{10}/v_{20} ratio increased, a higher input ratio was required in order for the v_3/v_2 ratio to take its maximum value of one. Interestingly, in the new network, this effect completely vanishes. The network is always able to reach the highest value for v_3/v_2 (assuming the input ratio is within

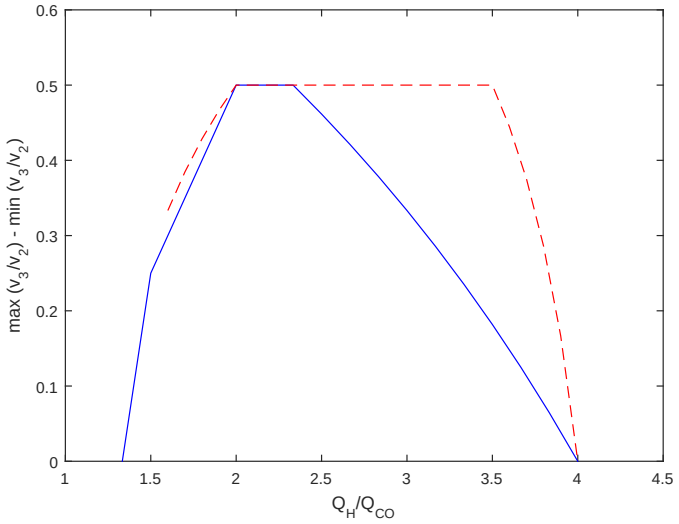

Figure 4.11: Plot of the difference between the minimum and maximum v_3/v_2 ratios for both networks i.e., the height of the region shown in Figure 4.10. The solid line corresponds to the old network, seen in Figure 3.1 and the dashed line corresponds to the new network, seen in Figure 4.7. The range for the new network, whilst it exists, is always at least as high as the range for the old network.

its required bounds) regardless of the value of v_{10}/v_{20}. Indeed, one of the key points of the new network is that FBA predicts that steady-states are always be achievable with $v_3/v_2 = 1$ and $1.6 \leq Q_H/Q_{CO} \leq 4$. This was a feature that was lacking in our previous network, with low input ratios requiring a lower value for v_3/v_2 in order to achieve steady-state.

4.7 Multiple steady-state regions

In section 4.3, we considered the region on our bifurcation diagram between the region where FBA predicted the possibility of steady-states, and the region where the ODEs exhibit behaviour 5. For our old network, this is the region where the ODEs exhibit behaviour 7 in the long timescale. For our new network, FBA predicts a larger region where steady-states could potentially exist, which leads to the shrinking of the region for behaviour 7. In fact, when we plot the FBA steady-state region on the same axes as the analytically-obtained boundary for behaviour 5, $\frac{Q_H}{Q_{CO}} = \frac{8k_3}{k_2+k_3}$, we find that there is an overlap. Figure 4.12 shows this overlapping region. Since this boundary for region 5 is independent of all rate parameters except k_2, or k_3, this means it is possible for us to pick a parameter set that satisfies the condition for behaviour 5, namely

$$\frac{Q_H}{Q_{CO}} \geq \frac{8k_3}{k_2 + k_3} \tag{4.7.1}$$

yet also shows a steady-state behaviour, as predicted by FBA. This suggests that one set of parameters could exhibit multiple long term behaviours. Specifically, it could show both the steady-state behaviour, or behaviour 5 depending on initial conditions.

To test this possibility, we choose a point in the overlapping region, and then use FBA to generate a set of steady-state fluxes for that region. The point we choose

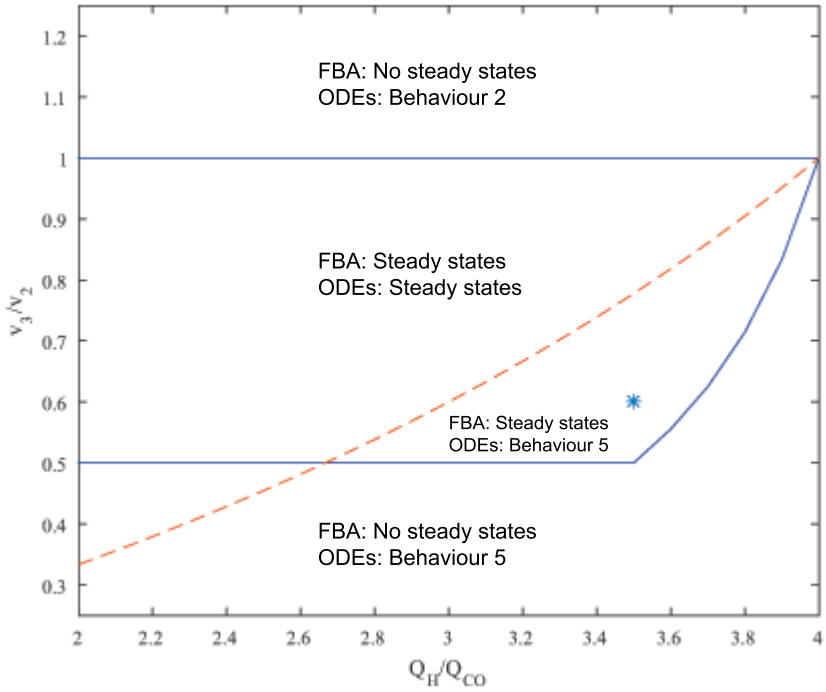

Figure 4.12: Diagram showing the crossover region between the steady-state region predicted by FBA (shown by the solid line) and the region that exhibits behaviour 5 found by analysing the ODEs (shown by the dashed line). The area contained by the solid line that is also under the dashed line is the crossover region, in which multiple long term behaviours exist. The star represents the point use to generate the FBA state described in equation (4.7.3)

is marked on Figure 4.12, and is a point when $Q_H/Q_{CO} = 3.5$, and $k_3/k_2 = 0.6$. Rearranging equation (4.7.1) gives

$$\frac{Q_H}{Q_{CO}} \geq \frac{8k_3/k_2}{1 + k_3/k_2}. \tag{4.7.2}$$

By inserting these values for Q_H/Q_{CO} and v_3/v_2 into equation (4.7.2), we see clearly that the condition for behaviour 5 is met; but according to FBA, a steady-state should also exist. By performing FBA we obtain a possible steady-state,

$$\begin{aligned}
v_1 &= 1.2152, & v_2 &= 1, & v_3 &= 0.6, & v_4 &= 0.6,\\
v_5 &= 0.0513, & v_6 &= 0.4, & v_7 &= 0.2203, & v_8 &= 0.1240,\\
v_9 &= 0.0215, & v_{10} &= 0.0192, & Q_{CO} &= 1.2152, & v_{12} &= 0,\\
v_{13} &= 0, & Q_H &= 4.2532, & out_{bio} &= 0.0032, & out_{lac} &= 0.0075,\\
out_{val} &= 0.0045, & out_{but} &= 0.2158, & out_{eth} &= 0.1432, & out_{ace} &= 0.0023.
\end{aligned} \tag{4.7.3}$$

If we were to fix all the concentrations in our ODEs to equal one, then our reaction rates, v_n would be equal to the parameters, k_n. This means that by setting our rate coefficients, k_n to equal these v_n calculated by FBA, we are able to guarantee at least one steady-state, the case when all concentrations are equal to one. We have also shown that this set of parameters matches the conditions for behaviour 5, and both behaviours should be possible.

To test this, we simulate the system of ODEs with parameters as described by equation (4.7.3) for different initial conditions. Figure 4.13 shows the evolution of the system with all initial concentrations set to equal zero. We see here that the concentrations do not tend to a steady-state in the long timescale. This long timescale behaviour is behaviour 5, as described in Table 4.4. On the other hand, if we set all initial concentrations to equal 0.6, we obtain the behaviour seen in Figure 4.14. This behaviour is the steady-state behaviour, in which all concentrations tend to one.

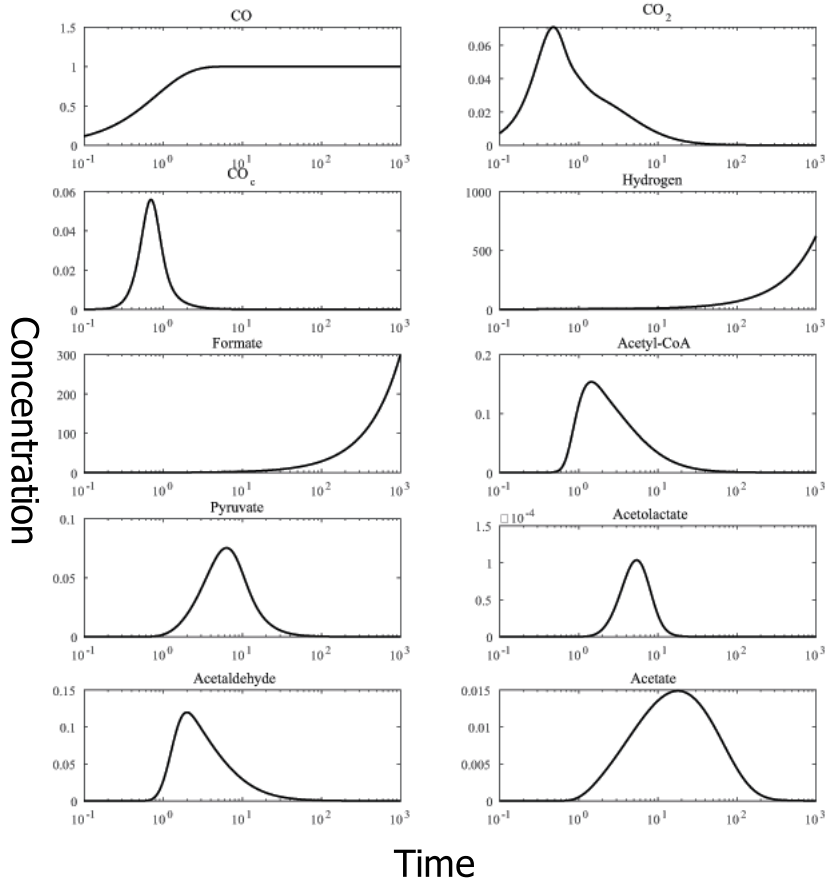

Figure 4.13: Plots of concentrations of metabolites in the ODE model against log time, with parameters set to be those seen in equations (4.7.3) and all initial concentrations equal to zero, showing the long timescale behaviour 5 described in Table 4.5.1.

By testing a variety of initial conditions, we see that most initial conditions tend to either behaviour 5, or the steady-state in which all concentrations are equal to one. This suggests that both of these states are stable. Since they are both stable, it is likely there is a third unstable state between these two, where a small perturbation causes the system to either converge to the steady-state in which all concentrations equal one, or switch into a linearly growing state. By taking equations (4.5.1) – (4.5.10), using parameters from equation (4.7.3), and by fixing the time derivatives to equal zero, we numerically solve the ODEs in steady-state. Obviously, one such solution to the steady-state equations is all concentrations equal to one, as we have already seen. We

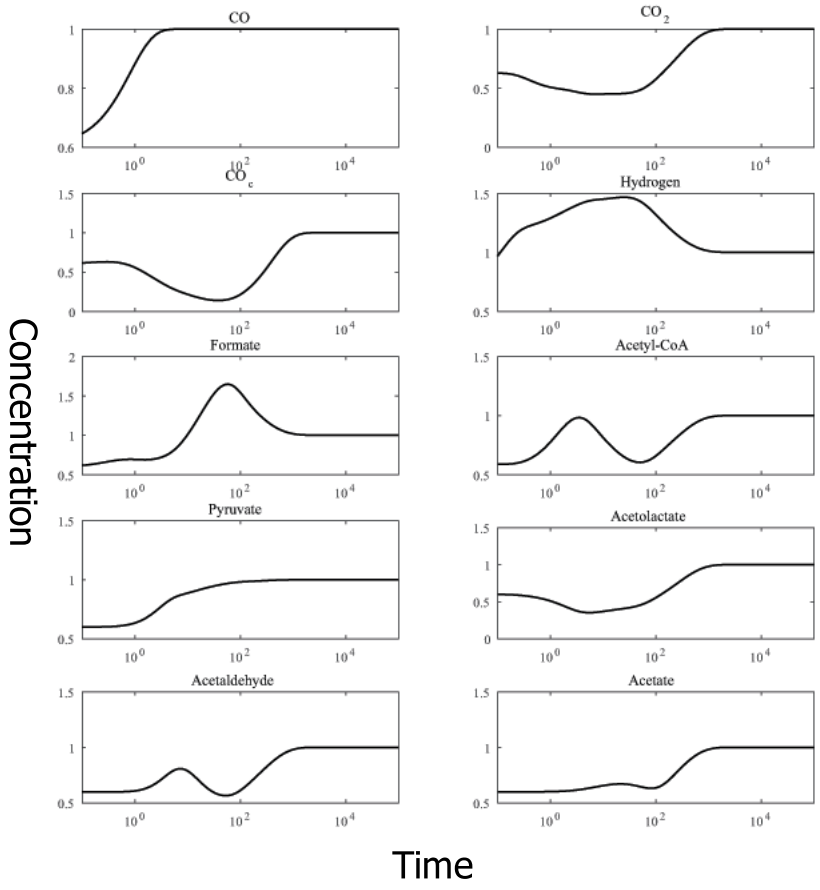

Figure 4.14: Plots of concentrations of metabolites in the ODE model against log time, with parameters set to be those seen in equations (4.7.3) and all initial concentrations equal to 0.6, showing all concentrations tending to steady sate.

also obtain a second solution

$$\begin{aligned} c_{CO}^{(0)} &= 1, & c_{CO_2}^{(0)} &= 0.198, & c_{CO_c}^{(0)} &= 0.0116, & c_F^{(0)} &= 3.4517 \\ c_{ACA}^{(0)} &= 0.2847, & c_{Py}^{(0)} &= 0.9194, & c_{Acl}^{(0)} &= 0.1731, & c_{Acd}^{(0)} &= 0.2542 \\ c_A^{(0)} &= 0.2847, & c_H^{(0)} &= 2.2279, \end{aligned} \tag{4.7.4}$$

where $c_n^{(0)}$ represents a steady-state concentration. This state is shown to be unstable by subjecting it to small perturbations. Figures 4.15 and 4.16 show the system starting very close to this unstable state, and converging to the stable steady-state or the unsteady-state respectively.

We note that if a concentration tends to infinity in the non-steady behaviour, its concentration in the unstable state is higher than one. If the concentration tends to zero in the non-steady behaviour, then its concentration in the unstable state is lower than one. In other words, every concentration for the unstable state lies between the concentration for the stable steady-state (equal to one), and that of the stable non-steady-state. Interestingly, the concentration of carbon monoxide, c_{CO}, tends towards one in all three states.

We have theoretical bounds on where this multistable behaviour occurs, but it is worth numerically analysing this case to find the exact boundaries. Using xppaut [20],

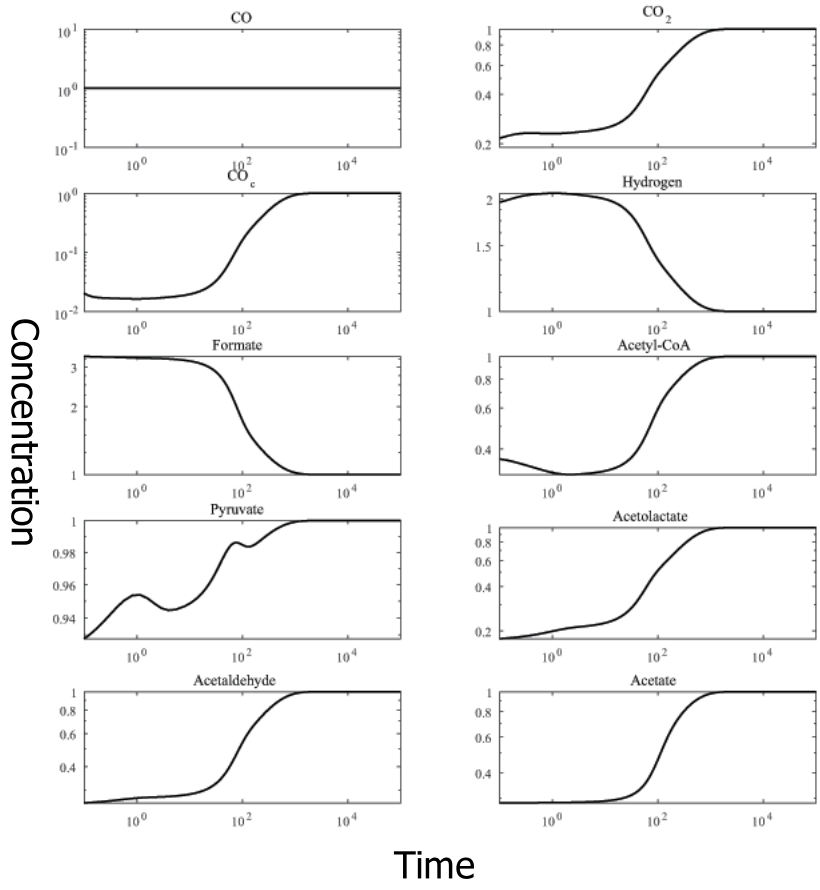

Figure 4.15: Log log plots of concentrations of metabolites in the ODE model against time, with parameters set to be those seen in equations (4.7.3), with initial conditions equal to those seen in 4.7.4, with a slight increase in the concentration of CO_2, we see the concentrations shifting from near the unstable steady-state to the stable steady-state.

we plot a bifurcation diagram for the system, seen in Figure 4.17. We choose to plot the concentration of formate for this diagram, because in the non-steady behaviour 5, it tends to infinity whilst in the steady-state behaviour it tends to a non-zero steady-state. We also expect that as k_3 increases towards one, the concentration of formate tends towards zero as it decays to zero in behaviour two, which we would expect the system to switch to as k_3 becomes greater than one. In Figure 4.17 we observe that for high values of k_3 we have a single stable steady-state behaviour. We also note the concentration of formate tending to zero as k_3 tends to one, as expected. As k_3 decreases past $k_3 = 0.8463$, we find the introduction of a second unstable steady-state. Then as k_3 decreases even further, past $k_3 = 0.5706$, both steady-states are lost via a saddle-node bifurcation, leaving the system only with the non-steady-state behaviour 5. We note that with the value $k_3 = 0.8463$,

$$\frac{8k_3}{k_2 + k_3} > 3.5. \tag{4.7.5}$$

This means the condition, (4.7.1), for behaviour five is not met, but we still have multistable behaviour. By examining the long time scale asymptotics for the unsteady-

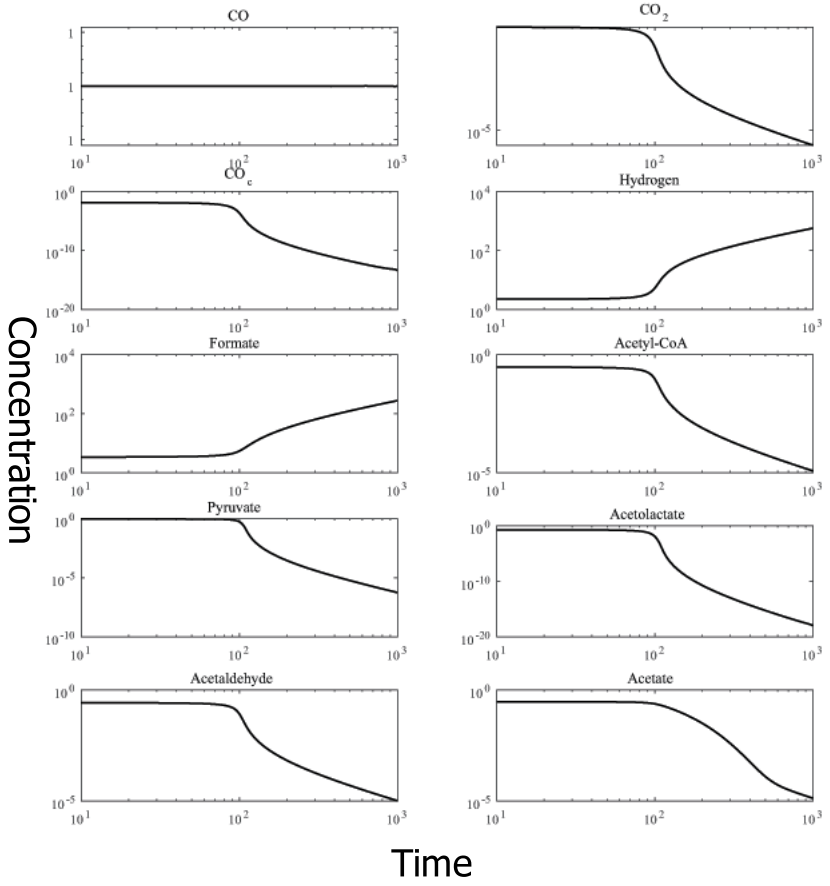

Figure 4.16: Log-log plots of concentrations of metabolites in the ODE model against log time, with parameters set to be those seen in equations (4.7.3), with initial conditions equal to those seen in 4.7.4, with a slight decrease in the concentration of CO_2. We see the concentrations remaining close to the unstable state for a period of time, before approaching behaviour 5, and then rapidly decreasing or increasing as prescribed by that behaviour.

state with these parameters, we see that the unsteady behaviour is actually behaviour 7. When $k_3 > \frac{7}{9}$, the unsteady behaviour is behaviour 7. We now have a complete picture of how the system behaves with these parameters as we vary k_3. For $k_3 < 0.5706$, we have only the non-steady-state behaviour 5. For $0.5706 < k_3 < \frac{7}{9}$, we have two steady-states, one stable and one unstable, and the stable non steady-state behaviour 5. For $\frac{7}{9} < k_3 < 0.8463$ we have the same two steady-states, and the stable non steady-state behaviour 7. Finally for $0.8463 < k_3 < 1$ we only have the steady-state behaviour.

4.8 Conclusions

We have constructed and compared steady-state regions for an FBA model and an ODE model for the same network. By comparing these regions, we found that the steady-state region predicted by FBA is larger than that predicted by the ODE model. In general, this always proves to be the case, as our FBA model is defined by the equations $S\mathbf{v} = \mathbf{0}$. The ODE model is based on the same stoichiometry, but also has

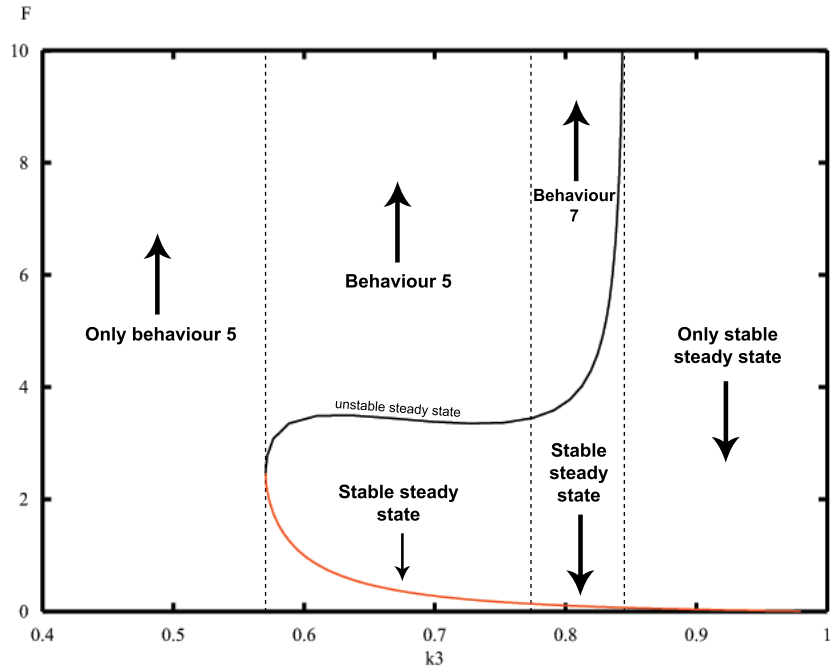

Figure 4.17: Bifurcation diagram for the system (4.5.1)–(4.5.10) using parameters described in 4.7.3. On the horizontal axis, the parameter k_3 is varied, and the vertical axis represents the concentration of formate. The lower branch is the stable steady-state, the upper branch is the unstable steady-state, leading to the stable non-steady-state above it, in which formate tends to infinity.

the additional set of relationships, $\mathbf{v} = \mathbf{v}(\mathbf{c})$, in which the values of the fluxes are also dependent on concentrations. Adding additional constraints such as these cannot increase the region in which $S\mathbf{v} = \mathbf{0}$.

We then made a modification to the network we were examining, so it became the network shown in Figure 4.7. Modifying the network in this way allowed it to correctly balance carbon throughout, and is therefore more biologically accurate. We then performed our asymptotic analysis on a new ODE model (4.5.1)–(4.5.10), to see if the set of behaviours seen in the old model, (3.3.3)–(3.3.12), were still applicable to this new network. We found that only two behaviours were no longer applicable, but equivalent behaviours existed for similar parameter schemes.

We also constructed a new region in parameter space in which FBA predicted the existence of steady-states. By comparing this to the predictions of our new ODE model, (4.5.1)–(4.5.10), we found that this region extended across multiple regions where the ODE models predicted the existence of non-steady-states. In particular, the FBA region extended past the analytically known boundary of region 5, suggesting that there could exist a parameter scheme where the ODE model, (4.5.1)–(4.5.10), could exhibit either steady or non-steady behaviours depending on initial conditions. We found that there was a region in which we saw two steady-states, one stable and one unstable, and the non-steady long term behaviour 5. By analysing these behaviours using xppaut, we were also able to find a region in which we had the two steady-states, and a non-steady long term behaviour 7, and a saddle node bifurcations leading to the regions where we only had steady-state behaviour, or the non-steady behaviour 5.

We were able to find this region of multiple long-term behaviour using a combination of asymptotic analysis on our ODE model, and FBAs predicted steady-state region

simply by comparing the analytically found boundaries from asymptotic analysis of the ODE model, with the computed steady-state boundaries from FBA. Finding this region using only the ODE model would have been challenging, but we do not yet know if it is possible to use FBA to predict the existence of multiple steady-states. It would be useful to look more closely at networks that permit multiple steady-state behaviours to see if these behaviours, and the bifurcations that lead to them, are found using solely flux balance methods. Whilst we could not observe the states with non-steady behaviour, as FBA requires mass balance and only provides steady-state solutions, it would be useful if the method could provide information on the possibility of multiple steady-states in a network.

Chapter 5

Using metabolic models to predict multiple steady-state behaviour

5.1 Introduction

In the previous chapter we found the model for our *C. autoethanogenum* network, under certain parameter schemes, could permit the existence of multiple steady-states. This model presented a system which is able to take one of three states depending on initial conditions, one stable steady-state, one unstable steady-state, and a non-steady-state in which several concentrations grow linearly in time. A more common form of multiple steady-states would be characterised by three steady-states, two stable, and one unstable state separating the two or vice versa.

It is not uncommon for biological systems to have the ability to exist in multiple different states, such as high flux or low flux states. In practice, even in a single bacterial colony, it is possible for different cells to be in different metabolic states. It is important for our models to be able to capture some of this behaviour. Mulukutla et al. [50] investigated a model for glycolysis and examined the role of fructose-6 Phosphate in the regulation of potential multiple steady-states. It was found that by including certain feedback loops in the model, the system could have multiple steady-states. Indeed, it was shown in some situations to have up to five different steady-states, three stable with two unstable separating them. Examination of the conditions required for the steady-states to exist suggested that the system would, under normal physiological conditions, exist in a high flux state. If there was a sudden and sustained lack of glucose however, the system would switch to a low flux state. This is interesting because it shows models can not only capture the multistable dynamics of metabolism, but also show the system switching from one state to another in response to external conditions, such as a lack of a particular nutrient.

Multistability in organisms is also linked to the study of antibiotic resistance in organisms, as shown by Amato et al. [1]. The switch from bacteria growing on one food source to another different food source (thus undergoing a switch from one metabolic steady-state to another) was linked to the creation of "persisters" [1], which are bacterial phenotypes that have developed an antibacterial resistance as a stress response.

We would like to examine systems that exhibit multiple steady-states, and analyse the differences between the states. We would also like to observe bifurcations, when a system switches from only allowing only one steady-state to a position where it permits multiple steady-states. Finally, we wish to find if other methods of modelling

X + 2Y	$\xrightarrow{k_1}$	3Y
X	$\xrightarrow{k_2}$	Y
Y	$\xrightarrow{k_3}$	Catalyst decay
External concentration	$\xleftrightarrow{k_4}$	X
External concentration	$\xleftrightarrow{k_5}$	Y

Table 5.1: The set of chemical reactions in our network as illustrated in Figure 5.1.

metabolism, such as FBA, are able to capture this multistable behaviour. To do this we first introduce a new network to analyse.

5.2 Autocatalysis system

A simple and well studied network that is known to exhibit multiple steady-states is an autocatalysis system. In this system there are two metabolites, X and Y, where Y is a catalyst. There are two reactions that transform X into Y, a direct uncatalysed reaction, and a reaction that uses two Ys as a catalyst. The system also has uptake/output reactions for both X and Y. Finally there is a second output reaction for Y representing it decaying into something not used by the system. The reactions of the system are fully shown in Table 5.1. A network diagram is also shown in Figure 5.1.

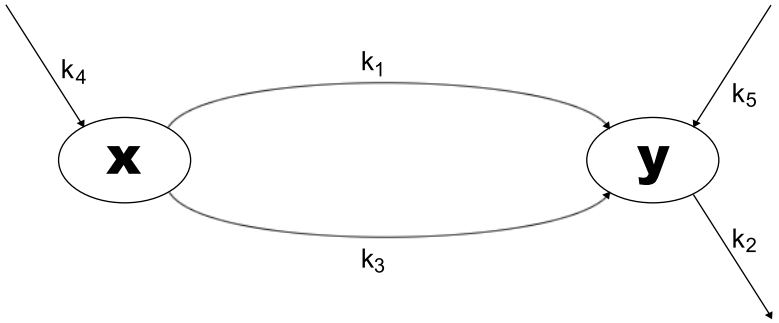

Figure 5.1: Network diagram for the autocatalysis network described by the equations in Table 5.1. Rate coefficients for each reaction and assumed directions are labelled.

Using the law of mass action to model the reaction rates, we formulate a set of two ODEs to model the dynamics of the network

$$\frac{dX}{dt} = k_4(X^* - X) - k_1XY^2 - k_2X, \tag{5.2.1}$$

$$\frac{dY}{dt} = k_5(Y^* - Y) + k_1XY^2 + k_2X - k_3Y, \tag{5.2.2}$$

with initial conditions

$$X(0) = X_0, Y(0) = Y_0. \tag{5.2.3}$$

Here, the $k_n, n \in [1,2,3,4,5]$ are the rate coefficients for the reactions, and X^* and Y^* represent the external concentrations of X and Y respectively, which are assumed to be constant. When examining this model, k_5 is commonly set to equal k_4, so that k_4 represents a rate coefficient for transport in and out of the cell. We also use this simplification in our analysis. We further simplify the model by non-dimensionalising it. A possible non-dimensionalisation for the system is

$$u = \frac{X}{X^*}, \quad v = \frac{Y}{X^*}, \quad \tau = tk_1X^{*2}, \tag{5.2.4}$$

which leads to the non-dimensional parameters

$$a = \frac{k_4}{k_1X^{*2}}, \quad b = \frac{k_2}{k_1X^{*2}}, \quad c = \frac{Y^*}{X^*}, \quad d = \frac{k_3}{k_1X^{*2}}. \tag{5.2.5}$$

In this, we scale all our reaction rates by the rate for the catalysed reaction k_1. Since this reaction is catalysed, we assume it is the quickest reaction in the system. Concentrations (including the constant external concentrations) are scaled by the external concentration of X, X^*.

Taking this non-dimensionalisation, we reduce the system down to a simpler form with only four parameters, a, b, c, and d for the two non-dimensional variables, u and v. The dynamics of this system are governed by the ODEs

$$\frac{du}{d\tau} = a(1-u) - uv^2 - bu, \tag{5.2.6}$$

$$\frac{dv}{d\tau} = a(c-v) + uv^2 + bu - dv. \tag{5.2.7}$$

with initial conditions

$$u(0) = u_0, v(0) = v_0. \tag{5.2.8}$$

By setting the time derivatives to zero, we obtain the nullclines for the system. By combining these two equations, we find a steady-state condition for u

$$F(u,a,b,c,d) = u^3 - 2(1+c)u^2 + \left((1+c)^2 + \frac{(a+d)^2}{a} + b\frac{(a+d)^2}{a^2}\right)u - \frac{(a+d)^2}{a} = 0. \tag{5.2.9}$$

This equation is a cubic, so for a fixed set of parameters, a, b, c, and d, there are up to three values of u that satisfy the equation to provide a steady-state. In fact, by Descartes' rule of sign, there are up to three positive values of u. This means there can exist up to three steady-states for this system. By fixing three of our parameters and varying the fourth, we are able to cause the system to undergo bifurcations from one steady-state to three steady-states. In our case, we choose a as our parameter to vary, and fix b, c, and d.

5.3 Bifurcation analysis

To begin, we choose a set of parameters that is known to allow multiple steady-states for a certain range of a, that is

$$b = 1 \times 10^{-3}, \quad c = 0.1, \quad d = 0.05.$$

Using xppaut [20] we plot a bifurcation diagram using a as our varied parameter. This bifurcation diagram is shown in Figure 5.2. We see a bifurcation pattern where as a increases, we move from one steady-state, to three steady-states, and then back to one again. This pattern shows two saddle nodes, one at $a = 0.2524$ and the second at $a = 0.2694$. This pattern is also sometimes referred to as a "breaking wave" pattern. When $a < 0.2524$ we have a single steady-state with a low u concentration. For a in the range $0.2524 < a < 0.2694$ we have three steady-states, a low u state, a high u state, and an unstable state with u between the high and low concentrations. At the saddle nodes, the unstable state converges to one of the stable steady-states (the high u state for $a = 0.2524$, and the low u state for $a = 0.2694$). For $a > 0.2694$, we have a single steady-state with a high u concentration. Also shown on the diagram for very low values of a is another region of unstable states for a in the range $0.006271 < a < 0.01388$. The values of a correspond to Hopf bifurcations, which give rise to periodic solutions. A periodic solution for $a = 0.01$ is shown in Figure 5.3

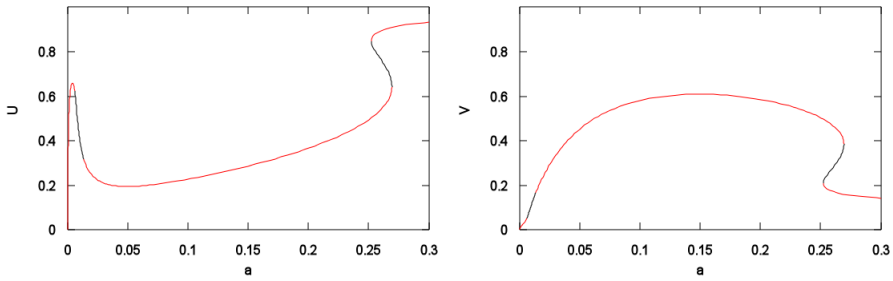

Figure 5.2: Bifurcation diagram for the bifurcation pattern with two saddle nodes observed for $b = 1 \times 10^{-3}$, $c = 0.1$, and $d = 0.05$. The horizontal axes are a, and the vertical axis are u and v respectively. The red branches are stable steady-states, whilst the black branch is the unstable steady-state.

5.3.1 Other bifurcation patterns

The pattern seen in Figure 5.2 with two saddle nodes is a commonly seen bifurcation pattern, but it is not the only one possible for this network. It can be shown that there are up to five different bifurcation patterns that this network can exhibit. Three of these patterns are found when $b = 0$, completely inhibiting the uncatalysed reaction. Using singularity theory it is possible to find the boundaries on c and d that lead to each different pattern, as shown by D'Anna et al. [16]. To do this, we need our steady-state condition, equation (5.2.9), which we call $F(u, a, b, c, d)$. In steady-state we have $F(u, a, b, c, d) = 0$. If we also require certain partial derivatives of this to equal zero, we find the boundaries in parameter space for each type of bifurcation diagram. If we require

$$F(u, a, b, c, d) = F_u(u, a, b, c, d) = F_{uu}(u, a, b, c, d) = 0 \tag{5.3.1}$$

then we have a set of three equations. Two of these are used to eliminate u and a, then the third gives us an equation for b, c, and d that defines one of these regions in parameter space. The same method allows us to find a region for the set of equations

$$F(u, a, b, c, d) = F_u(u, a, b, c, d) = F_a(u, a, b, c, d) = 0. \tag{5.3.2}$$

Since in this situation we have $b = 0$, the solutions to both these equations can be plotted on a plane to show the regions that lead to each bifurcation diagram. In one

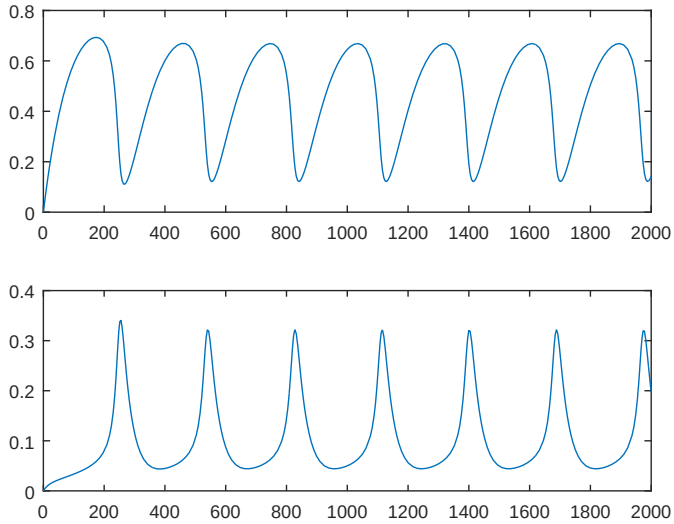

Figure 5.3: Numerical simulation for the ODEs (5.2.6)–(5.2.7) with parameters $a = 0.01$, $b = 1 \times 10^{-3}$, $c = 0.1$, $d = 0.05$ and initial conditons $u(0) = 0$, $v(0) = 0$. The system with these parameters exhibits a stable periodic solution.

region we find mushroom patterns, in which there are four saddle nodes. In the second we find isola patterns where there are two saddle nodes, but the unstable state converges to the same stable state at both bifurcation points. Finally, in the third we find a single steady-state for all values of a with no bifurcations.

Allowing a value of b that is greater than zero allows use to observe two additional bifurcation patterns, as well as the three seen with $b = 0$. Again using singularity theory, Kay et al. [34] were able to plot the regions in parameter space that lead to each bifurcation pattern. In this, since there is a non-zero b, there is a third parameter to plot. Rather than plotting a 3D diagram to show this, instead they plotted several 2D (c,d) planes for various values of b. In this way they were able to observe five different regions leading to five different bifurcation patterns. Including the three already seen when $b = 0$, there are the breaking wave pattern shown earlier, and a pattern containing both a breaking wave pattern and an isola pattern. It was also shown that for $b > 1/27$ there is only ever a single steady-state regardless of other parameter values. Examples of the mushroom and isola patterns are shown in Figure 5.4, and example parameters leading to those patterns are given in the caption. The pattern which is a combination of both isola and breaking wave patterns is only found in a very small set of parameters, and is not shown here. Observing the patterns for isola and mushrooms in Figure 5.4 we see that when there are three steady-states, these also seem to be a stable high u state (with low v), a stable high v state (with low u), and an unstable state in between the two. For both isola and mushrooms, we find that for both high and low values of a, we only attain a high u, low v state, with the high v, low u state only being attainable by choosing values of a between bifurcation points. For the isola, we require $0.02558 \leq a \leq 0.1134$. For the mushroom, the range is a little larger, with a needing to be in the range $0.01885 \leq a \leq 0.1281$. There are also some values of a that lead to two unstable states and a single stable state in both the isola and the mushroom. These are again due to Hopf bifurcations, and the existence of stable periodic solutions. Therefore we are also able to find stable periodic solutions existing with other unstable and stable steady solutions.

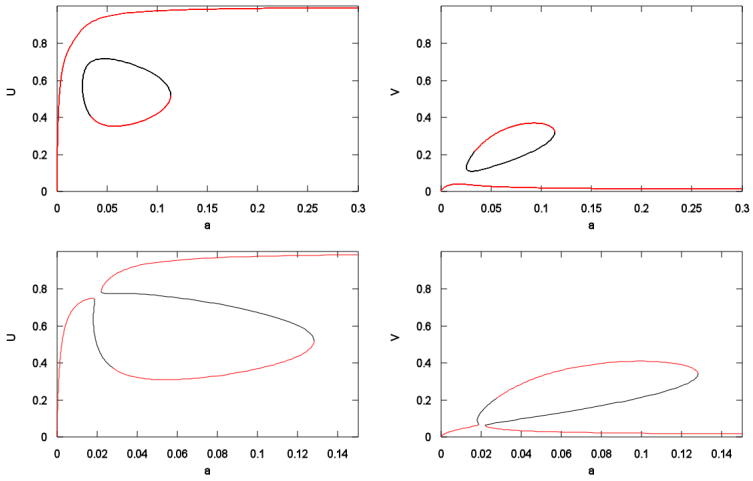

Figure 5.4: Bifurcation diagrams for the mushroom and isola patterns observed. The first row is the isola pattern seen when $b = 0.002$, $c = 0.01$, and $d = 0.06$. The second row is the mushroom pattern seen when $b = 0.002$, $c = 0.01$, and $d = 0.056$. The horizontal axes are a, and the vertical axes are u (left panels) and v (right panels). In all graphs, red lines represent stable branches and black lines represent unstable branches. Note the existence of multiple unstable branches at once on the isola and left half of the mushroom. In these instances, the system also permits a stable limit cycle, so the system can have three steady-states (one stable and two unstable) with a stable periodic solution between the two unstable branches.

Here, we have characterised the different steady-states as high or low concentration states, however often biological systems are though of in terms of high or low flux states. In addition, some techniques for modelling metabolism refer only to the fluxes in the system rather than the concentrations. In experiments, it may also be difficult to measure the concentrations of metabolites, whereas measuring certain flux values may prove to be much easier. For greater applicability it would be useful to consider the flux vector of the steady-states. That is, rather than characterising the states by their concentrations, we instead consider the reaction rates.

5.4 A flux perspective

In order to examine the flux vector of a steady-state, we must first calculate the reaction rates in our system. Our autocatalysis network has five reactions, listed in Table 5.1. The reaction rates for these are calculated from the set of ODEs (5.2.6)–(5.2.7) as

$$\begin{aligned} Q_u &= a(1-u), \quad Q_v = a(c-v), \quad q_1 = uv^2, \\ q_2 &= dv, \qquad\qquad q_3 = bu. \end{aligned} \tag{5.4.1}$$

The external transport reactions Q_u and Q_v are both reversible, since for high u and v respectively, they transport the metabolites out of the system rather than in. All other reactions are irreversible. A diagram for the network with the fluxes labelled is given in Figure 5.5

First, let us examine the three states seen in the breaking wave shown in Figure 5.2. We choose $a = 0.26$ to give us three steady-states to analyse. Figure 5.6 shows

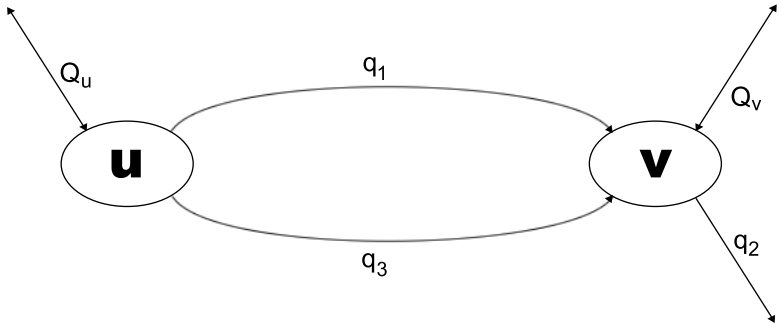

Figure 5.5: Network diagram for the autocatalysis system. Fluxes and directions are labelled for each reaction.

Description	u	v	q_1	q_2	q_3	Q_u	Q_v
High u steady-state	0.8933	0.1733	0.0268	0.0087	0.0009	0.0277	-0.0191
Low u steady-state	0.5389	0.4706	0.1193	0.0235	0.0005	0.1199	-0.0964
Unstable steady-state	0.7678	0.2787	0.0596	0.0139	0.0008	0.0604	-0.0465

Table 5.2: Concentrations and reaction fluxes for the ODEs (5.2.6)–(5.2.7) in steady-state with parameters $a = 0.26$, $b = 1 \times 10^{-3}$, $c = 0.1$, and $d = 0.05$. The first two listed states are stable, whilst the third is unstable

the numerical solutions of the ODEs (5.2.6)–(5.2.7) starting close to the unstable state and eventually converging to one of the stable steady-states. This demonstrates the existence of all three steady-states with this parameter scheme. We calculate values for the concentrations u and v for all three steady-states. We then substitute these values into the equations for the fluxes seen in equation (5.4.1). This allows us to calculate the flux for each reaction in every steady-state. Concentrations and fluxes for each steady-state are given in Table 5.2.

From these results, we see that the flux taking up v, Q_v, is negative for each of the three potential steady-states. For these parameters, the system always outputs v through Q_v, rather than taking it up. In fact, for every value of a that leads to these multiple steady-states in the breaking wave pattern we find that in steady-state the system always outputs v i.e. $Q_v < 0$. This is because the catalyst decay level, q_2 is low, so the system has an excess of v. Interestingly, the periodic solution seen in Figure 5.3 cycles between taking up v and outputting v.

The first state shown in Table 5.2 is the high u state. In this state we see the highest flux through the uncatalysed reaction q_3. We also see the lowest levels of catalyst decay, q_2, and the lowest overall levels of uptake of u and output of v. The second state is the low u state. In this state we see a higher flux through the system overall and greater output of v. It has the lowest overall use of the uncatalysed reaction, q_3, and the highest levels of catalyst decay, q_2. The third (unstable) state has concentrations somewhere between the two stable states, and the fluxes observed are also between the two previous stable states.

Whilst the overall levels of flux through the system are useful, it is perhaps more instructive to examine the relative levels of flux to understand the differences between the three states. For example, we see from the overall levels of flux that the low u state is the most effective for producing larger quantities of v in a short time, but this state

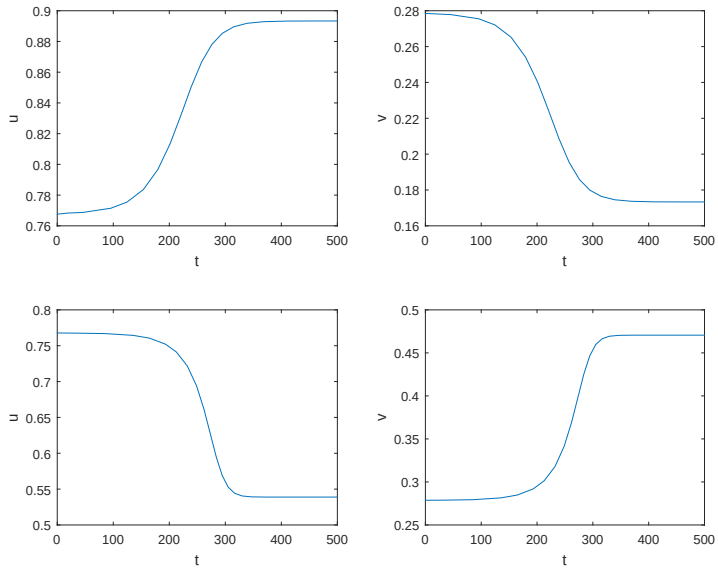

Figure 5.6: Numerical simulations for equations (5.2.6)–(5.2.7) with parameters $a = 0.26$, $b = 1 \times 10^{-3}$, $c = 0.1$, and $d = 0.05$ and initial conditions close to $u(0) = 0.77$, $v(0) = 0.28$. The top row shows u and v starting near the unstable steady-state, and tending towards the high u steady-state. The bottom row shows u and v starting near the unstable steady-state and tending towards the low u steady-state.

State	$\frac{q_1}{Q_u}$	$\frac{q_2}{Q_u}$	$\frac{q_3}{Q_u}$	$\frac{Q_v}{Q_u}$
High u steady-state	0.9678	0.3125	0.0322	-0.6875
Low u steady-state	0.9955	0.1963	0.0045	-0.8037
Unstable steady-state	0.9873	0.2307	0.0127	-0.7693

Table 5.3: Relative reaction fluxes for the ODEs (5.2.6)–(5.2.7) in steady-state, with parameters $a = 0.26$, $b = 1 \times 10^{-3}$, $c = 0.1$, and $d = 0.05$.

may be less efficient overall due to the increase in catalyst decay. Table 5.3 shows us the fluxes in steady-state scaled by Q_u. This enables us to see which reactions are more active in each state.

From the relative fluxes in Table 5.3 we find that, relative to Q_u, the most efficient production of v is from the low u state. The relative levels of catalyst decay are much smaller in this case. In fact, the relative levels of catalyst decay in the low u case are around 60% of those in the high u case. Therefore, over a long period of time, the low u state is actually more efficient at outputting v through Q_v. By examining the fluxes, rather than just the concentrations, we are able to see that whilst the high u state has a lower level of catalyst decay in a short time period, the lower u state is actually more efficient at output v through Q_v over a longer time scale.

Mode	q_1	q_2	q_3	Q_u	Q_v
1	1	0	0	1	-1
2	0	0	1	1	-1
3	0	1	0	0	1

Table 5.4: The set of elementary modes for the autocatalysis system.

5.4.1 Elementary modes

Flux vectors are only one way of visualising the flux data. Another method of this would be using elementary modes. We decompose our steady-state flux vectors into positive sums of elementary modes. In this way, rather than considering the values of the fluxes themselves, we analyse how much flux there is through certain pathways within the system. In a larger network, this could help us determine which pathways are highly active in a steady-state, and which pathways are less active, and therefore could potentially be inhibited with little effect to the overall behaviour of the system.

To calculate the elementary modes, we first require the stoichiometric matrix of the network. The stoichiometric matrix for the system defined by the reactions in Table 5.1 is

$$S = \begin{pmatrix} -1 & 0 & -1 & 1 & 0 \\ 1 & -1 & 1 & 0 & 1 \end{pmatrix}. \tag{5.4.2}$$

We remember that both Q_u and Q_v are considered to be reversible. Using CellNetAnalyzer [38], we compute the elementary modes for this system. There are three modes, which are shown in Table 5.4. The first two are routes with input through Q_u and output through Q_v (a negative flux). The third is a route from Q_v to q_2. These fluxes are illustrated in Figure 5.7. Each diagram shows the reactions active for each elementary mode, with directions shown. We note that none of the pathways lead to an output of u from the system, which tells us that if these reversibility conditions are correct, it is impossible for the system to exist in a steady-state that produces u. We also note that the elementary modes consider any flux from Q_u to q_2 to be a sum of one of the modes leading from Q_u to Q_v and the mode leading from Q_v to q_2. This does not necessarily mean that the system is outputting v via Q_v and then taking it up again, it is just how the elementary modes have come out in the simplest form.

Though in practice this is generally not the case, our elementary modes form a basis for the nullspace of the stoichiometric matrix S shown in equation (5.4.2). As our modes form a basis, we are able to uniquely decompose our steady-state flux vector into a positive sum of these elementary modes. We decompose our steady-state flux vectors found in Section 5.4 for $a = 0.26$. First, we write our modes as column vectors $\mathbf{m}_n$, where n is the mode number shown in Table 5.4. We have

$$\mathbf{m}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ -1 \end{pmatrix} \quad \mathbf{m}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ -1 \end{pmatrix} \quad \mathbf{m}_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{5.4.3}$$

The high u state can be written as

$$\mathbf{q}_{hi} = 0.0268\mathbf{m}_1 + 0.0009\mathbf{m}_2 + 0.0087\mathbf{m}_3. \tag{5.4.4}$$

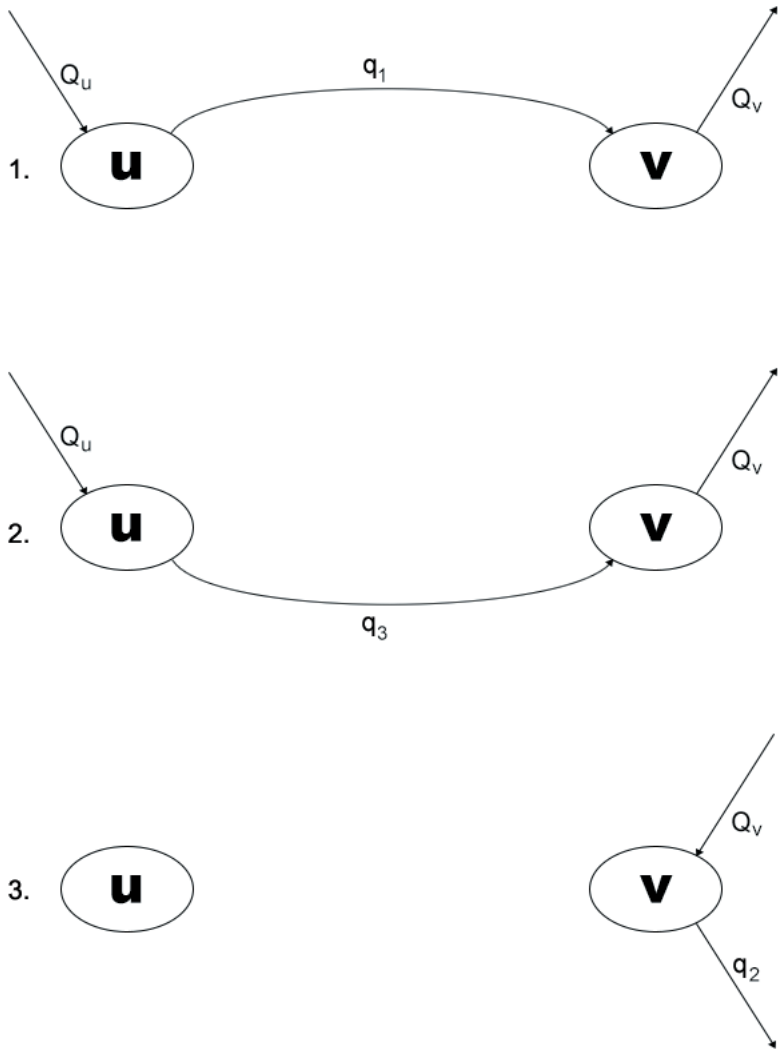

Figure 5.7: Network diagrams for each elementary mode in the system. For each mode only the active reactions are shown, along with their directions.

Similarly the low u state can be written as

$$\mathbf{q}_{low} = 0.1193\mathbf{m}_1 + 0.0005\mathbf{m}_2 + 0.0235\mathbf{m}_3, \tag{5.4.5}$$

and the unstable state can be written as

$$\mathbf{q}_{unst} = 0.0596\mathbf{m}_1 + 0.0008\mathbf{m}_2 + 0.0139\mathbf{m}_3. \tag{5.4.6}$$

As before it is clear that the low u state has a much higher utilization of the first mode and the catalyst decay mode $\mathbf{m}_3$. For clarity, we scale these sums so that the coefficient for $\mathbf{m}_1$ is equal to one. Then the high u state is

$$\mathbf{q}_{hi} = 1\mathbf{m}_1 + 0.0333\mathbf{m}_2 + 0.3229\mathbf{m}_3, \tag{5.4.7}$$

the low u state is

$$\mathbf{q}_{low} = 1\mathbf{m}_1 + 0.0045\mathbf{m}_2 + 0.1972\mathbf{m}_3, \tag{5.4.8}$$

and the unstable state is

$$\mathbf{q}_{unst} = 1\mathbf{m}_1 + 0.0129\mathbf{m}_2 + 0.2337\mathbf{m}_3. \tag{5.4.9}$$

These scaled sums provide similar results to those seen by just examining the flux. The absolute flux leaving the system is less clear, but we clearly see that the relative catalyst decay is much lower in the low u state, and the catalysed reaction in the $\mathbf{m}_1$ mode is heavily favoured over the uncatalysed reaction in the $\mathbf{m}_2$ mode. In this case, we see the information more clearly by simply looking at the active fluxes, however in other cases it may be more clear that alternative pathways are being used by different steady-states by looking directly at these elementary modes rather than the fluxes.

All the analysis so far has been performed on steady-states obtained from a full ODE model. Often it is the case for biological systems that such a model is not available. Obtaining parameters is difficult and accurate modelling of kinetics may also require more data than we have access to. Conversely, data for the structure of the system may be readily available. We would like to be able to observe and analyse these multiple steady-states using optimisation methods such as FBA.

5.5 Optimisation methods

Before moving directly to using flux balance analysis and linear constraints for optimisation, it would be useful to check that we are able to obtain information about the existence of steady-states of ODE models using optimisation methods. The simplest way to observe multiple steady-states via optimisation would be to optimise a system with one parameter we vary across a range. If we plot the optimised values of the concentrations against the value of the varied parameter, we would expect to see a "jump" in the solutions, representing a switch from a high u or v solution to a lower one, or vice versa. This would be a clear sign of the existence of multiple steady-states. In some cases the same solution may be the optimal solution for all values of the parameter. For example if we examined the maximum value for u in the isola case seen in Figure 5.4, we would find a continuous solution for all values of a. By searching for both minima and maxima we avoid this problem, since it is highly unlikely that one solution could take both the minimum and maximum values over the entire range of a.

A good choice of optimisation technique for this purpose is the method of Lagrange multipliers, which allows us to find both minimum and maximum solutions using the same set of equations. This technique also requires complete knowledge of the system, including kinetic parameters.

5.5.1 Lagrange multipliers

For Lagrange multipliers we require a function to maximise or minimise and some conditions to place on the variables. An obvious choice for the conditions is requiring mass balance. That is, we require the ODEs of our metabolic system to be in steady-state. We know that one of the key differences between states was the rate of catalyst decay, dv, so we choose this as our objective function that we wish to minimise or maximise. Then the initial setup for an optimisation problem is

$$\begin{aligned} \underset{u,v}{max/min} \quad dv \quad \text{such that: } & a(1-u) - uv^2 - bu = 0, \\ & a(c-v) + uv^2 + bu - dv = 0. \end{aligned} \tag{5.5.1}$$

To solve this optimisation problem, we introduce additional variables, λ_n, known as Lagrange multipliers. Since we have two constraints, we require two multipliers. We write the Lagrangian,

$$L = dv - \lambda_1[a(1-u) - uv^2 - bu] - \lambda_2[a(c-v) + uv^2 + bu - dv]. \tag{5.5.2}$$

To solve this problem we take partial derivatives with respect to u, v, λ_1, and λ_2. In order to find a minimum or a maximum, we require that these partial derivatives are all equal to zero. This gives a set of four equations for four unknowns which can be solved analytically or numerically. The four equations are

$$\frac{\partial L}{\partial u} = 0 = -\lambda_1(-a - v^2 - b) - \lambda_2(v^2 + b), \tag{5.5.3}$$

$$\frac{\partial L}{\partial v} = 0 = d - \lambda_1(-2uv) - \lambda_2(-a + 2uv - d), \tag{5.5.4}$$

$$\frac{\partial L}{\partial \lambda_1} = 0 = a(1 - u) - uv^2 - bu, \tag{5.5.5}$$

$$\frac{\partial L}{\partial \lambda_2} = 0 = a(c - v) + uv^2 + bu - dv. \tag{5.5.6}$$

We note that equations (5.5.5) and (5.5.6) simply recover the steady-state conditions. These two equations permit only one, two, or three solutions so we are sure there is a maximum and a minimum solution to these equations. We choose our parameters b, c, and d so that we observe a bifurcation as a is varied. Here we take $b = 0.002$, $c = 0.01$, and $d = 0.06$, which recovers the isola pattern seen in Figure 5.4. We then solve the equations (5.5.3)–(5.5.6) over a range of a. For most values of a, the steady-state conditions only permit one solution, which is both the minimum and the maximum. When the system permits multiple steady-states, we find two of those states, a maximum and a minimum. Figure 5.8 shows all solutions to these equations plotted on the same graph. In the range $0.02558 \leq a \leq 0.1134$ there are two solutions. Since we are trying to maximise or minimise the catalyst decay, dv, the solution with the highest v maximises dv, and the lower v solution minimises dv. This means, where it exists, we take the solution on the isola to be the maximised value for dv. We note that the maximum solution persists even in the region of the isola where both steady-states are unstable due to the Hopf bifurcation. The minimum solution is the continuous solution not on the isola. If we take another set of parameters b, c, and d, we find

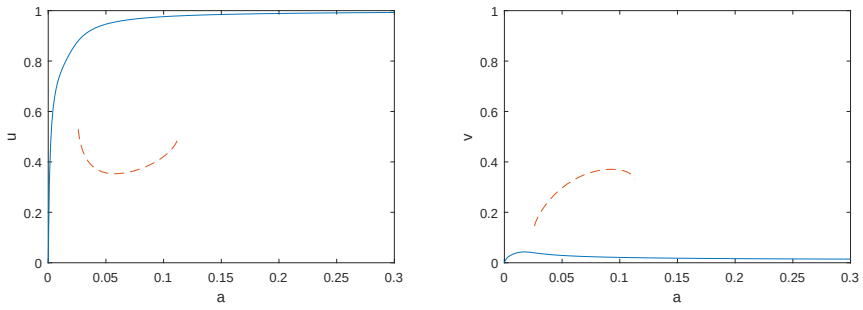

Figure 5.8: Solutions to the set of simultaneous equations (5.5.3)–(5.5.6) with $b = 0.002$, $c = 0.01$, $d = 0.06$, and a varied in the range $0 \leq a \leq 0.3$. Where two solutions exist, the dotted lines represent the second solution, which has a discontinuity to the areas with only a single solution.

another bifurcation pattern, such as pattern with two saddle nodes, seen in Figure 5.2 or the mushroom, seen in Figure 5.4. Both of these bifurcation patterns give rise to a discontinuity in both the minimum and maximum solutions, as shown in Figure 5.9. Here, for $0 \leq a < 0.2524$, the minimum and maximum solutions are equal. As a increases past the point $a = 0.2524$, the minimum solution jumps to the branch of steady-state solutions with a lower value of v. Then in the region $0.2524 \leq a \leq 0.2694$

the two solutions are separate. Finally as a increases past $a = 0.2694$, the maximum value solution jumps down to the lower v branch, as the higher v branch ceases to exist. In this way we have observed the discontinuous behaviour we hoped to find in an optimisation problem.

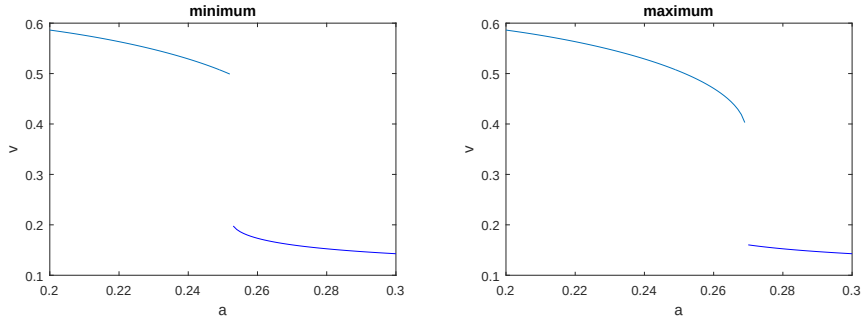

Figure 5.9: Solutions to the set of simultaneous equations (5.5.3)–(5.5.6) with $b = 1 \times 10^{-3}$, $c = 0.1$, $d = 0.05$, and a varied in the range $0.2 \leq a \leq 0.3$. On the left is the solution corresponding to the minimum of dv and on the right is the solution corresponding to the maximum of dv. Both solutions exhibit discontinuities.

In our examination of the relative fluxes in the steady-state in Section 5.4, we found that the lower v solution had the smaller overall catalyst decay. However, when the system was scaled so that $Q_u = 1$, the catalyst decay was lower in the high v solution. This shows the limitations of linear objective functions such as dv. Whilst Lagrange multipliers have provided the overall maximum or minimum, it may be more useful to use a more complicated objective function such as the ratio of catalyst decay to flux in

$$F(u, v) = \frac{dv}{a(1-u)},$$

though for nonlinear objective functions, care must be taken as they may turn out to be non-convex. With a non convex objective function, we may find our optimisation methods producing local minima rather than global minima.

We now know that it is possible to identify the existence of multiple steady-states of ODE models using optimisation methods. Now we would like to apply the same principle to flux balance analysis and identify the existence of multiple steady-states of flux models using FBA.

5.5.2 Flux balance analysis

Before specifically looking into FBA, it is worth noting that we can examine a metabolic structure and determine whether or not it could potentially contain unstable steady-states. Without unstable steady-states we cannot have multiple stable states, and we are unlikely to see any periodic solutions and limit cycles. Wilhelm [79] presented a technique to identify whether a system could potentially exhibit unstable states. By taking the system $S\mathbf{v} = \mathbf{0}$, and applying the Routh-Hurwitz stability criterion, we find whether or not the system could have an unstable state. If the system satisfies the criterion, it does not contain an instability causing structure, and therefore all steady-states will be stable.

Our stoichiometric matrix S is given in equation (5.4.2). For both Q_u and Q_v we set minimum values of $-\infty$ since the reactions are reversible. All other reactions take

a minimum value of zero. We allow the maximum internal reaction and catalyst decay rates to be infinite. The overall rate of flux through the system is largely governed by the rate of uptake, so we do not impose any maximal rates on the internal reactions so as to not unnecessarily stifle the uptake rates. Next we set the maximum uptake rates; for Q_u, we set this to equal a, which is the same parameter seen in the ODEs (5.2.6)–(5.2.7). Since we are looking for bifurcations, we vary a across a range of values to observe the bifurcations we already know to be there. We also set the maximum of Q_v to equal ac, where c is also the same parameter seen in (5.2.6)–(5.2.7). As c is an external concentration, measuring it should not be difficult in practice. Finally we choose an objective function. We continue to use $dv = q_2$ as our objective function, as we know it to be a key difference between the multiple steady-states observed in the ODE model.

Then we have the flux balance problem,

$$\begin{aligned} \underset{q_1,q_2,q_3,Q_u,Q_v}{max/min} \quad & q_2 \quad \text{such that: } S\mathbf{q} = \mathbf{0}, \\ & 0 \leq q_1, q_2, q_3 \leq \infty, \\ & -\infty \leq Q_u \leq a, \\ & -\infty \leq Q_v \leq ac. \end{aligned} \tag{5.5.7}$$

We set $c = 0.1$ as in our breaking wave example, and vary a across the range $0 \leq a \leq 10$ to attempt to observe a switch in the system. Figure 5.10 shows the results of this FBA. In this case the minimum value is given when $q_2 = 0$. Unfortunately, the desired behaviour is not observed. Since FBA picks the optimal steady-state, it is likely that the parameters required to achieve the state are not those that allow us to observe multiple steady-state behaviour. In fact, we see all the flux into the system directed towards catalyst decay, so the maximum value for q_2 is equal to $1.1a$ (that is, the maximum value for Q_u plus the maximum value for Q_u). It is worth noting that neither the minimum nor maximum flux vectors satisfying the optimisation 5.5.7 are unique. The maximum flux vector maintains one degree of freedom, and the minimum has two. In the case of the maximum solution, both external fluxes equal their maxima, so we end up with a system of four equations for five unknowns, leaving one degree of freedom. In the case of the minimum solution, q_2 is equal to its minimum value, so we have three equations for five unknowns, leaving two degrees of freedom. In metabolic systems, non-unique optimum solutions are common, due to the structure of stoichiometric matrices. Many metabolites are only used in one or two reactions, and most reactions only use a few metabolites, leading to sparse matrices with a large degree of degeneracy.

We have seen that our simple flux balance solution does not lead to results resembling the results we have seen from our ODEs, especially our Q_v flux which is generally used as an output in our ODE solutions, but is used as input for the flux balance solutions. We add more constraints to the system in order to restrict it to solutions more accurate to the ODE model.

In order to make our FBA model behave more like the ODE solutions, we must use some of our knowledge of the kinetics and reaction rates in that model. Specifically by examining the kinetics of Q_u, Q_v, q_3 and q_2, we construct an additional set of two linear constraints for our FBA. Using these additional constraints comes at a cost of needing to know more about the kinetics of the system and values of the parameters involved.

For the first constraint we have, from equation (5.4.1), $Q_u = a(1-u)$, and $q_3 = bu$. We rearrange both of these to write u in terms of fluxes and parameters, and equate them to find

$$1 - \frac{Q_u}{a} = u = \frac{q_3}{b} \tag{5.5.8}$$

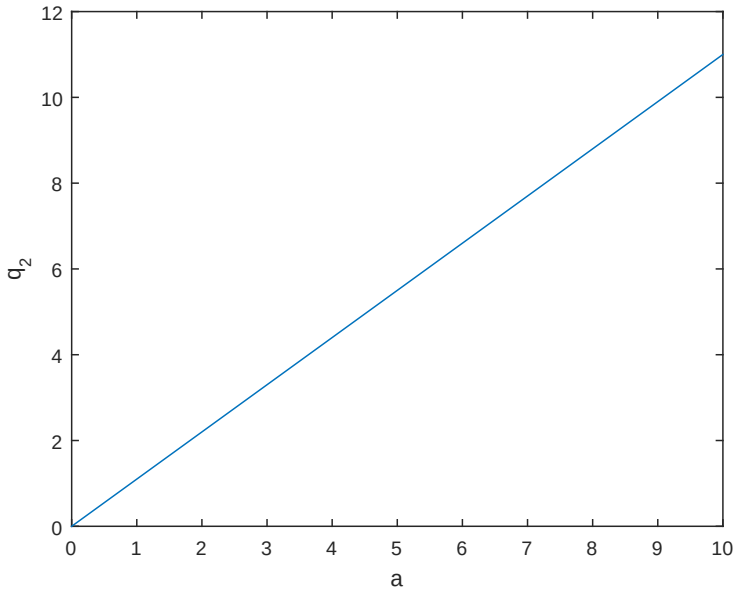

Figure 5.10: Maximum and minimum values for q_2 for the FBA problem described in equation (5.5.7) across a range of values for a. We see a linear growth in the maximum value for q_2 as a increases, whilst the minimum remains at zero.

leading to the constraint

$$\frac{Q_u}{a} + \frac{q_3}{b} = 1. \tag{5.5.9}$$

The second constraint is found in the same way, using the equations from equations (5.4.1) for Q_v and q_2. We have $Q_v = a(c - v)$, and $q_2 = dv$. We rearrange these in the same way, writing u in terms of fluxes and parameters, leading to the equality

$$c - \frac{Q_v}{a} = v = \frac{q_2}{d}, \tag{5.5.10}$$

which is then rearranged to find the constraint

$$\frac{Q_v}{a} + \frac{q_2}{d} = c. \tag{5.5.11}$$

These linear constraints are added to the optimisation problem described in equation (5.5.7) to give the new optimisation problem

$$\begin{aligned} \underset{q_1,q_2,q_3,Q_u,Q_v}{max/min} \quad & q_2 \quad such\ that : S\mathbf{v} = \mathbf{0}, \\ & \frac{Q_u}{a} + \frac{q_3}{b} = 1, \\ & \frac{Q_v}{a} + \frac{q_2}{d} = c, \\ & 0 \leq q_1, q_2, q_3 \leq \infty, \\ & -\infty \leq Q_u \leq a, \\ & -\infty \leq Q_v \leq ac. \end{aligned} \tag{5.5.12}$$
$$\tag{5.5.13}$$

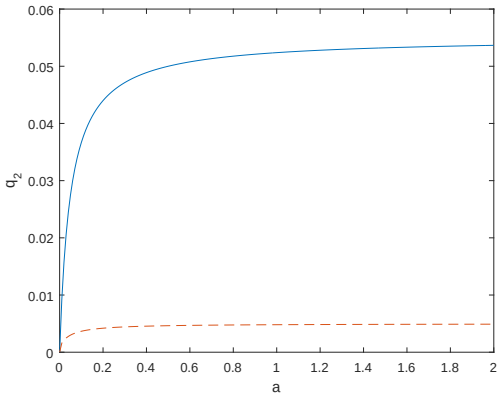

Figure 5.11: Maximum values for q_2 for the FBA problem described in equation (5.5.7) with the additional constraints (5.5.9), and (5.5.11) across a range of values for a. The solid line represents the maximum solution for q_2, whilst the dashed line represents the minimum solution. We see the both minimum and maximum values for q_2 asymptotically approaching maximum values as a increases.

Minimum and maximum values of q_2 plotted against a are given in Figure 5.11. Unfortunately, this still does not provide us with results similar to the multiple steady-states seen from Lagrange multipliers, seen in Figure 5.9. With the additional constraints we still see the optimal value of q_2 increasing as a increases, though this time it appears to approach a limit rather than growing linearly with a, however this limit does not appear to match the values for $q_2 = dv$ observed in the steady-states seen in the ODE. Contrary to our previous FBA results without the additional constraints given in equations (5.5.9) and (5.5.11), these optimal minimum and maximum solutions are unique. Here we have four constraints to begin with, and with optimisation at least one of our inequality constraints takes its extreme value, leading to at least a fifth equality governing five unknowns, guaranteeing a unique solution for a given objective function, however these unique solutions persist for all values of a, and unlike our ODE analysis (Section 5.3), we never see the maximum solution becoming equal to the minimum solution or vice versa (except in the trivial case where $a = 0$).

By examining the flux vectors for our FBA solutions and comparing them with those found via Lagrange multipliers (or simply those found via the ODEs), we see where FBA is going wrong in its predictions for steady-state flux modes, or rather, where FBA differs most from the predicted steady-states of the ODE system. Table 5.5 shows the minimum and maximum flux values given by each method for $a = 0.26$.

It is clear from this table that our flux balance solutions are allowing lower minimum and higher maximum values of q_2, though they remain on the same order of magnitude as those found via Lagrange multipliers. Interestingly we see that the minimum FBA solution actually uses Q_v as an input flux again. In our ODEs, this would indicate an extremely low concentration of v in the system, leading to $Q_v = a(c - v) > 0$. Whilst it may seem counter intuitive for the system to be taking up additional v whilst trying to minimise the flux out of v via q_2, it makes sense that a positive Q_v corresponds to a low concentration of v, which would then lead to a lower value for $q_2 = dv$.

Method	Minimum/maximum	q_1	q_2	q_3	Q_u	Q_v
FBA	Minimum	0	0.0044	0.001	0.001	0.0034
FBA	Maximum	0.26	0.0461	0	0.26	-0.2139
Lagrange multipliers	Minimum	0.0268	0.0087	0.0009	0.0277	-0.0191
Lagrange multipliers	Maximum	0.1193	0.0235	0.0005	0.1199	-0.0964

Table 5.5: Flux vectors that minimise or maximise q_2 obtained by either Lagrange multipliers or flux balance analysis. The Lagrange multipliers results are obtained by solving the system of equations (5.5.3)–(5.5.6). The FBA solutions are obtained from the linear programming problem equation (5.5.7) with the additional constraints from equations (5.5.9) and (5.5.11).

5.6 Discussion

The major difference between the FBA and the Lagrange multiplier solutions is the values for q_1 and Q_u. FBA allows a maximal input value for Q_u, whilst still allowing large fluxes out via v_1 and v_3. In our ODE system, this would not be permitted, as maximal input values for Q_u correspond to a zero concentration of u, which in turn would lead to zero flux out of u (and a non-steady-state). FBA does not have this limitation, even with our additional constraints, and this allows it to record much higher flux through the system than is allowed by the full ODE system seen in equations (5.2.6)–(5.2.7). Similarly in the minimum case, FBA allows a low input which would correspond to a high u and therefore higher q_1 and q_3 in our ODE system. Without an additional constraint linking these two fluxes to have some kind of inverse proportional relationship the system will never be in full agreement with the ODEs. A suitable constraint for this would have to be non linear in nature, for example $Q_u q_1 = w$ for some estimated w, and would therefore take away from the linearity of FBA. It would also require even more knowledge about the system than we may not have suitable access to. It would also add a fifth constraint to the system, which only has five unknowns, and the system would become fully determined, removing the need for optimisation methods. Indeed if we had all the information to form these constraints in the first place, it would be more instructive to create a full ODE model to determine the kinetics rather than using FBA to examine the potential steady-states.

From this, we see that whilst we are able to determine the possibility of multiple steady-states just by examining the structure of a metabolic network, it is unlikely that we will be able to examine and analyse when these multiple steady-states arise simply from this structure. In particular, it is not possible to produce an example of these multiple steady-states with only this structural data and flux balance analysis. Even if we add additional linear constraints by assuming extra knowledge about the dynamics of the system and values for the parameters that govern it, we are still unable to accurately replicate the steady-states from an ODE system. We cannot identify a discontinuity in our FBA solution that would typically be seen when looking at multiple steady-states. It may be possible to add additional non-linear constraints to find the behaviour we seek, but if possible we wish to avoid that in order to keep the computation simplicity of FBA as a linear programming technique. Even if we did add non linear constraints, we may find ourselves requiring more information than we have access to in order to observe the discontinuity in the solution. We observe multiple steady-states through optimisation methods such as Lagrange multipliers which requires a full ODE model of the system along with values for parameters; however, it seems that more simplistic methods are not able to recover the same results.

We have observed systems that exhibit multiple steady-states in a number of different bifurcation structures such as isolas and mushrooms. We have also seen that an organism's metabolism can have multiple steady-states, depending on external conditions and other stimuli [50]. Finally we also observed metabolic networks exhibiting non-steady-state long timescale behaviour, such as periodic solutions, seen in Figure 5.3. We also saw, in Chapter 3, our *C. autoethanogenum* network showing states in which concentrations grew indefinitely in time. The FBA model we have used up until now has is inadequate for modelling networks displaying these behaviours, as it relies on mass balance to provide solutions in steady-state. Instead we now look to an extension of FBA, dynamic flux balance analysis, which is able to model the dynamic behaviour of networks. By applying this to previously examined networks, we will consider whether it can predict these non-steady-states seen until now, as well as the multiple steady-state behaviours that FBA was unable to predict in Section 5.5.2.

Chapter 6

Dynamic flux balance analysis

6.1 Introduction

A key limitation of flux balance analysis is that it only models networks that are in steady-state, specifically when concentrations do not change in time, i.e. when $S\mathbf{v} = 0$, where S is the stoichiometric matrix, and $\mathbf{v}$ is a vector of fluxes. As we have seen with our *C. autoethanogenum* model however, not all metabolic networks tend to a steady-state. Sometimes they exhibit dynamic behaviour, such as the linear growth we have seen in Chapter 3, and even periodic solutions which were also seen for the autocatalysis model discussed in Section 5.2. All our models until now have also focused on continuous models, in which there is a continuous input into the system, whether by assuming a constant uptake rate, or by having a constant external concentration. In the case of a batch model, in which there is a limited amount of substrate for the model to uptake in the first place, much of the relevant behaviour will be transient behaviour. In order to observe this behaviour, we need methods which model the dynamics of these systems outside of steady-state. We have already used ODE systems to describe this kind of non-steady-state behaviour, but we would like to extend flux balance analysis to model dynamic behaviour. In extending FBA to describe dynamic behaviours, we use many of its strengths, such as not needing to estimate parameters, whilst removing one of its main weaknesses, the requirement for steady-state. This extension, as initially set out by Mahadevan et al. [43], is known as dynamic flux balance analysis (dFBA).

6.1.1 Forming a dFBA model

In FBA, the key constraint is $S\mathbf{v} = \mathbf{0}$, which forces the system to be in a steady-state. If we wish to model a network outside of steady-state, we must change this constraint. We change it to

$$\frac{d\mathbf{c}}{dt} = S\mathbf{v}, \tag{6.1.1}$$

where $\mathbf{c}$ is a vector of the concentrations of the metabolites in the network. It is worth noting now that whilst we write this derivative as a constraint, when we later come to solve this model, this derivative is instead considered as a numerical approximation, with the previous time step known, so that this instead becomes a constraint on the concentrations, $\mathbf{c}$, based on the values of $\mathbf{v}$, and $\mathbf{c}$ at the previous time step. We now include these concentrations as variables in our optimisation problem. We now have an ODE system for concentrations ($\mathbf{c}$) that depends on the fluxes ($\mathbf{v}$). These fluxes do not explicitly depend on the concentrations, however, so it is not like the full ODE models seen previously. When it comes to simulating the dFBA model, we find that

we are still able to use linear programming, but first we must create some additional constraints to account for the addition of the concentration variables, and the fact that both concentrations and fluxes vary in time.

Firstly, we add some physical constraints to the system. In addition to our usual FBA constraints providing maximum and minimum flux values, we should also consider adding equivalent constraints for the new concentration variables. Obviously, we must not have concentrations going below zero, since this is physically impossible, so we add the constraint

$$\mathbf{c} \geq \mathbf{0}. \tag{6.1.2}$$

We can also add maximum values for these concentrations, however it is unlikely that these are necessary. Since we have previously seen unbounded growth in an ODE model for the concentrations in our *C. autoethanogenum model*, it may even be beneficial to not include maxima, to allow for states when concentrations grow linearly in time.

If we solve the optimization problem, $\underset{\mathbf{v}}{max}\ \mathbf{a}^T\mathbf{v}$) as in FBA, with only constraints (6.1.1) and (6.1.2) and the usual FBA constraints for any of our previous networks, we do not find useful results. The system immediately jumps to a state in which the objective function is maximised. As long as this state doesn't cause a concentration to drop below zero, then the system remains in this state indefinitely. This does not tell us anything about the transient behaviour we wish to investigate, instead only showing us the long term behaviour we have already investigated. Whilst we are now able to see linear growth states, we still do not observe the transient behaviour with only these constraints. The problem is that, whilst concentrations depend on the fluxes, the fluxes have no dependency on the concentrations, other than not allowing elements of $\mathbf{c}$ to drop below zero. In order for the concentrations to be relevant, we add new concentration-based constraints to the model.

Whilst with FBA we attempted to avoid requiring the knowledge of any parameters at all, for dFBA, we require some in order to observe transient behaviour. In general, the constraints we add to the model are described by

$$v_n \leq g(\mathbf{c}), \tag{6.1.3}$$

where v_n is some flux, and $g(\mathbf{c})$ is some function dependent on the concentrations. This function is often non-linear. This $g(\mathbf{c})$ takes the form of some function describing the rate of flux in terms of the concentrations. For example, in a metabolic network, this could be Michaelis-Menten kinetics, or first order kinetics, as seen in our ODE models examined in Chapter 3. We add as many or as few of these constraints as we have information for. Generally, since we are using these flux balance methods to attempt to avoid having to obtain detailed kinetics for flux rates, we attempt to use as few of these concentration based constraints as possible. We also note that whilst this has been described as an inequality, it is also fine to use equality constraints here, which in some cases may be more useful.

Now that we have some fluxes depending on concentrations we will be able to see transient behaviour, as some of our fluxes now require concentrations to grow before they also grow. Often we see this happening much faster than we would see in a more detailed model and faster than we would expect in a real organism. We can choose to add an additional constraint on the rate of change of fluxes to slow down the overall system. Whilst this is not necessary, it may prove useful for some networks, so we will use it in all the models described in this chapter. A maximum rate of change for the fluxes is described as

$$-\mathbf{v}_{dt-max} \leq \frac{d\mathbf{v}}{dt} \leq \mathbf{v}_{dt-max}, \tag{6.1.4}$$

where $\mathbf{v}_{dt-max}$ is a vector of maximum values for the rate of change of flux. Effectively, this constraint limits the steepness of the flux changes. This slows the evolution of the fluxes, helping to account for some of the fluxes which may not have concentration based constraints.

Finally, since the concentrations are now variables, we are able to include them in the objective function. This means we generalise an objective function that is simply a linear combination of fluxes, $\underset{\mathbf{v}}{max/min} \quad \mathbf{a}^T\mathbf{v}$, to one that is a combination of both fluxes and concentrations, given by

$$\underset{\mathbf{c},\mathbf{v}}{max/min} \quad \mathbf{a}^T\mathbf{v} + \mathbf{b}^T\mathbf{c} \tag{6.1.5}$$

By combining all our constraints from equation (6.1.1) to (6.1.4), along with our FBA maximum and minimum flux values, our new objective function, and some initial conditions for both $\mathbf{c}$ and $\mathbf{v}$, we have a full dFBA model,

$$\begin{aligned} \underset{\mathbf{c},\mathbf{v}}{max/min} \quad \mathbf{a}^T\mathbf{v} + \mathbf{b}^T\mathbf{c} \quad \text{such that: } \frac{d\mathbf{c}}{dt} &= S\mathbf{v}, \\ \mathbf{v}_{min} \leq \mathbf{v} &\leq \mathbf{v}_{max}, \\ \mathbf{0} \leq \mathbf{c}&, \\ v_n \leq g(\mathbf{c})&, \\ -\mathbf{v}_{dt-max} \leq \frac{d\mathbf{v}}{dt} &\leq \mathbf{v}_{dt-max}, \\ \mathbf{v}(0) = \mathbf{v}_0, \quad \mathbf{c}(0) &= \mathbf{c}_0. \end{aligned} \tag{6.1.6}$$

Unlike for regular FBA, there is no way to know if a dFBA problem will have a well defined solution prior to attempting to solve it. This is especially problematic for large systems that take a long time to solve, as it is possible for the solver to fail after a large amount of time has already been invested in finding a solution. This must be kept in mind when forming dFBA models.

6.1.2 Growth rate

It is possible, and often useful, to consider the growth rate in a dFBA model. Some fluxes in the network may lead to the production of biomass, which in turn leads to growth. To do this, we introduce an additional variable, X, which describes the overall biomass concentration. The flux vector $\mathbf{v}$ specifically becomes the flux per unit of biomass. The rate of growth is given by the equation

$$\frac{dX}{dt} = \mu X, \tag{6.1.7}$$

where μ is some weighted some of the fluxes v_n that produce biomass, $\sum_n w_n v_n$, where the weight w_n describes how many units of biomass are produced per unit flux through the reaction v_n. We must also modify the ODE constraint slightly, becoming

$$\frac{d\mathbf{c}}{dt} = S\mathbf{v}X, \tag{6.1.8}$$

so that the rate of change of concentrations is now proportional to the overall biomass concentration.

Since we may wish to keep track of growth rate, and it may prove a useful choice as an objective function it is useful to add this extra variable and constraint to the dFBA model. In previous models, however, we have neglected to consider growth rate in this way, so in order to keep our dFBA models directly comparable with previous chapters, we do not consider the growth rate and overall biomass concentration in our models in this chapter.

6.1.3 Solving the model

Now that we have a full dFBA model, we must now consider how to solve it. Whilst the model (6.1.6) is presented as an optimization problem with an ODE constraint, when it comes to solving the model, it is more useful to think of it as the set of ODEs given by equation (6.1.1), where the values of the fluxes, $\mathbf{v}$, are determined by the optimization problem (6.1.6) without the ODE constraint. Then we approximate the ODEs, $\dot{\mathbf{c}} = S\mathbf{v}$, numerically, and at each time point solve the optimization problem to generate a new $\mathbf{v}$ which then determines the concentrations. The usual method to do this in dFBA is known as the static optimization approach. We discretise time into a number of time points, the difference between them being Δt, and for each time step solve the optimization problem

$$\begin{aligned} \underset{\mathbf{C},\mathbf{v}}{max/min} \quad \mathbf{a}^T\mathbf{v}_{t+\Delta t} + \mathbf{b}^T\mathbf{c}_{t+\Delta t} \quad \text{such that: } \mathbf{c}_{t+\Delta t} &= \mathbf{c}_t + \Delta t S\mathbf{v}_{t+\Delta t}, \\ \mathbf{v}_{min} \leq \mathbf{v}_{t+\Delta t} &\leq \mathbf{v}_{max}, \\ \mathbf{0} \leq \mathbf{c}_{t+\Delta t}&, \\ v_{t+\Delta t,n} \leq g(\mathbf{c}_t)&, \\ -\mathbf{v}_{dt-max} \leq \frac{\mathbf{v}_{t+\Delta t} - \mathbf{v}_t}{\Delta t} &\leq \mathbf{v}_{dt-max}. \end{aligned} \tag{6.1.9}$$

where $\mathbf{v}_t$ and $\mathbf{c}_t$ are the flux and concentration vectors at time t respectively. It must be noted that the concentration based coefficients have the flux at time $t + \Delta t$ depending on the concentrations at time t, i.e. from the previous time point. This is done because when the function $g(\mathbf{c})$ is nonlinear in $\mathbf{c}$, we will be unable to use linear programming to solve the model if we use $g(\mathbf{c}_{t+\Delta t})$. By using $g(\mathbf{c}_t)$, we are able to simply calculate the value of this function from the previous time point, allowing us to maintain linearity in our constraints (as the constraint then becomes $v_t \leq$ some constant calculated from the previous time steps results). This induces a slight lag in the maximum values of some fluxes, so we must choose a small Δt in order for this to have minimal effect on the system. The first time point is given as some initial condition, often one in which fluxes and concentrations are zero, however we could choose any initial condition that did not violate any of the known constraints.

Now that we are able to solve dFBA models, we model some metabolic networks using this technique.

6.2 Straight line system

Rather than jumping straight in to a full metabolic network, it is useful to first see how dFBA behaves on a simpler network. We start by examining a simple straight line network in which a metabolite is taken up, transformed into something else via some chemical reactions, and then outputted. Consider a network of three metabolites, A, B, and C. These metabolites are linked by a reaction that transforms A into B, and a reaction that transforms B into C. We also have an uptake of A and an output of C. In this way, there is a flow through the system into A, through B, and finally out of C. This network is summarised as

$$\xrightarrow{Q_A} A \xrightarrow{q_1} B \xrightarrow{q_2} C \xrightarrow{Q_C}, \tag{6.2.1}$$

We form the stoichiometric matrix for this network

$$S = \begin{pmatrix} -1 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}. \tag{6.2.2}$$

Intuitively, we see that a steady-state solution for this network is one in which $q_1 = q_2 = Q_A = Q_C$. If we performed regular FBA on this network, all our solutions would take the form $\mathbf{v}_{ss} = \lambda(1, 1, 1, 1)$, where λ is some scalar, and the elements of the vector represent one on the fluxes in the network. As it is a simple straight line network, we would not expect its long term behaviour to be non-steady, and therefore, we would expect our dFBA model to provide a solution that tends to a steady-state of this form.

6.2.1 The dFBA model

We now formulate a dFBA model to describe the dynamic behaviour of the network (6.2.1). Since we have three metabolites, we need to add three variables to the model to represent these concentrations: C_A, C_B, and C_C, then our vector of concentrations, $\mathbf{C} = (C_A, C_B, C_C)$. Combined with our vector of fluxes, $\mathbf{v} = (q_1, q_2, Q_A, Q_C)^T$, and our stoichiometric matrix from equation (6.2.2), we have the ODE constraint,

$$\frac{d\mathbf{C}}{dt} = S\mathbf{v}. \tag{6.2.3}$$

For the network (6.2.1), we also impose constraints for the uptake flux, Q_A, and the output flux, Q_C, that depend on the concentrations. In this case, we limit the maximum value of the output flux, Q_C, to be proportional to the concentration of C, and the maximum value of the input to be proportional to the difference between the internal concentration of A, and some external concentration that remains constant. In this way we create two additional constraints:

$$Q_A \leq k_1(C_A^* - C_A), \tag{6.2.4}$$
$$Q_C \leq k_2 C_C, \tag{6.2.5}$$

where k_1 and k_2 are rate coefficients and C_A^* is the constant external concentration of A. In addition to these concentration-dependent constraints, we also set minimum and maximum values for fluxes as in regular FBA. In this case, all fluxes except Q_A take a minimum value of zero, signifying that they are irreversible. We do not impose a minimum value for Q_A, as we see from equation (6.2.4), high internal concentrations of A causes Q_A to be negative. For this example, we see that the overall flux through the system is limited by the external concentration of A, C_A^*, so we do not need to set maximum values for any fluxes. We also limit concentrations to be greater than zero as usual. The final constraint for this model is the rate of change constraint on fluxes,

$$-\mathbf{v}_{dt-max} \leq \frac{d\mathbf{v}}{dt} \leq \mathbf{v}_{dt-max}. \tag{6.2.6}$$

for some vector $\dot{\mathbf{v}}_{max}$. Then by combining all these constraints along with some objective function that can depend on both fluxes and concentrations, we form the full dFBA model for this network,

$$\begin{aligned} \max_{\mathbf{C},\mathbf{v}} \quad \mathbf{a}^T\mathbf{v} + \mathbf{b}^T\mathbf{C} \quad \text{such that: } & \frac{d\mathbf{C}}{dt} = S\mathbf{v}, \\ & Q_A \leq k_1(C_A^* - C_A), \\ & Q_C \leq k_2 C_C, \\ & \mathbf{C} \geq \mathbf{0}, \\ & q_1, q_2, Q_C \geq \mathbf{0}, \\ & -\mathbf{v}_{dt-max} \leq \frac{d\mathbf{v}}{dt} \leq \mathbf{v}_{dt-max}, \end{aligned}$$

$$\mathbf{v}(0) = \mathbf{v}_0, \quad \mathbf{C}(0) = \mathbf{C}_0. \tag{6.2.7}$$

Then all that must be done to run a simulation is choose values for the parameters k_1, k_2, and C_A^*, an objective function, a maximum rate of change vector for our fluxes, and an initial condition. For now, we set the output flux Q_C to be our objective function by choosing $\mathbf{b} = \mathbf{0}$, and $\mathbf{a} = (0, 0, 0, 1)$. We set all the parameters k_1, k_2, and C_A^* equal to one for simplicity, and we also take all elements in our maximum rate of change vector to be equal to one. For our initial condition we start with all concentrations and fluxes set to equal zero. Then we simulate our model (6.2.7). The results of this simulation are shown in Figure 6.1. We see that very rapidly q_1, q_2, and Q_A all reach maximum values of one. From here the concentration C_c increases steadily, allowing the output flux, Q_C to increase too, also increasing towards a maximum value of one. We note that the concentrations of A and B remain at zero throughout the simulation. In particular C_A remains at zero, allowing the input flux Q_A to reach its maximum allowed value according to the constraint (6.2.4), which takes its maximum value when C_A is equal to zero.

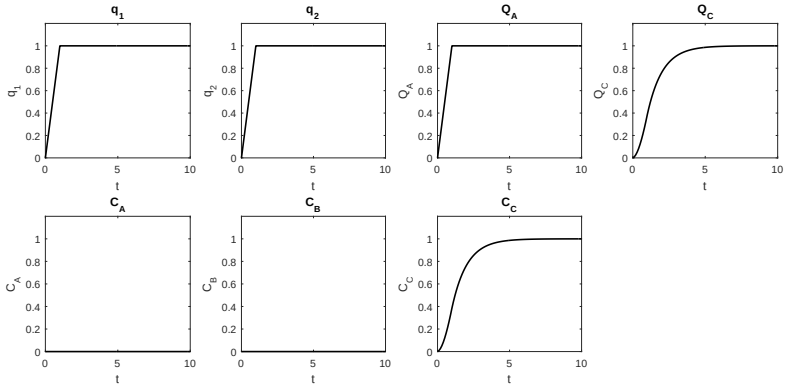

Figure 6.1: Plot of fluxes and concentrations against time in the simulation of equation (6.2.7) with all parameters and maximum flux rate of changes equal to one, and the objective function max Q_C. We have initial conditions $C_A(0) = c_B(0) = C_C(0) = 0$. The first row shows plots for the fluxes v_n, whilst the second row shows those of the concentrations, C_n.

We now modify some parts of our model to see whether it continues behaving as we expect. First we try different initial conditions. For maximum flux through the system, we require C_A to equal zero, as previously mentioned. We start with a non-zero initial condition for C_A. We would expect this to decay to zero to allow as much input as possible. Figure 6.2 shows this simulation, with the initial concentration of A, $C_A = 0.5$. As expected, this concentration decays to zero, with the system attaining the same steady-state as before. We see that q_1 and q_2 both initially increase to values above one, before dropping back down below one, finally settling exactly on one as the network reaches steady-state. The initial increase of Q_A is slower than the previous simulation, as it is limited by the initial concentration. Despite this change in the initial condition, the network still tends to the same steady-state, as expected. Raising the initial condition of C_A even higher produces the same effect but more pronounced, in some cases causing the output Q_C to increase past one, before decreasing back down to the same long term behaviour.

We next consider starting with an initial concentration of C. Figure 6.3 shows a simulation with an initial concentration of C, $C_C = 0.5$. We see that output flux Q_C

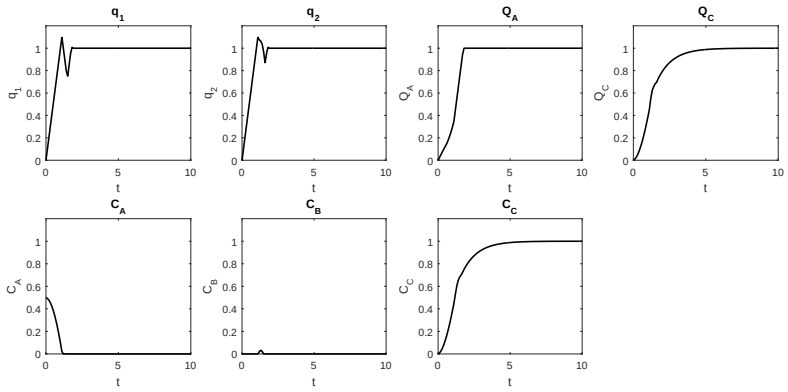

Figure 6.2: As in Figure 6.1, except we have initial conditions $\mathbf{v}(0) = \mathbf{0}$, $C_A(0) = 0.5$, $C_B(0) = 0$, and $C_C(0) = 0$. The first row shows the graphs for the fluxes, whilst the second row shows those of the concentrations.

starts increasing rapidly, but after reaching the point where it is equal to the decreasing concentration of C, it also drops slightly, until the value of q_2 becomes high enough to allow C_C to increase again. The remaining fluxes, q_1, q_2, and Q_A also do not increase smoothly or constantly as they did in previous simulations. There is also a minor build up of A early on, which is quickly removed. In the long term, the network still tends to the same steady-state.

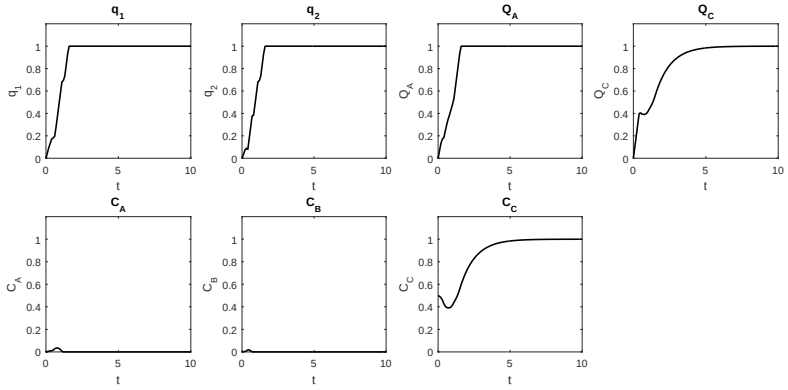

Figure 6.3: As in Figure 6.1, except we have initial conditions $\mathbf{v}(0) = \mathbf{0}$, $C_A(0) = 0$, $C_B(0) = 0$, and $C_C(0) = 0.5$. The first row shows the graphs for the fluxes, whilst the second row shows those of the concentrations.

Finally, let us consider the effect of modifying the rate of change constraints on the fluxes. Decreasing the values of $\mathbf{v}_{dt-max}$ uniformly causes the system to behave in the same way, simply over a longer timescale, as seen in Figure 6.4. Interestingly, the time taken does not increase exactly in line with the maximum rate of change decrease. In the simulation shown in Figure 6.4, the maximal rate of change is set to equal 0.1, which is one tenth of the previous value, however the output flux Q_C only takes twice as long to increase to its maximum state. The remaining fluxes, q_1, q_2, and Q_A do, however, take ten times as long to reach their steady-state values.

Rather than changing the maximum rate of change for all fluxes, we now instead consider different maximum rates of change for each flux. We might expect the overall

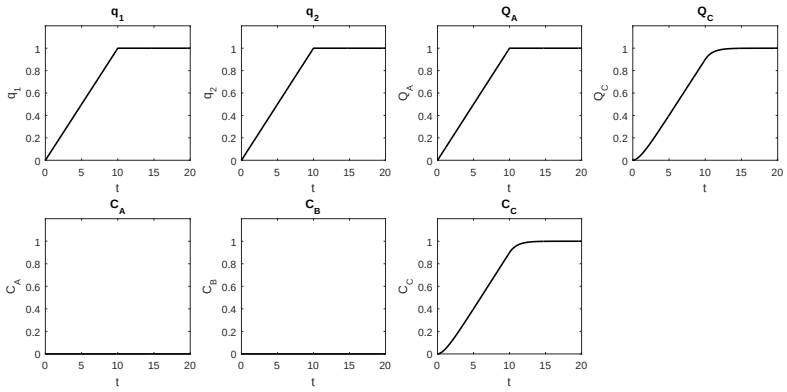

Figure 6.4: Plot of fluxes and concentrations against time in the simulation of equation (6.2.7) with all parameters equal to one, and the objective function max Q_C. The initial condition has $\mathbf{v}(0) = \mathbf{C}(0) = \mathbf{0}$. The maximum rates of change are $\mathbf{v}_{dt-max} = (0.1, 0.1, 0.1, 0.1)$. The first row shows the graphs for the fluxes, whilst the second row shows those of the concentrations.

behaviour of the system to be limited by the slowest changing flux, but the remaining fluxes could behave in different ways, depending on which flux is the "slow" flux. By setting a single fluxes maximum rate of change to 0.1 whilst keeping the rest at one, we consider if different slow fluxes have different effects on the overall behaviour. Figure 6.5 shows a simulation with q_1 as the slow flux. We see the input Q_A tries to increase as quickly as possible, but since q_1 increases slower, this leads to a build up of the concentration C_A, which limits the uptake rate. As q_1 increases, the system is able to remove the excess concentration of A, allowing Q_A to increase again. The remaining fluxes q_1, q_2, and Q_C behave in the same way as they did for the simulation when all fluxes were "slow". If we set q_2 to be our slow flux, we see a similar behaviour, shown in Figure 6.6. Both Q_A and q_1 increase quickly early on, but with the slow flux not able to match them, this causes an increase in concentrations of both A and B, which eventually decay to zero as the system tends back to the same steady-state.

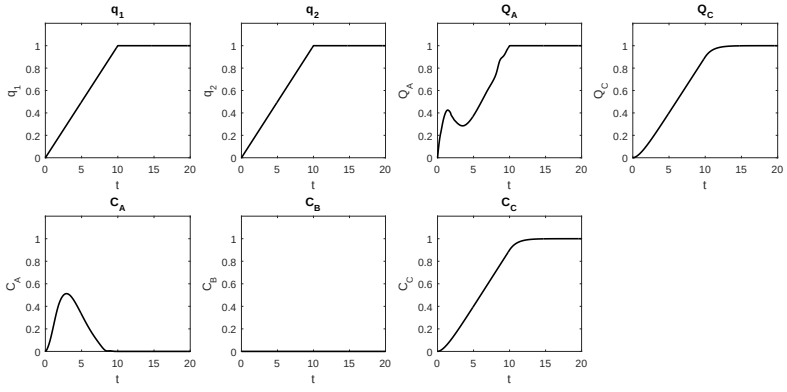

Figure 6.5: As in Figure 6.4, except the maximum rates of change are given by $\mathbf{v}_{dt-max} = (0.1, 1, 1, 1)$.

Throughout these simulations, the concentration of B has been largely inconsequential, only showing temporary increases before decaying back down to the zero

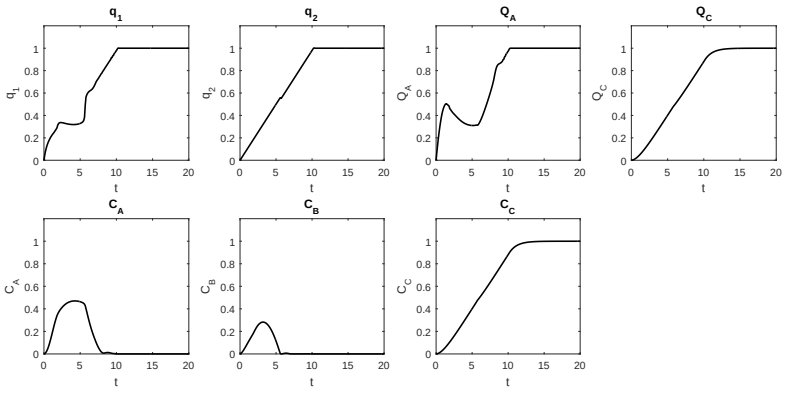

Figure 6.6: As in Figure 6.4, except the maximum rates of change are given by $\mathbf{v}_{dt-max} = (1, 0.1, 1, 1)$.

steady-state. In general, we have also seen the fluxes in and out of B, q_1 and q_2 respectively, being equal. This metabolite has almost always been in steady-state, with the flux in equal to the flux out. Since there are no concentration based constraints that rely on C_B, it could take any value at all with no consequence on the long term behaviour. If we extend this chain further, and include another intermediary metabolite between B and C (or between A and B), this additional metabolite exhibits the same behaviour, with its concentration being inconsequential. This suggests that for dFBA, when reactions are sequential, intermediates can be neglected unless there are constraints that are based on their concentrations. In this case rather than having $A \longrightarrow B \longrightarrow C$, we could simplify the network to $A \longrightarrow C$. This simplifies our model, including two fewer variables, with very little loss of detail.

6.3 An autocatalytic system

In the previous chapter we examined the possibility of using flux balance analysis to observe the existence of multiple steady-states. Whilst we were not able to do this with regular FBA methods, dFBA may be able to show the existence of multiple steady-states. To examine this possibility, we return to examining the autocatalysis network described in Section 5.2. This network has been shown to exhibit multiple steady-states under certain parameter conditions.

To reintroduce the network, we have two metabolites: X, the primary metabolite, and Y, the catalyst. The uptake rates for these metabolites are Q_x, and Q_y. We have two reactions that transform X into Y, the first a reaction using Y as a catalyst in which one X and two Y are used to create three Y, and an uncatalysed reaction in which X is changed into Y. Finally we also have an output reaction for Y, representing the decay of the catalyst. These reactions are summarised in Chapter 5, in Table 5.1.

As in previous examples, we begin creating our dFBA model by formulating the stoichiometric matrix for this network. This network has five reactions for two metabolites, resulting in the 2×5 stoichiometric matrix,

$$S = \begin{pmatrix} -1 & 0 & -1 & 1 & 0 \\ 1 & -1 & 1 & 0 & 1 \end{pmatrix}. \tag{6.3.1}$$

Together with the vector of concentrations, $\mathbf{c} = (x, y)$, where x and y are the concentrations of X and Y respectively, and the vector of fluxes,

$\mathbf{v} = (v_1, v_2, v_3, Q_x, Q_y)$ this forms the ODE constraint,

$$\frac{d\mathbf{c}}{dt} = S\mathbf{v}. \tag{6.3.2}$$

Next we consider constraints based on concentrations. We use the dimensionless form of the ODE model for this network, presented in equations (5.2.6)–(5.2.7), as a basis for these constraints. First our uptake fluxes, Q_x and Q_y are given in the ODE model as $Q_x = a(1-x)$ and $Q_y = a(c-y)$. Normally for dFBA, we would set these constraints to be inequalities, so that $Q_x \leq a(1-x)$ and $Q_y \leq a(c-y)$, however, in this case we are looking for evidence of multiple steady-states. Since we have previously seen a to be an important parameter when determining if multiple steady-states exist, we keep these as equality constraints to allow us more control over how the dFBA model behaves. We also add two additional constraints that depend on concentrations, in the form of $v_2 \leq dy$ and $v_3 \leq bx$. These parameters, b and d, are less important for determining the existence of multiple steady-states, so we allow these constraints to be inequality constraints as usual. This gives us the four constraints

$$Q_x = a(1-x), \tag{6.3.3}$$
$$Q_y = a(c-y), \tag{6.3.4}$$
$$v_2 \leq dy, \tag{6.3.5}$$
$$v_3 \leq bx. \tag{6.3.6}$$

Next we add maximum and minimum values on our fluxes and concentrations. As usual, all concentrations must be greater than or equal to zero, that is, $\mathbf{c} \geq \mathbf{0}$. We also have three fluxes which are irreversible, v_1, v_2, and v_3, which must also have a minimum value of zero. The uptake fluxes Q_x and Q_y are reversible, and indeed the reversibility of Q_y is known to be important for the multiple steady-state behaviour. The overall flux through the system is constrained by the external concentrations, allowing a maximum value for Q_x of a, and a maximum value for Q_y of ac. We choose maximum values for all fluxes and concentrations to be 100, though this is likely to prove unnecessarily high.

Finally we add in our usual maximum rate of change for flux values,

$$-\mathbf{v}_{dt-max} \leq \frac{d\mathbf{v}}{dt} \leq \mathbf{v}_{dt-max},$$

and now formulate the entire model as

$$\begin{aligned} \underset{\mathbf{c},\mathbf{v}}{max/min} \quad \mathbf{x}^T\mathbf{c} + \mathbf{y}^T\mathbf{v} \quad such\ that: \frac{d\mathbf{c}}{dt} &= S\mathbf{v}, \\ Q_x &= a(1-x), \\ Q_y &= a(c-y), \\ v_2 &\leq dy, \\ v_3 &\leq bx, \\ \mathbf{0} &\leq \mathbf{c}, \\ 0 &\leq v_1, v_2, v_3, \\ \mathbf{v}, \mathbf{c} &\leq 100, \\ -\mathbf{v}_{dt-max} \leq \frac{d\mathbf{v}}{dt} &\leq \mathbf{v}_{dt-max}, \\ \mathbf{v}(0) = \mathbf{v}_0, \quad \mathbf{c}(0) &= \mathbf{c}_0. \end{aligned} \tag{6.3.7}$$

It is important to note that if we wish to observe multiple steady-states, we must pick a set of parameters, (a, b, c, d), that permit their existence. From our previous analysis of the ODE model for this network in Section 5.3, we know that if we choose $b = 0.001$, $c = 0.01$, and $d = 0.05$, then for values of a within the bounds $0.2524 < a < 0.2694$, there are three possible steady-states, two stable and one unstable. We choose a value for a which is safely within that range, $a = 0.26$. Whilst it may not be possible to find the unstable state, we expect to be able to find some equivalent to the two stable steady-states.

In order to observe multiple steady-states, we set our parameters to those described in the previous paragraph and perform dFBA from two different initial conditions. First from a starting point with low concentrations and once from a state with higher starting concentrations. For our low initial conditions, we choose $x = 0.1$ and $y = 0.1$. We also set our initial fluxes according to the flux rates described in the dimensionless ODE system, so that we have

$$Q_x = a(1 - x) = 0.234, \tag{6.3.8}$$
$$Q_y = a(c - y) = 0, \tag{6.3.9}$$
$$v_1 = xy = 0.01, \tag{6.3.10}$$
$$v_2 = dy = 0.005, \tag{6.3.11}$$
$$v_3 = bx = 0.0001, \tag{6.3.12}$$

For our higher starting concentration, we choose $x = 0.65$ and $y = 0.38$. Then our initial flux state for this higher condition is

$$Q_x = 0.091, \tag{6.3.13}$$
$$Q_y = -0.0728, \tag{6.3.14}$$
$$v_1 = 0.247, \tag{6.3.15}$$
$$v_2 = 0.019, \tag{6.3.16}$$
$$v_3 = 0.00065. \tag{6.3.17}$$

We run the simulation with the objective function being maximising the rate of catalyst decay, that is maximising the flux v_2. When we previously examined the system, we found that one key difference between the three states was the level of catalyst decay, so it makes sense to use this flux as our objective function. If dFBA is able to predict the existence of multiple steady-states, we would expect them to appear even with the same objective function. That is, despite trying to maximise v_2, we would expect under some initial condition for the model to provide the low v_2 state. Similarly, we expect a higher v_2 state from some other initial condition, with the same objective function. Finally, we select a relatively small maximum rate of change of flux of 0.1.

The results of the two simulations are shown in Figures 6.7 and 6.8. We see that from the same model, we are able to achieve two qualitatively different results depending on the initial conditions. When we start in the low initial condition state, we tend towards a low y state, in which the concentration of X is near one, and the concentration of Y is near 0.1. However, when we start from the higher initial condition, the system tends towards a high y state, with the concentration of X near 0.25, and the concentration of Y near 0.7.

Here, dFBA has been able to predict the existence of multiple steady-states for a network. These states do not align exactly with the states provided by the ODE model for the same network. Table 6.1 shows the steady-state concentrations of x and y from both high and low y states from the ODE model as well as those found from dFBA.

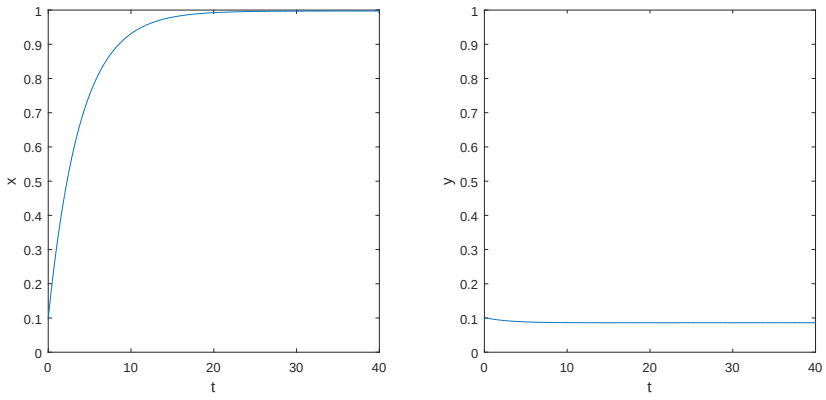

Figure 6.7: Plots of the concentrations of x and y against time in a simulation of the dFBA model described in equation (6.3.7) with $a = 0.26$, $b = 0.001$, $c = 0.1$, and $d = 0.05$, and a maximum rate of change of flux of 0.1. The initial condition for this simulation are $x = 0.1$, $y = 0.1$.

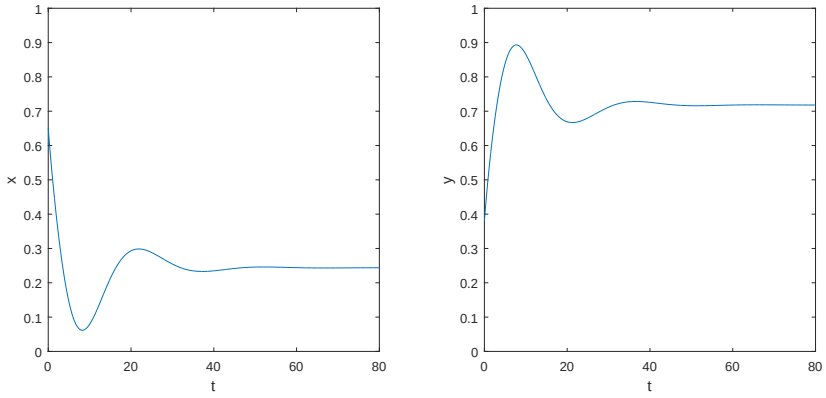

Figure 6.8: Plots of the concentrations of x and y against time in a simulation of the dFBA model described in equation (6.3.7) with $a = 0.26$, $b = 0.001$, $c = 0.1$, and $d = 0.05$, and a maximum rate of change of flux of 0.1. The initial condition for this simulation are $x = 0.65$, $y = 0.38$.

We notice that the low y states are in closer agreement than the high y states. This may be due to our objective function trying to maximise y. It is possible that if we choose to maximise x, we would obtain a high y state that is closer to the one seen in the ODE model. This proves to not be the case, however. By choosing our objective function to maximise x, we lose the high y state entirely and the system again tends to the low y state, regardless of initial conditions, as seen in Figure 6.9. For one objective function we were able to observe the existence of multiple steady-states using dFBA, however for another, this was not possible.

It is worth noting that the steady-state concentrations for the low y state are not the same across the two objective functions. The steady-state concentrations for this state with the objective function maximising x are $x = 1$, $y = 0.0907$. Even though the two steady-states are qualitatively similar, there is some variation between the two that we wouldn't expect to see in an ODE model. Since the difference between the steady-states for the two objective functions is so small, and the difference between the

State	Low y		High y	
Method	dFBA	ODE	dFBA	ODE
x	0.9973	0.8933	0.2437	0.5389
y	0.0861	0.1733	0.7182	0.4705

Table 6.1: Steady-state concentrations of x and y in the high and low y states, predicted by dFBA and ODE models.

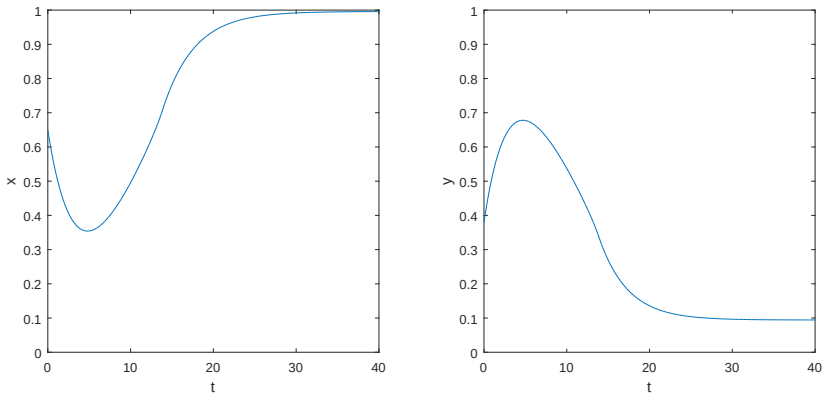

Figure 6.9: Plots of the concentrations of x and y against time in a simulation of the dFBA model described in equation (6.3.7) with $a = 0.26$, $b = 0.001$, $c = 0.1$, and $d = 0.05$, and a maximum rate of change of flux of 0.1. The initial condition for this simulation are $x = 0.65$, $y = 0.38$. In this case, the objective function is maximising the concentration of X, rather than that of Y.

high and low y states is so big, it does not matter too much in this situation. If we were to examine other systems, these slight variations could have much larger effects on the overall behaviour of the system, so it is worth examining these slight variations more closely, to see if there is some way to avoid them if possible, or if this is just a consequence of a different objective function.

6.3.1 The effects of multiple optimal solutions

To closer examine these slight variations, we modify the model slightly. We keep the network as described in Table 5.1, so the stoichiometric matrix remains as in equations (6.3.1). We also keep the parameters and concentration-based constraints the same. The slight variation we introduce is changing the maximum values of the flux which previously were all set to 100. We now increase these maxima, so that all elements of $\mathbf{v}_{max}$ are equal to 1000. Theoretically, this should not cause any difference to the behaviour of the system, as the fluxes are not able to attain such high values; however, we find that increasing the maximum values of the fluxes leads to a change in the steady-state concentrations, as shown in Table 6.2. This steady-state has a higher concentration of y than the base state, and therefore, a higher value of the objective function, v_2. This steady-state also persists across all initial conditions. To understand why two models which are fundamentally the same have two different behaviours we need to more closely examine what happens at each time step.

At each time point in the system, we solve a linear programming problem. In this

	Baseline	Objective function maximising Y	Maximum flux increase
x	0.9973	1.0000	0.7547
y	0.0861	0.0907	0.2896

Table 6.2: Steady-state concentrations of x and y for the model described in equation (6.3.7), using the initial condition $x = 0.1$, $y = 0.1$, with different variations in the model.

example, the problem at each time point is

$$\begin{aligned} \max_{\mathbf{c}_{t+\Delta t}, \mathbf{v}_{t+\Delta t}} v_{2,t+\Delta t} \quad & such\ that : \mathbf{c}_{t+\Delta t} = \mathbf{c}_t + \Delta t S \mathbf{v}_t, \\ & Q_{x,t+\Delta t} = 0.26(1 - x_t), \\ & Q_{y,t+\Delta t} = 0.26(0.1 - y_t), \\ & v_{2,t+\Delta t} \leq 0.05 y_t, \\ & v_{3,t+\Delta t} \leq 0.001 x_t, \\ & \mathbf{0} \leq \mathbf{c}_{t+\Delta t}, \\ & 0 \leq v_{1,t+\Delta t}, v_{2,t+\Delta t}, v_{3,t+\Delta t}, \\ & \mathbf{v}_{t+\Delta t}, \mathbf{c}_{t+\Delta t} \leq 1000, \\ & -\mathbf{v}_{dt-max} \leq |\frac{\mathbf{v}_{t+\Delta t} - \mathbf{v}_t}{\Delta t}| \leq \mathbf{v}_{dt-max}, \end{aligned} \tag{6.3.18}$$

where $\mathbf{c}_t$ is the vector of concentrations at time t, (x_t, y_t), $\mathbf{v}_t$ is the vector of fluxes at time t, $(v_{1,t}, v_{2,t}, v_{3,t}, Q_{x,t}, Q_{y,t})$ and Δt is the size of the time steps. Starting from an initial condition of $x = 0.5$, $y = 0.5$, we have an initial flux vector of $(v_1, v_2, v_3, Q_x, Q_y) = (0.25, 0.025, 0.0005, 0.13, -0.104)$. Taking $\Delta t = 1$ and all elements of $\mathbf{v}_{dt-max}$ to equal 0.1, we solve the optimisation problem for this single time step.

One solution to (6.3.18) is

$$\mathbf{c}_{t+\Delta t} = (0.48, 0.521),$$

and

$$\mathbf{v}_{t+\Delta t} = (0.15, 0.025, 0, 0.13, -0.104),$$

however this solution is not unique, there are multiple solutions that satisfy the constraints and maintain a maximised value for v_2. In Section 2.5.2, we introduced flux variability analysis, to characterise how much an individual flux in a FBA problem could vary whilst still maintaining an optimal solution. By applying the same techniques to this single dFBA time-step, we aim to characterise the variability of fluxes and concentrations at each time-step for this dFBA problem.

By adding the additional constraint $v_{2,t+\Delta t} = 0.025$, and using the objective function to find the solution that minimises or maximises each flux or concentration in turn, we find the variability ranges for the fluxes and concentrations. We find that whilst v_2, Q_x, and Q_y have no range of variability, v_1, v_3, and the concentrations x and y are all able to be varied whilst still maintaining the optimal value for v_2. Specifically, v_1 has the range of variability, $0.15 \leq v_1 \leq 0.35$, v_3 has the range $0 \leq 0.005$, the concentration x has the range $0.2895 \leq 0.48$, and the concentration y has the range $0.521 \leq y \leq 0.7215$.

This range of y is notable, because the concentration of y directly effects the maximum value of v_2 in the next time-step. If y takes its minimum value according to the

variability analysis, at the next time step, the maximum value of v_2 is 0.01303, whilst if y took its maximum value, the maximum value for v_2 would instead be 0.01804. This possible difference becomes more pronounced in subsequent time-steps, splitting the solutions into the high and low y steady-states seen in Figures 6.7 and 6.8. These results suggest that the multiple steady-state behaviour seen for this network, rather than being due to differing initial conditions, is a result of the variability allowed in v_1, v_2, Q_x and Q_y at each time-step.

In order to limit the effects of these variabilities, we introduce secondary objective functions, as set out by Gomez et al. [26]. For each time step, we solve an optimisation problem, like the one described in equation (6.3.18). This provides an optimal value for our primary objective function, $\mathbf{a}^T\mathbf{v} + \mathbf{b}^T\mathbf{c} = A_{max}$, in this case, $v_2 = v_{2,max}$. Next we form another optimisation problem with all the constraints of the first, but also includes the additional constraint, $\mathbf{a}^T\mathbf{v} + \mathbf{b}^T\mathbf{c} = A_{max}$. We choose a secondary objective function of the same form for this new optimisation problem, and solve again. In this way, dFBA chooses from the set of states that optimise the primary objective function, the state that optimises the secondary objective function. We could continue adding more objective functions in this way until we a unique solution that is optimal according to all the objective functions. Since we must perform all these optimisations in sequence, this means for n objective functions, we must form n optimisation problems and solve them. This could prove quite time consuming for larger systems, which are also the ones more at risk of these multiple objective functions, especially those with a high level of degeneracy.

As an example, we apply this method to our model in equation (6.3.7), with the first objective function of maximising v_2, and the second as described, maximising $v_1 + v_3$ to maximise the flux from x to y. Figure 6.10 shows the results of this model with two

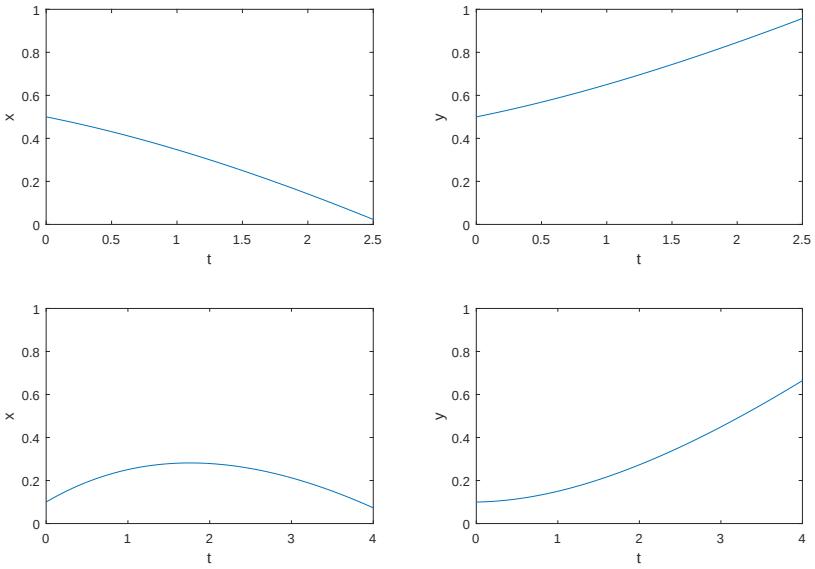

Figure 6.10: Graph of concentrations against time for the dFBA model described in equation (6.3.7) for the primary objective function of maximising v_2, and the secondary of maximising $v_1 + v_3$, for two different initial conditions. The top row shows the behaviour with the initial condition $x = 0.5$, $y = 0.5$, and the second row for $x = 0.1$, $y = 0.1$. Both initial conditions lead to the same long term behaviour with x decaying to zero and y increasing.

different initial conditions. We see that both models tend towards a high v_2 state, with the concentration of x decaying to zero, and y growing to a maximum.

By adding a secondary objective function to our model, we force the model to have only one final steady-state that depends on both objective functions. After eliminating the problem of these alternative optimal solutions, were are unable to observe these multiple steady-states for different initial conditions. Whilst dFBA is able to show the existence of both a high y and a low y state as expected, after including secondary objective functions, we are not able to observe the same dFBA model (including the same objective function) tending to two different states depending on the initial conditions. dFBA is able to provide some insight into the existence of multiple steady-states, however it is not able to predict differing initial conditions leading to these two different states.

6.4 Systems with periodic solutions

Until now, we have only used dFBA to examine systems with steady-state behaviour. Ideally, we wish to use dFBA to model a system exhibiting non-steady-state behaviour, such as periodic oscillations. To do this, it would be useful to analyse a system which exhibits these behaviours for all parameter values, unlike the autocatalytic system analysed in the previous section, which only exhibits periodic solutions for a small range of parameters. By choosing a system which observes these oscillations for any given set of parameter values, we are more likely to observe them in our dFBA.

Though flux balance techniques are normally only used for metabolic networks, it is possible to use them for any system that follows the correct structure, i.e. networks that can be written in the form $\dot{\mathbf{x}} = S\mathbf{v}(\mathbf{x})$. One such system of equations that follows this structure and exhibits periodic solutions is the Lotka-Volterra predator-prey equations

$$\frac{dx}{dt} = \alpha x - \beta xy, \tag{6.4.1}$$

$$\frac{dy}{dt} = \delta xy - \gamma y, \tag{6.4.2}$$

with some initial conditions, $x(0) = x_0$ and $y(0) = y_0$. This system of equations has two steady-states, an extinction state, where $x = y = 0$, and an equilibrium state, $x = \gamma/\delta$, $y = \alpha/\beta$. The extinction state $(0, 0)$ is unstable, whilst the equilibrium state is marginally stable. All other points belong to periodic orbits about the equilibrium point, examples of which are shown in Figure 6.11. Due to this, it is a good candidate for testing the ability of dFBA to observe periodic solutions to a network.

6.4.1 Forming the model

The Lotka-Volterra system can be considered as a metabolic network with three reactions, an input of X, an output of Y, and a third reaction that transforms X into Y. These reactions are given in Table 6.3. From these reactions we form the stoichiometric matrix

$$S = \begin{pmatrix} -\beta & \alpha & 0 \\ \delta & 0 & -\gamma \end{pmatrix}. \tag{6.4.3}$$

This stoichiometric matrix, (6.4.3), is similar to the one used in Section 6.2.1. Indeed the network is effectively a straight line system with no intermediary metabolites. In order for our new model behave differently from the previous one, we use different

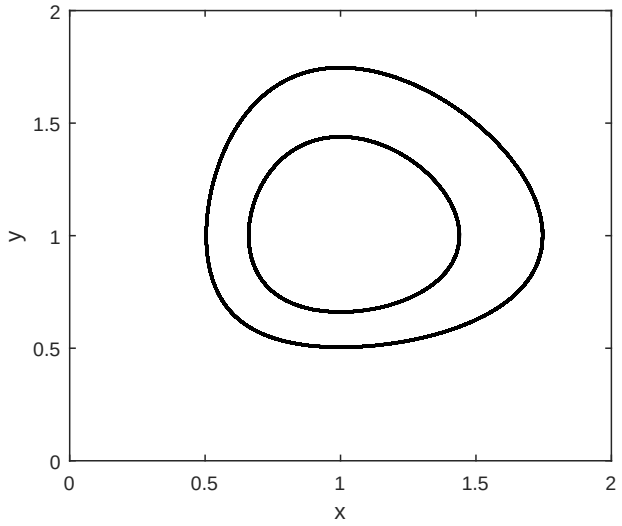

Figure 6.11: Plot of two separate periodic solutions for the Lotka-Volterra equations, (6.4.1)–(6.4.2) in the xy plane.

β X	$\xrightarrow{v_1}$	δ Y
External concentration	$\xrightarrow{v_2}$	α X
γ Y	$\xrightarrow{v_3}$	External concentration

Table 6.3: The metabolic network that is equivalent to the Lotka-Volterra predator-prey system.

concentration-based constraints. Before, the input flux was constrained by equation (6.2.4) but we now use

$$v_2 \leq x. \tag{6.4.4}$$

Similarly, we set the output of y to be constrained by

$$v_3 \leq y. \tag{6.4.5}$$

As usual, we set maximum and minimum values for our fluxes. In this example, all fluxes are irreversible, having minimum values of zero. Unlike previous examples, there is no factor limiting the input rate or overall flow through the network, so we impose maximum flux values of 100. We also add the usual constraints for concentrations (limiting their minimum value to be zero), and a maximum rate of change vector for the fluxes, $\mathbf{v}_{dt-max}$. As in our autocatalysis model, we set all the values of this vector to be equal for simplicity. By combining all our constraints with an objective function, we again formulate a dFBA model. In this case, we take the objective function to be maximising the output of y, v_3, so our model is

$$\begin{aligned} \underset{\mathbf{c},\mathbf{v}}{max/min} \quad v_3 \quad such\,that: \frac{d\mathbf{c}}{dt} &= S\mathbf{v}, \\ v_1 &\leq x, \\ v_3 &\leq y, \\ \mathbf{c} &\geq \mathbf{0}, \end{aligned}$$

$$
\begin{aligned}
0 \leq v_n \leq 100, \\
-\mathbf{v}_{dt-max} \leq \frac{d\mathbf{v}}{dt} \leq \mathbf{v}_{dt-max}, \\
\mathbf{v}(0) = \mathbf{v}_0, \quad \mathbf{c}(0) = \mathbf{c}_0.
\end{aligned} \tag{6.4.6}
$$

Now that we have our model, we choose values for α, β, γ, and δ in (6.4.1)–(6.4.2) . For simplicity, we take them to all equal one. This means the equilibrium state for the ODE model is $x = 1$, $y = 1$, and the equivalent equilibrium state for our dFBA model would be $\mathbf{v}_{ss} = \lambda(1, 1, 1)$, for some scalar λ. This requires concentrations of at least $x = \lambda$, and $y = \lambda$. Now that we have a complete model, we simulate it.

We start by testing the system using the two known steady-states as initial conditions, that is the extinction state of $x = 0$ and $y = 0$, and the equilibrium population, where $x = 1$ and $y = 1$, or rather, some linear multiple of this. It is obvious to see that if the system starts with the initial condition of $x = 0$, $y = 0$ then the system is unable to take up any additional x, and remains in the extinction state with all concentrations and fluxes equal to zero. If the system starts at the maximal point, that is the point at which all fluxes and concentrations take their assigned maximum value (in this case, 100), the system remains at this point. We see that this is indeed the optimal value for our objective function, and it is also an equilibrium state. Taking initial conditions that aren't either one of these points, however, leads to a problem. As seen in Figure 6.12 one of the concentrations tends towards zero with a sharp gradient. At this point

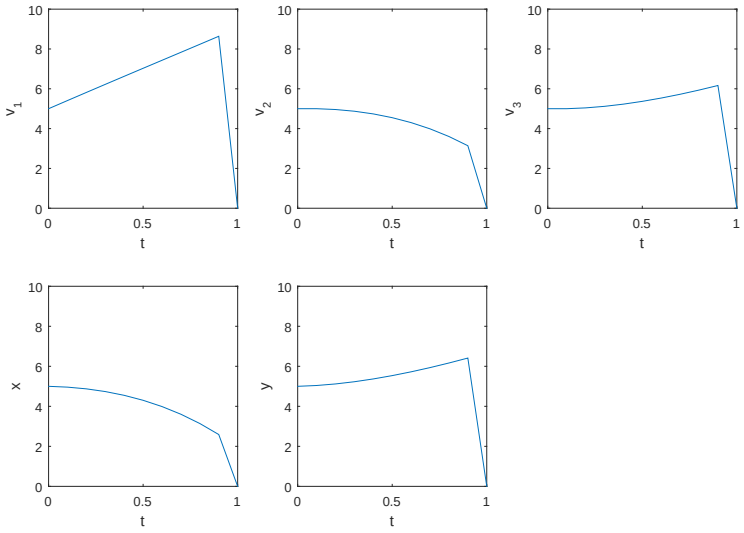

Figure 6.12: Graph of concentrations and fluxes against time for the dFBA model described in equation (6.4.6) with all fluxes and concentrations initially equal to five. The model fails at $t = 1$, as the concentration of x will decrease past zero at this time point.

the fluxes are such that in the next time step, this decreasing concentration will drop below zero. Due to the maximum rate of change that we have enforced on our fluxes, the system is unable to choose a point in the next time step that does not cause this concentration to drop below zero, and the system becomes infeasible at this time step. We need some way to deal with a system that is crashing in this manner in order to prevent concentrations below zero being recorded.

6.4.2 Dealing with crashes

There are a few methods to deal with these "crashes" in concentration. The easiest way is to relax or entirely remove the rate of change constraints on the fluxes at points when the system has become infeasible. In this way, we allow the fluxes to change enough to avoid a concentration decreasing below zero. However, the choice of how much we relax these constraints is rather arbitrary. Removing the constraints entirely may lead to different results than if we doubled the maximum rates of changes. We aim to give the system only as much freedom as it needs. Using the technique described by Gomez et al. [26], we introduce a new set of variables to the system, labelled $\mathbf{s}$. These variables represent how much the flux rate of change constraints are relaxed at each time step. The flux rate of change constraints now become

$$\frac{d\mathbf{v}}{dt} \leq \mathbf{v}_{dt-max} + \mathbf{s}, \tag{6.4.7}$$

where $\mathbf{v}$ is the flux vector, $\mathbf{v}_{dt-max}$ is a vector containing their maximum allowed values, and $\mathbf{s}$ is the vector of variables chosen by the linear programming, showing how much these rate of change constraints are being violated in a given time step. The elements of $\mathbf{s}$ are constrained to be positive. By changing our single linear programming problem into a two step linear programming, we assure that these additional variables, $\mathbf{s}$ are as small as possible whilst still maintaining our usual objective function. In the first step, we set our objective function to be minimising the sum of the elements of $\mathbf{s}$. In the case of the original system being feasible, this minimum value is zero (as each element cannot take a value less than zero). If the original system would be infeasible, then the sum of the elements takes a non-zero minimum value. In the second step, we fix the sum of these $\mathbf{s}$ elements to take its minimum value as found in the first step as an additional constraint, and then optimise for our usual objective function. In this way, we are able to ensure the flux rate of change conditions are violated as little as possible, and only when necessary. This allows us to run simulations that would otherwise become infeasible and thus have no solution.

Now that we are able to find solutions in situations we previously couldn't, we examine our Lotka-Volterra model outside of the two previously observed steady-state points. First we examine the system on other points of the dFBA equilibrium state, where $x = y$ and $v_2 = v_3$. As seen in Figure 6.13 despite starting in what is a steady-state (where the concentrations would not change in time), the method attempts to choose points off this line in order to attain a more optimal state of the objective function, that is, increasing v_1, with v_3 increasing later. This leads to a rapid decrease in the concentration of x, which in turn eventually leads to a sharp decline in v_1, and eventually the decay of the concentration of y, leading to the extinction state. This behaviour persists outside of the equilibrium vector, with almost all initial conditions (with the exception of the previously mentioned optimal state) tending towards this state.

6.4.3 An initial condition from the ODE model

We now attempt to use dFBA to predict periodic behaviour in a metabolic network. The Lotka-Volterra ODE system has two steady-states, and any other point in the (x, y) plane lies on a periodic orbit around the equilibrium point. In order to see if this periodic behaviour is observed with dFBA, we choose an initial condition that corresponds to a state taken by the ODE model. Given a point (x, y), we calculate the values of the fluxes at this point, and use that as our initial condition. We choose the point $x = 1.107$ and $y = 0.284$. We then calculate the values of each flux according to

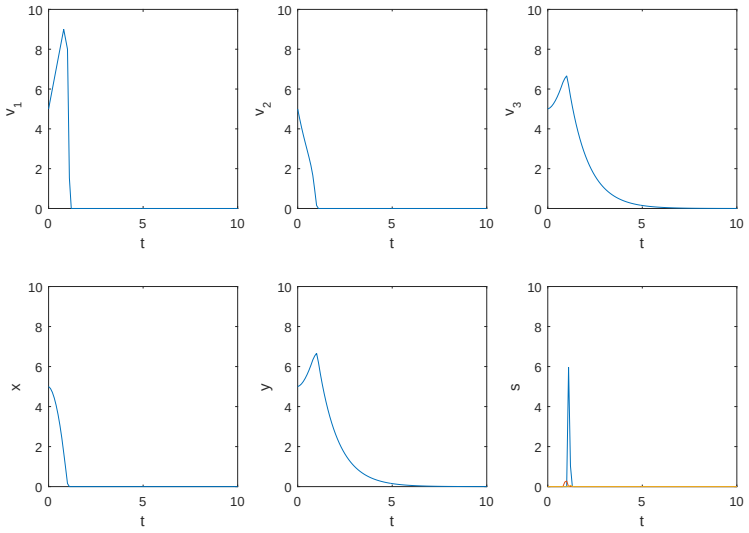

Figure 6.13: Graph of concentrations and fluxes against time for the dFBA model described in equation (6.4.6) with all fluxes and concentrations initially equal to five. We see a spike in the $\mathbf{s}$ variables at about $t = 1$, to prevent the system from becoming infeasible. After this happens, the concentration of x becomes zero, and the system decays to the extinction state.

the ODE model for these concentrations, to find $v_1 = xy = 0.3144$, $v_2 = x = 1.107$, and $v_3 = y = 0.284$. We also set $\mathbf{s} = \mathbf{0}$ for our initial condition, leading to the initial condition vector,

$$(v_1, v_2, v_3, x, y, s_1, s_2, s_3) = (0.3144, 1.107, 0.284, 1.107, 0.284, 0, 0, 0). \tag{6.4.8}$$

Using this initial condition, with the model given in equation (6.4.6), we obtain the solution shown in Figure 6.14. At first, this solution behaves similarly to the ODE model, with the increase in the concentration x up to a peak before it starts to decrease, followed by a peak in the concentration of y. However, unlike the ODE system, the concentration of x continues decreasing until it reaches zero. At this point, there is a "crash", which necessitates a non-zero value for $\mathbf{s}$. After this point, the system is unable to increase the concentration of x, and the concentration of y decreases towards the extinction state. In this case, the amplitude of the periodic oscillations in the ODE system was large, so it is worth examining a solution with a smaller amplitude, where we do not expect the concentration of x to decrease as rapidly. An initial condition for a state with lower amplitude oscillations is

$$(v_1, v_2, v_3, x, y, s_1, s_2, s_3) = (0.7310, 0.9352, 0.7816, 0.9352, 0.7816, 0, 0, 0). \tag{6.4.9}$$

However, as seen in Figure 6.15 the model shows similar behaviour even with this different initial condition. Once the concentration of x starts decaying, the system is not able to increase it again, so it crashes towards the extinction state. This is due to the objective function prioritising the output of y over everything else. In order for the system to start accumulating x again, it would require v_1 to be lowered, which would in turn lower the concentration of y, allowing for less output though v_2. We see that the dFBA model is not able to predict the existence of periodic oscillatory solutions for this network.

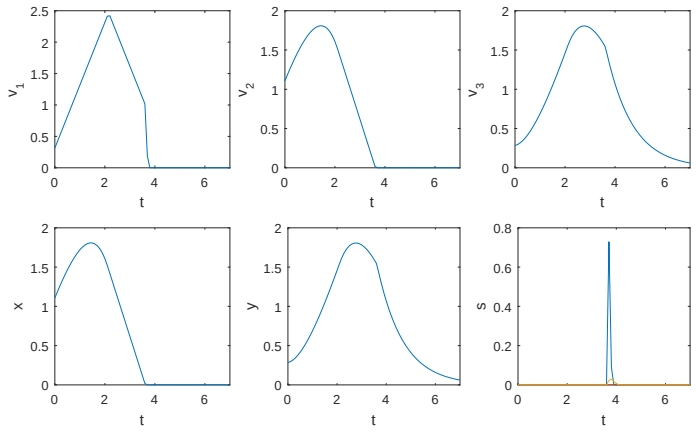

Figure 6.14: Graph of concentrations and fluxes and **s** variables for the model in equation (6.4.6), with initial condition from equation (6.4.8. We see a peak in one concentration before it starts to decay, and the other concentration then peaks and also decays. Rather than becoming periodic, the system crashes, and eventually decays to the extinction state seen previously.

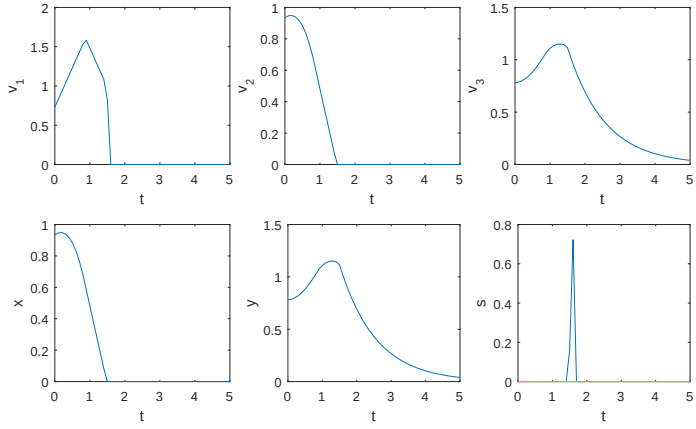

Figure 6.15: As in Figure 6.14, except with initial condition given by equation (6.4.9).

6.5 The *C. Autoethanogenum* network

Now that we have examined smaller systems and identified the potential problems, we apply dFBA to a larger metabolic network, such as the modified *C. autoethanogenum* network seen in Section 4.4. As before, we first form the stoichiometric matrix. The stoichiometric matrix concerning the internal reactions for our *C. autoethanogenum*

network is

$$S_{in} = \begin{array}{r} \\ CO \\ CO_2 \\ CO_c \\ Formate \\ Acetyl-CoA \\ Pyruvate \\ Acetolactate \\ Acetaldehyde \\ Acetate \\ Hydrogen \end{array} \begin{array}{c} \begin{array}{cccccccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_9 & v_{10} \end{array} \\ \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 2 & -2 & -2 & -4 & -2 & 0 & 0 & -2 & 0 & 0 \end{pmatrix} \end{array}, \tag{6.5.1}$$

and the stoichiometric matrix for the uptake and output reactions is

$$S_{ex} = \begin{array}{r} \\ CO \\ CO_2 \\ CO_c \\ Formate \\ Acetyl-CoA \\ Pyruvate \\ Acetolactate \\ Acetaldehyde \\ Acetate \\ Hydrogen \end{array} \begin{array}{c} \begin{array}{cccccccc} Q_{CO} & Q_H & out_{bio} & out_{lac} & out_{val} & out_{but} & out_{eth} & out_{ace} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & -2 & 0 & -2 & -2 & 0 \end{pmatrix} \end{array}. \tag{6.5.2}$$

Then the complete stoichiometric matrix is the augmented matrix

$$S = (S_{in}|S_{ex}). \tag{6.5.3}$$

We fix constraints on the uptake and output fluxes depending on concentrations. We have two uptake fluxes, Q_{CO} and Q_H. In our previous examination of this network, using both ODEs and FBA, we fixed these fluxes to be constants. To provide fair comparison, we continue fixing these fluxes to take constant values,

$$Q_H = H_{in}, \quad Q_{CO} = CO_{in}. \tag{6.5.4}$$

We constrain our output fluxes as in the ODE model, so that

$$out_{bio} = k_{15} c_{ACA} c_{Py}, \tag{6.5.5}$$
$$out_{lac} = k_{16} c_{Py} c_H^2, \tag{6.5.6}$$
$$out_{val} = k_{17} c_{Al}, \tag{6.5.7}$$
$$out_{but} = k_{18} c_{Al} c_H^2, \tag{6.5.8}$$
$$out_{eth} = k_{19} c_{Ad} c_H^2, \tag{6.5.9}$$
$$out_{ace} = k_{20} c_A. \tag{6.5.10}$$

We require, as usual, the concentrations and irreversible fluxes to be positive. We choose not to add maximum values for concentrations and fluxes in this case, so

$$\mathbf{c} \geq \mathbf{0}, \quad \mathbf{v} \geq \mathbf{0}. \tag{6.5.11}$$

Finally, we continue to enforce constraints on how much the fluxes $\mathbf{v}$ are allowed to change in a single time step. For our simulations, we allow these to be relatively relaxed, and fix the maximum change in a flux in a single unit of time to be one, so that

$$-1 \leq \frac{dv_n}{dt}, \frac{dQ_n}{dt}, \frac{dout_n}{dt} \leq 1. \tag{6.5.12}$$

By combining these constraints with as objective function, in this case maximising the production of biomass, out_{bio}, we form a dFBA model for *C. autoethanogenum*. This is the same objective function we used in previous FBA analysis. Then our dFBA model is

$$\begin{aligned} \max_{\mathbf{c},\mathbf{v}} \quad out_{bio} \quad such\ that: \frac{d\mathbf{c}}{dt} &= S\mathbf{v}, \\ Q_H &= H_{in}, \\ Q_{CO} &= CO_{in}, \\ out_{bio} &= k_{15}c_{ACA}c_{Py}, \\ out_{lac} &= k_{16}c_{Py}c_H^2, \\ out_{val} &= k_{17}c_{Al}, \\ out_{but} &= k_{18}c_{Al}c_H^2, \\ out_{eth} &= k_{19}c_{Ad}c_H^2, \\ out_{ace} &= k_{20}c_A, \\ \mathbf{c} &\geq \mathbf{0}, \\ \mathbf{v} &\geq \mathbf{0}, \\ -1 \leq \frac{dv_n}{dt}, \frac{dQ_n}{dt}, \frac{dout_n}{dt} &\leq 1, \\ \mathbf{v}(0) = \mathbf{v}_0, \quad \mathbf{c}(0) &= \mathbf{c}_0. \end{aligned} \tag{6.5.13}$$

6.5.1 Steady-state regions

First, to check that the model is producing the results we expect, we investigate whether it tends to a steady-state in parameter regimes where steady-states exist. In Section 4.6 we determined the region in which steady-states exist using FBA. We fix our parameters so that the model (6.5.13) resides in this region. We set $k_n = 1$ for $15 \leq n \leq 20$. To maintain the Q_H/Q_{CO} ratio that permits steady-states, we set $Q_H = 3$ and $Q_{CO} = 1$. Using these parameters along with the objective function of maximising biomass, we simulate the dFBA model (6.5.13). The results of this simulation are shown in Figure 6.16. We see that our concentrations all tend to steady-states, as expected, though some concentrations remain very low. We also note there is an initial growth and decay in the concentrations of hydrogen and carbon monoxide. This is likely due to both metabolites having constant input rates, Q_{CO} and Q_H, which dominate the system early on, before other fluxes and concentrations grow to steady-state. After this initial decay, the system levels out to the steady-state that we expect to see.

6.5.2 Linearly growing states

In some regions of parameter space, our ODE model tends to a non-steady-state in which one or more concentrations grows linearly or sublinearly in time, with others decaying to zero. It is worth examining whether dFBA predicts the existence of these non-steady-state regions. We would also like to see if it predicts concentrations growing linearly in time, as well as the products formed in these cases.

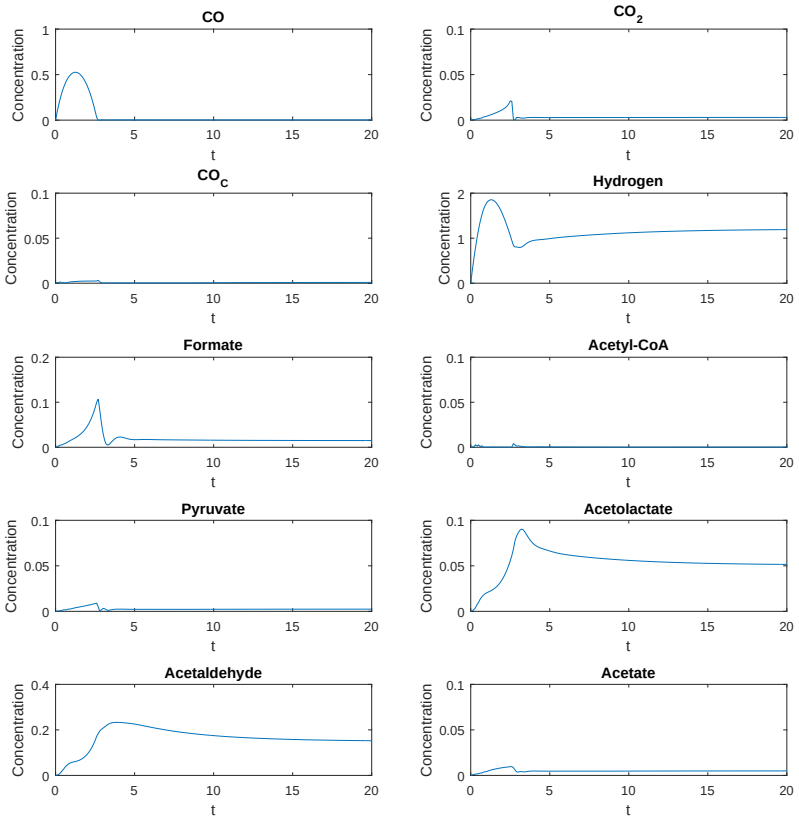

Figure 6.16: Simulation of the model described in equation (6.5.13), with all $k_n = 1$, $CO_{in} = 1$, and $H_{in} = 3$.

Our asymptotic analysis of the ODE system led to seven qualitatively different long term behaviours. A diagram of regions in parameter space where these behaviours are found is shown in Figure 6.17. Behaviour 1 is the steady-state behaviour, whilst the remaining six refer to various states with concentrations experiencing linear, or sublinear growth, and others decaying to zero. The long timescale behaviour of the concentrations is summarised in Table 6.4.

In our dFBA model, (6.5.13), we have direct control over the input levels for hydrogen and CO, so we are able to vary the input ratio to force the dFBA model into a state that corresponds to one of the behaviours found in our ODE model. As noted in Section 4.2.1, the ratio of two fluxes, v_2/v_3, is equal to the ratio of the two parameters, k_2/k_3, so by constraining the ratio of those two fluxes, as well as the input fluxes, we are able to force the dFBA model to act in any of the regions shown in Figure 6.17. The ratio v_2/v_3 is especially important for distinguishing between behaviours 5 and 6 since, when there is a high hydrogen input, our ODE model determines whether it is in behaviour 5 or 6 based on this ratio. To force this ratio to behave in the way we wish, we add one of two constraints. In the cases where we want $v_2 < v_3$, we add the constraint

$$v_2 - v_3 < 0, \tag{6.5.14}$$

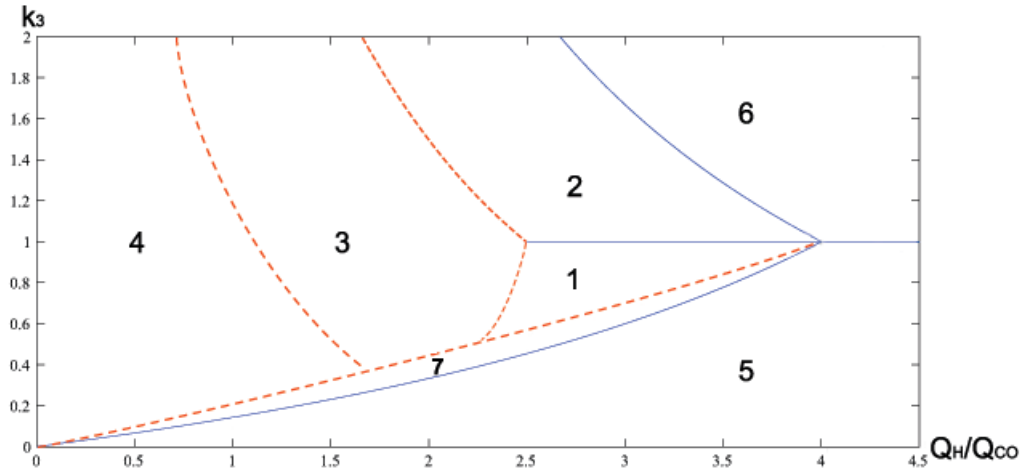

Figure 6.17: The regions in parameter space for the system of ODEs with regions numbered by observed behaviour. The horizontal axis is Q_H/Q_{CO}, with Q_H being varied, and the vertical axis is k_3. All other parameters are fixed to be one.

Metabolite	Regions of parameter space						
	1	2	3	4	5	6	7
CO	t^0	t^0	t^0	t^0	t^0	t^0	t^0
CO_2	t^0	t^0	$t^{0.5}$	t^1	t^{-2}	t^{-2}	t^{-1}
CO_c	t^0	t^1	t^1	t^1	t^{-5}	t^1	t^{-3}
Formate	t^0	t^{-1}	t^0	t^1	t^1	t^{-5}	t^1
Acetyl-CoA	t^0	t^0	t^0	t^{-1}	t^{-2}	t^{-2}	t^{-1}
Pyruvate	t^0	t^0	t^0	t^0	t^{-3}	t^{-4}	t^{-1}
Acetolactate	t^0	t^0	t^0	t^0	t^{-8}	t^{-10}	t^{-3}
Acetaldehyde	t^0	t^0	$t^{0.5}$	t^0	t^{-2}	t^{-2}	t^{-1}
Acetate	t^0	t^0	t^0	t^{-1}	t^{-2}	t^{-2}	t^{-1}
Hydrogen	t^0	t^0	$t^{-0.25}$	$t^{-0.5}$	t^1	t^1	$t^{0.5}$

Table 6.4: Scalings for each metabolite as $t \to +\infty$ in each regime, illustrated in Figure 6.17. A scaling of t^0 represents a metabolite tending to a nonzero steady-state. An exponent greater than zero represents a metabolite showing unbounded growth. A negative exponent represents a concentration decaying to zero.

and in the cases where we want $v_3 < v_2$, we add the constraint

$$v_3 - v_2 < 0. \tag{6.5.15}$$

Then by combining our dFBA problem described in equation (6.5.13) with one of either (6.5.14) or (6.5.15) for various values of CO_{in} and H_{in}, we force our system to be in one of the regions seen in Figure 6.17.

We have already observed steady-state behaviour in Section 6.5.1, but it is worth noting that in this behaviour we have $v_3 < v_2$, despite equation (6.5.15) not being included as a constraint in the dFBA model. We now consider other input ratios and constraints on v_2 and v_3 to try and observe the behaviours seen in other regions.

We start with the regions in which $v_3 < v_2$, that is, with equation (6.5.15) as an additional constraint to the dFBA model described in equation (6.5.13). As when observing the steady-state region, parameters k_{15} to k_{20}, and CO_{in} are all set to equal one. We then choose a range of values of H_{in} to observe how the input ratio affects the behaviour of the model. First we choose $H_{in} = 3.5$ to represent a high hydrogen level.

Figure 6.18 shows the results of the simulation with these constraints. With the hydrogen input H_{in} set to equal 3.5, and the constraint from equation (6.5.14), the dFBA should be modelling the region of parameter space in which the ODEs exhibited behaviour 5. We expect to see a linear growth in hydrogen and formate, with the decay to zero of all other concentrations. Whilst we do see the expected increase in hydrogen, formate does not increase as expected, and most concentrations stay at very low steady-states rather than decaying entirely. We have a decay of acetolactate and acetaldehyde, and to a lesser extent, carbon dioxide and acetate. Examining some of these low steady-states in more detail, we see that they aren't actually steady, and instead exhibit random spiking with a low amplitude, as seen for the concentration of acetyl-CoA in Figure 6.19

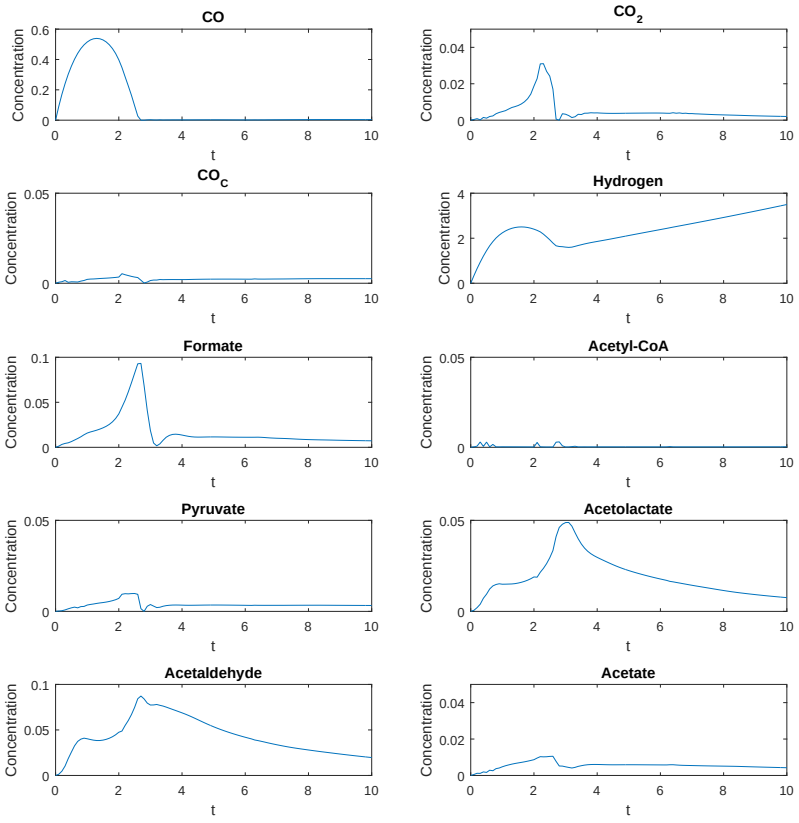

Figure 6.18: Plot of the concentrations of metabolites for the model given in equation (6.5.13), with $H_{in} = 3.5$ and $v_3 < v_2$.

We next examine a lower hydrogen case, with $H_{in} = 1$. Figure 6.20 shows the results of the dFBA simulation in this case. With this lower hydrogen input, we expect the system to exhibit either behaviour 3 or behaviour 4 from the ODEs. Instead we observe a combination of the two. We have the growth in carbon dioxide and formate from

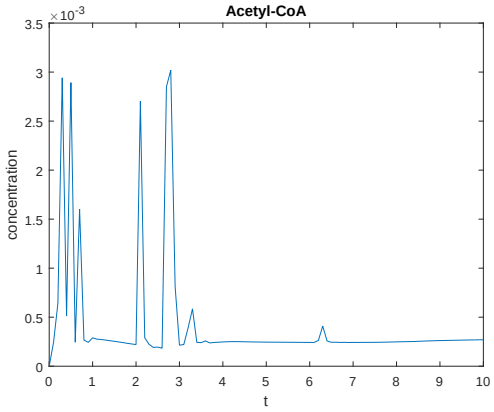

Figure 6.19: Plot of the concentration of acetyl-CoA for the model in equation (6.5.13), with $H_{in} = 3.5$ and $v_3 < v_2$, showing the random spiking in the metabolite concentration.

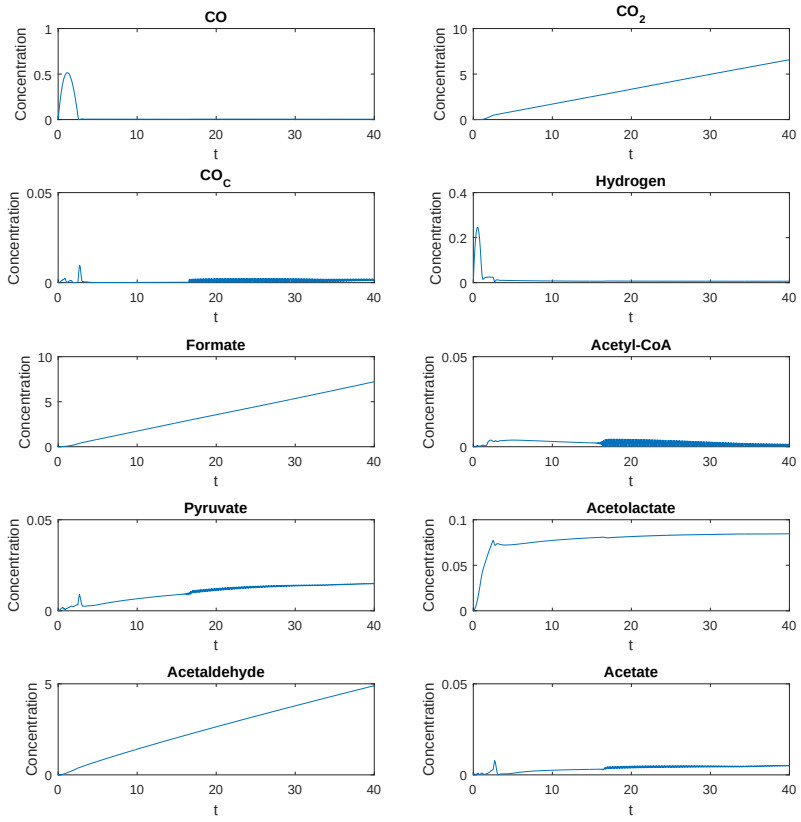

Figure 6.20: Plot of the concentrations of metabolites for the model given in equation (6.5.13), with $H_{in} = 1$ and $v_3 < v_2$.

behaviour 4, but also the growth in acetaldehyde from behaviour 3. Interestingly, there is no growth in the secondary source of carbon monoxide, CO_c, that was observed in

the ODEs for both behaviours. We note that the concentration of acetyl-CoA exhibits random spiking of a higher amplitude, but there is still a slow decay in the concentration as time goes on, which matches behaviour 4. However, we see a non-zero steady-state in the concentration of acetate, which is expected in behaviour 3. Pyruvate and acetolactate maintain steady-states as expected in both behaviours. It is interesting that our dFBA has provided us a set of balances that is a mixture of two of the states observed in our ODE model. It may be that dFBA has found a possible linear growth state for this ODE model that was not found in our asymptotic analysis. It turns out, however, that this is not the case. If these growth and decay results were seen in our ODE model, there is no leading balance in the equation for acetate. Discarding the non-leading terms would give us the equation

$$0 = -(k_9 + k_{20})c_A, \tag{6.5.16}$$

where c_A is the concentration of acetate. This could only hold true if the concentration of acetate decayed to zero, since all of k_9, k_{20}, and c_A are positive, however in this case there is a non zero steady-state for the concentration of acetate. Whilst this balance might be possible for a different set of ODEs modelling the same network, it is not possible for the our model.

We next examine regions in which $v_3 > v_2$, that is, when equation (6.5.14) is added included as a constraint in the dFBA model (6.5.13). As before, we begin with the high hydrogen level, where $H_{in} = 3.5$. Figure 6.21 shows the results of the dFBA simulation. Here, we expect to observe behaviour similar to behaviour 6 on the Table 6.4. The observed behaviour is closely matches the behaviour from our ODE models, with growth in the secondary carbon dioxide, CO_c and in hydrogen as expected. We also have carbon dioxide, acetolactate, acetaldehyde and acetate all decaying in time. Acetyl-CoA maintains a low concentration with low amplitude spiking, as seen in Figure 6.19). We also observe these low amplitude concentration spikes in formate.

Next, we model the system with a moderate level of hydrogen, with $H_{in} = 2.5$. With this level of hydrogen, we expect the system to behave as in behaviour 2 from Table 6.4. Figure 6.22 shows the results of our dFBA simulation in this case. We find almost every concentration tending to a non zero steady-state. There is linear growth in the concentration of the secondary source of carbon monoxide, CO_c, and the concentration of formate remains close to zero for the entire time frame. The long timescale behaviour here is close to the long timescale behaviour exhibited by the ODEs in this region of parameter space.

Finally, we examine the system with a low level of hydrogen, where $H_{in} = 1$. The results of the dFBA simulation in this low hydrogen case are shown in Figure 6.23. We have growth in carbon dioxide, CO_c, and acetaldehyde. It appears that the concentration of hydrogen is decaying, though in a non smooth way, as we see some spiking at around $t = 23$. Other concentrations, other than acetolactate, also exhibit these spikes, though the amplitude of these does not appear to change in time. Whilst these concentrations are not steady, they might be considered such, as the average value appears to be constant. Acetolactate exhibits a regular steady-state. If we take the oscillating concentrations to be nearly steady, we do find that the behaviour matches that of behaviour 3 for our ODE model. We do have the same concentrations growing indefinitely in time, and an overall decay in the levels of hydrogen in the long timescale.

Running the low hydrogen simulation again with neither equations (6.5.15) or (6.5.14) added as constraints, we find the results seen in Figure 6.24. These are similar to those seen in Figure 6.20, where we had equation (6.5.15) included as a constraint in the dFBA model, but we also have growth in the secondary carbon monoxide source

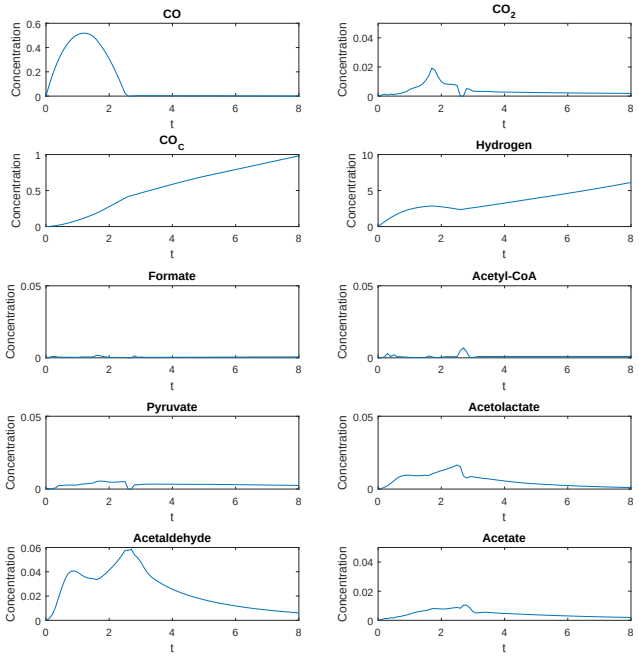

Figure 6.21: Plot of the concentrations of metabolites for the model given in equation (6.5.13), with $H_{in} = 3.5$ and $v_3 > v_2$.

that was expected from Table 6.4. This is also similar to the results with equation (6.5.14) with an additional growth in formate, seen in Figure 6.23.

It is interesting to note that even with the hydrogen input to equal zero, we cannot find behaviour that matches behaviour 4 in Table 6.4 perfectly. Specifically, we cannot prevent the unbounded growth in acetaldehyde seen in both of our low hydrogen dFBA models. Our ODE models predict that the creation of acetaldehyde is limited by the concentrations of acetate, acetyl-CoA and hydrogen. In behaviours 4 for the ODEs, all three of these metabolites decay to zero, however we see in our low hydrogen dFBA simulations, all three of these concentrations maintain steady-states, allowing for a constant production of acetaldehyde.

Across all of these behaviours we note that whilst linearly growing concentrations are generally predicted in agreement with the ODE results, dFBAs ability to predicted concentrations decaying to zero is more limited. In some cases, we find the previously observed problem with concentration "crashing" at zero, and the system struggling to change fluxes to avoid negative concentrations, whilst in other we see random spiking of the metabolite concentration, allowing the fluxes related to the concentration to maintain activity. As an example we observe the concentration of hydrogen in the low hydrogen regions. In our ODE model, we observe the concentration of hydrogen decaying smoothly to zero, but in the dFBA model, we see the concentration of hydrogen fluctuates at a low level, as shown in Figure 6.25.

Overall, there is a general agreement between the dFBA predictions for linearly growing concentrations, and those observed in the ODE model. All behaviours, except behaviours 2 and 4, are observed in the dFBA model, though on some occasions they may appear in regions we do not expect them to (such as behaviour 3 being observed when we have a near zero input of hydrogen, where we would expect to find behaviour

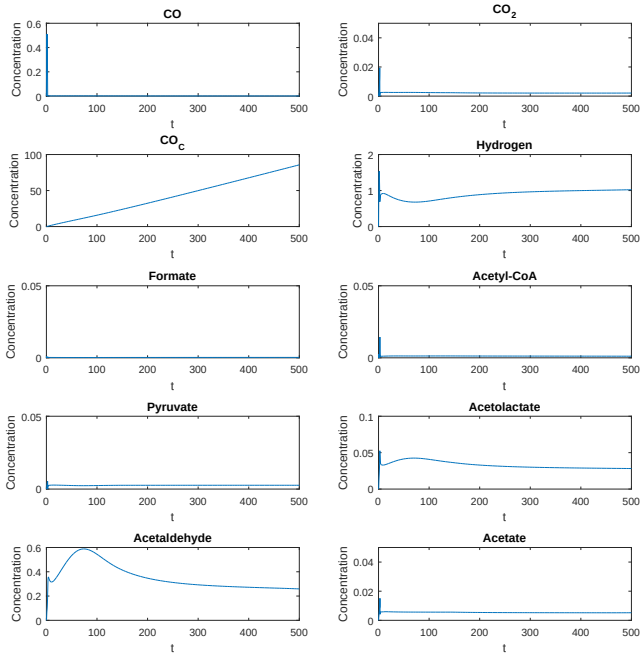

Figure 6.22: Plot of the concentrations of metabolites for the model given in equation (6.5.13), with $H_{in} = 2.5$ and $v_3 > v_2$.

4). We find problems with predicting concentrations decaying to zero, allowing potential behaviours that would not be permitted by our previous ODE model. The ability of dFBA to capture this non-steady-state behaviour is interesting because it suggests that the linear and sublinear growth behaviours observed in the ODE models from Chapter 4 may not be necessarily tied to the specific choices we make for modelling reaction rates, but may instead be more generic to the structure of the metabolic network. In particular we see a broad pattern of high hydrogen input leading to an unbounded growth for hydrogen, and low hydrogen regions leading to an unbounded growth of acetaldehyde. We also see in regions with $v_3 > v_2$ a growth in CO_c and CO_2, and in regions with $v_2 > v_3$ a growth in formate. These predictions are generally in line with those of our ODE model, in particular the switch between CO_2 growth and formate growth.

For the straight line model examined in Section 6.2, we noted that internal concentrations were largely irrelevant to the steady-state behaviour of the dFBA model, however, in our non-steady-state simulations we find the internal concentrations to be relevant and similar to those observed in the ODEs. It should be noted that this network has already been simplified by reducing long chains of reactions down into single reactions, as suggested at the end of Section 6.2.

6.6 Conclusions

Dynamic flux balance analysis is a powerful tool for examining a metabolic network in non-steady-states. We have been able to model systems tending to both steady-state and linear growth states. However, we were not able to use it to model systems

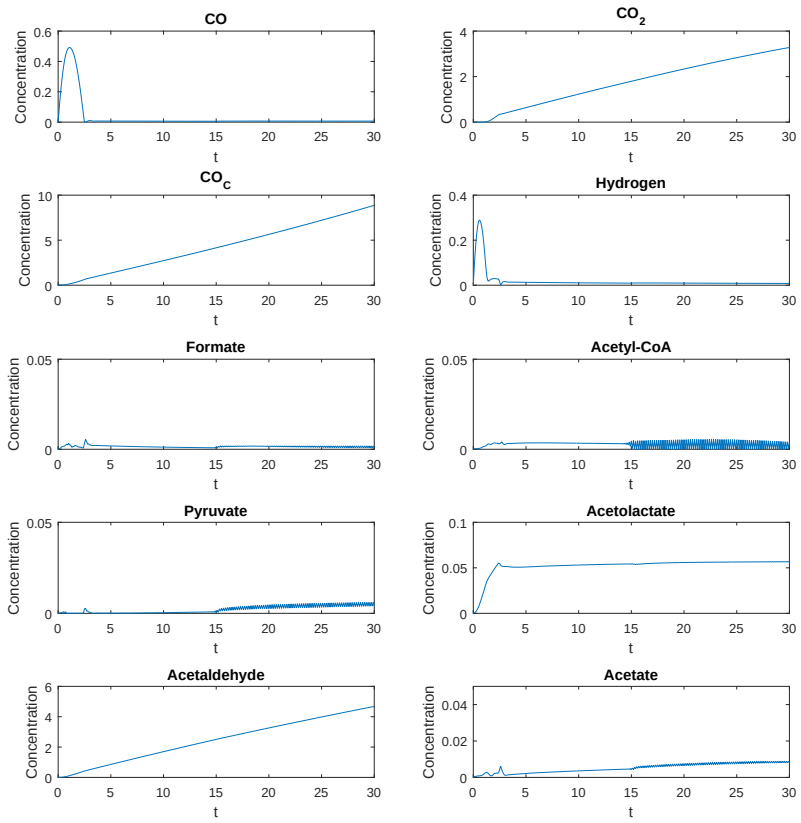

Figure 6.23: Plot of the concentrations of metabolites for the model given in equation (6.5.13), with $H_{in} = 1$ and $v_3 > v_2$.

with periodic oscillations, and we were unable to find multiple steady-states using the same objective function after addressing the issue of multiple optimal solutions. The technique has found use in modelling batch cultures, as seen in studies of *E. coli* by Hanly and Henson [30], and *S. cerevisiae* by Hjersted and Henson [31], in which dFBA was used to optimise fermentation in batch cultures. In particular, they were able to find medium conditions that optimised these processes. The technique is still relevant to continuous cultures. In Section 6.5, we showed the ability of dFBA to predict the linear growth states of a metabolic network with constant inputs. The results were comparable to those observed from our previous ODE model for the same system. The results obtained from dFBA are notable, because they have less restrictions on the flux values than our original ODE models. It may be that the growth states seen in our dFBA model represent behaviours more generic to the network, whilst those seen from our ODE model are more specific to that exact set of kinetics. We note that the state seen in Figure 6.20 would be impossible in our previously seen ODE system, but might be possible for a different set of kinetics matching the network.

dFBA models are able to predict and model the dynamics of a metabolic network. However, since they rely on only the stoichiometry of the network, and a small number of kinetic parameters, it is difficult to quantify exactly how a network is behaving through them. We are able to see the dynamic flux and concentration profiles, but we have little insight into the mechanical properties of the reactions. It would be useful

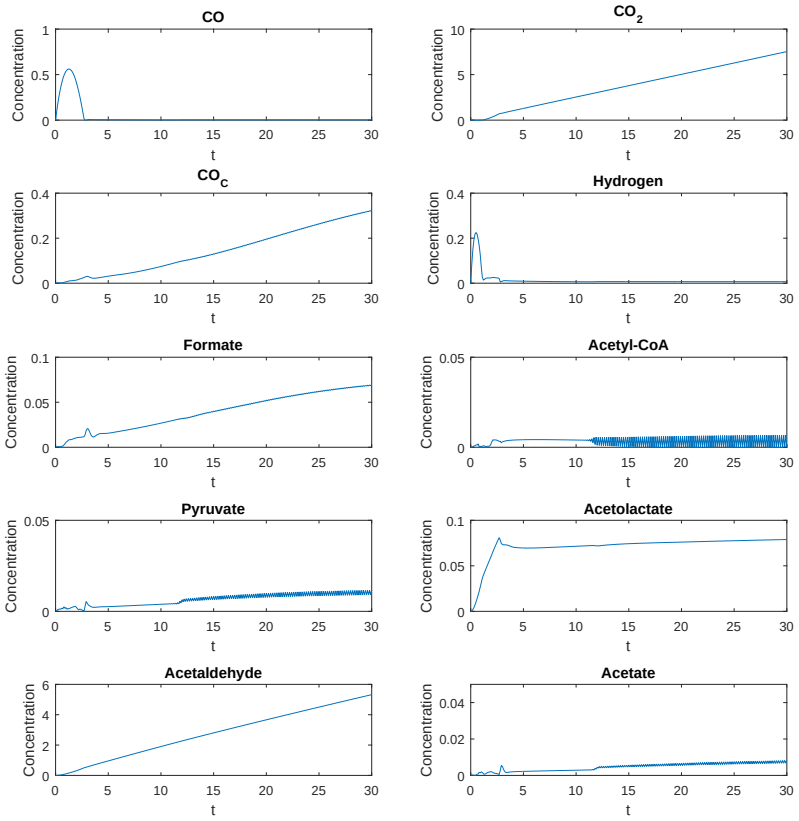

Figure 6.24: Plot of the concentrations of metabolites for the model given in equation (6.5.13), with $H_{in} = 1$ and no constraint on v_3 or v_2.

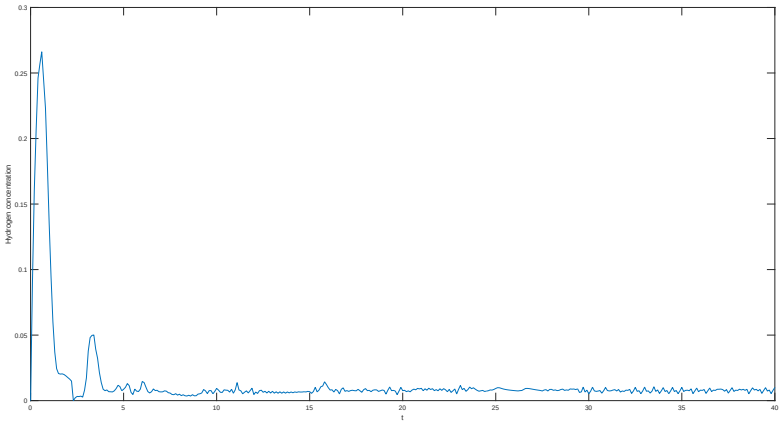

Figure 6.25: Plot of the concentration of hydrogen against time given by our dFBA model (6.5.13), with $H_{in} = 1$ and $v_3 > v_2$.

if we could produce an ODE model that behaves similarly to the dFBA model, and use this as an additional tool for analysis. One method that is able to parameterise

ODE models using experimental data is ensemble modelling. In the next chapter we look to use ensemble models in order to produce an ODE model that exhibits similar behaviour to a dFBA model.

Chapter 7

Ensemble modelling

7.1 Introduction

When modelling metabolism, one of the largest problems we encountered was finding parameter values. Through flux balance analysis and its extensions, we have attempted to circumvent this problem by using techniques that treat fluxes as variables in a linear system that do not rely on additional parameters. In Chapters 2 and 6 we found some success in analysing a metabolic network with these techniques, but we have struggled to observe key behaviours, such as multiple steady-states. If we wish to reliably find such behaviours, we are forced to return to non-linear ODE models. In order to do this we must have some method to choose parameters for our model. Experimental data to estimate these parameters can be difficult to generate, so we try and find methods that use as little data as possible to achieve the most useful possible results. One such technique is known as Ensemble modelling.

Ensemble modelling, as set out by Tran et al. [72] seeks to construct a set (or ensemble) of ODE models that all tend towards the same steady-state. Then, by perturbing the parameters in the model and performing a sensitivity analysis, the models are compared to existing data in order to find the ones that match. In this way, the ensemble of models is screened and filtered to find a smaller set of models that match the data, and hopefully, down to a single model that matches the data.

7.1.1 Building an ensemble

Building this ensemble of models involves two key steps. The first is deciding on a structure for our model, the second is parameterising that model. In its original form, ensemble modelling seeks to expand enzymatic reactions into their elementary steps, and model each step as a reversible reaction using the law of mass action. For example, the reaction

$$A + E \leftrightarrow B + E, \tag{7.1.1}$$

where A and B are metabolites, and E is an enzyme will often be expanded into the set of three reactions

$$A + E \overset{k_1}{\leftrightarrow} AE \overset{k_2}{\leftrightarrow} BE \overset{k_3}{\leftrightarrow} B + E. \tag{7.1.2}$$

Here, A attaches to the enzyme E, is changed into B whilst still attached, and then separates. All of these separate steps are considered reversible. This is not the only possible set of steps for a reaction, but it is a commonly seen one. We assume a fixed amount of enzyme E, T_E, which is considered the sum of the free enzyme and the complexes, i.e. $T_E = E + AE + BE$. The proportion of enzyme in each state is expressed by $e_F = E/T_E$, $e_A = AE/T_E$, and $e_B = BE/T_E$. Then we have a conservation law,

$e_F + e_A + e_B = 1$. We use these proportions as variables in our ODE models, along with our metabolite concentrations c_A and c_B respectively, and form a model of five equations for the reaction described in equation (7.1.2),

$$\frac{dc_A}{dt} = v_{in} - k_1 c_A e_F T_E + k_{-1} e_A T_E, \tag{7.1.3}$$

$$\frac{dc_B}{dt} = k_3 e_B T_E - k_{-3} c_B e_F T_E - v_{out}, \tag{7.1.4}$$

$$\frac{de_F}{dt} = -k_1 c_A e_F T_E + k_{-1} e_A T_E + k_3 e_B T_E - k_{-3} c_B e_F T_E, \tag{7.1.5}$$

$$\frac{de_A}{dt} = k_1 c_A e_F T_E - k_{-1} e_A T_E - k_2 e_A T_E + k_{-2} e_B T_E, \tag{7.1.6}$$

$$\frac{de_B}{dt} = k_2 e_A T_E - k_{-2} e_B T_E - k_3 e_B T_E + k_{-3} c_B e_F T_E, \tag{7.1.7}$$

with some initial conditions

$$\mathbf{c}(0) = \mathbf{c}_0, \mathbf{e}(0) = \mathbf{e}_0, \tag{7.1.8}$$

where $\mathbf{c} = (c_A, c_B)$ and $\mathbf{e} = (e_F, e_A, e_B)$.

Here, v_{in} and v_{out} represent some uptake and output fluxes for A and B respectively. k_n are the rate coefficients for reactions in the forward direction, assumed to be left to right, whilst k_{-n} are the rate coefficients for reactions in the backwards direction. The enzyme proportions have been scaled by a reference value for the total amount of enzyme in the system, and are therefore dimensionless. Assuming we have such data, we also scale the concentrations by some reference state, A_{ref} and B_{ref}, so that these variables also become dimensionless. This reference state is generally taken to be the steady-state the ODE tends to, so that in steady-state, the scaled concentrations, c_A/A_{ref} and c_B/B_{ref} both equal one.

Now that we have formulated a model for our reaction, we must now parameterise it. The key here is we wish to choose parameters that lead the model to tend to some specific steady-state. For this method, we specifically wish to have the reactions rates (fluxes) tending to some specific steady-state, the value of which we have determined from experimental data, or alternatively, using FBA. It is unlikely we would have specific flux data for the forward and backwards reactions (indeed, if we did, this method would be unnecessary), rather we would have some value for the net flux across the reaction, v_{net}. This net flux would be equal to the difference between the forward and backwards reactions, i.e. $v_{net} = v_n - v_{-n}$, where v_n is the flux in the forward direction, and v_{-n} is the flux in the backwards direction. Since each reaction is a chain of steps in sequence, in steady-state the net flux across each step is equal to the net flux across the entire reaction, due to mass balance. Then, in steady-state, we have

$$k_1 c_A e_F - k_{-1} e_A = V_{net}, \tag{7.1.9}$$

$$k_2 e_A - k_{-2} e_B = V_{net}, \tag{7.1.10}$$

$$k_3 e_B - k_{-3} c_B e_F = V_{net}. \tag{7.1.11}$$

Then for a given set of steady-state concentrations, c_n, and enzyme proportions, e_n, we generate sets of parameters, k_n and k_{-n} that satisfy these equations, and lead to models that are included in our ensemble. We also have the condition on our enzyme proportions, $e_A + e_B + e_F = 1$, so we must choose proportions that match these conditions. Finally, we generate parameters, k_n to satisfy the conditions (7.1.9)–(7.1.11). To multiple sets of k_n that lead to the same steady-state, we use the idea of reversibilities.

The reversibility of a reaction is the ratio of the fluxes in the forward and backwards directions. Explicitly, the reversibility, R_v is defined as

$$R_v = \begin{cases} \frac{v_F}{v_B} & \text{if } v_F < v_B \\ \frac{v_B}{v_F} & \text{if } v_B \leq v_F. \end{cases} \tag{7.1.12}$$

Since the flux in the dominant direction is always the denominator, the reversibility takes values between zero (in the case the reaction is completely irreversible), and one (in the case that the reaction happens at equal rates in each direction). In the case of $R_v = 1$, then the net flux across the reaction is equal to zero. In general, we ignore the possibility of R_v being equal to one. We select a range of reversibilities for each step in the reaction, and then use those to generate values for the forwards and backwards fluxes. For a given R_v and positive v_{net}, we find values for the forward and backwards fluxes, v_F and v_B,

$$v_F = \frac{v_{net}}{1 - R_v}, \quad v_B = \frac{v_{net} R_v}{1 - R_v}. \tag{7.1.13}$$

in the case of a negative v_{net}, we instead have

$$v_F = \frac{-v_{net} R_v}{1 - R_v}, \quad v_B = \frac{-v_{net}}{1 - R_v}. \tag{7.1.14}$$

By generating forward and backwards fluxes for a range of reversibilities, we are then able to generate rate parameters, k_n and k_{-n}. With metabolite concentrations c_A in steady-state equal to one, and for a given steady-state set of enzyme proportions, e_n, we calculate our rate coefficients. For example, for the first reaction in (7.1.2), we find that

$$k_1 = \frac{v_F}{e_F}, \quad k_{-1} = \frac{v_B}{e_A}. \tag{7.1.15}$$

We determine the rest of the parameters in the same way. By repeating this step for different reversibility values, we form multiple models that tend to the same steady-state, with different parameters and steady-state enzyme proportions. This set of models is our ensemble.

7.1.2 Generating perturbation data

Now that we have an ensemble of models, we want to narrow them down to find a subset which react to perturbations in the same way as our experimental data. The data we have to compare to is flux values from mutant strains, which either over- or under-express a particular enzyme. We generate similar data for our ensemble by perturbing the models in such a way as to represent this over- or under-expression of a given enzyme. To do this, we introduce new parameters into our ensemble of models. For each enzyme, we include an e_p parameter, representing the new total concentration of enzyme relative to the previous concentration. Specifically, if T_E represents the original total concentration of enzyme, and P_E represents the concentration of enzyme in the perturbed state, we have

$$e_p = \frac{P_E}{T_E}. \tag{7.1.16}$$

For an over-expressed state, this parameter is greater than one, and for an under-expressed state it is less than one. We multiply any terms in our ODEs that rely on this enzyme by our new parameter e_P. For example, equation 7.1.3 would become

$$\frac{dc_A}{dt} = v_{in} - k_1 c_A e_F e_P + k_{-1} e_A e_P. \tag{7.1.17}$$

We see that the v_{in} term, which is assumed to not rely on the enzyme E remains the same, however, the k_1 and k_{-1} terms are multiplied by the parameter, p_e, changing their values. In this way, we are able to represent a perturbed level of flux in the reaction, but our enzyme proportion variables, e_A, e_B, and e_F remain the same, and still adhere to the conservation relationship, $\sum_n e_n = 1$. We construct these perturbations for each flux in the network, and each model in the ensemble. Then, using the steady-state from the unperturbed model (with $c_A = c_B = 1$, and enzyme proportions as prescribed) as our initial condition, we simulate these perturbed ODEs until the concentrations reach a new steady-state. The flux values in these new steady-states should be different for each model in the ensemble, allowing us to compare these new states to the experimental data from mutant strains to see which models in our ensemble match. Generally, this perturbation data is given relative to the original flux. For a net flux in the system v_n, and a given enzyme E_m, let the relative change in the flux be given as $p_{n,m}$, where

$$p_{n,m} = \frac{v_{n,m,perturbed}}{v_{n,m,reference}} \tag{7.1.18}$$

where $v_{n,m,perturbed}$ is the net flux after perturbation, and $v_{n,m,reference}$ is the net flux before the perturbation, used to generate our ensemble model parameters initially.

7.1.3 Screening the ensemble for matches

Having generated the perturbation data, we now have to decide what models count as matching. With experimental data from a real organism, it is highly unlikely that we have a model that matches the data perfectly, but we still wish to narrow down our number of models to those that match closely enough. To do this, we score our data found by perturbing the enzyme levels. Our perturbation data gives the relative value of each net flux after a specific enzyme level has been perturbed, which we have called $p_{n,m}$, where n tells us which reaction in the network we are looking at, and m tells us which enzyme was perturbed. We then score them by how large this change is, and in which direction. Tran et al. [72] used five levels when scoring this data,

$$\text{Score}_{n,m} = \begin{cases} 2, & \text{if } 1.2 < p_{n,m}, \\ 1, & \text{if } 1.05 < p_{n,m} \leq 1.2, \\ 0, & \text{if } 0.95 \leq p_{n,m} \leq 1.05, \\ -1, & \text{if } 0.8 \leq p_{n,m} < 0.95, \\ -2, & \text{if } p_{n,m} < 0.8, \end{cases} \tag{7.1.19}$$

though this is just one possible choice. It is possible to widen or shrink these categories, or include more or less of them, as would be more relevant to specific data. In the case of relatively small fluxes, for example, it may be wise to choose larger category sizes, as a large $p_{n,m}$ may still only relate to a very small change in flux if $v_{n,m,reference}$ is small in equation (7.1.18). We score the experimental data we wish to find a model to match in the same way. Then it is a simple matter of comparing the scores of the experimental data with that from the ensemble of models to pick out those that match. In this way, we reduce the large ensemble down to a much smaller set of possible models. Specifically we hope to obtain a single model from this method, though often we may find ourselves with a small set of models, as in the case of Khazaei et al. [35] who used ensemble modelling to attempt to derive a model for a cancer cell. They were able to reduce their ensemble to a set of four models, using perturbation data for only four enzymes. They found high variation in the some parameters in their four remaining models, suggesting that with more data, some of these models could further be screened out.

Now that we have laid out the steps for ensemble modelling, we start applying it to metabolic networks, including the *C. autoethanogenum* network examined in previous chapters.

7.2 A single step reaction

Let us first consider a single metabolic reaction. We consider the simplest possible reaction, which is a reaction in which A is transformed to B in a single step. We also have an uptake reaction for A and an output reaction for B. The enzyme catalysing this reaction is E. The network for this single reactions is

$$\overset{Q_A}{\rightarrow} A \; ; \; A + E \overset{k_1}{\longleftrightarrow} B + E \; ; \; B \overset{Q_B}{\rightarrow}$$

We form the ODEs to represent this reaction,

$$\frac{dc_A}{dt} = v_{in} - k_1 c_A e_F + k_{-1} c_B e_F, \tag{7.2.1}$$

$$\frac{dc_B}{dt} = k_1 c_A e_F - k_{-1} c_B e_F - v_{out}, \tag{7.2.2}$$

$$\frac{de_F}{dt} = -k_1 c_A e_F + k_1 c_A e_F + k_{-1} c_B e_F - k_{-1} c_B e_F, \tag{7.2.3}$$

with initial conditions $c_A(0) = c_{A(0)}$, $c_B(0) = c_{B(0)}$ and $e_F(0) = e_{F(0)}$. c_A and c_B are the concentrations of A and B respectively, and e_F is the concentrations of free enzyme. Since E is used and produced in both forward and backward reactions, and is never stored in any complexes, its concentration remains constant in time, and does not need an ODE. In fact, since E is a constant, we combine it with the parameters, k_1 and k_{-1}, giving us the simpler model

$$\frac{dc_A}{dt} = v_{in} - k_F c_A + k_B c_B, \tag{7.2.4}$$

$$\frac{dc_B}{dt} = k_F c_A - k_B c_B - v_{out}, \tag{7.2.5}$$

$$\tag{7.2.6}$$

with initial conditions $c_A(0) = c_{A(0)}$ and $c_B(0) = c_{B(0)}$ Here, $k_F = k_1 e_F$ and $k_B = k_{-1} e_F$. Finally, let us specify v_{in} and v_{out}. We let the uptake flux v_{in} be proportional to the internal concentration of A, c_A, and some constant external concentration A^*, and Q_B be proportional to c_B, so that $Q_A = k_{in}(A^* - c_A) = A^{**} - k_{in} c_A$ where $A^{**} = k_{in} A^*$, and $Q_B = k_{out} c_B$. We consider these reactions to be known to be irreversible, and do not generate ensembles that vary the reversibilities of these two reactions so that we are only considering the backwards and forwards k_1 reactions in our ensemble. Then the general form of our model is

$$\frac{dc_A}{dt} = A^{**} - k_{in} c_A - k_F c_A + k_B c_B, \tag{7.2.7}$$

$$\frac{dc_B}{dt} = k_F c_A - k_B c_B - k_{out} c_B, \tag{7.2.8}$$

with initial conditions $c_A(0) = c_{A(0)}$ and $c_B(0) = c_{B(0)}$ For simplicity, let us set all parameters to one, and take initial conditions $c_A(0) = c_B(0) = 0$. A simulation for this is shown in Figure 7.1. Our steady-state concentrations are $c_A^{ss} = 2/3$ and $c_B^{ss} = 1/3$. The net flux across each reaction is equal to $1/3$. It is obvious to see that the net flux across all reactions must be the same in steady-state, due to mass balance.

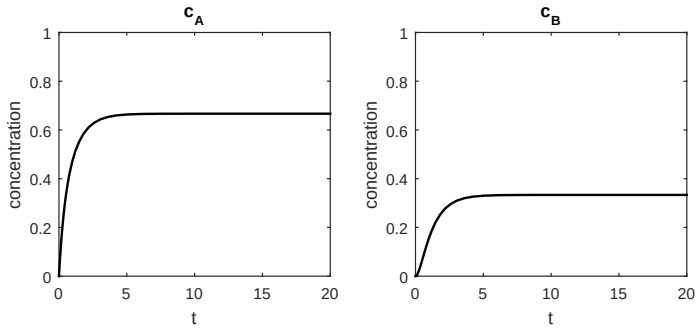

Figure 7.1: Simulation of the ODE model given in equations (7.2.7)–(7.2.8) with all parameters set to equal one and initial concentrations equal to zero.

Reversibility	k_F	k_B
0	0.3333	0
0.1	0.3704	0.0370
0.2	0.4167	0.0833
0.3	0.4762	0.1429
0.4	0.5556	0.2222
0.5	0.6667	0.3333
0.6	0.8333	0.5
0.7	1.1111	0.7778
0.8	1.6667	1.3333
0.9	3.3333	3

Table 7.1: Parameters generated by equation (7.1.13) for the ensemble of models described by equations (7.2.7)–(7.2.8). We see that for each k_B/k_F is equal to the reversibility, and $k_F - k_B$ is equal to the reference net flux, in this case, 1/3.

Let us consider a range of reversibilities for the k_1 reaction. We form an ensemble consisting of 10 models, in which we consider reversibilities between 0 and 0.9 in increments of 0.1. Each model has the same values for k_{in} and k_{out}, where $k_{in} = c_A^{ss} = 2/3$ and $k_{out} = c_B^{ss} = 1/3$, so that the steady-state is attained when $c_A = 1$ and $c_B = 1$. We then use equations (7.1.13) to generate k_F and k_B, which is also scaled by the relevant c_n^{ss}, for our range of reversibilities. These parameters are shown in Table 7.1. Every model with these parameters has the same steady-state, and in each case the net flux across all reactions is 1/3.

We now generate the perturbation data from our models. For now, we only perturb the k_1 reaction. Since we have lumped the (constant) enzyme concentration in with the rate coefficients k_F and k_B, in order to simulate a change in the enzyme levels, we multiply both k_F and k_B by the same factor. We simulate an increase in enzyme levels by doubling the rate coefficients, to simulate twice as much available enzyme. After doubling k_F and k_B, we simulate the ODEs for a time period long enough for the model to tend to a new steady-state. We then find the net flux values for this new steady-state and divide through by the original flux values, to find the relative change p_n, given by equation (7.1.18), in a given flux v_n. In particular, we find relative change for v_1 using the equation $p_1 = \frac{(k_F c_A - k_B c_B)}{1/3}$, where k_F and k_B are the perturbed parameters, and c_A and c_B are the new steady-state concentrations. Table 7.2 gives the relative change

Reversibility	Relative change, p
0	1.5
0.1	1.4286
0.2	1.3636
0.3	1.3043
0.4	1.25
0.5	1.2
0.6	1.1538
0.7	1.1111
0.8	1.0714
0.9	1.0345

Table 7.2: Perturbation data for the single step reaction, generated by using the parameters in 7.1, and doubling them to simulate an increase in enzyme concentration. The ODE system seen in equations (7.2.7)–(7.2.8) is then run to find a new steady-state, and the relative change in flux for v_1 is found.

in net flux for the k_1 reaction for each model in our ensemble.

We apply the same perturbation to our original ODE model, equations (7.2.7)–(7.2.8), which has parameters set to equal one. We double k_F and k_B, so that we have the parameter set $A^{**} = 1$, $k_{in} = 1$, $k_F = 2$, $k_B = 2$, and $k_{out} = 1$. A simulation with these parameters, and initial conditions set to equal zero is shown in Figure 7.2. We see that the new steady-state concentrations are $c_A = 0.6$ and $c_B = 0.4$. The new

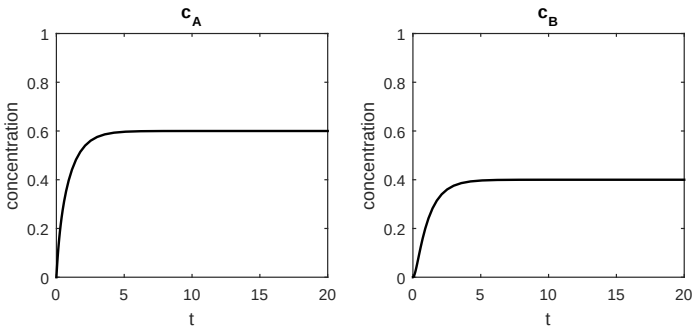

Figure 7.2: Simulation of the ODE model given in equations (7.2.7)–(7.2.8) after perturbation, so that $A^{**} = k_{in} = k_{out} = 1$, and $k_F = k_B = 2$ with initial concentrations set to equal zero.

relative change across the k_1 reaction is $p_1 = 2c_A - 2c_B = 0.4$. Dividing through by the original net flux of 1/3 gives us a relative change of 1.2. We see immediately from Table 7.2 that this corresponds to a reversibility of 0.5. We could rank our data as described in Section 7.1.3, but since we have already found an exact match for the data, this is unnecessary. From Table 7.1, we see the scaled parameters that relate to this reversibility are $k_F = 0.6667$ and $k_B = 0.3333$. To find the true unscaled parameters, we divide through by the original steady-state concentrations for $c_A^{ss} = 2/3$ and $c_B^{ss} = 1/3$, so that $k_F = 1$ and $k_B = 1$ as originally prescribed. Using ensemble modelling, we have been able to recover the original parameters used in the model for a single step reaction.

7.3 Matching a full metabolic model

7.3.1 A fully irreversible model

Now we turn our attention to a full metabolic network. For simplicity we continue to assume all reactions are single-step, meaning that free enzyme concentrations continue to be constant, and therefore the enzyme concentrations are incorporated into rate coefficients. This conveniently allows us to refer directly to the ODE model for *C. autoethanogenum* described in Chapter 4. First we examine the network in the case that all reactions are irreversible, and later we add additional reactions to our ODE model to simulate reversibility.

The ODE model we use remains unchanged from Chapter 4, consisting of ten equations, one for each important concentration, using the law of mass action to model reaction rates with constant input values for carbon monoxide and hydrogen. The model is

$$\frac{dc_{CO}}{dt} = Q_{CO} - k_1 c_{CO}, \tag{7.3.1}$$

$$\begin{aligned}\frac{dc_{CO_2}}{dt} &= k_1 c_{CO} - k_2 c_{CO_2} {c_H}^2 - k_3 c_{CO_2} {c_H}^2 \\ &\quad - k_5 c_{CO_2} c_{ACA} {c_H}^2 + k_7 {c_{Py}}^2 + k_{18} c_{Al} {c_H}^2,\end{aligned} \tag{7.3.2}$$

$$\frac{dc_{CO_c}}{dt} = k_3 c_{CO_2} {c_H}^2 - k_4 c_{CO_c} c_F {c_H}^4, \tag{7.3.3}$$

$$\frac{dc_F}{dt} = k_2 c_{CO_2} {c_H}^2 - k_4 c_{CO_c} c_F {c_H}^4 - k_6 c_F c_{ACA}, \tag{7.3.4}$$

$$\begin{aligned}\frac{dc_{ACA}}{dt} &= k_4 c_{CO_c} c_F {c_H}^4 - k_5 c_{CO_2} c_{ACA} {c_H}^2 - k_6 c_F c_{ACA} \\ &\quad - k_8 c_{ACA} {c_H}^2 - k_9 c_{ACA} - k_{15} c_{ACA} c_{Py},\end{aligned} \tag{7.3.5}$$

$$\frac{dc_{Py}}{dt} = k_5 c_{CO_2} c_{ACA} {c_H}^2 + k_6 c_F c_{ACA} - 2k_7 {c_{Py}}^2 - k_{15} c_{ACA} c_{Py} - k_{16} c_{Py} {c_H}^2, \tag{7.3.6}$$

$$\frac{dc_{Al}}{dt} = k_7 {c_{Py}}^2 - k_{17} c_{Al} - k_{18} c_{Al} {c_H}^2, \tag{7.3.7}$$

$$\frac{dc_{Ad}}{dt} = k_8 c_{ACA} {c_H}^2 + k_{10} c_A - k_{19} c_{Ad} {c_H}^2, \tag{7.3.8}$$

$$\frac{dc_A}{dt} = k_9 c_{ACA} - k_{10} c_A - k_{20} c_A, \tag{7.3.9}$$

$$\begin{aligned}\frac{dc_H}{dt} &= Q_H + 2k_1 c_{CO} - 2k_2 c_{CO_2} {c_H}^2 - 2k_3 c_{CO_2} {c_H}^2 \\ &\quad - 4k_4 c_{CO_c} c_F {c_H}^4 - 2k_5 c_{CO_2} c_{ACA} {c_H}^2 - 2k_8 c_{ACA} {c_H}^2 - 2k_{16} c_{Py} {c_H}^2 \\ &\quad - 2k_{18} c_{Al} {c_H}^2 - 2k_{19} c_{Ad} {c_H}^2.\end{aligned} \tag{7.3.10}$$

with initial conditions $\mathbf{c}(0) = \mathbf{c}_0$. We set all parameters to equal one, except $Q_H = 3$, and $k_3 = 0.8$. This is so that we know the ODE model tends to a steady-state, which is required for ensemble modelling. A simulation of the model with these parameters is shown in Figure 7.3. From this simulation we are able to draw the two pieces of information we need, the steady-state concentrations, given in Table 7.3, and the flux values calculated from these concentrations, given in Table 7.4.

We generate the set of ODE models that tend to this steady-state. We do not have any thermodynamic data to bound the reversibilities of our reactions, so instead we choose a spread of reversibilities for each reversible reaction. We choose four different levels for each reaction, 0, 0.25, 0.5 and 0.75 to cover a wide range of possibilities, from completely irreversible ($R = 0$), to highly reversible ($R = 0.75$). We do not examine

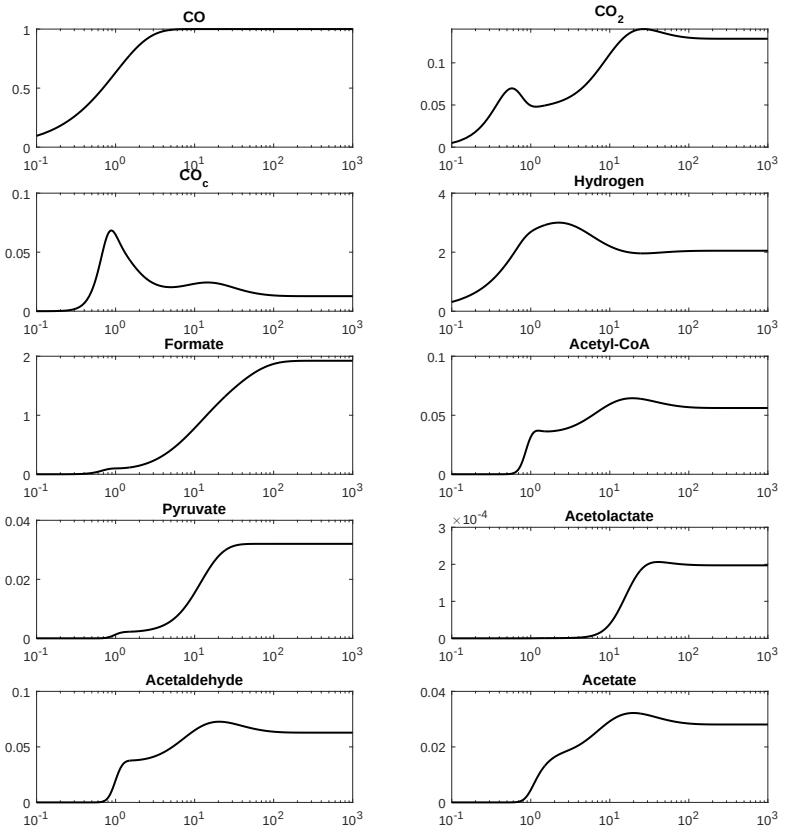

Figure 7.3: Simulation of the ODE model given in equations (7.3.1)–(7.3.10) with all parameters set to equal one, except $k_3 = 0.8$ and $Q_H = 3$ and initial concentrations of zero.

reversibilities of $R = 1$, as this would indicate a completely reversible reaction, which has a net flux of zero, meaning the reaction has no effect at steady-state. If we considered every reaction to be potentially reversible, we would have 4^{18} different models, which is of the order of 10^9 models. This would be prohibitively expensive to compute, so instead we assume some of the reactions in the model are always irreversible. Since the input reactions Q_H and Q_{CO} are constants in the model, we consider these two to be irreversible. We also consider all output reactions to be irreversible. This limits us to only considering ten reactions. This still gives us over 10^6 models to consider, so we would like to reduce this further still. The reactions producing carbon monoxide and formate from carbon dioxide often dictate whether or not the model is able to achieve steady-state. It is sensible to consider these two reactions as irreversible, as if we change the rate parameters for those equations k_2 and k_3, we may find the ODE

c_{CO}	c_{CO_2}	c_{CO_c}	c_F	c_{ACA}	c_{Py}	c_{Al}	c_{Ad}	c_A	c_H
1	0.1286	0.0127	1.9231	0.0561	0.032	0.0002	0.0628	0.0281	2.0488

Table 7.3: Steady-state concentrations for the simulation seen in Figure 7.3.

v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}
1	0.5398	0.4318	0.4318	0.0303	0.1080	0.0010	0.2356	0.0561	0.0281
Q_{CO}	Q_H	out_{bio}	out_{lac}	out_{val}	out_{but}	out_{eth}	out_{ace}		
1	3	0.0018	0.1344	0.0002	0.0008	0.2637	0.0281		

Table 7.4: Steady-state fluxes calculated from the concentrations from Table 7.3.

no longer tending to steady-state, which limits the effectiveness of ensemble modelling. Similarly, if we were to add a backwards reaction for the reaction in which carbon monoxide is changed into carbon dioxide (k_1), it would take the same form as the forward reactions for k_2 and k_3, which could lead to the same problems with the models not tending to steady-state. We therefore consider all three of these reactions to also be irreversible. Then we have only seven reactions remaining, represented by the rate coefficients k_4–k_{10}, which we vary the reversibilities of, leading to a large, but much more manageable, ensemble consisting of 4^7 models. Thus we write the general form of a model in our ensemble

$$\frac{dc_{CO}}{dt} = Q_{CO} - k_1 c_{CO}, \tag{7.3.11}$$

$$\frac{dc_{CO_2}}{dt} = k_1 c_{CO} - k_2 c_{CO_2} {c_H}^2 - k_3 c_{CO_2} {c_H}^2 - k_5 c_{CO_2} c_{ACA} {c_H}^2 + k_{-5} c_{Py} + k_7 {c_{Py}}^2 - k_{-7} c_{CO_2} c_{Al} + k_{18} c_{Al} {c_H}^2, \tag{7.3.12}$$

$$\frac{dc_{CO_c}}{dt} = k_3 c_{CO_2} {c_H}^2 - k_4 c_{CO_c} c_F {c_H}^4 + k_{-4} c_{ACA}, \tag{7.3.13}$$

$$\frac{dc_F}{dt} = k_2 c_{CO_2} {c_H}^2 - k_4 c_{CO_c} c_F {c_H}^4 + k_{-4} c_{ACA} - k_6 c_F c_{ACA} + k_{-6} c_{Py}, \tag{7.3.14}$$

$$\frac{dc_{ACA}}{dt} = k_4 c_{CO_c} c_F {c_H}^4 - k_{-4} c_{ACA} - k_5 c_{CO_2} c_{ACA} {c_H}^2 + k_{-5} c_{Py} - k_6 c_F c_{ACA} + k_{-6} c_{Py} - k_8 c_{ACA} {c_H}^2 + k_{-8} c_{Ad} - k_9 c_{ACA} + k_{-9} c_A - k_{15} c_{ACA} c_{Py}, \tag{7.3.15}$$

$$\frac{dc_{Py}}{dt} = k_5 c_{CO_2} c_{ACA} {c_H}^2 - k_{-5} c_{Py} + k_6 c_F c_{ACA} - k_{-6} c_{Py} - 2k_7 {c_{Py}}^2 + 2k_{-7} c_{CO_2} c_{Al} - k_{15} c_{ACA} c_{Py} - k_{16} c_{Py} {c_H}^2, \tag{7.3.16}$$

$$\frac{dc_{Al}}{dt} = k_7 {c_{Py}}^2 - k_{-7} c_{CO_2} c_{Al} - k_{17} c_{Al} - k_{18} c_{Al} {c_H}^2, \tag{7.3.17}$$

$$\frac{dc_{Ad}}{dt} = k_8 c_{ACA} {c_H}^2 - k_{-8} c_{Ad} + k_{10} c_A - k_{-10} c_{Ad} - k_{19} c_{Ad} {c_H}^2, \tag{7.3.18}$$

$$\frac{dc_A}{dt} = k_9 c_{ACA} - k_{-9} c_A - k_{10} c_A + k_{-10} c_{Ad} - k_{20} c_A, \tag{7.3.19}$$

$$\frac{dc_H}{dt} = Q_H + 2k_1 c_{CO} - 2k_2 c_{CO_2} {c_H}^2 - 2k_3 c_{CO_2} {c_H}^2 - 4k_4 c_{CO_c} c_F {c_H}^4 + 4k_{-4} c_{ACA} - 2k_5 c_{CO_2} c_{ACA} {c_H}^2 + 2k_{-5} c_{Py} - 2k_8 c_{ACA} {c_H}^2 + 2k_{-8} c_{Ad} - 2k_{16} c_{Py} {c_H}^2 - 2k_{18} c_{Al} {c_H}^2 - 2k_{19} c_{Ad} {c_H}^2. \tag{7.3.20}$$

with the initial conditions $\mathbf{c}(0) = \mathbf{c}_0$ Here, the k_n represent the forward rate coefficients, and the k_{-n} represent the backwards rate coefficients. These are generated in pairs using the reversibility values. The k_n rate coefficients without a matching k_{-n} are the rate coefficients for irreversible reactions. These are generated from the flux and concentration data provided from our original ODE models steady-state. If all the

k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}
1	0.5398	0.4318	0.4318	0.0303	0.1080	0.0010	0.2356	0.0561	0.0281
Q_{CO}	Q_H	k_{15}	k_{16}	k_{17}	k_{18}	k_{19}	k_{20}		
1	3	0.0018	0.1344	0.0002	0.0008	0.2637	0.0281		

Table 7.5: Parameter values for the ODEs (7.3.11)–(7.3.20) in our ensemble in which all reactions are fully irreversible.

reversibilities are zero ($k_{-n} = 0 \quad \forall n$) the model reduces to (7.3.1)–(7.3.10), which we use to generate the data. Therefore, there is at least one model in our ensemble that matches the data we use to screen the ensembles. We wish to see if only this model is found, or if there are others that match the data.

Now that we have our ensemble of models, we generate the perturbation data. As in previous examples, this is done by perturbing enzyme levels in the ensemble of ODEs. As in our single step reaction in Section 7.2, we consider the enzyme concentrations to be combined with the rate coefficients, so by increasing the forward and backward rate coefficients, we simulate an increase in enzyme. For this example, we simulate doubling the amount of free enzyme for each reaction in turn, by doubling the k_n parameter for that reaction and, for the reversible reactions, doubling the k_{-n} at the same time.

Compared to our single reaction from Section 7.2, this *C. autoethanogenum* model has many more ensembles and reactions, so we score the perturbation data, data, $p_{n,m}$, as described in Section 7.1.3. We use the scoring described in equation (7.1.19) for this. Then all of our scored data, $\text{Score}_{n,m}$, take values between -2 and 2, describing how large of a change they have undergone.

Since the model we aim to match is the case in which all reactions are irreversible, we have already generated the perturbation data for our ODE model as part of the ensemble. We extract the $p_{n,m}$ for this model, convert it to a $\text{Score}_{n,m}$ as described in equations (7.1.19) and compare it to the data for all other models in the ensemble to see if any others match. Since we have the $p_{n,m}$ values for all fluxes, n, we include them all in the comparison. In this case, with this full data set, there are no other models in our ensemble that match the behaviour of the irreversible case. This shows the scored perturbation data, $\text{Score}_{n,m}$, allows us to extract a single model out of many in the ensemble that tend to the same steady-state. The forward parameters of the matched model in our ensemble are given in Table 7.5, whilst all backwards parameters are equal to zero. We note that these parameters from Table 7.5 do not match those used in our initial ODE model, but they are equal to the flux values given in Table 7.4. By dividing these values by the relevant concentrations from Table 7.3, we obtain exactly the parameters used in our ODE model. For example, by dividing k_8 k_8 by $c_{ACA}{c_H}^2$, we recover the original parameters used, namely $k_8 = 1$. Doing this to all parameters from Table 7.5) we obtain exactly the original used in our ODE model, with all equal to one, except $k_3 = 0.8$, and $Q_H = 3$.

With perturbation data for all fluxes, we have been able to find exactly the parameters used in our ODE model. However, what if we cannot obtain data for all the fluxes? In some cases, we may not have experimental data for all of our internal fluxes, as these often prove difficult to find. In these cases, it may be that we have data for certain fluxes, such as the outputs. Interestingly, in the case where we screen using only the perturbation data for the output fluxes, i.e. using $p_{n,m}$, for $n = 15$–20, we are unable to narrow down the ensemble to a single model. Specifically, we reduce the ensemble down to a set of sixteen models. Every model in this set has the same reversibilities for v_5, v_7, v_8, v_9, and v_{10}, however, for v_4 and v_6, all reversibilities remain in the set. If we

c_{CO}	c_{CO_2}	c_{CO_c}	c_F	c_{ACA}	c_{Py}	c_{Al}	c_{Ad}	c_A	c_H
1	0.1215	0.0138	1.5909	0.0674	0.0316	0.0002	0.0591	0.0225	2.1011

Table 7.6: Steady-state concentrations for the ODE model with reversible reactions, seen in 7.4.

only have the perturbation data for the reactions that output a product, we are, in this case, unable to find any bounds on the reversibilities for v_4 and v_6. It is worth noting from this, that when using this method, it is entirely possible that it is not possible to narrow down the ensemble to a single model. In future, if we were to only use the perturbation data for the outputs on this model, it would not be worth generating models for each level of reversibility for k_4 and k_6, as we are not able to determine their reversibility this way. By doing this, we could reduce the number of models we have to test by a factor of sixteen, which would save us computation time, though we should remain aware that these reactions could still take any level of reversibility, and if we have additional perturbation data, we should still include them, as it may be possible to eliminate some of these models from the ensemble. For the rest of the chapter, we only consider the perturbation data for the outputs, and reduce our ensemble of models, so that we only consider the reversibilities of v_5, v_7, v_8, v_9, and v_{10}.

We have now examined a case where all reactions are irreversible, however we should also examine a case when some of the reactions in our ODE are reversible. It would also be useful to consider a case when the exact model we use is not in the ensemble. If the true model we are trying to match is not also contained exactly within the ensemble, there may not be a model that matches our perturbation data exactly, so we should find some way of determining the closest fit model.

7.3.2 A model with reversible reactions

The general form of our ensemble models, given in equations (7.3.11)–(7.3.20) is useful for this purpose, as it already has backwards reactions included. By choosing a set of parameters for this network, including some non-zero backwards parameters, we form a new network with some reversible reactions to match our ensemble data to. We keep the same rate coefficients as in the completely irreversible case, with all forward parameters equal to one, except $k_3 = 0.8$, and $Q_H = 3$. Choosing the k_8 and k_9 reactions as our reversible reactions, we set the backwards rate coefficients k_{-8} and k_{-9} to equal one, with all other k_{-n} set to equal zero. This model is our reference model. We simulate this reference model with the described parameters to find a new steady-state to construct an ensemble of models from. The results of this simulation are seen in Figure 7.4. We again use the concentrations of metabolites in these steady-states, seen in Table 7.6, to generate the steady-state fluxes. Since we use the net flux across a reaction for ensemble modelling, we note that $v_8 = k_8 c_{ACA} {c_H}^2 - k_{-8} c_{Ad}$ and $v_9 = k_9 c_{ACA} - k_{-9} c_A$. The reversibility of a reaction is calculated from the values of these forward and backwards fluxes too, and are worth noting now for later comparison. Most reactions are still irreversible and hence have a reversibility of zero. The reversibilities of v_8 and v_9 are

$$R_{v_8} = \frac{k_{-8} c_{Ad}}{k_8 c_{ACA} {c_H}^2} = 0.1986, \quad R_{v_9} = \frac{k_{-9} c_A}{k_9 c_{ACA}} = 0.3333. \tag{7.3.21}$$

We note that by using the same four levels of reversibility (0, 0.25, 0.5, and 0.75) to generate our ensemble, there is not an exact match in the set of models. This allows

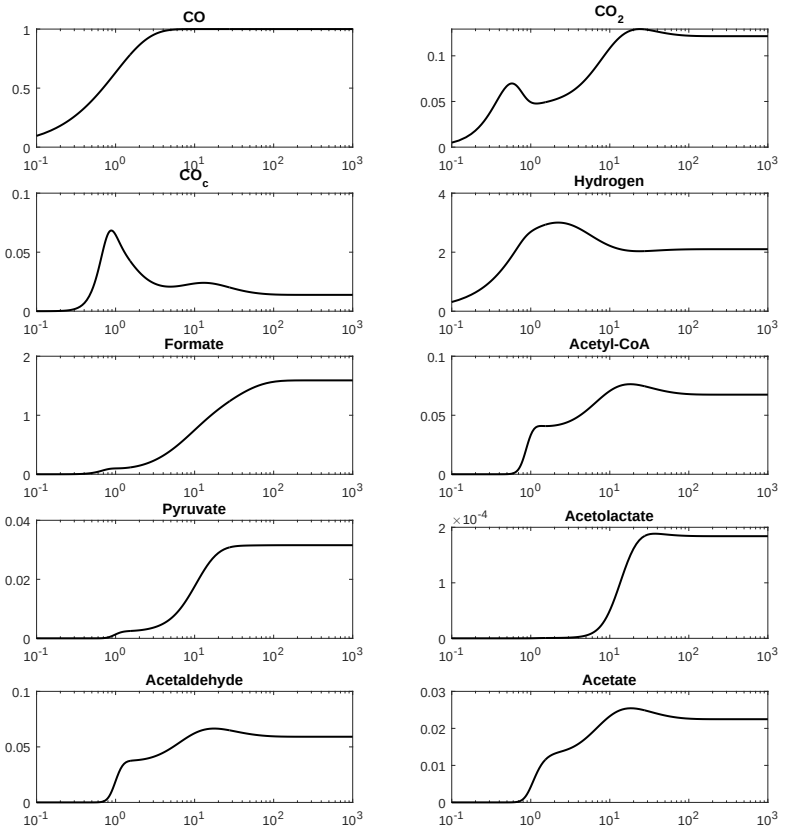

Figure 7.4: Simulation of the ODE model (7.3.11)–(7.3.20) with reversible reactions k_8 and k_9. All forward parameters are set to equal one, except $k_3 = 0.8$ and $Q_H = 3$. We also have non-zero backwards parameters, $k_{-8} = 1$ and $k_{-9} = 1$.

us to see if the method is able to find close models that match the perturbation data, or if there is not a model that matches exactly.

As usual, we generate our ensemble of models, and then perform a sensitivity analysis by perturbing the parameters in the model one reaction at a time to generate our $p_{n,m}$ values. We generate the perturbation data, $p_{n,m}$ for our reference model, convert them to $\text{Score}_{n,m}$ values, and compare these results with those from our ensembles to find a model from the ensemble that matches. In this instance we find that no model in our ensemble is able to match the scored perturbation data from our reference model exactly. There are two possibilities for why an exact match can't be found. The first is the structure we are using to generate our ensemble of models is incorrect. Since our reference model follows this structure exactly, we know this is not the case here. The other possibility is that we simply don't have a model in the ensemble close enough to match this scored perturbation data, which is true in this case. This could be fixed by generating more models in the ensemble for different levels of reversibility. This is not desirable, as we already have a large ensemble to examine. Rather than generating many more models, most of which we already know do not match, we should look in our current ensemble the find the "closest" fitting models. From these models we seek to gain some understanding of approximate levels of reversibility for the reactions, and use this information to generate a more targeted ensemble.

We must now decide on what constitutes a "close" fitting model. There are a few options here. First we could consider the raw perturbation data, $p_{n,m}$. We find the absolute differences between the perturbation data for each model in the ensemble, and the perturbation data from the reference model. Taking $p^{ref}_{n,m}$ to be the perturbation data for our reference model, and $p^{x}_{n,m}$ to be the perturbation data for the xth model in the ensemble, we let

$$E^{(1)}(x) = \sum_n \sum_m |p^x_{n,m} - p^{ref}_{n,m}|, \tag{7.3.22}$$

and then choose the model x that minimises $E^{(1)}(x)$. For our data, this gives a single model where this minimum $E(x)$ is equal to 1.1014. This model has reversibilities of 0.25 for both v_8 and v_9, with all other reversibilities equal to zero. We see by comparing to the true reversibilities seen in equation (7.3.21) that this approximate model from the ensemble is close. Indeed, in terms of absolute difference, these are the closest reversibilities that could be chosen from our ensemble.

Rather than using the pure perturbation data, we could instead use the scored data, $\text{Score}_{n,m}$ we have been using to eliminate models from the ensemble. We calculate a similar sum to that seen in equation (7.3.22), only using $\text{Score}_{n,m}$ rather than $p_{n,m}$. Then we generate the sum for each model,

$$E^{(2)}(x) = \sum_n \sum_m |\text{Score}^x_{n,m} - \text{Score}^{ref}_{n,m}|, \tag{7.3.23}$$

and select the model x that minimises $E^{(2)}(x)$. For our data, we are able to find a model where $E(x) = 1$. The minimum value being one means there is a model in the ensemble that only differs from the reference data for a single metabolite in a single perturbation, and this difference is only at a level of one. Interestingly, the model x that minimises $E^{(2)}(x)$ is not the same model that minimises $E^{(1)}(x)$. The model that minimises $E^{(2)}(x)$ has reversibilities of zero for v_5 and v_7, with reversibilities of 0.25 for v_8, v_9, and v_{10}. This model is further away from the reference model, as it sets v_{10} to be reversible, whereas in the reference model it is not. The model that minimises the raw data has a sum of two when looking at the levelled data. It has a second perturbation where it differs greatly enough from the reference data to take a different level. The difference in results suggests that whilst the model that minimises $E^{(1)}(x)$ may vary less from the reference data overall, it has more large $\text{Score}_{n,m}$ values that cause the ensemble modelling process to screen it out.

This process has found two potential closest models, which are similar, differing in only a single reaction. We could repeat the entire ensemble modelling process using a more targeted set of reversibilities, rather than some that evenly span the entire possible space between zero and one. In doing so we would hope to find a model, or set of models, that match the data exactly. If we were to then still find no model that matched exactly, we could again find a closest fitting model, and create a new set of reversibilities around that model, and repeat the process more times.

7.4 Matching a dFBA model

It is a trivial matter to match an ODE model which we know has the same structure as those generated during our ensemble modelling, as we are always able to find a model to match by refining our choice of reversibilities. In reality, however, our data will not come from such clear sources. It is worth examining data from a source which does not have a fixed ODE structure. In Chapter 6 we examined dynamic flux balance analysis (dFBA) as a technique for modelling dynamic behaviour in metabolic networks.

v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}
1	0.6072	0.4843	0.4843	0.0807	0.1229	0.0930	0.0874	0.1933	0.1876
Q_{CO}	Q_H	out_{bio}	out_{lac}	out_{val}	out_{but}	out_{eth}	out_{ace}		
1	3.2	0.0000	0.0176	0.0138	0.0792	0.2750	0.0057		

Table 7.7: Steady-state fluxes for the dFBA model simulated in Figure 7.5.

Whilst this technique uses a set of ODEs for modelling concentrations in a metabolic network, the ODEs do not follow a fixed mass action structure as seen in our ensemble modelling. The dFBA model should also produce dynamics that are in some sense optimal (according to a prescribed objective function), so it would be interesting to see if we could generate an ODE model from the data that behaves in this optimal manner. In particular, the model generated from this method could be considered to behave optimally in the presence of the small perturbations in parameter values used in the sensitivity analysis that screens the ensemble.

The first step is to run a dFBA simulation to steady-state, and then generate a set of ODE models that tend towards the same steady-state. To do this, we return to the dFBA model in Section 6.5. The general dFBA model for the *C. autoethanogenum* model is

$$\begin{aligned}
\max_{\mathbf{c},\mathbf{v}} \quad \mathbf{a}^T\mathbf{v} + \mathbf{b}^T\mathbf{c} \quad \text{such that } \frac{d\mathbf{c}}{dt} &= S\mathbf{v}, \\
Q_H &= H_{in}, \\
Q_{CO} &= CO_{in}, \\
out_{bio} &\leq k_{15}c_{ACA}c_{Py}, \\
out_{lac} &\leq k_{16}c_{Py}c_H^2, \\
out_{val} &\leq k_{17}c_{Al}, \\
out_{but} &\leq k_{18}c_{Al}c_H^2, \\
out_{eth} &\leq k_{19}c_{Ad}c_H^2, \\
out_{ace} &\leq k_{20}c_A, \\
\mathbf{c} &\geq \mathbf{0}, \\
\mathbf{v} &\geq \mathbf{0}, \\
-\mathbf{v}_{dt-max} \leq \frac{d\mathbf{v}}{dt} &\leq \mathbf{v}_{dt-max}, \\
\mathbf{c}(0) = \mathbf{c}_0,\ \mathbf{v}(0) &= \mathbf{v}_0. \qquad (7.4.1)
\end{aligned}$$

Here S is the stoichiometric matrix defined by equations (6.5.1)–(6.5.3), $\mathbf{v}$ is the vector of fluxes, and $\mathbf{c}$ is the vector of concentrations. In order for this model to tend to a steady-state, we must be careful in our choice of parameters. We choose $H_{in} = 3.2$, $CO_{in} = 1$, all $k_n = 1$, and all elements of $\mathbf{v}_{dt-max}$ equal to one. We also choose our objective function to be maximising the production of biomass, i.e. maximising the flux, out_{bio}. Plots of the concentrations for this simulation are shown in Figure 7.5. This model also directly gives values for the net fluxes which we need to generate the parameters for our ensemble of models. The net flux values in this steady-state are given in Table 7.7 From these flux values we generate the parameters for our ensemble of ODE models. Since this is the same metabolic network as our ODE models, we construct the ensemble in the same way. We consider v_1, v_2, and v_3 to be irreversible reactions, to prevent the system from switching to non-steady-state behaviour. The

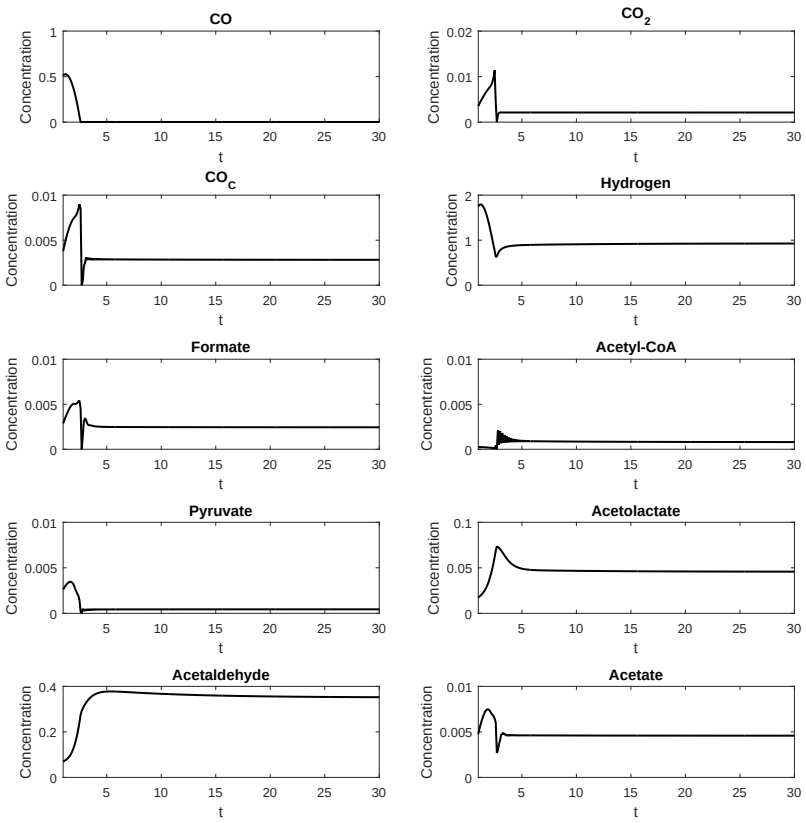

Figure 7.5: Simulation of the dFBA model described in equation (7.4.1). The objective function for this simulation is maximising biomass production. All k_n parameters are set to equal one, with $CO_{in} = 1$, and $H_{in} = 3$, initial concentrations equal to zero, and initial fluxes equal to zero except $Q_{CO}(0) = 1$ and $Q_H(0) = 3$.

output fluxes, out_n are fixed to be irreversible by the dFBA model. As in Section 7.3.2, we only consider the scored perturbation data, $\text{Score}_{n,m}$ for the output fluxes so we are unable to determine the reversibility of v_4 and v_6, and do not include reversibilities for these reactions. By considering the reversibilities of v_5, v_7, v_8, v_9, and v_{10} at four levels, 0, 0.25, 0.5 and 0.75, we generate an ensemble of $4^5 = 1024$ models. We generate the perturbation data for this ensemble using the same method as usual, but generating perturbations in the dFBA model is more difficult.

Most fluxes in the dFBA model do not have associated rate parameters. We could perturb the model by simply increasing the levels of a specific flux, but there is little to stop the model from simply reducing the flux back to its initial level and eventually settling to the same steady-state. Indeed, since the dFBA steady-state is optimal according to its objective function, in most cases it returns to its original state in the presence of perturbations. In some cases, these perturbations could lead to a different state, such as the Lotka-Volterra dFBA model examined in Section 6.4, in which perturbing the network from its optimal state lead to the extinction state, but in general we cannot rely on this. It would also not be possible impose useful restrictions on the perturbed flux, to prevent the model returning to its original state, without also impos-

	Perturbed parameter			
Flux	k_{15}	k_{18}	k_{19}	k_{20}
out_{bio}	2	0.5775	0.9859	0.9197
out_{lac}	1	0.2333	0.8787	1.1988
out_{val}	1	1.2791	1.0297	0.7684
out_{but}	1	0.9379	0.9267	1.0121
out_{eth}	1	1.0749	1.0527	0.9844
out_{ace}	1	0.9815	0.8743	1.8914

Table 7.8: Perturbation data for the dFBA model, (7.4.1). Output fluxes were perturbed by doubling the flux and associated parameter, then the dFBA model was simulated until it reached a new steady-state.

	Perturbed parameter			
Flux	k_{15}	k_{18}	k_{19}	k_{20}
out_{bio}	2	-2	0	-1
out_{lac}	0	-2	-1	1
out_{val}	0	2	0	-2
out_{but}	0	-1	-1	0
out_{eth}	0	1	1	0
out_{ace}	0	0	-1	2

Table 7.9: Perturbation data from Table 7.8 ranked according to equations (7.1.19)

ing a specific reversibility on it, which we do not wish to do. Instead, we perturb the parameters for the output fluxes, which are prescribed by the model. These reactions are assumed to be irreversible. As in Section 7.2, we assume the enzyme concentrations are combined with the rate coefficient, so we simulate an increase in enzyme levels by doubling these rate coefficients one at a time. We also double the relevant flux in the initial condition for dFBA, and run the simulation until it reaches a new steady-state. The steady-state flux values after perturbation divided by steady-state flux values prior to perturbation, $p_{n,m}$ for the output fluxes for each perturbed parameter are given in Table 7.8. Two perturbations out of six (namely k_{16} and k_{17}) cause the dFBA model to stop tending to a steady-state, instead moving to a linear growth state, and therefore cannot be used for screening the ensemble of models. This reduces us to only consider four perturbations. We note that out_{lac} undergoes a relatively large decrease when k_{18} is increased, suggesting that we should score our data as described in equation (7.1.19) so as to not specifically search for models from our ensemble with such a large decrease. Table 7.9 shows the data after this scoring. Of these 4 sets of ranked perturbation data, every model in our ensemble matches the set for the increase of k_{15}, 8 models match the data for k_{19}, and no models exactly match the data for k_{18} and k_{20}. The reversibilities of the eight models that match the perturbation data for k_{19} are shown in Table 7.10. Since no model matches all our data exactly, we need to find some model that matches our observed behaviour closest. It would be desirable for it to be one of the models that matches the data for k_{19}, so we restrict ourselves to choosing one of these models. We use the two methods described in Section 7.3.2 to find a closest fitting model. First, we find the model that minimizes the difference in the scored data, $\text{Score}_{n,m}$, between

	Flux				
Model number	v_5	v_7	v_8	v_9	v_{10}
683	0.5	0.5	0.5	0.5	0.5
747	0.5	0.75	0.5	0.5	0.5
767	0.5	0.75	0.75	0.75	0.5
939	0.75	0.5	0.5	0.5	0.5
942	0.75	0.5	0.5	0.75	0.25
959	0.75	0.5	0.75	0.75	0.5
1003	0.75	0.75	0.5	0.5	0.5
1023	0.75	0.75	0.75	0.75	0.5

Table 7.10: Reversibilities of models from the ensemble that match the perturbation data for k_{19} seen in Table 7.9

k_1	k_2	k_3	k_4	k_{-4}	k_5	k_{-5}	k_6	k_{-6}	k_7
1	0.6072	0.4843	0.4843	0	0.1614	0.0807	0.1229	0	0.3719
k_{-7}	k_8	k_{-8}	k_9	k_{-9}	k_{10}	k_{-10}	Q_{CO}	Q_H	k_{15}
0.2789	0.1748	0.0874	0.3866	0.1933	0.3751	0.1876	1	3.2	0.0000
k_{16}	k_{17}	k_{18}	k_{19}	k_{20}					
0.0176	0.0138	0.0792	0.2750	0.0057					

Table 7.11: Parameter values for the ODEs (7.3.11)–(7.3.20) in our ensemble in which all reactions are fully irreversible.

its own data and that seen in 7.9. That is, we find the model x that minimises $E^{(2)}(x)$ as described in equation (7.3.23). There are five models that minimize this difference, models 747, 767, 959, 1003, and 1023. These models all have the reversibility for v_{10} equal to 0.5, and all other reversibilities are either 0.5 or 0.75. We could use this as a base to construct a more targeted ensemble of models. It seems that in order to match the data for dFBA, we must have high reversibilities for these reactions. We also find the model that minimises $E^{(1)}(x)$ as described in equation (7.3.22). In this case, the model that minimises $E^{(1)}(x)$ is model 747. This model was also included in the set that minimises $E^{(2)}(x)$. This suggests that the best model to pick would be model 747, and the parameters for this model are given in Table 7.11.

7.5 Conclusions

We have been able to use ensemble modelling to match data from both reference ODE models and dFBA models with some success. In the case of the ODEs, we successfully reduced an ensemble of $4^7 = 16384$ models to a single model that matched the data in the case that the model was exactly contained in the ensemble. We did, however, find that when the model is *not* exactly contained in the ensemble, there is the possibility of no model matching the reference ODE exactly. In this case, we had to find another method to find the closest matching model. We did this by selecting the model whose perturbation data had the minimal difference from the perturbation data of the reference model.

When using ensemble modelling to try and match data from the dFBA model, we

reduced a set of $4^5 = 1024$ models to eight models that matched one set of perturbation data. Unfortunately, we again found that no model in the ensemble matched all of the perturbation data we generated. We turned to finding the model that minimised the difference in overall perturbation data again. This technique has also been used by Khodayari et al. [36], who formalised a method for finding the model that minimised the difference from the reference data. By adding one set of data at a time, and solving a minimisation at each step, they were able to reduce a large ensemble into a single closest fitting model. The found model was then validated by testing against a final set of mutant data, and it was found that 74% of fluxes in the closest fitting model were within experimental ranges, with the remaining 26% being within three standard deviations.

For our models, we have only simulated reaction rates using first order kinetics. This method has benefits of allowing us consider enzyme levels and proportions, as well as easily manipulate them algebraically to generate perturbation data. However many kinetic models of metabolism are not based on these structures. We could also use other ways of describing reactions rates, such as Michaelis-Menten, as described by Lee et al. [42]. Here, we have only used ensemble modelling for parameter estimation; however, the technique of building an ensemble of models has also been used to fit other features of a network, such as cell signalling networks, as explored by Schaber et al. [63], or other cellular feedback systems demonstrated by Kuepfer et al. [39].

Ensemble modelling is be a useful tool for dealing with uncertainty in a metabolic model. It allows us to find best fitting models for data sets, but also provides us with a set of other models that may also be of interest. In the case where screening does not reduce an ensemble to a single model, we look for similarities in the remaining models to provide a more general overview of how a network may behave. Building up a relevant ensemble of models is relatively simple, as data on metabolic networks are often available; however, the data required to reduce this large ensemble may not be complete. Indeed, we have seen that even for our network, it is not possible to reduce an ensemble to a single model without knowledge of internal fluxes after perturbation, as two pairs of rate coefficients (those of k_4 and k_6) cannot be determined using only the product formation rates. We may also struggle to build an ensemble that sufficiently spans the space of reversibilities, especially in the case of large networks with multiple steps included in reactions. This can be somewhat alleviated by incorporating thermodynamic data to limit possible values for the rate coefficients, as described by Tran et al. [72]. Overall, the method has proved useful for providing information on models without the need to estimate kinetic parameters exactly.

Chapter 8

Comparison of methods for modelling metabolism

8.1 Introduction

We have attempted to model metabolic networks using both kinetic models, and flux balance techniques. These techniques have been used to model a variety of networks, most prominently a metabolic network for *C. autoethanogenum*. We have examined networks with a variety of long term behaviours, including steady-states, linearly growing states, and oscillatory solutions. To conclude this work, we introduce one final simple network to test each method, and compare the results provided by each method. We will then also consider more complex long timescale behaviours, such as periodic solutions, and examine what effect these could have on the methods and in particular which methods could predict their existence. Finally we present our conclusions.

8.2 A simple loop network

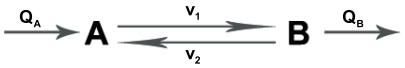

Figure 8.1: Network diagram for an example metabolic network featuring a loop. In this network, A is transformed into B in one reaction, and B is in turn transformed back into A though another reaction.

For the final network, we return to a network seen earlier in Chapter 2, shown again in Figure 8.1. In this network we have two metabolites, A and B. There is an uptake flux for A, Q_A, and an output flux for B, Q_B. Linking the two metabolites, we include a reversible reaction transforming A into B. Stoichiometrically, this reversible reaction is considered as two separate reactions, a forward reaction, v_A, transforming A into B, and a second backward reaction, v_B, transforming B back into A, however with a kinetic model we would have the option of modelling it using a single reversible term, or two irreversible terms. We examined this network in Chapter 2 in relation to metabolic flux analysis, showing that the internal fluxes, v_A and v_B, cannot be determined only with knowledge of the two external fluxes, giving only the condition that if

$$Q_A = Q_B = v_A - v_B, \tag{8.2.1}$$

then a steady-state is formed. As part of our comparison, we aim to see how much information about the internal reactions of the network can be found through each method.

8.2.1 A flux balance model

Let us begin by forming a flux balance model for this network. First we must generate the stoichiometric matrix for the network. This network is fairly simple, with only four reactions for two metabolites. The stoichiometric matrix is

$$S = \begin{matrix} \\ A \\ B. \end{matrix} \begin{matrix} \begin{matrix} v_A & v_B & Q_A & Q_B \end{matrix} \\ \begin{pmatrix} -1 & 1 & 1 & 0 \\ 1 & -1 & 0 & -1 \end{pmatrix} \end{matrix} \tag{8.2.2}$$

For a basic flux balance model, with a flux vector $\mathbf{v} = (v_A, v_B, Q_A, Q_B)^T$, the mass balance constraint $S\mathbf{v} = \mathbf{0}$ expands out to two equations,

$$Q_A - v_A + v_B = 0, \tag{8.2.3}$$
$$v_A - v_B - Q_B = 0. \tag{8.2.4}$$

From these two equations we are able to recover the condition for steady-state previously mentioned, i.e. $Q_A = v_A - v_B = Q_B$. All solutions provided by FBA satisfy this condition, as it is derived directly from the mass balance constraint. We form a general optimisation problem for this network that provides steady-state solutions via flux balance analysis. This general form is

$$\begin{aligned} \max_{\mathbf{v}} \quad \mathbf{a}^T\mathbf{v} \quad \text{such that:} \quad & Q_A - v_A + v_B = 0, \\ & v_A - v_B - Q_B = 0, \\ & \mathbf{v}_{min} \leq \mathbf{v} \leq \mathbf{v}_{max}, \end{aligned} \tag{8.2.5}$$

where $\mathbf{v}_{min}$ and $\mathbf{v}_{max}$ are vectors containing maximum and minimum values of each flux, and $\mathbf{a}$ is some vector of coefficients, such that $\mathbf{a}^T\mathbf{v} = a_1v_A + a_2v_B + a_3Q_A + a_4Q_B$ is some linear sum of the fluxes forming an objective function. As v_A and v_B are two fluxes representing one reversible reaction, these are both irreversible and take minimum values of zero. Both external fluxes (the uptake and output fluxes) are considered reversible. For now, let us set every maximum flux to be equal to one, and the minimum fluxes for reversible reactions to be equal to negative one. We take our objective function to be maximising the output flux, Q_B, with $\mathbf{a}^T = \mathbf{a}_1^T = (0, 0, 0, 1)$. Then the optimisation problem becomes

$$\begin{aligned} \max_{\mathbf{v}} \quad Q_B \quad \text{such that:} \quad & Q_A - v_A + v_B = 0, \\ & v_A - v_B - Q_B = 0, \\ & 0 \leq v_A, v_B \leq 1, \\ & -1 \leq Q_A, Q_B \leq 1, \end{aligned} \tag{8.2.6}$$

and our solution is the flux vector $\mathbf{v}_1^T = (1, 0, 1, 1)$. We have maximum values for Q_A, Q_B, and v_A, and a minimum value for v_B. This solution corresponds to a flux from the uptake of A, through v_A to the output of B, with no feedback through v_B at all.

Interestingly, if we change our objective function to be minimising the output flux, Q_B, i.e. maximising $-Q_B$, with $\mathbf{a}^T = \mathbf{a}_2^T = (0, 0, 0, -1)$, we find the solution $\mathbf{v}_2^T = (0, 1, -1, -1)$. Here, the solution has been reversed, with minimum values for Q_A, Q_B,

and v_A, and a maximum values for v_B. The network is instead taking up B, carrying it through the backward reaction v_B and outputting A. The network is able to operate in both directions due to the reversibility of both Q_A and Q_B. If either external reaction was irreversible, the second state could not exist. We refer to steady-states of the first kind, i.e. in which A is taken up and B is output as being in the forward direction, whilst steady-states in which B is taken up and A is output is considered the backwards direction.

Both states, $\mathbf{v}_1$ and $\mathbf{v}_2$, are unique optimal solutions for their respective optimisation problems, $\mathbf{a} = \mathbf{a}_1$ or $\mathbf{a} = \mathbf{a}_2$ but relaxing the maximum value constraints for the internal fluxes, v_A and v_B, allows alternate optimal solutions. As we have already stated, any steady-state solution for this network satisfies $Q_A = v_A - v_B = Q_B$. This means we can increase v_A to arbitrarily high values whilst still maintaining a steady-state, as long as we also increase v_B to compensate. For an optimal solution in FBA, for example maximising Q_B, as long as we maximise the flux through to Q_B first, we are free to choose a flux as large or small as we like for v_A, so long as any remaining flux to B that is not output through Q_B is fed back into A through v_B.

To do this, we note that there are two elementary modes for this system,

$$\mathbf{m}_1^T = (1, 0, 1, 1), \text{ and } \mathbf{m}_2^T = (0, 1, -1, -1), \tag{8.2.7}$$

representing the network in a forward state, and a backwards state. Whilst not an elementary mode, it is also useful to consider a third solution, which is the sum of the elementary modes,

$$\mathbf{m}_3 = \mathbf{m}_1 + \mathbf{m}_2 = (1, 1, 0, 0). \tag{8.2.8}$$

This solution is one in which the internal loop is active. Stoichiometrically, this is indistinguishable from a state in which the forward and backwards states both show the same level of activity, but it is useful to consider it as a separate state here. This allows us to write down a general solution for the flux balance problem seen in equation (8.2.5),

$$\mathbf{v} = \lambda_1(1, 0, 1, 1) + \lambda_2(0, 1, -1, -1) + \lambda_3(1, 1, 0, 0), \tag{8.2.9}$$

for $\lambda_n \geq 0$ In general, we would expect to only have flux in one direction, so that either λ_1 or λ_2 are equal to zero, but there is still some freedom of choice for the level of activity in the loop, λ_3. Often in FBA, we form a secondary optimisation problem when we have a non-unique solution to the original optimisation problem, in which we try to maximise or minimise a secondary objective function across the solutions that maximised or minimised the initial objective function. For our optimisation problem seen in equation (8.2.6) we form a secondary problem,

$$\begin{aligned} \underset{\mathbf{v}}{max} \quad \mathbf{a}^T\mathbf{v} \quad \text{such that:} \quad & Q_A - v_A + v_B = 0, \\ & v_A - v_B - Q_B = 0, \\ & 0 \leq v_A, v_B \leq v_{max}, \\ & -1 \leq Q_A, Q_B \leq 1, \\ & Q_B = Q_B^{max}, \end{aligned} \tag{8.2.10}$$

where Q_B^{max} is the maximised value of the primary objective function. A common choice for this secondary objective function is minimising the overall flux through the system. In this case, since Q_A and Q_B are fixed by the primary objective function, we only need to minimise the sum of v_A and v_B. This objective function chooses the solution in which $\lambda_3 = 0$, shutting off the internal loop entirely. Whilst this is one way of dealing with the freedom in solutions, it may not be accurate and does not provide us

with any information about the activity in the loop. Using only the first optimisation problem, (8.2.6), we are unable to find a unique steady-state flux for the network. By incorporating a second optimisation problem, such as the one in (8.2.10) allows us to find a single unique solution, however this relies on us choosing a secondary objective function. FBA has been able to provide us information on how a network can behave in steady-state, in the form of the elementary modes seen in equation (8.2.7), but finding a unique optimal flux vector requires multiple objective functions.

8.2.2 ODE models

Let us next examine an ODE model for the same network. This model consists of two ODEs, one for each concentration of A and B. Let the concentration of A be denoted by the variable c_A, and the concentration of B similarly be the variable c_B. Then we form the ODEs,

$$\frac{dc_A}{dt} = k_1(A^* - c_A) - k_A c_A + k_B c_B, \tag{8.2.11}$$

$$\frac{dc_B}{dt} = k_2(B^* - c_B) + k_A c_A - k_B c_B, \tag{8.2.12}$$

with initial conditions $c_A(0) = c_B(0) = 0$. We have six parameters for this network, where k_1 and k_2 are rate coefficients for the transfer of A and B in and out of the network, A^*, and B^* are constant external concentrations of A and B, and k_A and k_B are rate coefficients for the forward and backwards internal reactions. In terms of the fluxes used in the previous flux balance model, we have

$$v_A = k_A c_A, \tag{8.2.13}$$

$$v_B = k_B c_B, \tag{8.2.14}$$

$$Q_A = k_1(A^* - c_A), \tag{8.2.15}$$

$$Q_B = -k_2(B^* - c_B). \tag{8.2.16}$$

We note that in the expression for Q_B, a positive $k_2(B^* - c_B)$ corresponds to a negative Q_B, since the forwards direction for Q_B is taken to be an output.

We already see an immediate disadvantage in our ODE model, in that it would require the values for six parameters to simulate. Two of these parameters are external concentrations that should be easily measurable; however, the remaining four could be difficult to estimate, especially the two internal parameters, k_A and k_B.

For now we use example sets of parameters to display different behaviours of the network. In particular, we are able to force this ODE network to behave identically to the three different behaviours predicted by FBA. For now, to simplify the model, let us assume that $k_1 = k_2 = 1$, so that the reversible input and output fluxes are simply equal to the difference between the internal and external concentrations. For our first behaviour, let us inhibit the backwards reaction entirely by setting $k_B = 0$. Then for any choice of A^* and B^*, the ODE model predicts a steady-state in which all flux in the network goes in the forward direction, corresponding to the elementary mode $\mathbf{m}_1$ seen in equation (8.2.7). A simulation with $A^* = B^* = 1$ is shown in Figure 8.2.

If, instead, we were to choose to inhibit the forward reaction by fixing k_A to equal zero, but allowing a non zero k_B, then any choice of A^* and B^* would lead us a steady-state solution with all flux in the backwards direction corresponding to the $\mathbf{m}_2$ mode from equation (8.2.7). A simulation for this state with $A^* = B^* = 1$ is shown in Figure 8.3.

If we allow non zero values for both k_A and k_B, we are able to produce solutions with an overall flux in either direction depending on our choices of A^* and B^*, though these

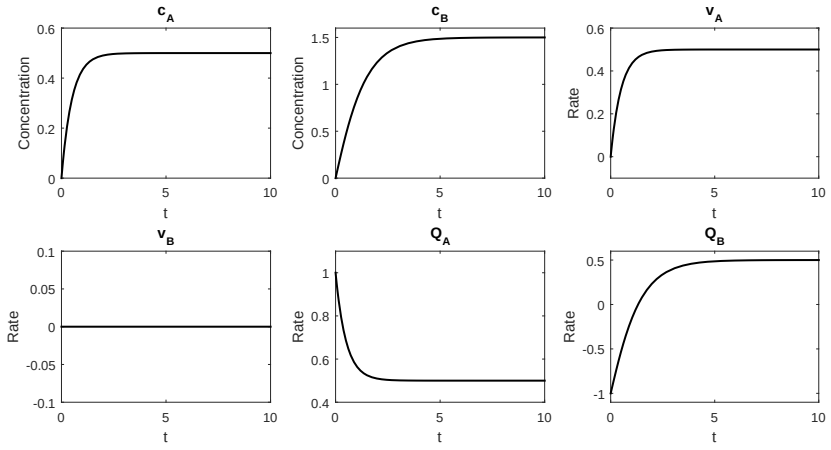

Figure 8.2: Plots of the simulation of the ODEs (8.2.11)–(8.2.12), with parameters, $k_1 = k_2 = k_A = A^* = B^* = 1$, and $k_B = 0$. The first two plots show the concentrations of metabolites in time, whilst the remaining four show the reaction rates (fluxes) in time. We see both Q_A and Q_B taking positive values, representing an uptake through Q_A and an output through Q_B. In this case, the fluxes in steady-state are $(v_A, v_B, Q_A, Q_B) = (0.5, 0, 0.5, 0.5) = 0.5\mathbf{m}_1$, where $\mathbf{m}_1$ is the elementary mode defined in equation (8.2.7).

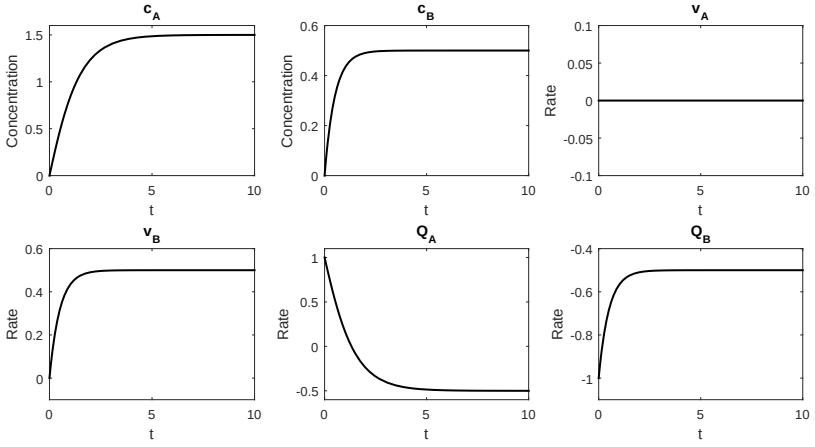

Figure 8.3: As in Figure 8.2, instead with $k_A = 1$ and $k_B = 0$. Here we have negative fluxes for Q_A and Q_B, showing an uptake through Q_B and an output through Q_A. In this case the steady-state fluxes are $(v_A, v_B, Q_A, Q_B) = (0, 0.5, -0.5, -0.5) = 0.5\mathbf{m}_2$, where $\mathbf{m}_2$ is the elementary mode defined in equations (8.2.7).

solutions do not correspond exactly to an elementary mode, as they are characterised as either $\lambda_1\mathbf{m}_1 + \lambda_3\mathbf{m}_3$ for a net flux in the forward direction, or $\lambda_2\mathbf{m}_2 + \lambda_3\mathbf{m}_3$ for a net flux in the negative direction, where $\mathbf{m}_n$ is a mode described in equation (8.2.7). In particular, for specific choices of A^*, B^*, k_A, and k_B we see the model predicting a steady-state that corresponds with the behaviour described by $\mathbf{m}_3$, in which neither Q_A or Q_B are active, and the network simply cycles flux across the loop. By choosing $A^* = B^* = k_A = k_B = 1$, we see this behaviour, shown in Figure 8.4.

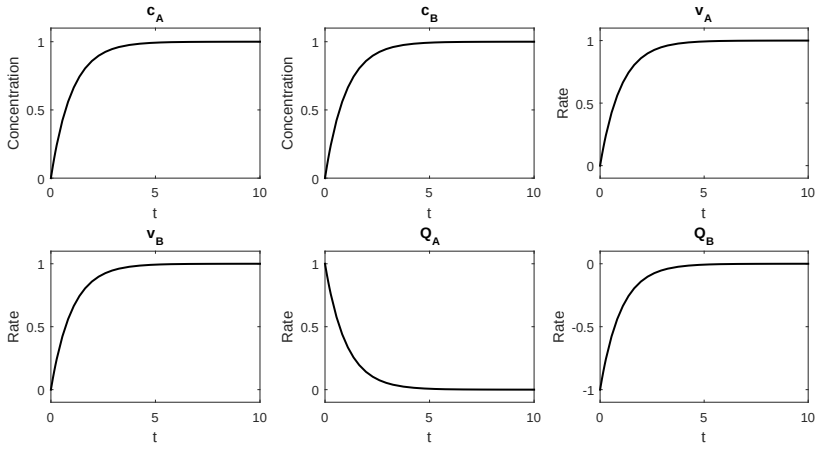

Figure 8.4: As in Figure 8.2, instead with both $k_A = 1$ and $k_B = 1$. We see zero values for the fluxes Q_A and Q_B, whilst v_A and v_B remain active, corresponding to only the internal loop being active. Here, the steady-state fluxes are $(v_a, v_B, Q_A, Q_B) = (1, 1, 0, 0)$. This is equivalent to $\mathbf{m}_1 + \mathbf{m}_2 = \mathbf{m}_3$, as defined in (8.2.9).

Interestingly, if we completely inhibit both forwards and backwards fluxes by setting $k_A = k_B = 0$, we find another solution in which no flux is active, i.e. $\mathbf{v} = (v_A, v_B, Q_A, Q_B) = (0, 0, 0, 0)$. This solution corresponds to the general form of the steady-state seen in equation (8.2.9) in which all λ_n are equal to zero, i.e. a zero flux solution. In general, these four described steady-state behaviours are the only ones that are exhibited by this system, confirming that the elementary modes generated in Section 8.2.1 are indeed the only steady-state behaviours possible, though FBA provided this same analysis with a great deal less work.

We now analyse the two ODEs, (8.2.11)–(8.2.12) in steady-state, in order to determine conditions on the parameters of the ODE that lead to each steady-state. We continue assuming that $k_1 = k_2 = 1$. Summing equations (8.2.11)–(8.2.12) give us

$$\frac{d(c_A + c_B)}{dt} = A^* - c_A + B^* - c_B.$$

Then setting time derivatives to equal zero gives the condition

$$A^* + B^* = c_A - c_B. \tag{8.2.17}$$

This tells us that in steady-state, the sum of the internal concentrations are equal to the sum of the external concentrations. Rearranging this equation to give c_B in terms of c_A, A^*, and B^*, and combining it with one of our initial ODEs in steady-state, we derive an expression for the value of c_A in steady-state. Then, by substituting this expression for c_A back into equation (8.2.17), we obtain a similar expression for the concentration c_B in steady-state. These expressions are

$$c_A^{ss} = \frac{A^* + k_B(A^* + B^*)}{1 + k_A + k_B}, \tag{8.2.18}$$

$$c_B^{ss} = \frac{B^* + k_A(A^* + B^*)}{1 + k_A + k_B}, \tag{8.2.19}$$

where c_A^{ss} and c_B^{ss} are the steady-state values of c_A and c_B respectively. We know that if we have a positive value for Q_A in steady-state, then we have flux in the forward

direction, whilst if Q_A takes a negative values, we have flux in the backwards direction. In the case that Q_A is equal to zero, the backwards and forwards fluxes are balanced, and there is activity only in the internal loop. If we wish to have $Q_A > 0$, then we have

$$Q_A = A^* - \frac{A^* + k_B(A^* + B^*)}{1 + k_A + k_B} = \frac{k_A A^* - k_B B^*}{1 + k_A + k_B} > 0, \tag{8.2.20}$$

leading to the condition

$$\frac{A^*}{B^*} > \frac{k_B}{k_A}. \tag{8.2.21}$$

We find the same results for $Q_A = 0$ and $Q_A < 0$. We are now able to predict which direction the flux flows in steady-state based on the parameters A^*, B^*, k_A, and k_B. If $\frac{A^*}{B^*} > \frac{k_B}{k_A}$ the flux is in the forward direction, if $\frac{A^*}{B^*} < \frac{k_B}{k_A}$ the flux is in the backwards direction, and if $\frac{A^*}{B^*} = \frac{k_B}{k_A}$, then there is no net production of A or B, and only the internal loop of the network is active. We find similar results if we start from Q_B instead of Q_A.

Whilst FBA was able to inform us of the different behaviours the network shows in steady-state, as it does not use any rate parameters, we would not be able to have any physical reference as to what conditions are required for the network to achieve that state. The ODE model has helped us see exactly how the network would react to different conditions, although knowledge of the parameters is required. These parameter values may not be known, which would limit the predictive power of our ODE model, limiting its overall usefulness.

8.2.3 dFBA models

We have now examined this network with flux balance and ODE models. The next technique we use to examine metabolic networks is dynamic flux balance analysis (dFBA). This method is an extension to the FBA that allows us to use flux balance techniques to predict dynamic behaviour in a model. It also uses some parameters from an ODE model to provide more accurate solutions. We now form a dFBA to try to observe these steady-state behaviours shown by flux balance analysis. As described in Chapter 6, we add the concentrations, c_A and c_B, as variables in our original FBA problem, and include two additional sets of constraints. The first is some constraints that depend on the concentrations. In this case we use our knowledge of the input and output fluxes from our ODEs (8.2.11)–(8.2.12), still with $k_1 = k_2 = 1$, to find

$$Q_A = A^* - c_A, \tag{8.2.22}$$
$$Q_B = c_B - B^*. \tag{8.2.23}$$

Rather than fixing these fluxes to be equal to their values in the ODEs, we allow some additional freedom to the method by setting these constraints to be inequality constraints, so that

$$Q_A \leq A^* - c_A, \tag{8.2.24}$$
$$Q_B \leq c_B - B^*. \tag{8.2.25}$$

We note, however, that these constraints are only useful when the overall network flux is in the forward direction. In the backward direction this would not prevent us from having very large negative fluxes without the required concentrations to make up for it. In simulations where we expect to find a large negative flux through Q_A and Q_B, we would instead use the set of constraints

$$-Q_A \leq c_A - A^*, \tag{8.2.26}$$

$$-Q_B \leq B^* - c_B. \tag{8.2.27}$$

The second constraint we add is the rate of change constraint, limiting how fast fluxes can change in a single time step. Finally, by replacing the FBA constraint $S\mathbf{v} = \mathbf{0}$ with the dFBA constraint $\dot{\mathbf{c}} = S\mathbf{v}$, where $\mathbf{c} = (c_A, c_B)$ is the vector of concentrations, we write down the general dFBA model for this network

$$\begin{aligned} \max_{\mathbf{c},\mathbf{v}} \quad \mathbf{a}^T\mathbf{v} + \mathbf{b}^T\mathbf{c} \quad \text{such that:} \quad \dot{c_A} &= Q_A - v_A + v_B, \\ \dot{c_B} &= v_A + v_B - Q_B, \\ \mathbf{0} \leq \mathbf{c} &\leq \mathbf{c}_{max}, \\ \mathbf{v}_{min} \leq \mathbf{v} &\leq \mathbf{v}_{max}, \\ \mathbf{v}_{dt-min} \leq \frac{d\mathbf{v}}{dt} &\leq \mathbf{v}_{dt-max}, \\ Q_A \leq A^* - c_A,\ Q_B &\leq c_B - B^*, \\ \text{or} & \\ -Q_A \leq c_A - A^*,\ -Q_B &\leq B^* - c_B, \\ \mathbf{c}(0) = \mathbf{0},\ \mathbf{v}(0) &= \mathbf{0} \end{aligned} \tag{8.2.28}$$

where $\mathbf{c}^T = (c_A, c_B)$ and $\mathbf{v}^T = (v_A, v_B, Q_A, Q_B)$. In every simulation of this model, we fix our initial concentrations to be $\mathbf{c} = (0, 0)$. These concentrations mean we need $Q_A = A^*$ and $Q_B = -B^*$ for our initial flux state, whilst v_A and v_B are free. We take them to both equal zero, so that our initial flux vector is $\mathbf{v} = (0, 0, A^*, -B^*)$.

First, let us consider solutions in the forward direction. To do this, we seek to maximise the output of B, that is maximising Q_B. As in previous examples in Chapter 6, the overall flux through the network is determined largely by the uptake flux, so we do not need to set any maximum values for flux or concentrations. Since we expect solutions in the forward direction, we use the first set of concentration based constraints, equations (8.2.22)–(8.2.23). Initially we take $A^* = B^* = 1$ as our choice of external concentrations. The results of this simulation are shown in Figure 8.5. We see the

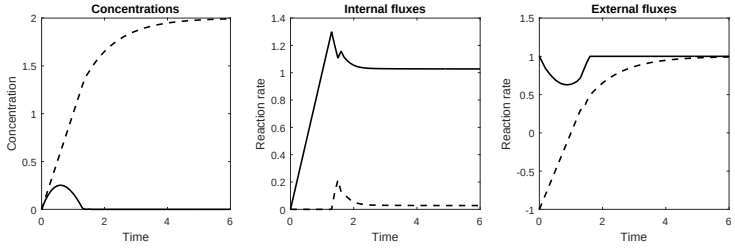

Figure 8.5: Simulation for the dFBA model described in equation (8.2.28). External concentrations are fixed so that $A^* = B^* = 1$, and the objective function is maximising Q_B. The first plot shows the concentrations, where the solid line is c_A, and the dashed line is c_B. The second plot is of the internal fluxes, where the solid line is v_A, and the dashed line is v_B. The final plot shows the external reaction rates, where the solid line is Q_A, and the dashed is Q_B.

concentration of B, c_B, smoothly increasing to a maximum value of two, allowing Q_B to rise to a maximum value whilst the concentration of A, c_A, increases slightly, before decreasing back down to zero to maintain the maximum value of Q_A. The internal reaction v_A increases linearly, before decreasing down to a value just higher than one, with v_B becoming slightly higher than zero, so that $v_A - v_B = 1 = Q_A$, matching the

condition for steady-state. We also note that $c_A + c_B = 2 = A^* + B^*$ in the steady-state, as predicted by equation (8.2.17), though $c_A = 0$, so all of the internal concentration is stored in B.

Next we increase B^* to see if this leads to the expected switch from the forward behaviour to the backward behaviour. Figure 8.6 shows one such simulation with $B^* = 5$. We see that the overall flux is still in the forward direction in steady-state, and again we have $v_A - v_B = 1 = Q_A$ and $c_A + c_B = 6 = A^* + B^*$, matching the two previously derived steady-state equations. We see in this case, c_B increases far past its steady-state value. This is because we begin with a large uptake of B (since we have a negative Q_B), and due to the maximum rate of change constraint, this flux does not increase quickly enough to prevent the large build up of B. When Q_B becomes positive, the network starts to output B. Eventually this excess of B is used up, and the system settles into the expected steady-state. We find that this behaviour continues no matter how much we increase the external concentration, B^*. In a dFBA model, it is possible to observe the forward behaviour for any ratio of external concentrations.

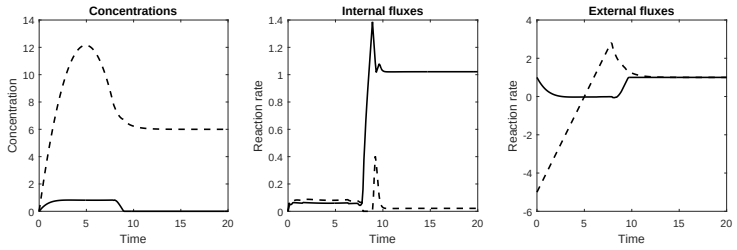

Figure 8.6: As in Figure 8.5, except $B^* = 5$.

We now seek states in which the overall flux through the network is in the backwards direction. To do this, we set a new objective function to be minimising Q_A, so that we maximise the output of A. We use the model (8.2.28) with the second set of constraints, (8.2.24)–(8.2.25). The results of this simulation are shown in Figure 8.7. They are very similar to those of the initial simulation where we attempted to maximise Q_B, but c_A and v_A increase to maximum values, rather than c_B and v_B. We also have Q_A decreasing to its minimum value, whilst Q_B starts off with its minimum value, increasing slightly as the concentration of B increases, before dropping back down to its minimum. We have v_B slightly greater than one, with v_A slightly greater than zero, so that $v_A - v_B = -1 = Q_A$ as required for steady-state. We also have $c_A + c_B = 2 = A^* + B^*$, though in contrast to the previous simulations where the flux was in the forward direction all of the internal concentration is stored in c_A, whilst before it was all in c_B.

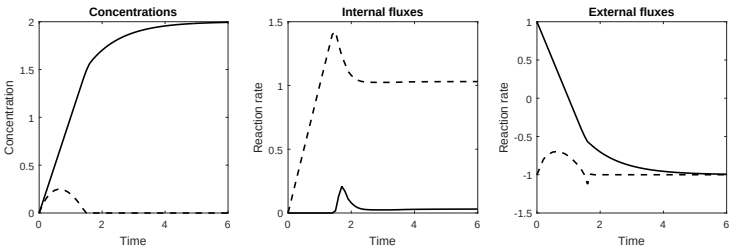

Figure 8.7: As in Figure 8.5, except the objective function is minimising Q_A.

As in the forward case, we can increase A^* arbitrarily, whilst still maintaining

this backwards state as long as the objective function is maximising the backwards flux. We see that, whilst for the ODEs there was a specific point, determined by equation (8.2.21), at which the solution switches from the forward behaviour to the backwards behaviour, this point does not exist in our dFBA model. The condition (8.2.21) determining whether the network exhibits forward or backward behaviour tells us that if $\frac{k_B}{k_A}$ is low enough, the ODEs are able to maintain the forward behaviour in the presence of a high B^*. That is to say, if the reactions in the loop are sufficiently biased in the forward direction, the overall forward behaviour is maintained even with high external concentrations of B. Since the dFBA model (8.2.28) does not offer any estimates of these two parameters, v_A and v_B do not depend on the concentrations, c_A and c_B, at all. This means that in the dFBA model, we can assume the value $\frac{k_B}{k_A}$ is as high or as low as required in order for the network to exhibit behaviour in either direction. In fact, we see that for dFBA, the objective function is the sole factor that decides whether the net flux through the network is forward or backwards. In regular FBA we see the same effect, the direction of the flux in the steady-state given by FBA is determined solely by the objective function. By not having fixed values for k_A or k_B, dFBA is able to show us that it is possible to have forward or backwards behaviour for any non-zero external concentrations A^* and B^*.

We note that in all our dFBA simulations, one of our metabolites tends to a concentration of zero in the steady-state. This was required for the uptake flux constrained by that concentration to take its maximum value. In our ODE model, a concentration of zero would also lead to one of the internal fluxes also being zero. This is not the case in the dFBA models, in which we maintain flux across the internal reactions despite having a concentration equal to zero. This allows the dFBA model to predict solutions in which we have the maximum possible flux through the system. These solutions are optimal, in that they maximise the objective function provided, however they could not explicitly be seen in the ODE model for the same network, unless we removed the dependence of the flux v_A on the concentration of A in the forward states, or the dependance of the flux v_B on the concentration of B in the case of the backwards states. Alternatively, if we let k_A in the ODEs (8.2.11)–(8.2.12) become arbitrarily large, our ODE steady-states would tend to the state seen in Figure 8.5. We could force our dFBA model to behave similarly to the ODE model by introducing concentration based constraints on v_A and v_B, but in doing this, we would fully determine the model, leaving no freedom of choice for the optimisation problem.

Finally, the dFBA model has predicted a small amount of activity in the internal loop in all cases. If we consider our steady-state flux vectors to be of the form $\mathbf{v} = \lambda_1\mathbf{m}_1 + \lambda_2\mathbf{m}_2 + \lambda_3\mathbf{m}_3$, where $\mathbf{m}_n$ refer to the modes descried in equation (8.2.7), then all of our dFBA solutions have a small, but non-zero, λ_3. There is still some freedom in this, and choosing a secondary objective function to increase the activity in this loop to a certain level, we could have λ_3 as small, or as large as we chose. This dFBA model, much like the FBA model, only informs us that there is could be some undetermined level activity in this loop. Without the addition of constraints to describe this loop, it is not possible to find a unique solution for this level of activity.

8.2.4 Ensemble modelling

Both flux balance methods, examined in Sections 8.2.1 and 8.2.3, have been unable to reasonably estimate the activity of the internal loop in the network. We need to find some estimate of the internal rate parameters in order to do this. In Chapter 7, we used ensemble modelling to find estimates of parameter values from perturbing parameters in a network. Specifically, by perturbing reactions to simulate a change in enzyme

levels, and measuring the response of each of the models in the ensemble, we are able to find parameterised ODEs that respond to these perturbations in the same way as a reference state. In particular, this method allows us to find reversibilities of reactions. If we were able to find an estimate of the reversibility of the internal reaction, we could use this reversibility to further tune our flux balance models.

Let us first write down a general model in our ensemble

$$\frac{dc_A}{dt} = A^{**} - k_1 c_A - k_A c_A + k_B c_B, \tag{8.2.29}$$

$$\frac{dc_B}{dt} = B^{**} - k_2 c_B + k_A c_A - k_B c_B, \tag{8.2.30}$$

with initial conditions $c_A(0) = c_{A(0)}$, and $c_B(0) = c_{B(0)}$. Here $A^{**} = k_1 A^*$, and $B^{**} = k_2 B^*$. As before we assume that the true values of k_1 and k_2 are both one, but we need to include the parameters in our ensemble to scale the fluxes correctly so that the steady-state is achieved when $c_A = c_B = 1$. This is similar to the single reaction we studied earlier in Section 7.2, but our output flux Q_B is now reversible. Let us treat A^* and B^* as known parameters. In this way, whilst both Q_A and Q_B are reversible reactions, we instead consider them each as two irreversible reactions. For Q_A we have an irreversible uptake with rate A^{**}, and an irreversible output with rate $k_1 c_A$. Similarly for Q_B we have an irreversible uptake of rate B^{**}, and an irreversible output with rate $k_2 c_B$. We have considered $v_A - v_B$ to be a single reversible reaction. We could consider both $v_A = k_A c_A - k_{-A} c_B$ and $v_B = k_B c_B - k_{-B} c_A$, but there is no structural difference in the model if we do this, since we would have

$$v_A - v_B = k_A c_A - k_{-A} c_B - k_B c_B + k_{-B} c_A = (k_A + k_{-B}) c_A - (k_{-A} + k_B) c_B$$

, which is the same model as described, with $k_A = k_A + k_{-B}$ and $k_B = k_{-A} + k_B$.

We choose a reference state from our ODE model, (8.2.11)–(8.2.12), with parameters

$$\begin{aligned} k_1 &= 1, \quad k_2 = 1, \\ k_A &= 1, \quad k_B = 1, \\ A^* &= 2, \quad B^* = 1. \end{aligned} \tag{8.2.31}$$

Simulating the ODEs with these parameters gives us a reference flux state with $Q_A = v_{net} = Q_B = 1/3$, where $v_{net} = v_A - v_B$ is the net flux across the loop. In our ensemble, this gives us parameters of $A^{**} = B^{**} = 1$, $k_1 = 5/3$, and $k_2 = 4/3$, with k_A and k_B are chosen such that

$$k_A = \frac{1}{3}\frac{1}{1-R}, \quad k_B = \frac{1}{3}\frac{R}{1-R}, \tag{8.2.32}$$

where R is the reversibility of the reaction.

In previous examples through Chapter 7, we sought data on the networks response to perturbations of internal enzyme levels. For this example, we have a simpler method to check which value of R produces a model that matches our reference ODE. We know that if we decrease the external concentration of A, A^*, or increase the external concentration of B, B^* past a threshold level, the network switches from forwards behaviour to backwards. By increasing the equivalent external concentration for our ensembles B^{**} incrementally, and seeing which model in the ensemble switches direction at a similar external concentration value, we find an estimate for the reversibility of this reaction. In this case, using an ensemble of ten models, with reversibilities evenly spaced between 0 and 0.9, we find the closest matching model has a reversibility of $R = 0.8$ this for this reaction. This R value indicates a relatively high level of reversibility, which

is expected given that our reference model has $v_A = v_B$. If we had values for c_A and c_B from our reference model, (8.2.11)–(8.2.12), we could then use this reversibility to calculate estimates for the parameters of our reference model (in this case, we would obtain exactly the parameters used for the reference model, given in (8.2.31)), but even without being able to calculate the parameters, the reversibility of a reaction is still useful information. A high reversibility means the backwards flux is close to the forwards flux, suggesting the net flux is small. In a biological network, a high backwards flux and small net flux would be unlikely, unless there was some necessity for it, such as the backwards reaction producing a useful secondary metabolic compound. If we find high reversibilities in a metabolic network, it is worth closely examining the backwards reaction, to discern any benefit for the organism.

Whilst neither flux balance method, from Sections 8.2.1 and 8.2.3, was able to find quantitative estimates on the individual internal fluxes, ensemble modelling has been able to provide a value for the ratio of the backward and forward fluxes in steady-state. Given estimates of internal concentrations, this method could further estimate values for the parameters of an ODE model modelling this network, which would allow us to model the full dynamics of this system. In this case, we achieved this simply by modifying an external concentration, though in some cases we may also require perturbing some internal reactions to narrow the ensemble down, as seen in our initial examples in Chapter 7.

8.3 Combining information from multiple methods

8.3.1 Ensemble modelling and FBA

Whilst each examined method gives different amounts of information on the behaviour of the network, if we are unable to form a complete ODE model, we are unlikely to be able to observe the full dynamic behaviour of a network. Often we are not able to form such a model, due to a lack of knowledge of the parameters involved in the reaction kinetics. In these cases we instead combine the techniques that we have to provide additional information about the network.

In Section 8.2.1, we used FBA to predict the steady-state behaviour of our simple loop network. We obtained general information about the steady-states this network could exhibit in the form of two elementary modes, $\mathbf{m}_1$ and $\mathbf{m}_2$ seen in equation (8.2.7). We also took a third mode (which was not elementary), $\mathbf{m}_3 = \mathbf{m}_1 + \mathbf{m}_2$. We were able to describe any steady-state as a sum of these modes, seen in equation (8.2.9), but we were unable to uniquely determine the λ_n in this situation. In particular, we may be able to fix λ_1 or λ_2 through our maximum flux values and objective function, but the value of λ_3 remained free in these situations.

Ensemble modelling provides us with the reversibilities of a given reaction. In particular, for a specific reference state for our loop network, we find an estimate for the value of v_B/v_A in the forward case, or v_A/v_B in the backwards case. If we restrict ourselves to the forward direction, this means we have a value for v_B/v_A. By using this result as a constraint in our FBA model, we are able to refine our FBA solutions to uniquely determine the behaviour inside this internal loop. If ensemble modelling provides us with the result $v_B/v_A = R$, we have $v_B - Rv_A = 0$, so our new FBA model for maximising the forward flux becomes

$$\max_{\mathbf{v}} \quad Q_B \quad \text{such that:} \quad Q_A - v_A + v_B = 0,$$

$$\begin{aligned} v_A - v_B - Q_B &= 0, \\ \mathbf{v}_{min} \leq \mathbf{v} &\leq \mathbf{v}_{max}, \\ v_B - Rv_A &= 0, \end{aligned} \tag{8.3.1}$$

where $\mathbf{v}_{min}$ and $\mathbf{v}_{max}$ are minimum and maximum values for the fluxes in the network respectively. In the forward case, we have λ_1 determined by the maximum values, $\mathbf{v}_{max}$, and λ_3 given by the $v_B - Rv_A = 0$ constraint. Specifically, we find $v_A = \lambda_1 + \lambda_3$ and $v_B = \lambda_3$, which combined with $v_B - Rv_A = 0$ gives us $\lambda_3 = R(\lambda_1 + \lambda_3)$, which is then rearranged to give

$$\lambda_3 = \frac{R\lambda_1}{1-R}. \tag{8.3.2}$$

In Section 8.2.4 we found an estimate of R for our reference ODE model to be $R = 0.8$. Using this value of R in our FBA model described in equation (8.3.1), with $\mathbf{v}_{min} = (0, 0, -\infty, -\infty)$ and $\mathbf{v}_{max} = (\infty, \infty, 1, 1)$ gives us the result

$$\mathbf{v} = (v_A, v_B, Q_A, Q_B) = (5, 4, 1, 1). \tag{8.3.3}$$

This flux vector is a linear multiple of the flux vector for the reference model used in ensemble modelling, seen in equations (8.2.11)–(8.2.12), with parameters $k_1 = k_2 = k_A = k_B = B^* = 1$, and $A^* = 2$. We could further refine our FBA model by lowering the maximum value of Q_A to be $1/3$, which would cause our FBA model to be an exact match.

Thus by combining results from ensemble modelling, in Section 8.2.4, with those from FBA, in Section 8.2.1 we are able to refine our FBA model further in order to choose more biologically relevant steady-state solutions. Sometimes this could lead to a uniquely determined solution for the internal fluxes, that could not be found using FBA alone.

8.3.2 Ensemble modelling and dFBA

In Section 7.4, we discussed the use of ensemble modelling to fit data from a dFBA model to an ODE model that behaved similarly in the presence of perturbations. Whilst we were unable to find a perfect fit, we were able to obtain estimates for the reversibilities of some reactions in our network in an "optimal" model. By applying the same methodology, we are able to draw similar conclusions as to what an "optimal" ODE model of the network pictured in Figure 8.1 would look like.

First, we need to produce a reference steady-state from our dFBA model seen in equation (8.2.28). We first consider an optimal flux in the forward direction. We choose our objective function to be maximising the flux in the forwards direction, that is maximising Q_B. As before, we do not need to set maximum values for our fluxes, since the overall level of flux through the network is constrained by the external concentrations, A^*, and B^* which we choose to both equal one in our reference state. This model has already been simulated and the results are seen in Figure 8.5. We have already seen from our dFBA analysis in Section 8.2.3 that the model does not switch direction regardless of how high we set B^*. By increasing B^{**} incrementally for each set of ODEs in our ensemble, we search for the model that maintains flux in the forwards direction for the highest values of B^{**}. This model would be the best fit for our dFBA data, and would be the model that maximises Q_B across all values of B^{**}. Using the same ensemble as in Section 8.2.4, we have ten models varying the reversibility of the internal loop between 0 and 0.9, we find that the lower the reversibility, the higher a value of B^* required for the network to switch direction. In particular, the irreversible case (when the reversibility is equal to zero) matches the dFBA model and does not

switch direction for any external concentration B^*. This corresponds to the parameters $k_A = v_B/c_B$ and $k_B = 0$. This result is somewhat obvious; the network would be most optimal at outputting B through Q_B when there is no backwards reaction, but it is still a useful result that has been provided by combining our dFBA model with an ensemble modelling method.

If we use the dFBA model (8.2.28) with the objective function of minimising Q_A rather than maximising Q_B, we obtain a reference state for a model that is most optimal in the reverse direction. A simulation of this model was shown in Figure 8.7. We form a new ensemble spanning the same range of reversibilities (0 to 0.9) for the internal loop reaction, but now the reference state is in the backwards direction, so rather than the reversibility being v_B/v_A as it was in the forward case, it is now v_A/v_B. To generate perturbation data we wish to match the ensemble to, we increase the external concentration of A, A^*, stepwise, to see when the model switches directions. As in the forwards case, we find that the dFBA model does not switch direction even in the presence of a very high external concentration A^*. This behaviour again matches the irreversible case, in which $v_A = 0$ implying $k_A = 0$.

For this single reaction, both the optimal forward direction model and the optimal backwards direction model require the internal reaction to be irreversible, with either the forward or backwards reaction being inhibited completely. Whilst for this simple network, these results are fairly obvious, in a larger network such behaviour might be more complex and difficult to understand. Whilst moderate values for reversibilities may not provide us with much information, extreme values (close to zero, or close to one) may give us useful insight into how an optimal version of a network might behave. A reversibility close to zero suggests that there is a very small backwards flux, which suggests that if there are any additional products produced by this backwards reaction, they are not vital for an optimal network. A reversibility close to one suggests that the overall net flux through a reaction is relatively low compared to the individual directions. As already suggested, a large reversibility may suggest the backwards reaction produces a useful additional product that the network needs. In the case of a dFBA model, this high reversibility with low net flux might suggest that the reaction itself is unnecessary for the optimal network, and provide a target for future metabolic engineering, such as trying to heavily inhibit that reaction. Such results could also be checked in our dFBA model, checking that solutions are still able to be found even without the presence of that reaction. This could also be checked in the simpler FBA version, checking that steady-state solutions still exist after the high reversibility reaction is removed.

Combining multiple techniques in this way allows us to draw more information out of a network, and even with a small amount of information on the kinetic behaviour of a network, we are able to draw interesting conclusions about the behaviour of the network, and even find potential targets for metabolic engineering.

8.4 More complex long term behaviour

8.4.1 Multiple steady-states

The loop network examined through the course of this chapter has only one long term behaviour, that of a single steady-state. In Chapter 5, we examined networks that permitted the existence of multiple steady-states for the same sets of parameters. In particular we analysed an autocatalysis system, shown in Figure 8.8, that permitted up to three steady-states for a given set of parameters, two stable and one unstable. The first stable state had a high concentration of x, and a low concentration of y. The

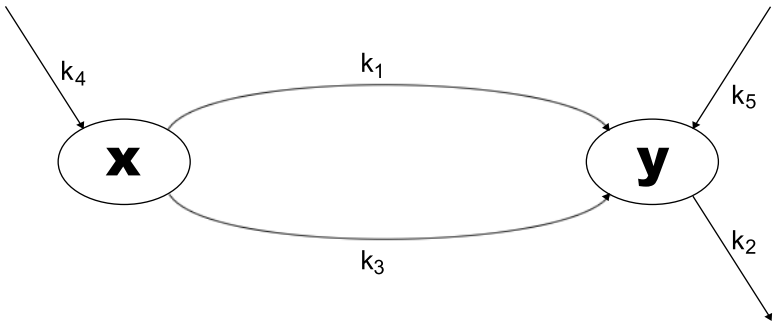

Figure 8.8: Network diagram for the autocatalysis network. We have two metabolites, X and Y, each with their own uptake reactions. We also have a separate output of Y, representing it decaying. There are two reactions transforming X into Y. The first, with rate k_1, is an autocatalysed reaction in which we have $X + 2Y \rightarrow 3Y$. The second, with rate k_3, is a simple uncatalysed reaction in which we simply have $X \rightarrow Y$.

other stable state had a lower concentration of x, and a higher concentration of y. The unstable state had intermediate concentrations of x and y between the concentrations of the stable states. We attempted to use various optimisation methods to find the existence of these multiple steady-states using FBA, without requiring full detail of the kinetics of the network. Unfortunately, we found that FBA was unable to predict these multiple steady-states effectively, as seen in Section 5.5. We also applied dFBA to this network find the multiple steady-state behaviours, as seen in Section 6.3. Whilst we were not able to predict the multiple states for different initial conditions as hoped, dFBA was able to predict the existence of both a state with a high concentration of x, and a state with a low concentration of x, though the low x state did not match up with the low x state predicted by the ODEs, and included a much higher value of the reaction representing the decay of y.

We did not apply the technique of ensemble modelling to this autocatalytic network, though the existence of multiple steady-states could provide an additional opportunity for use when screening models. In Chapter 7, we screened models by determining models in the ensemble that had the same response to perturbations in parameters the models we were using to provide reference data (in some cases another ODE model, and later a dFBA model). In Section 8.2.4, rather than perturbing each parameter in turn, we perturbed a single parameter multiple times to find the model from the ensemble that switched direction at the same time as our reference model. We could apply this same idea to multiple steady-states, screening our ensemble by seeking models that switch from one steady-state to another with the modification of parameters, or even initial conditions. The multiple steady-states could also prove problematic in the screening of models. If the reference model has parameters close to a bifurcation, there might exist a model in the ensemble that has the closest possible reversibilities, but resides on the other side of the bifurcation. This could lead to a model with similar parameters exhibiting a different behaviour after perturbation, causing it to be screened out, despite being a good match. When performing ensemble modelling, we should be aware of the possibility of multiple steady-states, and bifurcations in parameter space.

Overall we found that only an ODE model was able to adequately capture the behaviour of multiple steady-states, with dFBA being able to predict the existence of multiple states for the same parameter set, though not the initial conditions that lead

to them. We also found that the states predicted by dFBA were more extreme than those predicted by ODEs. In the case of the autocatalytic model, the high X state had a higher concentration of X in the dFBA model, and the low X state had a much lower concentration of X in the dFBA model than in the ODE model.

8.4.2 Dynamic long term behaviours

There are other long term behaviours that can be exhibited by an ODE model of metabolism. With our *C. autoethanogenum* ODE model, we saw multiple different states in which one or more metabolite concentrations grew linearly in time, whilst others decayed to zero. Another behaviour we saw was periodic solutions, most commonly oscillatory solutions in which the same concentrations increase to a maximum, and then decrease to a minimum, before increasing back to the maximum and repeating. Periodic behaviour was seen for specific parameter values of the examined autocatalysis network, but also for the adapted Lotka-Volterra model examined with dFBA in Section 6.4. Whilst this model is not strictly a metabolic network, it follows a similar structure.

Since the growth and oscillatory behaviours are dynamic, they cannot be described FBA, since FBA assumes steady-state behaviour.

The long term behaviour of the Lotka-Volterra network is interesting in that it has an infinite number of periodic solutions as well as two steady-state points. The model is

$$\frac{dx}{dt} = \alpha x - \beta xy, \tag{8.4.1}$$

$$\frac{dy}{dt} = \delta xy - \gamma y. \tag{8.4.2}$$

with initial conditions $x(0) = x_0$, and $y(0) = y_0$. These ODEs are normally used to model predator-prey populations, where x is the prey population, and y is the predator population. The two steady-state points are $(x, y) = (0, 0)$, i.e. an extinction state, and $(x, y) = (\gamma/\delta, \alpha/\beta)$, the equilibrium state. Every other point in the (x, y) plane, where neither x or y are equal to zero, are part of a periodic solution.

Whilst this model is a population model, we see that if we take

$$\mathbf{v} = (v_1, v_2, v_3) = (xy, x, y),$$

this model has the structure of a metabolic model with stoichiometric matrix

$$S = \begin{matrix} \\ -\beta \\ \delta \end{matrix} \begin{matrix} \begin{matrix} v_1 & v_2 & v_3 \end{matrix} \\ \begin{pmatrix} \alpha & 0 \\ 0 & -\gamma \end{pmatrix} \end{matrix} \begin{matrix} \\ x \\ y. \end{matrix} \tag{8.4.3}$$

In Section 6.4, we saw that we were unable to obtain periodic solutions for this network using dFBA. The dFBA model was able to accurately predicted the extinction state, and an equilibrium state (that was unique up to scalar multiples). However for all initial conditions, the model predicts the network tending to a steady-states rather than the periodic solutions we expect. In the Lotka-Volterra system, for all parameter choices, every point on the (x, y) plane that is not a steady-state and has neither x nor y equal to zero is part of a periodic solution, however dFBA is unable to find or maintain periodic solutions here. Hence we are unlikely to observe periodic solutions in other networks using dFBA. Other networks may require specific parameters in order to observe these periodic solutions, making it less likely that dFBA is able to predict

the existence of these solutions. In general, though we may be able to see periodic solutions using dFBA, we cannot expect dFBA to reliably find and describe them.

In Section 6.5, we used dFBA to successfully predict the existence and behaviour of states in which one or more metabolites increase linearly in time. By fixing certain fluxes to take certain values, we were able to predict linear growth in some cases. We compared these states in dFBA to those predicted by our ODEs and found that of the six linear growth states observed by our ODE model, we were able to match two exactly, and two more partially with a dFBA model. These states were independent of objective function choice, suggesting that they were general behaviour in the network, and the states found solely in our ODE model, and not in dFBA, may be a consequence of our specific choice of how we modelled flux dependence in the ODE system, and not directly on the structure of the network itself. We did note that the method had problems with metabolite concentrations decaying to zero, sometimes instead maintaining a very low steady-state, rather than a concentration of zero. Overall dFBA was useful for modelling some linear growth states, but not useful for modelling periodic behaviour.

Ensemble modelling also struggles with non-steady-state behaviours. We may not be able to reliably construct an ensemble of models that all tend to a specific periodic solution, certainly we could not guarantee that they even approach a solution with the same period, which would make automatically screening models impossible, and manually screening models would be prohibitively difficult for large ensembles. Similarly, with linearly growing states, we could not reliably construct an ensemble of models that grow at the same speed, especially after perturbation of parameters. We applied ensemble modelling to our *C. autoethanogenum*, but had to take special care when constructing the ensemble to not choose reversibilities for reactions that would lead to parameter schemes that caused the ODE model to tend to non-steady-states. In particular, it is possible to have a steady flux vector, whilst concentrations grow linearly in time. In this case, ensemble modelling may select a model that has the correct flux vector for a given perturbation, but with concentrations behaving completely differently to the reference state. As with multiple steady-states, linear growth states could lead to ensemble modelling choosing incorrect models, so where possible, we must avoid parameter choices that lead to non-steady-states.

Whilst FBA cannot be used to predict the specific dynamics of a non-steady-state behaviour, we are able to use it to predict the existence of non-steady-state behaviours. In Chapters 2 and 4, we applied FBA to our *C. autoethanogenum* network, and were able to predict the range of parameters for which steady-states exist. In particular, in Chapter 4 we determined a 2D region of flux space in which steady-states could exist. These fluxes were directly comparable to parameters in our ODE model, allowing us to provide strict bounds on these parameters that would lead to steady-state behaviour. Whilst FBA cannot be used to predict how the system behaves outside of these steady-state bounds, we can predict the existence of non-steady-state behaviour, which could then be examined further with a dynamic technique, such as dFBA. We also note in Chapter 4, we found the modified *C. autoethanogenum* model exhibiting two different steady-states for the same parameters, combined with a third linear growth state. Again, whilst this behaviour could not have been predicted by FBA alone, the behaviour was easily found to exist by combining the FBA results showing where steady-states exist, with a dynamic result from our ODE analysis in Chapter 3, showing that for the same parameters, a linear growth state would also exist.

It seems, again, that only a full ODE model is able to completely capture long term dynamic behaviour. In particular neither flux balance method is able to predict periodic solutions, though dFBA was able to predict linear growth states. Ensemble modelling becomes difficult in the presence of long term dynamic behaviours and care must be

Method	Data required	Information provided
FBA	Stoichiometry of the network	Steady state flux vectors
dFBA	Stoichiometry of the network Partial kinetic data	Dynamic flux and concentration vectors
Ensemble	Stoichiometry of the network Elementary reactions of the network Flux response to perturbations	Ensemble of models, and estimates of parameters within
ODEs	Stoichiometry of the network Complete kinetic data Estimates of parameters	Dynamic concentration vectors, from which flux vectors are calculated

Table 8.1: Data required by each model to be formed, and the information provided by each model.

	Method			
Long timescale behaviour	FBA	dFBA	Ensemble	ODE
Steady state	✓	✓	✓	✓
States with unbounded growth		✓	✓	✓
Periodic solutions			✓	✓
Multiple long timescale behaviours		✓	✓	✓

Table 8.2: Long timescale behaviours able to be predicted by each method. Note for dFBA, multiple long timescale behaviours can be predicted; however, we cannot consider different initial conditions leading to them. For ensemble modelling, all long term behaviours could be predicted, as it forms an ODE model, however non-steady-state behaviours can lead to problems with the method, as noted in Section 8.4.

taken when applying it to networks that we know could exhibit these behaviours.

8.5 Conclusions

We have now compared each method on a single model, to consider the data each model is able to provide. A summary of this is provided in Table 8.1. In general, it is clear that the more data we have to use to form a model, the more complete information we are able to get back from the model. FBA relies on only the stoichiometry of a network to form a model, but as a trade off, it only gives us steady-state flux vectors for the network. As we have seen in Section 8.2.1, these steady-state flux vectors may not be unique, instead providing us with a range of potential flux vectors that are all optimal according to a specific objective function. On the other end of the scale, a complete ODE model is able to provide a complete overview of the dynamics of a metabolic network. Unfortunately, forming a complete ODE model like this is difficult, requiring a vast knowledge of the kinetic properties of each given reaction, along with estimate of parameters, which we may not readily have access to. Our other two methods, dFBA and ensemble modelling allow us to gain information on the dynamic behaviour of a network whilst only requiring partial knowledge of these kinetics. dFBA allows us to observe the dynamics of the network with knowledge of the kinetics of only a few reactions, filling in the rest using the stoichiometry of the network, whilst ensemble modelling uses elementary reactions and data of response to perturbations in order to

estimate parameters, providing us with, in the best case, a single ODE model, or a small set of models that all behave in a similar manner.

Throughout this work we have also considered what types of long term behaviour can be predicted by each model. Table 8.2 summarises these behaviours. We note that only full ODE models have been seen to predict every type of long term behaviour seen by our metabolic networks. The models provided by ensemble modelling could exhibit all these long term behaviours, but non-steady behaviours can cause problems when constructing these models, as noted in Section 8.4. dFBA was able to predict unbounded growth and steady-states, and indicate the possibility of multiple long timescale behaviours, however we found it was not reliably able to predict periodic solutions. FBA, whilst able to indicate the possibility of non-steady behaviours in general by telling us when steady-states cannot exist, is not able to provide any information on how a network may behave when steady-states cannot exist.

Whilst we have seen that only ODE models are able to fully predict all kinds of dynamic behaviour that a metabolic network could exhibit, these models are not often feasible to produce. Full genome scale models for metabolism can contain hundreds of metabolites and even more reactions. Forming and simulating an ODE model on this scale would be highly challenging. In these cases, we are able to instead use FBA to quickly gain insight into the steady-state behaviour of such a model. A key strength of FBA is how the method is able to be scaled up to networks of this size and still provide results. In Chapter 4 we were able to determine a region of flux space in which steady-states could exist for our *C. autoethanogenum* model. Similar analysis can be performed even on larger networks. In particular we could consider the effects of a variety of ratios of input fluxes on the product formation rates within the network. In this way, even for large and complex networks, we are still able to draw insight into possible optimal product yields. We are also able to simulate gene knockouts with an FBA model. By inhibiting each reaction in turn, we find which reactions are necessary for the growth and survival of an organism, and which are unnecessary, providing potential targets for metabolic engineering, as seen by Edwards and Palsson [18] with regards to *E. coli*. Such analysis is performed very rapidly with FBA models, and provides a high degree of accuracy when compared to experimental data.

dFBA models have been applied with success to optimising product formation rates in batch cultures. Hjersted and Henson [31] used dFBA to model the production of ethanol in a fed-batch reactor by *S. cerevisiae*. By using dFBA to predict the maximal yields of ethanol, and performing a non-linear optimisation problem treating a variety of physical parameters as variables, such as feed flow rate, and initial glucose level, an optimal set of conditions for this batch culture was found. It is possible to efficiently simulate dFBA models multiple times with variations to physical parameters such as these, highlighting dFBA as a useful tool for this kind of optimisation. It is possible to form an ODE model to perform this analysis, however the difficulty in forming such a model makes it undesirable. Our own analysis of dFBA found use in its ability to predict states of unbounded growth in a metabolic network, seen in Section 6.5. FBA allowed us to determine which conditions would lead to steady-state solutions, but dFBA allowed us to quantify possible behaviours of the network outside of this steady-state region. By combining the information from the two techniques, we were able to build up a more complete picture of the behaviour of the network under all conditions, particularly those of the input ratio, Q_H/Q_{CO}.

Both FBA and dFBA have been useful in providing general information about whether a network attains steady-state in a long timescale, however there are questions about its ability to accurately predict these steady-states. In particular, both methods rely on an objective function which is chosen by the modeler. We have seen for both

FBA, in Section 2.4, and dFBA, as in Section 6.3, these objective functions have a strong effect on the solutions provided. Different objective functions often lead to completely different solutions for a network, and, whilst there are common choices for objective functions, there is no universal best choice objective function that provides a best fitting solution. We can examine a range of different objective functions, and compare them to experimental data, but there is no guarantee that the best objective function fitting for a specific organism is the best fitting objective function for a different organism. Indeed, there is no guarantee that the best fitting objective function applies to the same organism in different environmental conditions. Whilst FBA is useful for predicting the existence of steady-states, and survival in the presence of gene knockouts, the specific steady-state flux vectors provided by the model may not be a good representation of how the metabolic network actually behaves in biological conditions.

We also saw the problem of multiple optimal solutions. Whilst with FBA, we are able to quantify these multiple optimal solutions, using methods such as flux variability analysis [44], when it comes to dFBA, alternative optima found on early time points have large effects on the flux vectors and concentrations at the end time point. In particular, in Section 6.3, we showed that these alternative optima lead to completely different states. We applied dFBA to the autocatalytic network seen in Figure 8.8. The same model could tend to a state in which we had a high internal concentration of X and a near zero concentration of Y, or a near zero concentration of X and a high concentration of Y. This is evidence of multiple steady-states being possible within the network, however due to the multiple optimal solutions at each time step, it was not possible to determine which of the two states would be exhibited by the model in the long timescale. We found that minimal changes to the model, such as raising maximum flux values, led to the model switching between the two behaviours, which is not desirable. The problem of these alternative optima can be bypassed by using secondary objective functions, to choose a unique solution out of those that maximise the primary objective function. The choice of these secondary objective functions will also affects the solutions found by dFBA, just as the primary objective function does. As with the primary objective function, choice of secondary objective functions is arbitrary, and the problems only magnify as additional objective functions are added.

Ideally, when modelling a metabolic network, we would form a complete ODE model for the network, and attain full information on the dynamics of that network. Indeed, if we wish to have such information on the dynamics, we must attempt to form an ODE model, whether attempting to fully model the reactions, as seen by Chassagnole et al. [12], or by using a simplified model, such as the linlog kinetics used by Visser and Heijnen [76], or indeed, parameterising a model consisting of elementary reactions via ensemble modelling [72]. Sometimes these models may be infeasible to generate, or we may not even wish to model the dynamics of the network in such detail. In these cases, it is worth considering the flux balance methods presented in Chapters 2 and 6. If we wish to produce a genome scale model, this is an area in which FBA excels. If we want to examine how a particular network may respond to different food sources, FBA can provide insight. dFBA has been shown to be useful in optimising yields in batch cultures, and also in providing general insight into non-steady-state behaviours, though not in the case of periodic solutions. Whilst the exact steady-state solutions provided by FBA should not be considered perfect, its ability to provide conditions on fluxes that are required for steady-states, in particular the networks response to gene knockouts, with little computational effort is highly useful. Whilst a complete ODE model for metabolism is desirable, in some cases, it may prove unnecessary to form such a model. The information that can be provided by these simpler methods may be sufficient for a modelers purposes, and even in cases where it isn't, the benefits of not

needing to construct a full ODE model may outweigh the disadvantages presented by flux balance models.

Bibliography

[1] Amato, S., Orman, M., and Brynildsen, M. Metabolic control of persister formation in escherichia coli. *Molecular Cell 50*, 4 (2013), 475–487.

[2] Armitage, E., and Barbas, C. Metabolomics in cancer biomarker discovery: current trends and future perspectives. *Journal of Pharmaceutical and Biomedical Analysis 87* (2014), 1–11.

[3] Atsumi, S., Cann, A. F., Connor, M. R., Shen, C. R., Smith, K. M., Brynildsen, M. P., Chou, K. J. Y., Hanai, T., and Liao, J. C. Metabolic engineering of escherichia coli for 1-butanol production. *Metabolic Engineering 10* (2008), 305–311.

[4] Atsumi, S., Hanai, T., and Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature 451* (2008), 86–89.

[5] Audoly, S., Bellu, G., D'Angio, L., Saccomani, M., and Cobelli, C. Global identifiability of nonlinear models of biological systems. *IEEE Transactions on Biomedical Engineering 48*, 1 (2001), 55–65.

[6] Bailey, J. E. Toward a science of metabolic engineering. *Science 252* (1991), 1668–1675.

[7] Beard, D. A., Liang, S., and Qian, H. Energy balance for analysis of complex metabolic networks. *Biophysical Journal 83* (2002), 79 – 86.

[8] Beg, Q. K., Vazquez, A., Ernst, J., de Menezes, M. A., Bar-Joseph, Z., Barabási, A.-L., and Oltvai, Z. N. Intracellular crowding defines the mode and sequence of substrate uptake by escherichia coli and constrains its metabolic activity. *Proceedings of the National Academy of Sciences 104* (2007), 12663–12668.

[9] Cameron, D. C., and Tong, I.-T. Cellular and metabolic engineering. *Applied Biochemistry and Biotechnology 38* (1993), 105–140.

[10] Cann, A. F., and Liao, J. C. Production of 2-methyl-1-butanol in engineered escherichia coli. *Applied Microbiology and Biotechnology 81* (2008), 89–98.

[11] Chao, Y.-P., and Liao, J. C. Metabolic responses to substrate futile cycling in escherichia coli. *Journal of Biological Chemistry 269* (1994), 5122–5126.

[12] Chassagnole, C., Noisommit-Rizzi, N., Schmid, J. W., Mauch, K., and Reuss, M. Dynamic modeling of the central carbon metabolism of escherichia coli. *Biotechnology and Bioengineering 79* (2002), 53–73.

[13] Chen, D.-X., Coughenour, M., Knapp, A., and Owensby, C. Mathematical simulation of c4 grass photosynthesis in ambient and elevated co2. *Ecological Modelling 73*, 1-2 (1994), 63–80.

[14] Collas, F., Kuit, W., Clément, B., Marchal, R., López-Contreras, A. M., and Monot, F. Simultaneous production of isopropanol, butanol, ethanol and 2, 3-butanediol by clostridium acetobutylicum atcc 824 engineered strains. *AMB Express 2* (2012), 45.

[15] Connor, M. R., and Liao, J. C. Engineering of an escherichia coli strain for the production of 3-methyl-1-butanol. *Applied and Environmental Microbiology 74* (2008), 5769–5775.

[16] Covert, M., Schilling, C., and Palsson, B. Regulation of gene expression in flux balance models of metabolism. *Journal of theoretical biology 213*, 1 (2001), 73–88.

[17] Dai, Z., Dong, H., Zhu, Y., Zhang, Y., Li, Y., Ma, Y., et al. Introducing a single secondary alcohol dehydrogenase into butanol-tolerant clostridium acetobutylicum rh8 switches abe fermentation to high level ibe fermentation. *Biotechnol Biofuels 5* (2012), 44.

[18] D'Anna, A., Lignola, P. G., and Scott, S. K. The application of singularity theory to isothermal autocatalytic open systems: The elementary scheme $a+mb = (m+1)b$. *Proc Roy Soc Lond A: Mathe, Physi and Engi Sci 403* (1986), 341–363.

[19] de Noronha Pissara, P., Nielsen, J., and Bazin, M. Pathway kinetics and metabolic control analysis of a high-yielding strain of penicillium chrysogenum during fed batch cultivations. *Biotechnology and Bioengineering 51*, 2 (1996), 168–176.

[20] Edwards, J. S., and Palsson, B. O. The *escherichia coli* mg1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences 97* (2000), 5528–5533.

[21] Edwards, J. S., Ramakrishna, R., and Palsson, B. O. Characterizing the metabolic phenotype: A phenotype phase plane analysis. *Biotechnology and Bioengineering 77* (2002), 27–36.

[22] Ermentrout, B. *Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XPPAUT for Researchers and Students.* Software, Environments, and Tools. Society for Industrial and Applied Mathematics, 2002.

[23] Fell, D., and Small, J. Fat synthesis in adipose tissue. an examination of stoichiometric constraints. *Biochemical Journal 238*, 3 (1986), 781–786.

[24] Fell, D. A. Metabolic control analysis: a survey of its theoretical and experimental development. *Biochemical Journal 286* (1992), 313.

[25] Fell, D. A. Increasing the flux in metabolic pathways: a metabolic control analysis perspective. *Biotechnology and Bioengineering 58* (1998), 121–124.

[26] Frank, M., and Wolfe, P. An algorithm for quadratic programming. *Naval Research Logistics (NRL) 3*, 1-2 (1956), 95–110.

[27] Gibson, G., and Skett, P. *Introduction to drug metabolism.* Nelson Thornes, 2001.

[28] Gillespie, D. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications 188*, 1-3 (1992), 404–425.

[29] Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry 81*, 25 (1977), 2340–2361.

[30] Gomez, J., Höffner, K., and Barton, P. Dfbalab: a fast and reliable matlab code for dynamic flux balance analysis. *BMC Bioinformatics 15*, 1 (2014), 409.

[31] Grafahrend-Belau, E., Junkerand, A., Eschenröder, A., Müller, J., Schreiber, F., and Junker, B. Multiscale metabolic modeling: dynamic flux balance analysis on a whole-plant scale. *Plant Physiology 163*, 2 (2013), 637–647.

[32] Gross, L., Kirschbaum, M., and Pearcy, R. A dynamic model of photosynthesis in varying light taking account of stomatal conductance, c3-cycle intermediates, photorespiration and rubisco activation. *Plant, Cell & Environment 14*, 9 (1991), 881–893.

[33] Hanai, T., Atsumi, S., and Liao, J. C. Engineered synthetic pathway for isopropanol production in escherichia coli. *Applied and Environmental Microbiology 73* (2007), 7814–7818.

[34] Hanly, T., and Henson, M. Dynamic flux balance modeling of microbial cocultures for efficient batch fermentation of glucose and xylose mixtures. *Biotechnology and Bioengineering 108*, 2 (2011), 376–385.

[35] Hjersted, J., and Henson, M. Optimization of fed-batch saccharomyces cerevisiae fermentation using dynamic flux balance models. *Biotechnology Progress 22*, 5 (2006), 1239–1248.

[36] Ibarra, R., Edwards, J., and Palsson, B. Escherichia coli k-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature 420*, 6912 (2002), 186.

[37] Kauffman, K. J., Prakash, P., and Edwards, J. S. Advances in flux balance analysis. *Current Opinion in Biotechnology 14* (2003), 491–496.

[38] Kay, S., Scott, S., and Lignola, P.-G. The application of singularity theory to isothermal autocatalytic reactions: The influence of uncatalysed reactions. *Proc Roy Soc Lond A: Mathe, Physi and Engi Sci 409* (1987), 433–448.

[39] Khazaei, T., McGuigan, A., and Mahadevan, R. Ensemble modeling of cancer metabolism. *Frontiers in Physiology 3* (2012).

[40] Khodayari, A., Zomorrodi, A., Liao, J., and Maranas, C. A kinetic model of escherichia coli core metabolism satisfying multiple sets of mutant flux data. *Metabolic Engineering 25* (2014), 50–62.

[41] Kirschbaum, M. The sensitivity of c3 photosynthesis to increasing co2 concentration: a theoretical analysis of its dependence on temperature and background co2 concentration. *Plant, Cell & Environment 17*, 6 (1994), 747–754.

[42] Klamt, S., Saez-Rodriguez, J., and Gilles, E. Structural and functional analysis of cellular networks with cellnetanalyzer. *BMC Systems Biology 1*, 1 (2007), 1.

[43] Kuepfer, L., Matthias, P., Sauer, U., and Stelling, J. Ensemble modeling for analysis of cell signaling dynamics. *Nature Biotechnology 25*, 9 (2007), 1001.

[44] Lee, J., Goel, A., Ataai, M. M., and Domach, M. M. Flux adaptations of citrate synthase—deficient escherichia colia. *Annals of the New York Academy of Sciences 745* (1994), 35–50.

[45] Lee, J., Jang, Y.-S., Choi, S. J., Im, J. A., Song, H., Cho, J. H., Papoutsakis, E. T., Bennett, G. N., Lee, S. Y., et al. Metabolic engineering of clostridium acetobutylicum atcc 824 for isopropanol-butanol-ethanol fermentation. *Applied and Environmental Microbiology 78* (2012), 1416–1423.

[46] Lee, S., Phalakornkule, C., Domach, M. M., and Grossmann, I. E. Recursive milp model for finding all the alternate optima in lp models for metabolic networks. *Computers & Chemical Engineering 24* (2000), 711–716.

[47] Lee, Y., Rivera, J., and Liao, J. Ensemble modeling for robustness analysis in engineering non-native metabolic pathways. *Metabolic Engineering 25* (2014), 63–71.

[48] Mahadevan, R., Edwards, J. S., and III, F. J. D. Dynamic flux balance analysis of diauxic growth in escherichia coli. *Biophysical Journal 83* (2002), 1331 – 1340.

[49] Mahadevan, R., and Schilling, C. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering 5* (2003), 264 – 276.

[50] Mann, M. S., and Lütke-Eversloh, T. Thiolase engineering for enhanced butanol production in clostridium acetobutylicum. *Biotechnology and Bioengineering 110* (2013), 887–897.

[51] Martyushev, L. M., and Seleznev, V. D. Maximum entropy production principle in physics, chemistry and biology. *Physics Reports 426* (2006), 1–45.

[52] Metris, A., and Baranyi, S. G. J. Modelling osmotic stress by flux balance analysis at the genomic scale. *International Journal of Food Microbiology 152* (2012), 123–128.

[53] Michaelis, L., and Menten, M. L. Die kinetik der invertinwirkung. *Biochem. Z 49* (1913), 352.

[54] Moles, C., Mendes, P., and Banga, J. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Research 13*, 11 (2003), 2467–2474.

[55] Mulukutla, B., Yongky, A., Daoutidis, P., and Hu, W.-S. Bistability in glycolysis pathway as a physiological switch in energy metabolism. *PloS One 9*, 6 (2014), e98756.

[56] Nikerel, I., van Winden, W., van Gulik, W., and Heijnen, J. A method for estimation of elasticities in metabolic networks using steady state and dynamic metabolomics data and linlog kinetics. *BMC Bioinformatics 7*, 1 (2006), 540.

[57] Novere, N. L., and Shimizu, T. Stochsim: modelling of stochastic biomolecular processes. *Bioinformatics 17*, 6 (2001), 575–576.

[58] Orth, J. D., Thiele, I., and Palsson, B. O. What is flux balance analysis? *Nature Biotechnology 28* (2010), 245–248.

[59] Papoutsakis, E. Equations and calculations for fermentations of butyric acid bacteria. *Biotechnology and Bioengineering 26*, 2 (1984), 174–187.

[60] Parks, S., Chiche, J., and Pouysségur, J. Disrupting proton dynamics and energy metabolism for cancer therapy. *Nature Reviews Cancer 13*, 9 (2013), 611–623.

[61] Peskov, K., Mogilevskaya, E., and Demin, O. Kinetic modelling of central carbon metabolism in escherichia coli. *The FEBS Journal 279*, 18 (2012), 3374–3385.

[62] Plata, G., Hsiao, T., Olszewski, K. L., Llinás, M., and Vitkup, D. Reconstruction and flux-balance analysis of the plasmodium falciparum metabolic network. *Molecular Systems Biology 6* (2010).

[63] Rodriguez-Fernandez, M., Mendes, P., and Banga, J. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *Biosystems 83*, 2 (2006), 248–265.

[64] Rohwer, J. Kinetic modelling of plant metabolic pathways. *Journal of Experimental Botany 63*, 6 (2012), 2275–2292.

[65] Salway, J. *Metabolism at a Glance*. John Wiley & Sons, 2013.

[66] Savinell, J., and Palsson, B. Optimal selection of metabolic fluxes for in vivo measurement. ii. application to escherichia coli and hybridoma cell metabolism. *Journal of Theoretical Biology 155*, 2 (1992), 215–242.

[67] Savinell, J., and Palsson, B. Optimal selection of metabolic fluxes for *in vivo* measurement. i. development of mathematical methods. *Journal of Theoretical Biology 155*, 2 (1992), 201–214.

[68] Schaber, J., Flöttmann, M., Li, J., Tiger, C.-F., Hohmann, S., and Klipp, E. Automated ensemble modeling with modelmage: analyzing feedback mechanisms in the sho1 branch of the hog pathway. *PloS One 6*, 3 (2011), e14791.

[69] Schellenberger, J., Que, R., Fleming, R., Thiele, I., Orth, J., Feist, A., Zielinski, D., Bordbar, A., Lewis, N., Rahmanian, S., et al. Quantitative prediction of cellular metabolism with constraint-based models: the cobra toolbox v2. 0. *Nature Protocols 6*, 9 (2011), 1290.

[70] Schilling, C. H., Edwards, J. S., Letscher, D., Palsson, B. O., et al. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnology and Bioengineering 71* (2000), 286–306.

[71] Schilling, C. H., Letscher, D., and Palsson, B. O. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology 203* (2000), 229–248.

[72] Schuetz, R., Kuepfer, L., and Sauer, U. Systematic evaluation of objective functions for predicting intracellular fluxes in *escherichia coli*. *Molecular Systems Biology 3* (2007).

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

[74] Schuster, S., and Hilgetag, C. On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems 2*, 02 (1994), 165–182.

[75] Segrè, D., Vitkup, D., and Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences 99* (2002), 15112–15117.

[76] Tan, Y., and Liao, J. C. Metabolic ensemble modeling for strain engineers. *Biotechnology Journal 7* (2012), 343–353.

[77] Tran, L., Rizk, M., and Liao, J. Ensemble modeling of metabolic networks. *Biophysical Journal 95*, 12 (2008), 5606–5617.

[78] Usuda, Y., Nishio, Y., Iwatani, S., Dien, S. J. V., Imaizumi, A., Shimbo, K., Kageyama, N., Iwahata, D., Miyano, H., and Matsui, K. Dynamic modeling of escherichia coli metabolic and regulatory systems for amino-acid production. *Journal of Biotechnology 147* (2010), 17–30.

[79] Varma, A., and Palsson, B. O. Metabolic flux balancing: Basic concepts, scientific and practical use. *Bio/technology 12* (1994).

[80] Vaseghi, S., Baumeister, A., Rizzi, M., and Reuss, M. In vivodynamics of the pentose phosphate pathway in saccharomyces cerevisiae. *Metabolic Engineering 1*, 2 (1999), 128–140.

[81] Visser, D., and Heijnen, J. Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. *Metabolic Engineering 5*, 3 (2003), 164–176.

[82] Visser, D., Schmid, J., Mauch, K., Reuss, M., and Heijnen, J. Optimal re-design of primary metabolism in escherichia coli using linlog kinetics. *Metabolic Engineering 6*, 4 (2004), 378–390.

[83] Visser, D., van der Heijden, R., Mauch, K., Reuss, M., and Heijnen, S. Tendency modeling: a new approach to obtain simplified kinetic models of metabolism applied to saccharomyces cerevisiae. *Metabolic Engineering 2*, 3 (2000), 252–275.

[84] Wiechert, W., Möllney, M., Petersen, S., and de Graaf, A. A. A universal framework for ^{13} c metabolic flux analysis. *Metabolic Engineering 3* (2001), 265–283.

[85] Wilhelm, T. Analysis of structures causing instabilities. *Phys. Rev. E 76* (2007), 011911.

[86] Wittmann, C., and Portais, J.-C. *Metabolic Flux Analysis*. Wiley-VCH Verlag GmbH and Co. KGaA, Weinheim, Germany, 2013, pp. 285–312.

[87] Zhang, M., Eddy, C., Deanda, K., Finkelstein, M., and Picataggio, S. Metabolic engineering of a pentose metabolism pathway in ethanologenic zymomonas mobilis. *Science 267* (1995), 240–243.

[88] Zhu, Y., Song, J., Xu, Z., Sun, J., Zhang, Y., Li, Y., and Ma, Y. Development of thermodynamic optimum searching (tos) to improve the prediction accuracy of flux balance analysis. *Biotechnology and Bioengineering 110* (2013), 914–923.

Glossary

ATP/ADP Adenosine triphosphate (ATP) is a chemical used in many metabolic reactions to provide additional energy. Upon consumption, this chemical is generally converted into adenosine diphosphate (ADP). The cycling of these two chemicals is important in the overall behavior of a metabolic network.

Autocatalytic An autocatalytic reaction is one in which a reaction is catalysed by one of its products.

Biomass Biomass is the mass of a living biological organism. In a metabolic network, reactions that create biomass are those induce growth in the organism.

C_3 and C_4 metabolic pathways The C_3 and C_4 metabolic pathways are used in plants to processing CO_2 into usable forms. The C_3 pathway uses CO_2 directly in a reaction with water and ribulose bisphophate (a sugar molecule) to produce phosphoglycerate, whilst the C_4 pathway first processes CO_2 into malate before using it in this reaction.

CoA Coenzyme A (CoA) as an additional enzyme used in the synthesis of fatty acid. It is sometimes bound to other chemicals, such as in the form of acetyl-CoA.

Diauxic growth Diauxic growth is growth in two phases, generally in which an organism switches from one food source to another.

Flux The flux of a metabolic reaction is a measure of the activity across it.

Gibbs free energy Gibbs free energy is a measure of how much reversible work can be performed by a thermodynamic system. The change in Gibbs free energy through a reaction is a measure of how much energy has been lost to the environment and is no longer contained in the system.

Glycolysis Gylcolysis is a common biological process in which glucose is transformed into pyruvate for use in the TCA cycle.

MILP Mixed integer linear programming (MILP) is a modification of linear programming in which some vectors must take integer values.

Nullspace The nullspace, also known as the kernel, of a matrix, M, is the set of vectors $\mathbf{v}$ that satisfy $M\mathbf{v} = \mathbf{0}$.

ODE Ordinary differential equations (ODEs) are equations that include include functions with a single independent variable, and its derivatives.

Stoichimetric matrix The stoichiometric matrix of a metabolic network is a matrix containing information on the usage of metabolites with the network.

TCA Cycle The tricarboxylic acid (TCA) cycle, also known as the citric acid cycle or the Krebs cycle, is a series of chemical reactions found in all aerobic organisms to release energy stored in carbohydrates into the form of ATP.

Wildtype A wildtype bacteria is one as found in nature, prior to any genetic mutation in a lab.